Phase space localization theorem for ondelettes

Guy Battle
Mathematics Department, Texas A&M University, College Station, Texas 77843

(Received 13 December 1988; accepted for publication 31 May 1989)

It is proven that an orthonormal basis of ondelettes can never have exponential localization in both position space and momentum space.

I. INTRODUCTION

Orthonormal bases of ondelettes have been the most effective localizations of phase space to date. This is to be expected, since they are related to renormalization group ideas.[1,2] For example, Meyer and his co-workers have constructed an orthonormal basis of ondelettes that are Schwartz functions with compactly supported Fourier transforms.[3] More recently, an orthonormal basis of compactly supported class C^N ondelettes[4] have been constructed for arbitrary N. Since compact support is impossible in both position space and momentum space, it is natural to ask whether ondelettes with exponential falloff in both position space and momentum space exist. The *answer is no*, and this no-go theorem is proven here. The key to our proof is based on the intimate connection between smoothness properties and moment properties, which (we show) must hold for ondelettes.

Although one is usually interested in a basis of ondelettes, it makes sense to speak of an individual ondelette.

Definition: A square-integrable function φ on $\mathbb{R}^d$ is an *ondelette* if and only if the functions $\varphi_{L,n}$ defined by

$$\varphi_{L,n} = L^{-d/2}\varphi(L^{-1}x+n) \tag{1}$$

are mutually orthogonal for $L = 2^{-r}$, $r\in\mathbb{Z}$, and $n\in\mathbb{Z}^d$.

One of the reasons for the current interest in various bases of ondelettes is the amount of phase space localization the ondelettes can have.[3,4] Such orthonormal bases circumvent the strong uncertainty principle of Balian and Low,[5–7] so our no-go theorem for ondelettes should not be confused with their no-go theorem. They consider the rather different game of tiling phase space with functions of the form $e^{i2\pi m\cdot x}f(x+n)$, and their uncertainty principle states that for such a function f the standard deviations $\Delta_f x$ and $\Delta_f p$ in position space and momentum space, respectively, cannot both be finite.

Remark 1: Actually, our definition of "ondelettes" is more restrictive than the popular one, which includes frames[4] and continuous decompositions as well as orthonormal bases.

Remark 2: Very recently, Bourgain[8] has tiled phase space with functions of the form $f_m(x+n)$ having the property that $\Delta_{f_m}x$ and $\Delta_{f_m}p$ are not only both finite but also bounded uniformly in $m\in\mathbb{Z}^d$. The general idea is to obtain better phase space localization by sacrificing the discrete translational symmetry of the basis in the momentum directions while simultaneously preserving uniformity in the standard deviations. On the other hand, this new orthonormal basis does not match the degree of phase space localization attainable by ondelette bases. Indeed, Bourgain's phase space localization is optimal because there is no basis $f_m(x+n)$, such that

$$\int dx|f_{mn}(x)|^2[1+|x-\langle x\rangle_{f_{mn}}|^2]^{1+\epsilon}\leqslant c, \tag{2}$$

$$\int dp|\hat f_{mn}(p)|^2[1+|p-\langle p\rangle_{f_{mn}}|^2]^{1+\epsilon}\leqslant c \tag{3}$$

for some $\varepsilon>0$. This negative result is due to Steger.[9]

Yet another difference between our theorem and the Balian–Low theorem is that completeness of the basis plays a vital role in any proof of their necessarily stronger conclusion. Completeness plays *no* role in the proof of our ondelette theorem.

II. VANISHING MOMENTS

Before we prove the main lemma, we prove a special case for which the intuition is clear.

Lemma 1: Let φ be an ondelette for which $\hat\varphi(p)$ is continuous and bounded and integrable. Then the zeroth-order moment of φ must vanish.

Proof: We assume $\hat\varphi(0)\neq 0$ and show how orthogonality leads to a contradiction. Since $\hat\varphi(p)$ is integrable, we know that $\varphi(x)$ is a bounded continuous function that vanishes at infinity. Pick a fixed scale $L_0 = 2^{-r_0}$ small enough to guarantee $\varphi(L_0n_0)\neq 0$ for some d-tuple n_0 of integers, and set $x_0 = L_0n_0$. Thus

$$\int e^{ix_0\cdot p}\hat\varphi(p)dp = \varphi(x_0)\neq 0. \tag{4}$$

Now consider a small scale $L = 2^{-r}$ which will be chosen as small as we need in the end. Obviously x_0 lies in the finer lattice of points $L\mathbb{Z}^d$: we simply define $n_L = 2^{r-r_0}n_0$ and note that

$$Ln_L = x_0. \tag{5}$$

It follows from the orthogonality of φ and $\varphi_{L,-n_L}$ that

$$\int e^{ix_0\cdot p}\hat\varphi(p)\ \overline{\hat\varphi_{L,0}(p)}\,dp = 0. \tag{6}$$

On the other hand, the integral in question is just $L^{d/2}$ times

$$\int e^{ix_0\cdot p}\hat\varphi(p)\ \overline{\hat\varphi(Lp)}\,dp, \tag{7}$$

and by dominated convergence and the continuity of $\hat\varphi$, (7) approaches $\hat\varphi(0)\varphi(x_0)$ as $L\to 0$. This yields the desired contradiction, because (7) would have to be nonzero for some nonzero value of L. □

 0022-2488/89/102195-02$02.50

Remark: It is worth mentioning that the zeroth-order moment vanishes for the more general ondelettes as well, but for an entirely different reason. For frames and continuous decompositions[4] the completeness property is used to verify the property

$$\int dp |p|^{-1} |\hat{\varphi}(p)|^2 < \infty, \tag{8}$$

which seems essential to such expansions. We repeat that completeness is *not* used in the orthonormal case.

We can easily extend the proof of Lemma 1 to prove the more general lemma on vanishing moments.

Lemma 2: Let φ be an ondelette for which $\hat{\varphi}(p)$ is class C^{N+1} and $(1+|p|)^{N+1}\hat{\varphi}(p)$ is integrable. Then all moments of φ of order $\leqslant N$ must vanish.

Proof: By Lemma 1 we know the zeroth-order moment vanishes. Suppose we have shown that all moments of order $\leqslant k-1$ vanish, where k is an integer $\leqslant N$, and assume there are some nonzero k th-order moments. Thus

$$\hat{\varphi}(p) = \sum_{|\alpha|=k} \frac{1}{\alpha!} \frac{\partial^{|\alpha|}}{\partial p^\alpha} \hat{\varphi}(p) \Big|_{p=0} p^\alpha + R_k(p), \tag{9}$$

where $R_k(p)$ is the Taylor remainder and we use the standard multi-index notation. The zero set of the polynomial is an algebraic hypersurface in $\mathbb{R}^d$, so the bounded continuous function

$$\sum_{|\alpha|=k} \frac{(-i)^\alpha}{\alpha!} \overline{(D^\alpha\hat{\varphi})(0)} \frac{\partial^{|\alpha|}}{\partial x^\alpha} \varphi(x) = \int e^{ix\cdot p} \sum_{|\alpha|=k} \frac{1}{\alpha!} \overline{(D^\alpha\hat{\varphi})(0)} p^\alpha \hat{\varphi}(p) dp \tag{10}$$

cannot vanish identically because $\hat{\varphi}(p)$ is also continuous. We now pick a dyadic point x_0 at which (10) is nonzero in exactly the same way that we picked a dyadic point in the proof of Lemma 1. There is a d-tuple n_0 of integers and a scale $L_0 = 2^{-r_0}$ for which $x_0 = L_0 n_0$, and for smaller $L = 2^{-r}$, n_L is given by (5). Thus we have Eq. (6) and so—given that (10) does not vanish at x_0—all we need to show is that we also have

$$\int e^{ix_0\cdot p} \hat{\varphi}(p) \overline{\hat{\varphi}(Lp)} dp \neq 0. \tag{11}$$

Applying (9) to $\hat{\varphi}(Lp)$ we obtain

$$\int e^{ix_0\cdot p} \hat{\varphi}(p) \overline{\hat{\varphi}(Lp)} dp = \int e^{ix_0\cdot p} \overline{R_k(Lp)} \hat{\varphi}(p) dp + L^k \int e^{ix_0\cdot p} \sum_{|\alpha|=k} \frac{1}{\alpha!} \overline{(D^\alpha\hat{\varphi})(0)} p^\alpha \hat{\varphi}(p) dp, \tag{12}$$

but by Taylor's theorem we have

$$|R_k(Lp)| \leqslant cL^{k+1} |p|^{k+1} \sum_{|\alpha|=k+1} \sup_{|q|<c\epsilon_0} |(D^\alpha\hat{\varphi})(q)| \tag{13}$$

for $|p| \leqslant \epsilon_0 L^{-1}$. It also follows from (9) and the boundedness of $\hat{\varphi}(Lp)$ that we have a bound for large p, namely,

$$|R_k(Lp)| \leqslant c(1 + L^k |p|^k). \tag{14}$$

Since $|p|^{k+1}\hat{\varphi}(p)$ is integrable, we have the estimates

$$\int_{|p|>\epsilon_0 L^{-1}} |\hat{\varphi}(p)| dp \leqslant c\epsilon_0^{-k-1} L^{k+1}, \tag{15}$$

$$\int_{|p|>\epsilon_0 L^{-1}} |p|^k \hat{\varphi}(p) | dp \leqslant c\epsilon_0^{-1} L \tag{16}$$

as well. Combining (13)–(16) we easily conclude that the first term in (12) is $O(L^{k+1})$. As the second term is $\alpha_0 L^k$ with $\alpha_0 \neq 0$, we need only choose L small enough to realize (11). □

Theorem: If φ is an ondelette, then it cannot have exponential localization in both position space and momentum space.

Proof: If $\varphi(x)$ and $\hat{\varphi}(p)$ both have exponential decay, then in particular φ is a Schwartz function. It follows from Lemma 2 that *all* moments of φ vanish. But this means that $\hat{\varphi}(p)$ vanishes at $p=0$ to infinite order, and so $\hat{\varphi}(p)$ cannot be real analytic. This contradicts the exponential decay assumed for $\varphi(x)$. □

Note added: The author has learned that Y. Meyer was already aware of the connection between smoothness and vanishing moments but that no consequences were ever publicized.

ACKNOWLEDGMENT

The author would like to thank the referee for a couple of background remarks that have been included in the revision of this paper.

[1]K. Gawedzki and A. Kupiainen, Commun. Math. Phys. 77, 31 (1980).
[2]G. Battle, Commun. Math. Phys. **110**, 601 (1987).
[3]Y. Meyer, Séminaire Bourbaki **38**, 662 (1986).
[4]I. Daubechies, AT&T Bell Laboratories preprint, 1987.
[5]R. Balian, C.R. Acad. Sci. Paris **292**, 1357 (1981).
[6]F. Low, in *A Passion for Physics*, G. F. Chew Volume (World Scientific, Singapore, 1985).
[7]G. Battle, Lett. Math. Phys. **15**, 175 (1988).
[8]J. Bourgain, J. Funct. Anal. **79**, 136 (1988).
[9]J. Steger, unpublished.

Factor structure of the Tomimatsu–Sato metrics

Zoltán Perjés
Central Research Institute for Physics, H-1525 Budapest 114, P. O. Box 49, Hungary

(Received 8 February 1989; accepted for publication 26 April 1989)

It is shown that the factoring property of the Tomimatsu–Sato metrics follows from the structure of special Hankel determinants. A set of linear algebraic equations determining the factors is found. The factors of the first five Tomimatsu–Sato metrics are tabulated.

I. INTRODUCTION

Motivated by a recent result[1] that the Tomimatsu–Sato (TS) metrics can be factored over the field of integers, I sought such a representation for higher TS metrics. Ernst[2] and Hoenselaers[3] have made the first observation that the $\delta = 1$, 2, and 3 TS metric functions can be factored in the form of the spectral decomposition $[g_{ik}] = [O_{im}][\Lambda_{mn}][O_{kn}]$ where the matrix $\Lambda = [\Lambda_{mn}]$ is diagonal and $O = [O_{ik}]$ is unimodular. I obtain here the factors of the $\delta = 4$ and $\delta = 5$ TS metrics. The following pattern now emerges: The TS metric functions (4) with an *odd* δ can be written

$$\begin{aligned} A &= \lambda_1\rho^2 + \lambda_2\sigma^2, \\ B &= A + \rho\pi + \lambda_2\sigma\tau, \\ 2\delta C &= \lambda_1\rho\tau - \sigma\pi, \\ D &= \lambda_1(2B - A) + \pi^2 + \lambda_1\lambda_2\tau^2, \end{aligned} \tag{1}$$

where the factors ρ,σ,π, and τ are polynomials in p^2, q^2, a, and b with integer coefficients. The *even* TS metrics have a similar form,

$$\begin{aligned} A &= \rho^2 + \lambda_1\lambda_2\sigma^2, \\ B &= A + \rho\pi + \lambda_2\sigma\tau, \\ 2\delta C &= \rho\tau - \lambda_1\sigma\pi, \\ D &= \lambda_1(2B - A) + \lambda_1\pi^2 + \lambda_2\tau^2. \end{aligned} \tag{2}$$

In these expressions,

$$\lambda_1 = p^2 a, \quad \lambda_2 = q^2 b. \tag{3}$$

In Sec. III, the factoring property of the TS metric functions will be related to the structure of certain Hankel determinants. The homogeneous parts of π and τ yield 2δ polynomials, to be called the *primitive factors*. The knowledge of the primitive factors $\pi(\delta,r)$ and $\tau(\delta,r)$, where $r = 1,2,\ldots,\delta$, alone enables one to generate the δth TS solution. In Sec. IV, the dual polynomials, defined by a pairwise interchange of variables, are considered. The dual of a primitive factor can be represented as a linear combination of either $\pi(\delta,r)$ or $\tau(\delta,r)$. One can use these linear algebraic relations to determine the values of the primitive factors. The factors of the first five TS metrics will be given in Table I, and their primitive factors are listed in Table II.

II. TS METRICS

The TS metrics can be written in the form[3]

$$ds^2 = \frac{B}{\delta^2 p^{2\delta-2}(a-b)^{\delta^2-1}}\left(\frac{dy^2}{b} - \frac{dx^2}{a}\right) + g_{ik}\,dx^i\,dx^k, \quad i,k = 3,4, \tag{4}$$

where

$$g_{33} = bD/\delta^2 B, \quad g_{34} = 2q(bC/B), \quad g_{44} = A/B, \tag{5}$$

$$a = x^2 - 1, \quad b = y^2 - 1, \quad p^2 + q^2 = 1, \tag{6}$$

$$AD - \lambda_1 B^2 - 4\delta^2\lambda_2 C^2 = 0, \tag{7}$$

and δ labels the solution of the vacuum gravitational equations. The first few solutions in the family are due to Kerr ($\delta = 1$) and Tomimatsu and Sato ($\delta = 2$, 3, and 4) while the metric functions for an arbitrary positive integer δ were given by Yamazaki,[4]

$$B = A + G + H, \quad C = (p/2qb\delta)(Q + R - (\delta/pq)A), \tag{8}$$

$$\begin{aligned} A &= F(\delta^2), \\ G &= 2\sum_{r=1}^{\delta} c(\delta,r)F(\delta^2 - r), \\ H &= 2px\sum_{r=1}^{\delta} d(r)a^{r-1}\sum_{r'=r}^{\delta} c(\delta,r')F(\delta^2 - r'), \\ Q &= -\frac{2px}{pq}\delta\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta} q^2 b^{r'} a^{1-r'} g(\delta,r,r') \times F(\delta^2 - r), \\ R &= \frac{\delta}{pq}\sum_{r=1}^{\delta}\sum_{r'=1}^{\delta}(p^2 a^{r'} b^{1-r'} - q^2 b^{r'} a^{1-r'}) \times h(\delta,r,r')F(\delta^2 - r), \end{aligned} \tag{9}$$

with the numerical coefficients

$$\begin{aligned} c(\delta,r) &= \delta\frac{(\delta + r - 1)!}{(\delta - r)!(2r)!}2^{2r-1}, \\ d(r) &= (-1)^{r-1}\frac{(2r-2)!}{[2^{r-1}(r-1)!]^2}, \\ g(\delta,r,r') &= \frac{re(r)c(\delta,r)}{\delta^2}\sum_{t=r'}^{\delta}\frac{td(t-r'+1)c(\delta,t)}{r+t-1}, \\ h(\delta,r,r') &= \frac{rr'e(r)c(\delta,r)c(\delta,r')}{\delta^2(r+r'-1)}. \end{aligned} \tag{10}$$

Einstein's vacuum field equations are a set of linear alge-

 0022-2488/89/102197-04$02.50

TABLE I. Factors of TS metrics.

$\delta = 1$ (Kerr metric)

$$\rho = 1,$$
$$\sigma = 1,$$
$$\pi = 2(px+1),$$
$$\tau = 0.$$

$\delta = 2$

$$\rho = p^2a^2 + q^2b^2,$$
$$\sigma = 2(a-b),$$
$$\pi = 4[a + (a+2)(px+1)],$$
$$\tau = -4b(px+1).$$

$\delta = 3$

$$\rho = p^2a^4 + q^2b^4 - 2q^2b^2(a-b)(b-3a),$$
$$\sigma = p^2a^4 + q^2b^4 + 2p^2a^2(a-b)(a-3b),$$
$$\pi = (6a^4 + 32a^3 + 32a^2)p^3x + p^2a^2(18a^2 + 48a + 32)$$
$$+ 6q^2b^4(px+1),$$
$$\tau = 4b\{px[-6a(a-b) - 4a + 12b + 8] - 9a^2$$
$$+ 12ab + 12b + 8\}.$$

$\delta = 4$

$$\rho = F(8) + 20\lambda_1\lambda_2 ab(a-b)^4,$$
$$\sigma = 4(a-b)[p^2a^4(a^2 - 4ab + 5b^2) + q^2b^4(b^2 - 4ab + 5a^2)],$$
$$\pi = (px+1)[p^2a^4(8a^3 + 80a^2 + 192a + 128)$$
$$+ q^2b^4(120a^3 - 192a^2b + 48a^2 + 80ab^2 - 128ab$$
$$- 64a + 160b^2 + 256b + 128)]$$
$$+ p^2a^5(24a^2 + 80a + 64) + q^2ab^4(40a^2 - 96ab$$
$$- 48a + 80b^2 + 128b + 64),$$
$$\tau = 4b\{(px+1)\{p^2[-20a^6 + 48a^5b - 30a^4b^2$$
$$+ (a-b)(160a^3b - 32a^4 + 96a^3 + 160a^2b$$
$$+ 128a^2)] - 2q^2b^6\}$$
$$+ p^2(a-b)(-20a^5 + 60a^4b + 32a^4$$
$$+ 80a^3b + 64a^3)\}.$$

$\delta = 5$

$$\rho = (p^2a^6 + q^2b^6)^2 + b^2(b-a)[2q^4b^6(2b^3 - 18ab^2 + 45a^2b - 35a^3)$$
$$+ p^2q^2(175a^4b^5 - 545a^5b^4 + 713a^6b^3 - 535a^7b^2$$
$$+ 230a^8b - 50a^9)],$$
$$\sigma = (p^2a^6 + q^2b^6)^2 - a^2(b-a)[2p^4a^6(2a^3 - 18ba^2 + 45b^2a - 35b^3)$$
$$+ p^2q^2(175b^4a^5 - 545b^5a^4 + 713b^6a^3 - 535b^7a^2$$
$$+ 230b^8a - 50b^9)],$$
$$\pi = 2px[p^4(5a^{12} + 80a^{11} + 336a^{10} + 512a^9 + 256a^8)$$
$$+ p^2q^2(525a^8b^4 - 1920a^7b^5 + 480a^7b^4 + 2800a^6b^6$$
$$- 2560a^6b^5 - 800a^6b^4 - 1920a^5b^7 + 7040a^5b^6 + 6656a^5b^5$$
$$+ 2560a^5b^4 + 525a^4b^8 - 7680a^4b^7 - 5760a^4b^6 + 3072a^4b^5$$
$$+ 3840a^4b^4 + 2800a^3b^8 - 2560a^3b^7 - 10240a^3b^6$$
$$+ 6144a^3b^5 + 2800a^2b^8 + 5120a^2b^7 + 2560a^2b^6) + 5q^4b^{12}]$$
$$+ 2[p^4(25a^{12} + 200a^{11} + 560a^{10} + 640a^9 + 256a^8)$$
$$+ p^2q^2(875a^8b^4 - 3600a^7b^5 + 6300a^6b^6 - 5120a^5b^7 + 5600a^5b^6$$
$$+ 8960a^5b^5 + 4480a^5b^4 + 1575a^4b^8 - 9600a^4b^7 - 11200a^4b^6$$
$$+ 3840a^4b^4 + 4200a^3b^8 - 8960a^3b^6 - 6144a^3b^5 + 2800a^2b^8$$

TABLE I. (*continued*)

$$+ 5120a^2b^7 + 2560a^2b^6) + 5q^4b^{12}],$$
$$\tau = 4b\{px[p^2(-50a^{10} + 210a^9b - 140a^9 - 300a^8b^2 + 1380a^8b + 600a^8$$
$$+ 140a^7b^3 - 2600a^7b^2 + 1088a^7b + 1920a^7 + 1400a^6b^3$$
$$- 4880a^6b^2 - 2944a^6b + 1280a^6 + 3360a^5b^3 - 960a^5b^2$$
$$- 3072a^5b + 2240a^4b^3 + 1920a^4b^2) + q^2(-140a^4b^6 + 300a^3b^7$$
$$- 40a^3b^6 - 210a^2b^8 + 120a^2b^7 + 48a^2b^6 + 50ab^9$$
$$- 140ab^8 - 160ab^7 - 64ab^6 + 100b^9 + 280b^8 + 320b^7 + 128b^6)]$$
$$+ p^2(-125a^{10} + 700a^9b - 1125a^8b^2 + 2100a^8b + 1400a^8$$
$$+ 560a^7b^3 - 4800a^7b^2 + 2560a^7 + 2800a^6b^3 - 5600a^6b^2$$
$$- 4480a^6b + 1280a^6 + 4480a^5b^3 - 3072a^5b + 2240a^4b^3$$
$$+ 1920a^4b^2)$$
$$+ q^2(-175a^4b^6 + 400a^3b^7 - 315a^2b^8 + 100ab^9 + 100b^9$$
$$+ 280b^8 + 320b^7 + 128b^6)\}.$$

TABLE II. The primitive factors.

$\delta = 1$

$$\pi(1,1) = 1, \quad \tau(1,1) = 0.$$

$\delta = 2$

$$\pi(2,1) = a, \quad \tau(2,1) = -\tfrac{1}{2}b,$$
$$\pi(2,2) = 1, \quad \tau(2,2) = 0.$$

$\delta = 3$

$$\pi(3,1) = p^2a^4 + \tfrac{1}{3}q^2b^4, \quad \tau(3,1) = ab(-2a + \tfrac{8}{3}b),$$
$$\pi(3,2) = p^2a^3, \quad \tau(3,2) = b^2,$$
$$\pi(3,3) = p^2a^2, \quad \tau(3,3) = b.$$

$\delta = 4$

$$\pi(4,1) = p^2a^7 + q^2(5a^3b^4 - 9a^2b^5 + 5ab^6),$$
$$\pi(4,2) = p^2a^6 + q^2b^6,$$
$$\pi(4,3) = p^2a^5 + q^2b^5,$$
$$\pi(4,4) = p^2a^4 + q^2b^4,$$
$$\tau(4,1) = p^2(a-b)(-5a^5b + 11a^4b^2) - \tfrac{1}{4}b^3(p^2a^4 + q^2b^4),$$
$$\tau(4,2) = 6p^2(a-b)a^3b^2,$$
$$\tau(4,3) = \tfrac{5}{2}p^2(a-b)(a^3b + a^2b^2),$$
$$\tau(4,4) = 4p^2(a-b)a^2b.$$

$\delta = 5$

$$\pi(5,1) = p^4a^{12} + \tfrac{1}{5}q^4b^{12} + p^2q^2(35a^8b^4 - 144a^7b^5$$
$$+ 252a^6b^6 - \tfrac{1024}{5}a^5b^7 + 63a^4b^8),$$
$$\pi(5,2) = p^4a^{11} + p^2q^2(28a^5b^6 - 48a^4b^7 + 21a^3b^8),$$
$$\pi(5,3) = p^4a^{10} + p^2q^2(16a^5b^5 - 20a^4b^6 + 5a^2b^8),$$
$$\pi(5,4) = p^4a^9 + p^2q^2(7a^5b^4 - 14a^3b^6 + 8a^2b^7),$$
$$\pi(5,5) = p^4a^8 + p^2q^2(15a^4b^4 - 24a^3b^5 + 10a^2b^6),$$
$$\tau(5,1) = 2b\{p^2(-5a^{10} + 28a^9b - 45a^8b^2 + \tfrac{112}{5}a^7b^3)$$
$$+ q^2(-7a^4b^6 + 16a^3b^7 - \tfrac{63}{5}a^2b^8 + 4ab^9)\},$$
$$\tau(5,2) = 2b\{p^2(\tfrac{21}{2}a^8b - 24a^7b^2 + 14a^6b^3) + \tfrac{1}{2}q^2b^9\},$$
$$\tau(5,3) = 2b\{p^2(\tfrac{5}{2}a^8 - 10a^6b^2 + 8a^5b^3) + \tfrac{1}{2}q^2b^8\},$$
$$\tau(5,4) = 2b\{p^2(4a^7 - 7a^6b + \tfrac{7}{2}a^4b^3) + \tfrac{1}{2}q^2b^7\},$$
$$\tau(5,5) = 2b\{p^2(5a^6 - 12a^5b + \tfrac{15}{2}a^4b^2) + \tfrac{1}{2}q^2b^6\}.$$

braic relations for the unknown functions F,

$$\sum_{r=1}^{\delta} h(\delta,r,r')f(r+r'-1)F(\delta^2-r) = \delta^1_{r'}F(\delta^2), \quad r'=1,2,\ldots,\delta, \tag{11}$$

where

$$f(r) = p^2a^r + q^2b^r. \tag{12}$$

Yamazaki's solutions

$$F(\delta^2) = M_\delta/N_\delta \tag{13}$$

contain the Hankel determinant

$$M_\delta = \begin{vmatrix} f(1) & f(2)/2 & f(3)/3 & \cdots & f(\delta)/\delta \\ f(2)/2 & f(3)/3 & \cdots & & \vdots \\ f(3)/3 & \cdots & & & \\ \vdots & \vdots & & & \\ f(\delta)/\delta & \cdots & & & f(2\delta-1)/(2\delta-1) \end{vmatrix} \tag{14}$$

and the normalizing factor is

$$N_\delta = \begin{vmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & 1/\delta \\ \frac{1}{2} & \frac{1}{3} & \cdots & & \vdots \\ \frac{1}{3} & \cdots & & & \\ \vdots & \vdots & & & \\ 1/\delta & \cdots & & & 1/(2\delta-1) \end{vmatrix}.$$

The function

$$re(r)c(\delta,r)F(\delta^2-r) = (-1)^{r-1}\frac{\det(f(s+t-1)/(s+t-1))}{\det(1/(s+t-1))}, \tag{15}$$

where $r=1,2,\ldots,\delta$, $s=1,2\ldots,r-1,r+1,\ldots,\delta$ and $t=2,3,\ldots,\delta$, is the cofactor of $F(\delta^2)$ belonging to the element in the rth row and first column.

III. FACTOR STRUCTURE

It follows from Eqs. (1), (2), and (8) that

$$G+H=\rho\pi+\lambda_2\sigma\tau, \tag{16}$$

where the factors ρ and σ are homogeneous polynomials. Comparing (16) with Eqs. (9) we see that the function H contributes the terms linear in px while G is the sum of terms without a factor px. Hence we can further decompose (16):

$$\pi=\pi_0+px\pi_1, \quad \tau=\tau_0+px\tau_1, \tag{17}$$

where π_i and τ_i are polynomials in p^2, q^2, a, and b with integer coefficients. Both the function G and H is a linear combination of the homogeneous polynomials $F(\delta^2-r)$. Thus considering terms in G and H of like homogeneity degree, each of the polynomials $F(\delta^2-r)$ must factorize in the form

$$F(\delta^2-r)=\rho\pi(\delta,r)+\lambda_2\sigma\tau(\delta,r). \tag{18}$$

We have

$$\pi_0=2\sum_{r=1}^{\delta}c(\delta,r)\pi(\delta,r), \tag{19}$$

$$\tau_0=2\sum_{r=1}^{\delta}c(\delta,r)\tau(\delta,r). \tag{20}$$

The primitive factors $\pi(\delta,r)$ and $\tau(\delta,r)$ are polynomials in p^2, q^2, a, and b with rational coefficients. They completely determine the TS functions. The factors satisfy the relations

$$\sigma=\frac{1}{\delta}\sum_{r=1}^{\delta}rc(\delta,r)d(r)\pi(\delta,r)b^{r-1} \tag{21}$$

and

$$\rho=\frac{1}{\delta}\sum_{r=1}^{\delta}rc(\delta,r)d(r)[a^{r-1}\pi(\delta,r)-\lambda_2b^{r-1}\tau(\delta,r)], \quad \text{for } \delta \text{ odd},$$

$$\sum_{r=1}^{\delta}rc(\delta,r)d(r)\tau(\delta,r)a^r=0,$$

$$\sum_{r=1}^{\delta}rc(\delta,r)d(r)\pi(\delta,r)a^r=0, \quad \text{for } \delta \text{ even}. \tag{22}$$

The primitive factors of the first five TS metrics are listed in Table II.

IV. DUALS

Let $P\in\mathscr{P}$ be an arbitrary polynomial in a, b, px, and qy. We introduce the involutory automorphism of the ring $\mathscr{P}$ of these polynomials

$$\mathscr{C}: \mathscr{P}\to\mathscr{P} \tag{23}$$

defined by the mutual substitutions

$$px \Leftrightarrow -qy, \quad p^2\Leftrightarrow q^2, \quad a\Leftrightarrow b.$$

The image (or *dual*) under $\mathscr{C}$ of a polynomial P will be denoted

$$P^* = \mathscr{C}P.$$

The Hankel determinants M_δ are invariant under the involution $\mathscr{C}$:

$$M_\delta^* = M_\delta.$$

According to Eqs. (9), we have

$$A^*=A, \quad G^*=G, \quad R^*=-R. \tag{24}$$

In order to close the set of Yamazaki functions under the action of $\mathscr{C}$, we introduce the duals

$$I=H^*, \quad P=Q^*. \tag{25}$$

These potentials satisfy the algebraic constraint[4]

$$H^2+I^2-G^2=2AG. \tag{26}$$

The Ernst potential of a TS space-time, given by $\xi=(H+iI)/G$, is *self-dual*,

$$\xi = i\bar{\xi}^*. \tag{27}$$

The discrete symmetry properties of the factors can be obtained by using the decomposition (1), (2), and (18) of the polynomials. We get

$$\sigma^* = \rho, \quad \rho^* = \sigma, \quad \text{for } \delta \text{ odd} \tag{28}$$

and

$$\rho^* = \rho, \quad \sigma^* = -\sigma, \quad \text{for } \delta \text{ even}. \tag{29}$$

Furthermore,

$$2\delta\rho = \pi_0^* - \lambda_2\tau_0, \quad \text{for } \delta \text{ odd} \tag{30}$$

and

$$\begin{aligned} 2\delta\rho &= -\lambda_1\tau_0^* - \lambda_2\tau_0, \\ 2\delta\sigma &= \pi_0 - \pi_0^*, \end{aligned} \quad \text{for } \delta \text{ even}. \tag{31}$$

Some further properties of the factors are the following.

(i) The polynomial τ contains an overall factor $4b$.

(ii) The factors ρ and σ of an *odd* TS space-time have the form

$$\rho = f(\delta+1)^{(\delta-1)/2} + 2q^2(a-b)Z,$$
$$\sigma = f(\delta+1)^{(\delta-1)/2} - 2p^2(a-b)Z^*,$$

where the polynomial Z is of homogeneity degree $(\delta^2-3)/2$.

(iii) The factors ρ and σ of an *even* TS space-time have the structure

$$\rho = f(\delta)^{\delta/2} + (a-b)V,$$
$$\sigma = (a-b)W,$$

where the polynomial V is of degree $\delta^2/2 - 1$ and W is of degree $\delta^2/2 - 2$.

Property (ii) has been verified to hold for $\delta = 1, 3, 5, 7$ and property (iii) holds for $\delta = 2, 4$, and 6.

The 2δ primitive factors $\pi(\delta,r)$ and $\tau(\delta,r)$ and their duals satisfy the set of 2δ linear homogeneous algebraic relations

$$\begin{aligned} &\frac{1}{\delta}b^{r'-1}c(\delta,r')\pi^*(\delta,r') \\ &\quad = \sum_{r=1}^{\delta} a^{r-1}h(\delta,r,r')\pi(\delta,r), \\ &\frac{1}{\delta}b^{r'-1}c(\delta,r')\tau^*(\delta,r') \\ &\quad = -\sum_{r=1}^{\delta} a^{r-1}h(\delta,r,r')\tau(\delta,r), \\ &\quad \text{for } \delta \text{ odd}, \end{aligned} \tag{32}$$

and

$$\begin{aligned} &\frac{1}{\delta}b^{r'-1}c(\delta,r')\tau^*(\delta,r') \\ &\quad = -\sum_{r=1}^{\delta} a^{r-1}h(\delta,r,r')\pi(\delta,r), \\ &\frac{1}{\delta}b^{r'-1}c(\delta,r')\pi^*(\delta,r') \\ &\quad = -\sum_{r=1}^{\delta} a^{r-1}h(\delta,r,r')\tau(\delta,r), \\ &\quad \text{for } \delta \text{ even}, \end{aligned} \tag{33}$$

where $r' = 1,2,\ldots,\delta$. These equations have been found by intuition. Substituting back in Eqs. (11) and using the symmetry properties (30) and (31) of the factors, we find that the field equations are satisfied.

The ratios of the primitive factors can be computed from the linear algebraic equations (32), (33) and the corresponding dual equations. The arbitrary factor of proportionality does not enter the gravitational field quantities. The functions ρ and σ are given by

$$\rho = \delta[\pi^*(\delta,1) - \lambda_2\tau(\delta,1)], \quad \text{for } \delta \text{ odd}, \tag{34}$$

and

$$\begin{aligned} \rho &= -\delta[\lambda_1\tau^*(\delta,1) - \lambda_2\tau(\delta,1)], \\ \sigma &= \delta[\pi(\delta,1) - \pi^*(\delta,1)], \quad \text{for } \delta \text{ even}. \end{aligned} \tag{35}$$

An obvious advantage of the field equations (32) and (33) over (11) is that they are linear algebraic relations with *constant* coefficients for the functions $a^{r-1}\pi(\delta,r)$ and $a^{r-1}\tau(\delta,r)$. Since the matrix $[m_{ik}] = [\delta(h(\delta,i,k)/c(\delta,k))]$ is involutory, $\mathbf{m}^2 = \mathbf{1}$, and both systems (34) and (35) are symmetrically partitioned, their determinant vanishes. One can form linear superpositions of the solutions. The coefficients of superposition are subject to the duality relations

$$\pi^*(\delta,t) = \lambda_2\tau(\delta,t), \quad \text{for } \delta \text{ odd}, \tag{36}$$

and

$$\begin{aligned} &\lambda_1\tau^*(\delta,t) + \lambda_2\tau(\delta,t) = 0, \\ &\pi^*(\delta,t) = \pi(\delta,t), \quad \text{for } \delta \text{ even}, \end{aligned} \tag{37}$$

with $t = 2,3,\ldots,\delta$. These symmetry relations among the primitive factors follow from the power structure of Eqs. (28)–(31).

ACKNOWLEDGMENT

I am grateful to Professor Roy Kerr for having kindled my interest in TS structures, and for amendments.

[1]R. P. Kerr, in *Proceedings of the Mathematical Relativity Miniconference*, edited by R. Bartnik, Australian National University, 1989; R. P. Kerr and W. B. Wilson, in *Proceedings of the 5th Marcel Grossmann Meeting*, edited R. Ruffini (North-Holland, Amsterdam, 1989).

[2]F. Ernst, J. Math. Phys. **17**, 1092 (1976); C. Hoenselaers and F. J. Ernst, *ibid.* **24**, 1817 (1983).

[3]A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973).

[4]M. Yamazaki, in *Proceedings of the Second Marcel Grossmann Meeting*, edited by R. Ruffini (North-Holland, Amsterdam, 1982), p. 371. The functions *G, H, I, P, Q*, and *R* are Yamazaki's *2G, 2H, 2I, PA, QA*, and *RA*, respectively.

New similarity reductions of the Boussinesq equation

Peter A. Clarkson
Department of Mathematics, University of Exeter, Exeter, EX4 4QE, England
Martin D. Kruskal
Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08540 and Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903

(Received 25 April 1989; accepted for publication 17 May 1989)

Some new similarity reductions of the Boussinesq equation, which arises in several physical applications including shallow water waves and also is of considerable mathematical interest because it is a soliton equation solvable by inverse scattering, are presented. These new similarity reductions, including some new reductions to the first, second, and fourth Painlevé equations, cannot be obtained using the standard Lie group method for finding group-invariant solutions of partial differential equations; they are determined using a new and direct method that involves no group theoretical techniques.

I. INTRODUCTION

The Boussinesq equation

$$u_{tt} + au_{xx} + b(u^2)_{xx} + cu_{xxxx} = 0, \tag{1.1}$$

where a, b, and c are constants and subscripts denote differentiation, was introduced to Boussinesq in 1871 to describe the propagation of long waves in shallow water[1] (see, also, Ref. 2). The Boussinesq equation also arises in several other physical applications including one-dimensional nonlinear lattice waves,[3,4] vibrations in a nonlinear string,[5] and ion sound waves in a plasma.[6]

It is well known (and was even to Boussinesq) that the Boussinesq equation (1.1) has a bidirectional solitary wave solution

$$u(x,t) = -\frac{3(\gamma^2 + a)}{2b} \times \operatorname{sech}^2\left\{\frac{1}{2}\left(\frac{\gamma^2+a}{-c}\right)^{1/2}(x \pm \gamma t) + x_0\right\},$$

where γ and x_0 are constants.

Recently there has been considerable mathematical interest in the Boussinesq equation, primarily because its Cauchy problem (for initial data on the infinite line that decays sufficiently rapidly) is solvable by inverse scattering,[5] through a third-order scattering problem (see, also, Ref. 7).

The inverse scattering method was originally developed by Gardner *et al.*[8] in order to solve the Cauchy problem for the Korteweg–de Vries (KdV) equation. In effect, this method reduces the solution of the nonlinear partial differential equation to that of a linear integral equation, and the partial differential equation is usually then said to be *completely integrable*. Completely integrable partial differential equations generally possess almost all of the following remarkable properties: the existence of multisoliton solutions; an infinite number of independent conservation laws and symmetries, and recursion operators generating them; a bi-Hamiltonian representation; a prolongation structure; a Lax pair; Bäcklund transformations; the Hirota bilinear representation; the Painlevé property, etc. (cf. Ref. 9). However, the precise relationship between these properties has yet to be rigorously established.

In this paper we study similarity reductions of the Boussinesq equation. Without loss of generality we shall assume that $a = 0$, $b = \frac{1}{2}$, and $c = \pm 1$ in Eq. (1.1) since the equation

$$u_{tt} + \tfrac{1}{2}(u^2)_{xx} \pm u_{xxxx} = 0 \tag{1.2}$$

is equivalent to Eq. (1.1) after suitable rescaling and translation of the variables. If the quantities in the equation are to be interpreted as real, then the sign matters and we choose the plus sign from here on only for convenience, and leave the reader the trivial modifications required for the other sign. However, if the quantities are interpreted as complex, then the sign does not matter and our analysis is complete.

The classical method for finding similarity reductions of a given partial differential equation is to use the Lie group method of infinitesimal transformations (sometimes called the method of group-invariant solutions), originally developed by Lie[10] (see Refs. 11–14 for recent descriptions of this method). Though the method is entirely algorithmic, it often involves a large amount of tedious algebra and auxiliary calculations which are virtually unmanageable manually. Recently symbolic manipulation programs have been developed, especially in MACSYMA[15] and REDUCE,[16] in order to facilitate the determination of the associated similarity reductions. (See Ref. 17 for a review of the use of computer algebra to find symmetries of differential equations.)

Bluman and Cole[18] proposed a generalization of Lie's method which they called the "*nonclassical method of group-invariant solutions*," which itself has been generalized by Olver and Rosenau.[19] All these methods determine Lie point transformations of a given partial differential equation, i.e., transformations depending only on the independent and dependent variables.

Noether[20] recognized that Lie's method could be generalized by allowing the transformations to depend upon the derivatives of the dependent variable as well as the independent and dependent variables. The associated symmetries, called *Lie–Bäcklund symmetries*, can also be determined by an algorithmic method (see Refs. 13 and 21).

In a recent paper, Bluman *et al.*[22] introduce an algorithmic method which yields new classes of symmetries of a given partial differential equation that are neither Lie point nor Lie–Bäcklund symmetries.

 0022-2488/89/102201-13$02.50

A common characteristic of all these methods for finding symmetries and associated similarity reductions of a given partial differential equation is the use of group theory.

In this paper we develop a new method of deriving similarity reductions of partial differential equations and apply it to the Boussinesq equation (1.2). The unusual characteristic of this new method in comparison to the ones mentioned above is that it does not use group theory (though we hope that a group theoretic explanation of the method will be possible in due course[23]). The basic idea is to seek a reduction of a given partial differential equation in the form

$$u(x,t) = U(x,t,w(z(x,t))), \tag{1.3}$$

which is the most general form for a similarity reduction (cf. Bluman and Cole[11]). Substituting this into the partial differential equation and demanding that the result be an ordinary differential equation for $w(z)$ imposes conditions upon U and its derivatives that enable one to solve for U. For the Boussinesq equation (1.2), it turns out to be sufficient to take (1.3) in the special form

$$u(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)). \tag{1.4}$$

The outline of this paper is as follows: in Sec. II we describe the previously known (classical and nonclassical) similarity reductions of the Boussinesq equation; in Sec. III we present our new method for finding similarity reductions of a given partial differential equation and use it to obtain new similarity reductions of the Boussinesq equation (1.2); in Sec. IV we justify the use of the special form (1.4); and in Sec. V we discuss our results.

II. CLASSICAL AND NONCLASSICAL SIMILARITY REDUCTIONS

First we sketch the derivation of the classical similarity reductions of the Boussinesq equation using Lie group method as given by Bluman and Cole.[11] Consider the one-parameter (ε) Lie group of infinitesimal transformations in (x,t,u) given by

$$\xi = x + \varepsilon X(x,t,u) + O(\varepsilon^2), \tag{2.1a}$$

$$\tau = t + \varepsilon T(x,t,u) + O(\varepsilon^2), \tag{2.1b}$$

$$\eta = u + \varepsilon U(x,t,u) + O(\varepsilon^2), \tag{2.1c}$$

$$\eta_\xi = u_x + \varepsilon U^x + O(\varepsilon^2), \tag{2.2a}$$

$$\eta_{\xi\xi} = u_{xx} + \varepsilon U^{xx} + O(\varepsilon^2), \tag{2.2b}$$

$$\eta_{\xi\xi\xi\xi} = u_{xxxx} + \varepsilon U^{xxxx} + O(\varepsilon^2), \tag{2.2c}$$

$$\eta_{\tau\tau} = u_{tt} + \varepsilon U^{tt} + O(\varepsilon^2), \tag{2.2d}$$

where the functions U^x, U^{xx}, U^{xxxx}, and U^{tt} in (2.2) are determined from Eqs. (2.1) (cf. Ref. 11). The Boussinesq equation (1.2) is invariant under this transformation if

$$\eta_{\tau\tau} + \tfrac{1}{2}(\eta^2)_{\xi\xi} + \eta_{\xi\xi\xi\xi} = 0. \tag{2.3}$$

By (2.1) and (2.2), to first order in ε, this becomes

$$U^{tt} + uU^{xx} + u_{xx}U + 2u_xU^x + U^{xxxx} = 0. \tag{2.4}$$

Conditions on the infinitesimals $X(x,t,u)$, $T(x,t,u)$, and $U(x,t,u)$ are determined by equating coefficients of like derivatives of monomials in u_x and u_t and higher derivatives. Solving these "determining equations" yields the following:

$$X = \alpha x + \beta, \quad T = 2\alpha t + \gamma, \quad U = -2\alpha u, \tag{2.5}$$

where α, β, and γ are arbitrary constants (cf. Refs. 24 and 25). Similarity reductions are then obtained by solving the characteristic equations

$$\frac{dx}{X(x,t,u)} = \frac{dt}{T(x,t,u)} = \frac{du}{U(x,t,u)}.$$

Integration of these ordinary differential equations yields the following cases.

Case (a), $\alpha=0$: This is the traveling wave reduction $u(x,t) = f(z)$, $z = \gamma x - \beta t$, where $f(z)$ satisfies

$$\beta^2 f + \frac{1}{2}\gamma^2 f^2 + \gamma^4 \frac{d^2 f}{dz^2} = Az + B, \tag{2.6}$$

with A and B arbitrary constants of integration. For $\gamma = 0$, this is a form of the first Painlevé equation (cf. Ince[26])

$$\frac{d^2 w}{dz^2} = 6w^2 + z \tag{2.7}$$

(or the Weierstrass elliptic function equation for $A = 0$). This reduction of the Boussinesq equation to the first Painlevé equation is well known in connection with the Painlevé conjecture (cf. Refs. 27 and 28) for soliton equations.

Case (b), $\alpha\neq 0$: This is the scaling reduction

$$u(x,t) = \frac{g(z)}{[t + \gamma/(2\alpha)]}, \quad z = \frac{(x + \beta/\alpha)}{[t + \gamma/(2\alpha)]^{1/2}}, \tag{2.8}$$

where $g(z)$ satisfies

$$\frac{z^2}{4}\frac{d^2 g}{dz^2} + \frac{7z}{4}\frac{dg}{dz} + 2g + g\frac{d^2 g}{dz^2} + \left(\frac{dg}{dz}\right)^2 + \frac{d^4 g}{dz^4} = 0. \tag{2.9}$$

This can be solved in terms of solutions of the fourth Painlevé equation

$$\frac{d^2 w}{dz^2} = \frac{1}{2w}\left(\frac{dw}{dz}\right)^2 + \frac{3}{2}w^3 + 4zw^2 + 2(z^2 - a)w + \frac{b}{w}, \tag{2.10}$$

where a and b are arbitrary constants[29] (see also Appendix A).

However, there also exist similarity reductions of the Boussinesq equation that *cannot* be obtained by the classical Lie group method. As noted by several authors,[19,24,25,29] the Boussinesq equation (1.2) possesses the similarity reduction

$$u(x,t) = f(z) - 4\lambda^2 t^2, \quad z = x + \lambda t^2, \tag{2.11}$$

where λ is a constant and $f(z)$ satisfies

$$\frac{d^3 f}{dz^3} + f\frac{df}{dz} + 2\lambda f = 8\lambda^2 z + A, \tag{2.12}$$

with A a constant of integration. If, in (2.12), we make the transformation

$$f(z) = \eta(\xi) + 2\lambda\xi, \quad z = \xi - A/(8\lambda^2),$$

then $\eta(\xi)$ satisfies

$$\frac{d^3\eta}{d\xi^3} + \eta\frac{d\eta}{d\xi} + 2\lambda\left(\xi\frac{d\eta}{d\xi} + 2\eta\right) = 0. \tag{2.13}$$

Solutions of Eq. (2.13) are known to be related through a one-to-one transformation to solutions of the second Painlevé equation

$$\frac{d^2w}{dz^2} = 2w^3 + zw + a, \tag{2.14}$$

where a is an arbitrary constant[30]—see, also, Appendix A. [We remark that this equation also arises from the scaling reduction

$$u(x,t) = (-3\lambda t)^{-2/3}\eta(\xi), \quad \xi = x/(-3\lambda t)^{-1/3}$$

of the KdV equation

$$u_t + uu_x + u_{xxx} = 0$$

—see Ref. 27.]

The infinitesimals that give rise to the similarity reduction (2.11) of the Boussinesq equation are

$$X(x,t,u) = 2\lambda t, \quad T(x,t,u) = -1, \quad U(x,t,u) = 8\lambda^2 t, \tag{2.15}$$

which are clearly not a special case of (2.5). Since Eqs. (2.15) describe a Lie point transformation of the Boussinesq equation, Rosenau and Schwarzmeier[25] suggest it can be obtained using the nonclassical method of Bluman and Cole[18] (see, also, Ref. 11). This method involves more algebra and calculations than the classical Lie method; in fact, Olver and Rosenau[19] suggest that for some partial differential equations, the determining equations for these nonclassical symmetries might be too difficult to solve explicitly. The principal reason for this is that although the determining equations for the infinitesimals X, T, and U in the classical method are a linear system of equations (in X, T, and U), in the nonclassical method, they are a nonlinear system. Furthermore, for some equations, such as the linear heat equation, it is well known that the nonclassical method does not appear to yield any more similarity reductions than the classical Lie method does[18] (see, also, Ref. 31).

III. NEW SIMILARITY REDUCTIONS

In this section we seek reductions of the Boussinesq equation (1.2) in the form

$$u(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)), \tag{3.1}$$

where $\alpha(x,t)$, $\beta(x,t)$, and $z(x,t)$ are to be determined. [We shall show in Sec. IV why it is sufficient to seek a similarity reduction of the Boussinesq equation (1.2) in the form (3.1) rather than the more general form (1.3).]

Substituting (3.1) into (1.2) and collecting coefficients of monomials of w and its derivatives yields

$$\begin{aligned}\beta z_x^4 w'''' &+ [6\beta z_x^2 z_{xx} + 4\beta_x z_x^3]w''' + [\beta(3z_{xx}^2 + 4z_x z_{xxx}) + 12\beta_x z_x z_{xx} + 6\beta_{xx} z_x^2 + \alpha\beta z_x^2 + \beta z_t^2]w'' \\ &+ [\beta z_{xxxx} + 4\beta_x z_{xxx} + 6\beta_{xx} z_{xx} + 4\beta_{xxx} z_x + 2\alpha_x \beta z_x + 2\alpha\beta_x z_x + \alpha\beta z_{xx} + 2\beta_t z_t + \beta z_{tt}]w' \\ &+ [\beta_{xxxx} + 2\alpha_x \beta_x + \alpha\beta_{xx} + \alpha_{xx}\beta + \beta_{tt}]w + \beta^2 z_x^2 ww'' + \beta[4\beta_x z_x + \beta z_{xx}]ww' \\ &+ \beta^2 z_x^2 (w')^2 + [\beta_x^2 + \beta\beta_{xx}]w^2 + [\alpha_{tt} + \alpha\alpha_{xx} + \alpha_x^2 + \alpha_{xxxx}] = 0,\end{aligned} \tag{3.2}$$

where $' := d/dz$. In order that this equation be an ordinary differential equation for $w(z)$ the ratios of coefficients of different derivatives and powers of $w(z)$ have to be functions of z only. This gives a set of conditions for $\alpha(x,t)$, $\beta(x,t)$, and $z(x,t)$ for which any solution will yield a similarity reduction.

Remark 1: We use the coefficient of w'''' (i.e., βz_x^4) as the normalizing coefficient and therefore require that the other coefficients be of the form $\beta z_x^4 \Gamma(z)$, where Γ is a function of z to be determined.

Remark 2: We reserve uppercase greek letters for undetermined functions of z so that after performing operations (differentiation, integration, exponentiation, rescaling, etc.) the result can be denoted by the same letter [e.g., the derivative of $\Gamma(z)$ will be called $\Gamma(z)$].

Remark 3: There are three freedoms in the determination of α, β, z and w we can exploit, without loss of generality, that are valuable in keeping the method manageable: (i) if $\alpha(x,t)$ has the form $\alpha = \alpha_0(x,t) + \beta(x,t)\Omega(z)$, then we can take $\Omega \equiv 0$ [by substituting $w(z) \to w(z) - \Omega(z)$]; (ii) if $\beta(x,t)$ has the form $\beta = \beta_0(x,t)\Omega(z)$, then we can take $\Omega \equiv 1$ [by substituting $w(z) \to w(z)/\Omega(z)$]; and (iii) if $z(x,t)$ is determined by an equation of the form $\Omega(z) = z_0(x,t)$, where $\Omega(z)$ is any invertible function, then we can take $\Omega(z) = z$ [by substituting $z \to \Omega^{-1}(z)$].

We shall now proceed to determine the general similarity reductions of the Boussinesq equation using this method.

The coefficients of ww'' and $(w')^2$ yield the common constraint

$$\beta z_x^4 \Gamma(z) = \beta^2 z_x^2,$$

where $\Gamma(z)$ is a function to be determined. Hence, using the freedom mentioned in Remark 3(ii) above, we choose

$$\beta = z_x^2. \tag{3.3}$$

The coefficient of w''' yields

$$\beta z_x^4 \Gamma(z) = 4\beta_x z_x^3 + 6\beta z_x^2 z_{xx},$$

where $\Gamma(z)$ is another function to be determined. Hence using (3.3) and rescaling Γ, we have

$$z_x \Gamma(z) + z_{xx}/z_x = 0,$$

which upon integration gives

$$\Gamma(t) + \ln z_x = \Theta(t),$$

where $\Theta(t)$ is a function of integration. Exponentiated this becomes

$$z_x \Gamma(z) = \Theta(t) \tag{3.4}$$

(recall Remark 2). Integrating again gives

$$\Gamma(z) = x\Theta(t) + \Sigma(t),$$

with $\Sigma(t)$ is another function of integration. By Remark 3(iii), we have

$$z = x\theta(t) + \sigma(t), \tag{3.5}$$

where $\theta(t)$ and $\sigma(t)$ are to be determined. From Eqs. (3.3) and (3.5), we have

$$\beta = \theta^2(t). \tag{3.6}$$

The coefficient of w'' yields

$$\beta z_x^4 \Gamma(z) = \beta(3z_{xx}^2 + 4z_x z_{xxx}) + 12\beta_x z_x z_{xx} + 6\beta_{xx} z_x^2 + \beta(\alpha z_x^2 + z_t^2),$$

where $\Gamma(z)$ is to be determined, and by Eqs. (3.5) and (3.6) this simplifies to

$$\theta^4 \Gamma(z) = \alpha\theta^2 + \left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)^2.$$

Hence by Remark 3(i) above

$$\alpha = -\frac{1}{\theta^2(t)}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)^2. \tag{3.7}$$

Let us see how Eq. (3.2) looks with the simplifications as determined so far, viz. (3.5)–(3.7):

$$\theta^6\{w'''' + ww'' + (w')^2\} + \theta^2\left(x\frac{d^2\theta}{dt^2} + \frac{d^2\sigma}{dt^2}\right)w' + 2\theta\frac{d^2\theta}{dt^2}w - \frac{d^2}{dt^2}\left[\left\{\frac{1}{\theta}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)\right\}^2\right] + \frac{6}{\theta^4}\left[\frac{d\theta}{dt}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)\right]^2 = 0. \tag{3.8}$$

We continue to make this an ordinary differential equation for $w(z)$. Then the remaining coefficients yield

$$\theta^6\gamma_1(z) = \theta^2\left(x\frac{d^2\theta}{dt^2} + \frac{d^2\sigma}{dt^2}\right), \tag{3.9}$$

$$\theta^6\gamma_2(z) = 2\theta\frac{d^2\theta}{dt^2}, \tag{3.10}$$

$$\theta^6\gamma_3(z) = -\frac{d^2}{dt^2}\left[\left\{\frac{1}{\theta}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)\right\}^2\right] + \frac{6}{\theta^4}\left[\frac{d\theta}{dt}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)\right]^2, \tag{3.11}$$

with $\gamma_1(z)$, $\gamma_2(z)$, and $\gamma_3(z)$ to be determined. First, since $z = x\theta(t) + \sigma(t)$ and the right-hand side of Eq. (3.9) is linear in x, consequently $\gamma_1(z) = Az + B$, where A and B are constants, and so

$$\theta^4[A(x\theta + \sigma) + B] = x\frac{d^2\theta}{dt^2} + \frac{d^2\sigma}{dt^2}. \tag{3.12}$$

Equating coefficients of powers of x gives

$$\frac{d^2\theta}{dt^2} = A\theta^5, \tag{3.13}$$

$$\frac{d^2\sigma}{dt^2} = \theta^4(A\sigma + B). \tag{3.14}$$

It is then easily seen from Eqs. (3.10) and (3.11) that

$$\gamma_2(z) = 2A, \quad \gamma_3(z) = -2(Az + B)^2.$$

[The Boussinesq equation is special in that, having satisfied Eq. (3.9), Eqs. (3.10) and (3.11) are satisfied automatically; slight modifications of the equation would not have significantly affected the application of the method until this point when further restrictions, on $\theta(t)$ and $\sigma(t)$, would arise from (3.10)and (3.11), severely limiting the set of similarity reductions.]

We conclude that the general similarity reduction of the Boussinesq equation (1.2) is given by

$$u(x,t) = \theta^2(t)w(z) - \frac{1}{\theta^2(t)}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)^2, \tag{3.15a}$$

$$z(x,t) = x\theta(t) + \sigma(t), \tag{3.15b}$$

where $\theta(t)$ and $\sigma(t)$ satisfy Eqs. (3.13) and (3.14), and $w(z)$ satisfies

$$w'''' + ww'' + (w')^2 + (Az + B)w' + 2Aw = 2(Az + B)^2. \tag{3.16}$$

It can be shown that of all the equations of the form

$$w'''' + ww'' + (w')^2 + f(z)w' + g(z)w = h(z),$$

with $f(z)$, $g(z)$, and $h(z)$ analytic, (3.16) is the most general one having the Painlevé property, that is, having no solutions with movable singularities except poles. In general, (3.16) is equivalent to the fourth Painlevé equation; but, when $A = 0$, it is equivalent to the second Painlevé equation, and, when $B = 0$ as well, it is equivalent to either first Painlevé equation of the Weierstrass elliptic function equation—see Appendix A for details. We remark that it is *not* essential to our method that all ordinary differential equations arising from similarity reductions are equivalent to one of the Painlevé equations (or more generally possess the Painlevé property). The Boussinesq equation is a completely integrable soliton equation for which the Painlevé conjecture[28] asserts that every ordinary differential equation arising from a similarity reduction is necessarily of the Painlevé type, in agreement with our results.

Henceforth, new symbols appearing in an equation obtained by integration are generally understood to be arbitrary constants. Furthermore, whenever we set a constant to be a specific value without further explanation, it is implied that this is easily seen to be without loss of generality.

There are three cases to consider.

Case 1. $A = 0$, $B = 0$: In this case, the general solutions of Eqs. (3.13) and (3.14) are

$$\theta(t) = a_1 t + a_0, \quad \sigma(t) = b_1 t + b_0,$$

and the similarity reduction of the Boussinesq equation is

$$u(x,t) = (a_1 t + a_0)^2 w(z) - \left(\frac{a_1 x + b_1}{a_1 t + a_0}\right)^2, \tag{3.17a}$$

$$z = x(a_1 t + a_0) + b_1 t + b_0, \tag{3.17b}$$

where $w(z)$ satisfies

$$w'' + \tfrac{1}{2}w^2 = c_1 z + c_0. \tag{3.17c}$$

Equation (3.17c) is the same as Eq. (2.6) and so, as we remarked in Sec. II, it is equivalent to either the first Painlevé equation (2.7) or the Weierstrass elliptic function equation. We note also that the traveling wave reduction arises as the special case of (3.17) where $a_1 = 0$ and $b_1 \neq 0$. However, if $a_1 = 0$, then we set $a_1 = 1$, $a_0 = b_1 = b_0 = 0$, and obtain the similarity reduction

$$u(x,t) = t^2 w(z) - x^2/t^2, \quad z = xt, \tag{3.18}$$

where $w(z)$ satisfies Eq. (3.17c). This is a new reduction of the Boussinesq equation to the first Painlevé equation.

With z and w as invariants, Eqs. (3.18) define the point transformation group

$$(x,t,u) \to (\gamma^{-1}x, \gamma t, \gamma^2 u + (\gamma^2 - \gamma^{-4})x^2/t^2).$$

The infinitesimals associated with this are

$$X = -x, \quad T = t, \quad U = 2u + 6x^2/t^2, \tag{3.19}$$

which clearly are not a special case of the infinitesimals obtained by the classical Lie group method [cf. (2.5)].

Case 2. $A=0$, $B\neq 0$: In this case the general solution of Eqs. (3.13) and (3.14) are

$$\theta(t) = a_1 t + a_0,$$

$$\sigma(t) = \begin{cases} \frac{1}{30} B a_1^{-2}(a_1 t + a_0)^6 + b_1 t + b_0, & \text{if } a_1 \neq 0, \\ \frac{1}{2} B a_0^2 t^2 + b_1 t + b_0, & \text{if } a_1 = 0. \end{cases}$$

Case (a). $a_1=0$: The similarity reduction of the Boussinesq equation is

$$u(x,t) = a_0^2 w(z) - (Ba_0^2 t + b_1)^2/a_0^2, \tag{3.20a}$$

$$z = a_0 x + \tfrac{1}{2} B a_0^2 t^2 + b_1 t + b_0, \tag{3.20b}$$

where $w(z)$ satisfies

$$w''' + ww' + Bw = 2B^2 z + c_0. \tag{3.21}$$

Equation (3.21) is the same as Eq. (2.12) and so, as remarked in Sec. II, it is equivalent to the second Painlevé equation (2.14)—see, also, Appendix A. We set $a_0 = 1$, $b_1 = b_0 = 0$, in (3.20), in which case it just reduces to the "nonclassical" similarity reduction (2.11) (cf. Refs. 19, 24, 25, and 29).

Case (b). $a_1\neq 0$: The similarity reduction of the Boussinesq equation is

$$u(x,t) = (a_1 t + a_0)^2 w(z) - \left(\frac{a_1^2 x + \frac{1}{5} B(a_1 t + a_0)^5 + a_1 b_1}{a_1(a_1 t + a_0)} \right)^2, \tag{3.20a'}$$

$$z = x(a_1 t + a_0) + [B/30 a_1^2](a_1 t + a_0)^6 + b_1 t + b_0, \tag{3.20b'}$$

where $w(z)$ satisfies (3.21). We set $a_1 = 1$, $a_0 = b_1 = b_0 = 0$, and obtain

$$u(x,t) = t^2 w(z) - (x + \lambda t^5)^2/t^2, \quad z = xt + \tfrac{1}{6}\lambda t^6, \tag{3.22}$$

where $w(z)$ satisfies (3.21) (we have also set $B = 5\lambda$). This is another new reduction of the Boussinesq equation; this time to the second Painlevé equation (2.14). The infinitesimals associated with the transformation group defined by (3.22) are

$$X = -(x + \lambda t^5), \quad T = t,$$
$$U = 2u + 2(x + \lambda t^5)(3x - 2\lambda t^5)/t^2. \tag{3.23}$$

[We note that if $\lambda = 0$ in (3.22) and (3.23), they reduce to (3.18) and (3.19).]

Case 3. $A\neq 0$: In this case we can set $B = 0$ in Eq. (3.14). Multiplying Eq. (3.13) by $d\theta/dt$ and integrating gives

$$\left(\frac{d\theta}{dt}\right)^2 = \frac{1}{3} A\theta^6 + A_0, \tag{3.24}$$

where A_0 is a constant. There are two possibilites.

Case (a). $A_0=0$: Equation (3.24) has the solution

$$\theta(t) = c_0(t + t_0)^{-1/2}, \tag{3.25}$$

with $c_0^4 = 3/(4A)$. Substituting this into Eq. (3.14) and solving yields

$$\sigma(t) = c_1(t + t_0)^{3/2} + c_2(t + t_0)^{-1/2}.$$

Therefore we may set $t_0 = 0$, $c_0 = 1$, and $c_2 = 0$, and obtain the similarity reduction

$$u(x,t) = t^{-1} w(z) - \tfrac{1}{4} t^{-2}(x - 3c_1 t^2)^2,$$
$$z = xt^{-1/2} + c_1 t^{3/2}, \tag{3.26}$$

where $w(z)$ satisfies

$$w'''' + ww'' + (w')^2 + \tfrac{3}{4} zw' + \tfrac{3}{2} w = \tfrac{9}{8} z^2. \tag{3.27}$$

Note that the scaling reduction (2.8) arises as the special case of (3.26) with $c_1 = 0$. If $c_1 \neq 0$, this is a new similarity reduction, namely, to the fourth Painlevé equation, since if in (3.27) we make the transformation $w(z) = g(z) + z^2/4$, then $g(z)$ satisfies Eq. (2.9) and therefore Eq. (3.27) is also equivalent to the fourth Painlevé equation (2.10)—see, also, Appendix A.

Case (b). $A\neq 0$: Equation (3.24) can be solved in terms of Jacobian elliptic functions (cf. Ref. 32). Furthermore we may set

$$A_0 = k^2, \quad A = (k^2 + 1)/3k^2, \tag{3.28}$$

where k is a constant to be chosen. For this choice of constants, the transformation

$$\theta^2(t) = 1/[\eta^2(t) - A] \tag{3.29}$$

reduces (3.24) to the normal form

$$\left(\frac{d\eta}{dt}\right)^2 = (1 - \eta^2)(1 - k^2\eta^2), \tag{3.30}$$

provided that

$$k^2 = \tfrac{1}{2}(1 \pm i\sqrt{3}) \tag{3.31}$$

(which we may assume without loss of generality). The solution of (3.30) is the Jacobian elliptic function $\mathrm{sn}(t + t_0;k)$, and so

$$\theta(t) = (\mathrm{sn}^2(t + t_0;k) - (k^2 + 1)/3k^2)^{-1/2}. \tag{3.32}$$

Equation (3.14) becomes

$$\frac{d^2\sigma}{dt^2} = \frac{k^2 + 1}{3k^2} \theta^4 \sigma,$$

which has the solution

$$\sigma(t) = [\, C([(2 - k^2)/3k^2]t - k^{-2} E(t + t_0;k)) + D\,]\theta(t), \tag{3.33}$$

where $E(t + t_0;k)$ is the elliptic integral of the second kind given by

$$E(t + t_0;k) = \int_0^{t + t_0} [1 - k^2\, \mathrm{sn}^2(s;k)]\, ds$$

and C and D are arbitrary constants—we set $D = 0$.

Therefore we have the following similarity reduction:

$$u(x,t) = (\mathrm{sn}^2(t+t_0;k) - A)^{-1}w(z) - [\,C(\mathrm{sn}^2(t+t_0;k) - A) - \{x + C([(2-k^2)/3k^2]t - k^{-2}E(t+t_0;k))\} \times [\mathrm{sn}(t+t_0;k)\sqrt{(1-\mathrm{sn}^2(t+t_0;k))(1-k^2\,\mathrm{sn}^2(t+t_0;k))}/(\mathrm{sn}^2(t+t_0;k) - A)\,]^2, \tag{3.34a}$$

with

$$z = [x + C([(2-k^2)/3k^2]t - k^{-2}E(t+t_0;k))] \times (\mathrm{sn}^2(t+t_0;k) - A)^{-1/2}, \tag{3.34b}$$

and

$$k^2 = \frac{1}{2} \pm i\frac{\sqrt{3}}{2}, \quad A = \frac{k^2+1}{3k^2} = \frac{1}{2} \mp \frac{i}{2\sqrt{3}}, \tag{3.34c}$$

where $w(z)$ satisfies

$$w'''' + ww'' + (w')^2 + Azw' + 2Aw = 2A^2z^2. \tag{3.34d}$$

This is another new similarity reduction, again to the fourth Painlevé equation (2.10).

As for the other new similarity reductions given above [(3.18) and (3.22)], we can write down the infinitesimals associated with the transformation groups defined by (3.26) and (3.34). Again they are not special cases of those obtained by the classical Lie group method.

In all three cases we have obtained new similarity reductions of the Boussinesq equation more general than those previously obtained (though, interestingly, the resulting ordinary differential equations are the same). As mentioned above, these similarity reductions are associated with Lie point transformations (since they depend only on the independent and dependent variables and not upon the derivatives of the dependent variable). It remains an open question as to whether all these new similarity reductions and their associated transformations can be obtained using any of the other generalizations of the classical Lie method, such as the nonclassical method of Bluman and Cole[18] (cf. Ref. 23), and the method developed by Bluman *et al.*[22] However, even if theoretically they can be obtained by either of these methods it seems that our method is somewhat simpler to implement; in fact, it appears to be simpler than calculating the classical Lie point symmetries manually.

It can be shown that for the similarity reductions of the Boussinesq equation that cannot be obtained using the classical Lie group method, the associated group transformation does not map the Boussinesq equation into itself, whereas the similarity reductions obtained by the classical Lie group method do. For example, consider the similarity reduction

$$u(x,t) = t^2w(z) - (x+\lambda t^5)^2/t^2, \quad z = xt + \tfrac{1}{6}\lambda t^6. \tag{3.22}$$

The one-parameter (γ) group associated with this similarity reduction is given by

$$x \to \gamma^{-1}x + \tfrac{1}{6}\lambda\gamma^{-1}(1-\gamma^6)t^5, \tag{3.35a}$$

$$t \to \gamma t, \tag{3.35b}$$

$$u \to \gamma^2 u + \gamma^2(1-\gamma^{-6}) \times \{x^2/t^2 + \tfrac{1}{3}\lambda x t^3 + \tfrac{1}{36}\lambda^2 t^8(1-25\gamma^{-6})\}. \tag{3.35c}$$

This group maps solutions of the Boussinesq equation (1.2) into solutions of

$$u_{tt} + uu_{xx} + u_x^2 + u_{xxxx} = (\gamma^6 - 1)t^{-2}\Phi, \tag{3.36a}$$

where

$$\Phi := (x^2 + \tfrac{1}{3}\lambda x t^5 - \tfrac{2}{3}\lambda^2 t^{10})u_{xx} + 4(x+\lambda t^5)u_x + 2u + \tfrac{5}{3}\lambda t^6 u_{xt} - t^2u_{tt} + 6x^2/t^2 + 2\lambda x t^3 - \tfrac{112}{3}\lambda^2 t^8. \tag{3.36b}$$

If u is the similarity reduction (3.22), then it is easily seen that $\Phi \equiv 0$, i.e., the group (3.35) maps the Boussinesq equation (1.2) into the "perturbed Boussinesq equation" (3.36a), but (3.36b) is identically zero. Therefore the perturbed equation is identical to the Boussinesq equation when u is given by (3.22).

In order to understand why the perturbation Φ must vanish identically, consider the infinitesimals

$$X = -(x+\lambda t^5), \quad T = t, \quad U = 2u + 2(x+\lambda t^5)(3x - 2\lambda t^5)/t^2, \tag{3.23}$$

for the similarity reduction (3.22). The similarity reduction necessarily satisfies the invariant surface condition

$$X(x,t,u)u_x + T(x,t,u)u_t = U(x,t,u),$$

i.e.,

$$\psi := (x+\lambda t^5)u_x - tu_t + 2u + 6x^2/t^2 + 2\lambda x t^3 - 4\lambda^2 t^8 = 0. \tag{3.37}$$

It is easily shown that

$$\Phi = (x - \tfrac{2}{3}\lambda t^5)\psi_x + t\psi_t + \psi. \tag{3.38}$$

IV. JUSTIFICATION OF THE SPECIAL FORM (1.4)

We show here that it is sufficient to seek a similarity reduction of the Boussinesq equation (1.2) in the special form

$$u(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)), \tag{4.1}$$

rather than the more general form

$$u(x,t) = U(x,t,w(z(x,t))). \tag{4.2}$$

Substituting (4.2) into (1.2) yields

$$[\,U_{tt} + 2U_{tw}w'z_t + U_{ww}(w')^2z_t^2 + U_w(w''z_t + w'z_{tt})\,] + U[\,U_{xx} + 2U_{xw}w'z_x + U_{ww}(w')^2z_x^2 + U_w(w''z_x + w'z_{xx})\,] + U_x^2 + 2U_xU_ww'z_x + U_w^2(w')^2z_x^2 + U_{xxxx} + 4U_{xxxw}w'z_x + 6U_{xxww}(w')^2z_x^2 + 4U_{xwww}(w')^3z_x^3 + U_{wwww}(w')^4z_x^4 + 6U_{xxw}(w'z_{xx} + w''z_x^2) + 12U_{xww}[\,w'w''z_x^3 + (w')^2z_xz_{xx}\,]$$

$$+ 6U_{www}[(w')^2w''z_x^4 + (w')^3z_x^2z_{xx}] + 4U_{xw}(w'''z_x^3 + 3w''z_xz_{xx} + w'z_{xxx})$$
$$+ U_{ww}[\{4w'w''' + 3(w'')^2\}z_x^4 + 18w'w''z_x^2z_{xx} + (w')^2(4z_xz_{xxx} + 3z_{xx}^2)]$$
$$+ U_w[w''''z_x^4 + 6w'''z_x^2z_{xx} + w''(4z_x z_{xxx} + 3z_{xx}^2) + w'z_{xxxx}] = 0. \tag{4.3}$$

For this to be an ordinary differential equation in $w(z)$, the ratios of different derivatives of $w(z)$ must be functions of w and z. Using the coefficient of w''' (i.e., $U_wz_x^4$) as the normalizing coefficient, the coefficients of $w'w'''$ and $(w'')^2$ require that

$$U_wz_x^4\Gamma(w,z) = U_{ww}z_x^4, \tag{4.4}$$

where $\Gamma(w,z)$ is a function to be determined. Hence

$$\Gamma(w,z) = U_{ww}/U_w,$$

which after two integrations yields

$$U(x,t,w) = \Theta(x,t)\Gamma(w,z) + \Phi(x,t), \tag{4.5}$$

with $\Theta(x,t)$ and $\Phi(x,t)$ arbitrary functions (cf. Remarks 2 and 3 in Sec. III). Therefore it is sufficient to seek similarity reductions of the Boussinesq equation (1.2) in the form (4.1).

Therefore, if we seek a similarity reduction of the Boussinesq equation in the general form (4.2), we are naturally led to the special form (4.1). Although, for many partial differential equations such as the Boussinesq equation, it is sufficient to seek similarity reductions in the special form (4.2), for some others it may be necessary to transform the dependent variable before using (4.1); however, the assumption (4.2) leads naturally to the required transformation.

For example, consider the Harry–Dym equation (cf. Ref. 33).

$$u_t + 2(u^{-1/2})_{xxx} = 0, \tag{4.6}$$

which can be solved by inverse scattering[34] (see, also, Ref. 12) and is related to the Korteweg–de Vries and modified Korteweg–de Vries equations through hodograph transformations.[35] Let us seek a similarity reduction in the form (4.2). Substitution yields

$$U_t + U_ww'z_t - \tfrac{15}{4}U^{-7/2}(U_x + U_ww'z_x)^3$$
$$+ \tfrac{9}{2}U^{-5/2}(U_x + U_ww'z_x)[U_{xx} + 2U_{xw}w'z_x + U_{ww}(w')^2z_x^2 + U_w(w''z_x^2 + w'z_{xx})]$$
$$- U^{-3/2}[U_{xxx} + 3U_{xxw}w'z_x + 3U_{xww}(w')^2z_x^2 + U_{www}(w')^3z_x^3 + 3U_{xw}(w''z_x^2 + w'z_{xx})$$
$$+ 3U_{ww}\{w'w''z_x^3 + (w')^2z_xz_{xx}\} + U_w(w'''z_x^3 + 3w''z_xz_{xx} + w'z_{xxx})] = 0. \tag{4.7}$$

Using the coefficient of w''' (i.e., $U^{-3/2}U_wz_x^3$) as the normalizing coefficient, the coefficient of $w'w''$ requires that

$$U^{-3/2}U_wz_x^3\Gamma(w,z) = \tfrac{3}{2}U^{-5/2}U_w^2z_x^3 - U^{-3/2}U_{ww}z_x^3,$$

that is,

$$\Gamma(w,z) = -\tfrac{3}{2}U_w/U + U_{ww}/U_w, \tag{4.8}$$

where $\Gamma(w,z)$ is a function to be determined. Integrating twice yields

$$U^{-1/2}(x,t) = \Theta(x,t)\Gamma(w,z) + \Phi(x,t), \tag{4.9}$$

with $\Theta(x,t)$ and $\Phi(x,t)$ arbitrary functions (cf. Remark 2 in Sec. III). Hence it is sufficient to seek similarity reductions of the Harry–Dym equation (4.6) in the form

$$u^{-1/2}(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)).$$

Alternatively we could first make the transformation $v = u^{-1/2}$ and then seek similarity reductions in the form (4.1). Obvious as this transformation is, our method leads to it systematically.

V. DISCUSSION

In this paper we have developed a direct method for determining similarity reductions of a given partial differential equation. However, there are a number of open questions our method poses. First, what is the relationship (if any) between our method and other generalizations of the classical Lie method, such as those of Bluman and Cole[18] (cf. Ref. 23), Olver and Rosenau,[19] and Bluman *et al.*[22] ? In their generalization of the method of Bluman and Cole,[18] Olver and Rosenau[19] showed that in order to determine a group-invariant solution to a given partial differential equation, one could try *any* group of infinitesimal transformations whatsoever. Generally, for any specific group and any specific equation, there will be *no* solutions of the equation invariant under the group, and so the question becomes how does one determine *a priori* which groups will give meaningful similarity reductions? One possibility is that by seeking a reduction of a certain form (as done in this paper), one is naturally led to the appropriate group (i.e., the requirement that the similarity reduction reduce the partial differential equation to an ordinary differential equation is equivalent to the *side conditions* in the terminology of Olver and Rosenau[18]).

Second, what kind of "*symmetries*" of the Boussinesq equation are those we have obtained that are not found using the classcial Lie method? (They are "*weak symmetries*" in the terminology of Olver and Rosenau.[19]) As shown in Sec. III, the associated group of infinitesimal transformations does *not* map solutions of the Boussinesq equation into other solutions of the Boussinesq equation, but rather into solutions of other equations.

The idea of making the ansatz that a similarity reduction of a given partial differential equation have a particular form has been suggested previously in the literature. For example, (i), Gilding[36] seeks solutions of the porous media equation

$$u_t = (u^m)_{xx}\,, \quad m>1,$$

in the form

$$u(x,t) = \mu(t) f(z), \quad z = \rho(t)[x+\lambda(t)];$$

and (ii), Fushchlich, in a series of papers with various coauthors,[37] has obtained exact solutions of several nonlinear relativistic and nonlinear wave equations (including the nonlinear Dirac, Klein–Gordon, Maxwell, and Schrödinger equations) in three spatial and one temporal dimension, using their symmetry properties and seeking solutions in the form

$$u(x_0,x_1,x_2,x_3) = A(x_0,x_1,x_2,x_3)\,w(z_1,z_2,z_3) + B(x_0,x_1,x_2,x_3)\,,$$

where

$$z = (z_1(x_0,x_1,x_2,x_3), z_2(x_0,x_1,x_2,x_3), z_3(x_0,x_1,x_2,x_3))$$

are the new independent variables, $w(z_1,z_2,z_3)$ the new dependent variable, and $A(x_0,x_1,x_2,x_3)$ and $B(x_0,x_1,x_2,x_3)$ are determined.

We have applied the method to several other integrable equations including Burgers' equation

$$u_t + uu_x + u_{xx} = 0\,, \tag{5.1}$$

which can be mapped into the linear heat equation through the Cole–Hopf transformation[38]; the Korteweg–de Vries equation

$$u_t + uu_x + u_{xxx} = 0\,, \tag{5.2}$$

which can be solved by inverse scattering[8]; and the modified Korteweg–de Vries equation

$$u_t + u^2u_x + u_{xxx} = 0\,, \tag{5.3}$$

which also can be solved by inverse scattering.[39] However, for these three equations, the similarity reductions obtained are precisely the same as those obtained using the classical Lie method of infinitesimal transformations (for further details see Appendices B, C, and D, respectively, which also provide further examples of the application of our method).

There is much current interest in the mathematically and physically significant determination of similarity reductions of given partial differential equations. (In addition to the references mentioned above, the interested reader might also consult Refs. 40–43, and the references therein.) Our method is a practical and direct one for finding similarity reductions; it has generated similarity reductions that, to the best of our knowledge, are previously unknown. It seems probable that the method can be generalized to higher-order equations with more independent and dependent variables.

ACKNOWLEDGMENTS

This paper arose as a consequence of a lecture given by one of us (PAC) at the Gregynog Symposium on "Differential Equations," July 1987. We would like to thank the organizers for the invitations to attend and the U. K. Science and Engineering Research Council for supporting the meeting (Grant GR/D 80032). PAC would like to thank Peter Olver, Harvey Segur, Steve Williams, and Pavel Winternitz for helpful discussions and the SERC for partial financial support.

APPENDIX A: REDUCTION TO PAINLEVÉ EQUATIONS

In this appendix it is shown that of all the equations of the form

$$w'''' + ww'' + (w')^2 + f(z)w' + g(z) = h(z)\,, \tag{A1}$$

with $f(z)$, $g(z)$, and $h(z)$ analytic, the most general one having the Painlevé property, that is, having no solutions with movable singularities except poles, is given by

$$w'''' + ww'' + (w')^2 + (Az+B)w' + 2Aw = 2(Ax+B)^2\,, \tag{A2}$$

where A and B are arbitary constants. To show this we follow Ablowitz *et al.*[28] in seeking a solution of Eq. (A2) in the Laurent series form

$$w(z) = \sum_{j=0}^{\infty} w_j (z-z_0)^{j+p}\,, \tag{A3}$$

with z_0 an arbitrary constant, $w_0 \neq 0$ and $w_j, j \geqslant 0$, constants to be determined. Leading-order analysis shows that

$$w_0 = -12, \quad p = -2. \tag{A4}$$

Substituting into (A1) and equating coefficients of powers yields for $j \geqslant 1$ the recursion relation

$$\begin{aligned}(j+1)(j-4)&(j-5)(j-6)w_j \\ &+ \frac{1}{2}(j-4)(j-5)\sum_{k=1}^{j-1} w_k w_{j-k} \\ &= -\sum_{k=0}^{j-3} f_k (j-k-5) w_{j-k-3} \\ &\quad - \sum_{k=0}^{j-4} g_k w_{j-k-4} + h_{j-6}\,,\end{aligned} \tag{A5a}$$

where

$$f(z) = \sum_{k=0}^{\infty} f_k (z-z_0)^k, \text{ etc.} \tag{A5b}$$

(defining $w_j = 0$ for $j<0$, etc.). This determines w_j for $j \geqslant 1$ except for $j = 4,5,6$, which are the so-called resonances. For each resonance there is a compatibility condition that must be identically satisfied for Eq. (A1) to have a solution in the form (A3). From Eq. (A5) we obtain

$$w_1 = 0, \quad w_2 = 0, \quad w_3 = f_0\,. \tag{A6}$$

The compatibility conditions for $j=4$ and $j=5$ are

$$g_0 = 2f_1, \quad g_1 = 2f_2\,,$$

respectively. Since z_0 is arbitrary, necessarily

$$g(z) = 2\frac{df}{dz}, \quad \frac{dg}{dz} = \frac{d^2f}{dz^2}. \tag{A7}$$

These hold simultaneously if and only if

$$\frac{d^2f}{dz^2} = 0\,,$$

i.e.,

$$f(z) = Az + B\,, \quad g(z) = 2A\,, \tag{A8a}$$

with A and B arbitrary constants. The compatibility condition for $j=6$ is

$$h_0 = 2f_0^2 .$$

Thus

$$h(z) = 2(Az+B)^2 . \qquad \text{(A8b)}$$

Unless $f(z)$, $g(z)$, and $h(z)$ are as given in Eqs. (A8), the compatibility conditions are violated and so Eq. (A1) has the Painlevé property only if it has the special form (A2).

In order to complete the proof that Eq. (A2) has the Painlevé property, we show that no solution of it has a movable essential singularity by reducing it to known such equations.

Case (a). $A=0$, $B=0$: Integrating Eq. (A2) twice yields

$$\frac{d^2w}{dz^2} + \frac{1}{2} w^2 = c_1 z + c_0 . \qquad \text{(A9)}$$

If $c_1 = 0$, $w(z)$ is a Weierstrass elliptic function (cf. Ref. 32); otherwise (A9) is the first Painlevé equation (cf. Ref. 26). In either case, all solutions possess the Painlevé property (in fact, are meromorphic); hence no solution of Eq. (A9) has a movable essential singularity.

Case (b). $A=0$, $B\neq 0$: Integrating Eq. (A2) once yields

$$\frac{d^3w}{dz^3} + w\frac{dw}{dz} + Bw = 2B^2 z + c_2 . \qquad \text{(A10)}$$

Then make the transformation

$$w(z) = B^{2/3}W(Z) + Bz + c_2/2B,$$
$$Z = -(B^{1/3}z + \tfrac{1}{2}c_2 B^{-5/3}) , \qquad \text{(A11)}$$

which produces

$$\frac{d^3W}{dZ^3} + W\frac{dW}{dZ} - \left(2W + Z\frac{dW}{dZ}\right) = 0 . \qquad \text{(A12)}$$

Whitham (see Refs. 27 and 30) noted that solutions of this equation are related to solutions of the second Painlevé equation

$$\frac{d^2V}{dZ^2} = 2V^3 + ZV + \alpha , \qquad \text{(A13)}$$

with α an arbitrary constant. Actually, as shown by Fokas and Ablowitz,[30] there is a one-to-one correspondence between solutions of (A12) and (A13) given by

$$W(Z) = -6(V'(Z) + V^2(Z)) , \qquad \text{(A14a)}$$
$$V(Z) = [W'(Z) + 6a]/[2W(Z) - 6Z] , \qquad \text{(A14b)}$$

where $' := d/dZ$. [Equation (A14a) is just the scaling, or self-similar, reduction of the Miura transformation[44] relating solutions of the modified Korteweg–de Vries equation (5.3) to solutions of the Korteweg–de Vries equation (5.2).] All solutions of the second Painlevé equation possess the Painlevé property (in fact, are meromorphic); hence no solution of Eq. (A10) has a movable essential singularity.

Case (c). $A\neq 0$: The transformation

$$w \to \left(\frac{4A}{3}\right)^{1/2} w, \quad z \to \left(\frac{3}{4A}\right)^{1/4} z - \frac{B}{A},$$

takes (A2) to the form

$$w'''' + w'' + (w')^2 + \tfrac{3}{4} zw' + \tfrac{3}{2} w = \tfrac{9}{8} z^2 . \qquad \text{(A15)}$$

Hirota and Satsuma[45] show that there is a "Miura-type" transformation relating solutions of the modified Boussinesq equation

$$q_{tt} - q_t q_{xx} - \tfrac{1}{2} q_x^2 q_{xx} + q_{xxxx} = 0 , \qquad \text{(A16)}$$

to solutions of the Boussinesq equation (1.2) (see, also, Refs. 29 and 46). The Bäcklund transformation

$$v_x(x,t) = -q_t + \sqrt{3} q_{xx} - \tfrac{1}{2} q_x^2 , \qquad \text{(A17a)}$$
$$v_t(x,t) = \sqrt{3} q_{xt} + q_{xxx} - q_x q_t - \tfrac{1}{6} q_x^3 + \delta , \qquad \text{(A17b)}$$

where δ is a constant, is easily seen to take a solution q of the modified Boussinesq equation (A16) to a solution v of the potential Boussinesq equation

$$v_{tt} + v_x v_{xx} + v_{xxxx} = 0; \qquad \text{(A18)}$$

furthermore $u = v_x$ is a solution of the Boussinesq equation (1.2). The modified Boussinesq equation (A16) has the similarity solution (cf. Ref. 29)

$$q(x,t) = -\gamma \ln t + p(z), \quad z = xt^{-1/2} , \qquad \text{(A19)}$$

where $p(z)$ satisfies

$$\gamma + \frac{3z}{4} p' + \frac{z^2}{4} p'' + \left(\gamma + \frac{1}{2} zp'\right) p''$$
$$- \frac{1}{2} (p')^2 p'' + p'''' = 0 ,$$

with $' := d/dz$; and if we now make the transformation

$$p'(z) = -3^{3/4} Q(Z) - z, \quad Z = 3^{1/4} z/2 , \qquad \text{(A20)}$$

then $Q(Z)$ satisfies the fourth Painlevé equation

$$\frac{d^2Q}{dZ^2} = \frac{1}{2Q}\left(\frac{dQ}{dZ}\right)^2 + \frac{3}{2} Q^3 + 4ZQ^2$$
$$+ 2(Z^2 - \alpha)Q + \frac{\beta}{Q} , \qquad \text{(A21)}$$

with $\alpha = 8\gamma/(9\sqrt{3})$ and β an arbitrary constant (see, also, Ref. 46). The Boussinesq equation (1.2) and the potential Boussinesq equation (A18) possess the similarity reductions

$$u(x,t) = t^{-1} w(z) - x^2/4t^2 , \quad z = xt^{-1/2} , \qquad \text{(A22a)}$$
$$v(x,t) = t^{-1/2} r(z) , \quad z = xt^{-1/2} , \qquad \text{(A22b)}$$

where $w(z)$ satisfies Eq. (A15) and $r(z)$ satisfies

$$r'''' + r'r'' - \tfrac{1}{2}(r + zr') = 0 . \qquad \text{(A23)}$$

Therefore, Eqs. (A17)–(A22) show that if $Q(Z)$ is a solution of the fourth Painlevé equation, then

$$w(z) := -\frac{3\sqrt{3}}{2}\left(\frac{dQ}{dZ} + Q^2(Z) + 2ZQ(Z) + 3Z^2\right)$$
$$+ \frac{9\sqrt{3}}{8} \alpha - \sqrt{3} , \qquad \text{(A24a)}$$
$$Z := 3^{1/4} z/2 , \qquad \text{(A24b)}$$

is a solution of Eq. (A15).

What all this shows is that from any solution of the fourth Painlevé equation we can obtain a solution of (A15). To obtain the converse we substitute the similarity reductions (A19) and (A22) into the Bäcklund transformation (A17) and easily see that if $r(z)$ is a solution of (A23), then

$$Q(Z):= -3^{-3/4}\left(\frac{\frac{3}{2}(r-zr')+\sqrt{3}r''-z^3-2\sqrt{3}z}{r'+z^2+2\gamma-\sqrt{3}}\right),$$
$$Z=3^{1/4}z/2\,, \qquad \text{(A25)}$$
satisfies the fourth Painlevé equation (A21); furthermore solutions of Eqs. (A15) and (A23) are related by

$$w(z)=\frac{dr}{dz}+\frac{z^2}{4}. \qquad \text{(A26)}$$

Equations (A24)–(A26) provide a one-to-one relationship between solutions of Eq. (A15) and solutions of the fourth Painlevé equation. All solutions of the fourth Painlevé equation possess the Painlevé property, i.e., have no movable essential singularities (in fact, are meromorphic). Therefore no solution of Eq. (A15) has a movable essential singularity.

We remark that there is also a direct method to show that no solution of Eq. (A2) has a movable essential singularity. Making the transformation

$$w(z)=v'(z)-(Az+B)^2/A\,, \qquad \text{(A26')}$$

we obtain a fifth-order equation easily integrated twice to yield

$$v'''+\frac{1}{2}(v')^2-\frac{Az+B}{A}\,[(Az+B)v'-Av]=c_1z+c_2\,. \qquad \text{(A27)}$$

Multiplying by v'' and integrating again yields

$$\tfrac{1}{2}(v'')^2+\tfrac{1}{6}(v')^3-(1/2A)[(Az+B)v'-Av]^2 = (c_1/A)[(Az+B)v'-Av]+c_2v'+c_3\,. \qquad \text{(A28)}$$

This is equivalent (through rescaling and translation of the variables) to an equation given by Chazy,[47]

$$(y'')^2+4(y')^3+(zy'-y)^2+\alpha y'+\beta=0\,, \qquad \text{(A29)}$$

with α and β constants. According to Chazy, this is "an algebraic transformation of the fourth Painlevé equation" [Eq. (A29) is sometimes referred to as Chazy IV, cf. Refs. 29 and 48]. Furthermore, as shown by Chazy,[47] for any solution of (A29), $\exp\{\int^z y(s)ds\}$ is analytic except at the points $0, \infty$. Hence we conclude that no solution of Eq. (A26), and hence also of Eq. (A2), has a movable essential singularity.

APPENDIX B: BURGERS' EQUATION

In this appendix we outline how to determine the similarity reductions of Burgers' equation

$$u_t+uu_x+u_{xx}=0\,, \qquad \text{(B1)}$$

using the method developed in this paper. As with the Boussinesq equation (1.2), it suffices to seek similarity reductions in the special form

$$u(x,t)=\alpha(x,t)+\beta(x,t)w(z(x,t)). \qquad \text{(B2)}$$

Substituting (B2) and (B1) and collecting coefficients yields

$$\beta z_x^2w''+(2\beta_xz_x+\beta z_{xx}+\beta z_t+\alpha\beta z_x)w' +(\beta_{xx}+\beta_t+\alpha\beta_x+\alpha_x\beta)w +\beta^2z_xww'+\beta\beta_xw^2+\alpha_{xx}+\alpha_t+\alpha\alpha_x=0\,. \qquad \text{(B3)}$$

We use the coefficient of w'' as the normalizing coefficient. For this to be an ordinary differential equation, from the coefficient of ww' we get

$$\beta z_x^2\Gamma(z)=\beta^2z_x\,,$$

where $\Gamma(z)$ is to be determined. Using the freedom in Remark 3(i) in Sec. III, we take

$$\beta=z_x\,. \qquad \text{(B4)}$$

The coefficient of w^2 gives

$$\beta z_x^2\Gamma(z)=\beta\beta_x\,,$$

where $\Gamma(z)$ is to be determined. Using (B4), integrating twice, and using the freedoms in Remark 2 and 3(iii), we have

$$z=x\theta(t)+\sigma(t)\,,\quad \beta=\theta(t)\,, \qquad \text{(B5)}$$

where $\theta(t)$ and $\sigma(t)$ are to be determined. Equation (B3) simplifies to

$$\theta^3(w''+ww')+\theta\left\{\left(x\frac{d\theta}{dt}+\frac{d\sigma}{dt}\right)+\alpha\theta\right\}w' +\left\{\frac{d\theta}{dt}+\alpha_x\theta\right\}w+\alpha_{xx}+\alpha_t+\alpha\alpha_x=0\,. \qquad \text{(B6)}$$

This is an ordinary differential equation for $w(z)$ provided that

$$\alpha=-\frac{1}{\theta}\left(x\frac{d\theta}{dt}+\frac{d\sigma}{dt}\right), \qquad \text{(B7)}$$

$$\theta\frac{d^2\theta}{dt^2}-2\left(\frac{d\theta}{dt}\right)^2=A^2\theta^6\,, \qquad \text{(B8)}$$

$$\theta\frac{d^2\sigma}{dt^2}-2\frac{d\theta}{dt}\frac{d\sigma}{dt}=\theta^5(A^2\sigma+2B)\,, \qquad \text{(B9)}$$

with A and B arbitrary constants. Multiplying (B8) by $2\theta^{-2}\,d\theta/dt$ and integrating gives

$$\left(\frac{d\theta}{dt}\right)^2=A^2\theta^6+C^2\theta^4\,, \qquad \text{(B10)}$$

with C an arbitrary constant.

Therefore the general similarity reduction of Burgers' equation (B1) is given by

$$u(x,t)=\theta(t)w(z)-\frac{1}{\theta}\left(x\frac{d\theta}{dt}+\frac{d\sigma}{dt}\right),$$
$$z=x\theta(t)+\sigma(t)\,,$$

where $\theta(t)$ and $\sigma(t)$ satisfy (B9) and (B10).

There are four cases to consider.

Case 1. $A=0$, $C=0$: Here the solutions are

$$\theta(t)=\theta_0,\quad \sigma(t)=Bt^2+c_1t+c_2\,.$$

We set $\theta_0=1$ and obtain the similarity reduction,

$$u(x,t)=w(z)-2Bt-c_1,\quad z=x+Bt^2+c_1+c_2\,. \qquad \text{(B11)}$$

Case 2. $A\neq0$, $C=0$: We set $A=-\frac{1}{2}$ and $B=0$. Then

$$\theta(t)=(t-t_0)^{-1/2},$$
$$\sigma(t)=c_3(t-t_0)^{1/2}+c_4(t-t_0)^{-1/2}\,.$$

Setting $t_0=1$, $c_4=0$, we obtain

$$u(x,t)=t^{-1/2}w(z)+x/2t-\tfrac{1}{2}c_3\,. \qquad \text{(B12)}$$

Case 3. $A=0$, $C\neq0$: We set $C=-1$. Then

$$\theta(t)=(t-t_0)^{-1},$$
$$\sigma(t)=B(t-t_0)^{-2}+c_5(t-t_0)^{-1}+c_6\,.$$

Setting $t_0 = 1$, $c_5 = 0$, and $c_6 = 0$, we obtain

$$u(x,t) = t^{-1}w(z) + \frac{x}{t} + \frac{2B}{t^2}, \quad z = \frac{x}{t} + \frac{B}{t^2}. \tag{B13}$$

Case 4. $A\neq 0$, $C\neq 0$: We set $A^2 = -1$, $B = 0$, $C^2 = 1$. Then

$$\theta(t) = (t^2 \pm 1)^{-1/2}, \quad \sigma(t) = c_7 t + c_8(t^2 \pm 1)^{-1/2}.$$

Setting $c_8 = 0$, we obtain

$$u(x,t) = (t^2 \pm 1)^{-1/2}w(z) + \frac{xt - c_7}{t^2 \pm 1}, \quad z = \frac{x + c_7 t}{t^2 \pm 1}. \tag{B14}$$

The infinitesimals for Burgers' equation obtained using the classical Lie method are

$$X = \alpha x + \beta t + \gamma x t + \delta, \tag{B15a}$$

$$T = 2\alpha t + \gamma t^2 + \kappa, \tag{B15b}$$

$$U = -\alpha u + \gamma(x - tu) + \beta, \tag{B15c}$$

with α, β, γ, δ, and κ arbitrary constants (cf. Ref. 49). It is easily shown that all the similarity reductions obtained by our method (B11)–(B14) for Bergers' equation (B1) can also be obtained from these infinitesimals (cf. Ref. 49).

APPENDIX C: KORTEWEG–dE VRIES EQUATION

In this appendix we outline how to determine the similarity reductions of the Korteweg–de Vries equation

$$u_t + uu_x + u_{xxx} = 0, \tag{C1}$$

using the method developed in this paper. It suffices to assume the special form

$$u(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)). \tag{C2}$$

Substituting and collecting coefficients yields

$$\begin{aligned}&\beta z_x^3 w''' + (3\beta_x z_x^2 + 3\beta z_x z_{xx})w'' \\ &\quad + (3\beta_{xx} z_x + 3\beta_x z_{xx} + \beta z_{xxx} + \beta z_t + \alpha\beta z_x)w' \\ &\quad + (\beta_{xxx} + \beta_t + \alpha\beta_x + \alpha_x\beta)w + \beta^2 z_x ww' + \beta\beta_x w^2 \\ &\quad + \alpha_{xxx} + \alpha_t + \alpha\alpha_x = 0.\end{aligned} \tag{C3}$$

We use the coefficient of w''' as the normalizing coefficient. For this to be an ordinary differential equation, from the coefficient of ww' we get

$$\beta z_x^3 \Gamma(z) = \beta^2 z_x,$$

where $\Gamma(z)$ is to be determined. Using the freedom in Remark 3(i) in Sec. III,

$$\beta = z_x^2. \tag{C4}$$

The coefficient of w^2 gives

$$\beta z_x^3 \Gamma(z) = \beta\beta_x$$

where $\Gamma(z)$ is to be determined. Using (C4), integrating twice, and using the freedoms in Remarks 2 and 3(iii), we have

$$z = x\theta(t) + \sigma(t), \quad \beta = \theta^2(t), \tag{C5}$$

where $\theta(t)$ and $\sigma(t)$ are to be determined. Equation (C3) simplifies to

$$\begin{aligned}&\theta^5(w''' + ww') + \theta^2\left\{\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right) + \alpha\theta\right\}w' \\ &\quad + \left\{2\theta\frac{d\theta}{dt} + \alpha_x\theta^2\right\}w + \alpha_{xxx} + \alpha_t + \alpha\alpha_x = 0.\end{aligned} \tag{C6}$$

The conditions for this to be an ordinary differential equation give successively, from the coefficients of w', w, and 1,

$$\alpha = -\frac{1}{\theta}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right), \tag{C7}$$

$$\frac{d\theta}{dt} = A\theta^3, \tag{C8}$$

$$\theta\frac{d^2\sigma}{dt^2} - 2\frac{d\theta}{dt}\frac{d\sigma}{dt} = 2\theta^6(A^2\sigma + B), \tag{C9}$$

with B another arbitrary constant.

Therefore the general similarity reduction of the Kortweg–de Vries equation (C1) is

$$u(x,t) = \theta^2(t)w(z) - \frac{1}{\theta}\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right),$$

$$z = x\theta(t) + \sigma(t),$$

where $\theta(t)$ and $\sigma(t)$ satisfy (C8) and (C9).

There are two cases to consider.

Case 1. $A\neq 0$: We set $A = -\frac{1}{3}$, $B = 0$. Thus

$$\theta = (t - t_0)^{-1/3}, \quad \sigma(t) = c_1(t - t_0)^{2/3} + c_2(t - t_0)^{-1/3}.$$

We set $t_0 = 0$, $c_2 = 0$ and obtain the similarity reduction

$$u(x,t) = t^{-2/3}w(z) + \frac{x}{3t} - \frac{2}{3}c_1, \quad z = \frac{x + c_1 t}{t^{1/3}}. \tag{C10}$$

Case 2. $A = 0$. We set $\theta = 1$, and then

$$\sigma(t) = Bt^2 + c_3 t + c_4.$$

Now set $c_4 = 0$ and obtain the similarity reduction

$$u(x,t) = w(z) - 2Bt - c_3, \quad z = x + Bt^2 + c_1 t. \tag{C10$'$}$$

The infinitesimals for the Korteweg–de Vries equation obtained using the classical Lie method are

$$X = \alpha x + \beta t + \gamma, \quad T = 3\alpha t + \delta, \quad U = -2\alpha u + \beta, \tag{C11}$$

with α, β, γ, and δ arbitrary constants (cf. Ref. 16, p. 129, and Refs. 41–43). It is easily shown that both the similarity reductions (C10) and (C11) for the Korteweg–de Vries equation (C1) can be obtained from these infinitesimals (cf. Ref. 16, p. 196, and Ref. 43).

APPENDIX D: MODIFIED KORTEWEG–dE VRIES EQUATION

In this appendix we outline how to determine similarity reductions of the modified Korteweg–de Vries equation

$$u_t + u^2 u_x + u_{xxx} = 0, \tag{D1}$$

using the method developed in this paper. It suffices to assume

$$u(x,t) = \alpha(x,t) + \beta(x,t)w(z(x,t)). \tag{D2}$$

Substituting and collecting coefficients yields

$$\beta z_x^3 w''' + (3\beta_x z_x^2 + 3\beta z_x z_{xx})w'' + (3\beta_{xx} z_x + 3\beta_x z_{xx} + \beta z_{xxx} + \beta z_t + \alpha^2\beta z_x)w' + (\beta_{xxx} + \beta_t + \alpha^2\beta_x + 2\alpha\alpha_x\beta)w + \beta^3 z_x w^2 w' + \beta^2\beta_x w^3 + 2\alpha\beta^2 z_x ww' + (2\alpha\beta\beta_x + \alpha_x\beta^2)w^2 + \alpha_{xxx} + \alpha_t + \alpha\alpha_x = 0 . \tag{D3}$$

We use the coefficient of w''' as the normalizing coefficient. For this to be an ordinary differential equation for $w(z)$, from the coefficient of w^2w' we get

$$\beta z_x^3 \Gamma(z) = \beta^3 z_x ,$$

where $\Gamma(z)$ is to be determined. Using the freedom in Remark 3(i) of Sec. III,

$$\beta = z_x . \tag{D4}$$

The coefficient of w^3 gives

$$\beta z_x^3 \Gamma(z) = \beta^2 \beta_x ,$$

where $\Gamma(z)$ is to be determined. Using (D4), integrating twice, and using the freedoms in Remarks 2 and 3(iii), we have

$$z = x\theta(t) + \sigma(t), \quad \beta = \theta(t) , \tag{D5}$$

where $\theta(t)$ and $\sigma(t)$ are to be determined. The coefficient of ww' gives

$$\beta z_x^3 \Gamma(z) = 2\alpha\beta^2 z_x ,$$

where $\Gamma(z)$ is to be determined. Using (D4) and the freedom in Remark 3(i), we have

$$\alpha \equiv 0. \tag{D6}$$

Equation (D3) simplies to

$$\theta^4(w''' + ww') + \theta\left(x\frac{d\theta}{dt} + \frac{d\sigma}{dt}\right)w' + \frac{d\theta}{dt}w = 0 . \tag{D7}$$

This is an ordinary differential equation for $w(z)$ provided that

$$\frac{d\theta}{dt} = A\theta^4 , \tag{D8a}$$

$$\frac{d\sigma}{dt} = \theta^3(A\sigma + B) , \tag{D8b}$$

where A and B are arbitrary constants.

Therefore the general similarity reduction of the modified Kortweg–de Vries equation is

$$u(x,t) = \theta(t)w(z), \quad z = x\theta(t) + \sigma(t) ,$$

where $\theta(t)$ and $\sigma(t)$ satisfy Eqs. (D8).

There are two cases to consider.

Case 1. $A \neq 0$: We set $A = -\frac{1}{3}$, $B = 0$. Hence

$$\theta(t) = (t - t_0)^{-1/3}, \quad \sigma(t) = c_1(t - t_0)^{-1/3} . \tag{D9}$$

Setting $t_0 = 0$, $c_1 = 0$, we obtain the similarity reduction

$$u(x,t) = t^{-1/3}w(z), \quad z = xt^{-1/3} . \tag{D10}$$

Case 2. $A = 0$: Solving (D8),

$$\theta(t) = c_2, \quad \sigma(t) = Bt + c_3 .$$

Setting $c_2 = 1$, $c_3 = 0$, we obtain the similarity reduction

$$u(x,t) = w(z), \quad z = x + Bt. \tag{D11}$$

The infinitesimals for the modified Korteweg–de Vries equation obtained using the classical Lie method are

$$X = \alpha x + \beta, \quad T = 3\alpha t + \gamma, \quad U = -2\alpha u , \tag{D12}$$

with α, β, and γ arbitrary constants (cf. Ref. 42). It is easily shown that both the similarity reductions (D10) and (D11) for the modified Korteweg–de Vries equation (D1) can be obtained from these infinitesimals (cf. 42).

[1] J. Boussinesq, Comptes Rendus **72**, 755 (1871); J. Math. Pures Appl. **7**, 55 (1872).
[2] F. Ursell, Proc. Camb. Philos. Soc. **49**, 685 (1953).
[3] N. J. Zabusky, in *Nonlinear Partial Differential Equations*, edited by W. F. Ames (Academic, New York, 1967), pp. 233–258.
[4] M. Toda, Phys. Rep. **18**, 1 (1975).
[5] V. E. Zakharov, Sov. Phys. JETP **38**, 108 (1974).
[6] A. C. Scott, in *Bäcklund Transformations, Lecture Notes in Mathematics*, Vol. **515**, edited by R. M. Miura (Springer, Berlin, 1975), pp. 80–105.
[7] M. J. Ablowitz and R. Haberman, J. Math. Phys. **16**, 2301 (1975); P. J. Caudrey, Phys. Lett. A **79**, 264 (1980); Physica D **6**, 51 (1982); P. Deift, C. Tomei, and E. Trubowitz, Commun. Pure Appl. Math. **35**, 567 (1982).
[8] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[9] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981); F. Calogero and A. Degasperis, *Spectral Transform and Solitons. I* (North-Holland, Amsterdam, 1982).
[10] S. Lie, *Vorlesungen über Differentialgleichungen mit Bekannten Infinitesimalen Transformationen* (Teuber, Leipzig, 1891) (reprinted by Chelsea, New York, 1967).
[11] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations, Applied Mathematical Sciences*, Vol. 13 (Springer, Berlin, 1974).
[12] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[13] P. J. Olver, *Applications of Lie Groups to Differential Equations, Graduate Texts in Mathematics*, Vol. 107 (Springer, New York, 1986).
[14] P. Winternitz, in *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, Berlin, 1983), pp. 263–331.
[15] P. Rosenau and J. Schwarzmeier, Courant Institute Report, COO-3077-160, MF-94, 1979; B. Champagne and P. Winternitz, preprint CRM-1278, Montreal, 1985.
[16] F. Schwarz, Comput. **34**, 91 (1985).
[17] F. Schwarz, SIAM Rev. **30**, 450 (1988).
[18] G. W. Bluman and J. D. Cole, J. Math. Mech. **18**, 1025 (1969).
[19] P. J. Olver and P. Rosenau, Phys. Lett. A **114**, 107 (1986); SIAM J. Appl. Math. **47**, 263 (1987).
[20] E. Noether, Nachr. König. Gesell. Wissen Göttingen, Math. Phys. **K1**, 235 (1918) [Transport Theory Stat. Phys. **1**, 186 (1971).
[21] R. L. Anderson and N. H. Ibragimov, *Lie-Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979); N. H. Ibragimov, *Transformation Group Applied to Mathematical Physics* (Reidel, Boston, 1985).
[22] G. W. Bluman, S. Kumei, and G. J. Reid, J. Math. Phys. **29**, 806 (1988).
[23] Since this was written, D. Levi and P. Winternitz [J. Phys. A: Math. Gen. **22**, 2915 (1989)] have provided such an explanation by showing all these similarity solutions of the Boussinesq equation can be obtained using the nonclassical method of Bluman and Cole.[18]
[24] T. Nishitani and M. Tajiri, Phys. Lett. A **89**, 379 (1982).
[25] P. Rosenau and J. L. Schwarzmeier, Phys. Lett. A **115**, 75 (1986).
[26] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[27] M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).
[28] M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978); J. Math. Phys. **21**, 715 (1980); J. B. McLeod and P. J. Olver, SIAM J. Math. Anal. **14**, 488 (1983).
[29] G. R. W. Quispel, F. W. Nijhoff, and H. W. Capel, Phys. Lett. A **91**, 143 (1982).
[30] A. S. Fokas and M. J. Ablowitz, Phys. Rev. Lett. **47**, 1096 (1981); J. Math. Phys. **23**, 2033 (1982).
[31] W. F. Ames, *Nonlinear Partial Differential Equations in Engineering. Vol. II* (Academic, New York, 1972); J. M. Hill, *Solutions of Differential Equations by Means of One-parameter Groups, Research Notes in Mathematics*, Vol. 63 (Pitman, London, 1982).
[32] E. E. Whittaker and G. M. Watson, Modern Analysis (Cambridge U.P., Cambridge, 1927), 4th ed.
[33] M. D. Kruskal, in *Dynamical Systems, Lecture Notes in Physics*, Vol. 38, edited by J. Moser (Springer, Berlin, 1975), 310–354.

[34]M. Wadati, K. Konno, and Y. H. Ichikawa, J. Phys. Soc. Jpn. **47**, 1698 (1979).
[35]D. Levi, O. Ragnisco, and A. Sym, Phys. Lett. A **100**, 7 (1984); S. Kawamoto, J. Phys. Soc. Jpn. **54**, 2055 (1985); P. A. Clarkson, A. S. Fokas, and M. J. Ablowitz, SIAM J. Appl. Math. **49**, 1188 (1989).
[36]B. H. Gilding, J. Hydrol. **56**, 251 (1982).
[37]W. I. Fushchlich and A. G. Nikitin, *Symmetries of Maxwell's Equations* (D. Reidel, Dordrecht, 1987); W. I. Fushchlich and N. I. Serov, J. Phys. A: Math. Gen. **16**, 3645 (1983); **20**, L929 (1987); W. I. Fushchlich and W. M. Shtelen, *ibid.* **16**, 271 (1983); Phys. Lett. B **128**, 215 (1983); W. I. Fushchlich and I. M. Tsifra, J. Phys. A: Math. Gen. **20**, L45 (1987); W. I. Fushchlich and R. Z. Zhdanov, *ibid.* **21**, L5 (1988); Phys. Rep. **172**, 123 (1989).
[38]E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950); J. D. Cole, Quart. Appl. Math. **9**, 225 (1951).
[39]M. Wadati, J. Phys. Soc. Jpn. **32**, 1681 (1972); M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[40]G. Baumann and T. F. Nonnenmacher, J. Math. Phys. **28**, 1250 (1987); B. Champagne and P. Winternitz, *ibid.* **29**, 1 (1988); D. David, N. Kamran, D. Levi, and P. Winternitz, Phys. Rev. Lett. **55**, 2111 (1985); J. Math. Phys. **27**, 1225 (1986); B. Dorizzi, B. Grammaticos, A. Ramani, and P. Winternitz, *ibid.* **27**, 2848 (1986); L. Gagnon, B. Grammaticos, A. Ramani, and P. Winternitz, J. Phys. A: Math. Gen. **22**, 499 (1989); L. Gagnon and P. Winternitz, *ibid.* **21**, 1493 (1988); **22**, 469 (1989); Phys. Lett. A **134**, 276 (1989); Phys. Rev. A **39**, 296 (1989); A. M. Grundland, J. Harnard, and P. Winternitz, J. Math. Phys. **25**, 791 (1984); A. M. Grundland and J. A. Tuszyński, J. Phys. A: Math. Gen. **20**, 6243 (1987); A. M. Grundland, J. A. Tuszyński, and P. Winternitz, Phys. Lett. A **119**, 340 (1977); R. A. Leo, L. Martina, and G. Soliani, J. Math. Phys. **27**, 2623 (1986); R. A. Leo, L. Martina, G. Soliani, and G. Tondo, Prog. Theor. Phys. **76**, 739 (1986); R. A. Leo and G. Soliani, Nuovo Cimento B **96**, 89 (1986); Y. Matsuno, J. Math. Phys. **28**, 2317 (1987); A. Oron and P. Rosenau, Phys. Lett. A **118**, 172 (1986); M. Skierski, A. M. Grundland, and J. A. Tuszyński, Phys. Lett. **133**, 213 (1988); P. Winternitz, A. M. Grundland, and J. A. Tuszyński, J. Math. Phys. **28**, 2194 (1987); J. Phys. C: Solid State Phys. **21**, 4931 (1988).
[41]S. Kawamoto, J. Phys. Soc. Jpn. **52**, 4059 (1983).
[42]M. Lakshmanan and P. Kaliappan, J. Math. Phys. **24**, 795 (1983).
[43]M. Tajiri and S. Kawamoto, J. Phys. Soc. Jpn. **51**, 1678 (1982).
[44]R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[45]R. Hirota and J. Satsuma, Prog. Theor. Phys. **57**, 797 (1977).
[46]V. I. Gromak, Differ. Eqs. **23**, 506 (1987); [Diff. Urav. **23**, 760 (1988)].
[47]J. Chazy, Acta Math. **34**, 317 (1911).
[48]Q. R. W. Quispel and H. W. Capel, Physica A **117**, 76 (1983).
[49]M. Tajiri, S. Kawamoto, and K. Thushima, Math. Japon. **28**, 125 (1983).

Integrability and orbits in quartic polynomial potentials

Paul W. Cleary
Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia

(Received 8 November 1988; accepted for publication 7 June 1989)

The effect of enlarging the class of admissible movable singularities in Painlevé analysis to include all rational algebraic branch points is examined for a class of quartic polynomial potentials. Eight homogeneous quartic potentials are found in addition to the seven integrable cases given by weak Painlevé analysis. They are examined using various numerical techniques and Ziglin's theorem. Only one of them remains a candidate for integrability, which indicates that, in general, most rational algebraic branch points are incompatible with integrability. Movable logarithmic singularities also appear to be inconsistent with integrability. However, the remaining potentials are either regular or nearly regular for the energies examined, and are therefore still of interest for numerical purposes. The corresponding surfaces of section are found to be particularly simple in structure and belong to a small number of topologically distinct classes. Their stable and unstable periodic orbit structures are examined to provide information about their regularity and for use in Ziglin's theorem. There appears to be a correlation between the resonances of the stable periodic orbits and the order of the corresponding movable singularities.

I. INTRODUCTION

Integrable systems are the exception and not the rule. Most nonlinear systems possess chaotic regions of various sizes in their phase spaces, indicating the absence of a full complement of global isolating integrals of the motion. The problem is how to locate these integrable systems.

Ablowitz *et al.*[1] suggested that if a partial differential equation was soluble by inverse scattering transform methods, then every ordinary differential equation arising as a similarity solution is of Painlevé type. They then presented an algorithm that allows the user to determine whether an ordinary differential equation possesses the Painlevé property, that is, its only movable singularities are poles. Chang *et al.*[2] and Segur[3] applied such an algorithm, which we call Painlevé analysis, to the Hénon–Heiles and Lorenz systems. They found that all the resulting systems were integrable. This lead to the idea, which we will refer to as the Painlevé conjecture, that a system possessing the Painlevé property will be integrable. This conjecture has been used successfully by a number of authors, including Bountis *et al.*,[4] to predict integrable cases of a wide range of systems. A generalization of the original conjecture by Ramani *et al.*[5] and Grammaticos *et al.*,[6] who allowed limited types of movable algebraic singularities as well as poles, was able to predict a number of additional integrable cases.

The essence of the Painlevé conjecture is that if the solutions of the equations of motion have a particular movable singularity structure, then there exist a full complement of isolating integrals of the motion and the system is integrable. The continued success of various versions of the conjecture suggests that some form of it is correct. The question is what types of movable singularities should be admitted? The original variation used by Chang *et al.*[2] allowed only movable poles, while Ramani *et al.*[5] allowed some movable branch points. Conversely, Yoshida[7] showed that movable complex and irrational algebraic branch points were not consistent with integrability. It is the remaining class of movable rational algebraic branch points, not covered by weak Painlevé analysis, that are of interest here.

In this paper we examine a collection of quartic polynomial potentials whose only movable singularities are rational algebraic branch points, using Ziglin's theorem[8] and explicit calculation of additional integrals of the motion. Their integrability or lack thereof can also be explored by various numerical techniques, such as the method of surface of section and an examination of their periodic orbit structures. Of particular interest are the invariant curves of unstable periodic orbits that indicate regular behavior for the system. It will be found that all the regular quartic potentials examined have common properties and are naturally ordered by the resonances exhibited by their stable periodic orbits.

II. PAINLEVÉ ANALYSIS OF THE VERHULST POTENTIALS

The quartic polynomial potentials of interest here were first studied by Verhulst.[9] He examined the existence and stability of bifurcations of orbit families, using averaging procedures and modified Birkoff transformations. The derivation of this discrete-symmetric quartic potential as a truncated Taylor series appears in Verhulst.[9] The Verhulst potentials are of the form

$$V(x,z) = \frac{1}{2}(\omega_1^2 x^2 + \omega_2^2 z^2) - \left(\frac{A_1}{3}x^3 + A_2 x z^2\right) - \left(\frac{B_1}{4}x^4 + \frac{B_2}{2}x^2 z^2 + \frac{B_3}{4}z^4\right), \qquad (1)$$

where the coefficients ω_1^2, ω_2^2, A_1, A_2, B_1, B_2, and B_3 are all real. We will restrict ourselves to potentials with B_1, B_2, and $B_3 \leqslant 0$. The equations of motion are

$$\ddot{x} + \omega_1^2 x = A_1 x^2 + A_2 z^2 + B_1 x^3 + B_2 x z^2\,, \qquad (2a)$$

$$\ddot{z} + \omega_2^2 z = 2A_2 xz + B_2 x^2 z + B_3 z^3\,. \qquad (2b)$$

 0022-2488/89/102214-12$02.50

The homogeneous quartic potential

$$V(x,z) = \lambda_1 x^4 + 2x^2z^2 + \lambda_3 z^4\,, \tag{3}$$

is obtained from the full Verhulst potential (1) by setting all the lower-order terms to zero. We have defined $\lambda_1 = B_1/B_2$ and $\lambda_3 = B_3/B_2$.

A number of authors[2-4] have successfully applied standard Painlevé analysis to various systems, including the Hénon–Heiles and Lorenz systems, Toda lattices with various boundary conditions and homogeneous quartic potentials and found a range of integrable cases. They all required that the only movable singularities exhibited by the equations of motion be poles.

Ramani *et al.*[5] found an integrable quintic potential

$$V = x^5 + x^3y^2 + \tfrac{3}{16}xy^4\,, \tag{4}$$

with an independent second integral

$$I = \dot{y}(y\dot{x} - x\dot{y}) + \tfrac{1}{2}x^4y^2 + \tfrac{3}{8}x^2y^4 + \tfrac{1}{3}y^6\,.$$

Singularity analysis revealed a leading-order behavior of $(t-t_0)^{-2/3}$. This showed that integrable systems with more general singularity structures than poles do exist. This lead Ramani *et al.*[5] to propose a generalization, termed weak Painlevé analysis, allowing a very restricted type of rational algebraic branch point in addition to the previously allowed poles. The type of singularity allowed is closely related to the degree of the polynomial potential being examined. Grammaticos *et al.*[6] clarified this by saying that a polynomial potential had the weak Painlevé property if it had expansions in terms of

$$(t-t_0)^{1/r}\,,$$

where

$$r = \begin{cases} p, & \text{for } p \text{ odd,} \\ p/2, & \text{for } p \text{ even,} \end{cases}$$

where the polynomial potential has degree $p+2$. So for the quintic potential (4), with $p=3$, they looked for expansions in terms of $(t-t_0)^{-1/3}$, as well as those in terms of $(t-t_0)^{-1}$. For any polynomial potential there are always two types of singularity expansion to be examined. The first are poles and the second are a very specific type of rational algebraic branch point whose order is related to the degree of the potential. It is, however, not obvious how to define this extra allowed exponent for nonpolynomial potentials.

Yoshida[7] showed that the existence of irrational and complex algebraic branch points is inconsistent with a system being integrable. This leaves systems whose movable singularity structures belong to the vast class of all rational algebraic branch points. The single extra expansion allowed by weak Painlevé analysis is but a single member of this infinite class of possible expansions. Solutions with these singularity types have not been examined before and it is one of the purposes of this paper to determine which, if any, of these rational algebraic branch points, in addition to the two allowed by weak Painlevé analysis, are consistent with integrability for the class of quartic potentials above.

The standard Painlevé analysis algorithm presented by Ablowitz *et al.*[1] remains unchanged for all the variations and consists of the three following parts.

(a) Finding the dominant behavior of the solution near any movable singularity. This is accomplished by substituting the expressions

$$x = a(t-t_0)^p, \quad z = b(t-t_0)^q$$

into Eqs. (2) and finding all values of p and q for which the leading-order terms balance.

(b) Finding the resonances: These indicate which terms in the Laurent series expansions of the solutions have arbitrary coefficients. The expressions

$$x = a\Delta t^p + \alpha \Delta t^{p+r}, \quad z = b\Delta t^q + \beta\Delta t^{q+r}\,,$$

where $\Delta t = t - t_0$ are substituted into Eqs. (2) and the corresponding values of r are calculated for each pair of p and q values.

(c) Determining the constants of integration involves calculating all the coefficients in the Laurent series expansions occurring up to the final resonance. This indicates the existence or otherwise of logarithmic branch points. In the original application of the conjecture, Chang *et al.*[2] suggested that the existence of logarithmic branch points was inconsistent with integrability.

In standard Painlevé analysis only poles were allowed. All the leading-order powers p and q and resonances r were required to be integers. In weak Painlevé analysis they are allowed to be integer or integer multiples of $\frac{1}{2}$. Now we wish to allow the movable singularities to be any type of rational algebraic branch point. The leading-order powers p and q and the resonances r then take all rational values.

For Eqs. (2) there are five cases of balancing. The expressions for p, q, and the corresponding resonances and existence criterion are given in Table I for each of these cases. All the leading-order powers and resonances for all five cases are rational if the expressions

$$\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1+8/\lambda_1}, \quad \tfrac{1}{2} + \tfrac{1}{2}\sqrt{1+8/\lambda_3}, \quad \tfrac{3}{2} + \tfrac{1}{2}\sqrt{25+16\gamma} \tag{5}$$

are all rational numbers of the form m_1/n, m_2/n, and m_3/n, respectively, where m_1, m_2, m_3, and n are all integers and $\gamma = (2-\lambda_1-\lambda_3)/(\lambda_1\lambda_3 - 1)$. These three conditions arise from q from case 2b, p from case 3a, and the third resonance r_3 from case 1, respectively. We will define the order of each potential to be the integral denominator n. The Laurent expansion of the solution is then written in terms of $(t-t_0)^{1/n}$ and possesses a finitely branched movable singularity of order n. In weak Painlevé analysis only $n=2$ is allowed.

To determine the values of λ_1 and λ_3 that satisfy the rationality conditions (5), we choose a value of n, beginning with $n=1$, and systematically use all appropriate values of m_1, m_2, and m_3 and attempt to solve for λ_1 and λ_3. In most cases there is no solution. The first 14 sets of coefficients λ_1 and λ_3, with the lowest values of n satisfying these conditions, are given in Tables II and III. Those in Table III correspond to homogeneous quartic potentials of the form

$$V_n(x,z) = \lambda_1(n)x^4 + 2x^2z^2 + \lambda_3(n)z^4\,, \tag{6}$$

which we call $V_1,\ldots,V_{10}$. Of these, only V_1 can be found using weak Painlevé analysis. None of the potentials in Tables II or III possess any movable irrational or complex algebraic

TABLE I. The leading-order powers p and q and the resonances are given for each of the five possible cases of the leading-order terms balancing. Case 2b and 3b exist only under restricted circumstances. The last column of this table gives the number of arbitrary parameters in the resulting series solutions.

Case	p	q	r_1	r_2	r_3	r_4	Existence conditions	Number of parameters in solution
1	-1	-1	-1	$\frac{3}{2} \pm \frac{1}{2}\sqrt{25+16\gamma}$		4		4 or 3
2a	-1	$\frac{1}{2}+\frac{1}{2}\sqrt{1+8/\lambda_1}$	-1	0	4	$\cdots$		3
2b	-1	$\frac{1}{2}-\frac{1}{2}\sqrt{1+8/\lambda_1}$	-1	0	$\sqrt{1+8/\lambda_1}$	4	$1/\lambda_1 < 1$	4
3a	$\frac{1}{2}+\frac{1}{2}\sqrt{1+8/\lambda_3}$	-1	-1	0	4	$\cdots$		3
3b	$\frac{1}{2}-\frac{1}{2}\sqrt{1+8/\lambda_3}$	-1	-1	0	$\sqrt{1+8/\lambda_3}$	4	$1/\lambda_3 < 1$	4

branch points. There are more potentials with higher order ($n > 19$) rational algebraic singularities, which will not be examined here.

For each of the potentials in Table II it is necessary to determine whether or not it possesses any logarithmic singularities by calculating the constants of integration. For convenience we set $\omega_1^2 = 1$ and $\omega_2^2 = \omega^2$ by scaling the potential. The restrictions in Table II all arise from the requirement that there be no movable logarithmic singularities. Any potential with these λ_1 and λ_3 values, which does not satisfy the corresponding restrictions possesses logarithmic singularities. The corresponding homogeneous versions of these potentials had previously been found by Bountis *et al.*[4] and Ramani *et al.*[5] using standard and weak Painlevé analysis, respectively. Integrals of the motion for the full potentials in Table II will be given in the next section. Integrals were only found for those subclasses whose solutions contained no logarithmic singularities. This will be discussed in more detail later. Numerical surface of section calculations for these potentials are described in Sec. IV. Again, any potential possessing logarithmic singularities was found to have significant chaotic regions and was therefore nonintegrable.

Performing the logarithmic analysis on the ten homogeneous potentials, given in Table III, we find that V_1 and V_3 both possess movable logarithmic singularities. The surfaces of section for both potentials, discussed in Sec. IV, contain large chaotic regions and are nonintegrable. The existence of logarithmic singularities appears to be incompatible with integrability for these types of potentials. This may not be universally true for all systems. There certainly exist integrable one-dimensional systems with movable logarithmic singularities. The only movable singularities possessed by the remaining eight potentials in Table III are rational algebraic branch points of order $n > 2$.

TABLE II. Seven families of potentials obtained by using weak Painlevé analysis. The first three potentials have movable poles. The others have Laurent series expansions in $(t-t_0)^{1/2}$.

Potential	λ_1	λ_3	Restrictions	Resonance
U_1	1	1	$A_1 = 0, \quad A_2 = 0$	1:1
U_2	1	1	$A_1 = 3A_2, \quad B_2 = -2A_2^2$	1:1
U_3	$\frac{1}{3}$	$\frac{1}{3}$	$\omega^2 = 1, \quad A_1 = A_2$	1:1
U_4	$\frac{8}{3}$	$\frac{1}{3}$	$A_1 = 8A_2, \quad \omega^2 = \frac{1}{4} + A_2^2/B_2$	1:1
U_5	$\frac{8}{3}$	$\frac{1}{6}$	$\omega^2 = \frac{1}{4} + \frac{1}{2}A_2/B_2(A_1 - 6A_2)$	2:1
U_6	$\frac{1}{3}$	$\frac{8}{3}$	$\omega^2 = 4 + A_2/B_2(8A_1 - 5A_2)$	1:1
U_7	$\frac{1}{6}$	$\frac{8}{3}$	$\omega^2 = 4 + A_2/B_2(8A_1 - 3A_2)$	2:1

III. INTEGRALS OF THE MOTION

In Sec. II we found a series of potentials whose only movable singularities were rational algebraic branch points or poles. Ultimately, the only way to prove that a potential is integrable is to find all the required integrals of the motion. For two-dimensional Hamiltonian systems it is necessary to find a second integral independent of the Hamiltonian. In this section we give second integrals, which are quadratic in the velocities, for the families of potentials in Table II. To find integrals of the motion by direct methods one assumes a particular form for the integral and then determines conditions on all the coefficients of the velocity terms by requiring the Poisson bracket $\{H,G\}$ to vanish. This gives a collection of PDE's that must be solved for the coefficients. Hietarinta[10] discusses the subject of finding such additional integrals of the motion by direct methods and summaries all the previous results in an extensive review article.

The potential U_1 is integrable since it possesses a second integral of the form

TABLE III. Pairs of coefficients λ_1 and λ_3 for which the corresponding homogeneous quartic potential $V_n(x,z) = \lambda_1 x^4 + 2x^2z^2 + \lambda_3 z^4$ satisfies the first two steps of the generalization of Painlevé analysis. The value of q is the power of the first term in the Laurent series expansion of the solution in each case. The resonances of the stable nonaxial periodic orbits are also given. Note that μ_3 is the integrability coefficient for the inclined straight line periodic orbits. This is used in the application of Ziglin's theorem.

Potential	q	λ_1	λ_3	Resonance	μ_3
V_1	$-\frac{1}{2}$	$\frac{8}{3}$	$\frac{1}{136}$	3:1	$4\frac{3}{8}$
V_2	$-\frac{1}{4}$	$\frac{32}{5}$	$\frac{1}{10}$	2:1	28
V_3	$-\frac{1}{4}$	$\frac{32}{5}$	$\frac{32}{1517}$	4:1 or 5:1	$13\frac{2}{9}$
V_4	$-\frac{2}{13}$	$11\frac{4}{15}$	$\frac{1}{15}$	3:1	78
V_5	$-\frac{2}{19}$	$17\frac{4}{21}$	$\frac{1}{21}$	4:1	171
V_6	$-\frac{1}{10}$	$18\frac{2}{11}$	$\frac{32}{725}$	4:1	167.32
V_7	$-\frac{1}{10}$	$18\frac{2}{11}$	$\frac{9}{275}$	4:1	83.08
V_8	$-\frac{1}{10}$	$18\frac{2}{11}$	$\frac{1}{136}$	8:1	$40\frac{3}{8}$
V_9	$-\frac{1}{13}$	$24\frac{1}{7}$	$\frac{1}{28}$	4:1	325
V_{10}	$-\frac{1}{17}$	$32\frac{1}{9}$	$\frac{1}{36}$	5:1	561

$$G = B_2(x\dot{z} - z\dot{x})^2 + 2(\omega^2 - 1)(\dot{z}^2 + \omega^2 z^2 - B_2 z^2(x^2 + z^2)/2) .$$

This integral has previously been given by Hietarinta.[11] Bountis *et al.*[4] suggested that potentials of this form only possessed the Painlevé property for $\omega^2 = 1$, for which the potential becomes trivially separable in polar coordinates with the angular momentum as the second integral. However, this class of potentials does possess the standard Painlevé property for all values of ω^2.

The potential U_2 possesses a previously unknown second integral of the form

$$\begin{aligned} G = {} & 4A_2^2(z\dot{x} - x\dot{z})^2 + 4A_2\dot{z}(z\dot{x} - x\dot{z}) \\ & - A_2 z^2[A_2(x^2 + z^2) - x] \\ & + (4\omega^2 - 1)[\dot{x}^2 + x^2 - A_2 x(2x^2 + z^2) \\ & + A_2^2 x^2(x^2 + z^2)] , \end{aligned}$$

and is therefore integrable. The potentials U_1 and U_2 are the only potentials with $\lambda_1 = \lambda_3 = 1$ that possess no movable logarithmic singularities. They are also the only potentials with $\lambda_1 = \lambda_3 = 1$ that possess quadratic second integrals. This potential cannot be obtained from U_1 by any linear transformation.

The potential U_3 is separable under the canonical transformation to new variables $x + z$ and $x - z$, given in Aizawa and Saitô.[12] A second integral of the motion independent of the Hamiltonian is then

$$G = \dot{x}\dot{z} + xz - A_2(x^2 z + z^3/3) - (B_2/3)xz(x^2 + z^2) .$$

This potential is therefore integrable. Bountis *et al.*[4] found a restricted version of this potential with $A_1 = A_2 = 0$, using the standard Painlevé analysis algorithm. The full potential can be obtained from the restricted Bountis *et al.* version by a translation in the x direction. Again, the only potentials with $\lambda_1 = \lambda_3 = \frac{1}{3}$, for which an additional integral was found, are those that possess no movable logarithmic singularities.

The potential U_5 is integrable. The required second integral is

$$\begin{aligned} G = {} & \dot{z}(x\dot{z} - z\dot{x}) + \frac{(6A_2 - A_1)}{2B_2}\Big[\dot{x}^2 + x^2 \\ & - \frac{2}{3}A_1 x^3 - \frac{4}{3}B_2 x^4\Big] \\ & - \frac{1}{2}(4A_2 - A_1)x^2 z^2 + \frac{A_2}{4}z^4 \\ & + \left[\frac{B_2}{6}(2x^2 + z^2) + \left(\omega^2 - \frac{1}{2}\right)\right]xz^2 . \end{aligned}$$

This potential can be simplified by a suitable translation in the x direction. It can be rewritten as a linear combination of the integrable homogeneous combinatorial potentials x, $4x^2 + z^2$, $2x^3 + xz^2$, and $16x^4 + 12x^2z^2 + z^4$, which are given in Hietarinta.[11] Such a global coordinate transformation does not affect the integrability of the system.

New quadratic second integrals were found for two subclasses of the U_7 family of potentials. The first is given by

$$\omega^2 = 4, \quad A_1 = A_2 = 0, \quad \lambda_1 = \tfrac{1}{6}, \quad \lambda_3 = \tfrac{8}{3} ,$$

with integral

$$G = \dot{x}(z\dot{x} - x\dot{z}) + x^2 z[(B_2/6)(x^2 + 2z^2) - 1] .$$

The second is given by

$$\omega^2 = 1, \quad 2A_1 = A_2, \quad 3B_2 = -A_2^2, \quad \lambda_1 = \tfrac{1}{6}, \quad \lambda_3 = \tfrac{8}{3} ,$$

with the corresponding integral

$$\begin{aligned} G = {} & B_2\dot{x}(z\dot{x} - x\dot{z}) - 2A_1\dot{x}\dot{z} - (B_2 x + 2A_1) \\ & \times\left[xz - \frac{A_1}{3}z(3x^2 + 2z^2) - \frac{B_2}{6}xz(x^2 + 2z^2)\right] . \end{aligned}$$

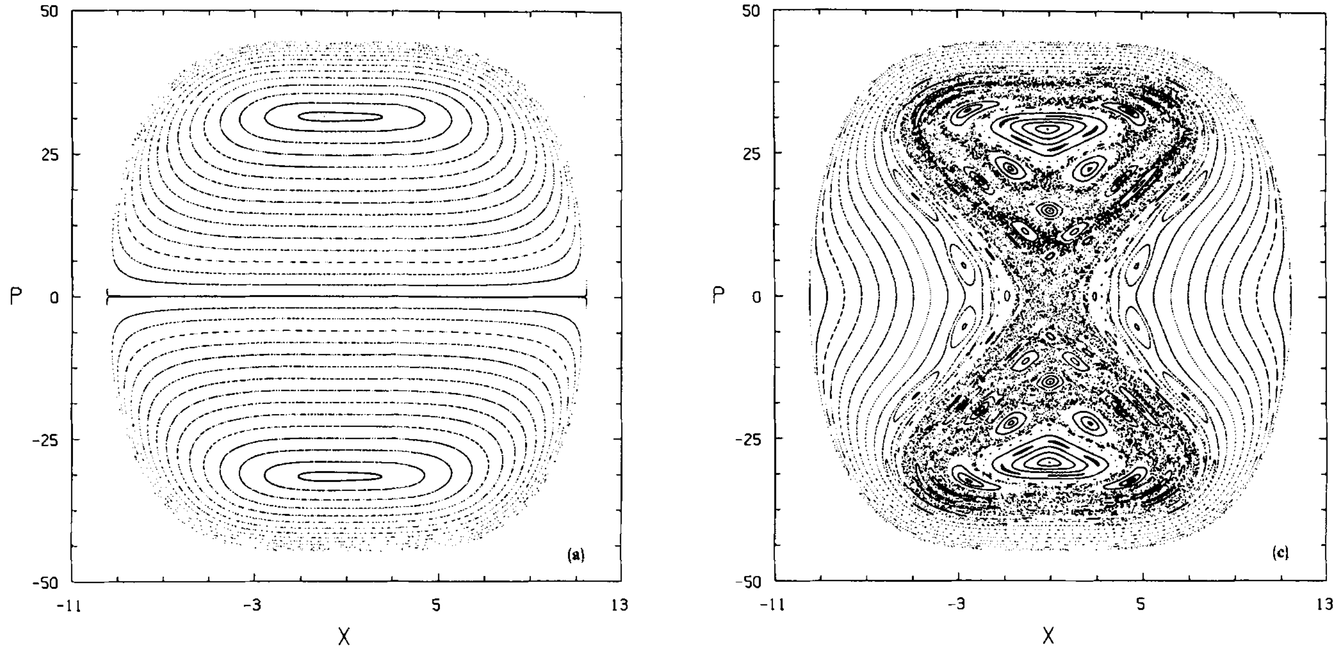
The above collection of potentials with quadratic second integrals is complete. Any other integrable Verhulst potential must have an integral that is quartic or higher order in the velocities. After appropriate translations all the above potentials and integrals belong to the four general classes of real potentials with quadratic second integrals given by Dorizzi *et al.*[13] They are presented here to verify that these potentials are in fact integrable.

The second integrals for the homogeneous potentials U_4 and U_6 are quartic in the velocities and are given by Ramani *et al.*[5] Therefore we do not expect the second integrals for the full U_4 and U_6 potentials to be quadratic in the velocities. The homogeneous U_5 and U_7 potentials, which are reflections of each other, have quadratic second integrals. However, for the full U_7 potential we only find two special cases possessing quadratic second integrals of the motion. This suggests that a general second integral is at least quartic in the velocities with two quadratic subcases for which the higher-order terms vanish. Four of the classes of potentials in Table II are integrable and some subset of each of the remaining three potentials are also integrable. The potentials U_4, U_5, and U_7 cannot be decomposed into linear combinations of known integrable homogeneous potentials, as was the case for U_5.

It should be noted that the same conditions arose from the requirement that the solutions possess no movable logarithmic singularities, as were found necessary for the existence of second integrals. No integrable potential had logarithmic singularities and every potential without logarithmic singularities had a second integral.

IV. SURFACES OF SECTION

The Verhulst potentials have four-dimensional phase spaces. Energy is conserved, so all orbits are constrained to lie on three-dimensional energy submanifolds. In order to study these systems, we lower the dimension of the problem to two by taking the intersection of the energy surface with the plane $z = 0$, simultaneously requiring $\dot{z} > 0$, to give a well-defined surface of section parametrized by x and $\dot{x}$.[14,15] Each time an orbit intersects the $z = 0$ plane with $\dot{z} > 0$, a point is placed in the surface of section. So instead of dealing with a four-dimensional phase space governed by a set of differential equations we have a two-dimensional surface of section with the Poincaré map determining the location of successive iterates. A periodic orbit appears in the surface of section as a finite collection of points. A quasiperiodic orbit, for which there is a locally conserved quantity, is represented in the surface of section by a smooth closed invariant curve



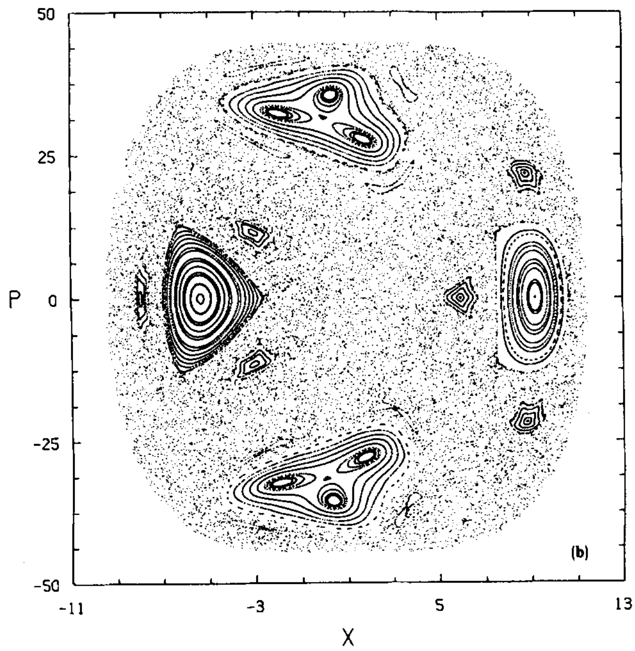
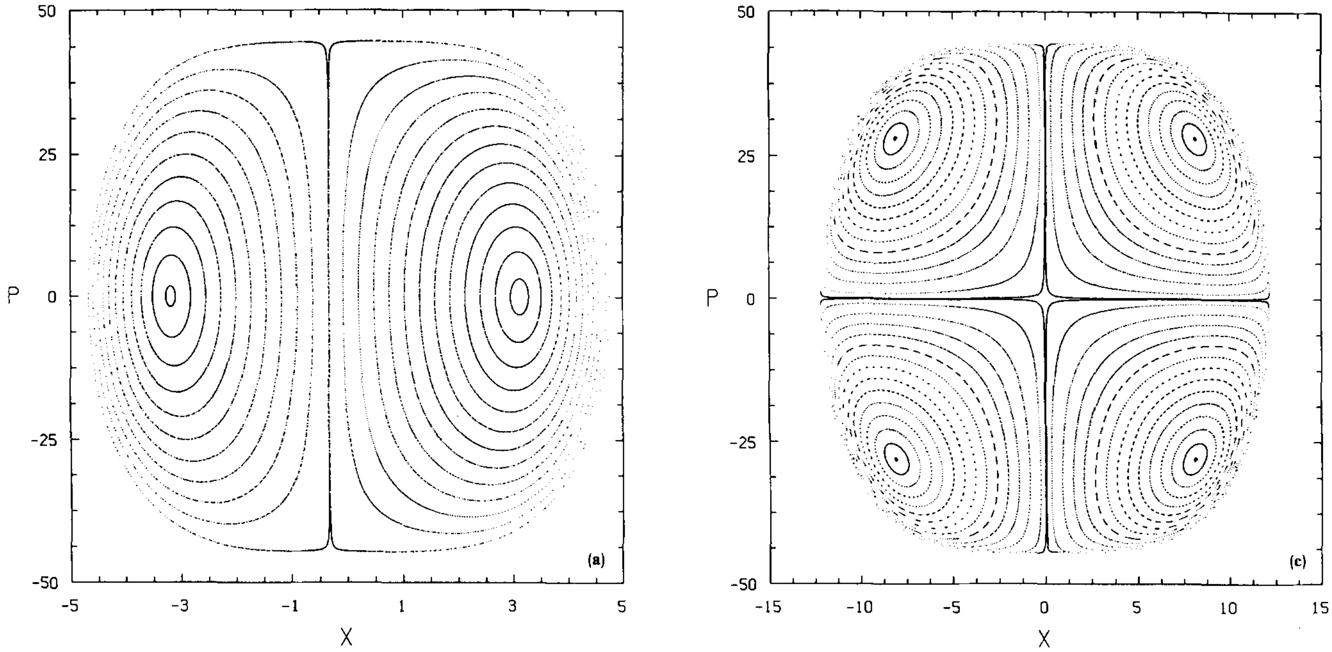


FIG. 1. (a) Surface of section for the U_3 potential with $\omega^2 = 1$ and $A_1 = 1$, $B_2 = -1$ for the energy $E = 1000$. (b) Surface of section for the perturbation of the U_3 potential with $\omega^2 = A_1 = A_2 = 1$, $B_1 = B_3 = -\frac{1}{3}$, and $B_2 = -\frac{1}{2}$ for the energy $E = 1000$. (c) Surface of section for the perturbation of the U_3 potential with $\omega^2 = A_1 = A_2 = 1$, $B_1 = -\frac{1}{3}$, and $B_2 = B_3 = -\frac{1}{2}$.

surrounding a fixed point. Chaotic orbits, for which only the energy is conserved, densely fill two-dimensional regions of the surface of section. We note here that the surface of section is always symmetric with respect to reflection in the x axis. We shall say that a potential is regular for the energy E if there are no chaotic regions in the corresponding surface of section.

The Hénon–Heiles system is obtained by setting $B_1 = B_2 = B_3 = 0$ in the general Verhulst potential (1). Hénon and Heiles[14] found that the existence of the second integral was dependent upon the energy. At low energies, such as $E = \frac{1}{12}$, the surface of section was regular. However, at higher energies, $E = \frac{1}{6}$, the invariant curves disintegrated producing large-scale chaos. For the quartic Verhulst potential there does not appear to be any such corresponding phenomenon. Even up to energies of the order of 10^3 or 10^4 the invariant curves do not break up. This apparent dependence of the second integral on the energy may be a consequence of the Hénon–Heiles potential not being bounded for all energies. All surfaces of section and all orbits in this paper will be calculated for $E = 1000$.

Surfaces of section were calculated for the potentials U_3, U_4, U_6, and U_7 to provide information about their integrability. The importance of the restrictions, given in Table II, to the integrability of each potential, was then investigated for each case by systematically breaking them, one at a time, and then in combinations, and examining the effects on the surface of section. Every perturbation to one of these potentials, which violated the conditions in Table II, possesses movable logarithmic singularities. The surface of section for every



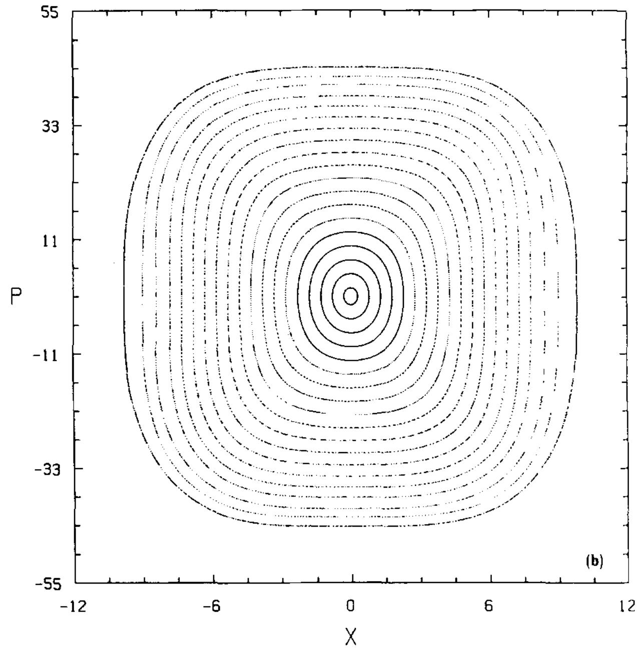


FIG. 2. (a) Surface of section for the U_5 potential with $\omega^2 = \frac{5}{12}$, $A_1 = 1$, $A_2 = \frac{1}{2}$ and $B_1 = -8$, $B_2 = -3$, and $B_3 = -1$. (b) Surface of section for the U_6 potential with $\omega^2 = 4$, $A_1 = A_2 = 0$ and $B_1 = -\frac{1}{3}$, $B_2 = -1$, and $B_3 = -\frac{8}{3}$. (c) Surface of section for the U_7 potential with $\omega^2 = 4$, $A_1 = A_2 = 0$ and $B_1 = -\frac{1}{6}$, $B_2 = -1$, and $B_3 = -\frac{8}{3}$.

one of these demonstrated nonintegrable behavior.

Figure 1(a) shows the surface of section for a typical U_3 potential. This structure is completely regular with no chaotic regions and no chains of islands. Figures 1(b) and 1(c) show surfaces of section for two different perturbations to the previous integrable potential and demonstrate the substantial effects that arise when breaking the restrictions given in Table II and introducing logarithmic singularities.

For the U_4 potentials, the condition $A_1 = 8A_2$ was found to be necessary to the integrability of the system, just as the condition $A_1 = A_2$ was for U_3. Any perturbation to the cubic terms always produced surfaces of section with chains of islands. Perturbations in λ_1 and λ_3 produced surfaces of section with substantial chaotic regions, regardless of whether the condition $A_1 = 8A_2$ was satisfied. Violating the condition on ω^2 gave surfaces of section with no noticeable chaotic regions, but with some chains of islands, indicating nonintegrability. All the surfaces of section belonging to the perturbations of the U_4 potentials always exhibited nonintegrable behavior.

Perturbations to the U_6 and U_7 potentials produced results similar to those obtained for U_3 and U_4 potentials. Figures 2(a), 2(b), and 2(c) are typical examples of the surfaces of section for the U_5, U_6, and U_7 potentials, respectively. Any potential not satisfying the restrictions in Table II has logarithmic singularities and always possessed features inconsistent with integrability. All the potentials examined, which satisfied the restrictions, did not have logarithmic singularities and looked numerically integrable. This supports the suggestion that the full U_4, U_6, and U_7

potentials are integrable. Their surfaces of section are all topologically very simple, unlike the surfaces of section produced by nonintegrable potentials, such as those shown in Figs. 1(b) and 1(c), whose solutions possess movable logarithmic singularities.

The potentials V_1 and V_3 both have substantial chaotic regions. Their surfaces of section are given in Figs. 3(a) and 3(b), respectively. They both possess movable logarithmic singularities in their solutions. Note that V_1 has the largest chaotic region and it has its logarithmic singularity in the fourth term of its Laurent series expansion; V_3 has the next largest and still significant chaotic region. Its logarithmic singularity is at sixth order. The size of the chaotic region appears to depend on the point at which the logarithmic singularities enter the Laurent series.

The other eight homogeneous potentials in Table III do not possess any movable logarithmic singularities. Any chaotic regions in their surfaces of section are all extremely small and surround hyperbolic points. Large-scale chaos was only found in conjunction with logarithmic singularities.

V. PERIODIC ORBIT STRUCTURE

It is useful to study the periodic orbits possessed by a potential. The application of Ziglin's theorem requires knowledge of the stability and resonances of the periodic orbits. The existence of rational invariant curves, consisting of an infinite number of unstable cycles and the absence of stable cycles, which generate island chains, provide important information about the regularity of the potential. Finally the stable periodic orbits form a framework for the entire dynamical structure of the phase space. They generate all the quasiperiodic orbits in state space or invariant curves in the surface of section.

Periodic orbits can be easily located numerically by determining the fixed points of the appropriate multiple of the Poincaré map $\mathbf{P}: S \to S$, where S is the surface of section. For $x \in S$, if $\mathbf{P}^{(n)}(x) = x$ then x is a fixed point of $\mathbf{P}^{(n)}$ or an n-cycle. The fixed point determination is carried out by a standard shooting method algorithm. The surface of section can be searched systematically for fixed points of a particular order n by performing the shooting method at each point on a grid of appropriate dimensions. The stability of a periodic orbit is determined by linear stability analysis. An orbit is stable if the two eigenvalues of the corresponding Jacobian matrix are complex conjugates of each other and of modulus one. If the eigenvalues are real, and one has an absolute value exceeding unity, then the orbit is unstable.[16]

For the purposes of classification, periodic orbits may be compared with Lissajous figures. These are obtained by orthogonal superposition of the harmonic oscillators,

$$x = a \sin(\omega_1 t + \phi_1), \quad z = b \sin(\omega_2 t + \phi_2) . \tag{7}$$

The resonance of an orbit is given by the ratio ω_1/ω_2 represented in the form $m{:}n$, where m and n are coprime integers. No significance will be attached to the order of the numbers m and n in the ratio. Here 1:1 resonance periodic orbits are the most common. The 1:1 resonance periodic orbits are of two types; *linear*—these straight line orbits can be either axial or inclined to the axes; *elliptic*—with major and minor axes perpendicular, but not necessarily coinciding with the x and z axes. Higher-order $m{:}n$ resonance periodic orbits will be classified as $m{:}n$ linear or elliptic periodic orbits by comparison with the 1:1 resonance periodic orbits. For example, Fig. 4(a) shows a 3:1 resonance linear orbit. This orbit has the property, as does the 1:1 resonance linear orbit, that the orbit touches the zero velocity curve (ZVC) at both ends. The particle proceeds along the orbit to an end point, stops, and then exactly retraces its path. There is no net angular momentum in any part in the orbit. In Fig. 4(b), we have a 5:1 resonance elliptic orbit, which again shares the properties of the 1:1 resonance analog, namely, that in each loop of the orbit there is a definite sense of rotation and a net angular momentum. These periodic orbits do not meet the ZVC anywhere. Linear orbits correspond to a phase difference $\phi_1 - \phi_2 = 0$, and elliptic orbits to a phase difference of $\phi_1 - \phi_2 = \pi/2$.

Consider any torus in the phase space that has a rational winding number. Any orbit on the torus is a closed curve and therefore periodic. There are an infinite number of these periodic orbits, all with the same period, densely covering the torus. All these periodic orbits are unstable. Taking the intersection of the torus with the plane $z = 0$ gives a corresponding smooth invariant curve in the surface of section.

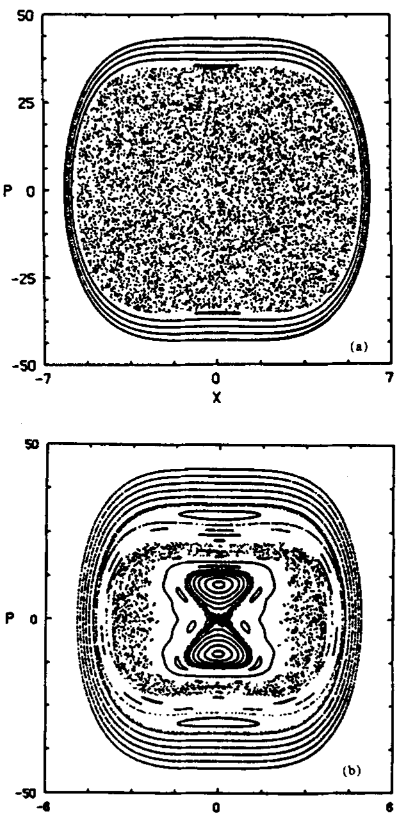


FIG. 3. (a) Surface of section for the homogeneous potential V_1. The large scale chaos is due to the existence of a logarithmic singularity at fourth order. (b) Surface of section for the homogeneous potential V_3. It has a logarithmic singularity at sixth order.

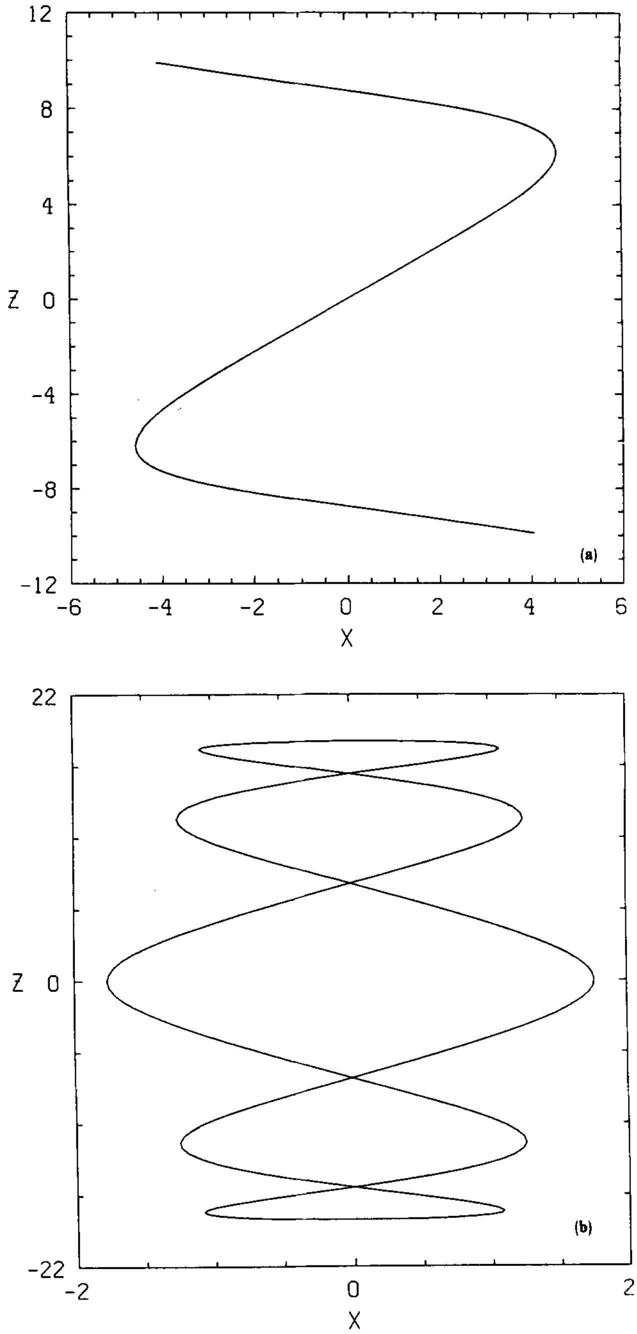


FIG. 4. (a) A 3:1 resonance linear periodic orbit. The stable nonaxial periodic orbits in the potential V_4 are of this type. (b) A 5:1 resonance elliptic periodic orbit. The stable nonaxial periodic orbits in the potential V_7 are of this type.

Such an invariant curve will be described as rational. Each of the periodic orbits on the torus produces an n-cycle in the surface of section, for some integer n. Therefore any rational invariant curve can be regarded as an infinite collection of unstable n-cycles. If a potential is integrable then there are an infinite number of rational invariant curves, since the winding number varies continuously from one invariant curve to the next. If an integrable potential is subjected to a nonintegrable perturbation, then some of these rational invariant curves disintegrate. When such an invariant curve of n-cycles disintegrates there remain only $2n$ points in the surface of section. These points form two n-cycles, one stable and one unstable. The points of the two cycles alternate, such that between any two points of the stable n-cycle is a point of the unstable n-cycle and similarly between any two points of the unstable n-cycle is a point of the stable n-cycle. The points belonging to the stable n-cycle are elliptic points and those belonging to the unstable n-cycle are hyperbolic points. The points on the stable n-cycle generate an island chain in the surface of section. For more details on elliptic and hyperbolic cycles see Hénon.[15]

If there exists a stable n-cycle, for a given potential, then there is a corresponding unstable cycle, whose points separate those of the stable one. This indicates that a rational invariant curve has disintegrated and that the potential is nonintegrable. If a systematic search reveals the existence of an invariant curve for each rational value of the winding number around each fixed point, then there are no stable cycles and the potential is regular at that energy. Therefore regularity can, in theory, be checked by calculating all the n-cycles in the surface of section and determining whether or not they form smooth invariant curves. Since it is not feasible to calculate all n-cycles for all values of n, checking whether cycles lie on invariant curves can be used as a guide to regularity when the surface of section does not have sufficient resolution to find the smaller island chains. Conversely, if there exists a single stable cycle then the potential is not integrable. The existence and location of periodic orbits can provide important information about the global integrability of the corresponding potentials.

All the potentials given in Table II, aside from U_5 and U_7, have 1:1 resonance periodic orbits. The remaining two have 2:1 resonance periodic orbits. The periodic orbits for the potentials given in Table III will now be explored. We shall determine the number and type of periodic orbits present in each potential and examine their stability. In each case there exist x- and z-axis orbits, defined as $z = \dot{z} = 0$ and $x = \dot{x} = 0$, respectively. These are always 1:1 resonance linear orbits whose stability depends on the potential.

The surface of section, shown in Fig. 5(a), for the potential

$$V_4(x,z) = 11\tfrac{4}{15}x^4 + 2x^2z^2 + \tfrac{1}{15}z^4, \quad (8)$$

is symmetric with respect to x and $\dot{x}$, so it suffices to systematically search only the first quadrant for periodic orbits. There are only three stable periodic orbits. The z-axis orbit is stable and generates the outer family of invariant curves. There are a pair of stable fixed points at $(\pm 1.873,0)$ in the surface of section, which are produced by 3:1 resonance elliptic orbits. The z-axis orbit, defined by $x = \dot{x} = 0$, is unstable and coincides with the hyperbolic point on the separatrices. The separatrices divide the invariant curves into three families, one set around each of the three stable periodic orbits. There is a second pair of unstable fixed points at $(0, \pm 39.723)$ that corresponds to 4:1 resonance elliptic orbits.

Earlier we discussed the existence of rational invariant curves surrounding the periodic orbits in integrable potentials. When such a potential is perturbed these curves are destroyed leaving a pair of periodic orbits, one stable and the

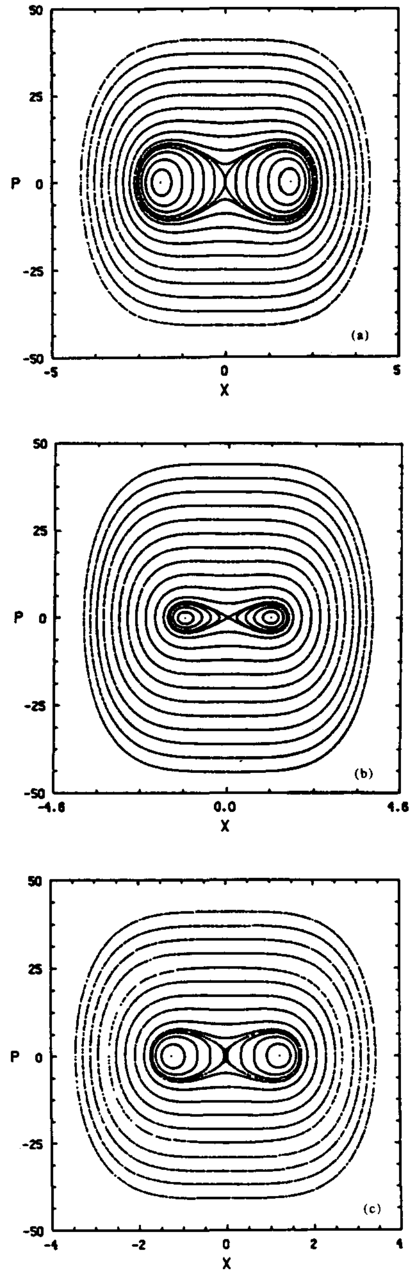


FIG. 5. (a) Surface of section for the homogeneous potential V_4. It is completely regular. This potential remains a candidate for integrability. (b) Surface of section for the homogeneous potential V_7. It is regular at $E = 1000$. (c) Surface of section for the homogeneous potential V_9.

other unstable. In the present case we have the required unstable orbits but there are *no* corresponding stable 4:1 resonance orbits anywhere in the phase space. The unstable 4:1 resonance elliptic orbits are not produced by the disintegration of a rational invariant curve. Furthermore, systematic searching revealed that there were no stable two or three-cycles anywhere in the surface of section. In fact, there were no two-cycles present at all. This potential appears to be completely regular at $E = 1000$.

Since V_4 is a good candidate for integrability we examined its surface of section for various other energies. No orbit bifurcations, chaotic orbits, or island chains were in evidence for the range of energies $E = 0.1$, 1, 10, 100, 1000, and 10 000. In all cases the surfaces of section were topologically identical. Again this is consistent with integrability. A systematic search for an integral of the motion that is sixth order in the velocities failed to reveal one. Integrals of lower order do not exist for this potential.[10] If this potential is integrable then the integral will be at least seventh or probably eighth order in the velocities.

The potential

$$V_7(x,z) = 18\tfrac{2}{11}x^4 + 2x^2z^2 + \tfrac{9}{275}z^4 \tag{9}$$

has surface of section in Fig. 5(b), and is topologically similar to the one obtained for the previous potential V_4. The principle differences are that the separatrix is now much more open, and the periodic orbits are different. Here the stable periodic orbits are the x-axis orbit and a pair of 4:1 resonance linear or W-shaped orbits at $(\pm 1.146,0)$. Again the z-axis orbit is unstable and is the hyperbolic point on the separatrix. There are a pair of unstable 5:1 resonance linear periodic orbits at $(0, \pm 30.944)$. There are no corresponding stable 5:1 resonance orbits. This potential does not appear to possess any island chains. In fact, there are only two unstable two-cycles at $(0, \pm 12.822)$. No stable two- or three-cycles were found anywhere in the surface of section. This potential appears to be regular at $E = 1000$.

The potential V_9, with surface of section in Fig. 5(c), is nonintegrable since there is a small chaotic region centered on the origin. Its surface of section is topologically very similar to the surfaces of section for the potentials V_4 and V_7, shown in Figs. 5(a) and 5(b). Its stable periodic orbits are the x-axis orbit and a pair of 4:1 resonance linear orbits, as was the case for the potential V_7. The unstable periodic orbits are the z-axis orbit and a pair of 5:1 resonance linear orbits at $(0, \pm 25.149)$ in the surface of section.

The homogeneous potential V_8 has surface of section shown in Fig. 6(a). There are four stable periodic orbits. The first pair are the x- and z-axis orbits. The second pair are 8:1 resonance elliptic orbits at $(0, \pm 18.077)$. This potential is almost regular. There are no visible chaotic regions but island chains do exist. The homogeneous potential V_2 is, like V_8, almost regular. Its surface of section is shown in Fig. 6(b). The potential has four stable periodic orbits; the x- and z-axis orbits and a pair of 2:1 resonance linear orbits on the x axis. There are two unstable 3:1 resonance linear orbits at $(0, \pm 35.950)$. Additionally there are a number of small stable cycles around the central fixed points, which generate chains of islands precluding integrability. Consider the corresponding Verhulst potential with quadratic terms present. For $\omega_2^2/\omega_1^2 > \frac{1}{3}$ this potential has surfaces of section that are topologically very similar to those in Fig. 5(a), with the two nonaxial fixed points corresponding to a pair of 2:1 resonance linear stable periodic orbits. As $\omega_2^2/\omega_1^2 \to \frac{1}{3}$ from above, the two nonaxial fixed points approach each other until precisely at the potential with $\omega_1^2 = 3\omega_2^2$ there is an orbit bifurcation point at the origin where the two 2:1 resonance periodic orbits merge and disappear. Additionally, the central z-axis orbit becomes stable producing surfaces of section such as that for the homogeneous version of this potential shown in Fig. 6(b). It should be noted that for all values of

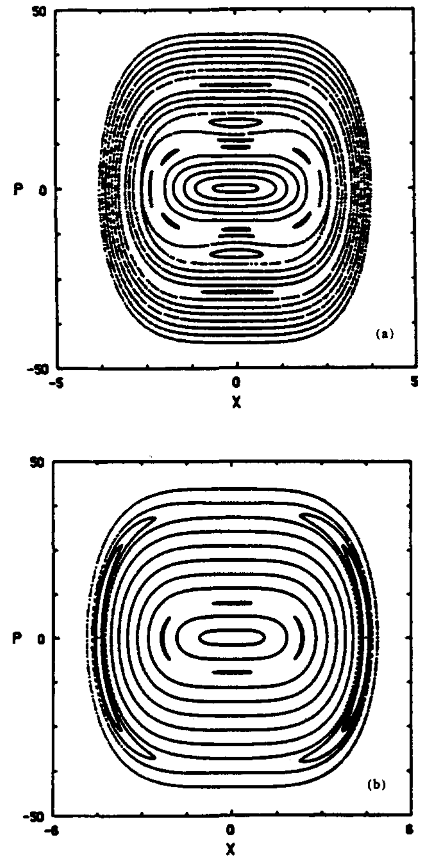


FIG. 6. (a) Surface of section for the homogeneous potential V_8. (b) Surface of section for the homogeneous potential V_2.

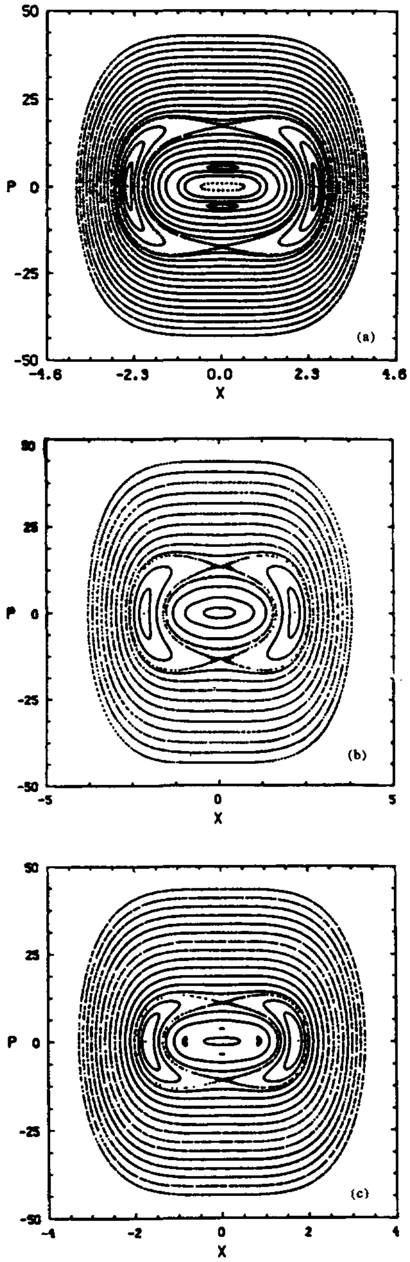


FIG. 7. (a) Surface of section for the homogeneous potential V_5. (b) Surface of section for the homogeneous potential V_6. (c) Surface of section for the homogeneous potential V_{10}.

ω_1^2 and ω_2^2 examined the potential is almost regular and may in fact be integrable for certain values of ω_2^2/ω_1^2.

The potentials V_5, V_6, and V_{10} are characterized by having surfaces of section, which are all topologically very similar; all the features of the surface of section for one of these potentials are reproduced by the others, even to the extent of having the same island chains around the stable z-axis orbit. Figures 7(a), 7(b), and 7(c) show the surfaces of section for the potentials V_5, V_6, and V_{10}, respectively. They all possess small chaotic bands centered on the separatrices and are almost regular. The homogeneous potential V_5 has four stable periodic orbits; the x- and z-axis orbits and a pair of 4:1 resonance axial or W orbits at $(\pm 2.478,0)$ in the surface of section. There are two unstable 4:1 resonance elliptic orbits at $(0, \pm 17.388)$ corresponding to the hyperbolic points on the separatrix separating the four sets of quasiperiodic orbits. The potential V_6 has precisely the same arrangement of stable 4:1 resonance axial periodic orbits. It is remarkable that two completely different potentials should have phase spaces with all their features, especially periodic orbits and rings of unstable fixed points, identical. The potential V_{10} has a pair of stable 5:1 resonance elliptic orbits instead of the 4:1 resonance linear orbits found in V_5 and V_6.

The resonances of the stable periodic orbits belonging to these potentials are given in Table III. There appears to be a reasonable correlation between the leading power q of the singularities and the resonance of the stable nonaxial periodic orbits in the corresponding potentials. Here V_1 and V_3 should be neglected in this comparison because of their logarithmic singularities.

VI. ZIGLIN'S THEOREM

There are a couple of different statements of Ziglin's theorem.[8] The relevant one, for two dimensions, quoted from Yoshida[18] is as follows: Assume that there exists an additional complex analytic integral $\Phi(\mathbf{p},\mathbf{q}) = \text{const}$, which is holomorphic in a neighborhood of the solution. If there exists a nonresonant matrix g_1 in the monodromy group G of the normal variation equations, then either (i) g_1 commutes with any other matrix g_2 in G or (ii) the eigenvalues of g_2 are i and $-i$. This rather abstract theorem can be used to relate integrability to the existence of exponentially unstable periodic orbits.

Yoshida[17,18] determined conditions under which a homogeneous potential V satisfies this theorem for the special case where the periodic orbits are exponentially unstable straight line or 1:1 resonance linear periodic orbits. Such an orbit can be parametrized as

$$x = c_1\Phi(t), \quad z = c_2\Phi(t) ,$$

where $\Phi(t)$ is the solution of the differential equation

$$\frac{d^2\Phi}{dt^2} + \Phi^{2m-1} = 0 ,$$

and the constants c_1 and c_2 are the solutions of

$$\frac{\partial V}{\partial x}(c_1,c_2) = c_1, \quad \frac{\partial V}{\partial z}(c_1,c_2) = c_2 .$$

Yoshida[17] defined an integrability coefficient, which for this type of potential is

$$\mu = \nabla^2 V(c_1,c_2) - (m-1) ,$$

where $\nabla^2 V$ is the Laplacian of V and m is the degree of the homogeneous potential V. He showed that if μ is in the region S then the system is nonintegrable and there is no additional integral of the motion. For our quartic potentials $m = 4$ and the region is

$$S = \{\lambda<0,1<\lambda<3,6<\lambda<10,15<\lambda<21,28<\lambda<36,45<\lambda<55,$$
$$66<\lambda<78,91<\lambda<105,120<\lambda<136,153<\lambda<171,190<\lambda<210,$$
$$231<\lambda<253,276<\lambda<300,325<\lambda<351,378<\lambda<406,435<\lambda<465,496<\lambda<528,561<\lambda<595,...\} .$$

There are three types of straight line orbits in our systems, for each of which we calculate a different integrability coefficient (1) $c_1 = 0$ for the z-axis orbit, which has $\mu_1 = 1/\lambda_3$; (2) $c_2 = 0$ for the x-axis orbit, which has $\mu_2 = 1/\lambda_1$; and (3) $c_1 \neq 0, c_2 \neq 0$ for the inclined linear orbits, such as found in the potential U_3, which have $\mu_3 = (1 - 2(\lambda_1 + \lambda_3) + 3\lambda_1\lambda_3)/\lambda_1\lambda_3 - 1$.

The values of μ_1 and μ_2 are easily found for each potential. The μ_3 values are given in Table III. Here μ_1 lies in the region S only for the potentials V_3 and V_7. Note that μ_2 does not lie in S for any of the ten potentials and μ_3 only lies in S for the potential V_6. We had already determined that V_3 and V_6 were nonintegrable. However, we have now eliminated V_7, which looked numerically integrable. We expect that for some energies the z-axis orbit will generate a chaotic orbit giving a surface of section topologically similar to that of the potential V_9. The potential V_3 has a chaotic orbit centered on the origin in the surface of section [Fig. 3(b)], which is generated by the z-axis orbit. This is consistent with the corresponding value of μ_1 lying in the region S. This potential also has a movable logarithmic singularity at sixth order in its solution.

All the nonaxial periodic orbits in all the potentials in Table III have resonances of type 2:1 or greater. That is, none of these periodic orbits are straight line solutions. Consequently the third of the cases used in this application of Ziglin's theorem is not relevant for the potentials in Table III. The x- and z-axis orbits are the only straight line solutions whose results should be checked. Therefore we disregard the result for V_6. Periodic orbit bifurcations can occur as the energy is varied. This may affect the stability of the x- and z-axis orbits, but will not produce inclined straight line orbits. All new orbit families will also be higher resonance types. Therefore it is important to determine the resonances of all the periodic orbits belonging to a potential before applying Yoshida's formulation of Ziglin's theorem to ensure that the results are meaningful. Ziglin's theorem should be applied in conjunction with an examination of the system's orbital structure. It is theoretically possible to numerically apply Ziglin's theorem to nonstraight line orbits. This is very difficult and will not be dealt with here.

It is surprising that only two of the homogeneous potentials in Table III are shown to be nonintegrable by this application of Ziglin's theorem. Seven of the remaining eight potentials are either integrable or close to integrable, satisfy Yoshida's formulation, and have only movable rational algebraic branch points in their solutions. It is also interesting to note that many of the potentials have one or more of their coefficients μ_1, μ_2, and μ_3 lying precisely on the boundary of the forbidden region S.

VII. CONCLUSION

Weak Painlevé analysis of the quartic polynomial Verhulst potential allowed us to identify seven integrable cases. Second integrals were given for four of these families and for subclasses of the remaining three. The corresponding surfaces of section are all particularly simple in structure and possessed no island chains or chaotic regions. Surfaces of section belonging to perturbations of these integrable potentials always exhibited nonregular behavior. All the nonaxial stable periodic orbits were 1:1 or 2:1 resonance.

The expansion of the class of admissible singularities in Painlevé analysis to include all rational algebraic branch points gave eight homogeneous potentials with negligible nonregular regions. Examination of the respective surfaces of section revealed that all these potentials, except for V_4 and V_7, possessed very small chaotic regions. Ziglin's theorem was used to show that V_7 is not integrable. Numerical searches reveal that there are no lower-order stable cycles or island chains anywhere in the surface of section of V_4. Surfaces of section were then calculated for this potential over a range of energies 0.1–10^4. They were all topologically identical and showed no orbit bifurcations, chaotic regions, or island chains. This potential still remains a candidate for integrability.

We conclude that, in general, most movable algebraic branch points are inconsistent with integrability. However, there are some exceptions. These include those used in weak Painlevé analysis, having expansions for the quartic potentials in terms of $\Delta t^{1/2}$, and possibly the potential V_4, with an expansion in terms of $\Delta t^{1/13}$, which may be integrable. Conversely, if one is interested only in potentials that are regular or nearly regular, so that they look numerically integrable, then allowing any type of movable rational algebraic branch points provides a significant number of such potentials.

Eight of the ten homogeneous potentials examined here are either regular or very close to regular for the energy examined. Only two have easily observable chaotic regions. They both possess movable logarithmic singularities at fourth and sixth order, respectively. Numerical calculations for a large number of nonhomogeneous as well as these two homogeneous quartic polynomial potentials suggest that for these types of systems movable logarithmic singularities are inconsistent with integrability. In fact, large-scale chaotic regions appear to be associated with the existence of movable logarithmic singularities and the point at which they occur in the series expansions of the solutions. Conversely, six of the remaining potentials possess very small chaotic orbits centered on hyperbolic points. These appear to be associated with movable rational algebraic branch points. Perhaps this is reasonable since logarithmic singularities are infinitely branched and the rational algebraic branch points examined have only a small finite number of branches.

The regular and nearly regular homogeneous potentials have surfaces of section, which can be divided into a small number of topologically distinct classes represented principally by those shown in Figs. 5(a) and 7(a). This suggests that their topology is very restricted and always very simple for these two-dimensional potentials. Many complex regular surfaces of section can be invented, but none of these structures arise from any of the potentials examined here.

As the order n of the singularities increases the corresponding λ_1 value of the potential increases and the λ_3 value decreases. This produces potentials that are increasingly elongated in the z direction. The stable periodic orbits exhibit higher and higher resonances producing multilobed quasiperiodic orbits. The existence of high-order resonance periodic orbits means that Yoshida's formulation of Ziglin's theorem cannot be used in isolation but should be used in conjunction with an examination of the periodic orbit structure. Only one of the eight potentials possessing only movable rational algebraic branch points was eliminated by Ziglin's theorem. Most of the integrability coefficients of other potentials occurred on the boundary of the forbidden region.

[1]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[2]Y. F. Chang, M. Tabor, and J. Weiss, J. Math. Phys. **23**, 531 (1982).
[3]H. Segur, "Solitons and inverse scattering transform," Lectures given at the International School of Physics, 'Enrico Fermi,' Varenna, Italy, 1980.
[4]T. Bountis, H. Segur, and F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).
[5]A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539 (1982).
[6]B. Grammaticos, B. Dorizzi, and A. Ramani, J. Math. Phys. **24**, 2289 (1983).
[7]H. Yoshida, Celestial Mech. **31**, 380 (1983).
[8]S. L. Ziglin, Funct. Anal. Appl. **16**, 181 (1983); **17**, 6 (1983).
[9]F. Verhulst, Philos. Trans. R. Soc. London **290**, 435 (1979).
[10]J. Hietarinta, Phys. Rep. **147**, 87 (1987).
[11]J. Hietarinta, Phys. Lett. A **96**, 273 (1983).
[12]Y. Aizawa, and N. Saitô, J. Phys. Soc. Jpn. **32**, 1636 (1972).
[13]B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **25**, 481 (1984).
[14]M. Hénon, and C. Heiles, Astron. J. **69**, 73 (1964).
[15]M. Hénon in, *Chaotic Behaviour of Deterministic Systems*, Les Houches, Session XXXVI, 1981, edited by G. Iooss, R. H. G. Helleman, and R. Stora (North-Holland, Amsterdam, 1983).
[16]J. D. Hadjidemetriou, Celestial Mech. **12**, 255 (1975).
[17]H. Yoshida, Physica D **21**, 163 (1986).
[18]H. Yoshida, Physica D **29**, 128 (1987).

Group cocycles, line bundles, and anomalies

Matthias Blau
Institut für Theoretische Physik, Universität Wien, Vienna, Austria and SISSA, Trieste, Italy

(Received 15 December 1988; accepted for publication 10 May 1989)

The relation between complex line bundles and certain group cocycles is explored in general to obtain explicit formulae for the transition functions and curvature of the determinant line bundle DET of a family of Dirac operators coupled to Yang–Mills fields. A covariant derivative on sections of DET is constructed which realizes the curvature and "minimally couples" to the integrated anomaly which thus appears as a "functional magnetic field" on gauge orbit space. The transcription of group cohomological (cocycles) into geometrical (line bundles) information is refined in such a way that the relevant cohomology groups can be computed in many cases, giving insight into the classification of lifts of principal group actions.

I. INTRODUCTION

Geometrical and topological methods have played a prominent role in recent developments in theoretical physics, and physicists have acquired familiarity with the standard notions of differential geometry and algebraic topology.

However, recently objects (called generalized associated bundles hereafter) have appeared in the physics literature,[1–6] about whose general structure little seems to be known. The purpose of this paper is to fill this gap and thus to perhaps pave the ground for further applications.

The basic idea is to replace representations of a Lie group G on a vector space V (thus defining ordinary associated vector bundles of a principal G-bundle P) by one-cocycles of G with values in suitable G-modules. The bundles obtained in this way could, of course, be abstractly regarded as arising from a special kind of G-bundles.[3,7] The methods and results of Secs. II–IV will, however, show that regarding them as associated in some way to a principal "parent" bundle has a number of succinct computational and conceptional advantages. The case we shall be interested in is the one of complex line bundles. The one-dimensional representations of $G(\mathrm{Hom}(G,\mathrm{GL}(1,\mathbf{C})))$ are contained in the larger set of cocycles $Z(G,\mathbf{C}^*(P))$ with values in the G-module $\mathrm{Map}(P,\mathbf{C}^*) = : \mathbf{C}^*(P)$. More precisely, the cocycle condition

$$f(p,g_1g_2) = f(p,g_1)f(pg_1,g_2),$$

satisfied by elements f of $Z(G,\mathbf{C}^*(P))$, reduces to

$$f(g_1g_2) = f(g_1)f(g_2)$$

on those f which are independent of P. We shall show how to associate a line bundle L_f on the base manifold M of P with every such cocycle f. The line bundles constructed in this way are thus true generalizations of ordinary associated line bundles. We explicitly construct the local data (sections, transition functions) which determines these bundles, and compute their first (real) Chern class (Secs. II and II). Then we show how the usual prescription for constructing covariant derivatives in ordinary associated bundles can be modified to accommodate the bundles considered here (Sec. IV).

It turns out that the "ordinary" connection is modified by a connection one-form proportional to the infinitesimal variation of f (i.e., a Lie-algebra cocycle) with transgresses to the first Chern class of L_f.

All this is, of course, quite reminiscent of anomalies in theories of chiral fermions interacting with non-Abelian gauge fields,[8] and indeed we shall show in Sec. VI that the determinant line bundle of a family of Dirac operators is of the form L_f for $P = \mathscr{A}$ (the space of vector potentials), $G = \mathscr{G}$ (the gauge group) and $f =$ the Wess–Zumino term.[9]

Our computation of the curvature then permits one to check[6] explicitly, that the curvature of the determinant line bundle and the integrated anomaly are related by transgression.

On the more mathematical side we use the relation between line bundles and cocycles to encode group cohomological into geometrical information to compute $H^1(G,\mathbf{C}^*(P))$ in a number of cases.

In particular, in Sec. V we shall prove the following theorem.

Theorem 1: If $H^2(P,\mathbf{Z}) = 0$,

$$H^1(G,\mathbf{C}^*(P)) = H^2(M,\mathbf{Z}).$$

This result was also derived in the particular case $P = \mathscr{A}$ mentioned in Ref. 5 and says that in this case all line bundles on M arise as L_f for some cocycle f. Furthermore, the relation between $H^1(G,\mathbf{C}^*(P))$ and the problem of classifying G-lifts to principal bundles[10] allows us to prove the next theorem.

Theorem 2: Let $P(M,G)$ be a principal G-bundle over a connected manifold M. Let $\mathscr{L}$ be a line bundle on P admitting a lift of the principal G-action on P. Then this lift is unique if either P is trivial or $H^2(M,\mathbf{Z}) = 0$.

Finally, the Appendix contains a technical lemma on partitions of unity which allows for a simplification of the calculations in Sec. VI.

II. COCYCLES AND LINE BUNDLES

Let $P = P(M,G,\Pi)$ be a principal G-bundle over a (paracompact, connected) manifold M, with projection Π: $P \to M, G$ a connected Lie group. Since P carries a natural (right) G-action, this is inherited by functions on P. In particular, therefore the Abelian group $\mathbf{C}^*(P)$ of complex valued nowhere vanishing functions on P is a G-module, and we

 0022-2488/89/102226-07$02.50

can define the cohomology of G with values in $\mathbf{C}^*(P)$ in the standard manner.[11] Zero-cochains are basically just elements f of $\mathbf{C}^*(P)$ and the coboundary operator δ acts on f by

$$\delta f(p,g) := f(pg)f(p)^{-1}. \tag{2.1}$$

Thus $f \in \mathbf{C}^*(P)$ is a cocycle iff it is G-invariant. On one-cochains $f(p,g)$, δ acts as

$$\delta f(p,g_1,g_2) = f(p,g_1)f(pg_1,g_2)f(p,g_1g_2)^{-1}. \tag{2.2}$$

In a similar way the action of δ is extended to higher cocycles, but this is all we will need here. The space of k-cocycles (k-coboundaries) will be denoted by $Z^k(G,\mathbf{C}^*(P))$ ($B^k(G,\mathbf{C}^*(P))$) and the cohomology groups $H^k(G,\mathbf{C}^*(P))$ are defined as usual by $H^k = Z^k/B^k$. We shall frequently abbreviate $H^1(G,\mathbf{C}^*(P))$ by $\mathbf{H}$.

Given an element f of Z^k we can use it to define an equivalence relation on $P \times \mathbf{C}$,

$$(p,c) \sim (pg, f(p,g)c), \quad g \in G. \tag{2.3}$$

This relation is indeed transitive, since f is a cocycle. The space of equivalence classes $[(p,c)]$—the quotient $(P \times \mathbf{C})/\sim$—has the structure of a complex line bundle over M, denoted by L_f. The local trivializability—which may perhaps not be obvious—will be explicitly proved in the next section.

As mentioned in the Introduction, bundles of this kind have recently appeared in the physics literature (mainly in relation with anomalies).[1–6] Their geometrical structure, however, was not further investigated.

If the one-cocycle f is independent of P, the cocycle condition $\delta f = 1$ implies that

$$f(g_1g_2) = f(g_1)f(g_2)$$

and thus that $f \in \mathrm{Hom}(G,\mathbf{C}^*)$. Since $\mathbf{C}^* = GL(1,\mathbf{C})$ the bundle L_f is in this case an ordinary associated complex line bundle to P via the representation f.

We shall now show that the assignment cocycles → line bundles descends to an assignment between cohomology classes and equivalence classes of line bundles in the sense that cohomologous cocycles lead to equivalent line bundles. This fact is implied by the following proposition.

Proposition 1: L_f is trivial iff f is trivial.

Before turning to its proof, let us note the following. In the particular case $P = \mathscr{A}$, this was also shown in Ref. 4. Furthermore, we shall make use of this proposition later in Sec. V, where we compute $H^1(G,\mathbf{C}^*(P))$, since it allows us to transform group-cohomological into (more tractable) geometrical information.

Proof of Proposition 1: If f is trivial, $f(p,g) = F(pg)/F(p)$ for some $F \in (\mathbf{C}^*(P))$. Then we can define a global nowhere vanishing section ψ of L_f (equivalently: a global section of the associated $\mathbf{C}^*$-bundle) by

$$\psi(m) := [(p,F(p))],$$

where $\Pi(p) = m$. This is indeed independent of the choice of $p \in \Pi^{-1}(m)$, since

$$[(p,F(p))] = [(pg, f(p,g)F(p))] = [(pg,F(pg))].$$

Conversely, if L_f is trivial, there exists a global nonvanishing section ψ: $M \to L_f$, which is always of the form

$$\psi(m) = [(p, f_\psi(p))]$$

for some $f_\psi: P \to \mathbf{C}^*$. Since ψ is well-defined we conclude that

$$f(p,g) = (f_\psi(pg)/f_\psi(p))$$

and hence that f is trivial. □

Already implicit in the proof above was the fact that sections ψ of L_f are generally in one-one correspondence with functions $f_\psi: P \to \mathbf{C}$ satisfying the equivariance condition

$$f_\psi(pg) = f(p,g)f_\psi(p) \tag{2.4}$$

via

$$\psi(m) = [(p, f_\psi(p))], \tag{2.5}$$

where the rhs is independent of p as a consequence of (2.4).

III. THE LOCAL GEOMETRY OF L_f

We shall now construct the local geometrical data that determine L_f (transition functions, local sections) from those of P and use them to derive an explicit formula for the curvature F_f of L_f and hence for a representative of the first Chern class $c_1(L_f)$ (in real cohomology). Let $\{U_\alpha\}$ be a locally finite, good[12] covering of M by open sets U_α and let s_α: $U_\alpha \to \Pi^{-1}(U_\alpha) \subset P$ be local trivializing sections of P, $g_{\alpha\beta}$: $U_\alpha \cap U_\beta \to G$ the corresponding transition functions. These define local sections ψ_α of L_f by

$$\psi_\alpha(m) := [(s_\alpha(m),1)] \tag{3.1}$$

and can be used to construct local trivializations, thus proving the local trivializability of L_f claimed in the previous chapter.

Defining $g_\alpha: \Pi^{-1}(U_\alpha) \to G$ by

$$s_\alpha(m)g_\alpha(p) = p, \quad p \in \Pi^{-1}(m),$$

we can express the ψ_α in the form (2.5)—i.e., via local equivariant functions—as

$$\psi_\alpha(m) = [(p, f_\alpha(p))] \tag{3.2}$$

with

$$f_\alpha(p) := f(s_\alpha(m), g_\alpha(p)). \tag{3.3}$$

Indeed we have

$$f_\alpha(pg) = f(p,g)f_\alpha(p), \tag{3.4}$$

where we have used

$$g_\alpha(pg) = g_\alpha(p)g \tag{3.5}$$

and the cocycle condition on f.

The local sections ψ_α (3.1) define the corresponding transition functions $\psi_{\alpha\beta}$ by

$$\psi_\alpha(m) = \psi_{\alpha\beta}(m)\psi_\beta(m).$$

Using $s_\alpha(m) = s_\beta(m)g_{\beta\alpha}(m)$ one computes

$$\begin{aligned}\psi_\alpha(m) &= [(s_\alpha(m),1)] \\ &= [(s_\beta(m)g_{\beta\alpha}(m),1)] \\ &= [(s_\beta(m), f(s_\alpha(m),g_{\alpha\beta}(m)))] \\ &= f(s_\alpha(m),g_{\alpha\beta}(m))\psi_\beta(m).\end{aligned}$$

Thus the transition functions of L_f are

$$\psi_{\alpha\beta}(m):=f(s_\alpha(m),g_{\alpha\beta}(m)). \quad (3.6)$$

Using $f(p,g)^{-1}=f(pg,g^{-1})$ and $g_{\beta\alpha}(m)g_\alpha(p)=g_\beta(p)$ one easily proves

$$\psi_{\alpha\beta}=\psi_{\beta\alpha}^{-1}$$

and

$$\psi_{\alpha\beta}\psi_{\beta\gamma}=\psi_{\alpha\gamma},$$

as behooves a set of transition functions.

This then completes the description of L_f in terms of local data, but in order to determine the first Chern class $c_1(L_f)$ from these (which must—in principle—be possible, since they contain the whole information on the bundle), we need as an additional input a partition of unity $\{\rho_\alpha\}$ subordinate to the locally finite covering $\{U_\alpha\}$: i.e., a set of *smooth* functions $\{\rho_\alpha\}$ with the properties

$$\operatorname{supp}\rho_\alpha\subset U_\alpha,$$
$$0\leqslant\rho_\alpha\leqslant 1,$$
$$\sum\rho_\alpha=1.$$

Such coverings and subordinate partitions of unity exist, since we assumed M to be paracompact. Given this, $c_1(L_f)$ —or rather a representative two-form $(1/2\pi i)F(L_f)$, where $F(L_f)$ is the "curvature" of L_f— can be expressed on U_α by[12]

$$F(L_f)_\alpha=\sum_\gamma d\rho_\gamma\wedge d\log\psi_{\gamma\alpha}, \quad (3.7)$$

where d is the exterior derivative on M. Note that $F(L_f)$ has a globally defined expression on P (where it is also exact), and to compute $\Pi^*F(L_f)$ and hence $\Pi^*c_1(L_f)$, we first observe that

$$\psi_{\gamma\alpha}=s_\alpha^*f_\gamma, \quad (3.8)$$

where f_α was defined in (3.3) and s_α was the local trivializing section of P on U_α. Using the fact that moreover

$$s_\alpha(\Pi(p))=pg_\alpha(p)^{-1},$$

by the definition of g_α we find

$$\Pi^*\psi_{\gamma\alpha}=f_\alpha^{-1}f_\gamma. \quad (3.9)$$

[Note that the functions $\Pi^*\psi_{\gamma\alpha}$ are the transition functions of the pullback bundle Π^*L_f on P. Equation (3.9) therefore shows that this bundle is trivial and we consequently expect $c_1(\Pi^*L_f)=\Pi^*c_1(L_f)$ to be trivial and $\Pi^*F(L_f)$ to be globally exact.]

Indeed we find

$$\Pi^*F(L_f)_\alpha=\left(d_p\sum_\gamma\Pi^*\rho_\gamma\, d_p\log\Pi^*\psi_{\gamma\alpha}\right)$$
$$=\left(d_p\sum_\gamma\Pi^*\rho_\gamma\, d_p\log f_\gamma\right)$$

(d_p is the exterior derivative on P), which is independent of α, and globally exact,

$$\Pi^*F(L_f)=d_p\Gamma, \quad (3.10)$$

with

$$\Gamma=\sum_\gamma\Pi^*\rho_\gamma\, d_p\log f_\gamma. \quad (3.11)$$

In Sec. VI—where we shall consider the case $P=\mathscr{A}$—we shall show in our discussion of the determinant line bundle, that the restriction of Γ to a fiber in P [the antitransgression of $F(L_f)$] gives the integrated anomaly[13] in accordance with the general expectation that the curvature of the determinant line bundle and the anomaly should be related in this way.[14–16]

In the next section we shall see how Γ (3.11) enters the construction of a covariant derivative on L_f (in particular—in the situation of Sec. VI—this will show how the connection on the determinant line bundle describes a "minimal coupling" to the anomaly).

IV. COVARIANT DERIVATIVES ON GENERALIZED ASSOCIATED LINE BUNDLES

Our aim will now be to define a connection in the generalized associated complex line bundle L_f. To see where the difficulty lies, let us recall how a connection on the principal bundle P induces a covariant derivative on the space of sections of an ordinary associated bundle E with standard fiber F. Given a section ψ of E and a vector field X on M, a new section $\nabla_X\psi$ is defined by the following steps:

Use the principal connection on P to lift X to a horizontal vector field $\tilde{X}$ on P;

Associate to ψ the corresponding equivariant function $f_\psi: P\to F$ such that $\psi(m)=[(p,f_\psi(p))]$;

Form $\tilde{X}f_\psi$ (the derivative of f_ψ along $\tilde{X}$).

By the equivariance of f_ψ and the right-invariance of $\tilde{X}$, $\tilde{X}f_\psi$ is again an equivariant F-valued function on P and thus defines a new section of E which we call $\nabla_X\psi$,

$$(\nabla_X\psi)(m):=[(p,\tilde{X}f_\psi(p))],\quad p\in\Pi^{-1}(m).$$

In our case, however, the equivariant function f_ψ corresponding to a section ψ of L_f will satisfy (3.7),

$$f_\psi(pg)=f(p,g)f_\psi(p).$$

Therefore $\tilde{X}f_\psi$ will only be equivariant up to a term proportional to $\tilde{X}f$ [this term is, of course, zero for ordinary associated bundles, where f is an element of $\mathrm{Hom}(G,\mathbf{C}^*)$ and therefore independent of P]. But lack of equivariance implies that $\tilde{X}f_\psi$ does not define a new section of L_f. We are thus led to look for a modification of the above prescription which preserves the equivariance (cf. also Ref. 2).

Let $\varphi=(\varphi_\alpha)$ be a section of L_f, with $\varphi_\alpha=\psi_{\alpha\beta}\varphi_\beta$, and let $h_\alpha(=f_{\varphi_\alpha})$ be the corresponding local equivariant functions. Define new functions $D_{\tilde{X}}^{\alpha}h_\alpha$ on $\Pi^{-1}(U_\alpha)$ by[17]

$$D_{\tilde{X}}^{\alpha}h_\alpha:=\tilde{X}h_\alpha-\Gamma_\alpha(\tilde{X})h_\alpha\quad(\text{no sum over }\alpha), \quad (4.1)$$

where

$$\Gamma_\alpha=d_p\log f_\alpha.$$

Then it is easy to see that indeed

$$(D_{\tilde{X}}^{\alpha}h_\alpha)(pg)=f(p,g)(D_{\tilde{X}}^{\alpha}h_\alpha)(p) \quad (4.2)$$

(i.e., the lack of equivariance of $\tilde{X}h_\alpha$ is precisely compensated by the lack of equivariance of Γ_α). Due to the equivariance of h_α and the invariance of $\tilde{X}$, $\tilde{X}h_\alpha$ transforms as

$$(\tilde{X}h_\alpha)(pg) = (\tilde{X}h_\alpha(p))f(p,g) + h_\alpha(p)\tilde{X}f(p,g)$$
$$= f(p,g)\,[\tilde{X}h_\alpha(p) + (\tilde{X}\log f(p,g))h_\alpha(p)]. \quad (4.3)$$

On the other hand, $d_p \log f_\alpha$ transforms as ($\Pi(p) = m$)

$$d_p \log f_\alpha(p,g) = d_p \log f(s_\alpha(m),g_\alpha(pg))$$
$$= d_p \log(f(s_\alpha(m),g_\alpha(p)g))$$
$$= d_p \log(f(p,g))f(s_\alpha(m),g_\alpha(p))$$
$$= d_p \log(f(p,g)) + d_p \log f_\alpha(p). \quad (4.4)$$

Putting (4.3) and (4.4) together we find (4.2).

Thus the "local" covariant derivative ∇_X^α given by

$$(\nabla_X^\alpha \varphi_\alpha)(m) := [(p,(D_X^\alpha h_\alpha)(p))] \quad (4.5)$$

is well defined. Furthermore, it is straightforward to check that

$$\nabla_X^\alpha \varphi_\alpha = \psi_{\alpha\beta} \nabla_X^\beta \varphi_\beta. \quad (4.6)$$

Hence $\nabla_X \varphi = (\nabla_X^\alpha \varphi_\alpha)$ is a well-defined new section of L_f, and the operator ∇_X defined in this way satisfies all the axioms of a covariant derivative. Equations (4.2) and (4.6) now show that we have arrived at our goal of defining a covariant derivative on the sections of L_f.

Regarding the operator $D^\alpha = d - \Gamma_\alpha$ instead as a covariant derivative on the sections of the trivial line bundle $P \times \mathbf{C}$ on P, we see that (according to the general prescription of extending connection potentials [Ref. 18, p. 68]) the Γ_α piece together to the connection potential

$$\Gamma = \sum \Pi^* \rho_\alpha \Gamma_\alpha,$$

where Γ is precisely the one-form on P obtained in Sec. III by pulling back $c_1(L_f)$ to P.

Thus the general recipe for finding a covariant derivative on L_f can also be expressed in the following way: Given the local expression (3.7) for the curvature of L_f, pull it back to obtain Γ via (3.10), (3.11). This Γ defines a covariant derivative on sections of $P \times \mathbf{C}$ which descends to a covariant derivative on $(P \times \mathbf{C})/\sim\; = L_f$.

Looking at it this way, the fact that (the pullback of) the curvature of L_f—computed by means of ∇_X—equals $d\Gamma$ is quite obvious.

V. CALCULATION OF $H^1(G,\mathbf{C}^*(P))$

As a by-product of our previous discussion we are now able to compute the cohomology group $H^1(G,\mathbf{C}^*(P))$ introduced in Sec. II in a number of cases. More generally, the groups $H^1(G,\mathrm{Map}(X,U(1)))$ (where X is a manifold carrying a G-action) were introduced in Ref. 10 to study the problem of lifting the G-action on X to automorphisms of torus-bundles on X. Explicit calculation of this group is, however, quite difficult in general, since it involves an intricate relationship between the topologies of G and X. In our case the additional structure provided by the fact that $P = X$ is a principal G-bundle allows us to compute it explicitly. The relation between $H^1(G,\mathbf{C}^*(P))$ and line bundles on M we have established so far will be refined in such a way that the computation of $H^1(G,\mathbf{C}^*(P))$ becomes geometrically accessible. We shall then also apply these results to the question of classifying G-lifts.

The first bit of information we need—and which we have already established in Ref. II—is the fact (Proposition 1) that L_f is trivial if f is trivial. Thus, if we can for some other (topological, geometrical) reason show that L_f has to be trivial, we can conclude the triviality of $\mathbf{H} := H^1(G,\mathbf{C}^*(P))$.

One such situation occurs if P is trivial because of the following proposition.

Proposition 2: If P is trivial, L_f is trivial.

Proof: Let s be a global trivializing section of $P, s: M \to P$. Then $\psi(m) := [(s(m),1)]$ is a nowhere vanishing section of L_f. Thus L_f is trivial.

According to the above remarks we thus have the following.

Corollary 1: If P is trivial, $H^1(G,\mathbf{C}^*(P))$ is trivial.

Since line bundles on M are classified by $H^2(M,\mathbf{Z})$ we can also deduce the next proposition.

Proposition 3: If $H^2(M,\mathbf{Z}) = 0$, $H^1(G,\mathbf{C}^*(P))$ is trivial.

Notice how $\mathbf{H}$ "feels" the triviality of P and the cohomology of M via the G-module $\mathbf{C}^*(P)$.

Another situation which is still tractable but in a somewhat less straightforward manner, finally giving rise to a nontrivial $\mathbf{H}$, occurs if $H^2(P,\mathbf{Z}) = 0$. Indeed we shall show below that then $\mathbf{H} = H^2(M,\mathbf{Z})$. Before proceeding to the proof let us make the following remarks.

(i) In particular, this result implies that in the case $H^2(P,\mathbf{Z}) = 0$ *every* complex line bundle on M is of the form L_f for some cocycle f.

(ii) In the case $P = \mathscr{A}$ (the space of connections on a principle bundle P'), $G = \mathscr{G}$ (the group of pointed vertical automorphisms of P') this result has been established in a very nice way in Ref. 5, where it was used to relate the group-cohomological and topological aspects of anomalies. This relation has been sharpened and made explicit in Ref. 6 (cf. also Sec. VI).

(iii) The result $H^1(G,\mathbf{C}^*(P)) = H^2(M,\mathbf{Z})$ is consistent with the results for $\mathbf{H}$ derived above. Indeed this is trivial for Proposition 3. As for Corollary 1 note that $H^2(P,\mathbf{Z}) = 0$ together with the triviality of P imply—by the Künneth formula applied to $P = M \times G$—that $H^2(M,\mathbf{Z}) = 0$.

Theorem 1: If $H^2(P,\mathbf{Z}) = 0$, then

$$H^1(G,\mathbf{C}^*(P)) = H^2(M,\mathbf{Z}).$$

Proof: We have already seen that for any L_f the pullback to P—$\Pi^* L_f$—is trivial (Sec. III). However, $H^2(P,\mathbf{Z}) = 0$ implies that all line bundles on P—and in particular the pullbacks of *all* line bundles on M—are trivial. This makes it plausible that we can recover *all* line bundles on M by "quotienting" $P \times \mathbf{C}$.

We shall now show how to construct an element of $Z^1(G,\mathbf{C}^*(P))$ from any line bundle L on M, and then prove the bijectivity (in cohomology) of this construction.

Thus let L be any line bundle on M with projection Π_L. Note that $\Pi^* L$ is trivial, and via the choice of a global non-vanishing section $\psi: P \to \Pi^* L$ we have $\Pi^* L \sim P \times \mathbf{C}$.

Since $\Pi^* L := \{(p,l) \epsilon P \times L : \Pi(p) = \Pi_L(l)\}$ ψ is of the form

$$\psi(p) = (p, f_\psi(p)) \qquad (*)$$

for some fiber-preserving bundle map f_ψ: $P \to L$. Since $f_\psi(pg)$ and $f_\psi(p)$ sit in the same fiber of L, there is a smooth function $\alpha_g: P \to \mathbf{C}^*$ such that

$$f_\psi(p) = \alpha_g(p) f_\psi(pg)$$

(the reason for putting α_g on the rhs of the above equation will become apparent below). Consistency of this relation requires that $\alpha(p,g) := \alpha_g(p)$ is a group one-cocycle with values in $\mathbf{C}^*(P)$; i.e., an element of $Z^1(G,\mathbf{C}^*(P))$, since

$$\alpha(p,gh) = \alpha(p,g)\alpha(pg,h).$$

In this way every line bundle on M determines a cocycle in $Z^1(G,\mathbf{C}^*(P))$ via a choice of trivialization. If a different trivializing section ψ' is chosen, then ψ is related to ψ' by $\psi = \varphi\psi'$ for some $\varphi \in \mathbf{C}^*(P)$. Defining the new cocycle α' by

$$f_{\psi'}(p) = \alpha'(p,g) f_{\psi'}(pg),$$

one finds

$$\alpha'(p,g) = \alpha(p,g)(\varphi(pg)/\varphi(p)).$$

Thus α' and α are cohomologous,

$$\alpha' = \alpha\delta\varphi$$

[cf. (2.1)] and every line bundle on M defines a cohomology class in $H^1(G,\mathbf{C}^*(P))$.

We shall show next that the cohomology class defined in this way is zero iff L is trivial, i.e., represents zero in $H^2(M,\mathbf{Z})$.

Assume first that α is trivial, i.e., that for a given trivialization ψ of Π^*L we have

$$f_\psi(p) = (F(pg)/F(p)) f_\psi(pg)$$

for some $F \in \mathbf{C}^*(P)$. Then

$$f_\psi(p)F(p) = f_\psi(pg)F(pg)$$
$$=: \sigma(m)$$

is obviously independent of $p \in \Pi^{-1}(m)$ and thus yields a global section σ: $M \to L$, which is nowhere vanishing. Thus L is trivial.

Conversely, if L is trivial, let σ: $M \to L$ be a global trivializing section. Then any trivializing section ψ of Π^*L is of the form $(*)$ with

$$f_\psi(p) = \sigma(m)F(p), \quad m = \Pi(p),$$

for some $F \in \mathbf{C}^*(P)$. Computing $f_\psi(pg)$ we find

$$f_\psi(p) = \alpha(p,g) f_\psi(pg) = \alpha(p,g)\sigma(m)F(pg)$$

and therefore

$$\alpha(p,g) = (F(p)/F(pg)) = \delta F^{-1}(p,g).$$

Thus if L is trivial, α is trivial.

The same method as above can be used to prove that equivalent line bundles on M give rise to the same cohomology class in $\mathbf{H}$. Thus the mapping h: $L \to \alpha$ gives rise to an injective group homomorphism,

$$h_*: H^2(M,\mathbf{Z}) \to H^1(G,\mathbf{C}^*(P)).$$

We shall now show finally that h_* is surjective. This will be done by showing that $h_*[L_f] = [f]$, where L_f is the generalized associated line bundle constructed from the cocycle f as in Sec. II.

The bundle map f_ψ: $P \to L_f$ corresponding to a trivialization of Π^*L_f is of the form

$$f_\psi(p) = [(p, \tilde{f}_\psi(p))]$$

for some $\tilde{f}_\psi \in \mathbf{C}^*(P)$.

Therefore, $f_\psi(pg)$ and $f_\psi(p)$ are related by

$$f_\psi(p) = [(p, \tilde{f}_\psi(p))] = [(pg, f(p,g)\tilde{f}_\psi(p))]$$

(by definition of the equivalence relation on $P \times \mathbf{C}$)

$$= [((pg), (\tilde{f}_\psi(p)/\tilde{f}_\psi(pg)) f(p,g)\tilde{f}_\psi(pg))]$$
$$= f(p,g)\delta\tilde{f}_\psi^{-1}(p,g) f_\psi(pg).$$

Thus $h_*[L_f] = [f]$. This shows that

$$h_*: H^2(M,\mathbf{Z}) \sim H^1(G,\mathbf{C}^*(P)) \qquad \square$$

is an isomorphism.

One final bit of information on $\mathbf{H}$ we can infer directly from Ref. 10. Namely, it is that $\mathbf{H} = H^1(G,\mathbf{Z})$ if G is connected and $H'(P,\mathbf{Z}) = 0$.

Again we can show (at least when M is simply connected) that this is compatible with the results we have derived above on the structure of $\mathbf{H}$. In particular if $H^2(M,\mathbf{Z}) = 0 = H^1(P,\mathbf{Z})$, the cohomology long exact sequence of the principal fibration $G \to P \to M$ implies $H^1(G,\mathbf{Z}) = 0$ and therefore Proposition 3 in this case. On the other hand, if P is trivial and $H^1(P,\mathbf{Z}) = 0$, this implies $H^1(G,\mathbf{Z}) = 0$ (Künneth formula) and hence Corollary 1. Finally, if $H^1(P,\mathbf{Z}) = 0 = H^2(P,\mathbf{Z})$, we have $H^1(G,\mathbf{Z}) = H^2(M,\mathbf{Z})$ (via transgression) and therefore we recover the result of Theorem 1 under the further assumption $H^1(P,\mathbf{Z}) = 0$.

Collecting our results in Table I, we see that the only case in which we have not been able to determine $\mathbf{H}$ explicitly is the one with P nontrivial and $H^1(P,\mathbf{Z})$, $H^2(P,\mathbf{Z})$ and $H^2(M,\mathbf{Z})$ all nonvanishing.

From Ref. 10 we know that $H^1(G,\mathrm{Map}(X,U(1)))$ classifies equivalence classes of lifts of the G-action on X to a $U(1)$-bundle on X (provided *one* lift exists). Similarly $H^1(G,\mathbf{C}^*(X))$ classifies lifts to $\mathbf{C}^*$-bundles (and hence to line bundles) on X; thus in particular $H^1(G,\mathbf{C}^*(P))$ classifies lifts of the principal G-action on P to automorphisms of $\mathbf{C}^*$-bundles on P. In the light of this fact and the above results, we have shown the following.

Theorem 2: Let $P(M,G)$ be a principal G-bundle over a connected manifold M. Let $\mathscr{L}$ be a line bundle on P admitting a lift of the principal G-action on P. Then this lift is unique if either P is trivial or $H^2(M,\mathbf{Z}) = 0$.

While this result could have undoubtedly been derived by other means as well, it illustrates nicely how group-cohomological can be transcribed into geometrical information.

TABLE I. Results of the computation of $H^1(G,\mathbf{C}^*(P))$.

$P(M,G)$	$H^1(G,\mathbf{C}^*(P))$
P trivial	0
$H^1(P,\mathbf{Z}) = 0$	$H^1(G,\mathbf{Z})$
$H^2(P,\mathbf{Z}) = 0$	$H^2(M,\mathbf{Z})$
$H^2(M,\mathbf{Z}) = 0$	0

VI. ANOMALIES, WESS–ZUMINO TERMS, AND THE DETERMINANT LINE BUNDLE

We now finally come to the main application of the techniques developed so far, namely to the construction and investigation of the determinant line bundle[14,19] of a family of Dirac operators coupled to non-Abelian gauge fields. Although these bundles have been around for some time, the results of the previous sections will allow us to be on more intimate terms with them.

Specializing Secs. II–IV to the case (mentioned several times already) $P=\mathscr{A}$, $G=\mathscr{G}$, $M=\mathscr{C}$ (the gauge orbit space) we shall now be able to see the following[6]:

(a) L_f— where f is the cocycle determined by the Wess–Zumino term[9,13]—is the determinant line bundle (cf. also Refs. 1,2) of the family of Dirac operators parametrized by $A\in\mathscr{A}$.

(b) The curvature of L_f (anti-) transgresses to the integrated anomaly obtained via the descent-equations[13] or perturbative calculations.[20] This establishes explicitly the equivalence between the topological (determinant line bundles, index theorem) and algebraic (BRS-cohomology) approaches to anomalies. This had already been done to some extent in Refs. 4 and 5, but our formalism allows us to be quite explicit about this.

(c) The connection on L_f constructed in Sec. IV provides the nice interpretation of the anomaly as a kind of "functional magnetic field" on the gauge orbit space, whose "field strength" is the curvature of the determinant line bundle. It would be interesting to see how this fact is related to the Fock-space picture, where the anomaly also shows up as a U(1) field[21] (Berry's phase[22]).

Finally, since the connection on L_f is in some sense natural, it ought to coincide with Quillen's connection[23] (or rather its generalization[24]), but I have been unable to show this.

Since all the computations have already been done in previous sections, we can be quite brief about these matters here.

Non-Abelian anomalies show up at one loop as lack of gauge invariance of the effective action,

$$W(A\cdot g)\neq W(A),$$

where

$$W(A)=\int d\psi\, d\bar{\psi}\exp\left[-\int\bar{\psi}\not{\partial}_A\psi\right],$$

$$\not{\partial}_A=\not{\partial}+\tfrac{1}{2}\not{A}(1+\gamma_5),$$

$$A\cdot g=g^{-1}Ag+g^{-1}\,dg.$$

However, the modulus of W can be shown to be gauge invariant. Thus

$$W(A^g)=\exp[2\pi i\omega(A,g)]\,W(A), \tag{6.1}$$

where ω is known as the Wess–Zumino term and $\exp 2\pi i\omega$ is a group cocycle in the sense of (2.2). Formally $W(A)$ is the determinant of the Dirac operator $\not{\partial}_A$, and comparing (6.1) with (2.4) and (2.5) shows that W actually defines a section of the line bundle

$$(\mathscr{A}\times\mathbf{C})/\sim\;=:L_f,$$

with

$$(A,c)\sim(A\cdot g,(\exp 2\pi i\omega(A,g))c).$$

Note that L_f is the determinant line bundle, which is thus trivial iff the Wess–Zumino term defines a trivial one-cocycle.

Completing the transgression of F (3.7) by restricting Γ (3.11) to the fiber via the fiber injection $i_A:\mathscr{G}\to\mathscr{A}$, $g\to A\cdot g$ and interpreting the result as a one-form on $\mathscr{G}$ gives

$$i_A^*\Gamma(g)=\sum\rho_\gamma([A])d_{\mathscr{G}}\,\log f(s_\gamma([A]),g_\gamma(A)\cdot g),$$

$$f(A,g)=\exp 2\pi i\omega(A,g), \tag{6.2}$$

where $[A]$ is the orbit $\{A\cdot g,\ g\in\mathscr{G}\}$ and $d_{\mathscr{G}}$ is the exterior derivative on $\mathscr{G}$.

To compare this with the integrated anomaly $[\int\omega^1_{2n}(A,X)$, where $X\in\mathrm{Lie}\,\mathscr{G}]$, we use the fact[13] that it can be obtained as the infinitesimal variation of the Wess–Zumino term (i.e., as a Lie-algebra cocycle).

To simplify the calculation, we choose a covering $\{U_\alpha\}$ in such a way that for a given $[A]$, $\rho_\gamma([A])=\delta_{\gamma\gamma_0}$. Although it may seem obvious that this is always possible, a proof of this is contained in the Appendix, since I was not aware of a reference in the literature and the proof is slightly technical. Then we find (using the equivariance of f and reinserting the factor $2\pi i$ we had omitted)

$$\int\omega^1_{2n}(A,X)=\frac{d}{dt}\,\omega(A,\exp tX)|_{t=0}$$

[by (6.2)]

$$=\frac{1}{2\pi i}\frac{d}{dt}\log f(A,\exp tX)|_{t=0}$$

$$=\frac{1}{2\pi i}\frac{d}{dt}\log f(s_{\gamma_0}(A\exp tX))|_{t=0}$$

$$=\frac{1}{2\pi i}\,i_A^*\Gamma(e)(X) \tag{6.3}$$

(where e is the identity-element of $\mathscr{G}$). Thus we have explicitly verified that the curvature of the determinant line bundle L_f antitransgresses to the integrated anomaly.

Pulling back the connection (Sec. IV) to $\mathscr{A}\times\mathbf{C}$, whose connection potential is Γ, one obtains a $\mathscr{G}$-equivariant flat and necessarily trivial connection there, which however restricts to a flat connection with nontrivial holonomy on the gauge orbits iff the anomaly is nontrivial [because of (6.3)], iff the cocycle is nontrivial (since $H^1(\mathscr{G})\sim H^2(\mathscr{A}/\mathscr{G})\sim\mathbf{H}$) iff L_f is nontrivial (Proposition 1).

APPENDIX: A LEMMA ON PARTITIONS OF UNITY

The purpose of this Appendix is to prove the following lemma, which we used in Sec. VI to simplify the calculation of the antitransgression.

Lemma: Let M be a paracompact smooth Hausdorff manifold and let $x\in M$ be a point of M. Then there exists a covering of M by open sets $\{W_\alpha\}_{\alpha\in I}$ and a partition to unity $\{\rho_\alpha\}$ subordinate to $\{W_\alpha\}$ such that $\rho_\alpha(x)=0\ \forall\alpha\in I$, $\alpha\neq\alpha_0$, and $\rho_{\alpha_0}(x)=1$.

Proof: Let $\{U_\alpha\}_{\alpha\in J}$ be any locally finite covering of M (this exists, since M is paracompact). Assume without loss of generality that $x \in M$ is covered by just two open sets U_1, U_2 (by local finiteness of $\{U_\alpha\}$ the procedure outlined below will just have to be repeated a finite number of times in the general case).

The strategy will be not to modify directly a partition of unity subordinate to $\{U_\alpha\}$, but rather to modify the covering itself in such a way that X will then be only covered by *one* open set W. A partition of unity subordinate to this new covering (which exists) will then have the desired property, since

$$\sum_\alpha \rho_\alpha(y) = 1, \quad \forall y \in M, \quad \operatorname{supp} \rho_\alpha \subset U_\alpha.$$

The crucial property we shall need of paracompact spaces is that they are normal, i.e., every two closed disjoint sets A, B can be separated by open disjoint sets O_A and O_B, formally

$$\forall A,B \subset M \text{ closed}, \quad A \cap B = \emptyset,$$

$$\exists O_A, O_B \text{ open such that:} \quad A \subset O_A, \quad B \subset O_B,$$

$$O_A \cap O_B = \emptyset.$$

Choose $V_x \subset U_1 \cap U_2$ such that $x \in V_x$ and V_x is open, and denote by $\overline{V}_x$ the closure of V_x. Assume now that W is an open set with the property $\overline{V}_x \subset W \subset U_1 \cap U_2$ (the existence of such a W will be shown below, using the normality of M). Let $\overline{V}_x^C$ be the complement of $\overline{V}_x$ in M; then $\overline{V}_x^C$ is open. Then one can convince oneself (by drawing pictures or by formal reasoning) that the union of the following three open sets

$$W, \quad W_1 := \overline{V}_x \cap U_1, \quad W_2 := \overline{V}_x^C \cap U_2,$$

is equal to the union of U_1 and U_2, and that x is only contained in W. A partition of unity subordinate to the new locally finite covering

$$\{U_\alpha, \alpha \in I, \alpha \neq 1,2; W, W_1, W_2\}$$

will then do the job.

Now we shall show that such a W can always be found. Since $A = \overline{V}_x$ and $B = (U_1 \cap U_2)^C$ are disjoint closed sets, we can—due to normality of M—find open sets O_A, O_B with

$$\begin{aligned} &\overline{V}_x \subset O_A, \\ &(U_1 \cap U_2)^C \subset O_B, \\ &O_A \cap O_B = \emptyset. \end{aligned} \tag{*}$$

Since $(U_1 \cap U_2)^C \subset O_B$, we have

$$O_B^C \subset U_1 \cap U_2, \tag{**}$$

and since O_A and O_B are disjoint, O_A is contained in the complement of O_B,

$$O_A \subset O_B^C. \tag{***}$$

Putting (*), (**), and (***) together, we find that

$$\overline{V}_x \subset O_A \subset U_1 \cap U_2.$$

Thus $O_A =: W$ is a possible choice. □

[1] J. Mickelsson, Phys. Rev. Lett. **57**, 2493 (1986); **55**, 2099 (1985).
[2] J. Mickelsson, Commun. Math. Phys. **110**, 173 (1987).
[3] R. Catenacci *et al.*, Phys. Lett. B **172**, 223 (1986).
[4] L. Bonora *et al.*, Commun. Math. Phys. **114**, 381 (1988).
[5] G. Falqui and C. Reina, Commun. Math. Phys. **102**, 503 (1985).
[6] M. Blau, Phys. Lett. B **209**, 503 (1988).
[7] M. F. Atiyah, *K-Theory* (Benjamin, New York, 1967).
[8] For a review cf. the contributions by R. Jackiw and B. Zumino in *Current Algebras and Anomalies*, edited by S. B. Treiman *et al.* (World Scientific, Singapore, 1985).
[9] J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971).
[10] A. Hattori and T. Yoshida, Jpn. J. Math. **2**, 13 (1976).
[11] H. Cartan and S. Eilenberg, *Homological Algebra* (Princeton U.P., Princeton, NJ, 1956).
[12] R. Bott and L. W. Tu, *Differential Forms in Algebraic Topology*, GTM 82 (Springer, New York, 1982).
[13] B. Zumino, Nucl. Phys. B **253**, 477 (1985).
[14] M. F. Atiyah and I. M. Singer, Proc. Natl. Acad. Sci. USA **81**, 2597 (1984).
[15] M. Martellini and C. Reina, Ann. Inst. H. Poincaré **43**, 443 (1985).
[16] R. Percacci, *Geometry of Nonlinear Field Theories* (World Scientific, Singapore, 1986), p. 58.

[17] This formula is to be read as follows: The h_α are the global equivariant functions of L_f restricted to U_α corresponding to the sections φ_α. Equation (4.6) shows that the new sections $D^\alpha h_\alpha$ (no sum) can be patched together consistently to give a new section $\nabla_x \varphi$ of L_f. This and the discussion following (4.6) show that this definition does not depend on the choice of local trivializations. I have, however, been unable to find an expression for ∇ without making recourse to such a choice. I am indebted to J. Mickelsson for critical remarks on this point.

[18] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I* (Wiley, New York, 1963).
[19] M. F. Atiyah and I. M. Singer, Ann. Math. **93**, 119 (1971).
[20] W. A. Bardeen, Phys. Rev. **184**, 1848 (1969).
[21] A. J. Niemi and G. W. Semenoff, Phys. Rev. Lett. **55**, 927 (1985).
[22] M. V. Berry, Proc. Roy. Soc. London, Ser. A **392**, 45 (1984).
[23] D. Quillen, Funk Anal. Prilozen **19**, 3 (1985).
[24] J. M. Bismut and D. S. Freed, Commun. Math. Phys. **106**, 159 (1986); **107**, 369 (1987).

Some classes of general solutions of the U(N) chiral σ models in two dimensions

Bernard Piette[a)]
Institut de Physique Theorique, Universite Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
Wojciech J. Zakrzewski
Department of Mathematical Sciences, University of Durham, Durham DH1 3LE, England

(Received 2 May 1988; accepted for publication 10 May 1989)

The factorization theorems of Uhlenbeck and Wood are used to derive various finite action solutions to the classical equations of motion of the Euclidean U(N) chiral model in two dimensions. They are obtained by adding a general basic uniton to solutions of the Grassmannian models. A brief comment is made on the properties of these solutions.

I. INTRODUCTION

In this paper we derive and study various finite action classical solutions of the U(N) chiral models in two Euclidean dimensions. These models have become increasingly popular in recent years as they possess nontrivial topological properties, which in turn are responsible for the existence of nontrivial solutions, and they appear to be two-dimensional analogs of physically relevant Yang–Mills theories. Moreover, they provide examples of harmonic maps and as such are also interesting from the mathematical point of view.

The models we are going to discuss are also called "principal chiral models," and are defined in terms of the Lagrangian density

$$L = \tfrac{1}{4}\mathrm{Tr}\,\partial_\mu Q^\dagger \partial_\mu Q, \tag{1.1}$$

where $Q^\dagger Q = QQ^\dagger = 1$.

The equations of motion (or strictly speaking the Euler–Lagrange equations, as we work in the Euclidean space) are given by

$$\partial_\mu (Q^\dagger \partial_\mu Q) = 0, \tag{1.2}$$

and to specify the problem completely we also have to state the boundary conditions satisfied by Q. The conditions we want to impose are those that come from the requirement of quantization in terms of part integrals—thus we require that

$$S = \int d^2x\, L < \infty .$$

The condition of the finiteness of the action effectively compactifies the two-dimensional Euclidean space, thus allowing us to take over the results derived in the case when the basic space is given by S^2. Moreover, this compactification introduces topology and is directly responsible for the discrete values of the action.

It has been known for some time that all solutions of Grassmannian models are also solutions of the chiral model [as the Grassmannian subspace is totally geodesic in U(N)][1]; at the same time not much has been known about other solutions. Recently, however, Uhlenbeck proved a very interesting factorization theorem.[2] Namely, she showed that all classical solutions of the chiral model are of the form

$$Q = K\prod_{i=1}^{k} (1 - 2R_i), \tag{1.3}$$

where k is some number (called by her the *uniton* number), K is a constant matrix, and R_i are projectors that satisfy some first-order differential equations. The theorem provides a convenient method of generating new solutions from old ones, the procedure that she called the *addition of a uniton*. Given a solution Q_0 of the model, she defines a uniton factor for this solution by $1 - 2R$, where R is a projector satisfying

$$\begin{aligned} &RA_-(1-R) = 0,\\ &(1-R)[\partial_- R + A_- R] = 0, \end{aligned} \tag{1.4}$$

where

$$A_\pm = \tfrac{1}{2}Q_0^\dagger \partial_\pm Q_0, \tag{1.5}$$

and $\partial_\pm$ denotes the derivative with respect to $x \pm iy$, respectively. We summarize her results by the following theorem.

Uhlenbeck's theorem: If Q_0 is a solution of the equations of motion of the U(N) chiral model—(1.2) and $1 - 2R$ is a uniton for this solution, then $Q = Q_0(1 - 2R)$ is another solution of this equation. Moreover, all solutions of the U(N) model can be constructed by adding less than N unitons to a constant solution.

If $Q_0 = K$, the equations reduce to $\partial_- RR = 0$, i.e., the equations for the instantons of the Grassmannian models. For $Q_0 \neq K$ we have more general solutions, which include noninstanton solutions of Grassmannian models and also non-Grassmannian solutions. It is important to note that the uniton number is, strictly speaking, not well defined. By this we mean that when we add a uniton to, say, a two-uniton solution, the resulting configuration may turn out to be equivalent to a one-uniton solution. For this reason, Uhlenbeck also defined the minimal uniton number as the minimal number of unitons that are needed to construct a given solution.

The main aspect of the Uhlenbeck construction is that it reduces the problem to having to solve a first-order nonlinear differential equation coupled with a nonlinear algebraic equation. This last equation admits two obvious solutions, namely, $RA_- = 0$ or $A_-(1-R) = 0$. Following Wood,[3] we will call their solutions, respectively, basic and antibasic uni-

[a)] Chercheur IISN, Belgium.

 0022-2488/89/102233-05$02.50

tons. Moreover, in his paper[3] Wood also managed to show that any uniton R can be decomposed into a product of basic unitons. Thus his theorem can be stated as follows.

Wood's theorem: Any uniton $1-2R$ corresponding to a given solution Q_0 can be factorized as

$$(1-2R)=(1-2R_1)(1-2R_2)\cdots(1-2R_k)$$

for some $k\leqslant N$, where $1-2R_1$ is a basic uniton for Q_0 and $1-2R_i$ are basic unitons for the solutions $Q_i = Q_0(1-2R_1)\cdots(1-2R_{i-1})$.

Therefore, we see that there are three types of unitons: the basic, the antibasic, and the remaining ones, which we shall call mixed. Moreover, each of the projectors appearing in these unitons can be of any rank.

Let us observe that when we try to add a basic or an antibasic uniton to a given solution, the Uhlenbeck equations (1.4) simplify. If we use I and H to describe, respectively, the image of A_- and A_+ and denote by $P(I)$ and $P(H)$ the corresponding projectors on these spaces, we can state the following proposition.

Proposition 1.1: For $1-2R$, a basic uniton corresponding to the solution Q_0, the Uhlenbeck equations (1.4) are equivalent to

$$RA_-=0,\qquad (1-R-P(I))\partial_-R=0. \tag{1.6}$$

Similarly, for the antibasic unitons $R=1-S$ and the Uhlenbeck equations (1.4) reduce to

$$A_-S=0,\qquad (1-S-P(H))\partial_+S=0. \tag{1.7}$$

Proof: The first equation in (1.6) is just the definition of a basic uniton. Using this definition and multiplying the second Uhlenbeck equation by $P(I)$ from the left we find

$$A_-R=-P(I)\partial_-R. \tag{1.8}$$

Substituting this result into the second Uhlenbeck equation we obtain the second equation (1.6). To prove the complete equivalence between Eqs. (1.6) and the Uhlenbeck equations (1.4) we have to show that if Eqs. (1.6) are satisfied, so are the Uhlenbeck equations. To prove this we consider the equation satisfied by $A_\pm$ (Ref. 2)

$$\partial_-A_+ + [A_-,A_+]=0,\quad A_-=-(A_+)^\dagger \tag{1.9}$$

and multiply it from the right by R. A simple algebraic manipulation then shows that if this equation is satisfied so is (1.8). Substituting this expression into the second equation (1.6) shows that the second Uhlenbeck equation is satisfied, thus completing the proof. The equivalence between the Uhlenbeck equations (1.4) and (1.7) can be proved in a very similar way. As an immediate consequence of Eq. (1.6) we see that $(1-P(I))$ and $P(H)$ are automatically projectors corresponding to, respectively, basic and antibasic unitons and that we can construct more general solutions of both types by considering projectors of smaller rank.

As we have seen, in order to add a uniton to a given solution Q_0, we have to compute the gauge field A_- corresponding to this solution. As a matter of fact, as Q_0 can be factorized, as in (1.3), it is quite easy to show that A_- is given by[4]

$$A_-=\sum_{i=1}^{k}\partial_-R_i. \tag{1.10}$$

In the sequel, we will use the following notation. For V, which is a matrix, we will denote by $P(V)$ the projector on the space it spans. When V is of maximal rank, this projector is given by

$$P(V)=V(V^\dagger V)^{-1}V^\dagger. \tag{1.11}$$

II. ONE-UNITON SOLUTIONS

As we have shown in Sec. I, the one-uniton solutions are of the form

$$Q=K(1-2R_1), \tag{2.1}$$

where R_1 satisfies

$$(1-R_1)\partial_-R_1=0. \tag{2.2}$$

These solutions are the so-called instanton solutions of Grassmannian models that have been known for some time.[5] The most general solutions for R_1 of this class are given by

$$R_1=P(F), \tag{2.3}$$

where F is a holomorphic matrix (i.e., whose entries are functions of only $x+iy$) of maximal rank.

It is important to note that a given one-uniton solution is not characterized by only one holomorphic matrix F. In fact, one can always take holomorphic linear combinations of the columns of F, i.e., replace F by $F'=FA$, where A is any invertible holomorphic square matrix of appropriate size, without altering the solution. From now on we will say that a projector R is holomorphic or antiholomorphic if it satisfies $\partial_-RR=0$ or $\partial_+RR=0$, respectively. As a consequence of the previous construction, holomorphic bases will play an important role in the sequel. For this reason, we will say that a set of rectangular matrices (of maximal rank)

$$V_1,V_2,\ldots,V_k \tag{2.4}$$

is an *orthogonal (anti)holomorphic basis sequence* if all the V_i are orthogonal to each other,

$$V_i^\dagger V_j=0,\quad i\neq j;\quad \sum_{i=1}^{k}P(V_i)=1, \tag{2.5}$$

and if $\mathbb{P}_j=\Sigma_{i=1}^{j}P(V_i)$ is (anti)holomorphic, and so it satisfies

$$(1-\mathbb{P}_j)\partial_\pm\mathbb{P}_j=0. \tag{2.6}$$

Moreover, we will say that this is an orthogonal (anti)holomorphic basis sequence of DZ type[6] if the V_i satisfy

$$V_i^\dagger\partial_\pm V_j=0 \tag{2.7}$$

for all i,j, such that $|i-j|\geqslant 2$. As a consequence of these properties, we can choose to normalize each V_i in such a way that

$$V_i=\Big(1-\sum_{j=1}^{i-1}P(V_j)\Big)F_i, \tag{2.8}$$

where F_i is some (anti)holomorphic matrix. This comes from the fact that all holomorphic projectors are of the type $P(F)$, where F is a holomorphic matrix.[7] Writing $\mathbb{P}_j=P(W_j)$ for some holomorphic W_j, we have by induction that $V_i = (1 - \mathbb{P}_{i-1}) W_i$. With this normalization, we have

$$V_i^\dagger \partial_- V_i = 0, \tag{2.9}$$

valid for all i. As this normalization will play an important role in what follows, we will call it the *natural (anti)holomorphic normalization*. When R is a holomorphic projector, then $1 - R$ is an antiholomorphic one, so that when read from the right to the left, the orthogonal holomorphic sequence becomes an antiholomorphic one. Moreover, when V is normalized in the natural holomorphic way, then $V(V^\dagger V)^{-1}$ has the natural antiholomorphic normalization. To add a uniton to the one-uniton solution, we will need the following proposition.

Proposition 2.1: When F is a holomorphic matrix of maximal rank and $R = P(F)$ corresponds to a one-uniton solution, then

$$((1 - P(I))F, I, H, G_R) \tag{2.10}$$

is an orthogonal holomorphic basis sequence of DZ type, where G_R is the orthogonal complement of the vectors (F,H).

Proof: In general, the matrix $(F, \partial_+ F)$ is not of maximal rank, but after some possible reordering of its columns we can always split $F = (F_1, F_2)$ into two parts, such that $(1 - P(F))\partial_+ F_1$ is of maximal rank. As a consequence we can write

$$\partial_+ F_2 = F_1 A + F_2 B + \partial_+ F_1 C \tag{2.11}$$

for some holomorphic matrices A, B, and C. So rather than taking F to construct R we can use $\check{F} = (F_1, G_L)$ where $G_L = F_2 - F_1 C$, and so $(1 - P(F))\partial_+ G_L = 0$. Now, using the fact that $P(F) = P(\check{F})$ and

$$A_- = \partial_- P(\check{F}) = \check{F}(\check{F}^\dagger \check{F})^{-1}((1 - P(\check{F}))\partial_+ F_1, 0)^\dagger, \tag{2.12}$$

we see that $G_L^\dagger A_- = (\partial_+ G_L)^\dagger (1 - P(\check{F})) = 0$ and that the rank of A_- is equal to the rank of $(1 - P(\check{F}))\partial_+ F_1$. As a consequence, we see that I is spanned by $(1 - P(G_L))F_1$ and we can write the sequence above as $(G_L, (1 - P(G_L))F_1, (1 - P(\check{F}))\partial_+ F_1 G_R)$. This sequence is obviously an orthogonal holomorphic basis sequence. Moreover, as $(1 - P(\check{F}))\partial_+ G_L = 0$ it is easy to check that this basis is of DZ type.

III. ADDING A BASIC UNITON

Wood's factorization theorem tells us that to construct all solutions of the U(N) model, all we have to do is to add successive basic unitons to the one-uniton solution. This appears to be a very difficult task. In our previous paper,[8] we reported the construction of all solutions of the U(3) and U(4) models. In the U(4) case we found that the construction of the general three-uniton configurations was rather difficult to perform. Nevertheless it is easy to observe that all solutions correspond to configurations that can be obtained by addition of one uniton to some Grassmannian solution. So rather than trying to construct two, three, and further general uniton solutions, we will restrict our construction to the addition of a general basic uniton to the already known general Grassmann solutions.[5,6]

The construction of these Grassmannian solutions is quite simple. We start by constructing an orthogonal holomorphic basis sequence of DZ type,

$$(Y_1, Y_2, Y_3, \ldots, Y_i, \ldots, Y_{2r}, Y_{2r+1}), \tag{3.1}$$

where the sets Y_1 and Y_{2r+1} may be empty. Defining $R_i = \Sigma_{j=1}^i P(Y_i)$, where $P(Y)$ is a projector (onto Y) and so satisfies $P^2 = P$, we know[5,6] that

$$Q = \prod_{i=1}^{2r} (1 - 2R_i) = (1 - 2R), \tag{3.2}$$

where

$$R = \sum_{i=1}^{r} P(Y_{2i}) \tag{3.3}$$

is a Grassmannian solution of the U(N) model (we shall prove this by induction as a by-product of the construction of more general solutions). Observe that when $r = 1$ and Y_1 is empty, we recover the instanton solutions described before.

We are now ready to add a basic uniton to these solutions, and as a by-product, prove once again (by induction) that (3.2) is a solution of the U(N) model. To proceed we observe that before we can solve Eq. (1.6), we must compute the A_- matrix for the solution (3.2). We find

$$A_- = \sum_{i=1}^{2r} \partial_- R_i. \tag{3.4}$$

Next we perform the "splitting," explained at the end of Sec. II and write, after an appropriate gauge transformation, $Y_i = (G_i, I_i)$, thus obtaining an orthogonal holomorphic basis sequence

$$(G_1, I_1, \ldots, G_{2r}, I_{2r}, G_{2r+1}), \tag{3.5}$$

which satisfies

$$G_i^\dagger \partial_- G_j = 0, \quad G_i^\dagger \partial_+ G_j = 0, \quad \text{for } i \neq j, \tag{3.6}$$

and

$$G_i^\dagger \partial_- I_j = 0, \quad I_i^\dagger \partial_+ G_j = 0, \quad \text{for } i \neq j. \tag{3.7}$$

To add a basic uniton to the solution (3.2) we have to solve Eqs. (1.6), the first of which can be solved by choosing a projector $\tilde{R}$ to be orthogonal to all I_i. To solve the second equation (1.6) we observe that we can rewrite it as $T \partial_- R = 0$ for some choice of T. Hence we split each nonempty set G_i into three parts,

$$G_i = (V_i, U_i, W_i), \tag{3.8}$$

such that the U_i's span the intersection of the two spaces onto which $\tilde{R}$ and T project and the V_i's (and, respectively, the W_i's) span the orthogonal complement of that intersection in the space on which $\tilde{R}$ (and, respectively, T) project. Thus we write, in full generality,

$$\tilde{R} = \sum_{i=1}^{2r+1} P(V_i) + P(V_M), \tag{3.9}$$

$$T = 1 - \tilde{R} - \sum_{i=1}^{2r} P(I_i) = \sum_{i=1}^{2r+1} P(W_i) + P(W_M), \tag{3.10}$$

where

$$V_M = \sum_{i=1}^{2r+1} U_i a_i,$$
$$W_M = \sum_{i=1}^{2r+1} U_i (U_i^\dagger U_i)^{-1} b_i, \tag{3.11}$$

and the a_i and b_i are some matrices that we can assume to be of maximal rank. In fact, had they not been of maximal rank, we would have found that one of the vectors spanned by the corresponding U_i would have to lie in V_i or W_i. We see that the second equation (1.6) reduces to

$$W_i^\dagger \partial_- V_j = 0 \tag{3.12}$$

for all i,j equal to $1,2,...,2r+1$ and M. Next we define

$$\mathbb{P}(V_i) = \sum_{j=1}^{i-1} P(Y_j) + P(V_i),$$
$$\mathbb{P}(U_i) = \mathbb{P}(V_i) + P(U_i), \tag{3.13}$$
$$\mathbb{P}(W_i) = \mathbb{P}(U_i) + P(W_i),$$

and observe that by construction

$$(1 - \mathbb{P}(W_i))\partial_- P(V_i) = 0,$$
$$(1 - \mathbb{P}(W_i))\partial_-(P(V_i) + P(U_i)) = 0. \tag{3.14}$$

However, we observe that (3.6) and (3.7) imply that

$$K_i^\dagger \partial_\pm \widetilde{K}_j = 0, \quad i \neq j, \tag{3.15}$$

where K and $\widetilde{K}$ stand for any V, W, or U. This means that (3.12) implies

$$U_i^\dagger \partial_- V_i = W_i^\dagger \partial_- V_i = W_i^\dagger \partial_- U_i = 0 \tag{3.16}$$

and we see that $\mathbb{P}(V_i)$, $\mathbb{P}(U_i)$, and $\mathbb{P}(W_i)$ are holomorphic, and that

$$(V_1, U_1, W_1, I_1, V_2, ..., W_{2r}, I_{2r}, V_{2r+1}, U_{2r+1}, W_{2r+1}) \tag{3.17}$$

is an orthogonal holomorphic basis sequence. Thus we are left with having to solve

$$W_M^\dagger \partial_- V_M = 0. \tag{3.18}$$

To proceed further we perform a gauge transformation, which brings the U_i to satisfy the condition of the natural holomorphic normalization, i.e., $U_i^\dagger \partial_- U_i = 0$ and compute

$$W_M^\dagger \partial_- V_M = \sum_{i,j} b_i^\dagger (U_i^\dagger U_i)^{-1} U_i^\dagger \partial_-(U_j a_j),$$
$$\sum_{i,j} b_i^\dagger \partial_- a_j = \sum_i b_i^\dagger \partial_- a_i. \tag{3.19}$$

Since the a_i have maximal rank, the above given expression vanishes if the a_i's are holomorphic.

If we now consider the particular example when all the sets U_i are empty, we find that our new solution takes the form

$$Q = (1 - 1R_0)(1 - 2R)$$
$$= \left(1 - 2\left(\sum_{i=1}^{r} P(\check{Y}_{2i}) + P(V_1)\right)\right), \tag{3.20}$$

where $\check{Y}_{2i} = (W_{2i}, I_{2i}, V_{2i+1})$. This solution is a Grassmannian solution computed from the orthogonal holomorphic basis sequence (OBS) $\widetilde{Y}_i = (W_i, I_i, V_{i+1})$, where $W_0 = I_0 = V_{2(r+1)} = 0$. Moreover, this new basis is easily shown to be of DZ type. Thus we see that the effect of our construction corresponds to the addition of one Y_1 in the sequence (3.1). We can now wonder whether all orthogonal holomorphic basis sequences (3.1) can be obtained by our construction. The answer to this question is positive as one can add successively all of the following basic unitons:

$$Q = \prod_{i=0}^{2r-1} (1 - 2R_i), \tag{3.21}$$

where

$$R_i = \sum_{j=1}^{2r-i} P(Y_j). \tag{3.22}$$

Let us notice that when the set Y_1 is empty, the last uniton in this sequence has to be dropped.

To construct all these solutions we need, it is easy to see that thus we have shown that our construction gives us all the Grassmannian solutions previously constructed by Din *et al.*[5] and by Sasaki.[9] To construct all these solutions we need to consider all holomorphic matrices F from which we must construct all possible orthogonal holomorphic basis sequences. Every step in such a construction adds a Y_i to the basis sequence, even though the space spanned by this sequence remains unchanged. The expressions (3.21) and (3.22) thus show how to construct all known Grassmannian solutions. We see that we have proved the following proposition.

Proposition 3.1: The most general basic uniton corresponding to the Grassmannian solution (3.2) is given by (3.9), where (3.17) is an orthogonal holomorphic basis sequence and V_M is given by (3.11), with all a_i being holomorphic.

On the other hand, when the sets U_i are not empty the new solution $Q = (1 - 2R_0)(1 - 2R)$ is non-Grassmannian as the projector $P(V_M)$, in general, does not commute with the projectors $P(U_i)$.

IV. AN EXPLICIT EXAMPLE

When we try to construct solutions of the Grassmannian models, it is convenient to use holomorphic bases.[5] Some of them have particularly useful properties that make them easy to use. Let us choose a holomorphic matrix F of maximal rank. Then we can define

$$P_+ F = \partial_+ F - F(F^\dagger F)^{-1} F^\dagger \partial_+ F \tag{4.1}$$

and use induction to define further vectors $P_+^k F$,

$$P_+^k F = P_+(P_+^{k-1} F). \tag{4.2}$$

This construction of our basis is equivalent to the Gram–Schmidt orthonormalization procedure of the sequence of analytic vectors $F, \partial_+ F, \partial_+^2 F,...$, thus showing that all $P_1^k F$, which correspond to different k's, are orthogonal to each other. Moreover, in addition they satisfy the following properties, which are essential in our construction[5]:

$$\partial_- P_+^k F = -P_+^{k-1} F \alpha_{k-1}^{-1} \alpha_k,$$
$$\partial_+ (P_+^k F \alpha_k^{-1}) = P_+^{k+1} F \alpha_k^{-1}, \tag{4.3}$$

where

$$\alpha_k = P_+^k F^\dagger P_+^k F. \tag{4.4}$$

As a consequence of the second equation, we see that the

$P^i_+ F$ have the natural holomorphic normalization and that

$$(F, P_+ F, P^2_+ F, \ldots, P^{2r}_+ F) \tag{4.5}$$

is an orthogonal holomorphic basis sequence. Defining projectors onto each of these vectors,

$$P_0 = P(F), \ldots, P_k = P(P^k_+ F), \tag{4.6}$$

we find that the sum of any of these projectors forms a solution of the type (3.2), in which the sum of consecutive projectors corresponds to a single projector $P(Y_i)$ of our previous construction. To make this clearer, let us consider, for example,

$$R = P_2 + P_7 + P_8 + P_9. \tag{4.7}$$

This solution corresponds to the choice $Y_1 = (F, P_+ F)$, $Y_2 = P^2_+ F$, $Y_3 = (P^3_+ F, P^4_+ F, P^5_+ F, P^6_+ F)$, $Y_4 = (P^7_+ F, P^8_+ F, P^9_+ F)$, and $Y_5 = (P^{10}_+ F, \ldots, P^{2r}_+ F)$. Using the relations (4.3) it is now easy to show that $I_1 = P_+ F$, $I_2 = P^2_+ F$, $I_3 = P^6_+ F$, and $I_4 = P^9_+ F$. An example of a basic uniton that we can add to this solution is given by

$$R = P_3 + P_{10} + P(V_M), \tag{4.8}$$

where

$$V_M = Fa + U_3 b + P^7_+ Fc, \tag{4.9}$$

and where a, b, and c are holomorphic matrices of maximal rank and U_3 is given by $U_3 = (1 - P_0 - P_1 - P_2 - P_3) \times (\partial^4_+ F, \partial^5_+ F)$. It is interesting to note that when we take the limit in which two of the three coefficients a, b, and c tend to zero, our solution goes over to a Grassmannian solution of the type (3.2). Thus we see that the non-Grassmannian $U(N)$ solutions we have computed have the surprising property that they interpolate between different Grassmannian solutions.

ACKNOWLEDGMENTS

Most of the work reported here was performed when one of us (B. P.) visted the University of Durham and another one (W. J. Z.) visted the Catholic University of Louvain-la-Neuve. They would like to thank the host institutions for their hospitality and J.-P. Antoine for making their collaboration possible. They would also like to thank M. Forger for many many constructive suggestions for the improvement of the manuscript.

B. P. would like to thank the Fonds National de la Recherche Scientifique, Belgium and the Royal Society, United Kingdom for the financial support of his visit to Durham organized under the European Science Exchange Programme scheme.

[1] H. Eichenherr and M. Forger, Nucl. Phys. B **164**, 528 (1980); erratum **282**, 745 (1987).
[2] K. Uhlenbeck, "Harmonic maps into Lie groups, Clasical Solutions of the Chiral Model," University of Chicago preprint, 1985, to be published in J. Differential Geom.
[3] J. C. Wood, "The explicit construction and parametrization of all harmonic maps from the two-sphere to the unitary group," Leeds University preprint No. 12, 1987.
[4] B. Piette, I. Stokoe, and W. J. Zakrzewski, Z. Phys. C **37**, 449 (1988).
[5] See, for example, W. J. Zakrzewski, J. Geom. Phys. **1**, 39 (1984).
[6] J.-P. Antoine and B. Piette, J. Math. Phys. **29**, 1687 (1988).
[7] A. J. Macfarlane, Phys. Lett. B **82**, 239 (1979).
[8] B. Piette and W. J. Zakrzewski, Nucl. Phys. B **300**, 207 (1988).
[9] R. Sasaki, Phys. Lett. B **130**, 69 (1984).

Local differential geometry of null curves in conformally flat space-time

H. Urbantke
Institut für Theoretische Physik, Universität Wien, Wien, Austria

(Received 6 April 1989; accepted for publication 17 May 1989)

The conformally invariant differential geometry of null curves in conformally flat space-times is given, using the six-vector formalism, which has generalizations to higher dimensions. This is then paralleled by a twistor description, with a twofold merit: first, sometimes the description is easier in twistor terms and sometimes in six-vector terms, which leads to a mutual enlightenment of both; and, second, the case of null curves in timelike pseudospheres or 2 + 1 Minkowski space could only be treated twistorially, making use of an invariant differential found by Fubini and Čech [*Geometria Proiettiva Differenziale* (Zanichelli, Bologna, 1926), Vol. 1; *Introduction à la Géométrie Projective Differentielle des Surfaces* (Gauthier–Villars, Paris, 1931)]. The result is the expected one: apart from the stated exceptional cases there is a conformally invariant parameter and two conformally invariant curvatures that, when specified in terms of this parameter, serve to characterize the curve up to conformal transformations.

I. INTRODUCTION

The local, Poincaré-invariant differential geometry of null curves in flat Minkowski space-time has been given by Bonnor[1]; we shall summarize it below by giving a two-component spinor version of it. According to it, nonstraight null curves are characterized by a Poincaré-invariant parameter and two "curvatures," which, when specified in terms of that parameter, fix the curve up to Poincaré transformations. Now the concept of a null curve is a conformally invariant one, so that it is natural to ask for a similar but conformally invariant treatment. The purpose of the present paper is to provide such a treatment. All necessary ingredients are in the literature; it is only necessary to put them together.

The result is that, "in general," a null curve "locally" has a conformally invariant parameter and two "curvatures," which fix the curve up to conformal transformations when specified as functions of that parameter. Here "local" means that the parameter intervals to be considered must not include certain exceptional points, while "in general" means that curves consisting of exceptional points only must receive a separate treatment.

The exceptional types of curves comprise (1) straight null curves ("null lines"), (2) conformal cubic null helices, and (3) curves on timelike hyperspheres. (In complexified space or for signature $++--$, one would have to add curves in totally isotropic two-planes; but we shall essentially stick to the real domain and signature $+---$.) For the first two cases there are no invariants and no invariant parameter, because the conformal group acts transitively on the sets (1) and (2), while the subgroup keeping any individual curve of these sets fixed acts multiply transitively locally. For curves of type (3) but other than (2), there is an invariant parameter and just one invariant to conformally characterize the curve.

These were the methods we employed: In part we were able to generalize and adapt the "$n+2$ vector formalism" of conformally Euclidean geometry[2] to the "null case" of pseudo-Euclidean geometry. On the other hand, the twistor correspondence[3] was also used and was of essential help to us in dealing with the exceptional cases; it enabled us to use results from textbooks (e.g., Ref. 4) on projective differential geometry. (In order to be brief, we have avoided more detailed geometric interpretations, although some motivations then remain unclear—the interested reader is referred to these textbooks.) The only adaption that had to be made here was to take care of the reality structure peculiar to the signature $+---$.

The plan of the paper is as follows. In Sec. II we summarize the formalism required and fix the notation; we also include a spinor version of Bonnor's theory. In Sec. III, the theory of null curves is developed using the six-vector formalism. In Sec. IV, the twistor correspondence for null curves is described. In Sec. V, a few concluding remarks concerning the relation between the formalism of Sec. II A and the manifestly conformally invariant formalism of the later sections are made. Not all calculations are given in detail, because they involve only elementary linear algebra and differential calculus in one variable.

II. SUMMARY OF FORMALISM

A. Bonnor's theory[1]

Consider, in flat Minkowski space $\mathbf{M}_4$ (signature $+---$; scalar product of four-vectors $x, y, \ldots$ denoted by $x\cdot y$; $x^2 := x\cdot x$) a null curve $x(\lambda)$ parametrized by λ. Then $x' := dx/d\lambda$ by definition satisfies $x'^2 = 0$, and we will assume x' to be future directed. It follows that $x'\cdot x'' = 0$, so x'' has to be spacelike or proportional to x'. Excluding the latter case, which has the direction of x' fixed, corresponding to a straight null line, we have, therefore, $x''^2 < 0$. Now changing the parameter as $\lambda = \lambda(\hat\lambda)$ monotonically, one finds

$$\left(\frac{d^2x}{d\hat\lambda^2}\right)^2 = \left(\frac{d^2x}{d\lambda^2}\right)^2 \left(\frac{d\lambda}{d\hat\lambda}\right)^4;$$

i.e., $(-\frac{1}{2}x''^2(\lambda))^{1/4}\, d\lambda$ is an invariant differential, whose integral can be used as a new, "natural" parameter, intrinsically defined by the curve. We assume that λ has already been chosen to be this parameter (whose "physical" interpretation was given by Synge[1]); we therefore have $x''^2 = -2$.

The pair x', x'' spans a null flag,[5] and, identifying the Minkowski vector space with the set of Hermitian elements of $\mathbf{S}\otimes\overline{\mathbf{S}}$, where $\mathbf{S}$ is complex-two-dimensional spin space and $\overline{\mathbf{S}}$ is the complex-conjugate space, we can choose a spinor $\xi(\lambda)\in \mathbf{S}$ such that $x'(\lambda) = \xi(\lambda)\otimes\bar{\xi}(\lambda)$ and such that the flag plane is given by $x'\wedge x''$. (See, e.g., Ref. 6 for the use of exterior algebra in describing the geometry of subspaces of a vector space.) Only a sign in ξ remains undetermined. Differentiating, we obtain $x'' = \xi'\otimes\bar{\xi} + \xi\otimes\bar{\xi}$. We may then use ξ together with ξ' as a basis for $\mathbf{S}$ (otherwise $\xi'\propto\xi \Rightarrow x''\propto x'$, which was excluded before) and, in fact, as a spin frame in the sense that $\varepsilon(\xi,\xi') = 1$, where ϵ is the spin "metric": this is consistent with $x''^2 = -2$. Differentiating again, we find $\varepsilon(\xi,\xi'') = 0$, so there is a complex-valued function I such that $\xi'' = I\xi$. This corresponds to Frenet's formulas, and Re I and Im I, considered as functions of the natural parameter λ, will fix the curve up to Poincaré transformations. They are the invariant curvatures of the curve and determine the infinitesimal null rotation that (ξ,ξ') undergoes when one proceeds from a point on the curve to its infinitesimal neighbor. The significance of the invariant condition Im $I = 0$ is easy to state: defining $m := \xi\otimes\bar{\xi}'$, we find $(m-\bar{m})' = (\bar{I}-I)x'$ and $i(m-\bar{m})\cdot x' = 0$; thus the condition Im $I = 0$ implies that $i(m-\bar{m})$ is a constant real vector and $x(\lambda)$ belongs to the fixed timelike hyperplane $i(m-\bar{m})\cdot x = \text{const}$. The subcase $I = 0$ will also be of interest later. Here we have $\xi(\lambda) = \xi_0 + \xi_0'\lambda$ and therefore

$$x(\lambda) = x_0 + x_0'\lambda + (x_0''/2)\lambda^2 + (x_0'''/6)\lambda^3,$$

which we shall refer to as the *cubic null helix*. It is the trajectory of the Killing field

$$(\tfrac{1}{2}x_0'''\cdot(x-x_0))x'' - (x_0''\cdot(x-x_0))\tfrac{1}{2}x_0''' + x_0'$$

passing through x_0, and λ is the canonical group parameter along the flow of this field which consists of simultaneous infinitesimal null rotations and null translations. We shall see later that this curve admits a larger group of conformal transformations.

B. Six-vector formalism

This formalism is used to describe conformally compactified Minkowski space. One considers a real six-dimensional vector space $\mathbf{V}_6$ (which we propose to call "sixtor space") together with a nonsingular symmetric bilinear form $X\cdot Y$ of signature $+ - - - + -$ on it, or rather the corresponding five-dimensional projective space $P(\mathbf{V}_6) =: \mathbf{P}_5$ whose points are the one-dimensional subspaces $\mathbf{R}X$ of $\mathbf{V}_6$, together with the quadric $\mathbf{Q}_4$ given by $X^2 := X\cdot X = 0$. Singling out two points on $\mathbf{Q}_4$ as representing infinity and the origin, Minkowski space can be imbedded into $\mathbf{Q}_4$ (conformally compactified Minkowski space) according to

$$x \to \mathbf{R}X = \mathbf{R}(x,(1-x^2)/2,\ (1+x^2)/2),$$

the scalar square being

$$X^2 := x^2 + ((1-x^2)/2)^2 - ((1+x^2)/2)^2.$$

From this, an interpretation of the points in $\mathbf{P}_5$ not on $\mathbf{Q}_4$ can also be given: for $H\in\mathbf{V}_6$, consider all $X\in\mathbf{V}_6$ satisfying $H\cdot X = 0$; those points of this hyperplane that belong to the imbedded Minkowski space form a (pseudo) hypersphere there if $H^2\neq 0$. A conformally invariant concept of central importance in the sequel are the null *lines* (light rays) in $\mathbf{M}_4$. Their description in the $\mathbf{V}_6$ formalism is given by two-dimensional subspaces of $\mathbf{V}_6$ that are totally isotropic, i.e., all its vectors X have $X^2 = 0$; projectively this means that null lines are given by lines of $\mathbf{P}_5$ that are entirely contained in $\mathbf{Q}_4$. (There are no higher-dimensional totally isotropic subspaces in the real $\mathbf{V}_6$.) For more geometrical details see Ref. 3.

We shall also employ a normalized determinant function $\widetilde{E}$ on $\mathbf{V}_6$, i.e., one whose value is ± 1 on any orthonormal basis. It can be used in the standard manner to define the Hodge $*$ operator in the exterior algebra $\Lambda\mathbf{V}_6$ over $\mathbf{V}_6$.

C. Twistor formalism

While the six-vector formalism can be immediately generalized to arbitrary dimensions and signatures, the twistor formalism is specifically tied to conformally flat four-spaces, the various signatures being related to rather different objects in twistor space. In addition to Ref. 3, we will rely here on Ref. 7. The basic ingredient is a complex four-dimensional vector space $\mathbf{T}$ (twistor space) together with a determinant function $\tilde{e}\in\Lambda^4\tilde{\mathbf{T}}$ (where $\tilde{\mathbf{T}}$ is the dual space of $\mathbf{T}$). Now $\tilde{e}$ and its dual $e\in\Lambda^4\mathbf{T}$ (called ω in Ref. 7) serve to define the dualization maps $\Lambda^p\mathbf{T}\to\Lambda^{4-p}\tilde{\mathbf{T}}$, $\Lambda^p\tilde{\mathbf{T}}\to\Lambda^{4-p}\mathbf{T}$ in the usual manner; they will be indicated by $_*$ (resp. $*$). The slight differences among various conventions involved in these definitions will be rather unimportant here since we are interested in most quantities only up to a nonvanishing factor. Namely, what matters here is the complex projective space $P(\mathbf{T})$ and its relation to conformal geometry. The basic step is to form $\Lambda^2\mathbf{T}$, a complex six-dimensional vector space which already comes with a nondegenerate symmetric bilinear form given by $F\cdot G := \langle {}_*F|G\rangle$, where $F, G\in\Lambda^2\mathbf{T}$ and $\langle\cdot|\cdot\rangle$ means the scalar product between the dual spaces $\Lambda^2\mathbf{T}$ and $\Lambda^2\tilde{\mathbf{T}}$. The Plücker condition $F^2 := F\cdot F = 0$ is known to be necessary and sufficient for the bivector F to be simple (or decomposable), i.e., of the form $t_1\wedge t_2$, where $t_1, t_2\in\mathbf{T}$ are not unique—rather, the subspace $\langle t_1, t_2\rangle$ spanned by them determines and is determined by the set of scalar multiples of F. Geometrically, the point $\mathbf{C}F\in P(\wedge^2\mathbf{T})$ then represents the straight line in $P(\mathbf{T})$ joining $\mathbf{C}t_1$ and $\mathbf{C}t_2$. If G is simple as well, $F\cdot G = 0$ means that the corresponding lines meet at a point. The idea is now to identify the real vector space $\mathbf{V}_6$ from Sec. II B with a suitable "real part" of $\Lambda^2\mathbf{T}$ such that the restriction of $F\cdot G$ to it yields a real-valued, nondegenerate form of the signature required. We now sketch the procedure for all signatures.

As one way to proceed, one could think of actually taking $\mathbf{T}$ as a real vector space and $\tilde{e}$ as real-valued; then $\Lambda^2\mathbf{T}$ is real. However, $F\to F\cdot F =: F^2$ turns out to have signature $+ + + - - -$. So the idea to identify "reality" already on the level of $\mathbf{T}$ itself fails for the other signatures.

A slight modification of this approach comes to mind next. The same thing as taking $\mathbf{T}$ to be real is to single out a real part of a complex $\mathbf{T}$ by taking its elements that are invariant under an "anti-involution of the *first* kind," i.e., an *antilinear* map $\mathscr{C}:\mathbf{T}\to\mathbf{T}$ satisfying $\mathscr{C}^2 = \mathrm{id}_{\mathbf{T}}$. Then the exterior square $\mathscr{C}\wedge\mathscr{C}$ is an antiinvolution of the first kind on $\Lambda^2\mathbf{T}$

that can be used to define a real part there. The modification now consists of taking instead of $\mathscr{C}$ an anti-involution $\mathscr{T}$ of the *second* kind: $\mathscr{T}^2 = -\mathrm{id}_{\mathbf{T}}$ (a "quaternionic structure" on **T**—see Ref 8). Then again $\mathscr{T} \wedge \mathscr{T}$ is an anti-involution of the *first* kind on $\Lambda^2\mathbf{T}$, whose invariant elements form a real six-dimensional vector space. By choosing $\tilde{e}$ suitably in relation to $\mathscr{T}$, $F\cdot G$ is real on these real elements; however, this time the signature of $F \to F^2$ turns out to be $\pm(+ + + + + -)$, as would be appropriate for the conformal geometry of Euclidean four-space. (This is why quaternions occur in the solution of the Euclidean instanton problem.[9]

The remaining signatures are obtained by considering an antilinear map $h: \mathbf{T} \to \tilde{\mathbf{T}}$; coupling its exterior square with the dualization * again leads to an antilinear map $\Lambda^2\mathbf{T} \to \Lambda^2\mathbf{T}$. This will be involutive (i.e., have square = multiple of identity) if h is Hermitian, i.e., if the corresponding sesquilinear form $(t_1|t_2) := \langle h(t_1)|t_2\rangle$ is Hermitian; and it is of the first (second) kind only if $(\cdot|\cdot)$ contains an even (odd) number of minuses in its signature. Now already for reasons of compactness of the invariance groups involved, the definite Hermitian case corresponds to signature $\pm(+ + + + + +)$ on the corresponding real part of $\Lambda^2\mathbf{T}$; but this signature does not occur in the six-vector formalism, where there is at least one minus sign.

The remaining possibility, a Hermitian form of signature $+ + - -$, is at the basis of the twistor formalism for real conformally flat space-time. As described in detail in Ref. 7, the anti-involution squares to identity iff the Hermitian form and the determinant function on **T** are suitably normalized relative to each other. Elements invariant under the anti-involution are called "self-adjoint" in Ref. 7, but we will call the corresponding points in projective space "real points." [Strictly speaking, if we now identify $\mathbf{V}_6$ of Sec. II B with this real part of $\Lambda^2\mathbf{T}$, then the *real* projective space $P(\mathbf{V}_6) = \{\mathbf{R}v | v \in \mathbf{V}_6\}$ can be *injected* into the complex projective space $P(\Lambda^2\mathbf{T})$ by assigning $\mathbf{R}v \to \mathbf{C}v$.

We have already noted that points $\mathbf{C}F \in P(\Lambda^2\mathbf{T})$ with $F^2 = 0$ correspond to lines in $P(\mathbf{T})$. We now add that to *real* points $\mathbf{R}F \in \mathbf{Q}_4$ there correspond lines of $P(\mathbf{T})$ that are completely contained in the "twistor null surface" $P\mathbf{N}$ given by $(t|t) = 0$; i.e., if F is simple and real, then $F = t_1 \wedge t_2$ where $(t_1|t_1) = (t_1|t_2) = (t_2|t_2) = 0$ and hence $(t|t) = 0$ for all t from the subspace $\langle t_1, t_2\rangle_{\mathbf{C}}$, which is then said to be totally isotropic. This follows by contracting the reality condition $h^{\wedge 2}F = {*F}$ or $h(t_1) \wedge h(t_2) = {*(t_1 \wedge t_2)}$ with t_1 and t_2, which are linearly independent. Conversely, if t_1 and t_2 span a totally isotropic two-space of **T**, then a suitable complex multiple of $t_1 \wedge t_2$ will satisfy the reality condition.[7]

This concludes our preparations.

III. NULL CURVES IN THE SIX-VECTOR FORMALISM

A. The conformal arc

A null curve in $\mathbf{M}_4$ has, be definition, null lines as its tangents. In the $\mathbf{P}_5$ picture, we have a curve on $\mathbf{Q}_4$ whose tangents, in the $\mathbf{P}_5$ sense, are contained in $\mathbf{Q}_4$. Thus if $X(\lambda)$ is a parametric description of the curve, the tangent is spanned by $X(\lambda)$ and $X'(\lambda)$ and we must have $X^2 = X\cdot X' = X'^2 = 0$; in the real domain we must have $X''^2 < 0$, however, if we exclude null straight lines from the class of curves to be considered.

It would now be tempting to fix the parameter just as in Sec. II A, but one has to remember that, in the six-formalism, there is the additional freedom of a λ-dependent scale factor $X(\lambda) \to f(\lambda)X(\lambda)$ that could spoil the parameter choice. Rather, we fix the scale by considering the quantity $\hat{X} := (-\frac{1}{2}X'')^{-1/2}X$, which is independent of the original scale and has $\hat{X}''^2 = -2$, and look for another possibility to fix the parameter invariantly. We assume that the scale has been chosen relative to the parameter λ in the manner just described, but we omit the hat on X. If we go over to another parameter $\hat{\lambda}$, we must rescale $X(\lambda(\hat{\lambda}))$ as

$$\hat{X}(\hat{\lambda}) := X(\lambda(\hat{\lambda}))\left(\frac{d\lambda}{d\hat{\lambda}}\right)^{-2}$$

in order that

$$\left(\frac{d^2\hat{X}}{d\hat{\lambda}^2}\right)^2 = -2;$$

we say that X has parameter weight -2. From this it follows that $X \wedge X'$, $X \wedge X' \wedge X''$,... have definite weights [namely, $-2 + (-1) = -3$, $-2 + (-1) + 0 = -3$,...]; in particular, the Wronskian $E(X, X', X'', X''', X^{\mathrm{IV}}, X^{\mathrm{V}})$ has weight 3, so that

$$|\tilde{E}(X,...,X^{V})|^{1/3}\, d\lambda = : d\sigma$$

is an invariant differential, while sign $\hat{E}(X,..., \mathrm{X}^{\mathrm{V}})$ is invariant under parameter changes where $d\lambda/d\hat{\lambda} > 0$. (We are restricting here to the connected component of the conformal group.)

There are now two cases: (1) If $d\sigma \neq 0$, we can use its integral σ as a new, conformally invariant parameter, defined up to an additive constant. It is called the *conformal arc* and can be used on open segments of the curve without zeros of $d\sigma$. (2) If $d\sigma = 0$, $X, X',..., X^V$ are linearly dependent, and we have to turn to the products mentioned above whose vanishing or nonvanishing has invariant meaning. Now generally $(X \wedge X' \wedge \cdots \wedge X^{(p)} = 0$ but $X \wedge \cdots \wedge X^{(p-1)} \neq 0$ on some open segment) $\Leftrightarrow (X \wedge X' \wedge \cdots \wedge X^{(p-1)})' \propto X \wedge \cdots \wedge X^{(p-1)}$, i.e., the $(p-1)$-plane spanned by $\mathbf{R}X$,..., $\mathbf{R}X^{(p-1)}$ in $\mathbf{P}_5$ does not change with λ and contains the segment. Then $p = 1$ means degeneration of the curve to a point, $p = 2$ corresponds to a null line, which we have already excluded; $p = 3$ does not allow anything new in the real case: this two-plane in $\mathbf{P}_5$ intersects $\mathbf{Q}_4$ at a quadric curve that degenerates into a null line (from $X^2 = X'^2 = 0$ it follows that $X\cdot X' = X'\cdot X'' = X\cdot X'' = 0$, so for any $Y = aX_0 + bX_0' + cX_0''$ we get $Y^2 = -2c^2$, so that $c = 0$ for $Y \in \mathbf{Q}_4$). A similar thing happens for $p = 4$: the three-plane in $\mathbf{P}_5$ intersects $\mathbf{Q}_4$ at a quadric two-surface that is (since $X''^2 \neq 0$ but $X\cdot X''' = 0$ in addition to the vanishing scalar products above) a two-dimensional null cone; there are obviously no null curves on it other than its generating null lines. Thus we see that in the real case of signature $\pm(+ - - - + -)$ only $p = 5$ remains, where the curve is contained in a hyperplane of $\mathbf{P}_5$; its intersection with $\mathbf{Q}_4$ represents a (pseudo) hypersphere in $\mathbf{M}_4$ as we have already mentioned. [One has only to note that its normal

$H: = *(X \wedge X' \wedge \cdots \wedge X^{IV})$ has $H^2 =$ the Gram determinant of $X,..., X^{IV} = (X''^2)^5 \neq 0$ (see Table I).]

Now a hypersphere in $\mathbf{M}_4$ containing real nonstraight null curves is conformally the same thing as a timelike hyperplane or Minkowskian three-space $\mathbf{M}_3$. What the $\mathbf{V}_6$ formalism is to $\mathbf{M}_4$, a $\mathbf{V}_5$ formalism is to $\mathbf{M}_3$ (the projective version of $\mathbf{V}_5 \subset \mathbf{V}_6$ is the hyperplane in $\mathbf{P}_5$ just mentioned). Hence we can generalize the considerations above to $\mathbf{M}_n$, using a $\mathbf{V}_N = \mathbf{V}_{n+2}$ formalism, and then specialize to $n = 3$, $N = 5$. The trouble, however, is that, in $\mathbf{V}_N$, $\tilde{E}(X,X',...,X^{(N-1)})$ has weight $3 + 4 + ... + N - 3 = (N/2)(N - 5)$, so that the construction of the invariant parameter σ does not work in the case $N = 5$ $(n = 3)$ only! Thus the analysis of null curves in $\mathbf{M}_n$ [$n \geqslant 4$, signature $\pm(+ - \cdots -)$] can be carried out in complete analogy to $n = 4$, and we still need to know how to continue with $n = 3$.

It is at this point where the twistor formalism can carry us further. Of course, it would have been possible to take the result from there, translate it back into the $\mathbf{V}_6$ formalism, and present it as a direct insight; but that would not be fair. So we admit the lack of direct insight and present the twistor approach in Sec. IV. It will provide us with an invariant parameter even in the $n = 3$ case, except for the conformal cubic null helices that are, by definition, all conformally equivalent to the one described in Sec. II A. One can therefore proceed essentially as with the parameter σ.

B. The conformal curvatures

Assume we can use σ. Then $X,...,X^V$ are linearly independent and can be used as a basis in $\mathbf{V}_6$. Hence we may expand X^{VI} in terms of them. The expansion coefficients, expressible in terms of scalar products, are conformal invariants, if σ is used as a parameter. By differentiation, we get similar expansions for $X^{(p)}$, $p \geqslant 6$. To find a list of independent invariants, we first complete a table of scalar products $X^{(p)} \cdot X^{(q)}$, putting $X'''^2 = : 2K$, $(X^{IV})^2 = : 2J$, $(X^V)^2 = :L$. Differentiating these equations of definition and also the relations $X^2 = X'^2 = 0$, $X''^2 = -2$, we find the results in Table I. As the fundamental invariants we can take K, J; all others are expressible in terms of them. For L this is so because we not only have $X'' = -2$ by our choice of scaling but also $\tilde{E}(X,...,X^V) = 1$ by choosing $\lambda = \pm\sigma$ as a parameter. This entails $1 = [\tilde{E}(X,...,X^V)]^2 =$ det (Table I), but we have

$$\tfrac{1}{16}\det(\text{Table I}) = -2L - 9K'^2 + 4K \times (K'' - K^2 - 2J),$$

and thus we find L in terms of K and J and their derivatives.

TABLE I. The scalar products $X^{(p)} \cdot X^{(q)}$.

q/p	0	1	2	3	4	5
0	0	0	0	0	-2	0
1	0	0	0	2	0	$2K$
2	0	0	-2	0	$-2K$	$-3K'$
3	0	2	0	$2K$	K'	$K''-2J$
4	-2	0	$-2K$	K'	$2J$	J'
5	0	$2K$	$-3K'$	$K''-2J$	J'	L

(Also in the case $d\sigma = 0$ [where det (Table I) = 0], L is expressible in terms of K and J, but the choice of fundamental invariants runs differently, as we shall see in Sec. IV.)

On differentiating the last column of Table I, we obtain the scalar products $X^{(p)} \cdot X^{VI}$ and from them the coefficients in the expansion of X^{VI} in terms of $X,...,X^V$. [Actually, from $\tilde{E}(X,...,X^V) = 1$ or 0, it follows that $\tilde{E}(X,..., X^{IV}, X^{VI}) = 0$ so that the expansion coefficient in front of X^V vanishes.] If $K(\sigma)$ and $J(\sigma)$ are given functions of σ, the expansion of X^{VI} gives a system of differential equations for X that can be solved and determines the curve uniquely up to conformal transformations.

From this it follows, for example, that a $d\sigma \neq 0$ curve that admits a one-parameter group of conformal transformations has $K =$ const, $J =$ const, and conversely.

For geometrical purposes it would be more appropriate to use other bases than the ones given by $X, X',...,X^V$, but we do not go into this here.

IV. NULL CURVES IN THE TWISTOR FORMALISM

A. Generalities

A curve in $\mathbf{M}_4$ or $\mathbf{Q}_4$ corresponds, in the twistor picture, to a family of straight lines in $P\mathbf{N} \subset P(\mathbf{T})$ parametrized by one real parameter λ. If the curve is null, (infinitesimally) consecutive points have null separation; hence consecutive lines of the family intersect. These points $\mathbf{C}t(\lambda)$ [where $t(\lambda) \in \mathbf{T}$] of intersection will, in general, form a curve in $P(\mathbf{T})$. By this term we mean a one real parameter set in $P(\mathbf{T})$, formerly sometimes called a "thread",[10] in contradiction to a (locally) holomorphic curve which is a one-complex parameter = two-real parameter submanifold of $P(\mathbf{T})$. Of course, if there exists a real analytic parametrization of the thread, we can imagine a local complex thickening by analytic continuation to complex values of the parameter. We shall apply the latter point of view in one case only, however: to the straight lines of $P(\mathbf{T})$ that appear, in particular, as tangents of our "threads." As a result of the lack of a concept of real points in $P(\mathbf{T})$, the only meaningful concept of a line tangent to the curve $\{\mathbf{C}t(\lambda)\}$ in $P(\mathbf{T})$ is the complex line joining $\mathbf{C}t(\lambda)$ and $\mathbf{C}\, dt(\lambda)/d\lambda$, formally described by $\mathbf{C}t \wedge t'$.

It is intuitively clear that the tangents of the curve obtained in $P(\mathbf{T})$ are just the lines of the family we started with, and since they are to be contained in $P\mathbf{N}$, our curve must also be contained in $P\mathbf{N}$; i.e., we must have $(t|t) = 0$, $(t|t') = 0$, $(t'|t') = 0$.

More formally, if we start from a curve $\mathbf{C}t(\lambda)$ in $P(\mathbf{T})$, we can form its tangents $\mathbf{C}X$, $X: = t \wedge t'$, which form a curve on $\mathbf{Q}_4^c \subset P(\Lambda^2\mathbf{T})$, where $\mathbf{Q}_4^c$ is the complexification of the $\mathbf{Q}_4$ introduced earlier. We then have $X' = t \wedge t''$, so $X^2 = X \cdot X' = X^2 = 0$, showing that $\mathbf{C}X(\lambda)$ is a null curve. This description of null curves has been known for some time. The author has encountered it on a very different occasion,[11] and it was studied more directly by Shaw.[12] If, in addition, $(t|t) = (t|t') = (t'|t') = 0$, the null curve will be real in our sense. Conversely, starting from a null curve $\mathbf{C}X(\lambda)$, from $X^2 = X \cdot X' = X'^2 = 0$ we conclude that X, X' are representable as $X = t \wedge u$, $X' = t \wedge v$, where t corresponds to the point of intersection of the lines corresponding to X and X' and which do intersect. Thus $0 = t\wedge X' = t\wedge(t'\wedge u + t\wedge u') = t\wedge t'\wedge u$. If t and t' are linearly independent in the parameter interval considered, $\{\mathbf{C}t(\lambda)\}$ will be a curve, and since now $u\in\langle t,t'\rangle$, we have $X\propto t\wedge t'$ ($u\propto t$ is excluded by $X\neq 0$), showing that $\mathbf{C}X(\lambda)$ are the tangents of that curve. The remaining case $t'\propto t$, where $\mathbf{C}t(\lambda)$ is one single point and the lines $\mathbf{C}X(\lambda)$ form a cone through it, can be excluded if we are to describe a real null curve other than a straight line on $\mathbf{Q}_4$. This is because then $\mathbf{C}t\in P\mathbf{N}$, where t can be assumed constant, $X(\lambda)\propto t\wedge u(\lambda)$, $(t|u) = (u|u) = 0$, but the lines on $P\mathbf{N}$ through a fixed point $\mathbf{C}t$, contained in the plane $(t|u) = 0$, correspond to the points of a real null line on $\mathbf{Q}_4$ only (see Ref. 3). (If we were to consider $\mathbf{M}_4$ with signature $+\ +\ -\ -$ or complexified $\mathbf{M}_4$, these cones would be relevant and would correspond to curves in totally null two-planes of $\mathbf{M}_4$ or $\mathbf{Q}_4$.)

Similarly, we may exclude planar curves in $P(\mathbf{T})$: since their tangents would have to belong to $P\mathbf{N}$, the whole plane containing the curve would have to belong to $P\mathbf{N}$, which is impossible on dimensional grounds. (Again, for signature $+\ +\ -\ -$ or complexified $\mathbf{M}_4$, planar curves would be relevant, representing curves in totally null two-planes of the second type in $\mathbf{M}_4$ or $\mathbf{Q}_4$.)

The net result of the preceding discussion now is that, from the twistor point of view, the geometry of real, nonstraight null curves has been translated to the geometry of *twisted* (i.e., nonplanar) curves in $P\mathbf{N}\subset P(\mathbf{T})$. Here we are in the lucky position of being able to copy the theory to a large extent from existing work on projective differential geometry[4]; we only have to add the restriction that the curve and its tangents belong to $P\mathbf{N}$. The essence of this theory, perhaps not in its most elegant form (which would be useful to read off more geometrical details), is as follows.

B. Invariant parameters and differential invariants

Let $\{\mathbf{C}t(\lambda)\}\subset P(\mathbf{T})$ be a twisted curve. The condition for this is, in analogy to what has been said in Sec. III A, that $t\wedge t'\wedge t''\wedge t'''\neq 0$ or

$$\tilde{e}(t,t',t'',t''')\equiv{}_*(t\wedge t'\wedge t''\wedge t''')\neq 0.$$

[Note, however, that here we are working in the algebra $\Lambda\mathbf{T}$ whereas there the relevant algebra was $\Lambda\mathbf{V}_6$. Some confusion might arise now because here we consider $\mathbf{V}_6$ as a subset of $\Lambda^2\mathbf{T}\subset\Lambda\mathbf{T}$, and we have tried to avoid confusion between $\Lambda(\Lambda^2\mathbf{T})$ and $\Lambda\mathbf{T}$ by using the symbols $\bigwedge$ and $\wedge$ for multiplication in the former and the latter, respectively. Thus, for $F, G\in\Lambda^2\,\mathbf{T}$, we have $F\bigwedge G = -G\bigwedge F$ while $F\wedge G = G\wedge F$!] This enables us to use t, t', t'', t''' as a basis in $\mathbf{T}$, and also to fix the freedom of complex scale factors, $t(\lambda)\to f(\lambda)t(\lambda)$, by requiring

$$\tilde{e}(t,t',t'',t''') = -1.$$

(The minus sign is introduced to cope with the reality condition and signature employed.) This can be achieved by taking

$$\hat{t}(\lambda) := (-\tilde{e}(t,\dots,t'''))^{-1/4}t(\lambda)$$

instead of $t(\lambda)$; the result is independent of the original scaling, and we will omit the hat from now on, understanding that this step has been done. (Geometrically, this gives no restriction if we consider the connected component of the conformal and projective group only.) Similarly to Sec. III A, we must take care of this normalization when a change in parametrization is made: we must rescale $t(\lambda(\hat{\lambda}))$ as

$$\hat{t}(\hat{\lambda}) := t(\lambda(\hat{\lambda}))\left(\frac{d\lambda}{d\hat{\lambda}}\right)^{-3/2}$$

in order that

$$\tilde{e}(\hat{t},\dots,d^3\hat{t}/d\hat{\lambda}^3) = -1;$$

i.e., t has parameter weight $-\frac{3}{2}$. From this it follows that $t\wedge t'$ and $t\wedge t'\wedge t''$ have definite weights [namely, $-\frac{3}{2}+(-\frac{1}{2}) = -2$, $(-\frac{3}{2})+(-\frac{1}{2})+\frac{1}{2} = -\frac{3}{2}$], corresponding to the fact that they have geometrical meaning: $\mathbf{C}t\wedge t'$ describes the tangent, while $\mathbf{C}t\wedge t'\wedge t''$ describes the osculating plane at $\mathbf{C}t$. However, $t\wedge t'\wedge t''\wedge t'''$ will have weight 0 due to our normalization; its dual cannot be used to find an invariant parameter as in Sec. III A.

We now proceed as in Sec. III B: we expand t^{IV} in terms of the basis t, t', t', t''' as

$$t^{\mathrm{IV}} = \alpha t + \beta t' + \gamma t'' + \delta t'''.$$

Differentiating $\tilde{e}(t,\dots,t''') = -1$, it follows that $\tilde{e}(t,t',t'',t^{\mathrm{IV}}) = 0$ so that $\delta = 0$. Similarly, the reparametrized $\hat{t}(\hat{\lambda})$ will give an expansion of $d^4\hat{t}/d\hat{\lambda}^4$ with coefficients $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, whose relation to α, β, and γ can be calculated in a straightforward manner. After some computation one obtains

$$\hat{\gamma} = \gamma\left(\frac{d\lambda}{d\hat{\lambda}}\right)^2 - 5\mathcal{S}(\lambda|\hat{\lambda}),$$

$$\hat{\beta} = \beta\left(\frac{d\lambda}{d\hat{\lambda}}\right)^3 + 2\gamma\frac{d\lambda}{d\hat{\lambda}}\cdot\frac{d^2\lambda}{d\hat{\lambda}^2} - 5\frac{d}{d\hat{\lambda}}\mathcal{S},$$

$$\hat{\alpha} = \alpha\left(\frac{d\lambda}{d\hat{\lambda}}\right)^4 + \frac{3}{2}\beta\left(\frac{d\lambda}{d\hat{\lambda}}\right)^2\frac{d^2\lambda}{d\hat{\lambda}^2} + \frac{3}{4}\gamma\left(\frac{d\lambda}{d\hat{\lambda}}\frac{d^3\lambda}{d\hat{\lambda}^3} - 2\left(\frac{d^2\lambda}{d\hat{\lambda}^2}\right)^2\right) - \frac{3}{2}\frac{d^2}{d\hat{\lambda}^2}\mathcal{S} - \frac{9}{4}\mathcal{S}^2,$$

where

$$\mathcal{S}(\lambda|\hat{\lambda}) := \frac{d^3\lambda/d\hat{\lambda}^3}{d\lambda/d\hat{\lambda}} - \frac{3}{2}\left(\frac{d^2\lambda/d\hat{\lambda}^2}{d\lambda/d\hat{\lambda}}\right)^2$$

is the well-known Schwarzian derivative.

It is convenient to rewrite the results for $\hat{\beta}$ and $\hat{\alpha}$ in the following manner:

$$\hat{\beta} - \frac{d\hat{\gamma}}{d\hat{\lambda}} \equiv \left(\beta - \frac{d\gamma}{d\lambda}\right)\left(\frac{d\lambda}{d\hat{\lambda}}\right)^3,$$

$$\hat{\alpha} - \frac{1}{2}\frac{d}{d\hat{\lambda}}\left(\hat{\beta} - \frac{d\hat{\gamma}}{d\hat{\lambda}}\right) - \frac{3}{10}\frac{d^2\hat{\gamma}}{d\hat{\lambda}^2} + \frac{9}{100}\hat{\gamma}^2 \equiv \left(\alpha - \frac{1}{2}\frac{d}{d\lambda}\left(\beta - \frac{d\gamma}{d\lambda}\right) - \frac{3}{10}\frac{d^2\gamma}{d\lambda^2} + \frac{9}{100}\gamma^2\right)\left(\frac{d\lambda}{d\hat{\lambda}}\right)^4.$$

This gives us two quantities of definite parameter weights (3 and 4). More accurately, if we consider a curve in the sense we are using it (a "thread"), then it follows that $\mathrm{Im}\,\hat{\gamma}$

$= \text{Im}\,\gamma(d\lambda/d\hat{\lambda})^2$, so that Im γ is a further quantity of definite weight ($=2$). This further weighted quantity is absent if we consider real $P(\mathbf{T})$ or local holomorphic curves in complex $P(\mathbf{T})$, since then in the first case γ is real and in the second $\lambda(\hat{\lambda})$ is complex-valued. However, it will turn out that for curves with all tangents on $P\mathbf{N}$, Im $\gamma = 0$ again. Similarly, for general "threads" the two complex, weighted quantities above actually give four real, weighted quantities after taking real and imaginary parts; but again, for curves with all tangents on $P\mathbf{N}$ it will turn out that the weight 4 quantity is real and the weight 3 quantity is pure imaginary.

We will now show that the weight 3 quantity corresponds exactly to the one constructed in Sec. III A. One way of seeing this is to form $X = t\wedge t'$, $X' = t\wedge t''$, $X'' = t\wedge t''' + t'\wedge t''$, and, using the expansion of t^{IV} repeatedly,

$$X''' = 2t'\wedge t''' + \beta X + \gamma X',$$
$$X^{\text{IV}} = 2t''\wedge t''' + 2\gamma t'\wedge t'' + (\beta' - 2\alpha)X + (\beta + \gamma')X' + \gamma X'',$$
$$X^{\text{V}} = 2(\gamma' - \beta)t'\wedge t'' + (\beta'' - 2\alpha' - \beta\gamma)X + (2\beta' - 4\alpha + \gamma'' - \gamma^2)X' + (\beta + 2\gamma')X'' + \gamma X'''.$$

Note that all these bivectors have definite parameter weight and, in particular, X'' has weight 0. Insertion into $\tilde{E}(X,...,X^{\text{V}})$ gives

$$\tilde{E}(X,...,X^{\text{V}}) = 8(\gamma' - \beta) \times \tilde{E}(t\wedge t', t\wedge t'', t\wedge t''', t'\wedge t''', t''\wedge t''', t'\wedge t'').$$

This already shows that the vanishing of $\beta - \gamma'$ is equivalent to the vanishing of $\tilde{E}(X,...,X^{\text{V}})$, since the six products $t\wedge t',...$ are linearly independent in $\Lambda^2\mathbf{T}$. [In fact, from the Hilbert *Nullstellensatz* one can deduce that $\tilde{E}(t\wedge t',...)$ is a numerical multiple of $[\tilde{e}(t,..., t''')]^3$, and this is consistent with the weights of $\tilde{E}(X,...)$ and $\beta - \gamma'$.] Another way is to consider

$$H := i(t\wedge t''' - t'\wedge t'' - \gamma t\wedge t')$$

satisfying

$$H' = i(\beta - \gamma')t\wedge t',$$
$$H\wedge X = 0 \ \ (\text{i.e., } H\cdot X = 0),$$
$$H\wedge H = 2t\wedge t'\wedge t''\wedge t''' \ \ (\text{i.e., } H^2 = -2).$$

(The factor i is for later purposes and is irrelevant for the moment.) The differential equation for H implies that H has parameter weight 0. Now if $\beta - \gamma' = 0$, then H is constant; hence X remains in the fixed hyperplane of $P(\Lambda^2\mathbf{T})$ given by $H\cdot X = 0$. The $\mathbf{M}_4$ interpretation of this has already been given. The gain in using the twistor formalism lies in the fact that it has provided us with the second, weight 4, quantity (which seems to go back to Fubini and Čech). If it does not vanish while $\beta = \gamma'$, the differential

$$|\alpha - \tfrac{3}{10}\gamma'' + \tfrac{9}{100}\gamma^2|^{1/4}\,d\lambda = d\tau$$

will supply another invariant parameter for the curve. We shall refer to it as the Fubini–Čech parameter. If the curve is referred to it, $d\lambda = d\tau$, both α and β are expressible in terms of γ, which remains the only independent invariant of the curve in this case.

If the Fubini–Čech differential also vanishes, it is convenient to fix the parametrization partially by requiring $\gamma = 0$, which leaves open fractional linear transformations, for which $\mathscr{S}(\lambda\,|\hat{\lambda}) = 0$. Relative to this class of ("projective") parameters, $\alpha = \beta = \gamma = 0$, so that $t^{\text{IV}} = 0$ and

$$t(\lambda) = t_0 + t'_0\lambda + \tfrac{1}{2}t''_0\lambda^2 + \tfrac{1}{6}t'''_0\lambda^3$$

is what is known as a "twisted cubic." When $X(\lambda) = t(\lambda)\wedge t'(\lambda)$ is formed and translated back to the $\mathbf{M}_4$ language, one obtains (a complex version, in general, of) a conformally transformed cubic null helix. (We will see this in terms of the invariants of Sec. II A in Sec. V.) We shall come back to it after discussing the reality constraints in general.

If, however, $\beta - \gamma' \neq 0$, we can use $|\beta - \gamma'|^{1/3}\,d\lambda = d\sigma$ to define an invariant parameter: in the context of curves in projective space it is called the "projective arc," and it is obvious that for curves in $P\mathbf{N}$ it will correspond to the conformal arc introduced in Sec. III A. What is lacking again is a discussion of the reality constraints. If the curve is referred to the projective arc, $d\lambda = d\sigma$, β becomes expressible in terms of γ, which, therefore, together with α forms the system of fundamental invariants of the curve ("projective curvatures").

Let us end this subsection by remarking that a similar development is possible for planar curves (and, dually, for cones): one obtains, in general, a projective arc and one single projective curvature, an exception being formed by conics (no projective arc, no invariants).[4] As we have seen, however, consideration of these objects would be necessary only for signature $++--$ in $\mathbf{M}_4$ or for complexified $\mathbf{M}_4$. In this paper we generally skip this topic and now turn to the reality constraints.

C. Reality constraints

For a twisted curve $\{\mathbf{C}t(\lambda)\}$ in $P(\mathbf{T})$, which together with its tangents is contained in $P\mathbf{N}$, we have already deduced from the reality condition on its tangents $X = t\wedge t'$ that $(t|t) = (t|t') = (t'|t') = 0$. Since λ is real and X is real, X' is real, and from $X' = t\wedge t''$ it also follows that $(t|t'') = (t''|t'') = 0$. [Note that the Hermitian property of $(\cdot|\cdot)$ allows us to deduce only $(t'|t) + (t|t') \equiv 2\,\text{Re}(t|t') = 0$ from $(t|t) = 0$, etc., by differentiation!] We now complete a table $(t^{(q)}|t^{(p)})$ of $(\cdot|\cdot)$ scalar products for $0 \leqslant p \leqslant 4$, $0 \leqslant q \leqslant 3$, putting $(t''|t''') =: iR$, $(t'''|t''') =: S$, $(t'''|t^{\text{IV}}) =: iT$ (where R, S, and T are real), exploiting all known (vanishing) scalar products as well as the normalization $\tilde{e}(t,t',t'',t''') = -1$ which implies det (Table II without last column) $= 1$ (see Table II). [A sign convention on $(\cdot|\cdot)$ has also been made that is not yet fixed by the relative normalization between $\tilde{e}$ and $(\cdot|\cdot)$.] To establish the relation of R, S, and T to α, β, and γ introduced earlier, we compare the last column of Table II with the values of

$$(t^{(p)}|t^{\text{IV}}) = (t^{(p)}|\alpha t + \beta t' + \gamma t''), \text{ for } p = 1,2,3.$$

We obtain

$$\gamma = R = \text{real},$$

TABLE II. The scalar products $(t^{(q)}|t^{(p)})$.

q/p	0	1	2	3	4
0	0	0	0	i	0
1	0	0	$-i$	0	$-iR$
2	0	i	0	iR	$-S+iR$
3	$-i$	0	$-iR$	S	$\frac{1}{2}S'+iT$

$$\beta-\gamma'=iS=\text{pure imaginary},$$

$$\alpha-\tfrac{1}{2}(\beta-\gamma')'=-T-R^2=\text{real},$$

verifying our earlier statement that the weight 3 (resp. 4) quantity is pure imaginary (resp. real). At the same time we see that we have two real fundamental invariants in the general case $S\neq 0$ and one in the special case $S=0$, except for the conformal cubic null helix, which has no invariant parameter and no invariants.

The explanation for this latter fact lies in the large invariance group the twisted cubic possesses, and the fact that all twisted cubics are projectively equivalent. This is conveniently described in twistor terms as follows. In the parametric representation given earlier, where t_0, t'_0, t''_0, t'''_0 satisfy the values of the $(\cdot|\cdot)$ scalar products given in Table II with $R=S=0$ together with $\tilde{e}(t_0,...)=-1$, we go over to a pair of "homogeneous parameters" s^1, s^2 by $\lambda\to s^2/s^1$, $t(\lambda)\to(s^1)^3 t(\lambda)$. Now think of the s^A as components of a vector s in a real two-dimensional vector space $\mathbf{S}$, referred to a basis $\{b_1,b_2\}$: $s=s^A b_A$. Form the symmetrized tensorial power $V^3\mathbf{S}^c$ of its complexification $\mathbf{S}^c$ and choose an isomorphism j: $V^3\mathbf{S}^c\to\mathbf{T}$ (which are both complex four-dimensional) mapping the product basis vectors $b_1^{\vee 3}$, $b_1^{\vee 2}\vee b_2$, $b_1\vee b_2^{\vee 2}$, and $b_2^{\vee 3}$ ($\vee$ indicates symmetric multiplication) to t_0, $t'_0/3$, $t''_0/6$, $t'''_0/6$, respectively. Then our curve and its complexification arises from composing with j the "Veronese imbedding" $s\to s^{\vee 3}$ of $\mathbf{S}$ into $V^3\mathbf{S}$ and $\mathbf{S}^c$ into $V^3\mathbf{S}^c$, as the image of $P(\mathbf{S})$ and $P(\mathbf{S}^c)$, respectively. The scalar product restrictions on t_0,... guarantee that $P(\mathbf{S})$ is mapped into $P\mathbf{N}$. The pullback of the Hermitian form $(t|t)$ by ($j\circ$Veronese) is the cube of an indefinite Hermitian form on $\mathbf{S}^c$ whose zero set on $P(\mathbf{S}^c)$ is the "Staudt chain" $P(\mathbf{S})\subset P(\mathbf{S}^c)$,[10] or, in simpler terms, real λ lead to a curve on $P\mathbf{N}$. Consider now an arbitrary (real) unimodular linear transformation A of $\mathbf{S}$, then $j\circ A^{\vee 3}\circ j^{-1}$ will not lead out of the image of $\mathbf{S}$. Therefore the cubic null helix has a three-parameter conformal invariance group (locally) isomorphic to $SL(2,R)$, acting triply transitively; thus explaining the absence of any conformally invariant parameter. [We have already written down a Killing field of $\mathbf{M}_4$ admitted by the example of a cubic null helix given in Sec. II A; two other independent conformal Killing fields admitted by it are given by

$$x-x_0+\tfrac{1}{4}[(x'_0\cdot(x-x_0))x'''_0-(x'''_0\cdot(x-x_0))x'_0]$$

and

$$(x'_0\cdot(x-x_0))x''_0-(x''_0\cdot(x-x_0))x'_0$$
$$+\tfrac{1}{3}[(x-x_0)^2x'''_0-2(x'''_0\cdot(x-x_0))(x-x_0)].$$

One can check that the Lie brackets of these three conformal Killing fields, suitably scaled, correspond to the Lie algebra of $SL(2,R)$.] The conformal equivalence of all twisted cubics on $P\mathbf{N}$ follows from the fact that all quadruples t_0, t'_0, t''_0, t'''_0 satisfying $\tilde{e}(t_0,...)=-1$ and the "antidiagonal" form of the $(\cdot|\cdot)$ scalar product table with $R=S=0$ can be transformed into each other by the group of complex linear transformations that preserve $\tilde{e}$ and $(\cdot|\cdot)$, i.e., the twistor group [$\approx SU(2,2)$].

D. Relation to the invariants of the six-vector formalism

We still have to relate the invariants obtained in this section to the ones of the $\mathbf{V}_6$ formalism. First note that $X(\lambda)$ is real in the sense of reality that we have introduced in Λ^2T. [So far we have used only the consequences $(t|t)=(t|t')=(t'|t')=0$ of the reality condition $h^{\wedge 2}X={*}X$ on $X=t\wedge t'$; however, taking the only nontrivial contraction (viz. with $t''\wedge t'''$) of this condition, we verify that $t\wedge t'$ itself, and not just a suitable complex multiple of it, satisfies the condition, due to our normalization $\tilde{e}(t,...,t''')=-1$.] Therefore all derivatives $X^{(p)}$ are real bivectors as well. From the differential equation it satisfies, we now also conclude that the bivector H introduced above is real, $H\in\mathbf{V}_6$, but not simple. From the definition $F\cdot G:=\langle{*}F|G\rangle={*}(F\wedge G)$ of the scalar product in $\Lambda^2\mathbf{T}$, we can express the $X^{(p)}\cdot X^{(q)}$ in terms of α, β, and γ or R, S, and T, using the list of the $X^{(p)}$ in terms of the $t^{(p)}$ given earlier. In this way we indeed reproduce Table I with the following identifications:

$$\tfrac{1}{2}K=\gamma=R,$$

$$\tfrac{1}{4}J=\alpha-\tfrac{1}{2}\beta'-\tfrac{3}{4}\gamma^2=\alpha-\tfrac{1}{2}(\beta-\gamma')'-\tfrac{1}{2}\gamma''-\tfrac{3}{4}\gamma^2$$
$$=-T-\tfrac{7}{4}R^2-\tfrac{1}{2}R''.$$

Instead of identifying L in terms of α, β, and γ, we identify the numerical factor

$$-8\widetilde{E}(t\wedge t',t\wedge t'',t\wedge t''',t'\wedge t'',t'\wedge t''',t''\wedge t''')$$

between $\widetilde{E}(X,...,X^{\text{v}})$ and $(\beta-\gamma')$. If e_1, e_2, e_3, and e_4 are from $\mathbf{T}$ and $e_{ik}:=e_i\wedge e_k$, we already remarked that $\tilde{e}(e_1,e_2,e_3,e_4)$ and $\widetilde{E}(e_{12},e_{13},e_{14},e_{23},e_{24},e_{34})$ vanish simultaneously, so

$$\widetilde{E}(e_{12},...,e_{34})=\varepsilon(\tilde{e}(e_1,e_2,e_3,e_4))^3,$$

for some $0\neq\varepsilon\in\mathbf{C}$, and we can determine ε from the requirements that $\widetilde{E}$ give ± 1 upon evaluation on a $\cdot$-orthonormal basis in $\mathbf{V}_6$ and that the $\tilde{e}$ have absolute value 1 when evaluated on a $(\cdot|\cdot)$-orthonormal basis in $\mathbf{T}$. Let e_1 e_2, e_3, and e_4 now be the unimodular orthonormal basis in $\mathbf{T}$ used in Ref. 7 and $\eta_1/\sqrt{2},...,\eta_6/\sqrt{2}$ be the orthonormal basis of $\mathbf{V}_6$ constructed from it in that paper. Let the orientation of $\mathbf{V}_6$ be given by $\eta_1,...,\eta_6$ in that order. Then from the expressions of η_1,... in terms of e_1,... given in Ref. 7 we deduce

$$1=\widetilde{E}(\eta_1/\sqrt{2},...)=i\widetilde{E}(e_{12},...,e_{34})=i\varepsilon\cdot 1^3;$$

hence $\varepsilon=-i$ and therefore

$$\widetilde{E}(X,...,X^{\text{v}})=-8(\beta-\gamma')(-i)(-1)^3=8S.$$

This completes the relationships between the twistor and six-vector formalism.

At the same time we have verified all the results claimed in the Introduction.

V. CONCLUDING REMARKS

In the preceding sections we have found conformally invariant parameters and a fundamental system of differential invariants for null curves in conformally flat space-time. The technique was to use a formalism that makes conformal invariance manifest. Now conformal invariance *a fortiori* means Poincaré invariance, and in Sec. II A we have recapitulated a formalism that gives a fundamental system of invariants for null curves under the Poincaré group. Therefore, an alternative approach to the problem of conformal invariants would be to make an ansatz for the conformal invariants in terms of Poincaré invariants and to work out the consequences of additional invariance under infinitesimal scale transformations and "conformal boosts." We do not carry this out here but only indicate how the result is derived in our formalism. We take the invariants from the manifestly conformally covariant formalism and insert, instead of $X(\lambda)$, the expression $(x,(1-x^2)/2,(1+x^2)/2)$, the invariant parameter of Sec II A, so that $x''^2 = -2$; then automatically

$$X''^2 = x''^2 + ((1-x^2)/2)''^2 - ((1+x^2)/2)''^2 = -2$$

is normalized properly. The equations $x' = \xi \otimes \bar{\xi}$, $\varepsilon(\xi,\xi') = 1, \xi'' = I\xi$ of Sec. II A allow us to express all products $x^{(p)} \cdot x^{(q)}$ in terms of I and its derivatives and thus all conformal invariants in tems of I and is derivatives.

Let us illustrate this first for the weight 3 quantity $\widetilde{E}(X,...,X^{\mathrm{V}})$. Adjusting the orientation of $\mathbf{M}_4$, given by its ε-tensor, properly relative to $\widetilde{E}$, we obtain

$$\begin{aligned}\widetilde{E}(X,...,X^{\mathrm{V}}) &= -2\varepsilon(x',x'',x''',x^{\mathrm{V}}) \\ &= -2(\varepsilon(x',x'',x''',x^{\mathrm{IV}}))',\end{aligned}$$

and, observing that $(\xi\otimes\bar{\xi} \pm \xi'\otimes\bar{\xi}')/\sqrt{2}$, $(\xi\otimes\bar{\xi}' + \xi'\otimes\bar{\xi})/\sqrt{2}$, and $(\xi\otimes\bar{\xi}' - \xi'\otimes\bar{\xi})/i\sqrt{2}$ form an orthonormal tetrad for $\mathbf{M}_4$, the determinant finally becomes $8\,\mathrm{Im}\,I'$ Thus we see that while $\mathrm{Im}\,I = 0$ has the only Poincaré-invariant significance, mentioned in Sec. II A, that the curve stays in a fixed timelike hyperplane, $\mathrm{Im}\,I = \mathrm{const}$, has the conformally invariant meaning that the curve stays in a fixed timelike pseudosphere $(x-a)^2 = -r^2$ or a timelike hyperplane (the *value* of the constant $\mathrm{Im}\,I$, related to the radius r, is *not* conformally invariant). (It is, of course, possible to verify these statements directly in the formalism of Sec. II A.)

Similarly, we find

$$4\gamma = 4R = 2K = X'''^2 = x'''^2 = 8\,\mathrm{Re}\,I,$$

$$\begin{aligned}&8[\alpha - \tfrac{1}{2}(\beta-\gamma')' - \tfrac{1}{2}\gamma'' - \tfrac{3}{4}\gamma^2] \\ &\quad = 2J = X^{\mathrm{IV}2} = x^{\mathrm{IV}2} = -32(\mathrm{Re}\,I)^2 - 8(\mathrm{Im}\,I)^2,\end{aligned}$$

which show that the conditions $I = 0$ and $\alpha = \beta = \gamma = 0$ are equivalent. Here we see without carrying out the twistor transformation explicitly that the cubic null helix, referred to the Bonnor parameter, corresponds to a twisted cubic in $P\mathbf{N} \subset P(\mathbf{T})$, referred to a projective parameter. More generally, these formulas allow us to go over from Bonnor's parameter to the conformal arc or the Fubini-Čech parameter and, using the transformation formulas $\gamma \to \hat{\gamma}$, $\alpha \to \hat{\alpha}$, to find the conformal invariants in terms of I. We do not write down the expressions explicitly, however.

ACKNOWLEDGMENT

I am indebted to R. Beig for pointing out Synge's paper[1] to me.

[1]W. B. Bonnor, Tensor (N.S.) **20**, 229 (1969). See, also, J. L. Synge, Nature **317** (24 Oct.), 675 (1985) for an interpretation of the Bonnor–Study–Vessiot parameter.
[2]W. Blaschke and G. Thomsen, *Differentialgeometrie* (Springer, Berlin, 1929), Vol. 3.
[3]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, 1986), Vol. 2; R. Penrose, J. Math. Phys. **8**, 345 (1967).
[4]G. Fubini and E. Čech, *Geometria proiettiva differenziale* (Zanichelli, Bologna, 1926), Vol. 1; *Introduction à la géomètrie projective differentielle des surfaces* (Gauthier–Villars, Paris, 1931); G. Bol, *Projektive Differentialgeometrie* (Vandenhoeck & Ruprecht, Göttingen, 1950), Vol. 1. These authors deal preferably with real space.
[5]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, 1984), Vol. 1.
[6]M. Crampin and F. A. E. Pirani, *Applicable Differential Geometry* (Cambridge U.P., Cambridge, 1987).
[7]W. Kopczyński and L. S. Woronowicz, Rep. Math. Phys. **2**, 35 (1971).
[8]I. Porteous, *Topological Geometry* (Cambridge U.P., Cambridge, 1981).
[9]M. F. Atiyah, "Geometry of Yang–Mills fields," Lezioni Fermiane, Accademia Nazionale dei Lincei, Pisa 1979; the first to notice the quaternion form of the standard BPST instanton was A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).
[10]J. L. Coolidge, *The Geometry of the Complex Domain* (Clarendon, Oxford, 1924). This author dismisses the *differential* geometry of threads in CP_3 as "presenting very little interest."
[11]H. Urbantke, Rep. Math. Phys. **21**, 111 (1985); also reprinted in *Gravitation and Geometry* (A volume in honor of I. Robinson), edited by W. Rindler and A. Trautman (Bibliopolis, Napoli, 1987).
[12]W. T. Shaw, Class. Quant. Grav. **2**, L113 (1985).

Twistors in 2+1 dimensions

R. S. Ward
Department of Mathematical Sciences, University of Durham, Durham, DH1 3LE, United Kingdom

(Received 22 March 1989; accepted for publication 31 May 1989)

The geometry of twistors for $(2+1)$-dimensional flat space-time is described. Functions on twistor space generate solutions of various field equations in space-time. As an illustrative example, it is shown what a sine–Gordon soliton looks like in this twistor description.

I. INTRODUCTION

Twenty-two years ago, a paper entitled "Twistor algebra," by Penrose, appeared in this journal.[1] It introduced twistors in $(3+1)$-dimensional Minkowski space-time. The analogous structure for $(2+1)$-dimensional flat space-time is in some ways simpler, but has not received much attention. It does, however, have many potential applications, some of which are mentioned below. This paper will explore some of the features of $(2+1)$-dimensional twistor theory.

The starting point is a two-dimensional real twistor space $\mathbb{N}$ (the points of which correspond to null planes in $\mathbb{R}^{2+1}$); functions on $\mathbb{N}$ correspond to solutions of various massless field equations on $\mathbb{R}^{2+1}$. A more geometrical description is obtained by complexifying $\mathbb{N}$, to yield a two-complex-dimensional twistor space $\mathbb{T}$. Roughly speaking, points in $\mathbb{T}$ correspond to directed timelike lines in $\mathbb{R}^{2+1}$.

Holomorphic vector bundles over $\mathbb{T}$ correspond to Yang–Mills–Higgs fields in $\mathbb{R}^{2+1}$ satisfying a set of nonlinear first-order equations (the hyperbolic analog of the Bogomolny equations for monopoles in $\mathbb{R}^3$). There are many reductions (i.e., special cases) of these equations which are of interest; examples include the Einstein vacuum equations with cylindrical symmetry, and $(1+1)$-dimensional soliton equations such as Korteweg–de Vries, nonlinear Schrödinger, and sine–Gordon. By way of example, the one-soliton solution of the sine–Gordon equation is described from this point of view. Of course, the equations just mentioned are already well understood. But the twistor picture may be useful in providing a unified geometrical description of all of them.

II. TWISTORS IN 2+1 DIMENSIONS

Let us begin by recalling one of the approaches to twistors in $(3+1)$-dimensional flat space-time $\mathbb{R}^{3+1}$. More details may be found in Refs. 2–4. One starts with the space N_5 of null geodesics in $\mathbb{R}^{3+1}$, which is five-dimensional; in fact, N_5 is $S^2\times\mathbb{R}^3$ as a real manifold. But it has some additional structure (arising from the conformal structure of space-time), namely a CR structure.[5] This is the structure inherited by a *real* hypersurface in *complex* manifold. So N_5 sits naturally inside a three-dimensional complex manifold T_3. Now T_3 is not uniquely determined by N_5; roughly speaking, that part of T_3 which lies close to N_5 is determined, but one can analytically continue away from this in many different ways. The simplest choice is to take T_3 to be the complex projective space $\mathbb{P}_3$, and this is the standard flat (projective) twistor space. This choice also effectively compactifies N_5 to $\overline{N}_5$, the space of null geodesics in compactified Minkowski space-time. The spaces $\overline{N}_5$ and $T_3=\mathbb{P}_3$ are homogeneous, being acted on transitively by the conformal group in $3+1$ dimensions, and its complexification, respectively.

A point of N_5 corresponds, of course, to a null geodesic in space-time. Points of $\mathbb{P}-\overline{N}_5$ can also be pictured in space-time: they correspond to twisting congruences of null lines (Robinson congruences); hence the name "twistor."

Let us turn now to $(2+1)$-dimensional flat space-time $\mathbb{R}^{2+1}$, and see how the situation differs. In this case, the space N_3 of null lines is three dimensional, but it does not have a natural CR structure. So one's first guess, that there should be a two-complex-dimensional twistor space in which N_3 sits as a real hypersurface, is wrong. Instead of N_3, the correct space to use is the space $\mathbb{N}$ of null *planes* in $\mathbb{R}^{2+1}$, which is two dimensional. In fact, $\mathbb{N}$ is $S^1\times\mathbb{R}$ as a manifold: if (t,x,y) are the usual space-time coordinates, then a null plane is given by an equation of the form

$$t+y\cos\theta+x\sin\theta=\omega, \tag{1}$$

where $(\theta,\omega)\in S^1\times\mathbb{R}$ are constant real numbers. This gives a one-to-one correspondence between null planes in $\mathbb{R}^{2+1}$ and points of $S^1\times\mathbb{R}$.

A more invariant description can be arrived at by making use of the fact that the identity-connected component of the $(2+1)$-Lorentz group O(2,1) is double-covered by SL(2,$\mathbb{R}$); so spinors are two-component real objects π_A, $A=0,1$. The space-time coordinates may be rearranged as a symmetric two-spinor,

$$x^{AB}=\begin{bmatrix} t+y & x \\ x & t-y \end{bmatrix}. \tag{2}$$

The space-time metric is

$$\begin{aligned} ds^2 &= \eta_{\mu\nu}\,dx^\mu\,dx^\nu \\ &= -dt^2+dx^2+dy^2 \\ &= -\tfrac{1}{2}dx^{AB}\,dx_{AB}. \end{aligned} \tag{3}$$

(Spinor indices are lowered with the alternating spinor ε_{AB}, as usual.)

An SL(2,$\mathbb{R}$)-invariant description of null planes is as follows. Let π_A be a nonzero real spinor (i.e., π_0 and π_1 not both zero) and ω a real number. These parameters determine a null plane by

$$\omega=x^{AB}\pi_A\pi_B. \tag{4}$$

Clearly (ω,π_A) are homogeneous coordinates for $\mathbb{N}$, in the sense that $(\lambda^2\omega,\lambda\pi_A)$, where λ is any nonzero real number,

 0022-2488/89/102246-06$02.50

determines the same null plane as (ω,π_A). If we put $\pi_0 = \cos\frac{1}{2}\theta$ and $\pi_1 = \sin\frac{1}{2}\theta$, then (4) reduces to (1).

Solutions of the wave equation in $\mathbb{R}^{2+1}$ can be generated from functions on $\mathbb{N}$: if $f = f(\omega,\theta)$ is a smooth function on $\mathbb{N}$, then

$$\varphi(t,x,y) = \int_0^{2\pi} f(t + y\cos\theta + x\sin\theta,\theta)d\theta \qquad (5)$$

is a solution of the wave equation $\Box\varphi = 0$. This is the Lorentzian version of Whittaker's[6] famous formula for solutions of the Laplace equation in $\mathbb{R}^3$. The discussion in Ref. 6, which uses a power-series argument, demonstrates that all *real-analytic* solutions of $\Box\varphi = 0$ can be obtained (locally) as in (5). To deal with nonanalytic solutions, one could adopt a purely "real" approach, and study the integral transform (5) from the point of view of real analysis. For example, such an approach is employed in Ref. 7, which deals with the closely related problem of the self-dual Maxwell equations in $\mathbb{R}^{2+2}$. Also, (5) is related to the Radon transform[8] between functions on $\mathbb{N}$ and functions on $\mathbb{R}^2$ (this $\mathbb{R}^2$ being thought of as an initial-data surface, such as $t = 0$, in $\mathbb{R}^{2+1}$). An alternative way involves working with the complex twistor space $\mathbb{T}$ (introduced below), and using cohomology and hyperfunctions, along the lines of Ref. 9. This subject will not be pursued further here; some remarks on cohomology are made in the next section.

Whether one stays with the real space $\mathbb{N}$, or works with its complexification $\mathbb{T}$, is partly a matter of taste. As far as nonlinear problems are concerned, the complex approach is more geometrical, and in some cases reduces to algebraic geometry; from now on, we shall use the complex framework.

The complex twistor space $\mathbb{T}$ is a two-dimensional complex manifold, a complexification of the two-dimensional real manifold $\mathbb{N}$. The most natural complexification is obtained by simply allowing the homogeneous coordinates (ω,π_A) to become complex. So $\mathbb{T}$ is $\mathbb{C}\times(\mathbb{C}^2 - \{0\})$, factored out by the equivalence relation

$$(\lambda^2\omega,\lambda\pi_A) \sim (\omega,\pi_A), \qquad (6)$$

where λ runs over the nonzero complex numbers. Now $\mathbb{C}^2 - \{0\}$ factored in this way is just the complex projective line $\mathbb{P}_1$. So $\mathbb{T}$ is a holomorphic line bundle over $\mathbb{P}_1$; from (6) it follows that $\mathbb{T}$ is in fact the holomorphic tangent bundle of $\mathbb{P}_1$.

This space $\mathbb{T}$ (sometimes known as "minitwistor" space) was used[10,11] in discussing monopole solutions on Euclidean three-space $\mathbb{R}^3$. In fact, one could begin with a purely complex description, defining $\mathbb{T}$ to be the space of all complex null planes in $\mathbb{C}^3$ (with respect to the complexified Euclidean metric); then different "reality structures" on $\mathbb{T}$ correspond to different signatures for the underlying real metric. The one used in Ref. 10 is relevant to the signature $+ + +$, whereas the one used here is relevant to $- + +$, and consists of the complex conjugation operation

$$(\omega,\pi_0,\pi_1) \mapsto (\bar{\omega},\bar{\pi}_0,\bar{\pi}_1). \qquad (7)$$

Since $\mathbb{T}$ is a bundle over $\mathbb{P}_1$, there are a natural family of curves in $\mathbb{T}$, namely the holomorphic sections of the bundle. These form a three-complex-parameter family (in fact, they

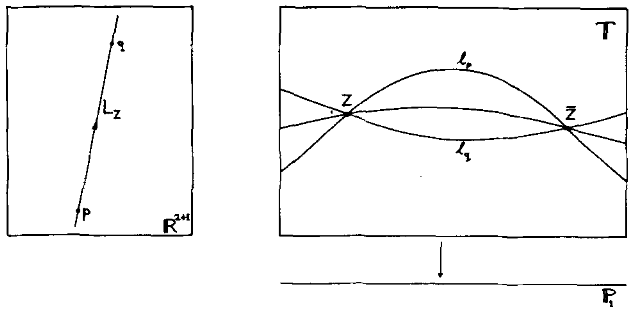


FIG. 1. The correspondence between $\mathbb{R}^{2+1}$ and T.

are parametrized by the space $\mathbb{C}^3$ referred to in the previous paragraph). The real sections, i.e., those preserved under the conjugation (7), are parametrized by $\mathbb{R}^{2+1}$. Let us denote by l_p, the real section corresponding to a point p in $\mathbb{R}^{2+1}$. See Fig. 1.

If p and q are two points in $\mathbb{R}^{2+1}$, then the two curves l_p and l_q in $\mathbb{T}$ will intersect either at one point (in which case p and q are null-separated) or at two points. In the latter case, the two points of intersection could either be real, i.e., each preserved by (7) (in which case p and q are spacelike separated), or complex, and conjugates of each other (in which case p and q are timelike separated). All this is easily deduced from Eq. (4). So we see that the causal structure of the space-time $\mathbb{R}^{2+1}$ is encoded into the geometry of $\mathbb{T}$.

What do the *points* of $\mathbb{T}$ correspond to in space-time? Again, Eq. (4) provides the answer: if ω and π_A are fixed, then (4) defines an affine subspace of $\mathbb{R}^{2+1}$. Three different cases must be distinguished. Let us assume that π_1 is nonzero (if not, interchange π_0 and π_1 in what follows), and write $\zeta = \pi_0/\pi_1$, $\nu = \omega/\pi_1{}^2$. So (4) becomes

$$\nu = (t + y)\zeta^2 + 2x\zeta + (t - y). \qquad (8)$$

If ν and ζ are both real, then (8) defines a null plane (as we already knew). If ζ is real but ν is complex, then (8) has no solution at all. Finally, if both ν and ζ are complex, then the solution of (8) is a timelike line in $\mathbb{R}^{2+1}$. The direction vector of this line is $v^{AB} = \pi^{(A}\bar{\pi}^{B)}$, where parentheses denote symmetrization; in (t,x,y) coordinates, this direction vector is

$$(1 + |\zeta|^2, \quad -\zeta - \bar{\zeta}, \quad 1 - |\zeta|^2). \qquad (9)$$

Clearly (ν,ζ) and $(\bar{\nu},\bar{\zeta})$ determine the *same* line. To elimi-

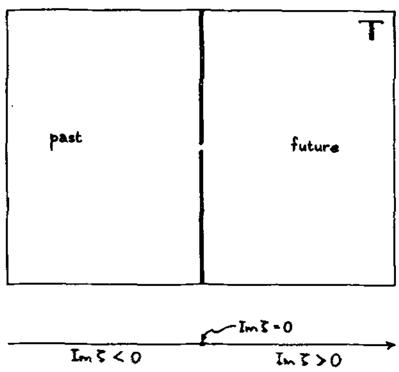


FIG. 2. The "causal structure" of T.

nate this redundancy, we may regard the line as being oriented: to the future if $\mathrm{Im}\,\zeta > 0$, and to the past if $\mathrm{Im}\,\zeta < 0$.

To sum up: the minitwistor space $\mathbb{T}$ is divided into two hemispheres $\mathrm{Im}\,\zeta > 0$ and $\mathrm{Im}\,\zeta < 0$, separated by an equator $\mathrm{Im}\,\zeta = 0$ (see Fig. 2). A point Z in the region $\mathrm{Im}\,\zeta > 0$ corresponds to a future-directed timelike line L_Z in $\mathbb{R}^{2+1}$ (each point on L_Z corresponds to a real section of $\mathbb{T}$ passing through both Z and $\bar{Z}$: see Fig. 1). Similarly, points in $\mathrm{Im}\,\zeta < 0$ correspond to past-directed timelike lines in $\mathbb{R}^{2+1}$. Most points on $\mathrm{Im}\,\zeta = 0$ do not correspond to anything in space-time, but some (namely those with ν real) correspond to real null planes.

III. MASSLESS FIELDS, COHOMOLOGY, AND COMPACTIFICATION

The integral formula (5) for solutions of $\Box\varphi = 0$ can be recast in terms of sheaf cohomology and complex contour integrals on $\mathbb{T}$. The ideas are the same as for four-dimensional space-time[2–4,12]; the following is a brief description.

Let M be a neighborhood of some section l_p in $\mathbb{T}$. Cover M with the two sets

$$U = \{\mathrm{Im}\,\zeta \geqslant 0\} \cap M, \quad \hat{U} = \{\mathrm{Im}\,\zeta \leqslant 0\} \cap M. \tag{10}$$

(Strictly speaking, we should take $U = \{\mathrm{Im}\,\zeta > -\varepsilon\} \cap M$, where ε is some positive number; and $\mathrm{Im}\,\zeta < \varepsilon$ in the definition of $\hat{U}$. In other words, U and $\hat{U}$ are "hemispheric" open sets whose intersection contains the equator $\mathrm{Im}\,\zeta = 0$.) Let $f = f(\omega, \pi_A)$ be a holomorphic function on $U \cap \hat{U}$, homogeneous of degree -2. [We say that f is homogeneous of degree n, and is a section of the sheaf $\mathcal{O}_{\mathbb{T}}(n)$, if

$$f(\lambda^2\omega, \lambda\pi_A) = \lambda^n f(\omega, \pi_A) \tag{11}$$

for all nonzero complex numbers λ.] This f, defined on the overlap region $U \cap \hat{U}$, represents an element of the sheaf cohomology group $H^1(M, \mathcal{O}_{\mathbb{T}}(-2))$. And it can be evaluated by a contour integral

$$\varphi(x^{AB}) = 2\oint f(x^{AB}\pi_A\pi_B, \pi_C)\pi_D\, d\pi^D \tag{12}$$

to give a solution of $\Box\varphi = 0$. The contour in (12) is the equator $\mathrm{Im}\,\zeta = 0$. If we parametrize this curve by $\zeta = \cot\frac{1}{2}\theta$, then (12) is identical to (5).

All this is easiest to handle rigorously if one deals with *holomorphic* fields φ on the complexified space-time $\mathbb{C}^3$: see, for example, Ref. 13. On real space-time, as mentioned previously, the situation is more complicated, since solutions may not be real-analytic.

If f is homogeneous of degree $-n-2$, where n is a positive integer, then (12) should be replaced by

$$\varphi_{AB\cdots D}(x) = 2\oint \pi_A\pi_B\cdots\pi_D f(x^{PQ}\pi_P\pi_Q, \pi_R)\pi_S\, d\pi^S \tag{13}$$

with n factors of π in the integrand. So $\varphi_{AB\cdots D}$ (which is totally symmetric) is a field of spin $\frac{1}{2}n$, and (13) is a solution formula for the massless free-field equation

$$\partial^{AE}\varphi_{AB\cdots D} = 0. \tag{14}$$

Here ∂_{AB} is the spinor version of the partial derivative $\partial_\mu = \partial/\partial x^\mu$ ($\mu = 0,1,2$). For functions homogeneous of degree greater than -2, one differentiates with respect to ω in order to reduce the degree. For example, if f has degree 0, then $\partial f/\partial\omega$ has degree -2, and so yields a scalar field. In fact, this case may be viewed differently: $\exp(f)$ is the patching function for a line bundle over M, and this bundle corresponds to a U(1) gauge field, plus a scalar field, satisfying a set of coupled equations (see the next section). This scalar field is (up to a multiplicative constant) the one generated by $\partial f/\partial\omega$.

As was remarked at the beginning of Sec. II, the standard twistor space compactifies naturally to a homogeneous space $\mathbb{P}_3$, and cohomology groups on $\mathbb{P}_3$ provide a description of massless fields on compactified Minkowski space-time; the whole setup is conformally invariant.[9,12] This is not the case for 2 + 1 dimensions and minitwistor space. Nevertheless, one can compactify $\mathbb{T}$, and try to extend the solution formula for $\Box\varphi = 0$ to this compactification. What happens is as follows.

Recall that $\mathbb{T}$ is fibered over $\mathbb{P}_1$, with each fiber being a copy of $\mathbb{C}$. To compactify, we add an extra section l_∞: the resulting compact space $\bar{\mathbb{T}}$ is still fibered over $\mathbb{P}_1$, but each fiber is now itself a copy of $\mathbb{P}_1$ (see Fig. 3). The space $\bar{\mathbb{T}}$ is a rational ruled surface that in algebraic geometry is denoted S_2.

We want to consider a region M in $\bar{\mathbb{T}}$ which is a neighborhood of a section l_p and which also contains l_∞. Let us take M to be $\bar{\mathbb{T}} - V$, where V is some small closed neighborhood of the set $\{\zeta \text{ real}, \omega = 0\}$ (see Fig. 3). The real sections in M (i.e., those which avoid V) are partitioned into two sets, corresponding to points inside the future and the past null cones of the origin O in $\mathbb{R}^{2+1}$. See Fig. 3, where p is inside the future null cone.

Now if $f(\omega, \pi_A)$ is holomorphic on $U \cap \hat{U}$ and homogeneous of degree -2 as before, then it yields a field φ which is real analytic inside the null cones, and which furthermore extends across $t = \infty$. By way of example, take $f(\omega, \pi_A) = \omega^{-1}$. Then doing the integral (5) or (12) gives

$$\varphi = (2\pi/t)(1 - r^2/t^2)^{-1/2} \tag{15}$$

(where $r^2 = x^2 + y^2$), the fundamental solution of $\Box\varphi = 0$. Notice that (5) is indeed real-analytic inside the future and past null cones, and across $t = \infty$ (where $t = +\infty$ and $t = -\infty$ are identified).

The solution (15) is singular on the null cone of the origin. But it is easy to modify it slightly, so as to obtain a

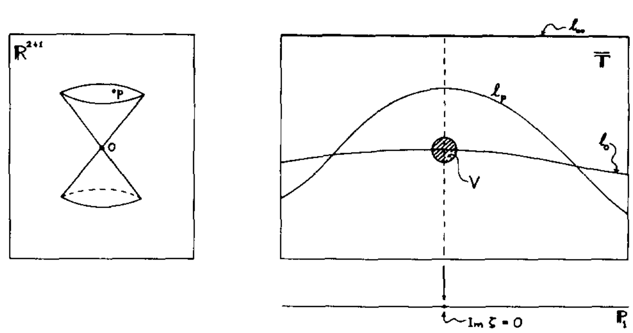


FIG. 3. Compactified minitwistor space $\bar{\mathbb{T}}$.

solution which is smooth on all of $\mathbb{R}^{2+1}$ and across $t=\infty$. For example,

$$f(\omega,\pi_A)=(\omega+i\pi_0^2+i\pi_1^2)^{-1}$$

leads to

$$\varphi=2\pi(t^2-r^2-1+2it)^{-1/2},$$

which is smooth and single valued on $\mathbb{R}^{2+1}$. Note that f is smooth on $\mathbb{N}$ (where ω and π_A are real); this is why the resulting field φ is smooth [cf. (5)].

So it appears that extending to the compactified minitwistor space $\overline{\mathbb{T}}$ corresponds to extending across infinity in space-time $\mathbb{R}^{2+1}$. However, unlike the standard twistor theory, the points at infinity ($t=\pm\infty$) do not correspond to holomorphic curves in the twistor space, at least not in any obvious geometrical way. This aspect should be investigated further.

IV. VECTOR BUNDLES AND YANG–MILLS–HIGGS FIELDS

Holomorphic vector bundles over the standard three-dimensional twistor space correspond to self-dual gauge fields in four-dimensional space-time (real or complexified).[3,4,14] The "mini" or reduced version of this is that bundles over $\mathbb{T}$ correspond to gauge fields in three-space, satisfying the Bogomolny equations. The form in which this correspondence is best known is the positive-definite one, relating to monopoles on Euclidean three-space.[10,11] In this section, we study the correspondence for gauge fields on $\mathbb{R}^{2+1}$.

The construction works for any gauge group, but for simplicity let us take the group to be SU(2). So a gauge potential A_μ is a one-form on $\mathbb{R}^{2+1}$ taking values in the Lie algebra su(2), i.e., each of A_0, A_1, and A_2 is an anti-Hermitian trace-free 2×2 matrix. The gauge field is

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu+[A_\mu,A_\nu].$$

In addition to the gauge field, one has an su(2)-valued scalar field Φ (the Higgs field). The Bogomolny equations are

$$D_\mu\Phi=\tfrac{1}{2}\epsilon_{\mu\alpha\beta}F^{\alpha\beta}, \tag{16}$$

where $\epsilon_{\mu\alpha\beta}$ is the alternating tensor. [We adopt the sign convention $\epsilon_{012}=-1$, with metric having signature $-++$, cf. (3).]

Equations (16) are hyperbolic, and so describe the time-evolution of the Yang–Mills–Higgs system in $\mathbb{R}^{2+1}$. Note that (16) implies the covariant wave equation

$$D_\mu D^\mu\Phi=0, \tag{17}$$

by virtue of the Bianchi identity on $F_{\mu\nu}$. In addition, (16) implies that

$$D^\mu F_{\mu\nu}=\tfrac{1}{2}\epsilon_{\alpha\beta\nu}[F^{\alpha\beta},\Phi]. \tag{18}$$

Equations (17) and (18) are the Euler–Lagrange equations obtained from the Lagrangian

$$L=\tfrac{1}{2}\,\mathrm{tr}\,F_{\mu\nu}F^{\mu\nu}-\mathrm{tr}(D_\mu\Phi)(D^\mu\Phi), \tag{19}$$

where tr denotes trace. In other words, the solutions of (16) are also solutions of the equations of motion obtained from (19). However, (19) is not exactly the usual Yang–Mills–Higgs Lagrangian, because in the latter the relative sign between the two terms is $+$ rather than $-$. One consequence of this change is that the conserved energy functional obtained from (19) is not positive-definite. Indeed, this energy density vanishes identically for solutions of (16). This matter will be returned to later, after the following description of how to construct solutions of (16).

Solutions of (16) on $\mathbb{R}^{2+1}$ may be generated from holomorphic rank-two vector bundles E over $\mathbb{T}$, satisfying

(A) for every real section of σ of $\mathbb{T}$, $E|_\sigma$ is trivial;

(B) det $E=\mathbb{1}$ and E has a "real structure." (20)

Condition (A) guarantees that the gauge field that will be constructed, is smooth (in fact real-analytic) on $\mathbb{R}^{2+1}$. Condition (B) guarantees that the gauge field will be su(2) valued, as opposed to merely gl(2,$\mathbb{C}$) valued; the explanation of exactly what (B) means will be given below.

Let us use an explicit construction, covering $\mathbb{T}$ with two patches U and $\hat{U}$, assuming E is trivial on each of these, and patching E together by

$$\hat{\psi}=F\psi, \tag{21}$$

where F is a 2×2 matrix of holomorphic functions on the overlap region $U\cap\hat{U}$. It is convenient to transform the coordinate ζ by a fractional linear transformation

$$\lambda=(\zeta-i)/(\zeta+i), \tag{22}$$

and to use λ instead of ζ as the coordinate on $\mathbb{P}_1$. The patches U and $\hat{U}$ used before [cf. (10)] become

$$U=\{|\lambda|\leqslant 1\},\quad \hat{U}=\{|\lambda|\geqslant 1,\ \text{including}\ \lambda=\infty\}. \tag{23}$$

The conjugation operation $\zeta\mapsto\bar{\zeta}$ is replaced by $\lambda\mapsto\bar{\lambda}^{-1}$. The fiber-coordinate ν [cf. (8)] is replaced by

$$\gamma=i\bar{z}\lambda+2t-iz\lambda^{-1}, \tag{24}$$

where $z=x+iy=x^1+ix^2$, and $t=x^0$. So $\gamma\lambda$ and $\gamma\lambda^{-1}$ serve as fiber-coordinates on $\mathbb{T}$, for $|\lambda|\leqslant 1$ and $|\lambda|\geqslant 1$, respectively. Note that (24) is preserved by complex conjugation, in the sense that

$$\overline{\gamma(x^\mu,\lambda)}=\gamma(x^\mu,\bar{\lambda}^{-1}).$$

We can now define what the conditions (B) in (20) mean: as conditions on the patching matrix F, they are

$$\det F=1,\quad F^\dagger=F, \tag{25}$$

where $F^\dagger$ is defined by

$$F^\dagger(\gamma,\lambda)=F(\bar{\gamma},\bar{\lambda}^{-1})^* \tag{26}$$

(here * denotes complex conjugate transpose).

The procedure for constructing A_μ and Φ is standard.[14] If one restricts $F(\gamma,\lambda)$ to a real section by imposing (24), then it is the patching matrix for a trivial bundle (by condition A), which means that it can be split:

$$F(i\bar{z}\lambda+2t-iz\lambda^{-1},\lambda)=\hat{H}(x^\mu,\lambda)H(x^\mu,\lambda)^{-1}, \tag{27}$$

where the matrix H is holomorphic for $|\lambda|\leqslant 1$, and $\hat{H}$ for $|\lambda|\geqslant 1$. Further, condition (B) enables us to impose $H^\dagger=\hat{H}^{-1}$ and det $H=1$. [Here $H^\dagger(x^\mu,\lambda):=H(x^\mu,\bar{\lambda}^{-1})^*$, cf. (26).] The two operators

$$\eth=\partial_z+\tfrac{1}{2}i\lambda^{-1}\partial_t,\quad \eth^\dagger=\partial_{\bar{z}}-\tfrac{1}{2}i\lambda\partial_t \tag{28}$$

each annihilate the expression (24), and hence annihilate the left-hand side of (27). Acting on (27) with $\eth$ therefore gives

$$\hat{H}^{-1}\eth\hat{H} = H^{-1}\eth H, \tag{29}$$

and each side of (29) must be linear in λ^{-1}. So $H^{-1}\eth H$ has the form

$$H^{-1}\eth H = \tfrac{1}{2}(A_x - iA_y) + \tfrac{1}{2}\lambda^{-1}(\Phi + iA_t), \tag{30}$$

for some anti-Hermitian trace-free matrices A_μ, Φ which depend only on x^μ (and not on λ). From (29), (30), and the unitarity condition $H^\dagger = \hat{H}^{-1}$ we get

$$H^{-1}\eth^\dagger H = \tfrac{1}{2}(A_x + iA_y) + \tfrac{1}{2}\lambda(\Phi - iA_t), \tag{31}$$

and the consistency condition for (30) and (31) is exactly the Bogomolny equation (16).

So holomorphic vector bundles over $\mathbb{T}$, satisfying (20), generate solutions of (16). Not all solutions arise in this way, however: in particular, holomorphic bundles lead to solutions that are real-analytic, whereas Eqs. (16), being hyperbolic, admit nonanalytic solutions. But it is certainly the case that one may obtain all real-analytic solutions in this way, at least locally (by considering holomorphic bundles over the neighborhood of a real line in $\mathbb{T}$). And the splitting procedure above works even if F is not analytic, so one can construct some nonanalytic solutions as well (although in this case they are not related to holomorphic bundles). This splitting, also known as the Birkhoff decomposition[15] or Riemann–Hilbert factorization, has wide applicability.

As was remarked previously, the Bogomolny equations (16), or the Yang–Mills–Higgs equations (17), (18), appear not to admit a local, positive definite, conserved energy density. One can rewrite (16) in a form which does admit an energy functional, but the price one pays for this is the loss of Lorentz invariance. Roughly speaking, one expresses A_μ and Φ in terms of first derivatives of an SU(2)-valued field J, and the first-order equations (16) then become second-order equations for J. The details are as follows.

Choose a gauge such that $A_t = A_y$ and $A_x = -\Phi$. Such a gauge exists for solutions of (16); in terms of the construction described above, it corresponds to choosing H such that $H|_{\lambda=1} = \mathbb{1}$ [cf. (30) and (31)]. Then J: $\mathbb{R}^{2+1} \to \mathrm{SU}(2)$ is taken to be a solution of

$$\begin{aligned} A_t &= A_y = \tfrac{1}{2}J^{-1}(J_t + J_y),\\ A_x &= -\Phi = \tfrac{1}{2}J^{-1}J_x. \end{aligned} \tag{32}$$

The integrability condition for (32) follows from (16), and in addition (16) implies an equation on J, namely

$$\eta^{\mu\nu}\partial_\mu(J^{-1}\partial_\nu J) + V_\alpha\epsilon^{\alpha\mu\nu}\partial_\mu(J^{-1}\partial_\nu J) = 0, \tag{33}$$

where V^α is the unit vector in the x-direction, i.e., $V_\alpha = (0,1,0)$. Conversely, given a solution J of (33), we can use (32) to get a solution of (16); so(16) and (33) are, in this sense, equivalent forms of the same thing. Equation (33) is a "chiral equation with torsion term" which admits a positive-definite conserved energy functional[16,17] (rewritten in terms of A_μ and Φ, this energy density would be nonlocal). The equation has soliton solutions, both lumplike and wavelike.[16–19]

V. REDUCTIONS AND THE SINE-GORDON EQUATION

One may reduce the Bogomolny equation (16) to 1 + 1 dimensions, by (roughly speaking) requiring the fields to be independent of one of the space-time coordinates. In fact, this can be done in several different ways. In this section, we shall concentrate on one example which serves to illustrate the general situation: the soliton solution of the sine–Gordon equation. But first, it is worthwhile to mention a few other examples.

In order to reduce from 2 + 1 to 1 + 1 dimensions, one assumes that the fields are constant along some Killing vector field in $\mathbb{R}^{2+1}$. There are several possibilities, depending on which Killing vector we choose. One choice is a null vector, say $V = \partial_t - \partial_y$. So in this case, the fields A_μ and Φ are assumed to be annihilated by V, and depend only on x and $t + y$. The reduced equations are then parabolic (with $t + y$ interpreted as "time"), and essentially reduce to either the nonlinear Schrödinger or the Korteweg–de Vries equation.[20] In other words, these well-known soliton equations are in effect special cases of (16).

Another possibility is obtained by using the Killing vector field $V = y\partial_t + t\partial_y$. Adopting the J description (33), we require that J be annihilated by V, i.e., that J be a function only of x and $\rho = (t^2 - y^2)^{1/2}$ (restricted to the region $t^2 - y^2 > 0$). Then (33) reduces to

$$\partial_x(J^{-1}\partial_x J) - \rho^{-1}\partial_\rho(\rho J^{-1}\partial_\rho J) = 0. \tag{34}$$

Also, let us take the gauge group to be SL(2,$\mathbb{R}$) rather than SU(2); this is achieved by an alteration of the reality condition (25). So J becomes an SL(2,$\mathbb{R}$)-valued matrix. Finally, impose the condition that J be symmetric [which is consistent with Eq. (34)]. Then (34) is effectively Einstein's vacuum equation for cylindrically-symmetric space-times (with x playing the role of time and ρ as the radial coordinate).[21] Numerous ways have been developed for generating solutions of these reduced Einstein equations, and the solutions can be interpreted in many different ways (for example, as cylindrical gravitational waves, as cosmological models, and as "gravitational solitons"). The twistor description provides a more geometric way of constructing and interpreting such solutions, and may prove to be useful.[21] (The corresponding structure for stationary axisymmetric space-times has already been studied in some detail.[22])

Let us return now to the sine–Gordon (sG) equation. To obtain this, we must reduce via a constant spacelike Killing vector, say $V = \partial_x$. In addition, the number of dependent variables is reduced by imposing algebraic constraints on the fields A_μ and Φ, consistent with Eq. (16). The situation may be summarized as follows.

Let g and φ be functions of t,y. Take A_μ and Φ to have the form

$$\begin{aligned} A_t &= -A_y = \tfrac{1}{2}ig\sigma_2,\\ A_x &= -\tfrac{1}{4}i(1 + \cos\varphi)\sigma_3 - \tfrac{1}{4}i(\sin\varphi)\sigma_1,\\ \Phi &= \tfrac{1}{4}i(1 - \cos\varphi)\sigma_3 - \tfrac{1}{4}i(\sin\varphi)\sigma_1, \end{aligned} \tag{35}$$

where σ_j are the Pauli matrices. Substitute (35) into the Bogomolny equations (16); the function g is easily eliminated, and the only equation that remains is the sG equation

$$\varphi_{yy} - \varphi_{tt} = \sin \varphi. \tag{36}$$

Note from (35) that φ and Φ are related by

$$-\operatorname{tr} \Phi^2 = \tfrac{1}{4}(1 - \cos \varphi). \tag{37}$$

Of course, one already knows how to construct solutions of the sG equation. The twistor picture merely gives an alternative (and more geometrical) view of the solutions. Let us see how to construct the one-soliton solution, for example.

The splitting procedure of Sec. IV can be carried out explicitly if the matrix F is upper triangular.[14] In this particular case, take

$$F(\gamma,\lambda) = \begin{bmatrix} \lambda & \Gamma(\gamma,\lambda) \\ 0 & \lambda^{-1} \end{bmatrix}, \tag{38}$$

where the function Γ is given by

$$\begin{aligned} \Gamma(\gamma,\lambda) &= Q^{-1}(h^{-1} + h^{\dagger}), \\ Q &= (\lambda - \alpha)(\lambda^{-1} - \alpha), \\ h &= \exp[\lambda^{-1}\gamma/(1-\alpha^2)] \end{aligned} \tag{39}$$

(α being a real number with $|\alpha| < 1$).

This patching matrix F does not satisfy the reality condition $F^{\dagger} = F$, but it is equivalent to one which does. Namely, if we multiply F on the left by

$$\begin{bmatrix} \lambda^{-1}Q & -h^{-1} \\ h & 0 \end{bmatrix}, \tag{40}$$

which is holomorphic on $\hat{U}$, then the resulting matrix *does* satisfy the reality condition. (Multiplying on the left by a matrix which is holomorphic and nonsingular on $\hat{U}$, and on the right by one which is holomorphic and nonsingular on U, simply amounts to a change of coordinates in the bundle, and does not affect A_μ or Φ.)

The gauge field generated by (38) can be expressed in terms of the function

$$\psi(t,x,y) = \oint \Gamma(i\bar{z}\lambda + 2t - iz\lambda^{-1},\lambda)\,d\lambda/(2\pi i\lambda), \tag{41}$$

which is a solution of the wave equation [cf. (12)]. The contour in (41) is $|\lambda| = 1$. Furthermore, there is a concise expression for tr Φ^2 in terms of ψ, namely

$$\operatorname{tr} \Phi^2 = \tfrac{1}{2}(\partial_t^2 - \partial_x^2 - \partial_y^2)\log \psi. \tag{42}$$

This is true no matter what Γ is. If we take Γ to be given by (39), then (41) yields

$$\psi = [2/(1-\alpha^2)]e^{-ix}\cosh[\beta(y - Vt)], \tag{43}$$

where

$$\begin{aligned} V &= -2\alpha/(1+\alpha^2), \\ \beta &= (1+\alpha^2)/(1-\alpha^2) = (1 - V^2)^{-1/2}. \end{aligned} \tag{44}$$

And then using (37) and (42) gives an expression for the sG field φ, namely

$$\cos \varphi = 1 - 2(\partial_y^2 - \partial_t^2)\log \cosh[\beta(y - Vt)], \tag{45}$$

which is the one-soliton solution of (36).

In fact, one can check that (38) does indeed generate the fields A_μ and Φ of (35), where φ is the one-soliton solution (45). This way of doing things is made slightly awkward by the gauge freedom in A_μ and Φ (corresponding to the freedom in the matrices H and $\hat{H}$ which split F). One may avoid this by using the gauge-invariant matrix J, and generate the appropriate solutions of (33) by using the "Riemann problem with zeros."[16,18,19]

Multisoliton solutions can be understood similarly in this picture. One needs to use an upper triangular matrix F as in (38), except that λ and λ^{-1} on the diagonal are replaced by λ^n and λ^{-n}, where n is a positive integer. In this case, there is a formula for tr Φ^2 as in (42), with ψ replaced by the determinant of an $n \times n$ matrix of functions, each generated by an integral like (41). The analogous formula for multisoliton solutions of the sG equation is, of course, well known. The details of all this have not yet been worked out, but it might be interesting to do so.

ACKNOWLEDGMENTS

I am grateful to the Durham University Research Foundation for research support, and to Robert Leese for useful conversations.

[1] R. Penrose, J. Math. Phys. **8**, 345 (1967).

[2] R. Penrose and R. S. Ward, "Twistors for flat and curved space-time," in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2, pp. 283–328.

[3] S. A. Huggett and K. P. Tod, *An Introduction to Twistor Theory* (Cambridge U.P., Cambridge, 1985).

[4] R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, 1986), Vol. 2.

[5] R. Penrose, Bull Amer. Math. Soc. **8**, 427 (1983).

[6] E. T. Whittaker, Math. Ann. **57**, 333 (1903); E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1965), Sec. 18.3.

[7] V. Guillemin and S. Sternberg, Lett. Math. Phys. **12**, 1 (1986).

[8] S. Helgason, *The Radon Transform* (Birkhäuser, Boston, 1980).

[9] R. O. Wells, Jr., Commun. Math. Phys. **78**, 567 (1981); T. Bailey, L. Ehrenpreiss, and R. O. Wells, Jr., Proc. R. Soc. London, Ser. A **384**, 403 (1982).

[10] N. J. Hitchin, Commun. Math. Phys. **83**, 579 (1982).

[11] R. S. Ward, Commun. Math. Phys. **79**, 317 (1981).

[12] M. G. Eastwood, R. Penrose, and R. O. Wells, Jr., Commun. Math. Phys. **78**, 305 (1981).

[13] M. K. Murray, Math. Ann. **272**, 99 (1985).

[14] R. S. Ward, Commun. Math. Phys. **80**, 563 (1981).

[15] A. Pressley and G. Segal, *Loop Groups* (Clarendon, Oxford, 1986), Secs. 8.1, 8.2.

[16] R. S. Ward, J. Math. Phys. **29**, 386 (1988).

[17] R. S. Ward, Nonlinearity **1**, 671 (1988).

[18] R. A. Leese, "Extended wave solutions in an integrable chiral model in (2 + 1) dimensions," J. Math. Phys. **30**, 2072 (1989).

[19] R. S. Ward, "Classical solutions of the chiral model, unitons, and holomorphic vector bundles," to appear in Commun. Math. Phys.

[20] L. J. Mason and G. A. J. Sparling, Phys. Letters A **137**, 29 (1989).

[21] N. M. J. Woodhouse, Class. Quant. Gravit. **6**, 933 (1989).

[22] R. S. Ward, Gen. Relativ. Gravit. **15**, 105 (1983); N. M. J. Woodhouse, Class. Quant. Gravit. **4**, 799 (1987); N. M. J. Woodhouse and L. J. Mason, Nonlinearity **1**, 73 (1988).

Multipole moments of axisymmetric systems in relativity

G. Fodor
Central Research Institute for Physics, P.O.B. 49, H-1525 Budapest 114, Hungary
C. Hoenselaers
Max-Planck-Institut für Astrophysik, 8046 Garching bei München, Federal Republic of Germany
Z. Perjés
Central Research Institute for Physics, P.O.B. 49, H-1525 Budapest 114, Hungary

(Received 15 March 1989; accepted for publication 10 May 1989)

An algorithm is developed for computing the nth gravitational multipole moment of an asymptotically flat, empty, stationary axisymmetric space-time. The moments are expressed in terms of the expansion coefficients of the Ernst potential on the axis of symmetry. The values of the first ten multipole moments are given.

I. INTRODUCTION

During the past two decades there has been slow but relentless progress in the theory of relativistic gravitational multipole moments. These moments are expected to be an effective means of generating solutions of the gravitational equations[1] as well as defining the space of relativistic states[2] once an efficient algorithm for computing them is found. The basic ideas were developed by Geroch.[3] He considered multipole moments in curved, static, asymptotically flat and empty space-times. The nth moment is given by the value at spatial infinity of a symmetric and traceless tensor $P^{(n)}_{i_1i_2\cdots i_n}(x^j)$ on the background three-space[4] of the timelike Killing trajectories.

Hansen[1] and Thorne[5] have generalized the notion of gravitational multipole moments to asymptotically flat stationary space-times. Their definitions are equivalent. A three-space $(\mathscr{M},h)$ with positive-definite metric h is called *asymptotically flat* if it can be conformally mapped to a manifold $(\widetilde{\mathscr{M}},\widetilde{h})$ and

(i) $\widetilde{\mathscr{M}} = \mathscr{M}\cup\Lambda$, where Λ is a single point;

(ii) $\Omega|_\Lambda = \Omega_{,i}|_\Lambda = 0, \quad \widetilde{D}_i\widetilde{D}_k\Omega|_\Lambda = \widetilde{h}_{ik}|_\Lambda$,

where $\widetilde{h}_{ik} = \Omega^2 h_{ik}$.

There are two sets of multipole moments describing mass and angular momentum, respectively, and they are given in terms of two potentials. These potentials are constructed from the norm $f = K^\mu K_\mu$ and curl $\psi_\mu = \epsilon_{\mu\nu\rho\sigma}K^\nu K^{\rho;\sigma}$ of the timelike Killing field K. From the vacuum Einstein equations $R_{\mu\nu}=0$, the curl is locally a gradient: $\psi_\mu = \psi_{;\mu}$. It is convenient to unify these potentials in the complex Ernst notation

$$\mathscr{E} = f + i\psi. \tag{1}$$

The complex gravitational potential

$$\xi = (1-\mathscr{E})/(1+\mathscr{E}) \tag{2}$$

is assigned the conformal weight $-\frac{1}{2}$, i.e.,

$$\widetilde{\xi} = \Omega^{-1/2}\xi. \tag{3}$$

The mass and the rotation potentials are then given by the real and imaginary parts, respectively,

$$\xi = \phi_M + i\phi_J. \tag{4}$$

The corresponding *complex* multipole tensors are defined[1] recursively:

$$\begin{aligned} &P^{(0)}(x^j) = \xi,\\ &P^{(1)}_i(x^j) = \xi_{,i},\\ &P^{(n+1)}_{i_1i_2\cdots i_{n+1}}(x^j) = \mathscr{C}\big[D_{i_{n+1}}P^{(n)}_{i_1\cdots i_n}\\ &\qquad - \tfrac{1}{2}n(2n-1)R_{i_1i_2}P^{(n-1)}_{i_3\cdots i_{n+1}}\big], \end{aligned} \tag{5}$$

where the symbol $\mathscr{C}$ denotes the operation of taking the symmetric and trace-free part. The values of the nth multipole moment are given by the smooth continuation of the conformal image of the nth multipole tensor to the point $\Lambda\in\widetilde{\mathscr{M}}$. [The conformal image $\widetilde{P}^{(n)}_{i_1i_2\cdots i_n}(x^j)$ is defined by the tilded version of Eqs. (5).]

A good candidate for the study of multipole moments is the class of axisymmetric gravitational fields. It has been shown[6] that these metrics are determined by the value of the complex gravitational potential on the axis of symmetry. It is the purpose of this paper to present an algorithm for computing the gravitational multipole moments in asymptotically flat, empty, stationary axisymmetric space-times. In Sec. II we summarize the theory of gravitational multipole moments in such space-times. The power series expansion of the potential $\widetilde{\xi}$ on the symmetry axis will be used in Sec. III for obtaining the multipole moments. In Sec. IV we shall give the results of the computation for the first ten moments.

II. MOMENTS IN AXISTATIONARY SPACE-TIMES

We first briefly review the theory of multipole moments in an asymptotically flat, empty, stationary, axisymmetric space-time.[1] We write the metric in the canonical form

$$\begin{aligned} ds^2 = {}& (1/f)[e^{2\gamma}(d\rho^2+dz^2)+\rho^2\,d\varphi^2]\\ &-f(dt-\omega\,d\varphi)^2, \end{aligned} \tag{6}$$

where the functions f, γ, and ω depend only on $x^1=\rho$ and $x^2=z$.

Since tensors on the symmetry axis $\rho=0$ are invariant under a rotation about the axis, the multipole moments $\widetilde{P}^{(n)}_{i_1\cdots i_n}|_\Lambda$ are necessarily multiples of the symmetric trace-free outer product of the axis vector n^a with itself. Hansen has defined the scalar moments by

$$P_n = (1/n!)\,\widetilde{P}^{(n)}_{i_1\cdots i_n}n^{i_1}\cdots n^{i_n}|_\Lambda\,. \tag{7}$$

Since on the axis $n^a = (0,1,0)$, the moments are

$$P_n = (1/n!)\,\tilde{P}^{(n)}_{2\cdots 2}|_\Lambda . \qquad (8)$$

Using (6), the metric on the three-dimensional manifold $\mathscr{M}$ is

$$h_{ij} = \begin{pmatrix} e^{2\gamma} & 0 & 0 \\ 0 & e^{2\gamma} & 0 \\ 0 & 0 & \rho^2 \end{pmatrix}. \qquad (9)$$

After the coordinate transformation

$$\bar{\rho} = \rho/(\rho^2 + z^2),\ \bar{z} = z/(\rho^2 + z^2),\ \bar{\varphi} = \varphi, \qquad (10)$$

the metric takes the form

$$h_{\bar{i}\bar{j}} = \frac{1}{\bar{r}^4}\begin{pmatrix} e^{2\gamma} & 0 & 0 \\ 0 & e^{2\gamma} & 0 \\ 0 & 0 & \bar{\rho}^2 \end{pmatrix}, \qquad (11)$$

where $\bar{r}^2 = \bar{\rho}^2 + \bar{z}^2$.

We drop the overbar from the coordinates. Let $\tilde{\mathscr{M}} = \mathscr{M} \cup \Lambda$, where Λ is the origin. Let the metric on $\tilde{\mathscr{M}}$ be $\tilde{h}_{ij} = \Omega^2 h_{ij} = r^4 h_{ij}$. One can easily verify that if $\gamma|_\Lambda = 1$, the space is asymptotically flat. By Eq. (3), $\tilde{\xi} = (1/r)\xi$. Then from the field equation for ξ,

$$(r^2\tilde{\xi}\tilde{\xi}^* - 1)\Delta\tilde{\xi} = 2\tilde{\xi}^*[r^2(\nabla\tilde{\xi})^2 + 2r\tilde{\xi}\nabla\tilde{\xi}\nabla r + \tilde{\xi}^2]. \quad (12)$$

The Ricci tensor is

$$\tilde{R}_{ij} = (1/D^2)\,(G_i G_j^* + G_i^* G_j), \qquad (13)$$

where $D = r^2\tilde{\xi}\tilde{\xi}^* - 1$, $G_1 = z\tilde{\xi}_1 - \rho\tilde{\xi}_2$, $G_2 = \rho\tilde{\xi}_1 + z\tilde{\xi}_2 + \tilde{\xi}$, and $G_3 = 0$. The Christoffel symbols, which are necessary for the covariant differentiation in (5), are $\Gamma^3_{13} = 1/\rho$, $\Gamma^1_{33} = -\rho e^{-2\gamma}$, $\Gamma^1_{11} = \Gamma^2_{12} = -\Gamma^1_{22} = \gamma_1$, $\Gamma^2_{22} = \Gamma^1_{12} = -\Gamma^2_{11} = \gamma_2$, and the remaining components are zero, where

$$\gamma_1 = (\rho/2)(\tilde{R}_{11} - \tilde{R}_{22}),\ \gamma_2 = \rho\tilde{R}_{12}. \qquad (14)$$

It follows from Eq. (12) that $\tilde{\xi}$ is uniquely determined by its values on the axis. Let

$$\tilde{\xi}(\rho = 0) = \sum_{n=0}^{\infty} m_n z^n.$$

It has been conjectured that $P_n = m_n$. This is obviously true for $n = 0$ and $n = 1$. For static solutions Hoenselaers has shown the conjecture to be true when $n = 2$. Furthermore Hansen[1] has given the moments of the Kerr solution as $P_n = m(ia)^n$, which conforms with the conjecture for all n. Singer[7] showed that $P_3 = m_3$. But Hauser[8] found the conjecture to be false for $n = 4$ and 5. Subsequently, Hoenselaers[9] has calculated the sixth and seventh moments.

III. GENERATING ALGORITHM

The aim of this section is to construct a recursive algorithm for generating the nth gravitational multipole moment P_n in terms of the expansion coefficients m_k. For obtaining $\tilde{P}^{(n)}_{i_1\cdots i_n}$, one has to calculate $\partial_1\tilde{P}^{(n-1)}_{i_1\cdots i_{n-1}}$ and $\partial_2\tilde{P}^{(n-1)}_{i_1\cdots i_{n-1}}$. Thus for the nth moment one needs $\partial_1^a\partial_2^b\tilde{\xi}|_\Lambda$, where $a + b \leqslant n$. Here $\tilde{\xi}$ can be written as

$$\tilde{\xi} = \sum_{i,j=0}^{\infty} a_{ij}\rho^i z^j, \qquad (15)$$

where $a_{0j} = m_j$. Putting this into (12) yields

$$\begin{aligned}(r+2)^2 a_{r+2,s} = {} & -(s+2)(s+1)a_{r,s+2} \\ & + \sum_{\substack{k+m+p=r \\ l+n+q=s}} a_{kl}a^*_{mn}\,[a_{pq}(p^2 + q^2 \\ & - 4p - 5q - 2pk - 2ql - 2) \\ & + a_{p+2,q-2}(p+2)(p+2-2k) \\ & + a_{p-2,q+2}(q+2)(q+1-2l)].\end{aligned} \quad (16)$$

Using this equation one can express the constants a_{ij} in terms of m_k. For the calculation of a given a_{ij} one needs nearly all such a_{kl} for which $i + j > k + l$, or for which $i + j = k + l$ and $k < j$. Equation (16) implies that $a_{ij} = 0$ if i is an odd number, which is necessary for $\tilde{\xi}$ to be analytic at Λ.

Define

$$P^{(n)}_{a,b,c} = \tilde{P}^{(n)}_{\underbrace{1\cdots1}_{a}\underbrace{2\cdots2}_{b}\underbrace{3\cdots3}_{c}}, \qquad (17)$$

where $a + b + c = n$. Invoking the forms of the Christoffel symbols, the recursive definition (5) of the tensors $\tilde{P}^{(n)}_{i_1\cdots i_n}$ takes the form

$$\begin{aligned}P^{(n)}_{a,b,c} = \frac{1}{n}\mathscr{C}\bigg\{ & a\frac{\partial}{\partial\rho}P^{(n-1)}_{a-1,b,c} + b\frac{\partial}{\partial z}P^{(n-1)}_{a,b-1,c} - \left[(a(a-1) + 2ab)\gamma_1 + 2ac\frac{1}{\rho}\right]P^{(n-1)}_{a-1,b,c} \\ & - [2ab + b(b-1)]\gamma_2 P^{(n-1)}_{a,b-1,c} + a(a-1)\gamma_2 P^{(n-1)}_{a-2,b+1,c} + b(b-1)\gamma_1 P^{(n-1)}_{a+1,b-2,c} \\ & + c(c-1)\rho e^{-2\gamma}P^{(n-1)}_{a+1,b,c-2} - (n - \tfrac{3}{2})\,[a(a-1)\tilde{R}_{11}P^{(n-2)}_{a-2,b,c} \\ & + 2ab\tilde{R}_{12}P^{(n-2)}_{a-1,b-1,c} + b(b-1)\tilde{R}_{22}P^{(n-2)}_{a,b-2,c}]\bigg\},\end{aligned} \qquad (18)$$

where now the symbol $\mathscr{C}$ denotes the trace-free part only.

For $\tilde{P}^{(n)}_{i_1\cdots i_n}$ we also get the correct value if we replace $\tilde{P}^{(k)}_{i_1\cdots i_k}$ by

$$\tilde{P}^{(k)}_{i_1\cdots i_k} + h_{(i_1 i_2}Q^{(k-2)}_{i_3\cdots i_k)}, \qquad (19)$$

where $k < n$, and $Q^{(k-2)}_{i_3\cdots i_k}$ is any tensor, since we get only additional terms that vanish when we take the trace-free part. As is shown in the Appendix, one can take the trace-free part by subtracting such a tensor. Define the tensors $S^{(n)}_{a,b,c}$ by writing S instead of P everywhere in Eq. (18), adding $h_{(i_1 i_2}Q^{(n-2)}_{i_3\cdots i_n)}$ on the right-hand side, and dropping the operation $\mathscr{C}$. Let

$$S^{(0)}_{0,0,0} = P^{(0)}_{0,0,0}, \quad S^{(1)}_{a,b,c} = P^{(1)}_{a,b,c}. \tag{20}$$

Then

$$P^{(n)}_{a,b,c} = \mathscr{C}(S^{(n)}_{a,b,c})^{(n)}_{a,b,c}. \tag{21}$$

We now show that $Q^{(k-2)}_{i_1\cdots i_k}$ can be chosen such that $S^{(n)}_{a,b,c} = 0$ if $c \neq 0$. This is obviously true for $n = 0$ and $n = 1$. Suppose that it is true for $k < n$. Obviously one has to choose $Q^{(n)}_{a,b,c}$ so that $Q^{(n)}_{a,b,c} = 0$ if $c \neq 0$. Then $S^{(n)}_{a,b,c} \neq 0$ only if $c = 0$ or $c = 2$. From $S^{(n)}_{a,b,2} = 0$ we get

$$Q^{(n-2)}_{a,b,0} = [(n-1)/\rho]e^{-2\gamma}S^{(n-1)}_{a+1,b,0}. \tag{22}$$

Then $S^{(n)}_{a,b,c} = 0$ if $c \neq 0$. Let $S^{(n)}_{a} = S^{(n)}_{a,b,0}$. The definition of the tensors $P^{(n)}_{a,b,c}$ takes the form

$$S^{(0)}_0 = \tilde{\xi}, \quad S^{(1)}_0 = \frac{\partial}{\partial z}S^{(0)}_0, \quad S^{(1)}_1 = \frac{\partial}{\partial \rho}S^{(0)}_0,$$

$$\begin{aligned} S^{(n)}_a = \frac{1}{n}\Big\{ & a\frac{\partial}{\partial \rho}S^{(n-1)}_{a-1} + (n-a)\frac{\partial}{\partial z}S^{(n-1)}_{a} + a\left[(a+1-2n)\gamma_1 - \frac{a-1}{\rho}\right]S^{(n-1)}_{a-1} \\ & + (a-n)(a+n-1)\gamma_2 S^{(n-1)}_a + a(a-1)\gamma_2 S^{(n-1)}_{a-2} \\ & + (n-a)(n-a-1)\left(\gamma_1 - \frac{1}{\rho}\right)S^{(n-1)}_{a+1} - \left(n-\frac{3}{2}\right)[a(a-1)\tilde{R}_{11}S^{(n-2)}_{a-2} \\ & + 2a(n-a)\tilde{R}_{12}S^{(n-2)}_{a-1} + (n-a)(n-a-1)\tilde{R}_{22}S^{(n-2)}_a]\Big\}, \end{aligned} \tag{23}$$

and

$$P^{(n)}_{a,b,c} = \mathscr{C}(S^{(n)}_a)^{(n)}_{a,b,c}, \tag{24}$$

where $\mathscr{C}$ denotes the trace-free part. For a given n $S^{(n)}_a$ has only $n+1$ components in contrast to the 3^n components of the tensor $\tilde{P}^{(n)}_{i_1\cdots i_n}$.

Let $\tilde{R}_{11} = \rho^2 R'_{11}$, $\tilde{R}_{12} = \rho R'_{12}$, $\tilde{R}_{22} = R'_{22}$, $\gamma_1 = \rho\gamma'_1$, and $\gamma_2 = \rho^2\gamma'_2$. Then it follows from (13) and (14) that R'_{11}, R'_{12}, R'_{22}, γ'_1, and γ'_2 are each power series in ρ^2 and z. Let

$$S^{(n)}_a = \rho^a Z^{(n)}_a. \tag{25}$$

Then (23) becomes

$$Z^{(0)}_0 = \tilde{\xi}, \quad Z^{(1)}_0 = \frac{\partial}{\partial z}Z^{(0)}_0, \quad Z^{(1)}_1 = \frac{\partial}{\partial \rho}Z^{(0)}_0,$$

$$\begin{aligned} Z^{(n)}_a = \frac{1}{n}\Big\{ & a\frac{1}{\rho}\frac{\partial}{\partial \rho}Z^{(n-1)}_{a-1} + (n-a)\frac{\partial}{\partial z}Z^{(n-1)}_{a} \\ & + a(a+1-2n)\gamma'_1 Z^{(n-1)}_{a-1} + (a-n)(a+n-1)\gamma'_2\rho^2 Z^{(n-1)}_a \\ & + a(a-1)\gamma'_2 Z^{(n-1)}_{a-2} + (n-a)(n-a-1)(\rho^2\gamma'_1 - 1)Z^{(n-1)}_{a+1} \\ & - \left(n-\frac{3}{2}\right)[a(a-1)R'_{11}Z^{(n-2)}_{a-2} + 2a(n-a)R'_{12}Z^{(n-2)}_{a-1}] + (n-a)(n-a-1)R'_{22}Z^{(n-2)}_a]\Big\}. \end{aligned} \tag{26}$$

One can easily verify that $Z^{(n)}_a$ again has the form of a power series in ρ^2 and z.

Since $S^{(n)}_{a,b,c} = 0$ if $c \neq 0$, so when one takes the trace-free part, h^{33}, which is singular at Λ, does not occur. Hence using the trace-free part of $S^{(n)}_a|_\Lambda$, one gets the right value for $P^{(n)}_{a,b,c}|_\Lambda$, so

$$P^{(n)}_{a,b,c}|_\Lambda = \mathscr{C}(S^{(n)}_a|_\Lambda)^{(n)}_{a,b,c}. \tag{27}$$

Using (25) we obtain that $S^{(n)}_a|_\Lambda = 0$ if $a \neq 0$. For the calculation of the scalar moments we only need $\tilde{P}^{(n)}_{2\cdots 2}|_\Lambda$. Thus we have to calculate the trace-free part of a tensor $S^{(n)}_{i_1\cdots i_n}$ for which only $S^{(n)}_{2\cdots 2} \neq 0$, and we need only the $2\cdots 2$ component of the result (Cf. Appendix). From the definition (7) of the scalar moments we get

$$P_n = [1/(2n-1)!!]S^{(n)}_0. \tag{28}$$

IV. THE COMPUTATION

The computation up to the mth moment consists of the following steps.

(1) The coefficients a_{ij} in the power series expansion (15) of $\tilde{\xi}$ are expressed in terms of m_i using the relation (16). We need the coefficients a_{ij} up to $i + j \leqslant m$.

(2) The components of the Ricci tensor $\tilde{R}_{ij}$ are obtained from (13) as polynomials of degree $m-2$ in ρ and z. The derivatives of the metric function γ are then given by Eqs. (14).

(3) Computation of the quantities $S_0^{(0)} = \bar{\xi}$, $S_0^{(1)} = (\partial/\partial z)S_0^{(0)}$, and $S_1^{(1)} = (\partial/\partial \rho)S_0^{(0)}$. The polynomial $S^{(0)}$ is required to degree m, but the polynomials $S_i^{(1)}$ are needed only to order $m-1$.

(4) Computation of the polynomials $S_a^{(n)}$ from the recursive relation (23). Their respective degrees are $m-n$. We need only the $S_a^{(n)}$'s for which $a \leqslant m-n$ and $a<n$.

(5) Evaluation of Eqs. (28) for the moments P_n.

The evaluation of the moments by computer requires a facility for differentiation and algebraic manipulation of polynomials.[10] To reduce the size of computations, it is often advantageous[11] to drop terms containing $\rho^i z^j$ whenever $i+j>n$ for some positive n. We introduce the convenient notation

$$M_{ij} = m_i m_j - m_{i-1} m_{j+1},$$

where $i > j+1$. The results of the computation of the first ten multipole moments are

$$P_0 = m_0,$$
$$P_1 = m_1,$$
$$P_2 = m_2,$$
$$P_3 = m_3,$$
$$P_4 = m_4 - \frac{1}{7} M_{20} m_0^*,$$
$$P_5 = m_5 - \frac{1}{21} M_{20} m_1^* - \frac{1}{3} M_{30} m_0^*,$$
$$P_6 = m_6 + \frac{1}{33} M_{20} m_0^{*2} m_0 - \frac{5}{231} M_{20} m_2^* - \frac{4}{33} M_{30} m_1^* - \frac{8}{33} M_{31} m_0^* - \frac{6}{11} M_{40} m_0^*,$$
$$P_7 = m_7 - \frac{3}{143} M_{20} m_0^{*2} m_1 + \frac{10}{429} M_{20} m_0^* m_1^* m_0 - \frac{5}{429} M_{20} m_3^* + \frac{15}{143} M_{30} m_0^{*2} m_0$$
$$- \frac{25}{429} M_{30} m_2^* - \frac{4}{39} M_{31} m_1^* - \frac{30}{143} M_{40} m_1^* - \frac{76}{143} M_{41} m_0^* - \frac{10}{13} M_{50} m_0^*,$$
$$P_8 = m_8 - \frac{1}{143} M_{20} m_0^{*3} m_0^2 + \frac{1}{11} M_{20} m_0^{*2} m_2 - \frac{2}{143} M_{20} m_0^* m_1^* m_1 + \frac{38}{3003} M_{20} m_0^* m_0 m_2^*$$
$$+ \frac{1}{273} M_{20} m_1^{*2} m_0 - \frac{1}{143} M_{20} m_4^* - \frac{24}{143} M_{30} m_0^{*2} m_1 + \frac{12}{143} M_{30} m_0^* m_1^* m_0$$
$$- \frac{14}{429} M_{30} m_3^* + \frac{3}{13} M_{31} m_0^{*2} m_0 - \frac{23}{429} M_{31} m_2^* + \frac{3}{13} M_{40} m_0^{*2} m_0 - \frac{15}{143} M_{40} m_2^*$$
$$- \frac{34}{143} M_{41} m_1^* - \frac{4}{13} M_{50} m_1^* - \frac{45}{143} M_{42} m_0^* - \frac{11}{13} M_{51} m_0^* - M_{60} m_0^*,$$
$$P_9 = m_9 - \frac{1}{221} M_{20} m_3 m_0^{*2} - \frac{1}{221} M_{20} m_5^* + \frac{2}{221} M_{20} m_0^{*3} m_0 m_1 - \frac{21}{2431} M_{20} m_0^{*2} m_1^* m_0^2$$
$$+ \frac{174}{2431} M_{20} m_0^* m_2 m_1^* + \frac{20}{2431} M_{20} m_0^* m_0 m_3^* - \frac{106}{17017} M_{20} m_0^* m_2^* m_1$$
$$- \frac{41}{17017} M_{20} m_1^{*2} m_1 + \frac{8}{2431} M_{20} m_1^* m_0 m_2^* - \frac{7}{221} M_{30} m_0^{*3} m_0^2$$
$$- \frac{18}{143} M_{30} m_0^* m_1^* m_1 + \frac{112}{2431} M_{30} m_0^* m_0 m_2^* + \frac{35}{2431} M_{30} m_1^{*2} m_0$$
$$- \frac{49}{2431} M_{30} m_4^* - \frac{24}{221} M_{31} m_0^{*2} m_1 + \frac{42}{221} M_{31} m_0^* m_1^* m_0 - \frac{7}{221} M_{31} m_3^*$$
$$- \frac{38}{221} M_{40} m_0^{*2} m_1 + \frac{42}{221} M_{40} m_0^* m_1^* m_0 - \frac{147}{2431} M_{40} m_3^* + \frac{7}{17} M_{41} m_0^{*2} m_0$$
$$- \frac{314}{2431} M_{41} m_2^* + \frac{7}{17} M_{50} m_0^{*2} m_0 - \frac{35}{221} M_{50} m_2^* - \frac{373}{2431} M_{42} m_1^* - \frac{87}{221} M_{51} m_1^*$$
$$- \frac{7}{17} M_{60} m_1^* - \frac{148}{221} M_{52} m_0^* - \frac{20}{17} M_{61} m_0^* - \frac{21}{17} M_{70} m_0^*,$$
$$P_{10} = m_{10} - \frac{70}{4199} M_{20} m_3 m_0^* m_1^* - \frac{63}{323} M_{30} m_3 m_0^{*2} - \frac{56}{4199} M_{30} m_5^*$$

$$-\frac{1}{323}M_{20}m_6^*-\frac{112}{4199}M_{20}m_2m_0m_0^{*3}+\frac{7}{4199}M_{20}m_0^{*4}m_0^3$$
$$-\frac{13}{4199}M_{20}m_0^{*3}m_1^2+\frac{42}{4199}M_{20}m_0^{*2}m_1^*m_0m_1-\frac{727}{138567}M_{20}m_0^{*2}m_0^2m_2^*$$
$$+\frac{38222}{969969}M_{20}m_0^*m_2m_2^*-\frac{37}{12597}M_{20}m_0^*m_1^{*2}m_0^2+\frac{274}{46189}M_{20}m_0^*m_0m_4^*$$
$$-\frac{146}{46189}M_{20}m_0^*m_1m_3^*+\frac{683}{57057}M_{20}m_2m_1^{*2}+\frac{262}{138567}M_{20}m_1^*m_0m_3^*$$
$$-\frac{106}{46189}M_{20}m_1^*m_2^*m_1+\frac{71}{138567}M_{20}m_0m_2^{*2}+\frac{294}{4199}M_{30}m_0^{*3}m_0m_1$$
$$-\frac{168}{4199}M_{30}m_0^{*2}m_1^*m_0^2+\frac{4144}{138567}M_{30}m_0^*m_0m_3^*-\frac{8960}{138567}M_{30}m_0^*m_2^*m_1$$
$$-\frac{3010}{138567}M_{30}m_1^{*2}m_1+\frac{112}{8151}M_{30}m_1^*m_0m_2^*-\frac{28}{323}M_{31}m_0^{*3}m_0^2$$
$$-\frac{264}{4199}M_{31}m_0^{*2}m_2-\frac{348}{4199}M_{31}m_0^*m_1^*m_1+\frac{4872}{46189}M_{31}m_0^*m_0m_2^*$$
$$+\frac{84}{2431}M_{31}m_1^{*2}m_0-\frac{938}{46189}M_{31}m_4^*-\frac{28}{323}M_{40}m_0^{*3}m_0^2$$
$$-\frac{28}{247}M_{40}m_0^*m_1^*m_1+\frac{4872}{46189}M_{40}m_0^*m_0m_2^*+\frac{84}{2431}M_{40}m_1^{*2}m_0$$
$$-\frac{1764}{46189}M_{40}m_4^*-\frac{321}{4199}M_{41}m_0^{*2}m_1+\frac{112}{323}M_{41}m_0^*m_1^*m_0-\frac{3626}{46189}M_{41}m_3^*$$
$$-\frac{42}{323}M_{50}m_0^{*2}m_1+\frac{112}{323}M_{50}m_0^*m_1^*m_0-\frac{392}{4199}M_{50}m_3^*+\frac{498}{4199}M_{42}m_0^{*2}m_0$$
$$-\frac{239}{2717}M_{42}m_2^*+\frac{210}{323}M_{51}m_0^{*2}m_0-\frac{924}{4199}M_{51}m_2^*+\frac{210}{323}M_{60}m_0^{*2}m_0$$
$$-\frac{70}{323}M_{60}m_2^*-\frac{1426}{4199}M_{52}m_1^*-\frac{182}{323}M_{61}m_1^*-\frac{168}{323}M_{70}m_1^*-\frac{1553}{4199}M_{53}m_0^*$$
$$-\frac{339}{323}M_{62}m_0^*-\frac{490}{323}M_{71}m_0^*-\frac{28}{19}M_{80}m_0^*.$$

Note that there exist algebraic identities among the quantities M_{ik} of the form

$$m_{a+2}M_{b+2,c}+m_cM_{a+3,b+1}=m_{b+1}M_{a+3,c},$$

where $a\geqslant b\geqslant c$. This makes it possible to rearrange some terms. We have not been able to find a closed expression for all the terms in these results. For the Kerr metric we have $M_{ik}=0$.

APPENDIX: TRACE-FREE PART OF A SYMMETRIC TENSOR

The trace-free part of a symmetric n-index tensor $T_{i_1i_2\cdots i_n}=T_{(i_1i_2\cdots i_n)}$ is obtained by adding multiples of symmetrized outer products of suitably many copies of the metric h_{ik} with the traces of the tensor T,

$$\mathscr{C}(T_{i_1\cdots i_n})_{i_1\cdots i_n}=T_{i_1\cdots i_n}+\sum_{k=1}^{[n/2]}A_k^{(n)}h_{(i_1i_2}h_{i_3i_4}\cdots h_{i_{2k-1}i_{2k}}\times h^{r_1r_2}h^{r_3r_4}\cdots h^{r_{2k-1}r_{2k}}T_{r_1\cdots r_{2k}i_{2k+1}\cdots i_n)}.\tag{A1}$$

The values of the coefficients $A_k^{(n)}$ are determined by the conditions satisfied by a symmetric and trace-free tensor $\mathscr{C}T$,

$$\mathscr{C}(T_{i_1\cdots i_n})_{i_1\cdots i_n}h^{i_1i_2}=0.\tag{A2}$$

Although these relations are valid generally in D dimensions, we are interested here only in the values of $A_k^{(n)}$ in $D=3$ dimensions. From (A2) we then get the recursion relation

$$A_{k+1}^{(n)}=-\frac{(n-2k)(n-2k-1)}{2(k+1)(2n-2k-1)}A_k^{(n)},$$

and we can define $A_0^{(n)}=1$. Thus we obtain the coefficients $A_m^{(n)}$ in the trace-free part (A1),

$$A_m^{(n)}=\frac{(-1)^m n!(2n-2m-1)!!}{2^m m!(n-2m)!(2n-1)!!}.\tag{A3}$$

As an application, we now consider a domain for which $h_{22}=h^{22}=1$ [This holds on the z axis of a regular axisymmetric space with metric (3)], and compute[12] the $22\cdots2$

component of the trace-free part of a symmetric tensor $T_{i_1\cdots i_n}$ with the only nonvanishing component $T_{22\cdots 2}$,

$$\mathscr{C}(T_{i_1\cdots i_n})_{22\cdots 2} = T_{22\cdots 2} + T_{22\cdots 2}\sum_{m=1}^{[n/2]} A_m^{(n)}$$

$$= \frac{T_{2\cdots 2}n!}{(2n-1)!!}\sum_{m=0}^{[n/2]}\frac{(-1)^m(2n-2m-1)!!}{2^m m!(n-2m)!}$$

$$= \frac{T_{2\cdots 2}n!}{(2n-1)!!}. \tag{A4}$$

[1]R. O. Hansen, J. Math. Phys. **15**, 46 (1974).
[2]R. Beig, Acta Phys. Austriaca **53**, 249 (1981).
[3]R. Geroch, J. Math. Phys. **11**, 2580 (1970).
[4]The metric of the background space of Killing trajectories is defined by the projection $h_{\mu\nu} = fg_{\mu\nu} + K_\mu K_\nu$ where local space-time coordinate indices are chosen $\mu,\nu,\ldots$. The integral curves of the vector field K (or points of the background three-space) will be labeled by $\{x^i\}$, $i = 1,2,3$. The commutator of the covariant derivatives of a three-vector v_i defines our sign convention for the curvature quantities by $(D_iD_j - D_jD_i)v_k = R_{ijkr}v^r$ and $R_{ik} = R^r_{\ irk}$.
[5]K. S. Thorne, Rev. Mod. Phys. **52**, 299 (1980).
[6]W. Simon and R. Beig, J. Math. Phys. **24**, 1163 (1983); F. J. Ernst, in *Solutions of Einstein's Equations, Techniques and Results*, edited by C. Hoenselaers and W. Dietz (Springer, Heidelberg, 1984).
[7]B. Singer, thesis Universität Würzburg, 1985.
[8]I. Hauser, private communication.
[9]C. Hoenselaers, in *Gravitational Collapse and Relativity*, edited by H. Sato and T. Nakamura (World Scientific, Singapore, 1986).
[10]At various stages of our work, we employed the algebraic language REDUCE [A. Hearn, Reduce user's manual (Rank, Santa Monica, 1983)] and POLYNOM (C. Hoenselaers, preprint 1982).
[11]When one uses a REDUCE code, this can be achieved by declaring WEIGHT $\rho = 1$, $z = 1$, WTLEVEL n. To enable differentiation again, one can cancel truncation by CLEAR ρ, z.
[12]We are indebted to J. Révai for evaluating the sum in (A4).

Poisson maps and canonoid transformations for time-dependent Hamiltonian systems

José F. Cariñena and Manuel F. Rañada
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

(Received 26 October 1988; accepted for publication 3 May 1989)

After a new presentation of the geometric theory of time-dependent systems in the Hamiltonian formulation, using Poisson structures, a characterization of canonoid transformations with respect to a dynamical vector field is given. The associated constants of motion and the generating functions of canonoid transformations are also studied. The theory is illustrated with several examples.

I. INTRODUCTION

The symplectic geometry has been shown to be the appropriate geometric setting for the description of autonomous systems in both the Hamiltonian and Lagrangian approaches, but conversely as far as time-dependent systems are concerned, there exist different alternative geometric approaches. Leaving aside the homogeneous formulations (see e.g., Ref. 1 and other references therein), the geometric approach developed in Ref. 2 seems to be the most often used. The main difference versus the time-independent approach is the geometric meaning of the time. While in that case the time is only locally defined and arises as being the parameter of the integral curves of the vector field responsible for the dynamics, in this time-dependent approach the time will be a new variable and curves obtained by a time reparametrization have to be considered as equivalent ones. That means that the relevant mathematical objects in the theory are no longer vector fields and its integral curves but one-dimensional distributions and its integral submanifolds. However, use is very often made of the possibility of choosing vector fields representatives of such distributions in such a way that the parameter of their integral curves reduces to the time variable, up to a constant time translation.

The geometric approach to the theory of canonical transformations in the time-dependent case is not so clear as in the time-independent one, mainly because the space is odd-dimensional, and the situation is worse for the theory of canonoid transformations. In order to elucidate the geometric meaning of the concept of canonoid transformation as introduced in Ref. 3, we first give in Sec. II an alternative approach for the description of time-dependent systems based on the product structure $\mathbf{R}\times M$, which allows us to define two supplementary distributions of vertical and horizontal fields and a one-to-one correspondence between these last fields and semibasic one-forms. By making use of these concepts, we introduce a Poisson bracket endowing $\mathbf{R}\times M$ with a Poisson structure. An identification is also given of those vector fields whose integral curves are obtained as solutions of Hamilton-like equations. Poisson maps and canonical transformations are analyzed from this new perspective of Poisson structures[4] in Sec. III.

We remark that not only canonical transformations in the phase space (those preserving the form of Hamilton equations of motion whatever the Hamiltonian H is) may be relevant for solving, or at least simplifying, a specific problem. On the contrary, given such a problem, there exist other transformations that will also preserve the Hamiltonian character for the equations for this particular Hamiltonian H but may not be so for another Hamiltonian function H'. They are said to be canonoid[3] with respect to H. The properties of such transformations have recently been investigated[5] in the time-independent case, using the tools of modern differential geometry. In this way, previous results obtained by Leubner and Marte[6] and by Negri *et al.*[7] were generalized and recovered in the language of symplectic manifolds, and furthermore, some additional interesting properties were obtained as, for instance, the existence of associated non-Noether constants of the motion.

One of the aims of this paper is to complete the geometric study of such canonoid transformations by covering the case of time-dependent systems. Consequently, the theory must be developed using as basic objects vector fields defined on odd-dimensional manifolds, and we must use constant rank presymplectic forms, i.e., degenerated closed two-forms, and contact structures in the sense of Ref. 2 instead of symplectic manifolds.

Once the theory of canonical transformations has been reviewed in this new approach, a suitable geometric definition of canonoid transformation with respect to a vector field may be introduced. This is made in Sec. IV, where its properties are also analyzed. The simplest case of a time-dependent two-dimensional system is used in Sec. V for an illustration of the physical meaning of the existence of canonoid transformations, and their consequences are explained; they give rise to constants of motion of a non-Noether theorem origin. The more general case of a higher-dimensional system is then presented in Sec. VI, and as well the theory of generating functions is developed in Sec. VII. Finally, several examples are collected in Sec. VIII.

II. TIME-DEPENDENT FORMALISM

Let (M,ω) be a $2n$-dimensional symplectic manifold, and let N denote the product manifold $N=\mathbf{R}\times M$. The natural coordinate on $\mathbf{R}$ will be denoted t and $\pi_1: \mathbf{R}\times M\to\mathbf{R}$ and $\pi_2: \mathbf{R}\times M\to M$ are the projections of N onto $\mathbf{R}$ and M, respectively.

The standard approach to time-dependent Hamiltonian dynamics uses the so-called[2] time-dependent vector fields X

 0022-2488/89/102258-09$02.50

in M and their suspensions to $\mathbb{R}\times M$. A time-dependent vector field in M is a map X: $\mathbb{R}\times M\to TM$ such that $X(t,m)\in T_m M$ for any $(t,m)\in\mathbb{R}\times M$. Actually, these fields are but vector fields along the projection π_2. For each such a field X, there is an associated vector field $\widetilde{X}\in\mathfrak{X}(N)$, such that

$$\pi_{2*(t,m)}[\widetilde{X}(t,m)] = X(t,m)$$

and $\langle dt,\widetilde{X}\rangle = 1$ (see e.g., Ref. 2). This vector field is called the suspension of X.

We recall that if ω is a two-form on a manifold P, then the set R_ω defined by $R_\omega = \{v\in TP / i(v)\omega = 0\}$ is called the characteristic set of ω. So a characteristic vector field is a vector field X such that $X_p\in R_\omega$ for all $p\in P$. Moreover, if ω is of constant rank then R_ω is a subbundle of TP, and if ω is also closed, then R_ω is integrable as well.

Given a time-dependent Hamiltonian function $H\in C^\infty(\mathbb{R}\times M)$, a contact form ω_H is defined by $\omega_H = \tilde{\omega} + dH\wedge dt$ where $\tilde{\omega}\in\Lambda^2(\mathbb{R}\times M)$ denotes the two-form defined in $\mathbb{R}\times M$ by the pullback $\tilde{\omega} = \pi_2{}^*(\omega)$. The dynamics is given by a field $\widetilde{X}_H$ that is the suspension to $\mathbb{R}\times M$ of a time-dependent field in M and is a characteristic field for ω_H. In this way, these fields $\widetilde{X}_H$ represent, in this odd-dimensional manifold, something similar to the Hamiltonian vector fields in a symplectic manifold. Nevertheless, this usual approach presents some difficulties for giving a correct interpretation of some fundamental concepts, as for instance, the meaning of locally Hamiltonian time-dependent vector fields or Poisson brackets of time-dependent functions.

We are now going to introduce an alternative formalism in which horizontal vector fields and semibasic forms will be used to define a Poisson structure in N and for a geometric definition of canonoid and canonical transformations.

The product structure $N = \mathbb{R}\times M$ and the natural chart for $\mathbb{R}$ permits us to define a vector field $\partial/\partial t$ that gives a basis for the $C^\infty(N)$-module of vertical (w.r.t. π_2) vector fields. Similarly, the one-form dt defines a $2n$-dimensional distribution and the vector fields in such a distribution will be called horizontal vector fields: they are annihilated by the one-form dt. We also recall that a one-form $\alpha\in\Lambda^1(N)$ is said to be semibasic (w.r.t. π_2) if the contraction of α with any vertical field vanishes, i.e., $i(\partial/\partial t)\alpha = 0$. We will denote $\mathfrak{X}^H(\mathbb{R}\times M)$ and $\Lambda^1_{\rm sb}(\mathbb{R}\times M)$, respectively, as the sets

$$\mathfrak{X}^H(\mathbb{R}\times M) = \{X\in\mathfrak{X}(\mathbb{R}\times M)\,|\,\langle dt,X\rangle = 0\}$$

and

$$\Lambda^1_{\rm sb}(\mathbb{R}\times M) = \left\{\alpha\in\Lambda^1(\mathbb{R}\times M)\,\middle|\,\alpha\left(\frac{\partial}{\partial t}\right) = 0\right\}.$$

If $\{\xi^i\}$, $i = 1,...,2n$, are local coordinates for the manifold M, the coordinate expressions for horizontal fields and semibasic forms in N in these coordinates are

$$X = f_i(t,\xi^j)\frac{\partial}{\partial\xi^i} \tag{2.1}$$

and

$$\alpha = \alpha_i(t,\xi^j)d\xi^i, \tag{2.2}$$

respectively.

Given a one-form, $\alpha\in\Lambda^1(N)$, we will denote $\alpha^{\rm sb}$ as the semibasic part of α, that is,

$$\alpha^{\rm sb} = \alpha - \left\{i\left(\frac{\partial}{\partial t}\right)\alpha\right\}dt. \tag{2.3}$$

In particular, if $F\in C^\infty(N)$, $d^{\rm sb}F$ will denote the semibasic one-form defined by

$$d^{\rm sb}F = (dF)^{\rm sb} = dF - \left\{i\left(\frac{\partial}{\partial t}\right)dF\right\}dt. \tag{2.4}$$

These properties can be extended to higher degree forms, so any k-form $\eta\in\Lambda^k(N)$ may be written as the sum of a semibasic k-form $\eta^{\rm sb}$ and a nonsemibasic k-form, and this splitting is unique. Also, $d^{\rm sb}\eta = (d\eta)^{\rm sb}$.

Two important algebraic identities to be used later are

$$d^{\rm sb}F\wedge dt = dF\wedge dt, \tag{2.5a}$$

$$i(X^h)d^{\rm sb}F = i(X^h)dF, \tag{2.5b}$$

for $F\in C^\infty(N)$ and any horizontal vector field $X^h\in\mathfrak{X}^H(\mathbb{R}\times M)$.

The form $\tilde{\omega} = \pi_2{}^*(\omega)$ is degenerated and its kernel is made up by π_2-vertical vectors. The equation $i(X)\tilde{\omega} = \alpha$ has a solution for $X\in\mathfrak{X}(N)$ if an only if α is semibasic; for instance, when α is the differential of the pullback through the projection π_2 of a function defined in M. Moreover, the solution is not unique but undetermined up to the addition of a vertical field. Nevertheless, we can define via $\tilde{\omega}$ a one-to-one linear map between the set of horizontal fields $\mathfrak{X}^H(\mathbb{R}\times M)$ and that of semibasic one-forms $\Lambda^1_{\rm sb}(\mathbb{R}\times M)$. So, if α is a semibasic one-form, the vector field X_α is the horizontal one such that $i(X_\alpha)\tilde{\omega} = \alpha$. In a similar way, for a field X in $\mathfrak{X}^H(\mathbb{R}\times M)$, the correspond one-form α_X is determined by

$$\alpha_X = i(X)\tilde{\omega} \tag{2.6}$$

and then $\langle\alpha_X,\partial/\partial t\rangle = 0$. This map is a linear bijection.

Consequently, if F: $\mathbb{R}\times M\to\mathbb{R}$ is a C^∞-function, then there is one horizontal vector field $X_F\in\mathfrak{X}^H(\mathbb{R}\times M)$ uniquely determined by the condition

$$i(X_F)\tilde{\omega} = d^{\rm sb}F. \tag{2.7}$$

Now we can give an intrinsic definition of the Poisson bracket in $\mathbb{R}\times M$ as follows.

Definition: Let (M,ω) be a symplectic manifold and $F,G\in C^\infty(\mathbb{R}\times M)$ be time-dependent functions. The Poisson bracket of F and G is the function

$$\{F,G\} = \tilde{\omega}(X_F,X_G), \tag{2.8}$$

where X_F and X_G are the horizontal vector fields associated with F and G, respectively.

Using the property (2.5b), we have

$$\tilde{\omega}(X_F,X_G) = i(X_G)d^{\rm sb}F = i(X_G)dF,$$

so that we can express $\{F,G\}$ in the alternative form

$$\{F,G\} = X_G(F) = -X_F(G). \tag{2.9}$$

The set $C^\infty(N)$ is so endowed with a Lie algebra structure. The map so defined is bilinear and skew-symmetric and for the Jacobi identity we will obtain first the value of $X_{\{G,F\}}$. If F and G are arbitrary functions in N, then using the identity

$$i([X_F,X_G])\tilde{\omega} = L_{X_F}\{i(X_G)\tilde{\omega}\} - i(X_G)L_{X_F}\tilde{\omega},$$

and the conditions defining X_F and X_G, we will obtain

$$i([X_F,X_G])\tilde{\omega} = L_{X_F}\left\{dG - \left\langle dG, \frac{\partial}{\partial t}\right\rangle dt\right\}$$
$$- i(X_G)d\left\{dF - \left\langle dF, \frac{\partial}{\partial t}\right\rangle dt\right\},$$

where we have also used that $L_{X_F}\tilde{\omega} = d\{i(X_F)\tilde{\omega}\}$, since $\tilde{\omega}$ is closed. Therefore

$$i([X_F,X_G])\tilde{\omega} = d(X_F G)$$
$$- \left[X_F\left\langle dG, \frac{\partial}{\partial t}\right\rangle - X_G\left\langle dF, \frac{\partial}{\partial t}\right\rangle\right]dt,$$

and taking into account that

$$X_F\left\langle dG, \frac{\partial}{\partial t}\right\rangle - X_G\left\langle dF, \frac{\partial}{\partial t}\right\rangle = i\left(\frac{\partial}{\partial t}\right)d\{\omega(X_G,X_F)\},$$

we see that

$$i([X_F,X_G])\tilde{\omega} = d^{\,\mathrm{sb}}\{G,F\},$$

and as the vector field $[X_F,X_G]$ is horizontal, the preceding relation shows that

$$[X_F,X_G] = X_{\{G,F\}}. \tag{2.10}$$

The Jacobi identity is a consequence of the fact that the two-form $\tilde{\omega}$ is closed. Thus $d\tilde{\omega}(X_F,X_G,X_H) = 0$, and making use of the property (2.10), this is equivalent to the Jacobi identity for F, G, and H.

Note that Eq. (2.10) means that the correspondence assigning the horizontal vector field X_F to the function F is an antihomomorphism of Lie algebras.

Moreover, the Poisson bracket defined by (2.8) gives a Poisson structure[4,8] on the manifold N. In fact, the Leibniz property,

$$\{F,G_1G_2\} = G_1\{F,G_2\} + G_2\{F,G_1\},$$

follows from the definition $\{F,G\} = -X_F(G)$, because X_F is a vector field. The skew-symmetric two-times contravariant tensor Λ defining the Poisson structure is defined by

$$\Lambda(dF,dG) = \{F,G\} = \tilde{\omega}(X_F,X_G) = -X_F(G). \tag{2.11}$$

Its rank is $2n$ and the kernel is generated by dt. The tensor Λ can be used for defining vector fields corresponding to one-forms in N and the vector fields associated to exact forms dF are the Hamiltonian vector fields X_F. The Casimir functions are the π_1-basic functions (pullback through the projection π_1 of functions defined by $\mathbb{R}$), namely, arbitrary functions of the time t. If $\{\xi^i\}$ with $i = 1,\ldots,2n$, are arbitrary local coordinates for M, then the local expression for the Poisson bracket and the vector field X_F, respectively, is given by[9]

$$\{F,G\} = \frac{\partial F}{\partial \xi^i} J^{ik} \frac{\partial G}{\partial \xi^k}, \tag{2.12}$$

$$X_F = \{\xi^i,F\}\frac{\partial F}{\partial \xi^i}, \tag{2.13}$$

where J^{ik} denotes the fundamental Poisson brackets

$$J^{ik} = \{\xi^i,\xi^k\}. \tag{2.14}$$

In particular, if $\{q^i,p_i\}$ are Darboux coordinates in M for the symplectic form ω, then the coordinate expression for the Poisson bracket $\{F,G\}$ is

$$\{F,G\} = \sum_{i=1}^{n} \frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i}. \tag{2.15}$$

As a first application of the above definition, we will deduce the geometrical characterization of the Poisson bracket theorem,[3] according to which a time evolution of a system is generated by some Hamiltonian if and only if for every pair of dynamical variables R, S, the well-known following relation holds,

$$\frac{d}{dt}\{R,S\} = \left\{\frac{dR}{dt},S\right\} + \left\{R,\frac{dS}{dt}\right\},$$

that in geometric terms reads

$$\Gamma(\{R,S\}) = \{\Gamma(R),S\} + \{R,\Gamma(S)\},$$

where Γ is the vector field giving the dynamics.

It has recently been proved[5] that for time-independent systems [that is, $\Gamma\in\mathfrak{X}(M)$], this relation is equivalent to Γ being locally Hamiltonian (that is, $L_\Gamma\omega = 0$). The following theorem presents the extension of this equivalence to the more general case of time-dependent systems.

Theorem: Let (M,ω) be a symplectic manifold and $\Gamma\in\mathfrak{X}(\mathbb{R}\times M)$ be a vector field. Then the relation

$$\Gamma(\{F,G\}) = \{\Gamma(F),G\} + \{F,\Gamma(G)\} \tag{2.16}$$

holds if and only if Γ preserves the horizontal distribution and $L_\Gamma\tilde{\omega}$ annihilates any pair of horizontal fields.

Proof: The Lie derivative of the Poisson bracket $\{F,G\}$ is given by

$$L_\Gamma\{F,G\} = (L_\Gamma\tilde{\omega})(X_F,X_G) + \tilde{\omega}([\Gamma,X_F],X_G)$$
$$+ \tilde{\omega}(X_F,[\Gamma,X_G]),$$

where X_F and X_G are the horizontal vector fields associated to the semibasic differentials $d^{\mathrm{sb}}F$ and $d^{\mathrm{sb}}G$.

For $\tilde{\omega}([\Gamma,X_F],X_G)$ we get

$$\begin{aligned}\tilde{\omega}([\Gamma,X_F],X_G) &= i(X_G)(L_\Gamma d^{\,\mathrm{sb}}F) - (L_\Gamma\tilde{\omega})(X_F,X_G)\\ &= i(X_G)d\Gamma(F) - \left\{i\left(\frac{\partial}{\partial t}\right)dF\right\}\{i(X_G)d\Gamma(t)\}\\ &\quad - (L_\Gamma\tilde{\omega})(X_F,X_G)\\ &= \{\Gamma(F),G\} - \left\{i\left(\frac{\partial}{\partial t}\right)dF\right\}\{i(X_G)d\Gamma(t)\}\\ &\quad - (L_\Gamma\tilde{\omega})(X_F,X_G),\end{aligned}$$

where we have used that $i(X_G)dt = 0$ and

$$i(X_G)d^{\,\mathrm{sb}}\Gamma(F) = i(X_G)d\Gamma(F).$$

A similar expression is obtained for $\tilde{\omega}(X_F[\Gamma,X_G])$, so, finally, $L_\Gamma\{F,G\}$ turns out to be

$$L_\Gamma\{F,G\} = \{\Gamma(F),G\} + \{F,\Gamma(G)\} - (L_\Gamma\tilde{\omega})(X_F,X_G)$$
$$+ \left[\left\{i\left(\frac{\partial}{\partial t}\right)dG\right\}i(X_F)\right.$$
$$\left. - \left\{i\left(\frac{\partial}{\partial t}\right)dF\right\}i(X_G)\right]d\Gamma(t).$$

Therefore, if we assume that Γ is such that $\langle dt,\Gamma\rangle$ only depends on t and $L_\Gamma\tilde{\omega}$ annihilates any pair of horizontal fields, then the relation (2.16) is true. Conversely, if this relation holds, necessarily $(L_\Gamma dt)(X_F)=0$ and $(L_\Gamma \tilde{\omega})(X_F,X_G)=0$ for any pair of functions. □

Notice that the first of these two conditions,

$$\text{Ker}\, L_\Gamma(dt)=\mathfrak{X}^H(\mathbb{R}\times M)$$

means that Γ preserves $\text{Ker}(dt)$, or in an equivalent way, that the horizontal distribution $\mathfrak{X}^H(\mathbb{R}\times M)$ is invariant under Γ, namely,

$$[\Gamma,\mathfrak{X}^H(\mathbb{R}\times M)]\subset\mathfrak{X}^H(\mathbb{R}\times M).$$

In relation to the second condition it seems that, for those vector fields preserving $\mathfrak{X}^H(\mathbb{R}\times M)$, the vanishing of the restriction to that distribution of $L_\Gamma\tilde{\omega}$ is the analog in $\mathbb{R}\times M$ of being locally Hamiltonian in a symplectic manifold. Remark that this means that $(L_\Gamma\tilde{\omega})^{\text{sb}}=0$.

A general vector field in $\mathbb{R}\times M$ is locally written in the form

$$\Gamma=a^k\frac{\partial}{\partial q^k}+b^k\frac{\partial}{\partial p_k}+c\frac{\partial}{\partial t}.$$

Then the first condition, that characterizes only the vertical part of Γ, tells us that the coefficient c must be a function of t along, $c=c(t)$, while the second one, that characterizes only the horizontal part, concerns the coefficients a^k and b^k by imposing the well-known relations $\partial a^k/\partial q^i = -\partial b^k/\partial p_i$.

The above theorem asserts that vector fields Γ, for which $\langle dt,\Gamma\rangle=c(t)\neq 1$, can be admissible for the description of time-dependent Hamiltonian dynamics. In this case, their integral curves will be parametrized by a parameter s that does not correspond with the time t, but $dt/ds=c(t)$ on the integral curves. The following proposition studies the particular case of vector fields for which $\langle dt,\Gamma\rangle=1$ and gives a more direct characterization of the locally Hamiltonian behavior for them.

Proposition: Let $\Gamma\in\mathfrak{X}(\mathbb{R}\times M)$ satisfy $\langle dt,\Gamma\rangle=1$, and define $\omega_\Gamma\in\Lambda^2(\mathbb{R}\times M)$ by

$$\omega_\Gamma=\tilde{\omega}+i(\Gamma)\tilde{\omega}\wedge dt.$$

Then

$$(L_\Gamma\tilde{\omega})(X^h,Y^h)=0,\quad \forall X^h,Y^h\in\mathfrak{X}^H(\mathbb{R}\times M)$$

is equivalent to $L_\Gamma\omega_\Gamma=0$.

Proof: If $L_\Gamma\tilde{\omega}$ annihilates any pair of horizontal fields, then we can write

$$L_\Gamma\tilde{\omega}=\delta\wedge dt,\quad \delta\in\Lambda^1(\mathbb{R}\times M).$$

But as $L_\Gamma\tilde{\omega}=d[i(\Gamma)\tilde{\omega}]$, we have $d[i(\Gamma)\tilde{\omega}]=\delta\wedge dt$, and therefore

$$\begin{aligned} d\omega_\Gamma &= d[\tilde{\omega}+i(\Gamma)\tilde{\omega}\wedge dt] \\ &= d\tilde{\omega}+(L_\Gamma\tilde{\omega})\wedge dt \\ &= 0. \end{aligned}$$

Since $\langle dt,\Gamma\rangle=1$, the vector field Γ is characteristic of ω_Γ, and consequently $L_\Gamma\omega_\Gamma=0$.

Conversely, let us assume that $L_\Gamma\omega_\Gamma=0$ holds. Then,

$$L_\Gamma\tilde{\omega}+[i(\Gamma)L_\Gamma\tilde{\omega}]\wedge dt=0,$$

and therefore $L_\Gamma\tilde{\omega}$ annihilates any pair of horizontal vector fields. □

III. POISSON MAPS AND CANONICAL TRANSFORMATIONS

The time-preserving Poisson transformations $\Phi: N\to N$ of (N,Λ), i.e., such that $\Phi^*t=t$, will play a relevant role in the time-dependent Hamiltonian formalism in much the same way as in the autonomous case. They are time-preserving diffeomorphisms $\Phi: N\to N$ such that $\Phi^*\Lambda=\Lambda$, or in other words, $\{F,G\}\circ\Phi=\{F\circ\Phi, G\circ\Phi\}$ for any pair of functions $F,G\in C^\infty(N)$. A characterization of such transformations in terms of the transformation properties of $\tilde{\omega}$ is given in the following theorem.

Theorem: A time preserving diffeomorphism $\Phi: N\to N$ is a Poisson map if and only if there exists one semibasic one-form κ such that

$$\Phi^*\tilde{\omega}=\tilde{\omega}+\kappa\wedge dt. \tag{3.1}$$

Proof: The condition for $\Phi: N\to N$ to be a Poisson map is written using the definition (2.8) of the Poisson bracket as follows:

$$\Phi^*[\tilde{\omega}(X_F,X_G)]=\tilde{\omega}(X_{F\circ\Phi},X_{G\circ\Phi}),\quad \forall F,G\in C^\infty(N),$$

or in a different, but equivalent, way,

$$\Phi^*[i(X_G)dF]=i(X_{G\circ\Phi})d(F\circ\Phi).$$

The left-hand side of this expression can also be rewritten in such a way that the preceding relation becomes

$$i(\Phi^{-1}{}_*X_G)d(F\circ\Phi)=i(X_{G\circ\Phi})d(F\circ\Phi),$$

and this shows that Φ is a Poisson map if and only if

$$\Phi^{-1}{}_*X_G=X_{G\circ\Phi},\quad \forall G\in C^\infty(N). \tag{3.2}$$

If Φ is time-preserving, the image under Φ of the definition (2.7) of the horizontal field X_F is

$$i(\Phi^{-1}{}_*X_F)\Phi^*\tilde{\omega}=d(F\circ\Phi)-\Phi^*\left(\left\langle dF,\frac{\partial}{\partial t}\right\rangle\right)dt. \tag{3.3}$$

Let us assume that there exists a semibasic one-form κ such that $\Phi^*\tilde{\omega}=\tilde{\omega}+\kappa\wedge dt$. By contracting this relation with the vector field $\partial/\partial t$, we see that $\kappa=-i(\partial/\partial t)\Phi^*\tilde{\omega}$. Then, using that $\Phi^{-1}{}_*X_F$ is horizontal since Φ is time preserving, (3.3) becomes

$$i(\Phi^{-1}{}_*X_F)\tilde{\omega}=d(F\circ\Phi)-\left\{i(\Phi^{-1}{}_*X_F)\kappa + \Phi^*\left(\left\langle dF,\frac{\partial}{\partial t}\right\rangle\right)\right\}dt. \tag{3.4}$$

Now, taking into account that $\kappa=-i(\partial/\partial t)\Phi^*\tilde{\omega}$, the term $i(\Phi^{-1}{}_*X_F)\kappa$ can be replaced on the right-hand side of (3.4) by

$$i(\Phi^{-1}{}_*X_F)i\left(\frac{\partial}{\partial t}\right)\Phi^*\tilde{\omega}=-(\Phi^*\tilde{\omega})\left(\frac{\partial}{\partial t'}\Phi^{-1}{}_*X_F\right),$$

and therefore (3.4) becomes

$$\begin{aligned} i(\Phi^{-1}{}_*X_F)\tilde{\omega} &= d(F\circ\Phi)-\Phi^*\left(\left\langle dF,\frac{\partial}{\partial t}\right\rangle\right)dt \\ &\quad -\left\langle d^{\text{sb}}F,\Phi_*\left(\frac{\partial}{\partial t}\right)\right\rangle dt \\ &= d^{\text{sb}}(\Phi^*F). \end{aligned}$$

Since Φ is time-preserving, the vector field $\Phi^{-1}{}_*X_F$ is horizontal, and in this way we see that $\Phi^{-1}{}_*X_F = X_{F\circ\Phi}$. Therefore Φ will be a Poisson map.

Conversely, let us assume that Φ is time-preserving and (3.2) is true for any function $G\in C^\infty(N)$. Taking into account that

$$i(\Phi^{-1}{}_*X_G)\tilde{\omega} = d^{\rm sb}(\Phi^*G) = d(\Phi^*G) - \left\langle d(\Phi^*G), \frac{\partial}{\partial t}\right\rangle dt$$

transforms under Φ into

$$i(X_G)\Phi^{-1*}\tilde{\omega} = dG - \Phi^{-1*}\left\langle d(\Phi^*G), \frac{\partial}{\partial t}\right\rangle dt, \qquad (3.5)$$

together with $i(X_G)\tilde{\omega} = d^{\rm sb}G$, we obtain that $\Phi^*\tilde{\omega} - \tilde{\omega}$ annihilates any pair of horizontal vector fields X_F, X_G and since they give a local basis for $\mathfrak{X}(N)$, there will exist a one-form δ such that $\Phi^*\tilde{\omega} - \tilde{\omega}$ can be written as $\Phi^*\tilde{\omega} - \tilde{\omega} = \delta\wedge dt$, and then $\kappa = -\Phi^*\delta$ is such that $\Phi^*\tilde{\omega} = \tilde{\omega} + \kappa\wedge dt$. □

Note that this transformation law can be rewritten as $(\Phi^*\tilde{\omega})^{\rm sb} = \tilde{\omega}$.

Lemma: Let α be a semibasic one-form. Then, $\alpha\wedge dt$ is closed if and only if α is the semibasic part of a closed form.

Proof: Let

$$\alpha = a_i(q^k,p_k,t)dq^i + b^i(q^k,p_k,t)dp_i,$$

the local expression of α in Darboux coordinates for $\tilde{\omega}$. Then, α is closed if and only if

$$\frac{\partial a_i}{\partial q^k} = \frac{\partial a_k}{\partial q^i}, \quad \frac{\partial b^j}{\partial p_k} = \frac{\partial b^k}{\partial p_j}, \quad \frac{\partial a_i}{\partial p_k} = \frac{\partial b^k}{\partial q^i},$$

and this means that there exists a function H such that

$$a_j = \frac{\partial H}{\partial q^j}, \quad b^j = \frac{\partial H}{\partial p_j},$$

from which we see that α locally coincides with $d^{\rm sb}H$. The converse is obvious because if $\alpha = \beta^{\rm sb}$ with β a closed one-form, then $\alpha\wedge dt = \beta\wedge dt$ and therefore $\alpha\wedge dt$ is closed. □

Let us remark that since ω is closed, the semibasic one-form κ arising in (3.1) is such that $d\kappa\wedge dt = 0$, and according to the previous lemma, this means that there will locally exist a function K_Φ such that $\kappa = d^{\rm sb}K_\Phi$.

The characterization of time-preserving Poisson diffeomorphisms in terms of $\tilde{\omega}$ supports the definition of canonical transformation, although some authors apply it only for the case corresponding to κ being an exact form. Actually, Eq. (3.1) of the above theorem corresponds to the property (S) of Ref. 2. Hereafter we will refer to such canonical transformations as Poisson maps. When we have a one-parameter subgroup Φ_S of such canonical transformations, then there will exist a family κ_S of semibasic one-forms such that $\Phi_S{}^*\tilde{\omega} = \tilde{\omega} + \kappa_S\wedge dt$ and therefore if X is the infinitesimal generator of Φ_S, then $L_X\tilde{\omega} = \zeta\wedge dt$, where ζ is the semibasic one-form $\zeta = ((d/ds)\kappa_S\,|_{S=0})^{\rm sb}$. The vector field X is horizontal because the Φ_S are time-preserving.

The form ζ also satisfies that $d\zeta\wedge dt = 0$ and therefore it is possible to find, at least locally, a function K such that $L_X\tilde{\omega} = dK^{\rm sb}\wedge dt$. This property is equivalent to saying that $L_X\tilde{\omega}$ vanishes on every pair of horizontal vector fields. On the other hand, the condition indicating that Φ_S is a subgroup of canonical transformation $L_X\Lambda = 0$, reduces to that of the Poisson bracket theorem because of

$$\begin{aligned}(L_X\Lambda)(dF,dG) &= L_X\{F,G\} - \Lambda\{d(XF),dG\} - \Lambda(dF,d(XG)) \\ &= L_X\{F,G\} - \{X(F),G\} - \{F,X(G)\}.\end{aligned}$$

A general vector field will be called locally Hamiltonian with respect to the Poisson structure Λ if $L_X\Lambda = 0$. Notice that in such a case $i(X)dt$ must only be a function of t, because the preceding relation shows that $\{i(X)dt,F\} = 0$, $\forall F\in C^\infty(\mathbb{R}\times M)$, and therefore $i(X)dt$ can only depend on t.

Let Γ be a vector field in $\mathbb{R}\times M$ such that $i(\Gamma)dt = 1$, α a semibasic one-form, and define $\omega_\alpha = \tilde{\omega} + \alpha\wedge dt$. Then $i(\Gamma)\omega_\alpha = 0$ if and only if $i(\Gamma)\tilde{\omega} = \alpha$ and $i(\Gamma)\alpha = 0$.

Proposition: Let α be a semibasic one-form in N. The two-form $\alpha\wedge dt$ is closed if and only if $d\alpha$ annihilates any pair of horizontal fields.

Proof: If $\alpha\wedge dt$ is closed, then $d\alpha\wedge dt = 0$. This means that $d\alpha$ is of the form

$$d\alpha = n\wedge dt, \quad n\in\Lambda^1_{\rm sb}(\mathbb{R}\times M),$$

and thus

$$(d\alpha)(X^h,Y^h) = 0, \quad \forall X^h,Y^h\in\mathfrak{X}^H(\mathbb{R}\times M).$$

Conversely, if

$$(d\alpha)(X^h,Y^h) = 0, \quad \forall X^h,Y^h\in\mathfrak{X}^H(\mathbb{R}\times M),$$

we can write $d\alpha = n\wedge dt$ and therefore $d\alpha\wedge dt = d(\alpha\wedge dt) = 0$. □

According to this proposition, if Γ satisfies $(L_\Gamma\tilde{\omega})(X^h,Y^h) = 0$, then $d\omega_\alpha = 0$ and the pair $(\mathbb{R}\times M,\omega_\alpha)$ is a contact manifold,[2] that is, the manifold is odd-dimensional and ω_α is a presymplectic form of maximal rank. Moreover, there will exist a closed one-form ζ_α such that $\alpha = (\zeta_\alpha)^{\rm sb}$. The one-form ζ_α can be locally expressed by the differential of a function H and therefore α is locally written as $\alpha = d^{\rm sb}H$ and ω_α as ω_H. The vector field Γ is then a characteristic field of ω_α and the coordinate expression of Γ in Darboux coordinates for $\tilde{\omega}$ is

$$\Gamma = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t}.$$

If Φ is a time-preserving Poisson map, then using $\Phi^*\tilde{\omega} = \tilde{\omega} + \kappa\wedge dt$, we obtain for any $(\mathbb{R}\times M,\omega_\alpha = \omega_\Gamma)$ a new contact structure $(\mathbb{R}\times M,\omega_\beta)$, with ω_β defined by

$$\begin{aligned}\omega_\beta &= \omega_{\Phi_*\Gamma} \\ &= \tilde{\omega} + \{(\Phi_*\Gamma)\tilde{\omega}\}\wedge dt\end{aligned}$$

and satisfying $\Phi^*(\omega_\beta) = \omega_\alpha$. Conversely, any time-preserving diffeomorphism Φ: $\mathbb{R}\times M\to\mathbb{R}\times M$ which satisfies $\Phi^*(\omega_{\Phi_*\Gamma}) = \omega_\Gamma$ for any vector field Γ such that $i(\Gamma)dt = 1$ is a Poisson map.

We can reformulate this last relation in a way that allows us to extend the above characterization to maps between different manifolds.

Let (M_a,ω_a) and (M_b,ω_b) be symplectic manifolds, M_{ab} the product manifold

$$M_{ab} = \mathbb{R}\times M_a\times M_b, \quad \pi_i\colon M_{ab}\to\mathbb{R}\times M_i,$$

the projections onto $\mathbb{R}\times M_i$, $i = a,b$, and

$$\Omega = \pi_a{}^*(\tilde\omega_a) - \pi_b{}^*(\tilde\omega_b),$$

where $\tilde\omega_i = \pi_2{}^*(\omega_i)$. If $V\in\mathfrak{X}(\mathbf{R}\times M_a\times M_b)$, we write Ω_V for the two-form $\Omega_V = \Omega + i(V)\Omega\wedge dt$. Then the time-preserving diffeomorphism Φ: $\mathbf{R}\times M_a \to \mathbf{R}\times M_b$ is a Poisson map if and only if $i_\Phi{}^*(\Omega_V) = 0$ for any vector field V of the form $V = \Gamma\times\Phi_*\Gamma$, with $\Gamma\in\mathfrak{X}(\mathbf{R}\times M_a)$, $i(\Gamma)dt = 1$, G_Φ is the graph of Φ, and i_Φ is the natural inclusion i_Φ: $G_\Phi \to \mathbf{R}\times M_a\times M_b$.

The two-form $i_\Phi{}^*(\Omega_V)$ is closed and its action on $\mathfrak{X}(G_\Phi)$ is given by

$$\begin{aligned} i_\Phi{}^*(\Omega_V)(Z_1,Z_2) &= \pi_a{}^*[\omega_a + i(\pi_{a*}V)\omega_a\wedge dt](Z_1,Z_2) \\ &\quad - \pi_b{}^*[\omega_b + i(\pi_{b*}V)\omega_b\wedge dt](Z_1,Z_2). \end{aligned}$$

As $\mathbf{R}\times M_a$ and G_Φ are diffeomorphic, the fields $Z_1,Z_2\in\mathfrak{X}(G_\Phi)$ may be written as $Z_1 = (X,\Phi_*X)$, and $Z_2 = (Y,\Phi_*Y)$. So

$$\begin{aligned} i_\Phi{}^*(\Omega_V)(Z_1,Z_2) &= \{[\omega_a + i(\pi_{a*}V)\omega_a\wedge dt] \\ &\quad - \Phi^*[\omega_b + i(\pi_{b*}V)\omega_b\wedge dt]\}(X,Y), \end{aligned}$$

and if V is of the form $V = \Gamma\times\Phi_*\Gamma$, i.e., the field V is tangent to G_Φ, we obtain

$$i_\Phi{}^*(\Omega_V)(Z_1,Z_2) = \{(\omega_a)_\Gamma - \Phi^*(\omega_b)_{\Phi_*\Gamma}\}(Z_1,Z_2).$$

Therefore the vanishing of $i_\Phi{}^*(\Omega_V)$ means $\Phi^*(\omega_b)_{\Phi_*\Gamma} = (\omega_a)_\Gamma$.

An equivalent approach is to characterize such Poisson maps by the condition $[i_\Phi{}^*(\Omega)]^{sb} = 0$. In any case, we can use the corresponding relation to prove, the same as for symplectic transformations, the existence of a generating function.

IV. CANONOID TRANSFORMATIONS

The study of canonoid transformations in the time-independent case has recently been carried out from the geometric point of view.[5,10] We recall that a cononoid transformation[3] with respect to a Hamiltonian function H is a diffeomorphism that preserves the form of Hamilton's equations for this particular H. Therefore the aforementioned property $\Phi^*(\omega_K) = \omega_H$, characterizing canonical transformations, does not hold.

We introduce next a geometric definition of canonoid transformation with respect to a vector field satisfying the conditions $i(\Gamma)dt = 1$, and

$$(L_\Gamma\tilde\omega)(X^h,Y^h) = 0, \quad \forall X^h,Y^h\in\mathfrak{X}^H(\mathbf{R}\times M),$$

or in an equivalent way $L_\Gamma\omega_\Gamma = 0$. In the particular case of Γ being globally defined as $\Gamma = X_H + \partial/\partial t$ [X_H is the horizontal vector field associated to $H\in C^\infty(\mathbf{R}\times M)$], then the definition will reduce to the concept of canonoid transformation w.r.t. the Hamiltonian H.

Definition: Let (M,ω) be a sympletic manifold and $\Gamma\in\mathfrak{X}(\mathbf{R}\times M)$ a vector field which satisfies $i(\Gamma)dt = 1$ and

$$(L_\Gamma\tilde\omega)(X^h,Y^h) = 0, \quad \forall X^h,Y^h\in\mathfrak{X}^H(\mathbf{R}\times M). \tag{4.1}$$

Then, a diffeomorphism $\Phi\in\mathrm{Diff}(\mathbf{R}\times M)$ is called a canonoid transformation with respect to the field Γ if it preserves the time and the transformed field $\Phi_*\Gamma\in\mathfrak{X}(\mathbf{R}\times M)$ is such that

$$(L_{\Phi_*\Gamma}\tilde\omega)(X^h,Y^h) = 0, \quad \forall X^h,Y^h\in\mathfrak{X}(\mathbf{R}\times M).$$

If Φ is canonoid with respect to Γ, then

$$(L_{\Phi_*\Gamma}\tilde\omega)(X^h,Y^h) = 0, \quad X^h,Y^h\in\mathfrak{X}^H(\mathbf{R}\times M),$$

and this means that the semibasic one-form $\beta\in\Lambda^1(\mathbf{R}\times M)$, defined by $\beta = i(\Phi_*\Gamma)\tilde\omega$, is such that the two-form ω_β defined by $\omega_\beta = \tilde\omega + \beta\wedge dt$ is closed. Therefore the pair $(\mathbf{R}\times M,\omega_\beta)$ turns out to be a contact manifold[2] with the transformed field $\Phi_*\Gamma$ generating the characteristic line bundle of ω_β.

Locally, there is a function K such that the form β is given by

$$\beta = d^{sb}K,$$

and therefore ω_β can be written as $\omega_\beta = \omega_K$ with $\omega_K = \tilde\omega + dK\wedge dt$. In coordinates, if $(q^1,\ldots,q^n,p_1,\ldots,p_n)$ is a set of Darboux coordinates in M for ω, we have the following expression for Γ:

$$\Gamma = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t},$$

and therefore if Φ is canonoid with respect to Γ, then the transformed field $\Phi_*\Gamma$ is given by

$$\Phi_*\Gamma = \frac{\partial K}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial K}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t}.$$

The first thing to be noticed is that the new local Hamiltonian function K is determined jointly by Γ and Φ and that, according to this definition, the associated two-form ω_K, that will be invariant along the integral curves of the field $\Phi_*\Gamma$, is not given by $\Phi^*(\omega_K) = \omega_H$ but its relation with ω_H will be, when obtained, something not so simple as a direct pullback.

Moreover, the transformed field is characteristic of the form ω_β. Since $i(\Phi_*\Gamma)\omega_\beta = \Phi^{-1*}[i(\Gamma)\Phi^*\omega_\beta]$, we see that the primitive field Γ is simultaneously a characteristic vector field for two different contact structures $(\mathbf{R}\times M,\omega_\alpha)$ and $(\mathbf{R}\times M,\Phi^*\omega_\beta)$. This property will be proved later to have very interesting consequences.

V. CANONOID TRANSFORMATIONS FOR TWO-DIMENSIONAL HAMILTONIAN DYNAMICAL SYSTEMS

Let Φ be a time-preserving diffeomorphism Φ: $\mathbf{R}\times M\to\mathbf{R}\times M$. Then Φ induces for every value of t a unique diffeomorphism Φ_t on M such that $\Phi_t(m) = \pi_2\circ\Phi(t,m)$.

If $\dim M = 2$, a two-form on M that is not zero at any point is a volume element and thus any two arbitrary nondegenerate two-forms must be proportional. Therefore there will exist a function $f_t\in C^\infty(M)$ associated to every Φ_t such that

$$\Phi_t{}^*(\omega) = f_t\omega. \tag{5.1}$$

This property will permit us to prove that, when $\dim M = 2$, it is possible to express the two-form $\Phi^*(\omega_\beta)$ in a form closely related with ω_α.

Since $\Phi^*(dt) = dt$, the pullback of ω_β is

$$\Phi^*(\omega_\beta) = \Phi^*(\tilde\omega) + \Phi^*(\beta)\wedge dt. \tag{5.2}$$

Using (5.1), we obtain that $\Phi^*(\tilde{\omega})$ is given by

$$\Phi^*(\tilde{\omega}) = f\tilde{\omega} + \omega',$$

where the function $f \in C^\infty(\mathbb{R}\times M)$ is defined by $f(t,m) = f_t(m)$ and the two-form ω' involves the dt dependence. This means that ω' can be written as $\omega' = \delta\wedge dt$, with $\delta\in\Lambda^1(\mathbb{R}\times M)$, and consequently it satisfies the equality

$$\left[i\left(\frac{\partial}{\partial t}\right)\omega'\right]\wedge dt = -\omega'.$$

Concerning the second term arising in (5.2), using the definition of β, we get

$$\Phi^*(\beta)\wedge dt = [i(\Gamma)\Phi^*(\tilde{\omega})]\wedge dt,$$

and therefore it turns out to be

$$\Phi^*(\beta)\wedge dt = f[i(\Gamma)\tilde{\omega}]\wedge dt + [i(\Gamma)\omega']\wedge dt.$$

So combining the two expressions, we obtain

$$\Phi^*(\omega_\beta) = f\omega_\alpha. \tag{5.3}$$

Locally, the form ω' reads

$$\omega' = g_1(q,p,t)dq\wedge dt + g_2(q,p,t)dp\wedge dt,$$

where (q,p) are local Darboux coordinates for ω in an open U of M and the two functions $q_i \in C^\infty(\mathbb{R}\times U)$, $i = 1,2$ are given by

$$g_1 = \frac{\partial Q}{\partial q}\frac{\partial P}{\partial t} - \frac{\partial Q}{\partial t}\frac{\partial P}{\partial q}$$

and

$$g_2 = \frac{\partial Q}{\partial p}\frac{\partial P}{\partial t} - \frac{\partial Q}{\partial t}\frac{\partial P}{\partial p},$$

respectively, where Q and P denote $Q = \Phi^*(q)$ and $P = \Phi^*(p)$, as usual. In this case, ω_α is written as ω_H, $H\in C^\infty(\mathbb{R}\times U)$, and Eq. (5.3) reduces to $\Phi^*(\omega_K) = f\omega_H$, a relation that replaces, for time-dependent two-dimensional dynamical systems, to the equality $\Phi^*(\omega_K) = \omega_H$, obtained in the case of the transformation Φ being canonical.

The relevance of this function f is pointed out in the following theorem.

Theorem: Let $\Phi\in\mathfrak{X}(\mathbb{R}\times M)$ be a vector field that satisfies (4.1) and $\langle dt,\Gamma\rangle = 1$. If $\omega_\alpha\in\Lambda^2(\mathbb{R}\times M)$ is the contact form associated to the field Γ, defined by $\omega_\alpha = \tilde{\omega} + [i(\Gamma)(\omega)]\wedge dt$, and $\Phi\in\mathrm{Diff}(\mathbb{R}\times M)$ is a time-preserving diffeomorphism, then Φ is a canonoid transformation with respect to the field Γ if and only if the function $f\in C^\infty(\mathbb{R}\times M)$ defined by $\Phi^*(\omega_\beta) = f\omega_\alpha$, where ω_β denotes the contact form associated to the transformed field $\Phi_*\Gamma$, is constant along the integral curves of Γ.

Proof: If Φ is canonoid with respect to Γ then we have $L_{\Phi_*\Gamma}(\omega_\beta) = 0$, and so

$$L_{\Phi_*\Gamma}(\omega_\beta) = \Phi^*[L_\Gamma(\Phi^*\omega_\beta)] = \Phi^*[L_\Gamma(f\omega_\alpha)] = 0.$$

Now, we can write

$$L_\Gamma(f\omega_\alpha) = \Gamma(f)\omega_\alpha + f(L_\Gamma\omega_\alpha),$$

thus since $L_\Gamma\omega_\alpha = 0$, we obtain $\Gamma(f) = 0$.

Conversely, assume $\Gamma(f) = 0$. Then,

$$\Gamma(f)\omega_\alpha = L_\Gamma(f\omega_\alpha) = L_\Gamma[\Phi^*(\omega_\beta)] = 0,$$

and the theorem is proved. □

In a similar way as was obtained for the time-independent case,[5] the function f, when expressed in local coordinates, turns out to be the Poisson bracket $\{Q,P\}$. This theorem corresponds to the geometrical approach of the time-dependent case of a result by Leubner and Marte.[6] It generalizes the canonical case for which f takes the value $f = 1$, and introduces a fundamental connection between canonoid transformations and the existence of constants of the motion.

VI. 2*n*-DIMENSIONAL HAMILTONIAN DYNAMICAL SYSTEMS

In this section we shall study the higher dimensional case, $\dim M = 2n > 2$. In this case, the two-form ω_α can be used for obtaining a volume element $\omega_\alpha^{\wedge n}\wedge dt$.

We then have

$$L_\Gamma(dt) = d(L_\Gamma t) = 0$$

and

$$L_\Gamma(\omega_\alpha^{\wedge n}) = n\omega_\alpha^{\wedge(n-1)}\wedge L_\Gamma\omega_\alpha = 0,$$

and so $\omega_\alpha^{\wedge n}\wedge dt$ is an invariant volume element for Γ. Observe that it follows in a similar way that $\omega_\beta^{\wedge n}\wedge dt$ is a volume element invariant under the transformed field $\Phi_*\Gamma$. Consequently we obtain that the $(2n+1)$-form $(\Phi^*\omega_\beta)^{\wedge n}\wedge dt$ satisfies

$$L_\Gamma[(\Phi^*\omega_\beta)^{\wedge n}\wedge dt] = 0.$$

Next we consider the family of two-forms

$$\omega_\lambda\in\Lambda^2(\mathbb{R}\times M),\quad \lambda\in\mathbb{R}$$

defined by $\omega_\lambda = \omega_\alpha - \lambda(\Phi^*\omega_\beta)$, and we denote by Ω_λ the volume element generated by ω_λ, $\Omega_\lambda = \omega_\lambda^{\wedge n}\wedge dt$.

Proposition: Ω_λ is an invariant volume element for Γ.

Proof: The Lie derivative of Ω_λ with respect to the field Γ takes the form

$$L_\Gamma\Omega_\lambda = [n\omega_\lambda^{\wedge(n-1)}\wedge L_\Gamma\omega_\lambda]\wedge dt + \omega_\lambda^{\wedge n}\wedge L_\Gamma(dt).$$

Then, since

$$L_\Gamma\omega_\lambda = L_\Gamma\omega_\alpha - \lambda L_\Gamma(\Phi^*\omega_\beta) = 0,$$

it follows that $L_\Gamma\Omega_\lambda = 0$. □

From the existence of such an invariant volume element, one is able to deduce some important properties for the transformation Φ.

Two different volume forms must be proportional, so there is a λ-dependent function $f_\lambda\in C^\infty(\mathbb{R}\times M)$ such that $\Omega_\lambda = f_\lambda(\omega_\alpha^{\wedge n}\wedge dt)$. This function f_λ is a polynomial of degree n in the λ parameter, since

$$\Omega_\lambda = \sum_{j=1}^{n}\binom{n}{j}(-1)^j\lambda^j[\omega_\alpha^{n-j}\wedge(\Phi^*\omega_\beta)^j]\wedge dt = \sum_{j=1}^{n} a_j\lambda^j\omega_\alpha^{\wedge n}\wedge dt.$$

Then, according to the above proposition, we see that

$$L_\Gamma[f_\lambda(\omega_\alpha^{\wedge n}\wedge dt] = \Gamma(f_\lambda)(\omega_\alpha^{\wedge n}\wedge dt) + f_\lambda L_\Gamma(\omega_\alpha^{\wedge n}\wedge dt) = 0,$$

and this shows that the function f_λ satisfies $L_\Gamma(f_\lambda)=0$. Using the form of f_λ just obtained, this last equation can also be written in the form

$$\sum_{j=1}^{n}\lambda^j L_\Gamma a_j=0,$$

and as a consequence of this, we wee that every one of these coefficients must be a constant of the motion $L_\Gamma a_j=0$, $\forall j=1,\dots,n$.

In order to present which ones are the coordinate expressions of these constants, we consider now the two simplest cases $n=1$ and $n=2$ and assume that $\alpha=dH$.

(i) If $n=1$, then

$$\Omega_\lambda=[\omega_H-\lambda(\Phi^*\omega_K)]\wedge dt$$

and we obtain $a_1=\{P,Q\}$.

(ii) If $n=2$, then

$$\begin{aligned}\Omega_\lambda&=[\omega_H-\lambda(\Phi^*\omega_K)]^2\wedge dt\\&=\omega_H^2\wedge dt-2\lambda\omega_H\wedge(\Phi^*\omega_K)\wedge dt\\&\quad+\lambda^2(\Phi^*\omega_K)^2\wedge dt,\end{aligned}$$

and we obtain

$$a_1=\{P_1,Q_1\}+\{P_2,Q_2\},$$
$$a_2=[q_1,q_2][p_1,p_2]+[q_1,p_1][q_2,p_2]+[q_1,p_2][q_2,p_1],$$

where [,] denotes the Lagrange brackets.

We have proved that to every time-dependent canonoid transformation, there will be n associated constants of the motion. The first of them, the coefficient a_1, generalizes the function f obtained for $n=1$ in Sec. V, and will always correspond to the sum of $\{P_i,Q_i\}$. The remaining constants of the motion will be different combinations of products of Lagrange brackets. These n constants are of non-Noetherian character and they can be nonindependent.

VII. GENERATING FUNCTIONS FOR CANONOID TRANSFORMATIONS

In the case of ω being an exact symplectic form, $\omega=-d\Theta$, and we have $\omega_H=-d\Theta_H$ with $\Theta_H=\widetilde{\Theta}-H\,dt$, $\widetilde{\Theta}=\pi_2{}^*\Theta$. In a similar way, $\omega_K=-d\Theta_K$, with $\Theta_K=\widetilde{\Theta}-K\,dt$. Now, using these one-forms, the property $L_\Gamma(\omega_H-\Phi^*\omega_K)=0$ becomes

$$d[L_\Gamma(\Theta_H-\Phi^*\Theta_K)]=0. \tag{7.1}$$

That is, $L_\Gamma(\Theta_H-\Phi^*\Theta_K)$ being closed is equivalent to Φ being canonoid with respect to Γ. This means the local existence of a function W such that

$$L_\Gamma(\Theta_H-\Phi^*\Theta_K)=dW. \tag{7.2}$$

Using local coordinates, we have

$$\frac{\partial^2H}{\partial q^j\partial p_k}p_k-\frac{\partial H}{\partial q^j}+P_k\frac{\partial}{\partial q^j}\{H,Q^k\}+\{H,P_k\}\frac{\partial Q^k}{\partial q^j}-\frac{\partial A_j}{\partial t}=\frac{\partial W}{\partial q_j}, \tag{7.3a}$$

$$\frac{\partial^2H}{\partial p_j\partial p_k}p_k+P_k\frac{\partial}{\partial p_j}\{H,Q^k\}+\{H,P_k\}\frac{\partial Q^k}{\partial p_j}-\frac{\partial B_j}{\partial t}=\frac{\partial W}{\partial p_j}, \tag{7.3b}$$

$$\frac{\partial^2H}{\partial t\partial p_k}p_k+P_k\frac{\partial}{\partial t}\{H,Q^k\}+\{H,P_k\}\frac{\partial Q^k}{\partial t}-\frac{\partial}{\partial t}(H-\Phi^*K)-\{H,\Phi^*K\}-\frac{\partial C}{\partial t}=\frac{\partial W}{\partial t}, \tag{7.3c}$$

where

$$A_j=P_k\frac{\partial Q^k}{\partial q^j},\quad B_j=P_k\frac{\partial Q^k}{\partial p_j},$$

and

$$C=P_k\frac{\partial Q^k}{\partial t}.$$

Given a canonoid diffeomorphism Φ, then the associated function $W(q,p,t)$ may be considered as its generating function. Same as for canonical transformations, this generating function can be determined up to an additive arbitrary function $f(t)$ of the time t alone. Reversing the procedure, given a function W, then every solution $Q=Q(q,p,t)$, $P=P(q,p,t)$ of the above equations represents a canonoid transformation for the Hamiltonian $H(q,p,t)$.

As a final comment concerning these equations, suppose that when studying the transformation determined by a certain function $W(q,p,t)$, we see that in that particular case there is a function F such that W can be expressed as the Lie derivative $W=L_\Gamma F$. Then we obtain that the solutions of (7.3) will represent a transformation that turns out to be not only canonoid for H, but also of canonical in general. Moreover, one can prove that if we write in Eqs. (7.3) $W=\{F,H\}+(\partial F/\partial t)$, then we obtain

$$\frac{\partial F}{\partial q^j}=p_j-A_j,\quad \frac{\partial F}{\partial p_j}=-B_j,$$
$$\frac{\partial F}{\partial t}=-(H-\Phi^*K)-C,$$

recovering in this way the equations that characterize the canonicity of a transformation.

VIII. EXAMPLES

The resolution of Eqs. (7.3) can be a formidable task in the general case. If we restrict ourselves to the so-called "fouling" transformations[11,12] (that is to say, time-preserving diffeomorphisms inducing the identity for the Darboux coordinates q^k) (see Ref. 13) for a time-independent Hamiltonian $H=H(q,p)$, then the new variables are $Q^k=q^k$, $P_k=P_k(q,p,t)$, and the above equations become

$$\frac{\partial^2H}{\partial q^j\partial p_k}(p_k-P_k)-\frac{\partial H}{\partial q^j}+\{H,P_j\}-\frac{\partial A_j}{\partial t}=\frac{\partial W}{\partial q^j}, \tag{8.1a}$$

$$\frac{\partial^2H}{\partial p_jp_k}(p_k-P_k)=\frac{\partial W}{\partial p_j}, \tag{8.1b}$$

$$\frac{\partial(\Phi^*K)}{\partial t} + \{\Phi^*K,H\} = \frac{\partial W}{\partial t}. \qquad (8.1c)$$

As an example, we will obtain the set of all those fouling transformations that are canonoid for the Hamiltonian $H = p^2/2$ of the free particle.

Equations (8.1a), (8.1b) are now

$$\frac{\partial W}{\partial q} = -p\frac{\partial P}{\partial q} - \frac{\partial P}{\partial t}, \qquad (8.2a)$$

$$\frac{\partial W}{\partial p} = p - P. \qquad (8.2b)$$

This system can be integrated and solved for P as a function $P = P(q,p,t)$ if and only if W satisfies the following partial differential equation,

$$p\frac{\partial^2 W}{\partial q \partial p} - \frac{\partial W}{\partial q} + \frac{\partial^2 W}{\partial p \partial t} = 0. \qquad (8.3)$$

The general solution of this equation is

$$W = \int\left[\int h(p,q-pt)dp\right]dp + p\frac{\partial f(q,t)}{\partial q} + \frac{\partial f(q,t)}{\partial t}, \qquad (8.4)$$

where $h(p,q-pt)$ and $f(q,t)$ are arbitrary differentiable functions of their arguments p and $q-pt$, and q and t, respectively.

With this expression for W, we find that the new momentum P is given by

$$P = p - \int h(p,q-pt)dp - \frac{\partial f(q,t)}{\partial q}$$

and that, therefore, the Poisson bracket $\{P,Q\}$ becomes

$$\{Q,P\} = 1 - h(p,q-pt). \qquad (8.5)$$

That is, $\{Q,P\}$ does not take the value $\{Q,P\} = 1$ obtained for the canonical case, but only disagrees with it by a function h depending on the three independent variables q, p, and t by means of $u_1 = p$ and $u_2 = q - pt$. Let us remark that these two functions are precisely the two independent constants of the motion associated to the Hamiltonian $H = p^2/2$. Moreover, in the case of $h(p,q-pt) = 0$, we obtain $Q = q$, $P = p - \partial f(q,t)/\partial q$, and the function W reduces to

$$W = p\frac{\partial f(q,t)}{\partial q} + \frac{\partial f(q,t)}{\partial t},$$

that can be written as $W = L_\Gamma f$ with

$$\Gamma = p\frac{\partial}{\partial q} + \frac{\partial}{\partial t}.$$

Consequently, according to the comments of the previous section, we see that this $f(q,t)$ which appears as an arbitrary function in the general solution (8.4), in the particular case of considering canonical transformations, turns out to be the generating function (in the canonical sense).

In short, for every function W of the form given by Eq. (8.4), we obtain a fouling transformation that is canonoid for $H = p^2/2$ and that has an associated Poisson bracket $\{Q,P\}$ that is a function depending on the two independent constants of the motion of the free particle. Conversely, from Eqs. (8.1a), (8.1b) we see that every canonoid transformation $Q = q$, $P = P(q,p,t)$ uniquely determines a function W of the form (8.4) up to an additive $f(t)$, and from Eq. (8.1c) we see that this ambiguity is reflected in the new Hamiltonian K that will also be determined up to the above $f(t)$.

As a further illustration, let us consider the family of solutions of the form

$$W = f(q)p + (1/t^2)h(q-pt) + g_1(p) + g_2(t), \qquad (8.6)$$

where $f(q)$, $h(q-pt)$, $g_1(p)$, and $g_2(t)$ are four arbitrary differentiable functions. We remark that when $h = g_2 = 0$, then W takes the form $W = f(q)p + g_1(p)$ and in this form, we recover the general form of the generating function for the time-independent canonoid transformations.[5] The new momentum P is given by

$$P = p - f(q) - g_1'(p) + (1/t)h'(q-pt),$$

and therefore the Poisson bracket $\{Q,P\}$ is

$$\{Q,P\} = 1 - g_1''(p) - h''(q-pt), \qquad (8.7)$$

depending of the two constants of the motion in an additive way. Moreover, Eq. (8.7) shows that if g_1 and h are lineal functions, then $\{Q,P\} = 1$ and the transformation is canonical. If g_1 and h are quadratic, then $\{Q,P\}$ = constant and the transformation corresponds to the so-called extended canonical transformations (canonical transformation plus a scale change).

We conclude this section with a final example. Let us consider now the transformation associated to the following function,

$$W = (p^2/2) + (1/6t^2)(q-pt)^3.$$

The equations giving the transformation are

$$Q = q, \quad P = (1/2t)(q-pt)^2.$$

The new Hamiltonian $K = K(Q,P,t)$ turns out to be

$$K = (1/t)[Q - \tfrac{2}{3}(2tP)^{1/2}]P,$$

and the new equations to be solved are

$$\frac{dQ}{dt} = \frac{1}{t}[Q - (2tP)^{1/2}], \quad \frac{dP}{dt} = -\frac{P}{t}.$$

ACKNOWLEDGMENTS

Partial financial support by CICYT (Comisión Interministerial Científica y Técnica) is acknowledged.

[1] M. Asorey, J. F. Cariñena, and L. A. Ibort, J. Math. Phys. **24**, 2745 (1983).
[2] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978).
[3] E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).
[4] A. Lichnerowicz, J. Diff. Geom. **12**, 253 (1977).
[5] J. F. Cariñena and M. F. Rañada, J. Math. Phys. **29**, 2181 (1988).
[6] C. Leubner and M. A. Marte, Phys. Lett. A **101**, 179 (1984).
[7] L. J. Negri, L. C. Oliveira, and J. M. Teixeira, J. Math. Phys. **28**, 2369 (1987).
[8] A. Weinstein, J. Diff. Geom. **18**, 523 (1983).
[9] P. J. Olver, *Application of Lie groups to Differential Equations* (Springer, New York, 1986).
[10] G. Marmo, E. J. Saletan, R. Schmid, and A. Simoni, Nuovo Cimento B **100**, 297 (1987).
[11] D. G. Currie and E. J. Saletan, J. Math. Phys. **7**, 967 (1966).
[12] Y. Gelman and E. J. Saletan, Nuovo Cimento B **18**, 53 (1973).
[13] As far as we know, the first paper studying these transformations is Ref. 11, but it seems that it was F. A. E. Pirani who introduced this terminology. We thank G. Marmo for information on the origin of this name.

Twistors for uniform acceleration

Gerald Harnett
Department of Mathematics, Western Washington University, Bellingham, Washington 98225

(Received 24 January 1989; accepted for publication 17 May 1989)

The skew tensor field and twistor descriptions of the momentum and angular momentum of a classical relativistic particle are adapted to describe a uniformly accelerated particle. An extension, natural within the twistor framework, of the usual duality between Poincaré momenta and symmetries yields osculating planes together with acceleration scalars as momenta of uniformly accelerated particles. The adapted twistor description leads to the construction of a local twistor attached to an arbitrary world line in a general space-time and a conformally invariant prescription for uniform acceleration.

I. INTRODUCTION

In special relativistic kinematics the angular momentum of a single particle about an origin is the bivector obtained as the product of position vector with four-momentum. Summing over several particles and allowing the space-time origin to vary yields a skew tensor field which completely encodes the kinematical structure of a system of particles, including the total four-momentum, the intrinsic angular momentum, and (in the case of nonzero mass) the center-of-mass line. Fields arising this way may be characterized by their particular affine dependence on position and also as certain types of twistors.[1,2] In this paper it is observed that the description of a uniformly accelerated particle may be similarly encoded into a skew tensor field, now with a quadratic position dependence and again with a natural twistor counterpart.

In Sec. II the kinematical structure of systems of particles in Minkowski space is reviewed and a skew tensor field that completely characterizes a uniformly accelerated particle is presented. No mention of twistors is made in Sec. II; readers unfamiliar with them may read this section profitably. In Sec. III the connection between Minkowski space symmetries, kinematics, and twistors is recalled and an extension of the usual duality between momenta and symmetries is obtained. In Sec. IV we find twistor counterparts of the skew tensor fields representing uniformly accelerated particles and characterize the resulting twistors. In the extended duality, the new momentum (or set of conserved quantities) arising in the case of a uniformly accelerated particle is seen to be essentially the osculating plane of the particle together with its acceleration scalar. The concluding discussion includes a construction, based on the twistors for uniform acceleration, of a local twistor on an arbitrary world line in a general space-time and a conformally invariant prescription for the notion of uniform acceleration. Also, the relations these twistors have with Liénard–Wiechert fields and the linearized curvature of the C-metric are described.

Throughout, Minkowski space M is viewed as an abstract affine space of points acted upon by a four-dimensional real vector space of displacements. The action of a displacement v^a on a point x is indicated by $x \mapsto x+v^a$. Otherwise, the notation for Minkowski space objects, including spinors, agrees with that of Penrose and Rindler.[3] In particular, $g_{ab} \equiv \epsilon_{AB}\epsilon_{A'B'}$ denotes the Lorentzian metric of signature $(+---)$, the alternating tensor e_{abcd} satisfies $e_{abcd}e^{abcd} = -24$, and the dual $*M_{ab}$ of a skew tensor M^{ab} is $*M^{ab} = \frac{1}{2}e^{ab}{}_{cd}M^{cd}$. For relevant material on twistors the reader is referred to Refs. 1 and 2, whose notation is used herein.

II. SKEW TENSOR FIELDS FOR UNIFORM ACCELERATION

In this section we first review the tensor field representation of "Poincaré kinematics"—the momentum and angular momentum of systems of particles in Minkowski space. Then uniformly accelerated particles are represented in a way that generalizes this (in the case in which the mass is nonzero and there is no intrinsic angular momentum).

The kinematical structure of a system of particles in M is given by a skew tensor field M^{ab} whose position dependence is

$$M^{ab}(x) = M^{ab}(o) - 2x^{[a}p^{b]}, \tag{1}$$

where p^a is the total four-momentum and x^a is the position vector of the point x with respect to a chosen origin o. If the four-momentum is timelike, then

$$M^{ab} = 2(p_c p^c)^{-1}(R^{[a}p^{b]} - \overset{*}{S}{}^{[a}p^{b]}), \tag{2}$$

where $R^a = p_b M^{ab}$ and $S^a = p_b {*M}^{ab}$. The vector $(p_c p^c)^{-1}R^a(x)$ is the displacement orthogonal to p^a from x to the world line of the relativistic center of mass and S^a is the Pauli–Lubanski spin vector. If k^a is a Killing vector field, then the quantity

$$\tfrac{1}{2}M^{ab}\nabla_a k_b + p_a k^a \tag{3}$$

is constant. As k^a ranges over the standard generators in an observer's coordinate system, the values obtained are the energy, the components of three-momentum and three-angular momentum, and the coordinates of the center of mass of the system of particles. In a Lagrangian approach, M^{ab} is in fact *defined* as an object dual to the Killing vector fields, with the above pairing.[4]

Now consider a uniformly accelerated particle in M, i.e., a pair (m,γ), where m is a positive constant (the mass) and $\gamma:\mathbf{R}\to M$ is a path with unit future timelike tangent $\dot{\gamma}^a$ and nonzero acceleration vector $\ddot{\gamma}^a$ such that the "acceleration bivector"

 0022-2488/89/102267-06$02.50

$$C^{ab} := 2m\dot{\gamma}^{[a}\ddot{\gamma}^{b]} \tag{4}$$

is constant. Note that the positive scalar acceleration a given by $a^2 = -\ddot{\gamma}_c\ddot{\gamma}^c$ may be written in terms of m and the acceleration bivector. Under these circumstances

$$\gamma(t_0+t) = o \dot{+} a^{-1}[(\sinh at)\dot{\gamma}^c(t_0) + (\cosh at)a^{-1}\ddot{\gamma}^c(t_0)],$$

where $o = \gamma_0 \dot{+} [-a^{-2}\ddot{\gamma}^a(t_0)]$ with $\gamma_0 := \gamma(t_0)$. The world line of the particle forms one branch of a hyperbola having the center o (to be employed as the origin) and lying in the timelike hyperboloid $x_cx^c = -a^{-2}$. The other branch will be called the "opposite world line." The path $s \mapsto \gamma(ms)$ is an integral curve of the Killing vector field

$$p^a(x) = m\dot{\gamma}^a(t_0) + C_b{}^a(x-\gamma_0)^b = C_b{}^a x^b. \tag{5}$$

The above formulas hold for an arbitrarily chosen point γ_0 on the world line.

The promised field that completely determines a uniformly accelerated particle (m,γ) and that generalizes (1) in the case of a single material particle (without spin) is given, in the above notation, by

$$\begin{aligned} M^{ab}(x) &= -2(x-\gamma_0)^{[a}p^{b]}(x) + \tfrac{1}{2}C^{ab}(x-\gamma_0)_c(x-\gamma_0)^c \\ &= -2x^{[a}p^{b]}(x) + \tfrac{1}{2}(x_cx^c - a^{-2})C^{ab}. \end{aligned} \tag{6}$$

The assertions among the following remarks may be proved directly with the above formulas; the results of the following sections may be helpful. (i) The first expression for M^{ab} is independent of the choice of γ_0 on the world line. In fact, we may choose γ_0 on the opposite world line. Thus M^{ab} vanishes on both branches of the hyperbola; these are the only loci where it vanishes. (ii) We have again, as in Eq. (1), $p^a = \frac{1}{3}\nabla_b M^{ab}$. (iii) If $p^a(x)$ is future timelike, then the field $R^a = p_b M^{ab}$ at the point x is $p_c(x)p^c(x)$ times the projection orthogonal to $p^a(x)$ of the displacement from x to the unique world line point on the past light cone of x. Thus $(p_cp^c)^{-1} R^a$ at x may be interpreted as the apparent position of the particle with respect to the observer at x determined by $p^a(x)$. The field p^a becomes null on two null hyperplanes which are approached asymptotically by the hyperbola and it vanishes on their intersection—a spacelike two-plane containing the point o and orthogonal to the osculating plane (determined by o and C^{ab}) of the hyperbola; p^a is past timelike on the opposite world line. (iv) The field $p_b{}^*M^{ab}$, corresponding to intrinsic spin in the unaccelerated case, vanishes identically. Where p_cp^c is nonzero, Eq. (2) with $S^a = 0$ holds. (v) We will see in Sec. III that there is a duality involving twistors which naturally extends the duality pairing (3). For a uniformly accelerated particle, the new conserved quantities that arise determine the locus of the osculating plane. (vi) At this point, it is clear that the particle (m,γ) may be completely recovered from M^{ab}. (The correct world line is the one on which p^a is future timelike.) (vii) Finally, we remark that the Liénard–Wiechert field F_{ab} of charge m on an *arbitrary* world line γ, when evaluated at a point $x \equiv \gamma(t) \dot{+} rl^a$ (with l^a null and $l_a\dot{\gamma}^a = 1$) on the future light cone of $\gamma(t)$, is $F_{ab}(x) = -r^{-3}M_{ab}(x)$.

III. TWO-TWISTORS

The twistor decription of Poincaré kinematics, i.e., of fields of the form (1), is well known.[1,2] In Sec. IV we will see that there is a related twistor description of uniformly accelerated particles, i.e., of fields of the form (6). In this section some relevant facts about general symmetric two-index twistors, hereafter called two-twistors, are presented. In the spirit of Noether's theorem concerning conserved quantities, or momenta, associated with symmetries, we start with the two-twistors most closely associated with symmetries and then pass to the twistors dual to these, some of which describe Poincaré kinematics and some of which describe uniformly accelerated particles.

A two-twistor $\mathscr{S}^{\alpha\beta}$ corresponds to a symmetric spinor field $\mathscr{S}^{AB}$ on Minkowski space satisfying the two-index twistor equation $\nabla_{A'}{}^{(A}\mathscr{S}^{BC)} = 0$. There is a corresponding skew real tensor field

$$S^{ab} := \mathscr{S}^{AB}\epsilon^{A'B'} + \bar{\mathscr{S}}^{A'B'}\epsilon^{AB} \tag{7}$$

satisfying a corresponding equation, the "tensor two-twistor equation":

$$\nabla^{(a}S^{b)c} - \nabla^{(a}S^{c)b} + g^{a[b}\nabla_d S^{c]d} = 0. \tag{8}$$

The expression on the lhs of (8) is simply $\frac{3}{2}$ times the projection of $\nabla^a S^{bc}$ onto a certain Lorentz-irreducible subspace of the space of tensors t^{abc} having the symmetry $t^{abc} = t^{a[bc]}$. Looking at the remaining irreducible parts leads to the equivalent equation

$$\nabla_c S^{ab} = -2g_c{}^{[a}j^{b]} + 2g_c{}^{[a}\overset{*}{k}{}^{b]},$$

where

$$j^a = \tfrac{1}{3}\nabla_b S^{ab}, \quad k^a = \tfrac{1}{3}\nabla_b{}^*S^{ab}. \tag{9}$$

If S^{ab} is a solution, then the fields j^a and k^a are Killing vector fields related by

$$l_{ab} \equiv \nabla_a k_b = -\tfrac{1}{2}\epsilon_{ab}{}^{cd}\nabla_c j_d.$$

The space of solutions to Eq. (8) is a 20-dimensional real vector space which gets mapped onto the ten-dimensional space of Killing vector fields via $S^{ab} \mapsto k^a$. In fact, the space of solutions is isomorphic to the space of (contravariant) two-twistors via the correspondence in (7)—and isomorphic to the space of *dual* two-twistors via (10).[5]

If $\mathscr{A}_{\alpha\beta}$ is a dual two-twistor there is again a corresponding spinor field $\mathscr{A}^{A'B'}$ and a skew real tensor field A^{ab}. To eventually match up with usual momentum and angular momentum quantities, consider the field

$$M^{ab} := -\tfrac{1}{2}{}^*A^{ab} = (i/2)(\mathscr{A}^{A'B'}\epsilon^{AB} - \bar{\mathscr{A}}^{AB}\epsilon^{A'B'}); \tag{10}$$

define vector fields p^a and q^a in terms of M^{ab}, exactly as j^a and k^a, respectively, are defined in terms of S^{ab} in (9) and set $r_{ab} := \nabla_a p_b$. The real duality pairing between the two-twistors $\mathscr{A}_{\alpha\beta}$ and $\mathscr{S}^{\alpha\beta}$ then becomes

$$\mathrm{Re}(\mathscr{A}_{\alpha\beta}\mathscr{S}^{\alpha\beta}) = \tfrac{1}{2}M^{ab}l_{ab} + p_ak^a - q_aj^a + \tfrac{1}{2}r_{ab}{}^*S^{ab}. \tag{11}$$

This will be demonstrated presently. The salient fact about formula (11) is that the combination of fields on the rhs is constant. With an appropriate interpretation of the fields it is a "conserved quantity."

Equation (11) is obtained as follows.[6] Recall that the spinor parts associated with a symmetric twistor $\mathscr{A}_{\alpha\beta}$ consist of the spinor field $\mathscr{A}^{A'B'}$—the "primary part"—and certain other spinor fields—the "projection parts"—derived from the twistor in an essentially algebraic way, but which may be obtained from the primary part by differentiation. This association is indicated by

$$\mathscr{A}_{\alpha\beta} \sim \begin{pmatrix} \mathscr{A}_{AB} & \mathscr{A}_{A}{}^{B'} \\ \mathscr{A}^{A'}{}_{B} & \mathscr{A}^{A'B'} \end{pmatrix},$$

where it turns out,

$$\mathscr{A}_{A}{}^{B'} = -(i/3)\nabla_{AA'}\mathscr{A}^{A'B'},$$

$$\mathscr{A}_{AB} = -\tfrac{1}{6}\nabla_{AA'}\nabla_{BB'}\mathscr{A}^{A'B'},$$

and $\mathscr{A}^{A'}{}_{B} = \mathscr{A}_{B}{}^{A'}$. (When the matrix elements are evaluated at a particular point $y\in M$ the symbol $\overset{y}{\sim}$ is used in place of $\sim$.) The projection spinor parts $\mathscr{S}_{B'}{}^{A}$ and $\mathscr{S}_{A'B'}$ of a two-twistor $\mathscr{S}^{\alpha\beta}$ are obtained by conjugating the above identities and replacing $\bar{\mathscr{A}}$ with $\mathscr{S}$. In terms of spinor parts we have

$$\mathscr{A}_{\alpha\beta}\mathscr{S}^{\alpha\beta} = \mathscr{A}^{A'B'}\mathscr{S}_{A'B'} + 2\mathscr{A}_{A}{}^{B'}\mathscr{S}_{B'}{}^{A} + \mathscr{A}_{AB}\mathscr{S}^{AB}.$$

Performing spinor-to-tensor translations of the spinor parts (e.g., $\mathscr{A}^{A'B'}\epsilon^{AB} = {}^*M^{ab} + iM^{ab}$, $\mathscr{A}^{AA'} = p^a - iq^a$, and $\mathscr{A}_{AB}\epsilon_{A'B'} = r_{ab} + i^*r_{ab}$) and taking the real part of the above expansion leads to formula (11). The imaginary part yields an expression which, for a fixed $\mathscr{A}_{\alpha\beta}$, can be made equal to the rhs of Eq. (11) by an appropriate choice of $\mathscr{S}^{\alpha\beta}$.

IV. TWO-TWISTORS FOR UNIFORM ACCELERATION

In this section we first comment on the twistor description of Poincaré kinematics and then show how uniformly accelerated particles, i.e., fields of the form (6), can also be described by certain two-twistors. In the remainder of the paper every simple skew twistor ($X_{\alpha\beta}$ or $X^{\alpha\beta}$, say) representing a point in M is assumed to be normalized with respect to the appropriate inifinity twistor $I^{\alpha\beta}$ or $I_{\alpha\beta} = \bar{I}_{\alpha\beta}$ (i.e., $X_{\alpha\beta}I^{\alpha\beta} = X^{\alpha\beta}I_{\alpha\beta} = 2$).

A skew field M^{ab} has the position dependence (1) (with p^a constant) if and only if $\nabla^{(a^*}M^{b)c} = 0$, which implies that ${}^*M^{ab}$ satisfies the tensor two-twistor equation (8). The structure is thus represented by a dual two-twistor $\mathscr{A}_{\alpha\beta}$—subject to a restriction involving the infinity twistor; namely,

$$P^{\alpha}_{\beta} \equiv I^{\alpha\lambda}\mathscr{A}_{\beta\lambda} = \bar{\mathscr{A}}^{\alpha\lambda}I_{\beta\lambda}.$$

(This guarantees that the primary part $P^{AB'}$ of the twistor P^{α}_{β} is real and constant. The field p^a that we have associated with a dual two-twistor [after Eq. (10)] is always the real part of this primary part. Further conditions are required for the primary part to be future pointing and timelike.) A few remarks about this well-known representation are appropriate here. First, the "kinematic twistor" $\mathscr{A}_{\alpha\beta}$ is given explicitly in terms of the four-momentum P^a, the Pauli–Lubanski spin vector S^a, and an arbitrary point o on the world line of the relativistic center of mass by

$$\mathscr{A}_{\alpha\beta} = 2O_{\kappa(\alpha}P^{\kappa}_{\beta)} + 2O_{\kappa\lambda}\mathbf{S}^{\kappa}_{\alpha}P^{\lambda}_{\beta} \overset{o}{\sim} \begin{pmatrix} 0 & P^{B'}_{A} \\ P^{A'}_{B} & 2\mathbf{S}^{A'}_{A}P^{AB'} \end{pmatrix}, \quad (12)$$

where $O_{\alpha\beta}$ represents the point o and $P_cP^c\mathbf{S}^{\alpha}_{\beta}$ is the unique trace-free twistor whose primary part is the Pauli–Lubanski spin vector. Next, a kinematic twistor with nonzero intrinsic spin may be obtained from a kinematic twistor with no spin—a "monopole two-twistor"—by "translating into the complex" [7] by $-i\mathbf{S}^a$. This is achieved twistorially by applying the transformation induced by the special linear transformation $\exp(\mathbf{S}^{\alpha}_{\beta}) = \delta^{\alpha}_{\beta} + \mathbf{S}^{\alpha}_{\beta}$ to the monopole twistor given by the first summand after the equality in (12). Finally, the spin vector may be extracted from the two-twistor by considering the trace-free part (or the primary part) of

$$\bar{\mathscr{A}}^{\alpha\lambda}\mathscr{A}_{\beta\lambda} = P_cP^c(2\mathbf{S}^{\alpha}_{\beta} - \tfrac{1}{2}\delta^{\alpha}_{\beta}). \quad (13)$$

(The inner product P_cP^c may be expressed twistorially if desired.)

Likewise, the dual of the skew field (6) that determines a uniformly accelerated particle satisfies the tensor two-twistor equation (8). Hence this field also arises from some two-twistor via (10). In the following we will exhibit the two-twistor explicitly in the manner of (12), obtain it from a monopole twistor in the manner that (12) is obtained—by applying an appropriate transformation, discuss the duality pairing (11) for this case, and characterize the two–twistors that represent uniformly accelerated particles.

The two-twistor will be exhibited in terms of invariant parameters of the uniformly accelerated particle, namely, the center o of the hyperbola of the world line—represented by $O_{\alpha\beta}$, the constant bivector $C^{ab} = 2m\dot{\gamma}^{[a}\ddot{\gamma}^{b]}$, and the scalar acceleration a. First note that C^{ab} may itself be represented by a two-twistor $\mathscr{C}^{\alpha\beta}$ whose only nonvanishing spinor part is the constant spinor $\mathscr{C}_{AB} = m\dot{\gamma}_{C'A}\ddot{\gamma}^{C'}_{B}$. In fact, if $\dot{\Gamma}^{\alpha}_{\beta}$ and $\ddot{\Gamma}^{\alpha}_{\beta}$ are the trace-free twistors whose primary parts are the constant vectors $\dot{\gamma}^{AA'}(t_0)$ and $\ddot{\gamma}^{AA'}(t_0)$, respectively, then $\mathscr{C}^{\alpha\beta} = m\dot{\Gamma}^{\alpha}_{\kappa}\ddot{\Gamma}^{\beta}_{\lambda}X^{\kappa\lambda}$, where $X^{\kappa\lambda}$ represents an arbitrary point in M. The two-twistor yielding the field (6) is then

$$\mathscr{A}_{\alpha\beta} = -2i\mathscr{C}^{\kappa\lambda}O_{\kappa\alpha}O_{\lambda\beta} - ia^{-2}\bar{\mathscr{C}}_{\alpha\beta}. \quad (14)$$

Again, the real part of the primary part of the twistor $P^{\alpha}_{\beta} = I^{\alpha\lambda}\mathscr{A}_{\beta\lambda}$ (which in the present case is equal to $-2i\mathscr{C}^{\lambda\alpha}O_{\lambda\beta}$) is the field p^a given in Eq. (5). [The imaginary part is the field $-q^a$ defined after (10).] The two-twistor's relation with the kinematic two-twistor (12) is better seen when it is expressed in terms of the twistor P^{α}_{β} and an arbitrary point $y = \gamma(t)$ on the world line—represented by $Y^{\alpha\beta}$. It turns out that

$$\mathscr{A}_{\alpha\beta} = 2Y_{\kappa(\alpha}P^{\kappa}_{\beta)} - 2i\mathscr{C}^{\kappa\lambda}Y_{\kappa\alpha}Y_{\lambda\beta}$$

$$\overset{y}{\sim} m\begin{pmatrix} -2i\dot{\gamma}^{A}_{C'}\ddot{\gamma}^{BC'} & \dot{\gamma}^{B'}_{A} \\ \dot{\gamma}^{A'}_{B} & 0 \end{pmatrix}. \quad (15)$$

This two-twistor was first obtained by "accelerating" the monopole two-twistor $2mO_{\kappa(\alpha}T^{\kappa}_{\beta)}$, where T^{α}_{β} is such that its primary part is a constant unit future timelike vector [to be identified with $\dot{\gamma}(t_0)$], that is, by applying to it the transformation induced by a special unitary transformation G of twistor space which yields a conformal transformation of M mapping the world line of the monopole to the world line of the uniformly accelerated particle. An appropriate such G is given by

$$G^{\alpha}_{\beta} = \exp(2iO^{\alpha\kappa}\ddot{\Gamma}^{\lambda}_{\kappa}O_{\lambda\mu})\exp\left(\frac{1}{2}ia^{-2}\ddot{\Gamma}^{\mu}_{\beta}\right)$$

$$\overset{o}{\sim}\begin{pmatrix} \delta^{A}_{B} & \frac{1}{2}ia^{-2}\ddot{\gamma}^{AB'}(t_0) \\ -2i\ddot{\gamma}_{A'B}(t_0) & \frac{1}{2}\delta^{B'}_{A'} \end{pmatrix},$$

where $\ddot{\Gamma}$ was introduced in the previous paragraph. Acting on M, this yields the composition of the translation

$$x \mapsto x + \tfrac{1}{2}a^{-2}\ddot{\gamma}^{c}(t_0)$$

with the special conformal transformation

$$x \mapsto o + [1 + 2\ddot{\gamma}_b(t_0)x^b - a^2x_bx^b]^{-1}[x^c + x_dx^d\ddot{\gamma}^c(t_0)].$$

The point o maps to $\gamma(t_0)$ and is the center of the resulting hyperbola. The transformation actually maps the monopole world line onto both branches of the hyperbola because the original world line, and hence the new one, is topologically a circle in the compactified Minkowski space $\mathscr{M}$, on which G acts. (Each of these circles is in fact the intersection with $\mathscr{M}$ of the *two-sphere* in complexified compactified Minkowski space on which the primary part of the respective two-twistor vanishes.[8]) The transformation is illustrated in Fig. 1 using a standard Penrose diagram of $\mathscr{M}$. With this perspective it is easy to see that the uniform acceleration field (6) vanishes only on the hyperbola, as claimed in the remarks after (6). The calculation of the action of G on the monopole two-twistor is somewhat long; it results in Eq. (14).

If $\mathscr{S}^{\alpha\beta}$ is a symmetric twistor, then

$$\mathrm{Re}(\mathscr{A}_{\alpha\beta}\mathscr{S}^{\alpha\beta}) = m\dot{\gamma}(t_0)_a k^a(\gamma_0) + \tfrac{1}{2}C_{ab}{}^*S^{ab}.$$

If $\mathscr{S}^{\alpha\kappa}I_{\beta\kappa} = 0$, then S^{ab} is constant and we obtain only the second term. Such $\mathscr{S}^{\alpha\beta}$ may thus be construed as the "symmetries" dual to the "momenta" which are essentially the osculating planes (and acceleration scalar) of uniformly accelerated particles.

Let us characterize the two-twistors which represent uniformly accelerated particles.[9] In what follows $\Omega_{\alpha\beta\gamma\delta}$ is the four-form on twistor space whose twistor conjugate

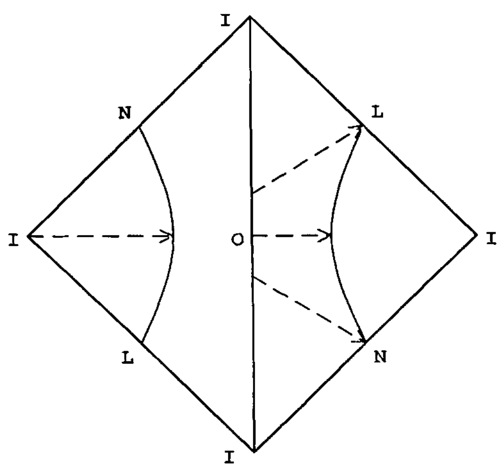


FIG. 1. "Acceleration" of a monopole world line in a Penrose diagram. Points with the same labels are identified.

$\Omega^{\alpha\beta\gamma\delta} = \overline{\Omega}^{\alpha\beta\gamma\delta}$ satisfies $\Omega^{\alpha\beta\gamma\delta}\Omega_{\alpha\beta\gamma\delta} = 24$. The determinant of a two-twistor $\mathscr{A}_{\alpha\beta}$ is given by

$$\Omega^{\kappa\lambda\mu\nu}\mathscr{A}_{\alpha\kappa}\mathscr{A}_{\beta\lambda}\mathscr{A}_{\gamma\mu}\mathscr{A}_{\delta\nu} = (\det \mathscr{A}_{\pi\rho})\Omega_{\alpha\beta\gamma\delta}. \qquad (16)$$

Theorem: Let $\mathscr{A}_{\alpha\beta}$ be a symmetric twistor satisfying the following conditions: (i) $\Delta^2 := \det \mathscr{A}_{\alpha\beta}$ is real and positive, (ii) $\mathscr{A}_{\alpha\lambda}\overline{\mathscr{A}}^{\beta\lambda} = -\Delta\delta^{\beta}_{\alpha}$, and (iii) $\mathscr{A}_{\alpha\beta}\mathscr{A}_{\kappa\lambda}I^{\alpha\kappa}I^{\beta\lambda} > 0$. Then $\mathscr{A}_{\alpha\beta}$ represents a uniformly accelerated particle, i.e., it is of the form (14).

Proof: For convenience we rewrite Eq. (14) as

$$\mathscr{A}_{\alpha\beta} = \mathscr{L}^{\kappa\lambda}O_{\kappa\alpha}O_{\lambda\beta} - \tfrac{1}{2}a^{-2}\overline{\mathscr{L}}_{\alpha\beta}. \qquad (17)$$

We begin by simply identifying the quantities involved in uniform acceleration in terms of $\mathscr{A}_{\alpha\beta}$. The twistor $\mathscr{L}^{\alpha\beta}$ [which is $-2iC^{\alpha\beta}$ in (14) and which determines the acceleration bivector (4)] is given by

$$\mathscr{L}^{\alpha\beta} = \mathscr{A}_{\kappa\lambda}I^{\alpha\kappa}I^{\beta\lambda},$$

the scalar $m^2/2$ is identified with Δ, the scalar $2(ma)^2$ is identified with the positive constant in condition (iii) [whose reality, we will see, is a consequence of (ii)], and the center of the resulting hyperbola is the point o represented by the simple skew twistor

$$O_{\alpha\beta} = (ma)^{-2}\mathscr{A}_{\alpha\kappa}\mathscr{A}_{\beta\lambda}I^{\kappa\lambda}. \qquad (18)$$

Next we show that these twistors and scalars have the required properties. First note that (ii) implies

$$\mathscr{A}_{\alpha\kappa}\mathscr{A}_{\beta\lambda}\Omega^{\kappa\lambda\gamma\delta} = \overline{\mathscr{A}}^{\gamma\kappa}\overline{\mathscr{A}}^{\delta\lambda}\Omega_{\alpha\beta\kappa\lambda}, \qquad (19)$$

as can be seen by contracting each side of identity (16) with $\overline{\mathscr{A}}^{\gamma\sigma}\overline{\mathscr{A}}^{\delta\tau}$. This in turn implies that the scalar in (iii) is real and the twistor $\mathscr{A}_{\alpha\kappa}\mathscr{A}_{\beta\lambda}I^{\kappa\lambda}$ represents a point in real compactified Minkowski space. The *nonvanishing* of the scalar in (iii) implies that this real point is in *affine* Minkowski space. The fact that the scalar is real and positive is necessary to obtain a bivector appropriate for a timelike osculating plane, as follows. In terms of the spinor parts of $\mathscr{A}_{\alpha\beta}$, the scalar in (iii) is $\mathscr{A}_{AB}\mathscr{A}^{AB}$. In the particular case considered in (14), this scalar was

$$-4\mathscr{C}_{AB}\mathscr{C}^{AB} = -C_{ab}C^{ab} = 2(ma)^2,$$

where $\mathscr{C}_{AB} = m\dot{\gamma}_{C'A}\ddot{\gamma}^{C'}_{B}$. Now the condition that the scalar $\mathscr{C}_{AB}\mathscr{C}^{AB}$ associated with a symmetric spinor $\mathscr{C}_{AB}$ is real and negative is necessary and sufficient for the associated skew tensor

$$C_{ab} = \mathscr{C}_{AB}\epsilon_{A'B'} + \overline{\mathscr{C}}_{A'B'}\epsilon_{AB}$$

to be "purely electric," i.e., of the form $C^{AB} = U^{[a}V^{b]}$ for some timelike U^a and spacelike V^a (orthogonal to U^a).[10] Thus condition (iii) is equivalent to the condition that the projection part $\mathscr{A}_{AB} = -2i\mathscr{C}_{AB}$ yield a bivector of the required type.

Finally, we show that our two-twistor may be written as claimed in Eq. (17). We already know that the normalized contravariant twistor representing the point o is simply the twistor conjugate of (18). Now we observe that the projection onto the two-dimensional subspace of twistor space determined by $I_{\alpha\beta}$, parallel to the two-dimensional subspace determined by $O^{\alpha\beta}$, is given by $J^{\alpha}_{\beta} = I^{\alpha\kappa}O_{\beta\kappa}$ and the complementary projection is given by $J'^{\alpha}_{\beta} = O^{\alpha\kappa}I_{\beta\kappa}$. Thus we have

$$\mathscr{A}_{\alpha\beta} = \mathscr{A}_{\alpha\lambda} J^{\lambda}_{\beta} + \mathscr{A}_{\alpha\lambda} J'^{\lambda}_{\beta}.$$

Now by using our expressions for the projections (and in turn for $O_{\alpha\beta}$ and its conjugate), we find that the two summands in this last identity are the two summands in (17), in the same order. (The first summand is unchanged under contraction with J^{α}_{κ} since its contraction with J'^{α}_{κ} vanishes.) Q.E.D.

Here is a further remark on condition (iii) in the theorem. Assume (i) and (ii) hold. If the scalar in (iii) vanishes, the twistor is not necessarily a monopole two-twistor (discussed at the beginning of this section), as one might expect. In fact, to obtain a monopole twistor, it is necessary and sufficient for $\mathscr{A}_{\alpha\kappa}\mathscr{A}_{\beta\lambda}I^{\kappa\lambda}$ to be proportional to $I_{\alpha\beta}$ [cf. Eq. (18)]. The situation is made clear by observing that for any nonsingular two-twistor $\mathscr{A}_{\alpha\beta}$, the combination on the lhs of Eq. (19) maps simple skew twistors into simple skew twistors and hence induces a transformation of complexified compactified Minkowski space. Equation (19) implies that real points are mapped to real points and hence, in particular, that the point at infinity represented by the infinity twistor is mapped to a real point. If this real point is in real affine Minkowski space, we have a two-twistor for uniform acceleration; if this point is the point at infinity itself, we have a monopole two-twistor. A third possibility—for which the scalar in condition (iii) is again zero—is that this real point is on the null cone at infinity; we do not know a simple interpretation for such two-twistors.

V. DISCUSSION

There are several related contexts in which twistors for uniform acceleration appear. One, already mentioned, is that of Liénard–Wiechert fields of charged particles. In the special case of a uniformly accelerated charged particle, the Liénard–Wiechert field may be obtained directly from (the inverse of) the appropriate two-twistor by means of a twistor contour integral.[1,2]

Another context is that of linear gravity. Weak-field curvatures (i.e., tensors on Minkowski space having Riemann curvature tensor symmetries and satisfying the linearized Bianchi identity) can also be obtained from two-twistors via certain contour integrals. From a kinematic two-twistor (12) one obtains a weak–field curvature corresponding to one of the Kerr family of solutions; from a two-twistor for uniform acceleration one obtains a weak-field curvature corresponding to the C-metric. This weak-field curvature has been studied by Robinson and Robinson.[11] Our construction of a twistor for uniform acceleration by "accelerating" a monopole two-twistor reflects the fact observed in Ref. 11 that this weak-field curvature is a conformal transformation of the linearized curvature of the Schwarzschild solution.

Dual two-twistors also appear as the "charges" resulting from the 20 linear gravity intergals of Penrose and Rindler[2] augmenting the usual ten two-surface integrals arising from total four-momentum and angular momentum.[12] We obtain an intepretation for six of the added and somewhat curious "ten vanishing integrals" in Ref. 2 by observing that if the weak-field curvature of the C-metric is used in the 20 integrals (with the two-surface linking one of the singular lines), then the resulting two twistor is a twistor for uniform acceleration. Thus six of the ten integrals that vanish under the conditions considered by Penrose and Rindler are the six independent conserved quantities arising from the parameters for uniform acceleration. (These six integrals were also constructed by Robinson and Robinson.[11])

Twistors for uniform acceleration may have significance in general relativity as well. A subtle modification of the procedure for obtaining weak-field gravitational charges leads to Penrose's quasilocal mass construction in general relativity.[2,13] Our identification and characterization of the two-twistors for uniform acceleration may be helpful in clarifying the interpretation of the two–surface twistors obtained in this context. This paper suggests that one need not insist that a two-surface two-twistor reduce to ten real independent quantities in order to obtain a satisfactory physical interpretation. In particular, this paper is a guide to what to look for in the case of the C-metric.

In general relativity we may also construct a *local* twistor on an arbitrary timelike world line γ (with $\dot\gamma_a\dot\gamma^a = 1$) simply by taking the matrix form of the two-twistor in (15) (with $m = 1$) at each point $\gamma(t)$ of the world line. This local twistor encodes in a certain way the geometry of the world line. For a straight or uniformly accelerated world line in flat space this local twistor is, of course, constant. In a general space-time the matrix of its local twistor derivative along the world line vanishes except in the upper left-hand block. This block is essentially the anti-self-dual part of the bivector

$$\dot\gamma_{[a}\dddot\gamma_{b]} - \dot\gamma_{[a}P_{b]c}\dot\gamma^{c}, \tag{20}$$

where $\dddot\gamma^a = \nabla_{\dot\gamma}\nabla_{\dot\gamma}\dot\gamma^a$ and $P_{ab} \equiv \frac{1}{12}Rg_{ab} - \frac{1}{2}R_{ab}$ is a common rearrangement of the Ricci tensor. A local twistor is said to undergo local twistor transport along a curve if its local twistor derivative along the curve vanishes. If γ is a geodesic, this occurs for the associated local two-twistor if and only if $\dot\gamma$ is an eigenvector of the Ricci tensor. The same is true if γ is uniformly accelerated in the sense that the Fermi–Walker derivative of its acceleration bivector vanishes. [This Fermi–Walker derivative is the first term in (20).] We observe that since the vanishing of the local twistor derivative is a conformally invariant condition, a conformally invariant notion of uniform acceleration is simply that the bivector in (20) vanish.

We have left open the possibility of interpreting two-twistors that satisfy some appropriate causality conditions, but which are neither kinematic twistors nor twistors representing uniformly accelerated particles. We have not been able to find any really satisfactory physical analog for such other two-twistors. However, if this *is* possible it seems that such a two-twistor would describe a semiclassical particle with a spin vector undergoing some kind of transport along a curve. It should also be mentioned that higher valence symmetric twistors lying in certain Poincaré-invariant subspaces of the appropriate twistor spaces represent multipole moments.[14] It is open to see whether symmetric twistors lying outside these subspaces represent some kind of uniformly evolving multipole moments.

ACKNOWLEDGMENT

This work was carried out in part while visiting at Oregon State University, Corvallis; I thank the Mathematics department there for its support.

[1]R. Penrose and M. A. H. MacCallum, Phys. Rep. **6**, 242 (1972).

[2]R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, 1986), Vol. 2.

[3]R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, 1984), Vol. 1.

[4]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1975), 4th ed., p. 40.

[5]See Ref. 2, pp. 55 and 77.

[6]The only equivalent formula that I know of in the literature–Eq. (3.19) of R. Penrose, in *J. C. Maxwell, the Sesquicentennial Symposium*, edited by M. S. Berger (Elsevier, Amsterdam, 1984), pp. 211–243—is incorrect. Inserting the derivative expressions for the spinor parts obtained later in the paragraph into the expansion of the twistor pairing that follows them corrects Penrose's equation in the two-index case.

[7]E. T. Newman and J. Winicour, J. Math. Phys. **15**, 1113 (1974).

[8]L. P. Hughston and T. R. Hurd, Phys. Rep. **100**, 273 (1983). See their discussion on p. 309. This fact is analogous to the fact that the primary part of a valence one twistor vanishes precisely on an "alpha plane" in complexified compactified Minkowski space, whose trace in real Minkowski space is a null geodesic—the world line of a massless particle.

[9]The analogous characterization of kinematic two-twistors (which are discussed in the second paragraph of Sec. IV) is given in Ref. 1, pp. 308–310.

[10]Reference 2, pp. 255–258.

[11]I. Robinson and J. R. Robinson, in *General Relativity, Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Clarendon, Oxford, 1972), pp. 151–166.

[12]See Ref. 4, p. 284. An extensive discussion of the momentum and angular momentum integrals is given in R. Sachs and P. G. Bergmann, Phys. Rev. **112**, 674 (1958).

[13]R. Penrose, Proc. R. Soc. London Ser. A **381**, 53 (1982).

[14]G. E. Curtis, Proc. R. Soc. London Ser. A **359**, 133 (1978).

The elastoplastic shock problem as an example of the resolution of ambiguities in the multiplication of distributions

Jean-François Colombeau
Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France

(Received 3 May 1988; accepted for publication 10 May 1989)

One-dimensional collisions between two isotropic solids, in which the equations of physics lead to "multiplications of distributions," are considered. Based on this example, a general physicomathematical method, to be adapted to each particular case, is proposed to resolve the ambiguity inherent in such products. This can be achieved with the aid of a new mathematical theory of generalized functions, which permits dealing with mathematical phenomena of a microscopic nature that govern products of distributions having singularities at the same point. This tool has recently been applied in various situations (in continuum mechanics) in which the equations of physics lead to "heuristic products of distributions." One obtains new (algebraic) formulas in the simplest cases, and new numerical schemes in more general cases. The key to the resolution of ambiguities lies in more precise statements of the laws of physics than are permitted within distribution theory, and have no analog in classical analysis, so that in general a resolution cannot be obtained from "formal calculations."

I. INTRODUCTION

We consider a frontal collision of two homogeneous isotropic solid layers with indefinite extension in the direction perpendicular to their common axis of symmetry, so that there is no "definite center."

We assume that the physical variables depend only on x and t; then the general system of elastoplasticity (Appendix 1 of Ref. 1) reduces to a one-dimensional system of five equations (plus a few other equations that are dissociated). Three of them express the conservations of mass, momentum, and energy; the two other ones are constitutive equations following from Hooke's law when the material is elastic, and from (for instance) a Mie–Gruneisen equation of state when the material is plastic.

Strong enough collisions produce shock waves. Then the system is wrought with several products appearing in the form of classically undefined (indeed really ambiguous) products of distributions. Indeed the usual formulation of nonlinear elasticity[2] does not hold globally in the case of rather strong collisions, such as the ones of projectiles on armor; there is not even a bijective correspondence between the stress and the strain. One states Hooke's law as an infinitesimal linear stress–strain relationship in a Lagrangian frame of reference following the medium, and then one expresses the full system in a fixed (Eulerian) frame of reference, since the Lagrangian frame is not convenient for numerical simulations. Although mathematically meaningless within distribution theory, the system of equations is certainly "correct" since it is successfully used by engineers for the design of armors and projectiles (Ref. 1).

There appear multiplications of distributions in many domains of physics: elasticity and elastoplasticity[1,3] shocks in fluids,[4] thermodynamics,[5] acoustics,[6–8] plasma physics,[9] relativity and astrophysics,[4,10] and quantum field theories.[11–13]

In this last subject multiplications of distributions have given rise to a wide literature for a long time. Those of the δ^2 kind appearing in the adiabatic limit[11,10] allow an obvious interpretation in our theory.[14] Those appearing in the formal perturbation series are much harder to interpret (renormalization theory) and have not been fully understood at present, especially in the case of nonrenormalizable theories. In this latter case it seems that a mathematical tool alone cannot resolve the ambiguities. But, perhaps, an intricate correlation between (new) mathematics and (new or not) physics, as presented in this paper for elastoplasticity, could resolve them. The goal of this paper is to present to physicists this possibility.

Using the mathematical theory of "new generalized functions" introduced in the literature,[1,15–17] one can state the system of equations of elastoplasticity in an original, but physically natural and mathematically correct, form. Then we obtain nonambiguous jump conditions that we compute explicitly; more generally one obtains new numerical schemes.

We believe that the main interest of this paper lies in that it presents a simple model of the use of this new method, which lies at the interface between pure mathematics, theoretical physics, and numerical analysis. It is clear that this method could be used in a wide variety of circumstances (nonlinear and linear problems in which there appear "heu-

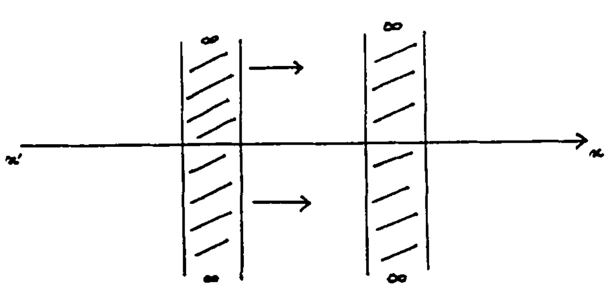


FIG. 1. Frontal collision of two homogeneous layers.

 0022-2488/89/102273-07$02.50

ristic multiplications of distributions," see Refs. 1,3,6,18–22 for some results obtained in this way). The ambiguities which appear in equations of physics cannot usually be resolved on purely mathematical grounds: by allowing more precise formulations of the physical postulates, our method permits to resolve them, at least in the cases that were studied up to now. The crucial point lies in that, without the new mathematical tool, the required more precise formulation of physics—to be discussed in each kind of application anyway—may be quite hidden, even on a purely formal level.

II. THE CLASSICAL STATEMENT OF A ONE-DIMENSIONAL SYSTEM OF ELASTOPLASTICITY, AND ITS MATHEMATICAL COMPLEXITY

In the case of strong collisions the solids are no longer Hookean: a linear stress strain relationship cannot hold since different strains can correspond to the same stress, at a given instant (the history of the collision has also to be taken into account). A constitutive equation is obtained by stating the differential form of Hooke's law in a (Lagrangian) frame of reference following the medium (the "Lamé constants" can depend on the state of the material). We use classical notations: $\rho =$ density, $v = (1/\rho) =$ specific volume, $I =$ internal energy, $e =$ total energy, $\vec{u} = (u_i)_{i<i<3} =$ velocity vector, $\overline{\Sigma} = (\sigma_{i,j})_{1<i<j<3} =$ stress tensor, $\overline{S} = (S_{i,j})_{1<i<j<3} =$ stress deviation tensor, $p = -\frac{1}{3}(\sigma_{1,1} + \sigma_{2,2} + \sigma_{3,3}) =$ pressure. Note that $e = I + \frac{1}{2}(\vec{u}\cdot\vec{u})$ and $\overline{S} = \overline{\Sigma} + pI$ if I is the identity 3×3 matrix. The system obtained is given in Ref. 1 (Appendix 1). In the one-dimensional case under consideration it reduces to (simplified notations $u = u_1, S = S_{1,1}$)

$$\rho_t + (\rho u)_x = 0, \quad \text{balance of mass,}$$

$$(\rho u)_t + (\rho u^2)_x + (p - S)_x = 0,$$

balance of momentum,

$$(\rho e)_t + [\rho e u + (p - S)u]_x = 0, \quad \text{balance of energy,}$$

$$S_t + uS_x - k^2(S)u_x = 0,$$

the deviation part of Hooke's law, (1)

$$p = \Phi(\rho, I),$$

constitutive equation (usually the isotropic part of Hooke's law in the elastic state, and a Mie–Gruneisen equation in the plastic state),

where $k^2 = (4/3)G$ (G is the shear modulus) is a function of $|S|$ and where Φ is a function. k^2 depends on $|S|$ and is null for $|S|$ large enough, say $|S| \geqslant S_0$. Then the material is a fluid and (1) reduces to the classical system of fluid dynamics: we say that the material is in the plastic state. The terms uS_x and $k^2(S)u_x$ put in evidence products of distributions of the kind $Y\cdot\delta$ ($Y =$ Heaviside function, $\delta =$ Dirac mass) when u and S are discontinuous simultaneously (case of shock waves). We do not take into account external forces (gravity), thermal effects, viscosity, and phenomena in the phase transition.

In the classical context of weak solutions of nonlinear PDE's the fourth equation in (1) cannot have discontinuous solutions; we refer to Ref. 23 for a standard textbook. In the context of our generalized solutions[1,15–17] this equation has solutions provided system (1) is stated in a suitably weak form.

It is explained in Refs. 1, 6, and 20 how systems in nonconservative form, like (1), have an infinite number of mathematically possible jump conditions. In Refs. 1 and 20 a method is explained to get rid of this ambiguity, in the case of systems involving only one constitutive equation.

III. A MORE PRECISE FORMULATION AND A RESOLUTION OF THE AMBIGUITIES

The mathematical tool is sketched for physicists in Refs. 6 and 20. It originated in a construction (of pure mathematics[1,15–17] of a differential algebra $\mathscr{G}(\Omega)$ ($\Omega =$ any open set in $\mathbf{R}^n$) containing the vector space $\mathscr{D}'(\Omega)$ of all distributions on Ω. The elements of $\mathscr{G}(\Omega)$ ("new generalized functions") have properties mimicking exactly those of the C^∞ functions (differentiation, multiplication, etc). The classical concept of equality splits necessarily into two concepts (both inducing on $\mathscr{D}'(\Omega)$ the classical equality): a strong one denoted by $=$ (and allowing exactly all the standard calculations) and a weak one denoted by $\approx$ [generalizing exactly the concept of distributional equality, and in general incoherent with the multiplication: $G_1 \approx G_2$ does not imply $GG_1 \approx GG_2$ for arbitrary $G, G_1, G_2, \in \mathscr{G}(\Omega)$].

At first, one states the system (1) with weak equalities $\approx$ since they correspond exactly to the usual concept of a weak solution in the distributional sense. Then one finds an infinite number of possible jump conditions for steady shocks, see Refs. 1 and 20. This corresponds faithfully to the classical ambiguity inherent in most nontrivial products of distributions. In contrast, if all equations in (1) are written with the strong equality, then one proves easily that the system cannot admit shock waves solutions, which is unacceptable, see Refs. 1 and 20.

The difference between the two concepts of weak and strong equalities lies in "microscopic phenomena" such as those occurring in the "width of a shock wave" (of the order of magnitude of a few mean free paths). Postulating that the laws of physics (balance of mass, momentum and energy) are valid within the "microscopic" width of the shock, we are led to state them with the strong equality. The constitutive equations have never been checked in a state of very fast deformation, such as the one inside the shock. This remark suggests to state them with the weak equality $\approx$ [intuitively $p \approx \phi(\rho, I)$ means that $p = \phi(\rho, I)$ in the classical sense outside the width of the shock wave, and that something like $|p - \phi(\rho, I)| < +\infty$ holds within this width]. Then the system (1) becomes

$$\begin{aligned} &\rho_t + (\rho u)_x = 0, \\ &(\rho u)_t + (\rho u^2)_x + (p - S)_x = 0, \\ &(\rho e)_t + [\rho e u + (p - S)u]_x = 0, \\ &S_t + uS_x - k^2(S)u_x \approx 0, \\ &p \approx \phi(\rho, I). \end{aligned} \tag{2}$$

In the case of a steady shock it follows from the first two equations in (2) (Refs. 1 and 20) that $v = 1/\rho$, u, and $p - S$ are represented by the "same Heaviside function" [in $\mathscr{G}(\mathbf{R})$

there are several generalized functions equal to 0 if $x<0$, to 1 if $x>0$, and whose jump at $x=0$ is of the kind of a classical discontinuity]. In this case we say that v, u, and $p-S$ vary in phase on a shock.

When the material is plastic, i.e., $k^2(S)=0$, then $S=\text{const}=S_0$ and we are in the case of hydrodynamics; system (2) has nonambiguous (classical) jump conditions; formulation (2) permits a new formulation of the system in terms of v, u, and p (which gives at once new numerical methods), see Refs. 1, 19, 21, 22, 24. Then system (2) appears in the form

$$\begin{aligned} &v_t+uv_x-vu_x=0,\\ &u_t+uu_x+vp_x=0,\\ &p_t+up_x+[(\gamma+1)p-F(v)-vF'(v)]u_x\approx 0,\\ &S=S_0, \end{aligned} \tag{3}$$

if the Mie–Gruneisen equation is stated in the form $p=\gamma\rho I-F(v)$, $\gamma>0$ and F a positive function of v.

When the material is elastic, i.e., $|S|<S_0$, when one adopts as constitutive equation $p\approx\phi(\rho,I)$ the isotropic part of Hooke's law (see Ref. 1 Appendix 1) and drops the balance of energy, then system (2) appears in the form

$$\begin{aligned} &v_t+uv_x-vu_x=0,\\ &u_t+uu_x+v(p-S)_x=0,\\ &S_t+uS_x-k^2(S)u_x\approx 0,\\ &p_t+up_x+a^2u_x\approx 0, \end{aligned} \tag{4}$$

($a>0$ is a constant: this last equation follows from the isotropic part of Hooke's law).
One usually considers that $k^2(S)=k^2$, $k>0$ constant. The strong equalities in the first two equations imply that u, v, and $p-S$ are represented by the same Heaviside function on a shock (see Refs. 1 and 20), but there exist ambiguities in the terms uS_x and up_x, since we do not know the individual behavior of S and p, relatively to u. Setting $\sigma=S-p$, (4) becomes

$$\begin{aligned} &v_t+uv_x-vu_x=0,\\ &u_t+uu_x-v\sigma_x=0,\\ &\sigma_t+u\sigma_x-(k^2+a^2)u_x\approx 0,\\ &p_t+up_x+a^2u_x\approx 0, \end{aligned} \tag{4'}$$

and the ambiguity lies only in the term up_x (from above, the term $u\sigma_x$ is no longer ambiguous). We propose to resolve this ambiguity, as follows, by a method of "transverse fictitious infinitesimal shock waves." For this we take into account that the real phenomenon is a three-dimensional one. Let us imagine an infinitesimal shock wave in the direction $y'y$ or $z'z$. One would obtain as above (strong equalities for the statement of the laws of balance of mass and momentum) that v and σ_{22} (respectively, v and σ_{33}) are in phase in this infinitesimal shock; since this holds also for v and σ_{11}, we obtain that $p=-\frac{1}{3}(\sigma_{11}+\sigma_{22}+\sigma_{33})$ and v are in phase in an infinitesimal shock. Thus all variables v,u,p,σ (and so S) are in phase in an infinitesimal shock.

Note: This reasoning does not apply for a noninfinitesimal shock, as will be obvious from the results in Secs. IV and V.

Remark: This result is based only on the strong equalities in the statement of balance of mass and momentum, and on the hypothesis of transverse fictitious infinitesimal shock waves. It does not depend on the constitutive equations and on the (weak or strong) statement of the balance of energy.

We shall show (by explicit formulas in a simplified case in the next section, then by numerical results in general) that this is sufficient to resolve the ambiguities due to the multiplications of distributions in the heuristic system (1).

IV. RESOLUTION BY EXPLICIT JUMP FORMULAS

For greater simplicity in explicit calculations we consider only shocks in which the density varies slightly in the neighborhood of a fixed value ρ_0 (this assumption is not always justified physically: in strong metallic shocks the relative variation of density can reach 0,1). Then as explained in Refs. 1 and 20, the first two equations in (2) may be replaced by

$$\rho_0(u_t+uu_x)+(p-S)_x\approx 0,$$

together with the assumption that u and $p-S$ are in phase on a shock. Setting $\rho_0=1/v_0$ we obtain, in the elastic case $|S|<S_0$), the system ($S=\sigma+p$)

$$\begin{aligned} &u_t+uu_x-v_0\sigma_x\approx 0,\\ &p_t+up_x+a^2u_x\approx 0\\ &\text{(isotropic form of Hooke's law)},\\ &\sigma_t+u\sigma_x-b^2u_x\approx 0(b^2=k^2+a^2), \end{aligned} \tag{5}$$

with u, p, σ in phase on infinitesimal shocks (and u, σ in phase on finite shocks). In the plastic case ($S=S_0$) one obtains from (3) the system

$$\begin{aligned} &u_t+uu_x+v_0p_x\approx 0,\\ &p_t+up_x+(dp+e)u_x\approx 0\\ &\quad\text{(Mie–Gruneisen equation of state)},\\ &S=S_0, \end{aligned} \tag{5'}$$

with u and p in phase on all shocks ($d,e=$ constant numbers).

Calculations of jump conditions for (5): We are going to show that the assumption that u, p, and σ are in phase on infinitesimal shocks implies, in the case of (5), that they are in phase also on finite shocks. Setting

$$w=\Delta w H(x-ct)+w_l,$$

with $w=u,\sigma,p$ in (5) ($H=$ Heaviside generalized function), one obtains at once the relations [one uses the relation $HH'\approx(\frac{1}{2})H'$ which follows from $H^2\approx H$ by differentiation]

$$\begin{aligned} &c-\frac{\Delta u}{2}=u_l-v_0\frac{\Delta\sigma}{\Delta u},\\ &c-\frac{\Delta u}{2}=u_l+a^2\frac{\Delta u}{\Delta p},\\ &c-\frac{\Delta u}{2}=u_l-b^2\frac{\Delta u}{\Delta\sigma}. \end{aligned} \tag{6}$$

Elimination of c gives

$$\frac{\Delta u}{\Delta p}=\pm\frac{b\sqrt{v_0}}{a^2},\quad \frac{\Delta\sigma}{\Delta u}=\pm\frac{b}{\sqrt{v_0}}. \tag{6'}$$

Therefore, if u,p,σ are in phase, then their jumps are proportional. Therefore a superposition of shocks in which u, p, and σ are in phase gives a shock in which u, p, and σ are still in phase (one has to be aware of the obvious fact that a superposition of two shocks in phase, but in which the variables have nonproportional jumps, leads to shock waves which are no longer in phase). Considering a shock as a superposition of infinitesimal shocks, the property that u, p, and σ are in phase on infinitesimal shocks implies in the case of (5) that u, p, and σ are also in phase on finite shocks.

Note: In this reasoning we have assumed that all infinitesimal shocks under consideration satisfy (6′) with the same sign + or −, which is the case in the physical situation under consideration. There is, further, some lack of mathematical rigor in this reasoning, since we treat an "infinite sum of infinitesimal shocks" like a "finite sum of finite shocks." Therefore a more refined mathematical analysis would be welcome. We content ourselves with the above fact to justify *that the variables u, p, and σ are in phase on (finite) shocks for the system (5).*

Now the ambiguities in the multiplications of distributions are resolved and we are able to compute explicitly jump formulas for (5) and (5′). For simplicity we consider only the case of a shock wave in which the respective values of (u,p,S) are (0,0,0) on the right-hand side and $(u_1,p_1,-S_0)$ on the left-hand side. This represents the shock wave produced in a target at rest by a projectile. This assumption is only a minor simplification to make the computations easier; exactly the same method applies when the values of (u,p,S) on both sides are arbitrary and, then, for the explicit solution of the Riemann problem by algebraic formulas: one finds elastoplastic shock waves like the one considered here, and also elastic precursors, see Fig. 5 and Ref. 1 (Appendix 4 of Chap. 3). It is convenient to sum up the system and assumptions (with $\sigma = S - p$)

$$u_t + uu_x - v_0\sigma_x \approx 0,$$
$$\sigma_t + u\sigma_x + (l(S)\sigma + m(S))u_x \approx 0,$$

where

$$l(S) = 0, \quad m(S) = -(k^2 + a^2)$$

if $|S| < S_0$ (isotropic part of Hooke's law), (7)

$$l(S) = \alpha, \quad m(S) = \beta$$

if $|S| = S_0$ [(α,β = constants) Mie–Gruneisen equation of state],

$$S_t + uS_x - k^2(S)u_x \approx 0,$$

where

$$k^2(S) = k^2,$$

if $|S| < S_0$ (deviation part of Hooke's law),

$$k^2(S) = 0,$$

if $|S| = S_0$ (no stress deviation in fluids),

u and $\sigma = S - p$ are in phase on (global) shocks, (7′)

u, σ, and p are in phase on the elastic part and on the plastic part of the shock. (7″)

The microscopic profile of the shock can be represented by Fig. 2.

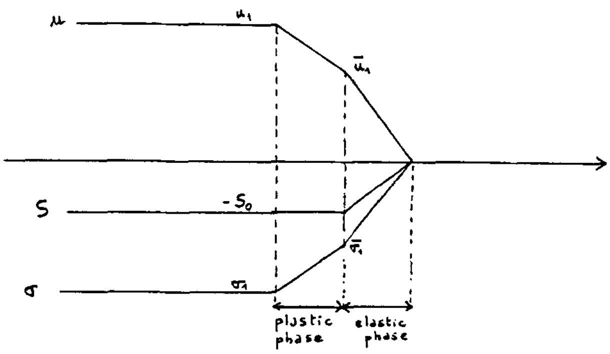


FIG. 2. Microscopic profile of the shock wave under consideration.

One has introduced the values $\bar{u}_1$ and $\bar{\sigma}_1$ corresponding to the elastoplastic transition. Let c be the shock velocity. Then (7″) leads to the following form for u, σ, S:

$$u(x,t) = (u_1 - \bar{u}_1)H(-x+ct) + \bar{u}_1K(-x+ct),$$
$$\sigma(x,t) = (\sigma_1 - \bar{\sigma}_1)H(-x+ct) + \bar{\sigma}_1K(-x+ct),$$
$$S(x,t) = -S_0K(-x+ct), \tag{8}$$

in which H, K are two Heaviside generalized functions, with $HK = 0$ since they represent nonoverlapping phenomena (elastic and plastic).

Equation (7′) implies that u/u_1 and σ/σ_1 are equal since they are the Heaviside functions (in the variable $-x+ct$) of u and σ in the global shock. This gives

$$(1 - \bar{u}_1/u_1)H + \bar{u}_1/u_1K = (1 - \bar{\sigma}_1/\sigma_1)H + \bar{\sigma}_1/\sigma_1K,$$

i.e.,

$$\bar{u}_1/u_1 = \bar{\sigma}_1/\sigma_1. \tag{9}$$

The first equation in (7) gives

$$c(u_1 - \bar{u}_1)H' + c\bar{u}_1K'$$
$$+ \{(u_1 - \bar{u}_1)H + \bar{u}_1K\}\{-(u_1 - \bar{u}_1)H' - \bar{u}_1K'\}$$
$$- v_0\{-(\sigma_1 - \bar{\sigma}_1)H' - \bar{\sigma}_1K'\} \approx 0,$$

i.e.,

$$cu_1 - (u_1 - \bar{u}_1)^2/2 - \bar{u}_1^2/2 + v_0\sigma_1 = 0,$$
$$cu_1 - u_1^2/2 + u_1\bar{u}_1 - \bar{u}_1^2 + v_0\sigma_1 = 0. \tag{10}$$

The second equation in (7) gives

$$c\sigma_1 - \tfrac{1}{2}[(u_1 - \bar{u}_1)(\sigma_1 - \bar{\sigma}_1) + \bar{u}_1\bar{\sigma}_1]$$
$$- \alpha/2(\sigma_1 - \bar{\sigma}_1)(u_1 - \bar{u}_1) + \bar{u}_1(k^2 + a^2)$$
$$- \beta(u_1 - \bar{u}_1) = 0.$$

Elimination of $\bar{\sigma}_1$ with (9) gives

$$c\sigma_1 - (\tfrac{1}{2} + \alpha/2)u_1\sigma_1 + (1+\alpha)\bar{u}_1\sigma_1$$
$$- (1 + \alpha/2)(\bar{u}_1^2/u_1)\sigma_1 + (k^2 + a^2 + \beta)\bar{u}_1$$
$$- \beta u_1 = 0. \tag{11}$$

The third equation in (7) gives

$$-cS_0K' + \bar{u}_1KS_0K' + k^2\bar{u}_1K' \approx 0,$$

i.e.,

$$cS_0 - \bar{u}_1S_0/2 - k^2\bar{u}_1 = 0. \tag{12}$$

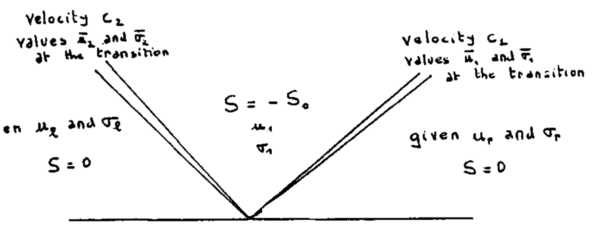


FIG. 3. Schematic representation of the solution of the Riemann problem calculated in Sec. IV: we have eight equations for the eight unknowns c_1, c_2, $\bar{u}_1, \bar{\sigma}_1, \bar{u}_2, \bar{\sigma}_2, u_1, \sigma_1$. An example is given in Fig. 4. For some values of u_l, σ_l, u_r, σ_r and of the coefficients one obtains elastic precursors (see Fig. 5), then the Riemann problem has a different solution.

Equations (9)–(12) contain the five unknowns $c, \bar{u}_1, \bar{\sigma}_1, u_1, \sigma_1$. One can more generally compute the complete solution of the Riemann problem in the form of Fig. 3.

These formulas prove that our original method easily gives algebraic formulas. They can be used to verify the adequacy of numerical methods. These explicit formulas for the solution of the Riemann problem can also be used to build new numerical codes.

V. RESOLUTION BY NUMERICAL METHODS

The ambiguity in multiplications of distributions appears from the fact that slightly different numerical schemes give different solutions (while they all give the same solutions in the case where the multiplications of distributions are not involved), see Ref. 1 (Appendix 2 of Chap. 3). But one has learned how to find numerical schemes expressing that some equations are stated with the strong equality[19] or that some variables vary in phase on infinitesimal shocks.[1,21,22] Then these schemes permit a general study of these systems. A few situations are given below, to illustrate and apply the study in Secs. III and IV.

The numerical results in Figs. 4 and 5 have been obtained from discretization techniques in Refs. 1 and 23 based on the method in Sec. III. A longer computation time or minor technical improvements would give very steep shocks. We prefer to reproduce the curves in Fig. 4 in order to observe easily the values $\bar{u}_1$ and $\bar{\sigma}_1$.

VI. CONCLUSION

The original mathematical tool permits a more precise formulation of physics, which resolves the ambiguities that usually appear when one attempts to solve classically, even formally, i.e., without mathematical rigor, problems involving "multiplications of distributions." This fact seems to us to be of wide interest due to the very large variety of physical problems involving such multiplications.

An interesting point is that this mathematical tool brings up new ideas and can be very well taken up on an intuitive basis; no deep mathematical study is required to use it successfully in physical situations; one only needs some familiarity with the basic ideas as explained above in Sec. III. The genuine difficulty in the resolution of ambiguities, in any particular situation, lies in the more precise way in which, with this new tool, one should formulate the equations. Up to now, in continuum mechanics, this has been done only by using classical ideas (the difference between the basic equations and the constitutive equations). In quantum field theory one may wonder whether classical ideas would be sufficient, or if new ideas would be needed for this more precise formulation.

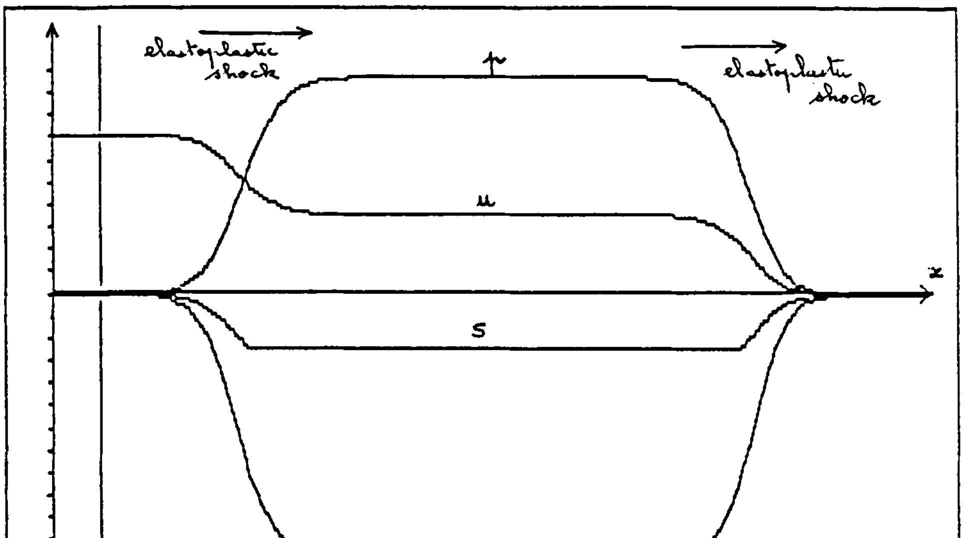


FIG. 4. *A numerical solution of the Riemann problem for system (7) at time $t > 0$.* In this case $k^2 = 4$, $a^2 = 9$, $\alpha = 0$, $\beta = -9$, $S_0 = 2$, $v_0 = 1$. The initial conditions at $t = 0$ are on the left side $x < 0$ (projectile) $u = 6$, $p = S = \sigma = 0$; and on the right side $x > 0$ (target) $u = p = S = \sigma = 0$. One observes two elastoplastic shock waves propagating to the right at different velocities. The right-hand side wave is the one theoretically depicted in Fig. 2 (in u, S, σ).

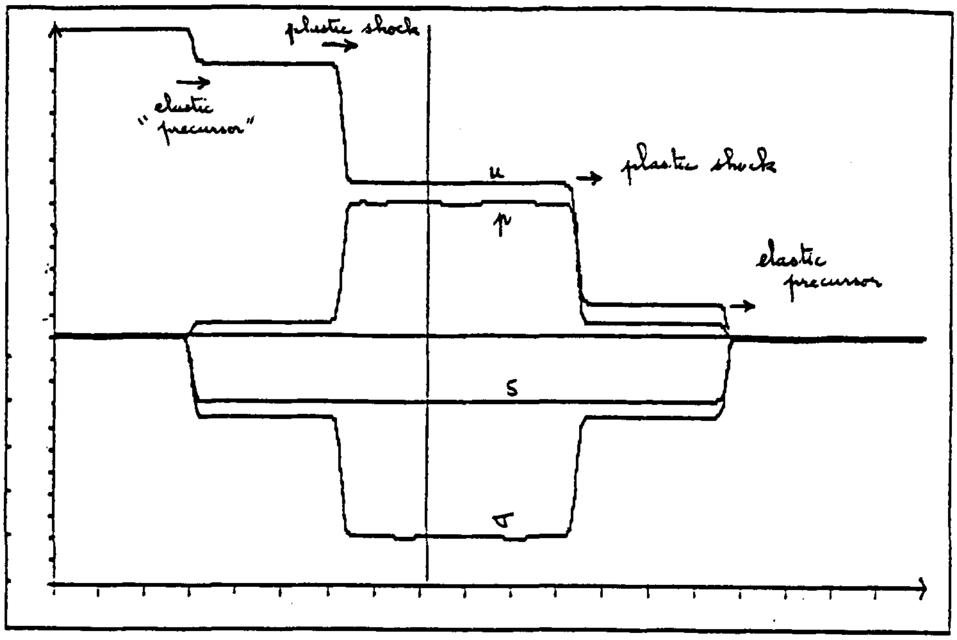


FIG. 5. *A numerical solution of the Riemann problem for system (7) at time $t > 0$.* In this case $k^2 = 4$, $a^2 = 1$, $\alpha = 0$, $\beta = -9$, $S_0 = 0{,}1$, $v_0 = 1$. The initial conditions at $t = 0$ are on the left side $x < 0$ (projectile) $u = 0{,}5$, $p = S = \sigma = 0$; on the right side $x > 0$ (target) $u = p = S = \sigma = 0$. One observes that "each shock as in Fig. 4" is dissociated into an "elastic precursor" and a "plastic shock wave." The following is used as empirical evidence of this dissociation in certain cases. In plane accidents involving a frontal collision it has been observed that passengers in the tail of the plane often survive. An elastoplastic shock wave (very destructive) would forbid any chance of survival. But in the case of a dissociation as above, the elastic precursor (not destructive) is reflected by the tail of the plane; then this tail is cut off at the place where the reflected elastic precursor meets the incoming plastic shock wave. Thus the tail escapes from the destructive plastic shock wave.

ACKNOWLEDGMENT

The author is indebted to the referee for various improvements in the text.

APPENDIX: MULTIPLICATION OF DISTRIBUTIONS IN MATHEMATICS AND PHYSICS

The formula $Y\delta = \frac{1}{2}\delta$ is certainly the more natural one concerning multiplication of Y and δ. Indeed, in our context, as soon as Y is a Heaviside generalized function one has $Y^2 \approx Y$ and by derivation $Y\delta \approx \frac{1}{2}\delta$ if $\delta = Y'$. This formula gives physically correct results in several cases (Refs. 1 and 20, even those in Fig. 5) but gives radically incorrect ones in the case of elastoplastic shock waves involving a phase transition [Figs. 2 and 4, (Ref. 20) Appendix B]. In our context the key of the paradox lies in that there are several Heaviside-like, Dirac-like, etc., functions. When $\delta \neq Y'$ then the product $Y \cdot \delta$ is not in general associated with $\frac{1}{2}\delta$.

In this case a mathematically correct multiplication of distributions (usually adopted by mathematicians) results in physically incorrect jump conditions. In our context there is a canonical inclusion $\mathscr{D}' \subset \mathscr{G}$. In $\mathscr{D}'$ there is one and only one Heaviside-like element Y_0, one and only one Dirac-like element δ_0; further $\delta_0 = Y_0'$. Thus $Y_0 \cdot \delta_0 \approx \frac{1}{2}\delta_0$: one recovers the formulas usually adopted by mathematicians. The observation that this formula might be inadequate is not paradoxical if one thinks of the different ways in which physicists and mathematicians conceive and use distributions.

For mathematicians the space $\mathscr{D}'$ is defined modulo an isomorphism (concerning all operations). Such an isomorphic copy of $\mathscr{D}'$ is canonically included in $\mathscr{G}$. It permits, via the multiplication in $\mathscr{G}$ and the association, a synthesis of most existing mathematical multiplications of distributions, see Refs. 14 and 25.

For physicists the space $\mathscr{D}'$ is considered as a reservoir of mathematical objects used to describe the physical world. In our context the use of the above subspace $\mathscr{D}'$ of $\mathscr{G}$ as such a reservoir may lead to mistakes in some cases involving "multiplications of distributions." Then the correct reservoir is $\mathscr{G}$ itself, which contains several Heaviside-like, Dirac-like, etc., functions.

In this way a nonambiguous mathematical multiplication of distributions can be reconciled with the well known fact that in physics multiplications of distributions such as $Y\delta$ or δ^2 can give different results according to the context.

In Part III of Ref. 15 the author attempted to explain (perturbative) renormalization in Q.F.T. by *adjusting the definition of the multiplication* (by changes in the definition of the auxiliary set A_q) in order to obtain the renormalized results. One was forced to this trick from the postulate (done in Ref. 15 from a too narrow interpretation of the new setting) that the free field operators were (vector valued) elements of $\mathscr{D}'$, and so precisely defined elements of $\mathscr{G}$ through the inclusion $\mathscr{D}' \subset \mathscr{G}$. An exactly similar postulate in continuum mechanics would amount to imposing the formula $Y\delta \approx \frac{1}{2}\delta$. The subsequent work done in Refs. 1, 6, 20, and 26 suggests that the correct viewpoint is to postulate only that the free field operators are (*a priori* unknown) elements of $\mathscr{G}$ which are associated with certain well defined elements of $\mathscr{D}'$ (the free field operators considered as distributions).

Then, in order to obtain the renormalized results, one should determine, from physical ground (as done in this paper, for instance by stating some equations with strong equality in $\mathcal{G}$), which precise elements of $\mathcal{G}$ the free (and also the interacting) field operators are. Attempts in particular cases show that the above trick of adjusting definitions amounts to such determination. This better method might be predictive, as it is in continuum mechanics.[1,6,19,20,27]

Shock wave solutions of systems in nonconservative form can illustrate the standard opinion (Richtmyer,[23] p. 37) that "no amount of mathematical reasoning can tell us which set of weak solutions has the right to be called a generalized solution." Nonconservative systems usually have an infinite number of possible weak solutions (in the sense of association in $\mathcal{G}$) which are equally acceptable from the mathematical viewpoint.[1,19,26] *In each physical situation a choice of one kind of solution, i.e., a resolution of ambiguities, can only be done on physical ground.*

[1]H. A. Biagioni, "Introduction to a Nonlinear Theory of Generalized Functions," preprint series Notas de Matematica, UNICAMP, 1988.

[2]Y. C. Fung, *A First Course in Continuum Mechanics* (Prentice–Hall, Englewood Cliffs, NJ, 1969).

[3]J. J. Cauret, "Analyse et developpement d'un code elastoplastique bidimentionnel," doctoral thesis, Bordeaux, 1986.

[4]C. K. Raju, "Junction conditions in general relativity," J. Phys. A. Math. Gen. **15**, 1785 (1982).

[5]D. Bedeaux, A. M. Albano, and P. Mazur, "Boundary conditions and non-equilibrium thermodynamics, Physica A **82**, 438 (1978).

[6]J. F. Colombeau, "Multiplications de distributions et acoustique," Rev. Acoust. **1**, 9 (1988).

[7]B. Poiree, "Les équations de l'acoustique linéaire et non linéaire," Acustica **57**, 5 (1985).

[8]B. Poiree, "Equations de perturbation et equations de passage associées," Rev. CETHEDEC **69**, 1 (1981).

[9]J. J. Lodder, "A simple model for a symmetrical theory of generalized functions V" Physica A **11**, 404 (1982).

[10]C. K. Raju, "Products and composition with the Dirac delta function," J. Phys. A. Math. Gen. **15**, 1785 (1982).

[11]N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York 1959).

[12]A. I. Akiezher and V. B. Berestetskii, *Quantum Electrodynamics* (Interscience, New York, 1965).

[13]I. B. Manoukian, *Renormalization* (Academic, New York, 1983).

[14]J. F. Colombeau and M. Oberguggenberger, "Generalized functions and products of distributions," preprint, 1987.

[15]J. F. Colombeau, *New Generalized Function and Multiplication of Distributions* (North-Holland, Amsterdam, 1984).

[16]J. F. Colombeau, *Elementary Introduction to New Generalized Functions* (North-Holland, Amsterdam, 1985).

[17]E. E. Rosinger, *Generalized Solutions of Nonlinear Partial Differential Equations* (North-Holland, Amsterdam, 1987).

[18]J. J. Cauret, J. F. Colombeau, and A. Y. Le Roux, "Discontinuous generalized solutions of nonlinear nonconservative hyperbolic equations," J. Math. Ana. Appl. **139**, 552 (1989).

[19]J. F. Colombeau and B. Perrot, "A numerical method for the solution of nonlinear systems of physics involving multiplications of distributions," preprint 1988.

[20]J. F. Colombeau and A. Y. Le Roux, "Multiplications of distributions in elasticity and hydrodynamics," J. Math. Phys. **29**, 315 (1988).

[21]J. F. Colombeau and A. Y. Le Roux, "Numerical techniques in elastodynamics, *Lecture Notes in Mathematics*, Vol. 1270 (Springer, New York, 1987), pp. 103–114.

[22]J. F. Colombeau and A. Y. Le Roux, "Numerical methods for hyperbolic systems in nonconservative form using products of distribution. Advances in Compter Methods for Partial Differential Equations VI," International Association for Mathematics and Computers in Simulation, 1987, pp. 28–37.

[23]R. D. Richtmyer, *Principles of Advanced Mathematical Physics* (Springer, Berlin, 1978).

[24]P. DeLuca, "Modélisation numérique en elastoplacité dynamique," doctoral thesis, Bordeaux, 1989.

[25]M. Oberguggenberger, "Products of distributions," J. Reine Agnew. Math. **365**, 1 (1986).

[26]J. F. Colombeau, A. Y. Le Roux, A. Noussair, and B. Perrot," Microscopic profiles of shock waves and ambiguities in multiplication of distributions," SIAM J. Numer. Anal., August 1989.

[27]Y. A. Barka, J. F. Colombeau, and B. Perrot, "A numerical modelling of the fluid-fluid acoustic dioptra," J. d'Acoustique, December 1989.

A class of exact plane wave solutions of the Maxwell–Dirac equations

A. Das and D. Kay[a)]
Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 4 October 1988; accepted for publication 31 May 1989)

Exact solutions of Maxwell–Dirac equations are investigated for which the Dirac field is of the type $\psi(x) = \alpha(p)e^{ip_\lambda x^\lambda}$. In the subclass where the mass parameter $m \neq 0$, there exists *no nontrivial solution* of the problem. In the subclass where the mass parameter $m = 0$, there exist infinitely many solutions inherent with arbitrary functions. Furthermore, every solution for $m = 0$ must have a null four-current vector field associated with it.

I. INTRODUCTION

In quantum electrodynamics, no nontrivial exact solution is known. The perturbative techniques involving Green's functions invariably lead to divergence difficulties. On the other hand, classical nonlinear field equations such as Einstein's vacuum field equations,[1] Einstein–Maxwell equations,[1] and Einstein–Maxwell–Dirac equations[2] have yielded plenty of exact solutions (without perturbative approaches). That is why we are motivated to investigate exact solutions of Maxwell–Dirac equations, which represent classical electrodynamics. Recently, the initial value problem for Maxwell–Dirac equations has been tackled by Flato, Simon, and Taflin.[3] In Sec. II, we write down the Maxwell–Dirac equations and an associated differential identity. We also prove that the four-current vector field $j^\lambda(x)$ is always nonspacelike, irrespective of the (real or complex) values of the mass parameter m.

In the case of the free Dirac equation, the plane wave solutions are the easiest ones to find. This fact prompts us to seek in Sec. III the class of exact solutions of Maxwell–Dirac equations such that the Dirac field is of the type $\psi(x) = \alpha(p)e^{ip_\lambda x^\lambda}$. This problem has been divided into two subclasses according to the mass parameter $m \neq 0$, or $m = 0$. In the first subclass ($m \neq 0$), it is proved that there exists *no nontrivial solution* of the problem. In the second subclass ($m = 0$), there exist infinitely many solutions involving *arbitrary constants* and *arbitrary functions.* All the solutions of the problem $m = 0$ are obtained and classified into four cases involving subcases. Every solution of this subclass must have an associated *null* four-current vector field $j^\lambda(x)$. This result is physically reasonable for the massless particles.

II. NOTATIONS AND FIELD EQUATIONS

The combined Maxwell–Dirac equations are studied in a flat (Minkowski) space-time manifold M. A Minkowski coordinate chart is used. (In the sequel, a mixed coordinate chart will be defined and used.) A space-time event is indicated by $x \equiv (x^0, x^1, x^2, x^3)$ where x^0 denotes the time coordinate. A greek index takes values from $\{0,1,2,3\}$ and a roman index takes values from $\{1,2,3\}$. The signature of the metric is assumed to be -2, so that the metric tensor $[\eta_{\mu\nu}] = \mathrm{diag}[1,(-1)^3]$. Einstein's summation convention is followed. The electromagnetic four-potential and field are denoted by $A^\mu(x)$ and $F_{\mu\nu}(x) \equiv \partial_\nu A_\mu - \partial_\mu A_\nu$, where the partial derivatives are denoted by ∂_μ. The four-component Dirac bispinor field and its Hermitian conjugate are indicated by $\psi(x)$ and $\psi^\dagger(x)$, respectively. The four 4×4 Dirac matrices are denoted by γ^μ. In this section, we assume that in the domain D of consideration, the potential functions A^μ are of the differentiability class $C^3(D)$ and the Dirac bispinor functions $\psi_1, \psi_2, \psi_3, \psi_4$ are of the class $C^2(D)$. In such a domain $D \subset M$, the combined Maxwell–Dirac equations (which are Poincaré covariant and gauge invariant) can be written as

$$\begin{aligned} M^\mu(x) &\equiv \partial_\nu F^{\mu\nu} - j_\mu(x) \\ &\equiv \partial_\nu F^{\mu\nu} - e\psi^\dagger(x)\gamma^0\gamma^\mu\psi(x) = 0, \\ D(x) &\equiv \{i\gamma^\mu[\partial_\mu + ieA_\mu(x)] - mI\}\psi(x) = 0, \quad (2.1) \\ F_{\mu\nu}(x) &\equiv \partial_\nu A_\mu - \partial_\mu A_\nu, \\ \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu &= 2\eta^{\mu\nu} I. \end{aligned}$$

Here e and m are the charge and mass parameters associated with the Dirac field and I stands for the 4×4 identity matrix. In this combined system of partial differential equations, there exists *one differential identity*, viz.,

$$\begin{aligned} \mathscr{I} &\equiv \partial_\mu M^\mu - ie[\psi^\dagger(x)\gamma^0 D(x) \\ &\quad - D^\dagger(x)\gamma^0\psi(x)] \equiv 0. \end{aligned} \quad (2.2)$$

Therefore, to make the system (2.1) determinate we have to impose one additional equation [which is not inconsistent with system (2.1)]. If we choose as that equation the Lorentz-gauge condition, then the combined system (2.1) goes over to

$$\begin{aligned} M^m(x) &\equiv \Box A^\mu - j^\mu(x) \equiv \partial^\nu\partial_\nu A_\mu \\ &\quad - e\psi^\dagger(x)\gamma^0\gamma^\mu\psi(x) \\ &= 0, \\ D(x) &\equiv \{i\gamma^\mu[\partial_\mu + ieA_\mu(x)] - mI\}\psi(x) = 0, \quad (2.3) \\ L(x) &\equiv \partial_\mu A^\mu = 0. \end{aligned}$$

The above system is invariant under the restricted gauge transformations:

[a)] Present address: Okanagan College, Penticton, British Columbia V5A 8E1, Canada.

 0022-2488/89/102280-05$02.50

$$\hat{\psi}(x) = \psi(x) e^{-ie\Omega(x)},$$
$$\hat{A}_\mu(x) = A_\mu(x) - \partial_\mu \Omega, \qquad (2.4)$$
$$\Box\Omega = 0.$$

We shall use the following Weyl representation of the 4×4 Dirac matrices:

$$\gamma^0 = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, \quad \gamma^k = \begin{bmatrix} 0 & -\sigma^k \\ \sigma^k & 0 \end{bmatrix}, \qquad (2.5)$$

where σ^k's are the 2×2 Pauli matrices and I stands for the 2×2 identity matrix.

Now we are in a position to state and prove the following theorem about a purely algebraic property of the Fermionic four-current vector field $j^\mu(x)$.

Theorem 2.1: Let the Dirac bispinor field $\psi(x)$ be defined (but not necessarily continuous) in a domain $D \subset M$. Then, the four-current vector field $j^\mu(x) \equiv e\psi^\dagger(x)\gamma^0\gamma^\mu\psi(x)$, is nowhere spacelike in D, irrespective of the (real or complex) values of the mass parameter m.

Proof: Here, the star stands for complex conjugation, and the vertical bar denotes the modulus of a complex number. By a straightforward computation we obtain

$$j_\mu(x) j^\mu(x) = 4e^2 |\psi_1^*(x)\psi_3(x) + \psi_2^*(x)\psi_4(x)|^2. \qquad (2.6)$$

Since, the right-hand-side of Eq. (2.6) is always a nonnegative real number, it follows that $j^\mu(x)$ is a nonspacelike four-vector field. ■

Now we shall introduce a mixed coordinate chart for the flat space-time M by the following coordinate transformation:

$$u \equiv \hat{x}^0 = \tfrac{1}{2}(x^0 + x^3),$$
$$v \equiv \hat{x}^3 = \tfrac{1}{2}(x^0 - x^3),$$
$$\rho \equiv \hat{x}^2 = \tfrac{1}{2}(x^1 + ix^2),$$
$$\rho^* \equiv \hat{x}^3 = \tfrac{1}{2}(x^1 - ix^2);$$
$$[J] \equiv \frac{\partial x^\rho}{\partial \hat{x}^\mu} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix},$$
$$\det[J] = -4i \neq 0. \qquad (2.7)$$

The coordinates u, v are usually called the *null* or *light-cone coordinates* and ρ, ρ^* are called the *complex-conjugate coordinates*. Under the coordinate transformation (2.7), the various tensor and spinor fields are assumed to transform as follows: $\hat{x} \equiv (u, \rho, \rho^*, v)$;

$$\hat{A}^\mu(\hat{x}) = \frac{\partial \hat{x}^\mu}{\partial x^\nu} A^\nu(x),$$
$$\hat{A}^u(\hat{x}) = \tfrac{1}{2}[A^0(x) + A^3(x)],$$
$$\hat{A}^v(\hat{x}) = \tfrac{1}{2}[A^0(x) - A^3(x)],$$
$$\hat{A}^\rho(\hat{x}) = \tfrac{1}{2}[A^1(x) + iA^2(x)],$$
$$\hat{A}^{\rho^*}(\hat{x}) = [\hat{A}^\rho(x)]^*;$$
$$[\hat{\eta}_{ab}] = [J]^T [\eta_{\nu\mu}] [J]$$
$$= 2\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix};$$
$$\hat{A}_u(\hat{x}) = 2\hat{A}^v(\hat{x}), \quad \hat{A}_v(\hat{x}) = 2\hat{A}^u(\hat{x}),$$
$$\hat{A}_\rho(\hat{x}) = -2A^{\rho^*}(\hat{x}), \quad \hat{A}_{\rho^*}(\hat{x}) = -2A^\rho(\hat{x});$$
$$\hat{\psi}(\hat{x}) \equiv \psi(x);$$
$$\hat{\sigma}^u = \tfrac{1}{2}(\sigma^0 + \sigma^3) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{\sigma}^v = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad (2.8)$$
$$\hat{\sigma}^\rho = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \hat{\sigma}^{\rho^*} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix};$$
$$\hat{\gamma}^u = \tfrac{1}{2}(\gamma^0 + \gamma^3) = \begin{pmatrix} 0 & \hat{\sigma}^v \\ \hat{\sigma}^u & 0 \end{pmatrix}, \quad \hat{\gamma}^v = \begin{pmatrix} 0 & \hat{\sigma}^u \\ \hat{\sigma}^v & 0 \end{pmatrix},$$
$$\hat{\gamma}^\rho = \begin{pmatrix} 0 & -\hat{\sigma}^\rho \\ \sigma^\rho & 0 \end{pmatrix}, \quad \hat{\gamma}^{\rho^*} = \begin{pmatrix} 0 & -\hat{\sigma}^{\rho^*} \\ \hat{\sigma}^{\rho^*} & 0 \end{pmatrix};$$
$$\hat{j}^u(\hat{x}) = e[|\hat{\psi}_1(\hat{x})|^2 + |\hat{\psi}_4(\hat{x})|^2],$$
$$\hat{j}^v(\hat{x}) = e[|\hat{\psi}_2(\hat{x})|^2 + |\hat{\psi}_3(\hat{x})|^2],$$
$$\hat{j}^\rho(\hat{x}) = e[\hat{\psi}_1^*(\hat{x})\hat{\psi}_2(\hat{x}) - \hat{\psi}_3^*(\hat{x})\hat{\psi}_4(\hat{x})],$$
$$\hat{j}^{\rho^*}(\hat{x}) = e[\hat{\psi}_2^*(\hat{x})\hat{\psi}_1(\hat{x}) - \hat{\psi}_4^*(\hat{x})\hat{\psi}_3(\hat{x})].$$

Dropping hats in the sequel, the Maxwell–Dirac equations (2.3) in the mixed coordinates, by Eqs. (2.7) and (2.8), go over into

$$x \equiv (u, v, \rho, \rho^*);$$
$$\Box \equiv \partial_u \partial_v - \partial_\rho \partial_{\rho^*};$$
$$M^u(x) \equiv \Box A^u - e[|\psi_1(x)|^2 + |\psi_4(x)|^2] = 0,$$
$$M^v(x) \equiv \Box A^v - e[|\psi_2(x)|^2 + |\psi_3(x)|^2] = 0,$$
$$M^\rho(x) \equiv \Box A^\rho - e[\psi_1^*(x)\psi_2(x) - \psi_3^*(x)\psi_4(x)],$$
$$M^{\rho^*}(x) = [M^\rho(x)]^* = 0;$$
$$D_1(x) \equiv [\partial_v + ieA_v(x)]\psi_3(x) - [\partial_\rho + ieA_\rho(x)]\psi_4(x) + im\psi_1(x) = 0, \qquad (2.9)$$
$$D_2(x) \equiv [\partial_u + ieA_u(x)]\psi_4(x) - [\partial_{\rho^*} + ieA_{\rho^*}(x)]\psi_3(x) + im\psi_2(x) = 0,$$
$$D_3(x) \equiv [\partial_u + ieA_u(x)]\psi_1(x) + [\partial_\rho + ieA_\rho(x)]\psi_2(x) + im\psi_3(x) = 0,$$
$$D_4(x) \equiv [\partial_v + ieA_v(x)]\psi_2(x) + [\partial_{\rho^*} + ieA_{\rho^*}(x)]\psi_1(x) + im\psi_4(x) = 0;$$
$$L(x) \equiv \partial_u A^u + \partial_v A^v + \partial_\rho A^\rho + \partial_{\rho^*} A^{\rho^*} = 0.$$

The preceding systems of equations is easier to solve than the equivalent system (2.3) in Minkowski coordinates.

III. A SPECIAL CLASS OF PLANE WAVE SOLUTIONS

In the case of the *free* Dirac equation, the exact plane wave solutions of the form $\psi(x) = \alpha(p)e^{ip_\mu x^\mu}$ are found and discussed in the standard textbooks. Fourier integrals involving these plane wave solutions yield more general exact solutions of the free Dirac equation. These considerations motivate us to seek the class of exact solutions of the Maxwell–Einstein equations such that the Dirac field is of the type $\psi(x) = \alpha(p)e^{ip_\mu x^\mu}$. This problem has to be divided into two subclasses according to $m \neq 0$, or else $m = 0$. *Complete solutions* can be found in both subclasses.

We shall define the notion of the trivial solution for subsequent use. The Dirac field $\psi(x)$ is called *trivial* in a domain D of flat space-time M, provided $\psi_1(x) = \psi_2(x) = \psi_3(x) = \psi_4(x) = 0$ for *all* x in D.

Now we are in a position to state and prove rigorous statements on the exact plane wave solutions of the type $\psi(x) = \alpha(p)e^{ip_\mu x^\mu}$.

Theorem 3.1: Let the potential functions $A^\mu \in C^2(D)$, and the Maxwell–Dirac equations (2.9) hold in a bounded domain $D \subset M$, with $e \neq 0$, $m \neq 0$. Let, furthermore, the Dirac bispinor field be of the type $\psi(x) = \alpha(p)e^{ip_\mu x^\mu}$, where $x = (u, v, \rho, \rho^*)$. Then, solutions of these equations exist in D only if the Dirac field is trivial $[\psi(x) \equiv 0]$.

Proof: Let us define

$$\Omega(x) \equiv (e)^{-1}[p_\mu x^\mu]$$
$$= (e)^{-1}[p_u u + p_v v + p_\rho \rho + p_{\rho^*}\rho^*]. \quad (3.1)$$

This function satisfies the wave equation

$$\Box\Omega = 0.$$

Therefore, we can make a gauge transformation [cf. Eq. (2.4)]

$$\hat{\psi}(x) = e^{-ie\Omega(x)}\psi(x) = e^{-ie\Omega(x)}\alpha(p)e^{ip_\mu x^\mu} = \alpha(p),$$
$$\hat{A}_\mu(x) = A_\mu(x) - \partial_\mu\Omega = A_\mu(x) - (e)^{-1}p_\mu. \quad (3.2)$$

Dropping the hats subsequently, and denoting the values $\alpha = \alpha(p)$ (which are independent of x), the system of equations (2.9) reduces to

$$M^u(x) \equiv \Box A^u - e(|\alpha_1|^2 + |\alpha_4|^2) = 0,$$
$$M^v(x) \equiv \Box A^v - e(|\alpha_2|^2 + |\alpha_3|^2) = 0,$$
$$M^\rho(x) \equiv \Box A^\rho - e(\alpha_1^*\alpha_2 - \alpha_3^*\alpha_4) = 0;$$
$$i(e)^{-1}D_1(x) \equiv \alpha_3 A_v(x) - \alpha_4 A_\rho(x) - (m/e)\alpha_1 = 0,$$
$$i(e)^{-1}D_2(x) \equiv \alpha_4 A_u(x) - \alpha_3 A_{\rho^*}(x) - (m/e)\alpha_2 = 0,$$
$$i(e)^{-1}D_3(x) \equiv \alpha_1 A_u(x) + \alpha_2 A_\rho(x) - (m/e)\alpha_3 = 0,$$
$$i(e)^{-1}D_4(x) \equiv \alpha_2 A_v(x) + \alpha_1 A_{\rho^*}(x) - (m/e)\alpha_4 = 0;$$
$$L(x) \equiv \partial_u A^u + \partial_v A^v + \partial_\rho A^\rho + \partial_{\rho^*} A^{\rho^*} = 0. \quad (3.3)$$

We shall first solve the Maxwell equations $\Box A^\nu = j^\nu$. We write $A^\nu(x) = A^\nu_h(x) + A^\nu_p(x)$ such that $\Box A^\nu_h(x) = 0$ and $\Box A^\nu_p = j^\nu$. Since the four-current vector field is a constant vector field, we can write the particular solutions as $A^\nu_p(x) = (1/2)j^\nu(uv - |\rho|^2)$. Furthermore, the homogeneous solutions can be written as $A^\nu_h(x) = \pm\frac{1}{2}W^\nu(x)$, where $W^\nu(x)$ are *arbitrary wave functions* in the domain D. The general solutions of Maxwell's equations can be written as

$$2A^u(x) = e(|\alpha_1|^2 + |\alpha_4|^2)(uv - |\rho|^2)$$
$$+ W^u(x) = A_v(x),$$
$$2A^v(x) = e(|\alpha_2|^2 + |\alpha_2|^2)(uv - |\rho|^2)$$
$$+ W^v(x) = A_u(x),$$
$$2A^\rho(x) = e(\alpha_1^*\alpha_2 - \alpha_3^*\alpha_4)(uv - |\rho|^2)$$
$$- W^\rho(x) = -A_{\rho^*}(x). \quad (3.4)$$

Now, substituting (3.4) into the Dirac equation $i(e)^{-1}D_1(x) = 0$, we obtain

$$e(uv - |\rho|^2)(|\alpha_1|^2\alpha_3 + \alpha_1\alpha_2^*\alpha_4)$$
$$+ [\alpha_3 W^u(x) - \alpha_4 W^{\rho^*}(x)] - (m/e)\alpha_1 = 0.$$

Operating by the d'Alembertian $\Box$ from the left on the above equation we get

$$\alpha_1(\alpha_1^*\alpha_3 + \alpha_2^*\alpha_4) = 0, \quad (3.5a)$$
$$\alpha_3 W^u(x) - \alpha_4 W^{\rho^*}(x) - (m/e)\alpha_1 = 0. \quad (3.5b)$$

Similarly, from the other three Dirac equations

$$-i(e)^{-1}D_2(x) = -i(e)^{-1}D_3(x)$$
$$= -i(e)^{-1}D_4(x) = 0,$$

we obtain

$$\alpha_2(\alpha_1^*\alpha_3 + \alpha_2^*\alpha_4) = 0, \quad (3.6a)$$
$$\alpha_4 W^v(x) - \alpha_3 W^\rho(x) - (m/e)\alpha_2 = 0, \quad (3.6b)$$
$$\alpha_3(\alpha_1\alpha_3^* + \alpha_2\alpha_4^*) = 0, \quad (3.7a)$$
$$\alpha_1 W^v(x) + \alpha_2 W^\rho(x) - (m/e)\alpha_3 = 0, \quad (3.7b)$$
$$\alpha_4(\alpha_1\alpha_3^* + \alpha_2\alpha_4^*) = 0, \quad (3.8a)$$
$$\alpha_2 W^u(x) + \alpha_1 W^\rho(x) - (m/e)\alpha_4 = 0. \quad (3.8b)$$

By contraposition, let us assume that the Dirac field is nontrivial, i.e.,

$$|\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2 + |\alpha_4|^2 > 0. \quad (3.9)$$

Then, from Eqs. (3.5a), (3.6a), (3.7a), and (3.8a), it follows that

$$\alpha_1^*\alpha_3 + \alpha_2^*\alpha_4 = 0 = \alpha_1\alpha_3^* + \alpha_2\alpha_4^*. \quad (3.10)$$

Now, multiplying (3.5b) by α_1^* and (3.8b)* by α_4, and adding, we obtain

$$(\alpha_1^*\alpha_3 + \alpha_2^*\alpha_4)W^u(x) + 0 - (m/e)(|\alpha_1|^2 + |\alpha_4|^2) = 0.$$

Using Eq. (3.10) and dividing by $-(m/e) \neq 0$, the above equation yields

$$|\alpha_1|^2 + |\alpha_4|^2 = 0. \quad (3.11)$$

Similarly, multiplying (3.7b) by α_3^* and (3.6b)* by α_2, and adding we get

$$(\alpha_1\alpha_3^* + \alpha_2\alpha_4^*)W^v(x) + 0 - (m/e)(|\alpha_2|^2 + |\alpha_3|^2)$$
$$= 0 + 0 - (m/e)(|\alpha_2|^2 + |\alpha_3|^2) = 0.$$

Dividing by $-(m/e)$, the preceding equation yields

$$|\alpha_2|^2 + |\alpha_3|^2 = 0. \quad (3.12)$$

Equation (3.11) plus (3.12), *contradicts* the strict inequality (3.9). Thus the theorem is proved. ■

In the above proof, $m \neq 0$ was used only in the latter part, in deriving Eq. (3.11) and (3.12). Therefore, Eq. (3.1)–(3.10) hold for every value of m, including $m = 0$. Unlike the subclass of $m \neq 0$, the zero mass subclass allows infinitely many nontrivial solutions. In fact, the zero mass subclass can be completely solved. We shall summarize a case in this subclass by the following theorem.

Theorem 3.2: Let the potential function $A^{\mu}\in C^2(D)$, and the Maxwell–Dirac equations (2.9) hold in a bounded domain $D\subset M$, with $e\neq 0$, $m=0$. Furthermore, let the Dirac bispinor field be of the type $\psi(x)=\alpha(p)e^{ip_\mu x^\mu}$ such that every component of $\alpha(p)$ is nonzero. In this case I, the general solutions of the equations exist and can be summarized [after the gauge transformation (3.2) and dropping hats] as follows.

Case I: All components nonzero, i.e.,

$$\alpha_1\neq 0,\quad \alpha_2\neq 0,\quad \alpha_3\neq 0,\quad \alpha_4\neq 0,$$

and

$$\alpha_1\alpha_3^* + \alpha_2\alpha_4^* = 0,$$

but otherwise arbitrary.

$$\begin{aligned}2A^u(x) &= e(|\alpha_1|^2+|\alpha_4|^2)(uv-|\rho|^2)\\ &\quad + W^u(x),\\ 2A^v(x) &= e(|\alpha_2|^2+|\alpha_3|^2)(uv-|\rho|^2)\\ &\quad + |\alpha_2/\alpha_1|^2 W^u(x),\\ 2A^\rho(x) &= e(\alpha_1^*\alpha_2-\alpha_3^*\alpha_4)(uv-|\rho|^2)\\ &\quad + (\alpha_2/\alpha_1)W^\rho(x),\\ 2W^u(x) &= -e\{(|\alpha_2|^2+|\alpha_3|^2)u^2\\ &\quad + |\alpha_1/\alpha_2|^2(|\alpha_2|^2+|\alpha_4|^2)v^2\\ &\quad - (\alpha_1/\alpha_2)(\alpha_1\alpha_2^*-\alpha_3\alpha_4^*)\rho^2\\ &\quad - (\alpha_1^*/\alpha_2^*)(\alpha_1^*\alpha_2-\alpha_3^*\alpha_4)\rho^{*2}\}\\ &\quad + \mathrm{Re}\left\{\int_{\mathbb{C}^2}[F(\lambda r+\lambda\sigma+\nu\sigma^*;\lambda,\nu)\right.\\ &\quad \left. + G(\nu r+\lambda\sigma+\nu\sigma^*;\lambda,\nu)]d\lambda\, d\nu\right\}.\end{aligned} \quad (3.13)$$

Here,

$$r\equiv|\alpha_2/\alpha_1|^2u+v,$$
$$\sigma\equiv v-(\alpha_2/\alpha_1)^*\rho,$$
$$\sigma^*\equiv v-(\alpha_2/\alpha_1)\rho^*;$$

F, G are arbitrary twice differentiable functions such that the improper integral over $\mathbb{C}^2$ is uniformly convergent.

Proof: Equation (3.10) implies $\alpha_1\alpha_3^*+\alpha_2\alpha_4^*=0$, which can be solved by setting

$$\alpha_2=\beta\alpha_1,\quad \alpha_3^*=-\beta\alpha_4^*, \quad (3.14)$$

where β is an arbitrary nonzero complex constant. The Lorentz-gauge condition $L(x)=0$ in (3.3), by (3.4), yields

$$\begin{aligned}&\partial_u W^u+\partial_v W^v-\partial_\rho W^\rho-\partial_{\rho^*}W^{\rho^*}\\ &\quad = -e\{(|\alpha_1|^2+|\alpha_2|^2)v+(|\alpha_2|^2+|\alpha_3|^2)u\\ &\qquad -(\alpha_1^*\alpha_2-\alpha_3^*\alpha_4)\rho^*-(\alpha_1\alpha_2^*-\alpha_3\alpha_4^*)\rho\}.\end{aligned} \quad (3.15)$$

Dirac equations (3.5b), (3.6b), (3.7b), (3.8b) (for $m=0$) go over to

$$\left.\begin{aligned}D_1'(x)&\equiv\alpha_3^*W^u(x)-\alpha_4^*W^\rho(x)=0,\\ D_2'(x)&\equiv\alpha_4W^v(x)-\alpha_3W^\rho(x)=0,\\ D_3'(x)&\equiv\alpha_1^*W^v(x)+\alpha_2^*W^\rho(x)=0,\\ D_4'(x)&\equiv\alpha_2W^u(x)+\alpha_1W^\rho(x)=0.\end{aligned}\right\} \quad (3.16)$$

We want to solve for the unknown functions $W^\mu(x)$ from the above system of linear equations. The rank of the coefficient matrix is two, due to the condition (3.10). Thus we have to solve only two independent equations such as $D_1'(x)=D_2'(x)=0$. Solving these we obtain

$$W^\rho(x)=-\beta W^u(x),\quad W^v(x)=|\beta|^2W^u(x),$$
$$\beta=(\alpha_2/\alpha_1)=-(\alpha_3/\alpha_4)^*. \quad (3.17)$$

Putting (3.17) into (3.4), we obtain the first three equations of (3.13) for the four-potential functions $A^\nu(x)$. Substituting (3.17) into (3.15), we get the following linear partial differential equation:

$$\begin{aligned}&(\partial_u+|\beta|^2\partial_v+\beta\partial_\rho+\beta^*\partial_{\rho^*})W^u\\ &\quad = -e[(|\alpha_2|^2+|\alpha_3|^2)u+(|\alpha_1|^2+|\alpha_2|^2)v\\ &\qquad -(\alpha_1\alpha_2^*-\alpha_3\alpha_4^*)\rho-(\alpha_1^*\alpha_2-\alpha_3^*\alpha_4)\rho^*].\end{aligned} \quad (3.18)$$

By the method of characteristic curves,[4] we obtain the general solution of (3.18) as

$$\begin{aligned}W^u(x)&=\mathrm{Re}[f(r,\sigma,\sigma^*)]-\frac{e}{2}\left\{(|\alpha_2|^2+|\alpha_3|^2)u^2\right.\\ &\quad +\frac{(|\alpha_1|^2+|\alpha_4|^2)v^2}{|\beta|^2}-\frac{(\alpha_1\alpha_2^*-\alpha_3\alpha_4^*)}{\beta}\rho^2\\ &\quad \left.-\frac{(\alpha_1^*\alpha_2-\alpha_3^*\alpha_4)}{\beta^*}\rho^{*2}\right\},\end{aligned} \quad (3.19)$$

$$r\equiv|\beta^2|u-v,\quad \sigma\equiv v-\beta^*\rho,$$

$$\sigma^*=v-\beta\rho^*,\quad p\equiv|\beta|^2u+v.$$

Here, f is an arbitrary, complex-valued, twice differentiable function of its arguments.

Now, W^u is a wave function. Therefore, from (3.19) we obtain

$$\begin{aligned}0=\Box W^u&=(\partial_u\partial_v-\partial_\rho\partial_{\rho^*})\{\mathrm{Re}[f(r,\sigma,\sigma^*)]\}\\ &=(-\partial_r^2+\partial_r\partial_\sigma+\partial_r\partial_{\sigma^*}-\partial_\sigma\partial_{\sigma^*})\{\mathrm{Re}[f(r,\sigma,\sigma^*)]\}.\end{aligned} \quad (3.20)$$

The above equation yields

$$\Box f=-\Box f^*,\quad \Box f=ih(r,\sigma,\sigma^*), \quad (3.21)$$

where h is a real-valued continuous function. Solving (3.21) we get

$$f(r,\sigma,\sigma^*)=i\int G(\cdots)h(\cdots)dr'\,d\sigma'\,d\sigma'^*+\phi(r,\sigma,\sigma^*), \quad (3.22)$$

where $G(\cdots)$ is the real-valued Green's function, and ϕ is an arbitrary complex-valued wave function. Therefore, we must have

$$\mathrm{Re}[f(r,\sigma,\sigma^*)]=\mathrm{Re}[\phi(r,\sigma,\sigma^*)],$$
$$\Box\phi=(-\partial_r^2+\partial_r\partial_\sigma+\partial_r\partial_{\sigma^*}-\partial_\sigma\partial_{\sigma^*})\phi=0. \quad (3.23)$$

We notice that the above differential operator is a homogeneous function of degree 2 in the first partial differential operators. Therefore, we use the ansatz

$$\phi(r,\sigma,\sigma^*)=F(\mu r+\lambda\sigma+\nu\sigma^*;\mu,\nu,\sigma), \quad (3.24)$$

where μ,λ,ν are complex parameters. The wave equation $\Box\phi = 0$ implies that

$$(-\mu^2 + \mu\lambda + \mu\nu - \lambda\nu)F'' = 0. \quad (3.25)$$

Assuming that F'' exists and is not necessarily zero, we have

$$\mu^2 - \mu(\lambda + \nu) + \lambda\nu = 0. \quad (3.26)$$

Solving the quadratic equation, we have $\mu = \lambda$ or $\mu = \nu$. Therefore, a linear combination of functions $F(\lambda r + \lambda\sigma + \nu\sigma^*; \lambda,\nu)$ and $G(\nu r + \lambda\sigma + \nu\sigma^*; \lambda,\nu)$ will solve the wave equation. The general solution[5] of the wave equation in (3.23) is

$$2\mathrm{Re}\,\phi(r,\sigma,\sigma^*) = \mathrm{Re}\left\{\int_{C^2} [F(\lambda r + \lambda\sigma + \nu\sigma^*;\lambda,\nu) + G(\nu r + \lambda\sigma + \nu\sigma^*;\lambda,\nu)]\,d\lambda\,d\nu\right\}. \quad (3.27)$$

Here, F and G are arbitrary, twice differentiable, complex-valued functions of their arguments such that the above improper integral converges uniformly. By Eqs. (3.19), (3.23), and (3.27), the last of the equations in (3.13) is proved and that completes the proof. ■

From Eq. (3.10) it is clear that there exists *no* solution in the case where *three* of α_1, α_2, α_3, α_4 are nonzero. In case II, there exists exactly *two* nonzero components of α. There are infinitely many solutions in this case.

In case III, exactly *one* of the components of α is nonzero and there are infinitely many solutions. In the last case for which Dirac field is trivial ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$), the Maxwell–Dirac equations reduce to Maxwell equations. That case will be ignored. In the following, all the solutions of cases II and III will be listed without proof.

Case II:

(i) $\alpha_3 = \alpha_4 = 0$, $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, but otherwise arbitrary. This subcase is obtained directly from Eq. (3.13) by setting $\alpha_3 = \alpha_4 = 0$.

(ii) $\alpha_1 = \alpha_2 = 0$, $\alpha_3 \neq 0$, $\alpha_4 \neq 0$, but otherwise arbitrary.

$$\begin{aligned}
2A^u(x) &= e|\alpha_4|^2(uv - |\rho|^2) + W^u(x),\\
2A^v(x) &= e|\alpha_3|^2(uv - |\rho|^2) + |\alpha_3/\alpha_4|^2 W^u(x),\\
2A^\rho(x) &= -e\alpha_3^*\alpha_4(uv - |\rho|^2) - (\alpha_3/\alpha_4)^* W^u(x),\\
2W^u(x) &\equiv -e\{|\alpha_3|^2u^2 + |\alpha_4/\alpha_3|^2v^2\\
&\quad - (\alpha_3/\alpha_3^*)(\alpha_4^*)^2\rho^2 - (\alpha_3^*/\alpha_3)(\alpha_4)^2\rho^{*2}\} \quad (3.28)\\
&\quad + \mathrm{Re}\left\{\int_{C^2}[F(\lambda r + \lambda\sigma + \nu\sigma^*;\lambda,\nu)\right.\\
&\quad \left. + G(\nu r + \lambda\sigma + \nu\sigma^*;\lambda,\nu)]\,d\lambda\,d\nu\right\},\\
r &\equiv |\alpha_3/\alpha_4|^2u + v,\quad \sigma \equiv v + (\alpha_3/\alpha_4)\rho,\\
\sigma^* &\equiv v + (\alpha_3/\alpha_4)^*\rho^*,
\end{aligned}$$

(iii) $\alpha_1 = \alpha_4 = 0$, $\alpha_2 \neq 0$, $\alpha_3 \neq 0$, but otherwise arbitrary.

$$\begin{aligned}
A^u(x) &= A^\rho(x) \equiv 0,\\
2A^v(x) &= e(|\alpha_2|^2 + |\alpha_3|^2)(uv - |\rho|^2) + W^v(x), \quad (3.29)\\
W^v(x) &\equiv -e(|\alpha_2|^2 + |\alpha_3|^2)(uv + |\rho|^2) + \mathrm{Re}[F(u,\rho)].
\end{aligned}$$

Here, F is an arbitrary function of the u variable and an arbitrary holomorphic function of the ρ variable.

(iv) $\alpha_2 = \alpha_3 = 0$, $\alpha_1 \neq 0$, $\alpha_4 \neq 0$, but otherwise arbitrary.

$$\begin{aligned}
A^v(x) &= A^\rho(x) \equiv 0,\\
A^u(x) &= -e(|\alpha_1|^2 + |\alpha_4|^2)|\rho|^2 + \mathrm{Re}[F(v,\rho)]. \quad (3.30)
\end{aligned}$$

Here F is an arbitrary function of the v variable and an arbitrary holomorphic function of the ρ variable.

Case III: Exactly *one* of the components of α is nonzero, but otherwise arbitrary. There are four subcases which can be directly obtained from Eqs. (3.29) and (3.30).

At this stage we have found all the nontrivial, local, plane-wave solutions [of the type $\psi(x) = \alpha(p)e^{ip_\mu x^\mu}$] of the Maxwell–Dirac equations with $m = 0$. Now we shall point out three common features of these solutions.

(i) Comparing Eq. (3.10) with (2.6) we can conclude that the four-current vector field $j^\mu(x)$ is *null* for every solution of this type.

(ii) Transforming back by the gauge transformation (3.2) and the coordinate transformation (2.8), it follows that exact solutions for the Dirac wave function are of the type:

$$\psi(x) = \begin{bmatrix}\alpha_1(p)\\ \alpha_2(p)\\ \alpha_3(p)\\ \alpha_4(p)\end{bmatrix} e^{ip_\nu x^\nu}.$$

Here, the only constraint on the coefficient functions is $\alpha_1(p)[\alpha_3(p)]^* + \alpha_2(p)[\alpha_4(p)]^* = 0$, and there is *no* restriction on the p_ν's at all. Therefore, the four-momentum components p_ν's *need not* satisfy the mass-shell constraint $\eta^{\mu\nu}p_\mu p_\nu - m^2 = 0$.

(iii) Since the gauge transformation (3.1) presupposes that $e \neq 0$, the solutions found in this section do not have limiting cases for $e \to 0$, unless $p_\mu \equiv 0$.

[1] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB Deutscher, Berlin, 1980).
[2] M. J. Hamilton and A. Das, J. Math. Phys. **18**, 2026 (1977).
[3] M. Flato, J. Simon, and E. Taflin, Commun. Math. Phys. **112**, 21 (1987).
[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1966), Vol. II, p. 62.
[5] H. Bateman, *Differential Equations* (Chelsea, New York, 1966), p. 273.

Generalized relations among N-dimensional Coulomb Green's functions using fractional derivatives

S. M. Blinder
Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109
E. L. Pollock
Physics Department, Lawrence Livermore National Laboratory, Livermore, California 94550

(Received 3 February 1989; accepted for publication 17 May 1989)

Hostler [J. Math. Phys. **11**, 2966 (1970)] has shown that Coulomb Green's functions of different dimensionality N are related by $G^{(N+2)} = \mathscr{O}G^{(N)}$, where $\mathscr{O}$ is a first-order derivative operator in the variables x and y. Thus all the even-dimensional functions are connected, as are analogously the odd-dimensional functions. It is shown that the operations of functional differentiation and integration can further connect the even- to the odd-dimensional functions, so that Hostler's relation can be extended to give $G^{(N+1)} = \mathscr{O}^{1/2}G^{(N)}$.

I. INTRODUCTION

Hostler showed in 1970 that Coulomb Green's functions of varying dimension N were related as follows[1–3]:

$$G^{(N+2)}(x,y,k) = -\frac{1}{\pi(x-y)}\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right)G^{(N)}(x,y,k),\quad N=1,2,3,\ldots. \tag{1.1}$$

Here x and y are the two coordinate variables

$$x,y \equiv r_1 + r_2 \pm r_{12} \tag{1.2}$$

and k is the wave number variable, such that, in atomic units ($\hbar = \mu = e = 1$),

$$E = \frac{\hbar^2 k^2}{2\mu} = \frac{k^2}{2},\quad \nu \equiv \frac{Z}{k}. \tag{1.3}$$

Thus the odd-dimensional functions $G^{(3)}, G^{(5)},\ldots$ are obtained by successive differentiation of $G^{(1)}$, while the even-dimensional functions follows analogously from $G^{(2)}$. We will show in this paper that the even- and odd-dimensional Coulomb Green's functions can be further connected to one another by the operations of fractional differentiation and integration.

By the N-dimensional Coulomb Green's function we understand the solution of the inhomogeneous differential equation:

$$\left(\frac{1}{2}k^2 + \frac{1}{2}\nabla_N^2 + \frac{Z}{\mathbf{r}_N}\right)G^{(N)}(\mathbf{r}_N, \mathbf{r}'_N, k) = \delta^{(N)}(\mathbf{r}_N - \mathbf{r}'_N), \tag{1.4}$$

which is not to be confused with the solution to Poisson's equation in N-dimensional space.

II. RESUME OF THE FRACTIONAL CALCULUS

The monograph of Oldham and Spanier[4] gives a definitive presentation of the fractional calculus. A brief heuristic account of some relevant results will suffice to make this paper self-contained.

Multiple differentiation in the complex plane can be represented by Cauchy's integral formula:

$$f^{(n)}(z) = \frac{n!}{2\pi i}\oint \frac{f(\zeta)d\zeta}{(\zeta - z)^{(n+1)}}, \tag{2.1}$$

for a contour enclosing $\zeta = z$. A possible generalization of (2.1) to derivatives of nonintegral order q defines

$$f^{(q)}(z) = \frac{\Gamma(q+1)}{2\pi i}\int_C \frac{f(\zeta)d\zeta}{(\zeta - z)^{q+1}}. \tag{2.2}$$

For $q \neq n$, $\zeta = z$ becomes a branch point. Let the contour C be taken counterclockwise around z and extending on both sides of a branch cut to a lower limit $\zeta = a$. The values $a = 0$ (Riemann) and $a = -\infty$ (Liouville) are the most common. For $q < 0$, (2.2) reduces to the Riemann–Liouville definition of a fractional derivative, viz.,

$$f^{(q)}(z) = \frac{1}{\Gamma(-q)}\int_a^z \frac{f(\zeta)d\zeta}{(z-\zeta)^{q+1}} \equiv {}_aD_z^q f(z). \tag{2.3}$$

The case $q = -\frac{1}{2}$ is called the semi-integral:

$${}_aD_z^{-1/2} f(z) = \frac{1}{\sqrt{\pi}}\int_a^z \frac{f(\zeta)d\zeta}{(z-\zeta)^{1/2}}. \tag{2.4}$$

For $q > 0$ (and $\neq n$) the singularity at $\zeta = z$ can be removed by integration by parts. Thus the semiderivative, with $q = \frac{1}{2}$, is given by

$${}_aD_z^{1/2} f(z) = \frac{1}{\sqrt{\pi}}\frac{f(a)}{(z-a)^{1/2}} + \frac{1}{\sqrt{\pi}}\int_a^z \frac{f'(\zeta)d\zeta}{(z-\zeta)^{1/2}}. \tag{2.5}$$

We will actually require the limit value $a = +\infty$. For appropriately behaved $f(z)$:

$${}_\infty D_z^{-1/2} f(z) = \frac{i}{\sqrt{\pi}}\int_z^\infty \frac{f(\zeta)d\zeta}{(\zeta - z)^{1/2}}, \tag{2.6}$$

and

$${}_\infty D_z^{1/2} f(z) = \frac{i}{\sqrt{\pi}}\int_z^\infty \frac{f'(\zeta)d\zeta}{(\zeta - z)^{1/2}}. \tag{2.7}$$

III. INTEGRAL REPRESENTATION OF N-DIMENSIONAL GREEN'S FUNCTION

The Coulomb Green's function in N-dimensional space can be expanded as a sum of partial waves as follows[5]:

 0022-2488/89/102285-03$02.50

$$G^{(N)} = \frac{\Gamma(N/2)}{2\pi^{N/2}(N-2)} \sum_{L=0}^{\infty} (2L+N-2) C_L^{N/2-1}(\cos\theta) G_L^{(N)}, \tag{3.1}$$

where $C_L^\nu(z)$ is a Gegenbauer (ultraspherical) polynomial,

$$C_L^\nu(z) = (-)^L \frac{\Gamma(L+2\nu)}{L!\Gamma(2\nu)} {}_2F_1(-L, L+2\nu;\nu+1/2;(1+z)/2). \tag{3.2}$$

The partial-wave retarded Green's functions are given by[6]

$$G_L^{(N)}(r_1,r_2,k) = (ik)^{-1}(r_1r_2)^{(1-N)/2}\Gamma(L+N/2-1/2-i\nu) \times M_{i\nu}^{L+N/2-1}(-2ikr_<) W_{i\nu}^{L+N/2-1}(-2ikr_>), \quad N=3,4,5,..., \tag{3.3}$$

where M and W are Whittaker functions as defined by Buchholz.[7,8]

Using Buchholz's integral representation for the above product of Whittaker functions,

$$G_L^{(N)} = -2(-i)^{2L+N-2}(r_1r_2)^{1-N/2} \int_0^\infty dq\, e^{2i\nu q} e^{ik(r_1+r_2)\coth q} J_{2L+N-2}(2k\sqrt{r_1r_2}\,\mathrm{csch}\, q), \tag{3.4}$$

the summation in (3.1) can be carried out using the Neumann series[9]:

$$\left(\frac{kz}{2}\right)^{\mu-\nu} J_z(kz) = k^\mu \sum_{n=0}^{\infty} \frac{\Gamma(\mu+n)}{n!\Gamma(\nu+1)} {}_2F_1(\mu+n,-n;\nu+1;k^2)(\mu+2n)J_{\mu+2n}(z), \tag{3.5}$$

with the identifications $n=L$, $k=\cos(\theta/2)$, $z=2k\sqrt{r_1r_2}\,\mathrm{csch}\, q$, $\mu=n-2$ and $\nu=(N-1)/2$. The result is the following integral representation for $G^{(N)}$ (see Ref. 10):

$$G^{(N)}(x,y,k) = (2\pi)^{1/2-N/2}(-i)^N k^{N/2-1/2}\eta^{3/2-N/2} \times \int_0^\infty dq (\mathrm{csch}\, q)^{N/2-1/2} e^{2i\nu q} e^{ik\xi\coth q} J_{N/2-3/2}(k\eta\,\mathrm{csch}\, q), \quad N=1,2,3,..., \tag{3.6}$$

where

$$\xi \equiv r_1+r_2 = (x+y)/2, \quad \eta \equiv 2r_1r_2\cos(\theta/2) = \sqrt{xy}. \tag{3.7}$$

The above result for $N=2$ follows by a separate derivation. The case $N=1$ corresponds to Meixner's one-dimensional Coulomb system[11]

$$G^{(1)} = i\eta\int_0^\infty dq\,\mathrm{csch}\, q e^{2i\nu q} e^{ik\xi\coth q} J_1(k\eta\,\mathrm{csch}\, q) = (ik)^{-1}\Gamma(1-i\nu) M_{i\nu}^{1/2}(-iky) W_{i\nu}^{1/2}(-ikx), \tag{3.8}$$

with the closed form following from Buchholz' integral representation. For $N=2$,

$$G^{(2)} = -\frac{1}{\pi}\int_0^\infty dq\,\mathrm{csch}\, q e^{2i\nu q} e^{ik\xi\coth q} \cos(k\eta\,\mathrm{csch}\, q), \tag{3.9}$$

which can be reduced to a series of Whittaker functions,

$$G^{(2)} = -\frac{1}{i\pi k\eta} \sum_{m=-\infty}^{\infty} \Gamma\left(|m|+\frac{1}{2}-i\nu\right) M_{i\nu}^{|m|}(-iky) W_{i\nu}^{|m|}(-ikx), \tag{3.10}$$

but no further reduction to a closed form is known.

IV. RELATIONS AMONG DIFFERENT DIMENSIONALITIES

Hostler's operator [cf. Eq. (1.1)], when applied to a function of ξ and η [cf. Eq. (3.7)], reduces as follows:

$$\mathscr{O} \equiv -\frac{1}{\pi(x-y)}\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial y}\right) = \frac{1}{2\pi\eta}\left(\frac{\partial}{\partial \eta}\right)_\xi = \frac{1}{\pi}D_{\eta^2}. \tag{4.1}$$

By the well-known derivative formula for Bessel functions,[12]

$$\left(\frac{1}{z}\frac{d}{dz}\right)^n z^{-\nu}J_\nu(z) = (-)^n z^{-\nu-n}J_{\nu+n}(z). \tag{4.2}$$

Identifying z with $k\eta\,\mathrm{csch}\, q$, we have

$$D_{\eta^2}^n \eta^{-\nu} J_\nu(k\eta\,\mathrm{csch}\, q) = (-k\,\mathrm{csch}\, q/2)^n \eta^{-\nu-n} J_{\nu+n}(k\eta\,\mathrm{csch}\, q). \tag{4.3}$$

Applying Hostler's operator succesively to the integral representation (3.6) then gives the odd-dimensional Green's function

$$G^{(2N+1)} = \mathcal{O}^N G^{(1)} \tag{4.4}$$

and analogously, for even N,

$$G^{(2N+2)} = \mathcal{O}^N G^{(2)}. \tag{4.5}$$

The identity (4.2) can be reexpressed as follows (with $z \to \sqrt{z}$):

$$D_z^n z^{-\nu/2} J_\nu(\sqrt{z}) = (-\tfrac{1}{2})^n z^{-(\nu+n)/2} J_{\nu+n}(\sqrt{z}). \tag{4.6}$$

Taking $n = 1$ and integrating between the limits a and z, we find

$$\zeta^{-\nu/2} J_\nu(\sqrt{\zeta})\Big]_a^z = -\frac{1}{2}\int_a^z \zeta^{-(\nu+1)/2} J_{\nu+1}(\sqrt{\zeta}). \tag{4.7}$$

For $\nu \geqslant 0$, the lower boundary term in (4.7) vanishes for $a = +\infty$. Thus the analog of (4.6) for negative n (multiple integration) can be written

$$_\infty D_z^{-n} z^{-\nu/2} J_\nu(\sqrt{z}) = (-2)^n z^{-(\nu-n)/2} J_{\nu-n}(\sqrt{z}). \tag{4.8}$$

It is now suggested that (4.6) and (4.8) might be generalized to fractional n. For the semi-integral, Eq. (4.8) with $n = \frac{1}{2}$, use (2.6) and evaluate the integral.[13] The result is

$$_\infty D_z^{-1/2} z^{-\nu/2} J_\nu(\sqrt{z}) = \frac{i}{\sqrt{\pi}}\int_z^\infty \frac{\zeta^{-\nu/2} J_\nu(\sqrt{\zeta})}{(\zeta - z)^{1/2}} d\zeta$$

$$= i\sqrt{2} z^{-\nu/2+1/4} J_{\nu-1/2}(\sqrt{z}). \tag{4.9}$$

Likewise, Eq. (4.6) works for $n = \frac{1}{2}$. One can therefore write the square root of Hostler's operator as

$$\mathcal{O}^{1/2} = -(1/\sqrt{\pi})\, _\infty D_{\eta^2}^{1/2} \tag{4.10}$$

such that

$$\mathcal{O}^{1/2} G^{(N)} = G^{(N+1)}, \quad \mathcal{O}^{N/2} G^{(1)} = G^{(N+1)},$$

$$N = 1,2,3,\ldots . \tag{4.11}$$

This does not, incidentally, provide a closed form for $G^{(2)}$ since the semiderivative still involves either an integral or an infinite sum.

For $Z = 0$, the above reduce to free-particle Green's functions. In particular,

$$G_{\rm FP}^{(1)} = (ik)^{-1}[e^{ik(x-y)/2} - e^{ik(x+y)/2}],$$

$$G_{\rm FP}^{(2)} = -(i/2) H_0^{(1)}(kR), \tag{4.12}$$

$$G_{\rm FP}^{(3)} = -e^{ikR}/2\pi R,$$

where $R \equiv r_{12} = (x - y)/2$. It can be verified that the Hostler operator and its square root also transform among the functions (4.12) in accord with (4.11).

ACKNOWLEDGMENT

I would like to thank Professor John G. Loeser for introducing me to the fractional calculus.

[1]L. C. Hostler, J. Math. Phys. 11, 2966 (1970).

[2]The original derivation of the Coulomb Green's function gave this relation between $G^{(3)}$ and $G^{(1)}$: L. C. Hostler, J. Math. Phys. 5, 591 (1964).

[3]The same operator appeared in connection with the Coulomb density matrix: Yu. N. Demkov and I. V. Komarov, Transactions of Leningrad State University, Series 2, No. 10, pp. 18–28, 1965.

[4]K. B. Oldham and J. Spanier, *The Fractional Calculus* (Academic, New York, 1974). See, also *Fractional Calculus and its Applications*, edited by B. Ross (Springer, New York, 1975), pp. 1–36.

[5]See Ref. 1, Eq. (3).

[6]S. M. Blinder, J. Math. Phys. 25, 905 (1984), Eq. (4.4).

[7]H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969). See especially the integral representation, p. 86, Eq. (5c).

[8]For compactness of notation, we write $M_\kappa^{\mu/2}(z)$ in place of $\mathcal{M}_{\kappa,\mu/2}(z)$ and $W_\kappa^{\mu/2}(z)$ in place of $W_{\kappa,\mu/2}(z)$. See S. M. Blinder, J. Math. Phys. 22, 306 (1981).

[9]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966), p. 140, Sec. 5.21, Eq. (3).

[10]An equivalent result is given in Ref. 1, Eq. (5).

[11]J. Meixner, Math. Z. 36, 677 (1933).

[12]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Natl. Bur. Stand., Washington, DC, 1972), p. 361, Eq. (9.1.30).

[13]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 703, Eq. 6.592, 10.

Free Green's function in a harmonic oscillator basis

B. Silvestre-Brac and C. Gignoux
Institut des Sciences Nucléaires, 53 Avenue des Martyrs, F-38026 Grenoble-Cédex, France
Y. Ayant
Laboratoire de Spectrométrie Physique, B.P. 87-38042 Saint Martin d'Hères-Cédex, France

(Received 28 February 1989; accepted for publication 17 May 1989)

A closed and analytical formula is given for the free Green's function in a harmonic oscillator basis. It is very useful for solving the Lippman–Schwinger equation for the scattering of two clusters within the resonating group method formula.

I. INTRODUCTION

For some problems in atomic or nuclear physics, one is led to solve scattering equations describing collisions of complex systems formed with identical particles. The exact solution for such a problem is presently not feasible once the number of particles exceeds 4. Among the approximate methods introduced to solve the general problem, the resonating group method (RGM) proposed a long time ago by Wheeler[1] is very attractive and powerful. This method and a related one, the generator coordinate method (GCM), have been extensively used in atomic and nuclear scattering.[2] In the RGM, the N-body problem is transformed into a system of coupled channel equations for the wave functions relative to the various partitions defining the scattered clusters. The coupling potentials between distinct channels are strongly nonlocal.

II. RESONATING GROUP METHOD

It is not our aim to enter into detail concerning this method but we shall give the basic equations necessary for understanding the philosophy of our paper. The trial RGM wave function looks like

$$|\Psi(1,2,...,N)\rangle = \sum_c \mathscr{A}\,|\{[\Psi_{c_1}(1,2,...,N_1)\Psi_{c_2}(N_1+1,...,N)]_c\chi_c\}\rangle, \tag{1}$$

where the sum runs over the various channels c defined by the chosen eigenmodes c_1 and c_2 for the clusters 1 (containing particles $1,2,...,N_1$) and 2 (containing particles $N_1+1,...,N$) as well as the various intrinsic couplings (color, spin, isospin, etc.) symbolically denoted by [ ] and the total angular momentum coupling denoted by { }. The state vector $|\chi_c\rangle$ describes the relative motion between the clusters c_1 and c_2.

The variational RGM principle only acts on the relative function χ_c while the cluster states ψ_{c_1} and ψ_{c_2} are supposed frozen once and for all. The Schrödinger equation in the Hilbert space spanned by the RGM function (1) gives rise to the well known Hill–Wheeler equation[3]

$$\sum_{c'} [EN_{cc'} - H_{cc'}]|\chi_{c'}\rangle. \tag{2}$$

The norm $N_{cc'}$ and energy $H_{cc'}$ kernels are, in general, very complicated as a result of the presence of the antisymmetizer $\mathscr{A}$ in the wave function. It is usual to split the contributions of $\mathscr{A}$ into two parts: the one called direct (D) comes from terms in $\mathscr{A}$ analogous to unity, and the other called exchange (E) comes from all other terms. The direct terms are relatively easy to calculate; while the exchange terms are very cumbersome. Expressing the Hamiltonian as the sum of intrinsic cluster Hamiltonians plus the relative kinetic energy K plus the intercluster interaction V_{12}, the Hill–Wheeler equation (2) is transformed into

$$[\varepsilon_c - K_c^{(D)}]|\chi_c\rangle = \sum_{c'} V_{cc'}(\varepsilon_{c'})|\chi_{c'}\rangle, \tag{3}$$

where $\varepsilon_c = E - E_{c_1} - E_{c_2}$ is the relative intercluster energy and

$$V_{cc'}(\varepsilon_{c'}) = K_{cc'}^{(E)} + V_{12_{cc'}}^{(D)} + V_{12_{cc'}}^{(E)} - \varepsilon_{c'}N_{cc'}^{(E)} \tag{4}$$

is some kind of effective potential, usually nonlocal. The form (3) is quite similar to a Schrödinger equation with coupled channels, although $|\chi_c\rangle$ does not represent a probability amplitude because of the nonorthogonality of the basis. Anyhow, from Eq. (3) one can use the well known methods of dealing with scattering problems. In particular, the Lippman–Schwinger equation relates the transition matrix T to the potential matrix V [in this case V is given by (4) and is energy dependent]:

$$T = V + VG_0^+T. \tag{5}$$

All the physical information concerning the scattering can be obtained from the transition operator T.

Usually the Lippman–Schwinger equation is solved in configuration or in momentum space in which the free propagator $G_0^+ = (E + i\varepsilon - K)^{-1}$ takes a particularly simple form. In that case, G_0^+ has a pole and this may lead to some numerical difficulties. However, the point is that, within the RGM framework, the exchange kernels are never calculated directly in the configuration or momentum space. Most of the time the cluster functions ψ_{c_1} and ψ_{c_2} are approximated by Gaussians, and the more natural way to calculate the kernels is to expand χ_c on peaked Gaussians.[4] Recently we proposed[5] an alternative method based on harmonic oscillator (HO) functions ϕ_{nl}. The advantage is that we are not limited to ground state clusters, but radial excitations can be considered as well. We proved in a previous paper[6] that the exchange kernels

$$H_{cc'}^{(E)}(n,n';l) = \langle\phi_{nl}|H_{cc'}^{(E)}|\phi_{n'l}\rangle$$

can always be calculated *exactly* in such a basis with the help of Brody–Moshinsky coefficients. Moreover, it can be shown in some cases, and it is expected on general grounds, that the exchange quantities are very rapidly convergent in the HO basis. In practice they are roughly zero once $n,n' \geqslant n_\alpha \cong 4$. In other words, the nonlocal exchange potentials are a very limited sum of separable (in HO basis) potentials. The only problem comes from $V_{12}^{(D)}$ in (4), which is, in general, local and cannot always be expanded on HO in a very convergent way. A possibility in that case is to use the propagator $(E + i\varepsilon - K - V_{12}^{(D)})^{-1}$ in the Lippman–Schwinger equation. In a number of interesting situations (electromagnetic interactions of two neutral objects, strong interaction of two colorless systems, etc). this fortunately will not be necessary since $V_{12}^{(D)}$ vanishes identically. Thus in the RGM a natural basis for evaluating the kernels is the harmonic oscillator one. In solving the Lippman–Schwinger equation, one must manage matrices of order $n_c \times n_\alpha$ (n_c is the number of channels), which is around a few tenth. Coming back to the coordinate or momentum space makes necessary (i) additional calculations to express the kernels from the HO basis to the new basis, and (ii) the discretization of the r or p axis into $n_p \cong 100$ points. This will result in dealing with matrices of order $n_c \times n_p$, which is typically of several hundred. From these remarks one sees that there is a great advantage to solving the Lippman–Schwinger equation in the HO basis directly.

III. MATRIX ELEMENTS OF THE FREE PROPAGATOR

But to perform such a program we absolutely need the expression of the free propagator G_0^+ in the harmonic oscillator basis. As far as we know this analytical expression was never published and the main topic of this paper is to derive it in a closed form.

Thus we are faced with the problem of evaluating

$$G^+_{0\,nn'l}(E,b) = \lim_{\varepsilon\to 0} \langle \varphi_{nlm}(b)|(E + i\varepsilon - K)^{-1}|\varphi_{n'lm}(b)\rangle. \qquad (6)$$

The kinetic energy operator $K = \mathbf{p}^2/2\mu$ is invariant under rotation and it is the same angular momentum lm that appears in the bra and in the ket. Let b be the size parameter for the harmonic oscillator wave function; it is more convenient to define the space vectors $\mathbf{r}$ in units of b, $\mathbf{x} = \mathbf{r}/b$, and the wave vectors $\mathbf{k}$ in units of b^{-1}, $\mathbf{q} = b\,\mathbf{k}$. The HO wave functions are defined in the configuration space with the usual Moshinsky[7] phase conventions:

$$\langle \mathbf{x}|nlm\rangle = \varphi_{nlm}(\mathbf{x}) = [U_{nl}(x)/x]\,Y_{lm}(\hat{x}),$$
$$U_{nl}(x) = \left[\sqrt{2(n!)/\Gamma(n+l+\tfrac{3}{2})}\right] \times x^{l+1} e^{-x^2/2} L_n^{l+1/2}(x^2), \qquad (7)$$
$$L_n^{l+1/2}(x^2) = \sum_{s=0}^{n} (-)^s \frac{\Gamma(n+l+\frac{3}{2})}{(n-s)!\,\Gamma(s+l+\frac{3}{2})} \frac{x^{2s}}{s!}$$

are the Laguerre polynomials.

In the momentum representation $|\mathbf{q}\rangle$ the HO wave functions [Fourier transforms of $\phi_{nlm}(\mathbf{x})$] have the same form as in the coordinate representation but differ by an important phase factor. Explicitly

$$\langle \mathbf{q}|nlm\rangle = (2\pi)^{-3/2} \int e^{-i\mathbf{q}\mathbf{x}} \varphi_{nlm}(\mathbf{x})\,d\mathbf{x} = (-i)^{2n+l} \varphi_{nlm}(\mathbf{q}). \qquad (8)$$

It is more convenient to use the momentum representation to evaluate the Green's function. For typographical reasons we will omit the (+) (outgoing propagator) and (0) (free propagator) superscript indices in all that follows having always in mind that we deal with the free outgoing propagator. One can write

$$G_{nn'l}(E,b) = (-1)^{n+n'}(2\mu b^2/\hbar^2) \times \lim_{\varepsilon\to 0} \int \varphi^*_{nlm}(\mathbf{q})\varphi_{n'lm}(\mathbf{q})[q_0^2 - q^2 + i\varepsilon]^{-1} d\mathbf{q}, \quad (9)$$

with $q_0^2 = 2\mu b^2 E/\hbar^2$. The integration over angles is trivial and we are left only with the radial integration

$$A = \lim_{\varepsilon\to 0} \int_0^\infty U_{nl}(q)U_{n'l}(q)[q_0^2 - q^2 + i\varepsilon]^{-1}\,dq. \quad (10)$$

Using the well known property

$$\lim_{\varepsilon\to 0}[q_0^2 - q^2 + i\varepsilon]^{-1} = \text{P.P.}[q_0^2 - q^2]^{-1} - i\pi\delta(q_0^2 - q^2), \qquad (11)$$

the Green's function can be split into two parts: the real part G^R coming from the principal part P.P. and the imaginary part G^I (present only for scattering problems $q_0^2 > 0$) coming from the delta function. The imaginary part is easy to calculate; the real part is more involved. One has to evaluate the following integral:

$$R = \text{P.P.} \int_0^\infty e^{-q^2} q^{2l+2} \times L_n^{l+1/2}(q^2) L_{n'}^{l+1/2}(q^2)[q^2 - q_0^2]^{-1}. \qquad (12)$$

Let us remark that $L_n^{l+1/2}(q^2)L_{n'}^{l+1/2}(q^2) = P_N(q^2)$ is a polynomial of degree $N = n + n'$. The trick is thus to isolate the singularity of (12) in a much simpler integral. The original integral can be decomposed into two terms R_1 and R_2. Explicitly

$$R = R_1 + R_2,$$

with

$$R_1 = \text{P.P.} \int_0^\infty e^{-q^2}[q^{2l+2}P_N(q^2) - q_0^{2l+2}P_N(q_0^2)] \times [q^2 - q_0^2]^{-1}\,dq \qquad (13)$$

and

$$R_2 = q_0^{2l+2}P_N(q_0^2)\,\text{P.P.}\int_0^\infty e^{-q^2}[q^2 - q_0^2]^{-1}\,dq.$$

In R_1 the pole disappears and the resulting integrals are standard. We are left with polynomials in q_0^2 whose coefficients b_k are purely geometrical and *can be computed and tabulated once and for all.*

The integral R_2 containing the pole is related to the complex error function erf(z).[8] After some manipulations, one gets

$$\text{P.P.} \int_0^\infty e^{-q^2}[q^2 - q_0^2]^{-1}\, dq$$

$$= \frac{i\pi}{2q_0} e^{-q_0^2} \operatorname{erf}(iq_0) = -\frac{\sqrt{\pi}}{q_0} F(q_0), \tag{14}$$

where

$$F(x) = e^{-x^2} \int_0^x e^{t^2}\, dt$$

is the real Dawson integral.

Now everything is complete. Let us summarize the results:

$$G_{nn'l}(E,b) = G^R_{nn'l}(E,b) + iG^I_{nn'l}(E,b), \tag{15}$$

$$G^I_{nn'l}(E,b) = (-1)^{n+n'+1} \frac{2\mu b^2}{\hbar^2} \frac{\pi}{2q_0} U_{nl}(q_0) U_{n'l}(q_0), \tag{16}$$

$$G^R_{nn'l}(E,b) = (-1)^{n+n'+1} \frac{2\mu b^2}{\hbar^2} \left\{ \frac{i\pi}{2q_0} U_{nl}(q_0) U_{n'l}(q_0) \operatorname{erf}(iq_0) + \sum_{k=0}^{l+n+n'} b_k(n,n',l) q_0^{2k} \right\}, \tag{17}$$

with

$$b_k(n,n',l) = 2^{2k+1-n-n'} \left(\frac{n!n'!(2n+2l+1)!(2n'+2l+1)!}{(n+l)!(n'+l)!} \right)^{1/2}$$

$$\times \sum_{p=0}^{n} \sum_{p'=0,\, p+p'>k-l+1}^{n'} (-1)^{p+p'} \frac{(2l+2p+2p'-2k)!(p+l)!(p'+l)!}{(p+p'+l-k)!(2p+2l+1)!(2p'+2l+1)!(n-p)!(n'-p')!p!p'!}. \tag{18}$$

The geometrical coefficients $b_k(n,n',l)$ fulfill the relations

$$b_k(n,n',l) = b_k(n',n,l),$$

$$b_k(n,n',l) = 0, \quad \text{if } k > n + n' + l,$$

and q_0 is given by (9).

IV. CONCLUSION

The expressions (16) and (17) give the free propagator in the HO basis; the pole has been eliminated analytically and this is a great advantage. This procedure allows us to solve the Lippman–Schwinger equation in the HO basis; this will be very well suited for scattering equations of Hill–Wheeler type with kernels rapidly converging in the harmonic oscillator basis. In fact, this method was applied with success in the description of nucleon–nucleon interaction in terms of quarks,[9] of $QQ\bar{q}\bar{q}$ multiquarks,[10] or of the dilambda system.[11]

[1] J. A. Wheeler, Phys. Rev. **52**, 1083 (1937).

[2] B. Giraud, J. Hocquenghem, and A. Lumbroso, Phys. Rev. C **7**, 2274 (1974); T. Matsuse, M. Kamimura, and Y. Fukushima, Prog. Theor. Phys. **53**, 706 (1975).

[3] D. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).

[4] M. Oka and K. Yazaki, *Quarks and Nuclei*, edited by W. Weise in Int. Rev. Nucl. Phys. (World Scientific, Singapore) **1**, 489 (1984); A. Faessler, F. Fernandez, G. Lübeck, and K. Shimizu, Phys. Lett. B **112**, 201 (1982).

[5] C. Gignoux and B. Silvestre-Brac, Int. Rep. ISN 86-42, Grenoble.

[6] B. Silvestre-Brac, J. Phys. **46**, 1087 (1985).

[7] M. Moshinsky, Nucl. Phys. **13**, 104 (1959).

[8] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[9] B. Silvestre-Brac, J. Carbonell, and C. Gignoux, Phys. Lett. B **179**, 9 (1986).

[10] S. Zouzou, B. Silvestre-Brac, C. Gignoux, and J. M. Richard, Z. Phys. C **30**, 457 (1986).

[11] B. Silvestre-Brac, J. Carbonell, and C. Gignoux, Phys. Rev. D **36**, 2083 (1987).

Asymptotic transition probabilities

J. H. Arredondo R.[a)]
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autonoma de México, Apdo. Postal 20-726, México, D.F., Mexico

(Received 17 June 1987; accepted for publication 3 May 1989)

The impact parameter model for the scattering of two heavy particles and a light one is studied. The asymptotic behavior of the transition probability is studied when the relative velocity of the heavy particles goes to zero. In particular, rigorous proof is given from first principles, within the context of the model, of the validity of Massey's criterion.

I. INTRODUCTION

In this paper we consider the impact parameter model for the scattering of a light particle and two heavy ones. In this model one assumes that the heavy particles are infinitely massive and that their motion along a classical trajectory is not affected by the light particle. In this approximation the influence of the heavy particles on the light one is represented by time-dependent potentials. We will neglect magnetic effects and will take the potentials of interaction between the light particle and the heavy ones to be separable.

For any positive integer m, let $L^2(R^m)$ be the Hilbert space consisting of all complex valued Lebesgue measurable square integrable functions on R^m. For $t\in R^1$, $H(t)\equiv H_v(t)$ is the following self-adjoint operator[1] in $L^2(R^m)$ with domain $H_2(R^m)$ (see Ref. 2), the Sobolev space of order 2:

$$H_v(t) = -\tfrac{1}{2}\Delta - \lambda_1 V - \lambda_2 V_\rho , \tag{1.1}$$

where Δ is the Laplace operator,[3]

$$\Delta = \sum_{i=1}^{m} \frac{\partial^2}{\partial x_i^2},$$

with derivatives in the distribution sense. Also, λ_1 and λ_2 are real, $\lambda_1 > \lambda_2 > 0$, and V is the rank 1 operator,

$$V\varphi = g(g,\varphi) , \tag{1.2}$$

for all $\varphi\in L^2(R^m)$, and g a fixed function in $L^2(R^m)$, $\|g\| = 1$, with $(\cdot,\cdot)$ denoting the scalar product in $L^2(R^m)$ antilinear on the factor on the left. Moreover, for $g\in L^2(R^m)$, we define

$$V_\rho \phi = g_\rho (g_\rho,\phi) , \tag{1.3}$$

with

$$g_\rho(\mathbf{x}) = g(\mathbf{x} - \rho(t)) . \tag{1.4}$$

Here $\rho(t)$ is a function from R^1 into R^m, two times continuously differentiable and such that, for some $t_\pm > 0$,

$$|[\rho(t) - \mathbf{c}_\pm - \mathbf{v}_\pm t]| \leqslant C_0/|t|^\alpha, \quad \forall \pm t > t_\pm ,$$
$$\left|\frac{d^k}{dt^k}[\rho(t) - \mathbf{c}_\pm - \mathbf{v}_\pm t]\right| \leqslant C_k v^k, \quad \forall \pm t > t_\pm \text{ and } k = 1,2 . \tag{1.5}$$

Here $\alpha > 0$, and $\mathbf{c}_\pm$ and $\mathbf{v}_\pm$ are vectors in R^m with $\mathbf{v}_\pm \neq 0$. Furthermore,

$$0 < v \equiv |\mathbf{v}_+| = |\mathbf{v}_-| \leqslant 1 \tag{1.6}$$

and

$$\left|\frac{d^k}{dt^k}\rho(t)\right| \leqslant v^k, \quad -t_- \leqslant t \leqslant t_+ \text{ and } k = 1,2 . \tag{1.7}$$

We will assume that for some fixed positive constant M, and B being any of $t_\pm$, $|\mathbf{c}_\pm|$, or C_k, $k = 0,1,2$, one has

$$B \leqslant M . \tag{1.8}$$

The state vector of the light particle satisfies the Schrödinger equation[4]

$$i\frac{\partial}{\partial t}\Psi(t) = H_v(t)\Psi(t) . \tag{1.9}$$

The solution to (1.9) is given by a two parameter family of unitary operators[5]

$$U_v(t,s), \quad t,s\in R^1 . \tag{1.10}$$

Let H_i, $i = 1,2$, be the following self-adjoint operators with domain $H_2(R^m)$:

$$H_i = -\tfrac{1}{2}\Delta - \lambda_i V . \tag{1.11}$$

Since V is rank 1, from the essential spectrum Weyl's theorem,[6] the essential spectrum of H_i, $i = 1,2$, is $[0, +\infty)$. It follows by explicit calculation that either $H_i \geqslant 0$ or H_i has at most one negative eigenvalue $-Q_i$ with multiplicity 1. We assume in this paper that this negative eigenvalue $-Q_i$ exists for H_i, $i = 1,2$. Note that since $\lambda_1 > \lambda_2$ it is enough to make the assumption for H_2, and that for given g one can find such a λ_2.

Let P_1 be the orthogonal projector onto the nondegenerate ground state of H_1 (i.e., the orthogonal projector onto the subspace generated by the eigenvector for the negative eigenvalue corresponding to the minimum of the spectrum of H_1). The existence of the Möller wave operators

$$\Omega_\pm(v) = \text{s-lim}_{t\to\pm\infty} U_v(0,t)e^{-itH_1}P_1 \tag{1.12}$$

is proved in Howland,[7] Yajima,[8] and Hagedorn[9] in the case of local potentials. We give in Lemma II.3 the simple proof in our case of rank 1 interactions (see, also, Arredondo[10]).

Let ϕ, $\|\phi\| = 1$, be a ground state of H_1, and assume that initially the light particles is in the state given by ϕ. Then the probability that after the interaction the light particle remains in the state ϕ is given by[6,11]

$$P(v) = |(\phi,\Omega_+(v)^*\Omega_-(v)\phi)|^2$$
$$= |(\Omega_+(v)\phi,\Omega_-(v)\phi)|^2 . \tag{1.13}$$

We denote by F the Fourier transform as an unitary

[a)] Present address: Departamento de Matemáticas, Universidad Autonoma Metropolitana, Av. Michoacán y la Purísima, Iztapalapa, C.P. 0934 Apdo. Postal 55-534, México, D.F., Mexico.

 0022-2488/89/102291-06$02.50

operator in $L^2(R^m)$. We have proved the following theorem.

Theorem I: Let $H_v(t)$ be defined by (1.1)–(1.8) with g and Fg in $H_2(R^m)$. Then

$$|P(v)-1| = O(v), \quad v\to 0. \tag{1.14}$$

Here v is defined by formula (1.6).

The well known fundamental experimental criterion of Massey[12,13] states that transitions between states of different energy are improbable if $\hbar v/d \ll 1$, where d is the range of the interaction and $v=|\mathbf{v}|$ is the relative velocity between the particles in the scattering process.

Theorem I gives, apparently for the first time, a rigorous theoretical proof from first principles of the validity of Massey's criterion.

II. PROOF OF THEOREM I

Lemma II.1: Let g and ρ be as in Theorem I. Then if $\lambda_1>\lambda_2>0$, the ground state energy, $-E(t)$, of $H_v(t)$ is nondegenerate and lies for all $t\in R^1$ in the interval $[-E_0, -Q_1]$, where $-E_0$ is the ground state energy of the self-adjoint operator $-\frac{1}{2}\Delta-(\lambda_1+\lambda_2)V$. Furthermore, $-E(t)$ is a twice continuously differentiable function of t. Moreover, there can exist other nonpositive eigenvalues only in the interval $[-Q_2,0]$. The eigenvector for $-E(t)$ can be taken as

$$\Psi(t)=F^{-1}Q_t^{-1}(p^2/2+E(t))^{-1}(c(t)\tilde{g}_\rho+\tilde{g}), \tag{2.1}$$

with

$$c(t)=\frac{\lambda_2(\tilde{g}_\rho,\tilde{g}(p^2/2+E(t))^{-1})}{1-\lambda_2\|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2}, \tag{2.2}$$

and Q_t^{-1} a normalization factor in order that $\|\Psi(t)\|=1$. Here F^{-1} denotes the inverse of the Fourier transform function $\tilde{}$. Moreover,

$$\frac{1}{\lambda_1+\lambda_2}\leqslant\left|\left|\tilde{g}\left(\frac{p^2}{2}+E(t)\right)^{-1/2}\right|\right|^2\leqslant\frac{1}{\lambda_1}, \quad \forall t\in R^1, \tag{2.3}$$

and

$$Q_t\equiv\left|\left|\left(\frac{p^2}{2}+E(t)\right)^{-1}(c(t)\tilde{g}_\rho+\tilde{g})\right|\right|$$
$$\geqslant\left(1-\frac{\lambda_2}{\lambda_1}\right)\left|\left|\left(\frac{p^2}{2}+E_0\right)^{-1}\tilde{g}\right|\right|, \quad \forall t\in R^1. \tag{2.4}$$

Proof: Let $\Psi(t)$ be an eigenvector with eigenvalue $-E(t)$ and $E(t)>0$:

$$H_v(t)\Psi(t)=-E(t)\Psi(t). \tag{2.5}$$

If we apply the Fourier transform[5] F on both sides of Eq. (2.5) and take the scalar product with $\tilde{g}=Fg$ and $\tilde{g}_\rho=Fg_\rho$, we get, in matrix notation,

$$\begin{bmatrix}1-\lambda_1\|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2 & -\lambda_2(\tilde{g},\tilde{g}_\rho(p^2/2+E(t))^{-1})\\ -\lambda_1(\tilde{g}_\rho,\tilde{g}(p^2/2+E(t))^{-1}) & 1-\lambda_2\|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2\end{bmatrix}\begin{bmatrix}(g,\Psi)\\(g_\rho,\Psi)\end{bmatrix}=0. \tag{2.6}$$

The equation has a nontrivial solution if and only if

$$\left(1-\lambda_2\left|\left|\tilde{g}\left(\frac{p^2}{2}+E(t)\right)^{-1/2}\right|\right|^2\right)\left(1-\lambda_1\left|\left|\tilde{g}\left(\frac{p^2}{2}+E(t)\right)^{-1/2}\right|\right|^2\right)=\lambda_1\lambda_2\left|\left(\tilde{g},\tilde{g}_\rho\left(\frac{p^2}{2}+E(t)\right)^{-1}\right)\right|^2. \tag{2.7}$$

We show now that the ground state $-E(t)$ satisfies $-E(t)\leqslant -Q_1<-Q_2$. Note that $-Q_1$ is strictly negative.

Let us take a fixed $E>0$ independent of t. We define

$$D_E(t)=(\lambda_1+\lambda_2)a+\lambda_1\lambda_2(f(t)-a^2), \tag{2.8}$$

with

$$a=\|\tilde{g}(p^2/2+E)^{-1/2}\|^2 \tag{2.9}$$

and

$$f(t)=|(\tilde{g},\tilde{g}_\rho(p^2/2+E)^{-1})|^2. \tag{2.10}$$

We have in this case

$$\sup_{-\infty<t<\infty} D_E(t)\leqslant(\lambda_1+\lambda_2)a, \tag{2.11}$$

$$\inf_{-\infty<t<\infty} D_E(t)=D_E(\pm\infty)=(\lambda_1+\lambda_2)a-\lambda_1\lambda_2a^2. \tag{2.12}$$

Note that in (2.11) the equality will be fulfilled if and only if $\rho(t)$ is a curve through the origin. Then $-E$ is an eigenvalue for $H_v(t)$ if and only if

$$D_E(\pm\infty)\leqslant 1\leqslant\sup_{-\infty<t<+\infty}D_E(t)\leqslant(\lambda_1+\lambda_2)a. \tag{2.13}$$

Furthermore $(\lambda_1+\lambda_2)a-\lambda_1\lambda_2a^2=1$ if and only if $a=1/\lambda_1$ or $a=1/\lambda_2$. We also have

$$\sup_a(\lambda_1+\lambda_2)a-\lambda_1\lambda_2a^2$$
$$=(\lambda_1+\lambda_2)a-\lambda_1\lambda_2a^2|_{a=(\lambda_1+\lambda_2)/2\lambda_1\lambda_2}$$
$$=(\lambda_1+\lambda_2)^2/4\lambda_1\lambda_2>1. \tag{2.14}$$

Then if $-E<0$ is an eigenvalue of $H_v(t)$ for some $t\in R^1$, either

$$1/(\lambda_1+\lambda_2)\leqslant a\leqslant 1/\lambda_1 \quad\text{or}\quad a\geqslant 1/\lambda_2. \tag{2.15}$$

Since a is a monotone decreasing function of E, we obtain that either $-E_0\leqslant -E(t)\leqslant -Q_1$ or $-Q_2\leqslant -E(t)<0$, where $-E_0$ is the ground state of $-\frac{1}{2}\Delta-(\lambda_1+\lambda_2)V$. Note that since $\lambda_1>\lambda_2$ then $-Q_1<-Q_2$, and that by explicit calculation

$$\|\tilde{g}(p^2/2+Q_1)^{-1/2}\|^2=1/\lambda_1, \tag{2.16}$$

$$\|\tilde{g}(p^2/2+E_0)^{-1/2}\|^2=1/(\lambda_1+\lambda_2). \tag{2.17}$$

From (2.9) and (2.15)–(2.17) one obtains (2.3).

We prove now that there is only one eigenvalue of multiplicity 1 in $[-E_0, -Q_1]$. Let $h(t)$ be a continuous function from R^1 into R^m such that $h(t)\equiv\rho(t)$ if $|t|>t_0$, for some $t_0>0$ and $h(0)=\mathbf{0}$. Now consider $\overline{H}_v(t)$ defined as $H_v(t)$

with $h(t)$ instead of $\rho(t)$. We prove by explicit calculation that

$$\dim \overline{P}_{[-E_0,-Q_1]}(0) = 1\,, \tag{2.18}$$

where $\overline{P}_A(t)$ is the spectral projector of $\overline{H}_v(t)$ associated with the set A. By a continuity argument as in Arredondo,[14] we can prove that

$$\dim \overline{P}_{[-E_0,-Q_1]}(t) = 1\,, \tag{2.19}$$

for every $t \in R^1$. Consider again $H_v(t)$ as in (1.1) with our original $\rho(t)$. From (2.19) we know that, for $|t| > t_0$,

$$\dim P_{[-E_0,-Q_1]}(t) = \dim \overline{P}_{[-E_0,-Q_1]}(t) = 1, \tag{2.20}$$

where $P_A(t)$ denotes the spectral projector of $H_v(t)$ associated with A. Using the same continuity argument as in Arredondo,[14] one proves that (2.20) is true for all $t \in R^1$. Then,

$$\dim P_{[-E_0,-Q_1]}(t) = 1, \quad \text{for all } t \in R^1\,. \tag{2.21}$$

Moreover, the eigenvalue corresponding to the ground state, $-E(t)$, is two times continuously differentiable by standard perturbation theory.[15]

It follows by (2.7) and (2.16) that $E(t)$ must satisfy

$$E(t) - Q_1 = \frac{|(\tilde{g},\tilde{g}_\rho(p^2/2+E(t))^{-1})|^2}{(1/\lambda_2 - \|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2)\|\tilde{g}(p^2/2+E(t))^{-1/2}(p^2/2+Q_1)^{-1/2}\|^2}\,. \tag{2.22}$$

Note that since for the ground state eigenvalue, $1/(\lambda_1+\lambda_2) \leqslant a \leqslant 1/\lambda_1$, we conclude that

$$C_1 \leqslant 1/\lambda_2 - \|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2 \leqslant C_2\,, \tag{2.23}$$

for some $C_1, C_2 > 0$, and every $t \in R^1$.

By (2.6) we see that

$$(\tilde{g}_\rho,\tilde{\Psi}) = \frac{\lambda_1(g,\Psi)(\tilde{g}_\rho,\tilde{g}(p^2/2+E(t))^{-1})}{1-\lambda_2\|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2}\,. \tag{2.24}$$

Taking the Fourier transform in $H_v(t)\Psi(t) = -E(t)\Psi(t)$ we obtain (2.1) and (2.2) from (2.24). We now prove (2.4). From (2.2), (2.3), and (2.7) one obtains

$$|c(t)| = \left(\frac{1/\lambda_1 - \|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2}{1/\lambda_2 - \|\tilde{g}(p^2/2+E(t))^{-1/2}\|^2}\right)^{1/2} \leqslant \lambda_2/\lambda_1 < 1\,.$$

From this one obtains (2.4) easily. This proves the lemma.□

Lemma II.2: Let g and ρ be as in Theorem I. Let $\Psi(t)$ and $-E(t)$ be the ground state and its corresponding eigenvalue of the time-dependent self-adjoint operator $H_v(t)$. Then one has the following estimations:

$$\left|\frac{d^k}{dt^k}E\left(\frac{t}{v}\right)\right| \leqslant \frac{C_M}{t^2}, \quad \left\|\frac{d^k}{dt^k}\Psi\left(\frac{t}{v}\right)\right\| \leqslant \frac{C_M}{t^2}, \quad \forall |t| > T_M \text{ and } k = 1,2\,, \tag{2.25}$$

$$\sup_{t\in R^1}\left|\frac{d^k}{dt^k}E\left(\frac{t}{v}\right)\right| \leqslant C_M, \quad \sup_{t\in R^1}\left\|\frac{d^k}{dt^k}\Psi\left(\frac{t}{v}\right)\right\| \leqslant C_M, \quad k=1,2\,. \tag{2.26}$$

Here C_M and T_M are positive constants depending only on M as given in (1.8).

Proof: Statement (2.26) follows easily by noting (1.5)–(1.7) and the formulas given below. We prove only (2.25).

We want to show at first that

$$\left|\frac{d}{dt}E\left(\frac{t}{v}\right)\right| \leqslant \frac{C_M}{t^2}, \quad |t| > T_M\,. \tag{2.27}$$

Without loss of generality we can assume that $\rho(t)$ is of the form

$$\rho(t) \equiv (\rho_1(t),\underbrace{0,0,0,\ldots,0}_{m-1 \text{ times}})\,. \tag{2.28}$$

We note that

$$|\rho_1(t/v)| \geqslant \tfrac{1}{2}\cdot|t|, \quad |t| > T_M\,,$$
$$\left|\frac{d^k}{dt^k}\rho\left(\frac{t}{v}\right)\right| \leqslant C_{1,M}, \quad \forall t \in R^1 \text{ and } k = 1,2. \tag{2.29}$$

One can see that

$$-\frac{1}{v}\cdot E'\left(\frac{t}{v}\right) \equiv -\frac{d}{dt}E\left(\frac{t}{v}\right) = v^{-1}\left(\Psi\left(\frac{t}{v}\right),H'_v\left(\frac{t}{v}\right)\Psi\left(\frac{t}{v}\right)\right). \tag{2.30}$$

From (2.30) we obtain the expression for $-(d/dt)E(t/v)$:

$$-\lambda_2 i v^{-1}\left[\left(\tilde{g}_\rho\left(\rho'\left(\frac{t}{v}\right)\cdot\mathbf{p}\right),\tilde{\Psi}\left(\frac{t}{v}\right)\right)\left(\tilde{\Psi}\left(\frac{t}{v}\right),\tilde{g}_\rho\right) - \left(\tilde{g}_\rho,\tilde{\Psi}\left(\frac{t}{v}\right)\right)\left(\tilde{\Psi}\left(\frac{t}{v}\right),\tilde{g}_\rho\left(\rho'\left(\frac{t}{v}\right)\cdot\mathbf{p}\right)\right)\right], \tag{2.31}$$

where $\cdot$ denotes the dot product in R^m. From (2.29) and (2.1)–(2.4) one has

$$v^{-1}\left|\left(\tilde{g}_\rho\left(\rho'\left(\frac{t}{v}\right)\cdot\mathbf{p}\right),\tilde{\Psi}\left(\frac{t}{v}\right)\right)\right| \leqslant d_1\left|c\left(\frac{t}{v}\right)\right| + d_2\left|v^{-1}\int e^{i\rho(t/v)\cdot\mathbf{p}}\left(\rho'\left(\frac{t}{v}\right)\cdot\mathbf{p}\right)|\tilde{g}(\mathbf{p})|^2\left(\frac{p^2}{2}+E\left(\frac{t}{v}\right)\right)^{-1}d^m p\right|\,. \tag{2.32}$$

We estimate now the second term on the right side of (2.32). This term is equal to

$$\frac{d_2}{\rho_1^2(t/v)}\left|v^{-1}\rho_1'\left(\frac{t}{v}\right)\int\frac{\partial^2}{\partial p_1^2}e^{i\rho(t/v)\cdot\mathbf{p}}p_1|\tilde{g}(\mathbf{p})|^2\left(\frac{p^2}{2}+E\left(\frac{t}{v}\right)\right)^{-1}d^m p\right|$$
$$= \frac{1}{\rho_1^2(t/v)}\left|v^{-1}\rho_1'\left(\frac{t}{v}\right)\int e^{i\rho(t/v)\cdot\mathbf{p}}\frac{\partial^2}{\partial p_1^2}p_1|\tilde{g}(\mathbf{p})|^2\left(\frac{p^2}{2}+E\left(\frac{t}{v}\right)\right)^{-1}d^m p\right| \leqslant \frac{d_3}{t^2}, \quad |t| > T_M, \tag{2.33}$$

where we have used (2.29) and the hypothesis on g.[16,17] Similarly, one can prove by using (2.2) and (2.3) that

$$|c(t/v)|\leqslant C_M/t^2, \quad |t|>T_M . \tag{2.34}$$

One obtains (2.27) from (2.30)–(2.34). We now prove (2.25) for $(d/dt)\Psi(t/v)$. From (2.1)–(2.4) one can see that

$$\frac{d}{dt}\Psi\Big(\frac{t}{v}\Big)=F^{-1}\Big[\frac{d}{dt}Q_{t/v}^{-1}\cdot G\Big(\frac{t}{v}\Big)+Q_{t/v}^{-1}\frac{d}{dt}G\Big(\frac{t}{v}\Big)\Big].$$

Here F^{-1} denotes the inverse of the Fourier transform function and

$$G(t)\equiv(p^2/2=E(t))^{-1}(c(t)\tilde{g}_p+\tilde{g}).$$

The derivatives of $Q_{t/v}^{-1}$ behave for $|t|>T_M$ as the derivatives of $E(t/v)$ and of $c(t/v)$. Here one uses (2.4). The L^2 norm of the derivatives of $G(t/v)$ also behaves in this manner. By using (2.3) and an estimation as in (2.33) one can see that the first derivative of $c(t/v)$ behaves as $O(1/t^2)$. Since we have already seen that $(d/dt)E(t/v)$ is of this order, we obtain (2.25) for $k=1$. Now we note that

$$-\frac{d^2}{dt^2}E\Big(\frac{t}{v}\Big)=\Big(\Psi\Big(\frac{t}{v}\Big),\Big[\frac{d^2}{dt^2}H_v\Big(\frac{t}{v}\Big)\Big]\Psi\Big(\frac{t}{v}\Big)\Big)+\Big(\frac{d}{dt}\Psi\Big(\frac{t}{v}\Big),\Big[\frac{d}{dt}H_v\Big(\frac{t}{v}\Big)\Big]\Psi\Big(\frac{t}{v}\Big)\Big)+\Big(\Psi\Big(\frac{t}{v}\Big),\Big[\frac{d}{dt}H_v\Big(\frac{t}{v}\Big)\Big]\frac{d}{dt}\Psi\Big(\frac{t}{v}\Big)\Big). \tag{2.35}$$

From (2.29) and the hypothesis on the function g we get

$$\Big|\Big|\frac{d^q}{dt^q}H_v\Big(\frac{t}{v}\Big)\Big|\Big|\leqslant C, \quad q=1,2. \tag{2.36}$$

From (2.36) and the just proved estimation for $(d/dt)\Psi(t/v)$ we obtain that the last two terms on the right side of Eq. (2.35) are of order $1/t^2$. To estimate the first term on the right side of (2.35) one can proceed as above: we note that

$$\frac{d^2}{dt^2}H_v(t)=F^{-1}\sum_{r+s=2}\Big(\frac{d^r}{dt^r}e^{-i\rho(t)\cdot\mathbf{p}}\tilde{g},\cdot\Big)\frac{d^s}{dt^s}e^{-i\rho(t)\cdot\mathbf{p}}\tilde{g}F. \tag{2.37}$$

Here F denotes the Fourier transform. Furthermore, if $r=1,2$,

$$\frac{d^r}{dt^r}e^{-i\rho(t)\cdot\mathbf{p}}=e^{-i\rho(t)\cdot\mathbf{p}}\sum_{\beta_1+2\beta_2=r}(i)^{\alpha_{\beta_1\beta_2}}d_{\beta_1\beta_2}(\rho'\cdot\mathbf{p})^{\beta_1}(\rho''\cdot\mathbf{p})^{\beta_2}. \tag{2.38}$$

Here β_i, $i=1,2$ are non-negative integers. Therefore, as in (2.32)–(2.34) with (2.1)–(2.4) and (2.29) we get

$$\Big|\Big(\tilde{g}\Big(\frac{d^r}{dt^r}e^{-i\rho(t/v)\cdot\mathbf{p}}\Big),\tilde{\Psi}\Big(\frac{t}{v}\Big)\Big)\Big|$$
$$\leqslant d_1\Big|c\Big(\frac{t}{v}\Big)\Big|+\frac{d_2}{t^2}\Big|\int e^{i\rho(t/v)\cdot\mathbf{p}}\frac{\partial^2}{\partial p_1^2}e^{-i\rho(t/v)\cdot\mathbf{p}}\Big(\frac{d^r}{dt^r}e^{i\rho(t/v)\cdot\mathbf{p}}\Big)|\tilde{g}(\mathbf{p})|^2\Big(\frac{p^2}{2}+E\frac{t}{v}\Big)^{-1}d^m p\Big|\leqslant\frac{d}{t^2}, \quad |t|>T_M, \tag{2.39}$$

where we have used (1.5)–(1.7), (2.38), and the hypothesis on the function g. By using (2.35)–(2.39) we then prove that $(d^2/dt^2)E(t/v)$ decays as $O(1/t^2)$, for all $|t|>T_M$. A very similar argument as the one given for the first derivative of $\Psi(t/v)$ can be used to prove that $d^2/dt^2\Psi(t/v)$ is of order $1/t^2$. This proves the lemma. □

We prove some equalities before we state some results on the adiabatic theorem.[18]

Let $P(t)$ be the projector onto the subspace generated by the ground state $-E(t)$ for $H_v(t)$. The operator differential equation

$$\dot{X}(t)=iA(t)X(t), \tag{2.40}$$

where

$$A(t)\equiv-i(\dot{P}(t)P(t)-P(t)\dot{P}(t)), \tag{2.41}$$

and with initial value

$$X(T)=P(T), \quad T\in R, \tag{2.42}$$

has the unique solution given by

$$U_A(t,T)P(T). \tag{2.43}$$

Here $U_A(t,s)$ is the two parameter family of unitary operators[5] that solves (1.9) with $-A(t)$ instead of $H_v(t)$. From the equality $P^2(t)=P(t)$ one obtains

$$P\dot{P}P=0, \quad \dot{P}=i(AP-PA).$$

It follows then that $W(t)=P(t)U_A(t,T)$ is also a solution to (2.40)–(2.42). This is only possible if

$$W(t)\equiv P(t)U_A(t,T)=U_A(t,T)P(T). \tag{2.44}$$

We take

$$\overline{\Psi}(t)=e^{-i\beta(t)}\Psi(t), \tag{2.45}$$

$$\beta(t)=-\operatorname{Im}\int_{-\infty}^{t}(\dot{\Psi}(s),\Psi(s))ds. \tag{2.46}$$

It follows that

$$(\dot{\overline{\Psi}},\overline{\Psi})=0. \tag{2.47}$$

We get from (2.40)–(2.47) that $W(t)\overline{\Psi}(T)$ and $\overline{\Psi}(t)$ satisfy the following initial value vector differential equation:

$$\dot{\varphi}(t)=\dot{P}(t)\varphi(t), \quad \varphi(T)=\overline{\Psi}(T).$$

Therefore, we must have

$$W(t)\overline{\Psi}(T)=\overline{\Psi}(t), \quad \forall T,t\in R. \tag{2.48}$$

Now let the time-dependent operator $H_{v/\gamma}(t)$ be defined with (1.1) by

$$H_{v/\gamma}(t)=H_v(t/\gamma) \tag{2.49}$$

and

$$U_{v/\gamma}(t,s),\quad t,s\in R^1, \tag{2.50}$$

and two parameter family of unitary operators associated with $H_{v/\gamma}(t)$ by (1.9). We take the adiabatic transformation for $t,s\in R^1$ and $\gamma\neq 0$,[19–26]

$$\Pi_\gamma(t,s)=\exp\left\{-i\gamma\int_s^t E(s)ds\right\}U_{v/\gamma}(\gamma t,\gamma s)\,. \tag{2.51}$$

We also define

$$\Upsilon(t)=(H_v(t)+E(t))^{-1}(1-P(t))\,. \tag{2.52}$$

By using (2.40)–(2.44) and (2.51) one obtains

$$\frac{d}{ds}(\Pi_\gamma(T,s)W(s))=\Pi_\gamma(T,s)\dot W(s)\,. \tag{2.53}$$

We take $T\equiv-\tau$ in this identity and we apply it to $\overline{\Psi}(-\tau)$. Integrating then from $-\tau$ to τ and noting (2.48) we get

$$\Pi_\gamma(-\tau,\tau)\overline{\Psi}(\tau)-\overline{\Psi}(-\tau)=\int_{-\tau}^{\tau}\Pi_\gamma(-\tau,s)\dot{\overline{\Psi}}(s)ds\,. \tag{2.54}$$

Now we integrate this last integral by parts,[19,21–23]

$$\begin{aligned}\Pi_\gamma(-\tau,\tau)\overline{\Psi}(\tau)-\overline{\Psi}(-\tau)&=(i\gamma)^{-1}\int_{-\tau}^{\tau}\left[\frac{d}{ds}\Pi_\gamma(-\tau,s)\right]\Upsilon(s)\dot{\overline{\Psi}}(s)ds\\&=(i\gamma)^{-1}\Pi_\gamma(-\tau,\cdot)\Upsilon(\cdot)\dot{\overline{\Psi}}(\cdot)|_{-\tau}^{\tau}\\&\quad+(-i\gamma)^{-1}\int_{-\tau}^{\tau}\Pi_\gamma(-\tau,s)\frac{d}{ds}[\Upsilon\dot{\overline{\Psi}}](s)ds.\end{aligned} \tag{2.55}$$

We make the change of variable $s=r/v$ and take $\gamma\equiv 1$, $\tau\equiv t/v$. From (2.51) and (2.55) we get

$$\begin{aligned}\limsup_{t\to+\infty}&\left\|U_v\left(\frac{t}{v},-\frac{t}{v}\right)\overline{\Psi}\left(-\frac{t}{v}\right)-\exp\left(iv^{-1}\int_{-t}^{t}E\left(\frac{q}{v}\right)dq\right)\overline{\Psi}\left(\frac{t}{v}\right)\right\|\\&\leqslant\limsup_{t\to+\infty}v\int_{-t}^{t}dr\left\|\frac{d}{dr}\left[\Upsilon\left(\frac{r}{v}\right)\frac{d}{dr}\overline{\Psi}\left(\frac{r}{v}\right)\right]\right\|\\&=O(v).\end{aligned} \tag{2.56}$$

Here we have used Lemma II.2. Furthermore, for the derivative of Υ one can see that

$$\begin{aligned}\left\|\frac{d}{dr}\Upsilon\left(\frac{r}{v}\right)\right\|&\equiv\left\|\Upsilon\left(\frac{d}{dr}H_v\left(\frac{r}{v}\right)+\frac{d}{dr}E\left(\frac{r}{v}\right)\right)\Upsilon\left(\frac{r}{v}\right)\right.\\&\quad\left.-\Upsilon\left(\frac{r}{v}\right)\frac{d}{dr}P\left(\frac{r}{v}\right)-\frac{d}{dr}P\left(\frac{r}{v}\right)\Upsilon\left(\frac{r}{v}\right)\right\|\\&\leqslant C\cdot(Q_1-Q_2)^{-1},\end{aligned} \tag{2.57}$$

for some positive constant C. Here we have used Lemma II.1, Lemma II.2, and (2.36). Therefore (2.56) is proved.

Lemma II.3: The Möller wave operators

$$\Omega_\pm(v)=\operatorname*{s-lim}_{t\to\pm\infty}U_v(0,t)e^{-itH_1}P_1$$

exist.

Proof: From (1.9),

$$\frac{d}{dt}U_v(0,t)e^{-itH_1}\phi=iU_v(0,t)g_\rho(g_\rho,\phi)e^{itQ_1}, \tag{2.58}$$

where ϕ is the normalized ground state of H_1:

$$H_1\phi=-Q_1\phi,$$

$$\phi=\frac{1}{\|\tilde g(p^2/2+Q_1)^{-1}\|}\left(-\frac{\Delta}{2}+Q_1\right)^{-1}g\,.$$

It follows, for $|r|>|\tau|>T_M>0$, that

$$\begin{aligned}\|[U_v(0,r)e^{-irH_1}&-U_v(0,\tau)e^{-i\tau H_1}]\phi\|\\&\leqslant\frac{1}{\|\tilde g(p^2/2+Q_1)^{-1}\|}\times\int_\tau^r ds\left|\left(\tilde g_\rho,\tilde g\left(\frac{p^2}{2}+Q_1\right)^{-1}\right)\right|\\&\leqslant C|\mathbf{v}|^{-2}\cdot(|\tau|^{-1}-|r|^{-1}),\end{aligned} \tag{2.59}$$

by the hypothesis on the function g. Then the left side of this last equation tends to zero as $\tau,r\to\pm\infty$. This proves the lemma. □

Proof of (1.14): Letting $r\to\pm\infty$ in (2.59) we obtain

$$\begin{aligned}\|[\Omega_\pm(v)-U_v(0,\tau)e^{-i\tau H_1}]\phi\|&\leqslant\frac{1}{\|\tilde g(p^2/2+Q_1)^{-1}\|}\times\int_\tau^{\pm\infty}ds\left|\left(\tilde g_\rho,\tilde g\left(\frac{p^2}{2}+Q_1\right)^{-1}\right)\right|\\&\leqslant C|\mathbf{v}|^{-2}|\tau|^{-1}.\end{aligned} \tag{2.60}$$

By using (2.60), we get

$$\begin{aligned}\big|\,|(\Omega_+^*(v)\Omega_-(v)\phi,\phi)|-|(\phi,\phi)|\,\big|&\leqslant\limsup_{t\to+\infty}\left\{\left|(\Omega_-(v)\phi,\Omega_+(v)\phi)-\left(U_v\left(0,-\frac{t}{v}\right)e^{itv^{-1}H_1}\phi,U_v\left(0,\frac{t}{v}\right)e^{-itv^{-1}H_1}\phi\right)\right|\right.\\&\quad+\left|\left[\left(U_v\left(\frac{t}{v},-\frac{t}{v}\right)\phi,\phi\right)-\left(U_v\left(\frac{t}{v},-\frac{t}{v}\right)\overline{\Psi}\left(-\frac{t}{v}\right),\phi\right)\right]e^{2itv^{-1}Q_1}\right|\\&\quad+\left|\left(U_v\left(\frac{t}{v},-\frac{t}{v}\right)\overline{\Psi}\left(-\frac{t}{v}\right)-\exp\left(iv^{-1}\int_{-t}^{t}E\left(\frac{q}{v}\right)dq\right)\overline{\Psi}\left(\frac{t}{v}\right),\phi\right)e^{2itv^{-1}Q_1}\right|\\&\quad\left.+\left|\exp\left(-iv\int_{-t}^{t}E\left(\frac{q}{v}\right)dq\right)\left(\overline{\Psi}\left(\frac{t}{v}\right)-\bar\phi,\phi\right)e^{2itv^{-1}Q_1}\right|\right\}\\&\leqslant\limsup_{t\to+\infty}\left\|U_v\left(\frac{t}{v},-\frac{t}{v}\right)\overline{\Psi}\left(-\frac{t}{v}\right)-\exp\left(iv^{-1}\int_{-t}^{t}E\left(\frac{q}{v}\right)dq\right)\overline{\Psi}\left(\frac{t}{v}\right)\right\|.\end{aligned} \tag{2.61}$$

Here $\bar{\phi}$ differs from ϕ only by a phase:

$$\bar{\phi} \equiv e^{-i\beta}\phi , \tag{2.62}$$

$$\beta \equiv -\operatorname{Im} \int_{-\infty}^{+\infty} (\dot{\Psi}(s),\Psi(s)) ds . \tag{2.63}$$

It follows from (2.1)–(2.4), (2.22), (2.45), (2.46), (2.62), and (2.63) that

$$\lim_{t\to+\infty} \|\overline{\Psi}(t/v) - \bar{\phi}\| = O(1/t) . \tag{2.64}$$

Similarly, $\overline{\Psi}(t/v) \to \phi, t \to -\infty$. Therefore, from the adiabatic estimate (2.56) and (2.61)–(2.64) we get

$$\big| \, |(\Omega_+^* (v)\Omega_-(v)\phi,\phi)| - |(\phi,\phi)| \, \big| = O(v), \quad v \to 0. \tag{2.65}$$

This proves (1.14).

ACKNOWLEDGMENTS

This work was part of my doctoral thesis at the Institute for Theoretical Physics at the University of Vienna.

I wish to express my sincere gratitude to Heide Narnhofer and Walter Thirring, for all they have taught me, and for their kind hospitality during my studies at Vienna.

I also wish to express my sincere gratitude to Ricardo Weder, from the University of Mexico, for all he has taught me, for his suggestions during the preparation of this paper, and for his critical reading of the final draft.

[1]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I (Academic, New York, 1972).
[2]R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).
[3]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).
[4]A. Böhm, *Quantum Mechanics* (Springer, New York, 1979).
[5]M. Reed and B. Simon, *Methods of Modern Mathematical Analysis*, Vol. II (Academic, New York, 1975).
[6]W. Thirring, *A Course in Mathematical Physics*, Vol. III (Springer, New York, 1981).
[7]J. S. Howland, Math. Ann. **207**, 315 (1974).
[8]K. Yajima, Commun. Math. Phys. **75**, 153 (1980).
[9]G. A. Hagedorn, Ann. Inst. H. Poincaré, Sect. A **36**, 19 (1982).
[10]J. H. Arredondo, Comunicaciones Técnicas, Serie Naranja, IIMAS–UNAM, Vol. 460, 1987.
[11]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[12]H. S. W. Massey, Rep. Prog. Phys. **12**, 248 (1949).
[13]J. B. Delos, Rev. Mod. Phys. **53**(2), 287 (1981).
[14]J. H. Arredondo, Comunicaciones Técnicas, Serie Naranja, IIMAS–UNAM, Vol. 459, 1987.
[15]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976).
[16]E. M. Stein, *Singular Integrals and Differentiability Properties of Functions* (Princeton U.P., Princeton, NJ, 1970).
[17]K. Yosida, *Functional Analysis* (Springer, Berlin, 1978).
[18]M. Born and V. Fock, "Beweis des Adiabatensatzes," Z. Phys. **51**, 165 (1928).
[19]T. Kato, J. Phys. Soc. Jpn. **5**, 435 (1950).
[20]L. D. Landau and E. M. Lifschitz, *Lehrbuch der Theoretischen Physik III* (Academic, Berlin, 1971).
[21]L. M. Garrido, J. Math. Phys. **5**, 355 (1964).
[22]S. J. Sancho, Proc. Phys. Soc. London **89**, 1 (1966).
[23]J. E. Avron, R. Seiler, and L. G. Yaffe, Preprint Nr. 158, Princeton University, 1987.
[24]A. Lenard, Ann. Phys. NY **6**, 261 (1959).
[25]G. Nenciu, Commun. Math. Phys. **82**, 121 (1981).
[26]B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).

Symmetry preserving deformations of generalized Taub–NUT space-times to all orders in perturbation theory

Vincent Moncrief
Department of Physics and Department of Mathematics, Yale University, P.O. Box 6666, New Haven, Connecticut 06511

(Received 14 February 1989; accepted for publication 19 April 1989)

Generalized Taub–NUT (Newman–Unti–Tamburino) space-times have compact Cauchy horizons and (generically) admit one (spacelike) Killing field in their globally hyperbolic regions. A large family of such vacuum space-times can be defined on circle bundles over $K \times R$, where K is a compact two-manifold, with the circular fibers of the bundle being defined by the orbits of the Killing field. For the simplest case of product circle bundles the symmetry preserving vacuum perturbations of such backgrounds to arbitrarily high order in perturbation theory are considered. The analytic form of the general solution of the nth-order perturbation equations for all n is derived under the restriction that the perturbations considered preserve the (one Killing field) symmetry of the background. The evolution equations are treated first and then the constraint equations are imposed, recovering along the way the well-known result that linearization instabilities arise only if one attempts to perturb from one (Killing) symmetry class to another. Gauge transformations, decompositions, and the natural symplectic structure associated with the perturbation formalism are also discussed. The possibility of extending these results to the case of symmetry breaking perturbations and of using the results to derive the asymptotic behavior of solutions near their singular boundaries is briefly discussed.

I. INTRODUCTION

One of the main open problems in classical general relativity is the understanding of space-time singularities. Even if quantum effects should ultimately be shown to modify the nature (or even the existence) of the singularities predicted by Einstein's theory, it seems likely that this should happen only at distance scales on the order of the Planck length or beyond. Thus one would expect that the approach to the singular state predicted by general relativity should remain valid up to corresponding large values of space-time curvature. The domain of applicability of the classical theory would thus extend from Planck scales to cosmic distance scales, a range of at least 50 orders of magnitude.

One of the main difficulties in studying singularities is that of producing reliable approximation methods for solving Einstein's equations near a region of divergent curvature. In this paper we develop a new approach to this problem by considering the perturbations, to arbitrarily high order, of a family of non-curvature-singular space-times, the generalized Taub–NUT (Newman–Unti–Tamburino) space-times discussed in earlier work.[1,2] The generalized Taub–NUT space-times all have compact Cauchy horizons at the boundaries of their globally hyperbolic regions and each has at least one Killing vector field spacelike in the globally hyperbolic region and null on the Cauchy horizon. In the analytic case (to which we restrict our attention) the metrics of these space-times can all be expressed in terms of certain convergent power series expansions about the horizon surfaces themselves. They comprise, on any given allowed manifold, an infinite-dimensional family of inequivalent solutions of the Einstein equations. Though infinite-dimensional this family contains (roughly speaking) only half the number of free functions one would expect to have in the general solutions of Einstein's equations of the chosen symmetry type. This is hardly surprising—even within the chosen symmetry class one expects the generic solution to exhibit a curvature singularity instead of a Cauchy horizon at the boundary of its maximal Cauchy development. Unfortunately, however, the method used to construct the generalized Taub–NUT solutions (a slight extension of the Cauchy–Kowalewski theorem) does not seem to be directly applicable to obtaining the general solution within the given symmetry class. This also is hardly surprising since the generic solution, in view of its expected curvature singularity, seems unlikely to admit an expression in terms of convergent expansions about its singular boundary. Of course, one can always expand the general solution about a nonsingular (Cauchy) hypersurface in the globally hyperbolic region but, unless one is miraculously able to sum the infinite series expressions explicitly, this technique is unlikely to shed much light on the nature of the singularities such general solutions are expected to include.

In this paper we consider the sequence of linear problems generated by perturbing Einstein's equations, to arbitrarily high order, about an arbitrary generalized Taub–NUT background space-time. For the present we only consider perturbations that remain within the given symmetry class (of one spacelike Killing field) but we believe that the same methods are equally applicable to the general problem of nonsymmetric perturbations. Our main results include the determination of the form of the general solution of

 0022-2488/89/102297-18$02.50

the nth-order perturbation equations for arbitrarily large n. In a sense we accomplish this by showing that one can "factor out" the singular parts of the nth-order perturbation functions and determine the analytic coefficients of these singular factors by means of the extended Cauchy–Kowalewski theorem alluded to above. At first order our technique is quite analogous to the classical Frobenius method for solving an ordinary, linear differential equation near a regular singular point of that equation. At higher orders the technique is similar except that we now encounter inhomogeneous, singular terms that arise from all the lower-order perturbations. Remarkably we find that the Frobenius-type approach can be extended to handle these singular source terms and to determine the form of the general solution of the nth-order perturbation equations for arbitrary n.

The plan of this paper is as follows. In Sec. II we recall the main features of our (background) generalized Taub–NUT space-times and derive the form of the general solution of the nth-order perturbed evolution equations for perturbations that remain within the chosen symmetry class. In Sec. III we impose the perturbed constraint equations and discuss gauge tranformations, decompositions, and the natural symplectic structure for our perturbation formalism. In Sec. IV we briefly discuss several possible extensions and generalizations of our work. In the future we hope to show how one can collect together the "dominant singular terms" from each order and sum the resultant truncated series to determine the asymptotic behavior of the perturbed solutions near their singular boundaries.

II. SOLVING THE nth-ORDER EVOLUTION EQUATIONS

A. Generalized Taub–NUT space-times

The main property we desire for our background solutions is that they admit compact Cauchy horizons. Compactness corresponds to the physically interesting boundary condition of a closed universe whereas the existence of a Cauchy horizon ensures (at least in the analytic case) that we can express the background metric in terms of convergent power series expansions about the horizon surface itself. A compact Cauchy horizon has a null geodesic generator passing through each of its points and lying entirely in the horizon surface. If the null generators are all closed curves then the horizon surface must have the structure of either a circle bundle or a Seifert manifold (which is covered by a circle bundle).[3,4] In these cases Isenberg and the author have shown that any analytic vacuum (or electrovacuum) space-time with such a horizon necessarily admits a Killing vector field null on the horizon (and thus tangent to its generators) and spacelike in a globally hyperbolic region neighboring the horizon (where its integral curves are also closed). Thus a globally hyperbolic region neighboring the horizon has the natural geometrical structure of a circle bundle or a space covered by such a bundle. In particular, the Taub regions of the Taub–NUT solutions are (nontrivial) S^1 bundles over $S^2 \times R$.

Isenberg and the author are attempting to extend this result with a proof of the conjecture that if the generators of the compact horizon are not closed then there is nevertheless a Killing field tangent to the horizon generators and furthermore the dimension of the full isometry group for the space-time must be greater than or equal to 2. The last assertion follows from the fact that the isometry groups of compact Riemannian manifolds are necessarily compact whereas the isometry group generated by a single Killing field (with nonclosed orbits) acting on a suitably chosen Cauchy surface in the globally hyperbolic region would be noncompact. If the stated conjecture is true then clearly the circle bundle horizons (and those covered by circle bundles) would provide the "largest" families of vacuum space-times admitting compact Cauchy horizons since the other families would necessarily have higher-dimensional isometry groups.

Because we wish to treat space-times having only one-dimensional isometry groups and because the perturbation of spatially compact vacuum space-times from one symmetry class to another is beset with linearization instabilities[5,6] we find it most convenient to choose background solutions admitting only one Killing vector field and thus belonging to one of the S^1-bundle (or Seifert) families mentioned above. We therefore consider vacuum space-times that are S^1 bundles over $K \times R$, where K is an arbitrary, compact, orientable two-manifold. For simplicity we shall restrict our attention to the trivial bundles $K \times R \times S^1 \to K \times R$, though the same methods could certainly be applied (by suitably patching together local trivializations) to handle the nontrivial bundles as well (see Ref. 2 for the treatment of the nontrivial bundle $S^3 \times R \to S^2 \times R$). We shall refer to such vacuum space-times having S^1-symmetric, compact Cauchy horizons as generalized Taub–NUT (GTN) space-times, although, strictly speaking, that term ought to be reserved for the bundle $S^3 \times R \to S^2 \times R$.

In the following we let $\{x^a, a = 1,2\}$ represent local coordinates on the compact two-manifold K, x^3 (defined mod 2π) represent an angle coordinate on the circle, and $x^0 = t \in R$ represent the "time."

The Lorentzian metrics on ${}^{(4)}V = K \times R \times S^1$ are expressible in the form

$$\begin{aligned} ds^2 &= {}^{(4)}g_{\mu\nu}\, dx^\mu\, dx^\nu \\ &= e^{-2\phi}[-N^2\, dt^2 + g_{ab}\, dx^a\, dx^b] \\ &\quad + t^2 e^{2\phi}[k\, dx^3 + \beta_a\, dx^a]^2, \end{aligned} \tag{2.1}$$

where k is a nonzero constant and $\partial/\partial x^3$ is a Killing vector field. By analogy with the well-known Kaluza–Klein–Jordan reduction we may view ϕ, $\beta_a\, dx^a$, and $(-N^2\, dt^2 + g_{ab}\, dx^a\, dx^b)$ as a scalar field, one-form, and Lorentzian metric, respectively, induced on the base manifold $K \times R$ by the space-time metric on $K \times R \times S^1$. For simplicity we have imposed the coordinate condition of zero shift vector field, which corresponds to dropping the time component of the one-form field and the two-dimensional shift field of the Lorentzian metric induced on $K \times R$.

The line element (2.1) degenerates at $t = 0$. However, if we reexpress it through the change of coordinates

$$t' = t^2, \quad x^{3'} = x^3 - (1/k)\ln t, \quad x^{a'} = x^a, \tag{2.2}$$

then we can easily show that the transformed metric is analytic and Lorentzian on a neighborhood $\mathscr{N}=K\times S^1\times(-\lambda,\lambda)$ of the surface $t'=0$ provided

(i) $\phi(t',x^{a'})$, $N(t',x^{a'})$, $\beta_a(t',x^{b'})dx^a$, and $g_{ab}(t',x^{c'})dx^{a'}dx^{b'}$ are analytic on $\mathscr{N}$;

(ii) $N>0$ and g_{ab} is positive definite on $\mathscr{N}$;

(iii) $((N^2-e^{4\phi})/4t')$ is analytic on $\mathscr{N}$.

The transformed metric has the form

$$\begin{aligned}ds^2 &= {}^{(4)}g_{\mu'\nu'}\,dx^{\mu'}\,dx^{\nu'}\\ &= (-e^{-2\phi}/4t')(N^2-e^{4\phi})(dt')^2\\ &\quad + e^{-2\phi}g_{ab}\,dx^{a'}\,dx^{b'}\\ &\quad + e^{2\phi}t'[k\,dx^{3'}+\beta_a\,dx^{a'}]^2\\ &\quad + e^{2\phi}\,dt'[k\,dx^{3'}+\beta_a\,dx^{a'}].\end{aligned}\tag{2.3}$$

For such a metric it is easy to show that

(iv) the surface $t'=0$ is a null hypersurface with $\partial/\partial x^{3'}$ tangent to its null generators;

(v) the Killing field $\partial/\partial x^{3'}$ is spacelike in the region $t'>0$ but timelike in the region $t'<0$—its orbits there being closed timelike curves.

Space-times satisfying the conditions (i)–(iii) above are globally hyperbolic in the regions $t'>0$ (which were covered by the original charts with either $t>0$ or $t<0$), have Cauchy horizons diffeomorphic to $K\times S^1$ at $t'=0$, and have closed timelike curves through every event in their acausal extensions $t'<0$. For each such space-time, a second, inequivalent extension through the Cauchy horizon can be defined by introducing the chart

$$t'=t^2,\quad x^{3'}=x^3+(1/k)\ln t,\quad x^{a'}=x^a \tag{2.4}$$

instead of (2.2) and proceeding as before. These two extensions correspond to the well-known pair of extensions for the Taub space-time that cannot simultaneously be accommodated within a Hausdorff manifold.

The Einstein equations for a metric of the form (2.1) are written out explicitly in Eqs. (2.4)–(2.6) of Ref. 1. As discussed in detail in that reference one may prove the existence of analytic solutions of Einstein's equations having all the properties (i)–(v) above by imposing a suitable coordinate condition to fix the lapse function N and applying the extended Cauchy–Kowalewski theorem sketched there and proved in detail in Ref. 2. Every choice of analytic initial data $\{\mathring{\phi}, \mathring{\beta}_a, \mathring{g}_{ab}\}$ $(0,x^c)$ specified over K (with $\mathring{\phi}$ a function, $\mathring{\beta}_a\,dx^a$ a one-form, and $\mathring{g}_{ab}\,dx^a\,dx^b$ a Riemannian metric on K) determines a unique, analytic solution of Einstein's equations having all the properties (i)–(v) above provided the lapse function is chosen to satisfy both conditions (i)–(iii) above and the condition

$$\text{(vi)}\quad (N/\sqrt{{}^{(2)}g})_{,t}=0, \tag{2.5}$$

where ${}^{(2)}g$ is the determinant of g_{ab}. These restrictions lead to the requirement that

$$N=(e^{2\mathring{\phi}}/\sqrt{{}^{(2)}\mathring{g}})\sqrt{{}^{(2)}g}, \tag{2.6}$$

which fixes N uniquely.

The rigid coordinate conditions described above were chosen originally to simplify the form of the Einstein evolution equations and are not strictly necessary for the analysis to follow. Any choice of lapse and shift that still permits one to apply the extended Cauchy–Kowalewski theorem would probably work just as well. For simplicity, however, we shall retain the coordinate conditions of zero shift and a lapse satisfying (vi) above in the present discussion.

Many of the solutions determined by data $\{\mathring{\phi}, \mathring{\beta}_a, \mathring{g}_{ab}\}$ prescribed on K are isometric to one another. In order to characterize the space of inequivalent generalized Taub–NUT space-times it is convenient to transform the metrics to a canonical gauge. For this purpose we consider, for each such space-time, the group of analytic diffeomorphisms that preserve the horizon at $t'=0$ and commute with the isometry generated by the Killing field $\partial/\partial x^{3'}$. Infinitesimal generators of such diffeomorphisms are vector fields ${}^{(4)}X$ that are tangent to the null hypersurface at $t'=0$ and that commute with $\partial/\partial x^{3'}$.

We can express such vector fields as

$${}^{(4)}X=t'Y\frac{\partial}{\partial t'}+X^{a'}\frac{\partial}{\partial x^{a'}}+X^{3'}\frac{\partial}{\partial x^{3'}}, \tag{2.7}$$

where $(Y,X^{a'},X^{3'})$ are analytic in $(t',x^{a'})$. The infinitesimal gauge transformation of ${}^{(4)}g$ induced ${}^{(4)}X$ is given by the Lie derivative $\mathscr{L}_{{}^{(4)}X}{}^{(4)}g$. In order that ${}^{(4)}X$ preserve the coordinate conditions we have imposed on ${}^{(4)}g$ it is necessary (and sufficient) that ${}^{(4)}X$ be required to satisfy the evolution equations [Eqs. (3.4) of Ref. 1]

$$\begin{aligned}Y_{,t} &= 2\frac{N}{\sqrt{{}^{(2)}g}}\frac{\partial}{\partial x^a}\left(\frac{\sqrt{{}^{(2)}g}}{N}\Delta^{a'}\right),\\ (t\Delta^{a'})_{,t} &= (N^2/2)t\,g^{ab}Y_{,b},\\ (X^{3'}+Y)_{,t} &= -2\beta_a X^{a'}{}_{,t},\end{aligned}\tag{2.8}$$

where

$$\Delta^{a'}=(1/t)(X^{a'}-\mathring{X}^{a'})$$

and $\mathring{X}^{a'}$ are the initial values of X^a (i.e., $X^{a'}|_{t'=0}$). As in Ref. 1 we can apply the extended Cauchy–Kowalewski theorem to prove the existence and uniqueness of solutions of Eqs. (2.8) that are analytic and even in t (hence also analytic in $t'=t^2$) for arbitrary analytic initial data $(\mathring{Y}, \mathring{X}^{a'}\,\mathring{X}^{3'})$ prescribed at $t=0$.

The infinitesimal gauge tranformations of (ϕ,β_a,g_{ab},N) induced by such a vector field are determined by computing $\mathscr{L}_{{}^{(4)}X}{}^{(4)}g$ and are given in Eqs. (3.5) of Ref. 1. These perturbations reduce, as $t\to 0^+$, to the expressions [also given in Eqs. (3.6) of Ref. 1]

$$\begin{aligned}\delta\mathring{\phi} &= \tfrac{1}{2}\mathring{Y}+\mathscr{L}_{{}^{(2)}\mathring{X}}\mathring{\phi},\\ \delta\mathring{\beta}_a &= \tfrac{1}{2}(\mathring{X}^{3'}+\mathring{Y})_{,a}+(\mathscr{L}_{{}^{(2)}\mathring{X}}\mathring{\beta})_a,\\ \delta\mathring{g}_{ab} &= \mathring{Y}\mathring{g}_{ab}+(\mathscr{L}_{{}^{(2)}\mathring{X}}\mathring{g})_{ab},\\ \delta\mathring{N} &= \mathring{Y}\mathring{N}+\mathscr{L}_{{}^{(2)}\mathring{X}}\mathring{N},\end{aligned}\tag{2.9}$$

where

$${}^{(2)}\mathring{X}=\mathring{X}^{a'}\frac{\partial}{\partial x^{a'}}=\mathring{X}^{a'}\frac{\partial}{\partial x^a}$$

and $\mathscr{L}_{^{(2)}\mathring{X}}$ signifies the Lie derivative with respect to this vector field on K.

These infinitesimal gauge transformations of the Cauchy horizon data clearly consist of (i′) an infinitesimal diffeomorphism of K generated by ${}^{(2)}\mathring{X}$; (ii′) an "electromagnetic" gauge transformation of β_a generated by

$$\mathring{Z} \equiv \tfrac{1}{2}(\mathring{X}^{3'} + \mathring{Y}),$$

and (iii′) a "conformal" transformation of $(\mathring{\phi}, \mathring{g}_{ab})$ generated by $\mathring{Y}$. The "conformal" transformation is just such as to preserve the tensor $e^{-2\mathring{\phi}}\mathring{g}_{ab}$ and in general these gauge transformations preserve the regularity condition $\mathring{N} = e^{2\mathring{\phi}}$ discussed earlier.

Equations (2.9) define the Lie algebra of a group of transformations acting on the space of Cauchy horizon data. To see the structure of this Lie algebra let $(Y, {}^{(2)}X, Z)$ and $(\hat{Y}, {}^{(2)}\hat{X}, \hat{Z})$ with $Z \equiv \frac{1}{2}(X^{3'} + Y)$ and $\hat{Z} = \frac{1}{2}(\hat{X}^{3'} + \hat{Y})$ be any two such infinitesimal transformations and compute their commutator. The result is a transformation of the same type with a generator given by

$$\begin{aligned}&(Y^*, {}^{(2)}X^*, Z^*)\\&\quad = (\mathscr{L}_{^{(2)}X}\hat{Y} - \mathscr{L}_{^{(2)}\hat{X}}Y, [{}^{(2)}X, {}^{(2)}\hat{X}], \mathscr{L}_{^{(2)}X}\hat{Z} - \mathscr{L}_{^{(2)}\hat{X}}Z).\end{aligned} \tag{2.10}$$

The associated group has two obvious, commuting, Abelian subgroups determined by generators of the type $(Y,0,0)$ ("conformal" transformations) and the type $(0,0,Z)$ ("electromagnetic" gauge transformations). In addition there is the non-Abelian subgroup with generators of the type $(0, {}^{(2)}X, 0)$. This last subgroup is clearly just the diffeomorphism group of K.

As is well known, every Riemannian metric on K is conformal to a metric of constant curvature. The scalar curvature of the transformed metric will be a positive constant if $K \approx S^2$, zero if $K \approx T^2$, and a negative constant if K is a higher genus two-manifold. We can always choose the conformal factor so that the transformed metric $\mathring{g}^*_{ab}$ satisfies (say)

$$\int_K \sqrt{{}^{(2)}\mathring{g}^*} = 4\pi,$$

in which case the scalar curvature, ${}^{(2)}R({}^{(2)}\mathring{g}^*)$, will assume a (constant) value fixed by the Gauss–Bonnet theorem and depending only upon the genus of K. Without disturbing this gauge condition we can apply an "electromagnetic" gauge transformation to make $\mathring{\beta}_a$ divergence-free (with respect to the new, constant curvature metric $\mathring{g}^*_{ab}$). Next we can, without disturbing these conditions, apply a diffeomorphism of K to bring the constant curvature metric into a canonical form. For example, if $K \approx S^2$ one could require that

$$\mathring{g}^*_{ab}\, dx^a\, dx^b = d\theta^2 + \sin^2\theta\, d\phi^2$$

whereas, if $K \approx T^2$ one could require that the (flat) metric $\mathring{g}^*_{ab}$ have constant components and thus, in view of the condition

$$\int_K \sqrt{{}^{(2)}\mathring{g}^*} = 4\pi,$$

depend on only two real parameters. More generally, for higher genus two-manifolds, the moduli spaces of conformal equivalence classes of Riemannian metrics are parametrized by $3g - 3$ complex parameters for genus $g \geqslant 2$. Thus in a suitably chosen canonical gauge $\mathring{g}^*_{ab}$ could be uniquely specified in terms of $6g - 6$ real parameters.

Finally, without disturbing the foregoing conditions one can apply gauge transformations generated by the "conformal isometries" of $\mathring{g}^*_{ab}$. More precisely, if ${}^{(2)}\mathring{X}^a$ is any conformal Killing field of $\mathring{g}^*_{ab}$ and we set

$$\mathring{Y} = -\frac{1}{\sqrt{{}^{(2)}\mathring{g}^*}}\frac{\partial}{\partial x^a}(\sqrt{{}^{(2)}\mathring{g}^*}\mathring{X}^a), \tag{2.11}$$

then

$$\delta\mathring{g}^*_{ab} = \mathring{Y}\mathring{g}^*_{ab} + (\mathscr{L}_{^{(2)}\mathring{X}}\mathring{g}^*)_{ab} = 0,$$

and one can check that the condition that $\mathring{\beta}^*_a$ have vanishing divergence with respect to $\mathring{g}^*_{ab}$ is left undisturbed provided we set $\mathring{Z} = \frac{1}{2}(\mathring{X}^{3'} + \mathring{Y}) = 0$.

The conformal isometry group of $(K, \mathring{g}^*_{ab})$ is a Lie group of dimension 6 if $K \approx S^2$, dimension 2 if $K \approx T^2$ (and coincides with the isometry group since $\mathring{g}^*_{ab}$ is flat), and dimension zero if K is a higher genus two-manifold. Thus for $K \approx S^2$ or T^2 we can take the quotient by the corresponding group action and further reduce the space of inequivalent generalized Taub–NUT space-times.

To summarize, we can always choose a canonical gauge for the Cauchy horizon data $\{\mathring{\phi}, \mathring{\beta}_a, \mathring{g}_{ab}\}$ such that (a) $\mathring{g}_{ab}$ is a constant curvature metric on K depending only on zero (if $K \approx S^2$), 2 (if $K \approx T^2$), or $6g - 6$ (if K has genus $g \geqslant 2$) real parameters; (b) $\mathring{\beta}_a$ has zero divergence with respect to $\mathring{g}_{ab}$; and (c) there is a residual gauge subgroup action of dimension 6 (if $K \approx S^2$) or dimension 2 (if $K \approx T^2$) generated by the conformal Killing fields of $(K, \mathring{g}_{ab})$, which acts on the data $(\mathring{\phi}, \mathring{\beta}_a, \mathring{g}_{ab})$.

B. First- and second-order perturbations

We want to consider perturbations of the Einstein evolution equations for ϕ, β_a, and g_{ab} [cf. Eqs. (2.4) of Ref. 1] about an arbitrary generalized Taub–NUT background. To generate the perturbation equations of arbitrary order one imagines having a one parameter family of exact solutions $\{\phi, \beta_a, g_{ab}\}(t, x^c, \epsilon)$ and differentiates the exact equations with respect to the parameter ϵ arbitrarily many times, setting $\epsilon = 0$ (the background value) at the end. The nth-order equations consist of a linear second-order operator acting on the nth-order perturbations,

$$\{\phi^{(n)}, \beta_a^{(n)}, g_{ab}^{(n)}\} \equiv \left\{\frac{\partial^n\phi}{\partial\epsilon^n}, \frac{\partial^n\beta_a}{\partial\epsilon^n}, \frac{\partial^n g_{ab}}{\partial\epsilon^n}\right\}\Bigg|_{\epsilon=0} \tag{2.12}$$

and an inhomogeneous "source" term formed from perturbations of the lapse function and (for $n > 1$) from all the lower-order perturbations. For simplicity we retain the coordinate conditions of zero shift and of $(N/\sqrt{{}^{(2)}g})_{,t} = 0$, which means that we only allow perturbations of the lapse function N generated by differentiations of

$$\frac{N}{\sqrt{{}^{(2)}g}} = \left(\frac{N}{\sqrt{{}^{(2)}g}}\right)\Bigg|_{\epsilon=0} e^{2\lambda(x^a,\epsilon)}, \tag{2.13}$$

where $\lambda(x^a, \epsilon)$ is an arbitrary analytic function of the indicated arguments with $\lambda(x^a, 0) = 0$.

The main tool we require is the extended Cauchy–Kowalewski (CK) theorem sketched in Ref. 1 and proved in detail in Ref. 2. Since the applications of this theorem to the present problem follow the same pattern employed in those earlier papers (once we have factored out the singularities) we shall simply state the results of applying the theorem without giving repetitious details. Just as for the background solutions, all of the coefficient functions determined by means of the extended CK theorem will prove to be both analytic in (x^a, t) and even in t (and thus also analytic in the time variable $t' = t^2$ defined above). For convenience we shall refer to any function analytic and even in t as a *regular* function. The analyticity of the solutions will follow directly from the application of the CK theorem whereas the evenness in t requires a more detailed consideration of the specific evolution equations—one shows inductively that all odd t derivatives of the analytic solutions vanish at $t = 0$.

The main result we intend to prove may be stated as follows. The solutions of the nth-order perturbation equations are expressible (suppressing the tensor indices on $\beta^{(n)} \equiv \beta_a^{(n)}$ and $g^{(n)} \equiv g_{ab}^{(n)}$ for simplicity) as

$$\phi^{(n)} = \sum_{k=0}^{(n-1)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} [(\ln t) a_{n-2k}^{(n)} + a_{n-2k-1}^{(n)}],$$

$$\beta^{(n)} = \sum_{k=0}^{(n-1)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} \times \left[(\ln t) b_{n-2k}^{(n)} + \frac{b_{n-2k-1}^{(n)}}{t^2}\right], \tag{2.14}$$

$$g^{(n)} = \sum_{k=0}^{(n-1)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} [(\ln t) c_{n-2k}^{(n)} + c_{n-2k-1}^{(n)}],$$

if n is odd (i.e., $n = 1,3,5,\ldots$), and

$$\phi^{(n)} = \sum_{k=0}^{(n-2)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} [(\ln t) a_{n-2k}^{(n)} + a_{n-2k-1}^{(n)}] + \frac{a_0^{(n)}}{t^n},$$

$$\beta^{(n)} = \sum_{k=0}^{(n-2)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} \left[(\ln t) b_{n-2k}^{(n)} + \frac{b_{n-2k-1}^{(n)}}{t^2}\right] + \frac{b_0^{(n)}}{t^n}, \tag{2.15}$$

$$g^{(n)} = \sum_{k=0}^{(n-2)/2} \frac{(\ln t)^{n-2k-1}}{t^{2k}} [(\ln t) c_{n-2k}^{(n)} + c_{n-2k-1}^{(n)}] + \frac{c_0^{(n)}}{t^n},$$

if n is even (i.e., $n = 2,4,6,\ldots$). Here each of the coefficient functions $\{a_l^{(n)}, b_l^{(n)}, c_l^{(n)};\ l = 0,1,\ldots,n;\ n = 1,2,3,\ldots\}$ will be shown to be regular [i.e., analytic in (t, x^a) and even in t] in a neighborhood of $t = 0$, the horizon surface of the background solution. The solutions at each order are determined uniquely by the solutions from all the preceding orders up to the addition of an arbitrary solution of the first-order ($n = 1$) equations.

The first-order perturbations, according to the above, have the form

$$\phi^{(1)} = (\ln t) a_1^{(1)} + a_0^{(1)},$$
$$\beta^{(1)} = (\ln t) b_1^{(1)} + b_0^{(1)}/t^2, \tag{2.16}$$
$$g^{(1)} = (\ln t) c_1^{(1)} + c_0^{(1)},$$

and, as we shall see, the freely specifiable data will be the initial values

$$\left\{ a_1^{(1)}, a_0^{(1)}, b_0^{(1)}, \frac{\partial^2 b_0^{(1)}}{\partial t^2}, c_1^{(1)}, c_0^{(1)} \right\}\Bigg|_{t=0}, \tag{2.17}$$

prescribed as analytic tensor fields on K and the first-order perturbation of the lapse function computed from Eq. (2.13). Thus at nth order the only freedom is that of adding an arbitrary solution of the first-order equations, which in turn is uniquely determined by the data indicated above. Later when we impose the nth-order constraint equations some of this "free data" will itself be restricted. The data indicated above are, aside from the arbitrary perturbation of the lapse function, just the 12 independent functions one would expect to have in solving the six second-order evolution equations for $\{\phi, \beta_a, g_{ab}\}$, but they are not conventional Cauchy data since we are choosing to specify the "initial conditions" on a Cauchy horizon instead of on a Cauchy surface of the background space-time.

The difference in form of the perturbations $\{\phi^{(n)}, g^{(n)}\}$ from the perturbations $\{\beta^{(n)}\}$ is traceable to the difference between the singular time derivative operators that occur in the evolution equations for these quantities. The operator $\partial^2/\partial t^2 + (1/t)\partial/\partial t$, which acts on the quantities $\phi^{(n)}$ and $g^{(n)}$, annihilates any field of the form $\{(\ln t) a_1(x^b) + a_0(x^b)\}$, whereas the operator $\partial^2/\partial t^2 + (3/t)\partial/\partial t$, which acts on the $\beta^{(n)}$, annihilates any field of the form $\{(1/t^2) b_1(x^c) + b_0(x^c)\}$. These facts determine the form of the freely specifiable data in the perturbation equations for these quantities. The interaction terms between these variables force the additional $(\ln t)$ term in the first-order perturbation of β_a and, at higher orders, these interactions drive the higher-order logarithmic singularities and the higher-order singularities in $1/t^2$, which appear in the general solution for the nth-order perturbations. If one turns off the β_a field altogether (which is always allowed on the trivial bundles $K \times R \times S^1 \to K \times R$) then the general solution for the remaining perturbations simplifies to

$$\phi^{(n)} = (\ln t)^n a_n^{(n)} + (\ln t)^{n-1} a_{n-1}^{(n)} + \cdots + (\ln t) a_1^{(n)} + a_0^{(n)}, \tag{2.18}$$
$$g^{(n)} = (\ln t)^n c_n^{(n)} + (\ln t)^{n-1} c_{n-1}^{(n)} + \cdots + (\ln t) c_1^{(n)} + c_0^{(n)},$$

where the coefficients $\{a_k^{(n)}, c_k^{(n)}\}$ are all analytic and even in t.

To analyze the perturbation equations we shall employ a convenient schematic form for the evolution equations. Since ϕ and g_{ab} occur in somewhat parallel ways in the evolution equations we can compress them into a single symbol, which, with indices suppressed, we shall call ψ. Thus we let ψ stand for the pair (ϕ, g_{ab}) and β stand for β_a as before. Taking account of the coordinate condition, which we have imposed to yield $(N/\sqrt{^{(2)}g})_{,t} = 0$, we find that the evolution equations can be written schematically as

$$\psi_{,tt} + (1/t)\psi_{,t} = A(\psi,N) + B(\psi,N)t^2\, d\beta\, d\beta + C(\psi)\psi_{,t}\psi_{,t} + D(\psi)t^2\beta_{,t}\beta_{,t}, \tag{2.19}$$
$$\beta_{,tt} + (3/t)\beta_{,t} = E(\psi,N)d\beta + F(\psi,N)\partial\, d\beta + G(\psi)\psi_{,t}\beta_{,t},$$

where $d\beta$ is the exterior derivative of the one-form β [i.e., $d\beta = \frac{1}{2}(\beta_{a,b} - \beta_{b,a})dx^b \wedge dx^a$], A, B, E, and F are certain analytic expressions in $\{\phi, g_{ab}, N\}$ and their first and second spatial derivatives and where C, D, and G are similar expressions in ϕ and g_{ab} alone. The second spatial derivatives of β are indicated explicitly by $\partial d\beta$ whereas the spatial derivatives of ψ and N are contained in the quantities $A,...,G$ and thus suppressed in the above notation.

Linearizing Eqs. (2.19) about a regular background solution and taking account of the coordinate condition (2.13) when perturbing the lapse function N we obtain first-order perturbation equations of the schematic form

$$\begin{aligned}\psi^{(1)}_{,tt} + (1/t)\psi^{(1)}_{,t} &= R\cdot\psi^{(1)} + t^2R\, d\beta^{(1)} + tR\psi^{(1)}_{,t} + t^3R\beta^{(1)}_{,t} + R,\\ \beta^{(1)}_{,tt} + (3/t)\beta^{(1)}_{,t} &= R\cdot\psi^{(1)} + R\cdot d\beta^{(1)} + tR\psi^{(1)}_{,t} + tR\beta^{(1)}_{,t} + R,\end{aligned} \tag{2.20}$$

where $R\cdot$ stands generically for a linear (spatial) differential operator with regular coefficients and R stands for a regular multiplicative operator or a regular additive inhomogeneous term (induced by perturbations of the lapse function).

We now seek solutions of Eqs. (2.20) of the form [cf. Eqs. (2.16) above]

$$\begin{aligned}\psi^{(1)} &= (\ln t)a^{(1)}_1 + a^{(1)}_0,\\ \beta^{(1)} &= (\ln t)b^{(1)}_1 + b^{(1)}_0/t^2,\end{aligned} \tag{2.21}$$

with coefficients $\{a^{(1)}_k, b^{(1)}_k\}$, which are regular. For convenience we write $b^{(1)}_0 = \mathring{\gamma}_0(x^a) + t^2\gamma_1$, where $\mathring{\gamma}_0$ is analytic in $\{x^a\}$ and independent of t (as signified by the overhead naught) and where γ_1 is regular. Thus we write

$$\beta^{(1)} = (\ln t)b^{(1)}_1 + \mathring{\gamma}_0(x^a)/t^2 + \gamma_1 \tag{2.22}$$

and

$$\psi^{(1)} = (\ln t)a^{(1)}_1 + a^{(1)}_0 \tag{2.23}$$

and substitute these forms into the linearized evolution equations (2.20). We organize the resulting system by collecting together all terms having, as an overall factor, a common power of $(\ln t)$. At this ($n = 1$) level only the two powers $(\ln t)^k$, $k = 1,0$, occur. To obtain solutions we demand that the coefficients of each power of $(\ln t)$ vanish separately. The vanishing of the coefficients of $(\ln t)$ is equivalent to the requirement that Eqs. (2.20), with $\psi^{(1)}$ and $\beta^{(1)}$ replaced by $a^{(1)}_1$ and $b^{(1)}_1$ and the inhomogeneous R terms dropped, be satisfied. But one can apply the extended CK theorem to this system and prove the existence of regular solutions $\{a^{(1)}_1, b^{(1)}_1\}$ that are uniquely determined by the arbitrarily specified, analytic initial data $\{a^{(1)}_1, b^{(1)}_1\}|_{t=0}$. Having thus killed all the terms with the factor $(\ln t)$ we find that the linearized equations reduce to

$$\begin{aligned}a^{(1)}_{0\,,tt} &+ (1/t)a^{(1)}_{0\,,t} + (2/t)a^{(1)}_{1\,,t}\\ &= R\cdot a^{(1)}_0 + tR(a^{(1)}_{0\,,t} + a^{(1)}_1/t) + t^2R\, d(\gamma_1 + \mathring{\gamma}_0/t^2)\\ &\quad + t^3R\,(\gamma_{1,t} + b^{(1)}_1/t - (2/t^3)\mathring{\gamma}_0) + R,\\ \gamma_{1,tt} &+ (3/t)\gamma_{1,t} + (2/t)b^{(1)}_{1\,,t} + (2/t^2)b^{(1)}_1\\ &= R\cdot a^{(1)}_0 + R\cdot d(\gamma_1 + \mathring{\gamma}_0/t^2) + tR(a^{(1)}_{0\,,t} + a^{(1)}_1/t)\\ &\quad + tR\,(\gamma_{1,t} + b^{(1)}_1/t - (2/t^3)\mathring{\gamma}_0) + R.\end{aligned} \tag{2.24}$$

These have the form of Eqs. (2.20) (with $\{a^{(1)}_0, \gamma_1\}$ in place of $\{\psi^{(1)}, \beta^{(1)}\}$) supplemented with some additional inhomogeneous terms arising from $a^{(1)}_1, b^{(1)}_1$, and $\mathring{\gamma}_0$. Recalling that $a^{(1)}_1$ and $b^{(1)}_1$ are regular (and thus even in t) one sees that all of the inhomogeneous contributions are regular except for the terms $(2/t^2)b^{(1)}_1$, occurring on the left-hand side, and $[R\cdot d(\mathring{\gamma}_0/t^2) - 2R\mathring{\gamma}_0/t^2]$ occurring on the right-hand side of the equation for γ_1. We can kill the singularity provided by these terms, leaving only an additional regular inhomogeneity, by requiring that the initial data for $b^{(1)}_1$ (which was arbitrary up to this point) be fixed by the equation

$$2b^{(1)}_1|_{t=0} = [R\cdot(d\mathring{\gamma}_0) - 2R\mathring{\gamma}_0]|_{t=0}, \tag{2.25}$$

for any choice of $\mathring{\gamma}_0(x^a)$. With this restriction, Eqs. (2.24) become amenable to the extended CK theorem, which guarantees that regular solutions $\{a^{(1)}_0, \gamma_1\}$ exist for arbitrary, analytic initial data $\{a^{(1)}_0, \gamma_1\}|_{t=0}$.

Thus we obtain solutions of the linearized equations uniquely determined by the independent data $\{a^{(1)}_1, a^{(1)}_0, \gamma_1, \mathring{\gamma}_0(x^a)\}|_{t=0}$ with $b^{(1)}_1$ fixed by Eqs. (2.20) and the initial condition (2.25). The independent data consists of two analytic functions, two analytic one-forms and two analytic, symmetric tensor fields that one can prescribe arbitrarily on the two-manifold K. By treating the linearized equations as a Hamiltonian system one can show that this free data consists of canonically conjugate pairs of variables, as we shall see in Sec. III D. When we impose the linearized constraints some of this "free data" will, of course, be fixed in terms of the remaining data. The constraints will be dealt with in Sec. III.

Before turning to the general inductive proof it may be useful to sketch how the pattern of solution continues to the second order. The second-order perturbation equations consist of equations of the form (2.20), with $\{\psi^{(1)}, \beta^{(1)}\}$ replaced by $\{\psi^{(2)}, \beta^{(2)}\}$, supplemented by additional inhomogeneous terms that arise from quadratic terms in the first-order perturbations. These additional inhomogeneities consist of terms of the form $[R(\ln t)^2 + R(\ln t) + R/t^4]$, occurring in the $\psi^{(2)}$ equation, and $[R(\ln t)^2 + (R/t^2)(\ln t) + R/t^4]$, occurring in the $\beta^{(2)}$ equation. Following the inductive hypothesis [(2.14) and (2.15)] we seek solutions of the form

$$\begin{aligned}\psi^{(2)} &= (\ln t)^2a^{(2)}_2 + (\ln t)a^{(2)}_1 + a^{(2)}_0/t^2,\\ \beta^{(2)} &= (\ln t)^2b^{(2)}_2 + (\ln t)b^{(2)}_1/t^2 + b^{(2)}_0/t^2,\end{aligned} \tag{2.26}$$

with regular coefficients $\{a^{(2)}_k, b^{(2)}_k\}$. For convenience we also write

$$a_0^{(2)} = \mathring{\alpha}_0^{(2),0}(x^b) + t^2\alpha_1^{(2),0}$$
$$b_1^{(2)} = \mathring{\gamma}_0^{(2),1}(x^b) + t^2\gamma_1^{(2),1} \qquad (2.27)$$
$$b_0^{(2)} = \mathring{\gamma}_0^{(2),0}(x^b) + t^2\gamma_1^{(2),0}$$

where $\{\mathring{\alpha}_0^{(2),0}(x^b), \mathring{\gamma}_0^{(2),1}(x^b), \mathring{\gamma}_0^{(2),0}(x^b)\}$ are analytic and independent of t and where $\{\alpha_1^{(2),0}, \gamma_1^{(2),1}, \gamma_1^{(2),0}\}$ are analytic and even in t. Thus we substitute the forms

$$\psi^{(2)} = (\ln t)^2 a_2^{(2)} + (\ln t) a_1^{(2)} + \alpha_1^{(2),0} + \mathring{\alpha}_0^{(2),0}(x^b)/t^2,$$
$$\beta^{(2)} = (\ln t)^2 b_2^{(2)} + (\ln t)[\gamma_1^{(2),1} + \mathring{\gamma}_0^{(2),1}(x^b)/t^2] + [\gamma_1^{(2),0} + \mathring{\gamma}_0^{(2),0}(x^b)/t^2] \qquad (2.28)$$

into the second-order perturbation equations and collect together all the terms with the factors $(\ln t)^k$, where $k = 2,1,0$. Again we attempt to solve the equations by requiring that the coefficients of each independent power of $(\ln t)$ vanish separately. Demanding that the coefficients of $(\ln t)^2$ vanish leads to equations of the form (2.20), with $\{\psi^{(1)}, \beta^{(1)}\}$ replaced by $\{a_2^{(2)}, b_2^{(2)}\}$. The extended CK theorem applies and yields regular solutions for arbitrary initial data $\{a_2^{(2)}, b_2^{(2)}\}|_{t=0}$. Next we demand that the coefficients of $(\ln t)$ vanish. This leads to equations of the same type for $\{a_1^{(2)}, \gamma_1^{(2),1}\}$ except that there are some additional regular inhomogeneities arising from $\{a_2^{(2)}, b_2^{(2)}, \mathring{\gamma}_0^{(2),1}(x^b)\}$ and, in the $\gamma_1^{(2),1}$ equation, a collection of inhomogeneous terms with the *singular* coefficient $1/t^2$. This singularity can be killed, leaving only an additional regular inhomogeneity, by requiring that the (heretofore arbitrary) initial data for $b_2^{(2)}$ be constrained to satisfy an equation of the form

$$4 b_2^{(2)}|_{t=0} = [R\cdot d(\mathring{\gamma}_0^{(2),1}(x^b)) - 2R\mathring{\gamma}_0^{(2),1}(x^b) + R\,]|_{t=0}. \qquad (2.29)$$

With this restriction the extended CK theorem becomes applicable and assures the existence of regular solutions determined uniquely by arbitrary, analytic initial data $\{a_1^{(2)}, \gamma_1^{(2),1}\}|_{t=0}$ and arbitrarily chosen, analytic $\mathring{\gamma}_0^{(2),1}(x^b)$.

Finally we demand that the remaining terms in the second-order perturbation equations vanish. This also leads to inhomogeneous generalizations of Eqs. (2.20) for the quantities $\{\alpha_1^{(2),0}, \gamma_1^{(2),0}\}$ with inhomogeneities of the form $R + r/t^2 + r/t^4$ occurring in each of the equations. Here $r = r(x^b)$ stands generically for a time-independent, analytic factor and R is regular as before. One can kill the singular terms by imposing the conditions

$$4\,\mathring{\alpha}_0^{(2),0} = r,$$
$$-2\,\mathring{\gamma}_0^{(2),1} = r,$$
$$2\,a_2^{(2)}|_{t=0} = \{R\cdot\mathring{\alpha}_0^{(2),0} - 2R\mathring{\alpha}_0^{(2),0} + r\}|_{t=0}, \qquad (2.30)$$
$$2\,\gamma_1^{(2),1}|_{t=0} = \{-2b_2^{(2)} + R\cdot\mathring{\alpha}_0^{(2),0} + R\cdot(d(\mathring{\gamma}_0^{(2),0})) - 2R\mathring{\alpha}_0^{(2),0} + R\mathring{\gamma}_0^{(2),1} - 2R\mathring{\gamma}_0^{(2),0} + r\}|_{t=0},$$

leaving equations that are again amenable to the extended CK theorem. The latter have regular solutions uniquely determined by arbitrary, analytic data $\{\alpha_1^{(2),0}, \gamma_1^{(2),0}\}|_{t=0}$. This procedure leaves $\{\alpha_1^{(2),0}, \gamma_1^{(2),0}, a_1^{(2)}, \mathring{\gamma}_0^{(2),0}\}|_{t=0}$ completely arbitrary, which of course simply corresponds to the freedom to add an arbitrary solution of the first-order equations to any solution of the second-order equations.

We can describe the above procedure in a more logically ordered way as follows. First, choose the free data $\{\alpha_1^{(2),0}, \gamma_1^{(2),0}, a_1^{(2)}, \mathring{\gamma}_0^{(2),0}\}|_{t=0}$ arbitrarily; then use the first three of Eqs. (2.30) together with Eq. (2.29) to determine $\{\mathring{\alpha}_0^{(2),0}, \mathring{\gamma}_0^{(2),1}, a_2^{(2)}, b_2^{(2)}\}|_{t=0}$; and, finally, use the fourth of Eqs. (2.30) to fix $\gamma_1^{(2),1}|_{t=0}$. Having fixed all the initial data in such a way as to guarantee the cancellation of the singularities described above, solve in order for each of the pairs $\{a_2^{(2)}, b_2^{(2)}\}$, $\{a_1^{(2)}, \gamma_1^{(2),1}\}$, and $\{\alpha_1^{(2),0}, \gamma_1^{(2),0}\}$ by applying the extended CK theorem to the corresponding (desingularized) evolution equations.

C. Perturbations of arbitrary order

We now complete the proof of the inductive hypothesis, given in Eqs. (2.14) and (2.15), for the form of the nth-order perturbations. Since the previous section has already treated the first and second orders we need to consider all the orders $n \geqslant 3$ in the following.

We first need to determine the form of the inhomogeneous "source" terms in the nth-order equations that are induced by all the lower-order perturbations. Assuming the inductive hypothesis to be valid up through order $n-1$, it is straightforward to compute the form of the inhomogeneous terms generated in the nth-order equations. The results of that rather lengthy computation may be stated briefly as follows. Let $S_\psi^{(n)}$ represent the source term in the nth-order perturbation equation for ψ and $S_\beta^{(n)}$ represent the corresponding source term in the nth-order perturbation equation for β. These sources are certain polynomial expressions in the lower-order perturbations $\{\psi^{(k)}, \beta^{(k)}\}$ (for $k = 1,2,\ldots,n-1$) and their first and second derivatives and in the perturbations $\alpha^{(1)},\ldots,\alpha^{(n)}$ of the function α, which determines the lapse function N [cf. Eq. (2.13)]. For convenience we include the nth-order perturbation of α as a part of the source since the quantity $\alpha^{(n)}$ is specified arbitrarily and thus contributes a fixed inhomogeneity to the nth-order equations for $\{\psi^{(n)}, \beta^{(n)}\}$. For all $n \geqslant 3$ one finds that

$$S_\psi^{(n)} = (\ln t)^n R + (\ln t)^{n-1} R + \sum_{k=1}^{(n-1)/2} \left[\frac{(\ln t)^{n-2k} R}{t^{2k+2}} + \frac{(\ln t)^{n-2k-1} R}{t^{2k+2}}\right], \qquad (2.31)$$

$$S_\beta^{(n)} = (\ln t)^n R + (\ln t)^{n-1} \frac{R}{t^2} + \sum_{k=1}^{(n-1)/2} \left[\frac{(\ln t)^{n-2k} R}{t^{2k+2}} + \frac{(\ln t)^{n-2k-1} R}{t^{2k+4}}\right],$$

if n is odd, and

$$S_\psi^{(n)} = (\ln t)^n R + (\ln t)^{n-1} R + \sum_{k=1}^{(n-2)/2} \left[\frac{(\ln t)^{n-2k} R}{t^{2k+2}} + \frac{(\ln t)^{n-2k-1} R}{t^{2k+2}}\right] + R/t^{n+2}, \qquad (2.32)$$

$$S_\beta^{(n)} = (\ln t)^n R + (\ln t)^{n-1} R/t^2 + \sum_{k=1}^{(n-2)/2} \left[\frac{(\ln t)^{n-2k} R}{t^{2k+2}} + \frac{(\ln t)^{n-2k-1} R}{t^{2k+4}} \right] + R/t^{n+2}$$

if n is even. As before, R stands generically for a regular function.

The nth-order perturbation equations take the form

$$\psi^{(n)}{}_{,tt} + (1/t)\psi^{(n)}{}_{,t} = R\cdot\psi^{(n)} + t^2 R d\beta^{(n)} + tR\psi^{(n)}{}_{,t} + t^3 R\,\beta^{(n)}{}_{,t} + S_\psi^{(n)}, \tag{2.33}$$

$$\beta^{(n)}{}_{,tt} + (3/t)\beta^{(n)}{}_{,t} = R\cdot\psi^{(n)} + R\cdot d\beta^{(n)} + tR\psi^{(n)}{}_{,t} + tR\,\beta^{(n)}{}_{,t} + S_\beta^{(n)},$$

where the notation used is the same as that introduced for the first-order equations (2.20). If we now substitute the conjectured form for $\{\psi^{(n)},\beta^{(n)}\}$ [cf. Eqs. (2.14) and (2.15)] into the above equations and collect all the terms having the common logarithmic factors $(\ln t)^k$ (for $k = n, n-1,\ldots,1,0$) we may attempt to solve these equations by requiring that the coefficients of each independent power of $(\ln t)$ vanish separately.

Beginning with the coefficients of $(\ln t)^n$ we obtain the equations

$$a_n^{(n)}{}_{,tt} + (1/t)a_n^{(n)}{}_{,t} = R\cdot a_n^{(n)} + Rta_n^{(n)}{}_{,t} + t^2 R db_n^{(n)} + Rt^3 b_{n,t}^{(n)} + R, \tag{2.34}$$

$$b_n^{(n)}{}_{,tt} + (3/t)b_n^{(n)}{}_{,t} = R\cdot db_n^{(n)} + Rtb_n^{(n)}{}_{,t} + R\cdot a_n^{(n)} + tRa_n^{(n)}{}_{,t} + R.$$

The extended CK theorem applies to this system and guarantees the existence of regular solutions for arbitrary, analytic initial data $\{a_n^{(n)},b_n^{(n)}\}|_{t=0}$.

Turning to the coefficients of $(\ln t)^{n-1}$ we find it convenient to reexpress the (regular) function $b_{n-1}^{(n)}$ as

$$b_{n-1}^{(n)} = \mathring{\gamma}_0^{(n),n-1} + t^2\gamma_1^{(n),n-1}, \tag{2.35}$$

where $\gamma_1^{(n)n-1}$ is regular and where $\mathring{\gamma}_0^{(n),n-1}$ is analytic and *independent of* t (as signified by the overhead "naught"). Demanding that the coefficients of $(\ln t)^{n-1}$ vanish leads to the equations

$$\begin{aligned} &a_{n-1,tt}^{(n)} + (1/t)a_{n-1,t}^{(n)} + (2n/t)a_n^{(n)}{}_{,t} \\ &= R\cdot a_{n-1}^{(n)} + R(ta_{n-1,t}^{(n)} + na_n^{(n)}) + Rd(\mathring{\gamma}_0^{(n),n-1} + t^2\gamma_1^{(n),n-1}) + Rt^3\gamma_{1,t}^{(n),n-1} - 2R\mathring{\gamma}_0^{(n),n-1} \\ &\quad + nRt^2 b_n^{(n)} + R, \end{aligned} \tag{2.36}$$

$$\begin{aligned} &\gamma_{1,tt}^{(n),n-1} + \frac{3}{t}\gamma_{1,t}^{(n),n-1} + \frac{2n}{t^2}b_n^{(n)} + \frac{2n}{t}b_{n,t}^{(n)} \\ &= R\cdot d\gamma_1^{(n),n-1} + Rt\gamma_{1,t}^{(n),n-1} \\ &\quad + R\cdot a_{n-1}^{(n)} + R(ta_{n-1,t}^{(n)} + na_n^{(n)}) \\ &\quad + nRb_n^{(n)} + \frac{1}{t^2}[R\cdot d\mathring{\gamma}_0^{(n),n-1} + R - 2R\mathring{\gamma}_0^{(n),n-1}]. \end{aligned}$$

To cancel the singularity of order $1/t^2$ in the second equation we require that the (heretofore arbitrary) initial data for $b_n^{(n)}$ be fixed by the condition

$$2nb_n^{(n)}|_{t=0} = \{R\cdot d\mathring{\gamma}_0^{(n),n-1} + R - 2R\mathring{\gamma}_0^{(n),n-1}\}|_{t=0}. \tag{2.37}$$

With this in force (for arbitrary $\mathring{\gamma}_0^{(n),n-1}$) the above system becomes amenable to the extended CK theorem, which assures the existence of regular solutions determined by arbitrary analytic initial data $\{a_{n-1}^{(n)},\gamma_1^{(n),n-1}\}|_{t=0}$. The quantity $\mathring{\gamma}_0^{(n),n-1}$ also remains arbitrary at this point.

Continuing in this manner we write

$$\begin{aligned} a_{n-2k}^{(n)} &= \left(\sum_{l=0}^{k-1} t^{2l}\mathring{\alpha}_l^{(n),n-2k}\right) + t^{2k}\alpha_k^{(n),n-2k}, \\ b_{n-2k}^{(n)} &= \left(\sum_{l=0}^{k-1} t^{2l}\mathring{\gamma}_l^{(n),n-2k}\right) + t^{2k}\gamma_k^{(n),n-2k}, \\ a_{n-2k-1}^{(n)} &= \left(\sum_{l=0}^{k-1} t^{2l}\mathring{\alpha}_l^{(n),n-2k-1}\right) + t^{2k}\alpha_k^{(n),n-2k-1}, \\ b_{n-2k-1}^{(n)} &= \left(\sum_{l=0}^{k} t^{2l}\mathring{\gamma}_l^{(n),n-2k-1}\right) + t^{2k+2}\gamma_{k+1}^{(n),n-2k-1}, \end{aligned} \tag{2.38}$$

where $k = 1,2,\ldots,(n-1)/2$, if n is odd, and $k = 1,2,\ldots,(n-2)/2$, if n is even. For n even, we also write

$$\begin{aligned} a_0^{(n)} &= \left(\sum_{l=0}^{n/2-1} t^{2l}\mathring{\alpha}_l^{(n),0}\right) + t^n\alpha_{n/2}^{(n),0}, \\ b_0^{(n)} &= \left(\sum_{l=0}^{n/2-1} t^{2l}\mathring{\gamma}_l^{(n),0}\right) + t^n\gamma_{n/2}^{(n),0}, \end{aligned} \tag{2.39}$$

which extends the pattern of the first two of Eqs. (2.38) to the case $k = n/2$. In the above quantities $\{\mathring{\alpha}_l^{(n),m},\mathring{\gamma}_l^{(n),m}\}$ are all taken to be analytic and *independent of* t (signified as before by the overhead "naught") whereas the $\{\alpha_k^{(n),n-2k},\gamma_k^{(n),n-2k}, \alpha_k^{(n),n-2k-1},\gamma_{k+1}^{(n),n-2k-1}\}$ are expected to prove regular in the subsequent analysis.

Requiring that the coefficients of $(\ln t)^{n-2k}$ vanish leads to the equations

$$\begin{aligned} &t^{2k}\left(\alpha_{k,tt}^{(n),n-2k} + \frac{1}{t}\alpha_{k,t}^{(n),n-2k}\right) + \sum_{l=0}^{k-1} 4(l-k)^2 t^{2l-2}\mathring{\alpha}_l^{(n),n-2k} \\ &\quad + 2(n-2k+1)ta_{n-2k+1,t}^{(n)} + (4-4k)(n-2k+1)a_{n-2k+1}^{(n)} + (n-2k+1)(n-2k+2)a_{n-2k+2}^{(n)} \\ &= t^{2k}\{R\cdot\alpha_k^{(n),n-2k} + Rt\alpha_{k,t}^{(n),n-2k} + t^2 Rd\gamma_k^{(n),n-2k} + t^3 R\gamma_{k,t}^{(n),n-2k}\} + \sum_{l=0}^{k-1} t^{2l}R\cdot\mathring{\alpha}_l^{(n),n-2k} \\ &\quad + \sum_{l=0}^{k-1}(2l-2k)t^{2l}R\mathring{\alpha}_l^{(n),n-2k} + \sum_{l=0}^{k-1} t^{2l+2}Rd\mathring{\gamma}_l^{(n),n-2k} + \sum_{l=0}^{k-1}(2l-2k)t^{2l+2}R\mathring{\gamma}_l^{(n),n-2k} \\ &\quad + (n-2k+1)t^2 Ra_{n-2k+1}^{(n)} + (n-2k+1)Rt^2 b_{n-2k+1}^{(n)} + R/t^2, \end{aligned} \tag{2.40}$$

$$t^{2k}\left(\gamma_{k,tt}^{(n),n-2k}+\frac{3}{t}\gamma_{k,t}^{(n),n-2k}\right)+\sum_{l=0}^{k-1}4(l-k)(l-(k-1))t^{2l-2}\mathring{\gamma}_l^{(n),n-2k}$$
$$+[2(n-2k+1)/t]b^{(n)}_{n-2k+1,t}+(2-4k)[(n-2k+1)/t^2]b^{(n)}_{n-2k+1}+(n-2k+1)(n-2k+2)b^{(n)}_{n-2k+2}$$
$$=t^{2k}\{R\cdot d\gamma_k^{(n),n-2k}+Rt\gamma_{k,t}^{(n),n-2k}+R\cdot\alpha_k^{(n),n-2k}+Rt\alpha_{k,t}^{(n),n-2k}\}+\sum_{l=0}^{k-1}t^{2l}R\cdot d\mathring{\gamma}_l^{(n),n-2k}$$
$$+\sum_{l=0}^{k-1}(2l-2k)t^{2l}R\mathring{\gamma}_l^{(n),n-2k}+\sum_{l=0}^{k-1}t^{2l}R\cdot\mathring{\alpha}_l^{(n),n-2k}+\sum_{l=0}^{k-1}(2l-2k)t^{2l}R\mathring{\alpha}_l^{(n),n-2k}$$
$$+(n-2k+1)Rb^{(n)}_{n-2k+1}+(n-2k+1)t^2Ra^{(n)}_{n-2k+1}+R/t^2,$$

where $k=1,2,...,(n-1)/2$, if n is odd, and where $k=1,2,...,n/2$, if n is even (with $n\geqslant 3$ in either case).

We should like to divide these equations by t^{2k} and apply the extended CK theorem to find regular solutions. To do this, however, we must first cancel the inhomogeneities of order t^{-2}, $t^0,...,t^{2k-2}$ that occur in both equations, since otherwise division by t^{2k} would not yield a regular "source." If $k=1$, it is easy to see that we can always choose $\mathring{\alpha}_0^{(n),n-2}$ to cancel the terms of order t^{-2} in the equation for $\alpha_1^{(n),n-2}$. If $k\geqslant 2$, then a careful inspection of Eqs. (2.40) reveals that one can always choose the quantities

$$\begin{aligned}&\mathring{\alpha}_l^{(n),n-2k}, \quad l=0,1,...,k-1,\\ &\mathring{\gamma}_l^{(n),n-2k}, \quad l=0,...,k-2,\end{aligned} \tag{2.41}$$

in such a way as to cancel the inhomogeneous terms of order $t^{-2},...,t^{2k-4}$ in the $\alpha_k^{(n),n-2k}$ equations and the corresponding terms of order $t^{-2},...,t^{2k-6}$ in the $\gamma_k^{(n),n-2k}$ equations. One first solves for the $l=0$ coefficients $\{\mathring{\alpha}_0^{(n),n-2k}, \mathring{\gamma}_0^{(n),n-2k}\}$ and then proceeds successively to higher values of l. For all $k\geqslant 1$, we must still cancel the terms of order t^{2k-2} in the $\alpha_k^{(n),n-2k}$ equations and the terms of order t^{2k-4} and t^{2k-2} in the $\gamma_k^{(n),n-2k}$ equations.

Assume for the moment that we can cancel these remaining inhomogeneities (we shall show below how this is accomplished). Then one can apply the extended CK theorem to obtain regular solutions for arbitrary, analytic initial data $\{\alpha_k^{(n),n-2k},\gamma_k^{(n),n-2k}\}|_{t=0}$. The analytic quantity $\mathring{\gamma}_{k-1}^{(n),n-2k}$ would also remain arbitrary at this point. Recalling Eqs. (2.38) we see therefore that the coefficients of t^{2k} in $a^{(n)}_{n-2k}$ and the coefficients of t^{2k-2} and t^{2k} in $b^{(n)}_{n-2k}$ would remain arbitrary through this stage of the argument.

Returning to the mainstream of the proof we now demand that the coefficients of $(\ln t)^{n-2k-1}$ in the nth-order perturbation equations vanish separately. This leads to the equations

$$t^{2k}\left(\alpha_{k,tt}^{(n),n-2k-1}+\frac{1}{t}\alpha_{k,t}^{(n),n-2k-1}\right)+\sum_{l=0}^{k-1}4(l-k)^2t^{2l-2}\mathring{\alpha}_l^{(n),n-2k-1}$$
$$+[2(n-2k)/t]a^{(n)}_{n-2k,t}-(4k/t^2)(n-2k)a^{(n)}_{n-2k}+(n-2k)(n-2k+1)a^{(n)}_{n-2k+1}$$
$$=t^{2k}\{R\cdot\alpha_k^{(n),n-2k-1}+Rt\alpha_{k,t}^{(n),n-2k-1}+t^2Rd\gamma_{k+1}^{(n),n-2k-1}+t^3R\gamma_{k+1,t}^{(n),n-2k-1}\}$$
$$+\sum_{l=0}^{k-1}t^{2l}R\cdot\mathring{\alpha}_l^{(n)n-2k-1}+\sum_{l=0}^{k-1}(2l-2k)t^{2l}R\mathring{\alpha}_l^{(n)n-2k-1}+\sum_{l=0}^{k}t^{2l}Rd\mathring{\gamma}_l^{(n),n-2k-1}$$
$$+\sum_{l=0}^{k}(2l-(2k+2))t^{2l}R\mathring{\gamma}_l^{(n),n-2k-1}+(n-2k)Ra^{(n)}_{n-2k}+(n-2k)t^2Rb^{(n)}_{n-2k}+R/t^2, \tag{2.42}$$
$$t^{2k+2}\left(\gamma_{k+1,tt}^{(n),n-2k-1}+\frac{3}{t}\gamma_{k+1,t}^{(n),n-2k-1}\right)+\sum_{l=0}^{k-1}4(l-k)(l-(k+1))t^{2l-2}\mathring{\gamma}_l^{(n),n-2k-1}$$
$$+2(n-2k)tb^{(n)}_{n-2k,t}+(n-2k)(2-4k)b^{(n)}_{n-2k}+(n-2k)(n-2k+1)b^{(n)}_{n-2k+1}$$
$$=t^{2k+2}\{R\cdot d\gamma_{k+1}^{(n),n-2k-1}+Rt\gamma_{k+1,t}^{(n),n-2k-1}+R\cdot\alpha_k^{(n),n-2k-1}+Rt\alpha_{k,t}^{(n),n-2k-1}\}+\sum_{l=0}^{k}t^{2l}R\cdot d\mathring{\gamma}_l^{(n),n-2k-1}$$
$$+\sum_{l=0}^{k}(2l-(2k+2))t^{2l}R\mathring{\gamma}_l^{(n),n-2k-1}+\sum_{l=0}^{k-1}t^{2l+2}R\cdot\mathring{\alpha}_l^{(n),n-2k-1}+\sum_{l=0}^{k-1}(2l-2k)t^{2l+2}R\mathring{\alpha}_l^{(n),n-2k-1}$$
$$+(n-2k)Rt^2b^{(n)}_{n-2k}+(n-2k)Rt^2a^{(n)}_{n-2k}+R/t^2,$$

where $k=1,...,(n-1)/2$, if n is odd, and $k=1,...,(n-2)/2$, if n is even (and $n\geqslant 3$ in either case).

We should like to divide Eqs. (2.42) by t^{2k} and t^{2k+2} and apply the extended CK theorem to prove the existence of regular solutions. To do this, however, we must first cancel the inhomogeneities of order $t^{-2},t^0,...,t^{2k-2}$ that occur in the equation for $\alpha_k^{(n),n-2k-1}$ and the inhomogeneities of order $t^{-2},t^0,...,t^{2k}$ that occur in the equation for $\gamma_{k+1}^{(n),n-2k-1}$ since otherwise the remaining "source" terms (after division by t^{2k} and t^{2k+2}, respectively) would not be regular. A careful inspection of Eqs. (2.42), however, reveals that we can always choose the quantities

$$\mathring{\alpha}_l^{(n),n-2k-1}, \quad l=0,\dots,k-1, \qquad \mathring{\gamma}_l^{(n),n-2k-1}, \quad l=0,\dots,k-1, \tag{2.43}$$

in such a way as to cancel the inhomogeneous terms of order $t^{-2},\dots,t^{2k-4}$ occurring in both of these equations. One first solves for the quantities $\{\mathring{\alpha}_0^{(n),n-2k-1},\mathring{\gamma}_0^{(n),n-2k-1}\}$ and then proceeds successively to higher values of l. We must still cancel the inhomogeneities of order t^{2k-2} in the equation for $\alpha_k^{(n),n-2k-1}$ and those of order t^{2k-2} and t^{2k} in the equation for $\gamma_{k+1}^{(n),n-2k-1}$. If we can succeed in canceling these remaining inhomogeneous terms then we shall be able to apply the extended CK theorem to prove the existence of regular solutions of Eqs. (2.42) determined by arbitrary, analytic initial data $\{\alpha_k^{(n),n-2k-1},\gamma_{k+1}^{(n),n-2k-1}\}|_{t=0}$. In addition, the quantities $\mathring{\gamma}_k^{(n),n-2k-1}$ would also remain arbitrary at this point. Recalling Eqs. (2.38) we thus see that the coefficients of t^{2k} in $a_{n-2k-1}^{(n)}$ and the coefficients of t^{2k} and t^{2k+2} in $b_{n-2k-1}^{(n)}$ would therefore still remain arbitrary at this stage in the argument.

Now we shall show that one can cancel the remaining troublesome inhomogeneities by appropriately choosing the free data that rests at our disposal. We have yet to cancel the terms of order t^{2k-2} in the equation for $\alpha_k^{(n),n-2k}$ and the terms of order t^{2k-4} and t^{2k-2} in the equation for $\gamma_k^{(n),n-2k}$. However, the coefficient of t^{2k-2} in the quantity $a_{n-2k+2}^{(n)}$, which occurs on the left-hand side of Eq. (2.40) with a constant coefficient, remains at our disposal for $k\geqslant 1$. We therefore choose this heretofore free data to cancel the remaining singularities in the $\alpha_k^{(n),n-2k}$ equations. Furthermore, since the coefficients of t^{2k-2} and t^{2k} in $b_{n-2k+1}^{(n)}$ remain at our disposal for $k\geqslant 1$, we can choose these in such a way that the terms

$$(1/t)b_{n-2k+1,t}^{(n)} + [(1-2k)/t^2]b_{n-2k+1}^{(n)}$$

which occur on the left-hand side of Eq. (2.40) with a nonzero constant coefficient, cancel the inhomogeneities of order t^{2k-4} and t^{2k-2} in the $\gamma_k^{(n),n-2k}$ equation.

Finally, we need to cancel the inhomogeneous terms of order t^{2k-2} in the equation for $\alpha_k^{(n),n-2k-1}$ and those of order t^{2k-2} and t^{2k} in the equation for $\gamma_{k+1}^{(n),n-2k-1}$. However, the coefficient of t^{2k-2} in $a_{n-2k+1}^{(n)}$ [which occurs with nonvanishing coefficient on the left-hand side of Eq. (2.42)] remains at our disposal. We therefore choose this data to cancel the remaining troublesome singularity in the $\alpha_k^{(n),n-2k-1}$ equation. Furthermore the coefficients of t^{2k} and t^{2k-2} in $b_{n-2k}^{(n)}$ remain at our disposal. One can choose this (heretofore arbitrary) data in such a way that the quantity $tb_{n-2k,t}^{(n)} + (1-2k)b_{n-2k}^{(n)}$, which occurs with nonvanishing coefficient on the left-hand side of Eq. (2.42), cancels the singularities of order t^{2k-2} and t^{2k} in the $\gamma_{k+1}^{(n),n-2k-1}$ equation.

III. SOLVING THE PERTURBED CONSTRAINT EQUATIONS

A. Constraints and Bianchi identities

The Hamiltonian and momentum constraints (written $\mathscr{H}$ and $\mathscr{H}_i$, respectively) of the usual Arnowitt, Deser, and Misner (ADM) formalism[7] are given explicitly (in the unprimed coordinate chart) by

$$\begin{aligned}
\mathscr{H} &= te^{\phi}\sqrt{{}^{(2)}g}\,\Big\{\frac{1}{4N^2}\Big[g^{ac}g^{bd}g_{cd,t}g_{ab,t} - \frac{4}{t}g^{cd}g_{cd,t} - g^{cd}g_{cd,t}g^{ef}g_{ef,t}\Big] + \frac{2}{N^2}\Big(\phi_{,t}\phi_{,t} + \frac{2}{t}\phi_{,t}\Big) \\
&\quad + \frac{1}{2}\frac{e^{4\phi}t^2}{N^2}g^{cd}\beta_{c,t}\beta_{d,t} - {}^{(2)}R + 2g^{ab}\phi_{,a}\phi_{,b} + \frac{t^2e^{4\phi}g^{ac}g^{bd}}{4}(\beta_{a,b}-\beta_{b,a})(\beta_{c,d}-\beta_{d,c})\Big\}, \\
\mathscr{H}_3 &= -\frac{\partial}{\partial x^a}\Big[\frac{t^3\sqrt{{}^{(2)}g}}{N}e^{4\phi}g^{ab}\beta_{b,t}\Big], \\
\mathscr{H}_a &= \beta_a\mathscr{H}_3 - 2\Big\{\nabla_b\Big[\frac{\sqrt{{}^{(2)}g}}{2N}(g^{bd}tg_{ad,t} - t\delta_a^b g^{cd}g_{cd,t})\Big] \\
&\quad - \Big(\frac{2\sqrt{{}^{(2)}g}}{N}\Big)\phi_{,a}(t\phi_{,t}+1) + \frac{\sqrt{{}^{(2)}g}}{N^2}N_{,a} + \frac{1}{2}\frac{t^3\sqrt{{}^{(2)}g}}{N}e^{4\phi}g^{bd}\beta_{d,t}(\beta_{a,b}-\beta_{b,a})\Big\}.
\end{aligned} \tag{3.1}$$

Here ${}^{(2)}g = \det(g_{ab})$, g^{ab} is the inverse of g_{ab}, ∇_a is the covariant derivative with respect to g_{ab}, and ${}^{(2)}R$ is the scalar curvature of this metric.

Equations of motion for the quantities $\{\mathscr{H},\mathscr{H}_i\}$ are implied by the Bianchi identity and, when our evolution equations are imposed, take the form[8]

$$\begin{aligned}
&\frac{\partial\mathscr{H}}{\partial t} + \Big(\frac{1}{t} - \phi_{,t} + \frac{1}{2}g^{ab}g_{ab,t}\Big)\mathscr{H} \\
&\quad = 2e^{\phi}g^{ab}N_{,b}(\mathscr{H}_a - \beta_a\mathscr{H}_3) \\
&\qquad + e^{\phi}N(g^{ab}(\mathscr{H}_a - \beta_a\mathscr{H}_3))_{,b}, \\
&\frac{\partial\mathscr{H}_3}{\partial t} = 0, \\
&\frac{\partial\mathscr{H}_a}{\partial t} = N_{,a}(e^{-\phi}\mathscr{H}) + \nabla_a(Ne^{-\phi}\mathscr{H}).
\end{aligned} \tag{3.2}$$

These equations have a slightly different form from those usually presented because we have derived them from a Hamiltonian in which the lapse function N depends explicitly upon the dynamical variables [through the coordinate condition (2.6) above] and the shift vector is zero. Any regular solution of the evolution equations [Eqs. (2.4) of Ref.

1] that, in addition, satisfies the regularity condition $\mathring{N} = e^{2\phi}$ (discussed in Sec. II A above) yields analytic expressions for $\{\mathscr{H}, \mathscr{H}_i\}$ that, moreover, vanish as $t \to 0$. However, the extended CK theorem applies to the system (3.2) and shows that the unique analytic solution of these equations having vanishing initial data is, in fact, identically vanishing. This result was used in proving the basic existence and uniqueness theorem for generalized Taub–NUT space-times given by Theorem (2) of Ref. 1.

B. Solving the linearized constraint equations

Recalling the results of Sec. II B let us write the first-order perturbations of $\{\phi, g_{ab}, \beta_a\}$ in the form

$$\phi^{(1)} = (\ln t) h^{(1),1} + h^{(1),0}, \quad g_{ab}^{(1)} = (\ln t) h_{ab}^{(1),1} + h_{ab}^{(1),0},$$
$$\beta_a^{(1)} = (\ln t) h_a^{(1),1} + (1/t^2) h_a^{(1),0}, \tag{3.3}$$

where, as before $\{h^{(1),1}, h^{(1),0}, h_{ab}^{(1),1}, h_{ab}^{(1),0}, h_a^{(1),1}, h_a^{(1),0}\}$ are each analytic and even in t (i.e., regular). Recalling also Eq. (2.13) we see that the first-order perturbation of the lapse function N can be written

$$\delta N = 2N\lambda^{(1)} + \tfrac{1}{2} N g^{ab} g_{ab}^{(1)}, \tag{3.4}$$

where $\lambda^{(1)}$ is an arbitrary analytic, time-independent function on K.

Linearizing the constraint equations $\mathscr{H} = \mathscr{H}_i = 0$ about an arbitrary generalized Taub–NUT background solution and substituting the perturbations (3.3) and (3.4) into the linearized expressions we find that the necessary and sufficient conditions for the vanishing of $\delta\mathscr{H}$ and $\delta\mathscr{H}_i$ as $t \to 0$ are

$$(4h^{(1),1} - g^{ab} h_{ab}^{(1),1})|_{t=0} = 0,$$
$$\left(\frac{\partial}{\partial x^a}\left[\frac{2\sqrt{^{(2)}g}}{N} e^{4\phi} g^{ab} h_b^{(1),0}\right]\right)\Bigg|_{t=0} = 0,$$
$$\Bigg\{2e^{-2\phi}\sqrt{^{(2)}g}\,\frac{\partial}{\partial x^a}\left(-2\lambda^{(1)} + 2h^{(1),0} - \frac{1}{2} g^{cd} h_{cd}^{(1),0}\right) - \nabla_b\left[\frac{\sqrt{^{(2)}g}}{N}(g^{bd} h_{ad}^{(1),1} - \delta_a^b g^{cd} h_{cd}^{(1),1})\right]$$
$$+ \frac{4\sqrt{^{(2)}g}}{N}\phi_{,a} h^{(1),1} + \frac{2\sqrt{^{(2)}g}}{N} e^{4\phi} g^{bd}(\beta_{a,b} - \beta_{b,a}) h_d^{(1),0}\Bigg\}\Bigg|_{t=0} = 0. \tag{3.5}$$

In fact, conditions (3.5) are actually sufficient to guarantee the vanishing of the linearized constraints $\{\delta\mathscr{H}, \delta\mathscr{H}_i\}$ for all t (in the interval of existence of the perturbations). To see this we first consider the terms in $\{\delta\mathscr{H}, \delta\mathscr{H}_i\}$ that contain the factor $(\ln t)$. The coefficients of this factor are precisely the linearized constraint operators acting on the (regular) quantities $\{h^{(1),1}, h_{ab}^{(1),1}, h_a^{(1),1}\}$, which occur (as coefficients of $\ln t$) in the perturbations (3.3). However, as we showed in Sec. II B, these quantities are regular solutions of the homogeneous, linearized evolution equations. It follows from linearizing Eqs. (3.2) about a GTN background and applying the extended CK theorem to this system that the linearized constraints (evaluated upon such a regular perturbation) vanish identically if and only if they vanish at $t = 0$. The vanishing at $t = 0$, however, is ensured by the first of the three conditions (3.5) above. Thus the coefficients of $\ln t$ in the linearized constraints $\{\delta\mathscr{H}, \delta\mathscr{H}_i\}$ vanish separately (for all t in the interval of existence of the perturbations) leaving purely analytic expressions for these quantities, which, moreover, vanish at $t = 0$ by virtue of the three conditions (3.5) imposed above. Once again the extended CK theorem may be applied to the linearized form of Eqs. (3.2) to conclude that $\delta\mathscr{H}$ and $\delta\mathscr{H}_i$ vanish identically on the domain of existence of the perturbations.

Thus the solution of the linearized constraints reduces precisely to the solutions of Eqs. (3.5) that constrain the choice of initial data at $t = 0$. The first of these equations merely fixes the trace of the (heretofore unrestricted) quantity $h_{ab}^{(1),1}|_{t=0}$ in terms of $h^{(1),1}|_{t=0}$. The second of Eqs. (3.5) merely requires that the (heretofore unconstrained) one-form $((e^{4\phi}/N) h_b^{(1),0})|_{t=0}$ have vanishing divergence (relative to the metric $g_{ab}|_{t=0}$). The third of Eqs. (3.5) can [upon making use of the background regularity condition

$$Ne^{-2\phi}|_{t\to 0} \to 1$$

and the first of Eqs. (3.5)] be reexpressed in the form

$$\Bigg\{\frac{1}{2}\frac{\hat{g}_{ac}}{\sqrt{^{(2)}\hat{g}}}\hat{\nabla}_b(\gamma^{bc}) - e^{2\phi}\hat{g}^{bd}(\beta_{a,b} - \beta_{b,a}) h_d^{(1),0}$$
$$- \frac{\partial}{\partial x^a}\Bigg(-2\lambda^{(1)} + h^{(1),1} + 2h^{(1),0}$$
$$- \frac{1}{2} g^{cd} h_{cd}^{(1),0}\Bigg)\Bigg\}\Bigg|_{t=0} = 0, \tag{3.6}$$

where

$$\gamma^{bc} = \sqrt{^{(2)}g}(g^{ac} g^{bd} h_{ad}^{(1),1} - \tfrac{1}{2} g^{bc} g^{ad} h_{ad}^{(1),1}),$$
$$\hat{g}_{ab} = e^{-2\phi} g_{ab}, \tag{3.7}$$
$$\sqrt{^{(2)}\hat{g}} = e^{-2\phi}\sqrt{^{(2)}g},$$

and where $\hat{\nabla}_b$ signifies the covariant derivative with respect to $\hat{g}_{ab}$.

The quantity γ^{ab} is a symmetric, traceless tensor density defined on the two-manifold K (we have suppressed the restriction to $t = 0$ to simplify the notation). Any such field can be uniquely decomposed into L^2-orthogonal summands of the form[9,10]

$$\gamma^{ab} = \gamma^{ab\,\mathrm{TT}} + \sqrt{^{(2)}\hat{g}}(\hat{\nabla}^a Y^b + \hat{\nabla}^b Y^a - \hat{g}^{ab}\hat{\nabla}_c Y^c), \tag{3.8}$$

where $\hat{\nabla}^b \equiv \hat{g}^{bc}\hat{\nabla}_c$, etc., and where the "transverse traceless" summand $\gamma^{ab\,\mathrm{TT}}$ satisfies

$$\hat{g}_{ab}\gamma^{ab\,\mathrm{TT}} = \hat{\nabla}_b \gamma^{ab\,\mathrm{TT}} = 0. \tag{3.9}$$

The vector field $Y = Y^a \partial/\partial x^a$ is determined by solving the linear elliptic equation obtained by computing the divergence of Eq. (3.8). By standard linear elliptic methods[11] one shows that this equation always has a solution unique up to the addition of a conformal Killing field of $(K, \hat{g}_{ab})$. Since only the conformal Killing form of Y occurs in Eq. (3.8) the decomposition of γ^{ab} is unique even if Y is not.

For any compact, orientable Riemannian two-manifold $(K, \hat{g}_{ab})$ the space of transverse traceless symmetric tensor (densities) is finite dimensional and represents (roughly speaking) the tangent space to Teichmüller space at the point represented by $\hat{g}_{ab}$.[12] In particular if $K \approx S^2$ then $\gamma^{ab\,\mathrm{TT}}$ vanishes identically, if $K \approx T^2$ then $\gamma^{ab\,\mathrm{TT}}$ belongs to a two-dimensional space, and if K is diffeomorphic to a manifold of genus $g \geqslant 2$ then $\gamma^{ab\,\mathrm{TT}}$ belongs to a $(6g-6)$-dimensional space.

Substituting the decomposition (3.8) into the constraint equation (3.6) one gets

$$\begin{aligned}\{\tfrac{1}{2}\hat{g}_{ac}\hat{\nabla}_b[\sqrt{{}^{(2)}\hat{g}}(\hat{\nabla}^b Y^c + \hat{\nabla}^c Y^b - \hat{g}^{bc}\hat{\nabla}_d Y^d)]\}|_{t=0} \\ = \{e^{2\phi}\sqrt{{}^{(2)}g}g^{bd}(\beta_{a,b} - \beta_{b,a})h^{(1),0}_{d} + \sqrt{{}^{(2)}\hat{g}}(-2\lambda^{(1)} \\ + h^{(1),1} + 2h^{(1),0} - \tfrac{1}{2}g^{cd}h^{(1),0}_{cd})_{,a}\}|_{t=0}. \qquad (3.10)\end{aligned}$$

The self-adjoint linear elliptic operator on the left-hand side of this equation has precisely the conformal Killing fields of $(K, \hat{g}_{ab})$ as its kernel. By standard linear elliptic theory, therefore, Eq. (3.10) has a solution (unique only up to the addition of a conformal Killing field) if and only if the source term on the right-hand side satisfies the integrability conditions

$$\begin{aligned}\int_K \{e^{2\phi}\sqrt{{}^{(2)}g}g^{bd}Z^a(\beta_{a,b} - \beta_{b,a})h^{(1),0}_{d}\}|_{t=0} \\ + \int_K \Big\{\sqrt{{}^{(2)}\hat{g}}(\hat{\nabla}_a Z^a)\Big(2\lambda^{(1)} - h^{(1),1} \\ - 2h^{(1),0} + \frac{1}{2}g^{cd}h^{(1),0}_{cd}\Big)\Big\}\Big|_{t=0} = 0, \qquad (3.11)\end{aligned}$$

for all $Z = Z^a \partial/\partial x^a$ such that

$$(\hat{\nabla}^a Z^b + \hat{\nabla}^b Z^a - \hat{g}^{ab}\hat{\nabla}_c Z^c)|_{t=0} = 0. \qquad (3.12)$$

Since the transverse (i.e., divergence-free) part of

$$((e^{4\phi}/N)h^{(1),0}_{b})|_{t=0} = (e^{2\phi}h^{(1),0}_{b})|_{t=0}$$

remains at our disposal along with $\lambda^{(1)}$, $h^{(1),1}$, etc., we can always satisfy the finite number of integrability conditions by imposing the integral conditions in Eq. (3.11) upon this free data.

Solving Eq. (3.10) for Y uniquely determines the "longitudinal" part of γ^{ab} in the decomposition (3.8) and thus completes the solution of the linearized constraint equations.

C. Singular gauge transformations and gauge conditions

In Sec. II A we considered analytic diffeomorphisms of the background space-times generated by vector fields of the form

$${}^{(4)}X = t'Y\frac{\partial}{\partial t'} + X^{a'}\frac{\partial}{\partial x^{a'}} + X^{3'}\frac{\partial}{\partial x^{3'}}, \qquad (3.13)$$

where $(Y, X^{a'}, X^{3'})$ were required to be analytic in $(t', x^{a'})$. Since we are now studying singular perturbations of these same backgrounds we must also consider singular gauge transformations (i.e., infinitesimal space-time diffeomorphisms) since the latter can masquerade as nontrivial singular perturbations. To do this we relax the requirement that $(Y, X^{a'}, X^{3'})$ be nonsingular and demand only that ${}^{(4)}X$ preserve the coordinate conditions that we have imposed on ${}^{(4)}g$ throughout.

This requirement leads to the following system of differential equations for the vector field components $(Y, X^{a'}, X^{3'})$:

$$\begin{aligned} tY_{,t} &= \frac{2N}{\sqrt{{}^{(2)}g}}\frac{\partial}{\partial x^a}\left(\frac{\sqrt{{}^{(2)}g}}{N}X^{a'}\right) + 4\lambda^{(1)}, \\ X^{a'}{}_{,t} &= (tN^2/2)g^{ab}Y_{,b}, \qquad (3.14) \\ (X^{3'} + Y)_{,t} &= -2\beta_a X^{a'}{}_{,t}, \end{aligned}$$

where $\lambda^{(1)}$ is, as before, an arbitrary time-independent analytic function of the $\{x^a\}$. Since both $\lambda^{(1)}$ and $N/\sqrt{{}^{(2)}g}$ are independent of t we can time differentiate the first of Eqs. (3.14) and appeal to the second of these equations to reexpress the result as a wave equation for the function Y:

$$Y_{,tt} + \frac{1}{t}Y_{,t} = \frac{N}{\sqrt{{}^{(2)}g}}\frac{\partial}{\partial x^a}(\sqrt{{}^{(2)}g}Ng^{ab}Y_{,b}). \qquad (3.15)$$

By means of the same techniques developed in Sec. II B we can show that the general solution of Eq. (3.15) (in the analytic case) has the form

$$Y = (\ln t)y_0 + y_1, \qquad (3.16)$$

where y_0 and y_1 are both regular and determined uniquely by their initial values prescribed at $t = 0$.

Substituting the result (3.16) into the second of Eqs. (3.14) and reexpressing the result slightly leads to an equation of the form

$$X^{a'}{}_{,t} = t(\ln t)\delta^a_{(0)} + t\delta^a_{(1)}, \qquad (3.17)$$

where $\delta^a_{(0)}$ and $\delta^a_{(1)}$ are both regular. Integrating this one finds that $X^{a'}$ has the form

$$X^{a'} = t^2(\ln t)\lambda^{a'}_{(0)} + \lambda^{a'}_{(1)}, \qquad (3.18)$$

where $\lambda^{a'}_{(0)}$ and $\lambda^{a'}_{(1)}$ are both regular and given explicitly by the formulas

$$\begin{aligned} t^2\lambda^{a'}_{(0)} &= \int_0^t s\delta^a_{(0)}(s, x^b)\,ds, \\ \lambda^{a'}_{(1)} &= \int_0^t s(\delta^a_{(1)} - \lambda^{a'}_{(0)})(s, x^b)\,ds + \mathring{\lambda}^{a'}_{(1)}(x^b), \end{aligned} \qquad (3.19)$$

where $\mathring{\lambda}^{a'}_{(1)}(x^b)$ is arbitrary, analytic initial data for $\lambda^{a'}_{(1)}$ (and hence for $X^{a'}$).

Substituting this result into the third of Eqs. (3.14) and reexpressing the result slightly leads to an equation of the form

$$(X^{3'} + Y)_{,t} = t(\ln t)\rho_{(0)} + t\rho_{(1)}, \qquad (3.20)$$

where $\rho_{(0)}$ and $\rho_{(1)}$ are regular. Integrating this as above leads to the expression

$$X^{3'} + Y = t^2(\ln t)\gamma_{(0)} + \gamma_{(1)}, \qquad (3.21)$$

where $\gamma_{(0)}$ and $\gamma_{(1)}$ are regular and given by

$$t^2\gamma_{(0)} = \int_0^t s\rho_{(0)}(s,x^b)ds,$$
$$\gamma_{(1)} = \int_0^t s(\rho_{(1)} - \gamma_{(0)})(s,x^b)ds + \mathring{\gamma}_{(1)}(x^b), \tag{3.22}$$

in which $\mathring{\gamma}_{(1)}(x^b)$ is arbitrary, analytic initial data for $\gamma_{(1)}$ (and hence for $X^{3'} + Y$).

The wave equation imposed on Y together with the equation for $X^{a'}$ suffice to ensure that the quantity

$$4\lambda^{(1)} = tY_{,t} - \frac{2N}{\sqrt{^{(2)}g}}\frac{\partial}{\partial x^a}\left(\frac{\sqrt{^{(2)}g}}{N}X^{a'}\right) \tag{3.23}$$

is indeed independent of t as required in the first of Eqs. (3.14). Thus we need only evaluate Eq. (3.23) at $t\to 0$ to determine the relationship between $\lambda^{(1)}$ and the free data $\{y_0, y_1, X^{a'}, X^{3'} + Y\}|_{t=0}$. The result is

$$4\lambda^{(1)} = \left\{y_0 - \left(\frac{2N}{\sqrt{^{(2)}g}}\frac{\partial}{dx^a}\left(\frac{\sqrt{^{(2)}g}}{N}X^{a'}\right)\right)\right\}\Bigg|_{t=0}. \tag{3.24}$$

Having obtained the general form of a vector field that preserves the assumed coordinate conditions we can compute the induced gauge transformations of the first-order perturbations $\{\phi^{(1)}, g^{(1)}_{ab}, \beta^{(1)}_a, \delta N\}$. Since these quantities are determined by data prescribed at $t = 0$ it suffices to compute the gauge transformations of this "initial data." Recalling the parametrization of the first-order perturbations given in Eq. (3.3) one finds the induced gauge transformations are given by

$$\begin{aligned}
\delta h^{(1),1}|_{t=0} &= (y_0/2)|_{t=0},\\
\delta h^{(1),0}|_{t=0} &= ((y_1/2) + \mathring{X}^{a'}\phi_{,a})|_{t=0},\\
\delta h^{(1),1}_{ab}|_{t=0} &= (g_{ab}y_0)|_{t=0},\\
\delta h^{(1),0}_{ab}|_{t=0} &= (g_{ab}y_1 + (\mathscr{L}_{(2)X'}g)_{ab})|_{t=0},\\
\delta h^{(1),1}_{a}|_{t=0} &= 0,
\end{aligned} \tag{3.25}$$

and, if we express $h^{(1),0}_a$ in the form

$$h^{(1),0}_a = \mathring{\gamma}^{(1),0}_a(x^b) + t^2\gamma^{(1),1}_a, \tag{3.26}$$

where $\mathring{\gamma}^{(1),0}_a$ is analytic and time independent and where $\gamma^{(1),1}_a$ is regular,

$$\begin{aligned}
\delta\mathring{\gamma}^{(1),0}_a &= 0,\\
\delta\gamma^{(1),1}_a|_{t=0} &= \{\tfrac{1}{2}(X^{3'} + Y)_{,a} + (\mathscr{L}_{^{(2)}\mathring{X}'}\beta)_a\}|_{t=0}.
\end{aligned} \tag{3.27}$$

Here $\mathscr{L}_{^{(2)}\mathring{X}'}$ designates the Lie derivative with respect to the vector field $\mathring{X}^{a'}\,\partial/\partial x^a$.

If, recalling Eq. (3.4), we express the perturbed lapse function as

$$\delta N = (\ln t)n_{(0)} + n_{(1)}, \tag{3.28}$$

then we find that $n_{(0)}$ and $n_{(1)}$ undergo the gauge transformations

$$\begin{aligned}
\delta n_{(0)}|_{t=0} &= (Ny_0)|_{t=0}\\
\delta n_{(1)}|_{t=0} &= (Ny_1 + (N/2)y_0 + \mathscr{L}_{^{(2)}\mathring{X}'}N)|_{t=0}.
\end{aligned} \tag{3.29}$$

If we let ω represent the linear perturbation of $Ne^{-2\phi}$, i.e.,

$$\omega = \delta Ne^{-2\phi} - 2Ne^{-2\phi}\phi^{(1)}, \tag{3.30}$$

then one can easily show, using the results above, that ω undergoes the gauge transformation

$$\delta\omega|_{t=0} = \tfrac{1}{2}y_0|_{t=0}. \tag{3.31}$$

The foregoing results could be used to impose suitable gauge conditions upon certain of the unconstrained initial data. Rather than pursue that issue here, however, we shall use the above results to study the possible existence of additional Killing fields of the background space-time metric $^{(4)}g$ and the characterization of such Killing fields in terms of the background initial data. The possible occurrence of such extra Killing fields is important for the study of the higher-order perturbations of the constraint equations since it is directly related to the "linearization stability problem" for the constraints as discussed in Refs. 5 and 6.

A vector field $^{(4)}X$ induces gauge transformations of $^{(4)}g$ through Lie differentiation, $\mathscr{L}_{^{(4)}X}\,{}^{(4)}g$. If $^{(4)}X$ is a Killing field of $^{(4)}g$ it thus induces purely vanishing gauge transformations of $^{(4)}g$. From the first of Eqs. (3.25) it follows that a Killing field must have $y_0|_{t=0} = 0$. However, y_0 is uniquely determined (as a regular solution of the wave equation) from $y_0|_{t=0}$ and thus, in this case, must vanish identically. This in turn implies $\delta^a_{(0)} \equiv 0$ in Eq. (3.17) and thus that $\lambda^{a'}_{(0)} \equiv 0$ in Eq. (3.18). Thus Y and $X^{a'}$ both reduce to regular quantities. This in turn implies that $\rho_{(0)} \equiv 0$ in Eq. (3.20) and hence that $\gamma_{(0)} \equiv 0$ in Eq. (3.21). Thus $X^{3'} + Y$ also reduces to a regular quantity. It follows that any Killing field $^{(4)}X$ of $^{(4)}g$ must take the nonsingular form already considered in Sec. II A. The gauge transformations induced by such a vector field are therefore regular solutions of the linearized evolution equation that vanish identically if and only if they have vanishing initial data. Recalling Eq. (2.9) we thus see that $^{(4)}X$ is a Killing field of $^{(4)}g$ if and only if its initial data $\{Y, X^{a'}, X^{3'}\}|_{t=0}$ satisfy

$$\begin{aligned}
&\{\tfrac{1}{2}Y + \mathscr{L}_{^{(2)}\mathring{X}'}\phi\}|_{t=0} = 0,\\
&\{\tfrac{1}{2}(X^{3'} + Y)_{,a} + (\mathscr{L}_{^{(2)}\mathring{X}'}\beta)_a\}|_{t=0} = 0,\\
&\{Yg_{ab} + (\mathscr{L}_{^{(2)}\mathring{X}'}g)_{ab}\}|_{t=0} = 0,\\
&\{YN + \mathscr{L}_{^{(2)}\mathring{X}'}N\}|_{t=0} = 0.
\end{aligned} \tag{3.32}$$

Since $(N - e^{2\phi})|_{t=0} = 0$, the last of these four equations is redundant, being equivalent to the first.

Equations (3.32) always admit the "trivial" solution $Y = X^{a'} = 0$, $X^{3'} = 1$, since $^{(4)}X = \partial/\partial x^{3'}$ is always a Killing field of our background space-times. We shall see in Sec. III E, however, that the occurrence or nonoccurrence of "nontrivial" Killing fields plays a key role in the analysis of the higher-order perturbed constraint equations. Roughly speaking, there are no obstructions to solving the higher-order constraint equations unless nontrivial Killing fields of $^{(4)}g$ exist. If such Killing fields do exist then additional integrability conditions must be imposed upon the first-order perturbations in order to proceed to a solution of the higher-order constraint equations. These are simply the "linearization stability" constraints that arise when one attempts to perturb from one (Killing) symmetry class to another (cf. Refs. 5 and 6).

D. Symplectic structure and canonically conjugate perturbations

The evolution equations discussed in Sec. II are Hamiltonian, having been derived from an ADM variational principle after a suitable choice of lapse and shift. It follows from general properties of Hamiltonian systems that one can define a symplectic form that yields a conserved, antisymmetric contraction of any pair of solutions of the corresponding linearized equations of motion. This form will be conserved whether or not we choose to impose the linearized constraint equations, provided our perturbations leave the Hamiltonian itself fixed. More precisely, we must require that the function λ, occurring in the coordinate condition [cf. Eq. (2.13)]

$$N/\sqrt{{}^{(2)}g} = (N/\sqrt{{}^{(2)}g})|_{\epsilon=0}e^{2\lambda(x^a,\epsilon)}, \tag{3.33}$$

remain unperturbed from its (vanishing) original value if we wish to have a conserved symplectic form for arbitrary pairs of solutions of the linearized evolution equation. This is so because a nonvanishing variation of λ induces a corresponding variation of the Hamiltonian itself through its induced perturbation of the lapse function.

One can evaluate the symplectic form on any pair of perturbations ${}^{(4)}h$ and ${}^{(4)}h'$ that satisfy the linearized evolution equations about a given background ${}^{(4)}g$. Since the resulting quantity, here designated by $\omega({}^{(4)}h, {}^{(4)}h')$, is conserved, it has a limit as $t\to 0$ even for singular perturbations (provided we take $\lambda^{(1)} = \lambda^{(1)\prime} = 0$ in lieu of imposing the linearized constraints).

A lengthy but straightforward evaluation of the symplectic form yields [in the notation of Eqs. (3.3) and (3.26) and taking $\lambda^{(1)} = \lambda^{(1)\prime} = 0$]

$$\omega({}^{(4)}h,{}^{(4)}h') = \int_K \frac{\sqrt{{}^{(2)}g}}{N}\{4(h^{(1),0}h^{(1),1\prime} - h^{(1),0\prime}h^{(1),1}) + 2e^{4\phi}g^{ab}(\gamma_a^{(1),0}\gamma_b^{(1),1\prime} - \gamma_a^{(1),0\prime}\gamma_b^{(1),1})$$
$$+ \tfrac{1}{2}(g^{ac}g^{bd} - g^{ab}g^{cd})(h_{ab}^{(1),0}h_{cd}^{(1),1\prime} - h_{ab}^{(1),0\prime}h_{cd}^{(1),1})\}|_{t=0}. \tag{3.34}$$

If we identify the quantities

$$\{h^{(1),0},h_{ab}^{(1),0},\gamma_a^{(1),1}\}|_{t=0} \tag{3.35}$$

as the canonical "coordinates" of the linearized evolution equations (at $t=0$) then their conjugate "momenta" are evidently given by

$$\left\{\frac{4\sqrt{{}^{(2)}g}}{N}h^{(1),1}, \frac{\sqrt{{}^{(2)}g}}{2N}(g^{ac}g^{bd} - g^{ab}g^{cd})h_{cd}^{(1),1}, -2\frac{\sqrt{{}^{(2)}g}}{N}e^{4\phi}g^{ab}\gamma_b^{(1),0}\right\}\Bigg|_{t=0}. \tag{3.36}$$

These two sets of quantities are precisely the free data we found for the general solution of the linearized evolution equations in Sec. II B. The canonical "momenta" provide the coefficients of the singular terms in the general first-order perturbation whereas the canonical "coordinates" yield the regular terms in the perturbation.

If for ${}^{(4)}h'$ we substitute a pure gauge perturbation with parameters y_1, $\gamma_{(1)} = X^{3\prime} + Y$, $\mathring{X}^{a\prime}\,\partial/\partial x^a$, and (since we are taking $\lambda^{(1)} = \lambda^{(1)\prime} = 0$)

$$y_0|_{t=0} = \frac{2N}{\sqrt{{}^{(2)}g}}\frac{\partial}{\partial x^a}\left(\frac{\sqrt{{}^{(2)}g}}{N}\mathring{X}^{a\prime}\right), \tag{3.37}$$

then $\omega({}^{(4)}h, {}^{(4)}h')$ reduces, after some simplification, to

$$\omega({}^{(4)}h,{}^{(4)}h_{\text{gauge}}) = \int_K \Bigg\{y_1\frac{\sqrt{{}^{(2)}g}}{2N}(g^{ab}h_{ab}^{(1),1} - 4h^{(1),1}) + \left(\frac{\gamma_{(1)}}{2} + \beta_c\mathring{X}^{c\prime}\right)\nabla_a\left[-2\frac{\sqrt{{}^{(2)}g}}{N}e^{4\phi}g^{ab}h_b^{(1),0}\right]$$
$$+\mathring{X}^{a\prime}\Bigg[e^{-2\phi}\sqrt{{}^{(2)}g}(g^{bc}h_{bc}^{(1),0} - 4h^{(1),0})_{,a} - \left(\frac{2\sqrt{{}^{(2)}g}}{N}e^{4\phi}g^{bc}h_c^{(1),0}(\beta_{a,b} - \beta_{b,a})\right)$$
$$-4\frac{\sqrt{{}^{(2)}g}}{N}\phi_{,a}h^{(1),1} + \nabla_c\left[\frac{\sqrt{{}^{(2)}g}}{N}(h_{ab}^{(1),1}g^{bc} - \delta_a^c g^{bd}h_{bd}^{(1),1})\right]\Bigg]\Bigg\}\Bigg|_{t=0} \tag{3.38}$$

Comparing this with Eq. (3.5) we see that the gauge parameters are canonically conjugate to the linearized constraints as one should have expected.[13]

If, of course, as we eventually intend to do, we impose the linearized constraints upon the perturbations, then we can also permit perturbations in the lapse and shift (e.g., permit nonvanishing $\lambda^{(1)}$, $\lambda^{(1)\prime}$) and still have a conserved symplectic contraction $\omega({}^{(4)}h, {}^{(4)}h')$. A more complete statement of this fact, together with a discussion of its geometrical significance, is given in Sec. III A of Ref. 1 and thus need not be repeated here.

E. Solving the *n*th order constraint equations

It is well known from linearization stability analysis[5,6] that obstructions to solving the higher-order perturbed constraint equations (on a compact Cauchy surface) arise precisely whenever one attempts to perturb from one (Killing) symmetry class to another. The obstructions take the form of certain second-order integral restrictions upon the first-order perturbations one must impose, in addition to the linearized constraints, in order to be able to continue solving the constraints to higher order. These second-order conditions are precisely the vanishing of the perturbational conserved quantities associated with the Killing symmetries of the background solution which the perturbation is breaking. If the perturbations considered preserve all the Killing symmetries of the background solution then the second-order conditions are identically satisfied and no obstructions prevent the solution of the constraints to higher order. In particular, if the background has no Killing symmetries at all then obstructions are absent.

Our background solutions all share the Killing symmetry generated by $Y = \partial/\partial x^3$ but, at the same time, our perturbations have all been constrained to preserve that symmetry. Thus we should expect that obstructions to solving the higher-order constraint equations will occur only if our background solution admits some additional "nontrivial" Killing symmetry (generated by one or more Killing fields linearly independent of $\partial/\partial x^3$). The necessary and sufficient conditions upon the background data prescribed at $t = 0$ for the occurrence of such additional Killing symmetry were given at the end of Sec. III C above. We shall see explicitly below that the absence of such additional Killing fields is indeed precisely sufficient to exclude obstructions to solving the higher-order constraint equations.

Ordinarily one imagines solving the perturbed constraints on a Cauchy hypersurface of the background spacetime but, for our purposes, it is more desirable to solve them at $t = 0$, the Cauchy horizon of the background. In this respect we shall follow the pattern already developed for the treatment of the evolution equations and the linearized constraint equations.

Suppose for the moment, however, that we have already solved the perturbed constraint equations up through order n on a Cauchy surface, $t = t_0 > 0$, of the background. Then Eqs. (3.2) and their perturbations up through order n may be used successively to prove that the perturbed constraints remain satisfied $\forall t > 0$ such that (in particular) $t \leqslant t_0$. This follows from the uniqueness result in the ordinary Cauchy–Kowalewski theorem and the fact that identically vanishing peturbed constraints are clearly a particular solution with the right initial conditions. Thus the perturbed constraints vanish $\forall 0 < t \leqslant t_0$ if and only if they vanish on the Cauchy surface $t = t_0$.

However, for any $0 < k \leqslant n$, we may compute the kth-order perturbed constraints directly by differentiating the exact expressions (3.1) k times with respect to ϵ and substituting expressions (2.14) and (2.15) [as well as the perturbations of (2.13)] for the fundamental perturbations. Superficially it is clear that these kth-order perturbations of the constraints will each contain terms proportional to $(\ln t)^l$ for each $l = 0,1,...,k$. Contributions to these terms come from both the linear terms in the kth-order perturbations of the fundamental fields and (for $k > 1$) from the nonlinear terms in the lower-order perturbations. A typical term in the kth-order perturbed constraints has the form $(\ln t)^l\, r/t^m$, where r is analytic in t and x^a. However, it is straightforward to show that such expressions cannot, in fact, vanish on an interval $(0 < t \leqslant t_0)$ unless the coefficient of each independent logarithmic power vanishes separately. Otherwise one could arrive at a contradiction by first multiplying the perturbed constraints by a suitable power of t (to clear the factors in the denominator), then differentiating sufficiently many times with respect to t and demanding that the resulting expressions have vanishing limits as $t \to 0$.

Thus to satisfy the kth-order constraints we must achieve the vanishing of the coefficient of each logarithmic power, $(\ln t)^l$, for $l = 0,1,2,...,k$. However, the only freedom we have to adjust the kth order perturbation is that of adding an arbitrary solution of the homogeneous (first-order) perturbed evolution equations to any particular solution of the kth order perturbation equations. But such first-order perturbations contain only the lowest two logarithmic powers of t (corresponding to $l = 1,0$) and thus (entering linearly as they do) contribute only to these two powers of $\ln t$ in the expressions for the perturbed constraints. Clearly, for $k \geqslant 2$, there is no freedom to cancel the terms in $(\ln t)^l$, for $l \geqslant 2$. On the other hand, we know, from the Cauchy surface argument given above, that (barring linearization instabilities) it is always possible to solve the kth-order constraints for arbitrarily large k.

This apparent contradiction is avoided if and only if the terms in $(\ln t)^l$ (for $2 \leqslant l \leqslant k$) vanish automatically as a consequence of the perturbed evolution equations and the perturbed constraint equations (up through order $k - 1$) which we assume to have already been imposed. For the same reason the coefficients of the negative powers of t that multiply $(\ln t)^l$, for $l = 1,0$, and that cannot be canceled by adjustment of the free data must also vanish automatically as a consequence of the evolution and lower-order constraint equations. Thus the kth-order perturbed constraints automatically reduce to expressions involving only the same powers of $(\ln t)$ and of t as those encountered in the study of the first-order constraints. Any other result would contradict the *a priori* known solvability of the kth-order constraint equations.

Since the free data at kth order has the same form as a first-order perturbation [cf Eq. (3.3)] we shall use the same notation introduced at first order to designate this data. Thus we let $\{\phi^{(1)}, g_{ab}^{(1)}, \beta_a^{(1)}\}$ represent an arbitrary solution of the first-order (homogeneous) perturbed evolution equations which we may add to any particular solution of the (inhomogeneous) kth-order equations.

The dominant surviving term in the kth-order perturbed constraints is one of order t^{-1} in the Hamiltonian constraint. It takes the form

$$(4h^{(1),1} - g^{ab}h_{ab}^{(1),1})|_{t=0} = \text{source}, \tag{3.39}$$

where "source" stands generically for an analytic inhomogeneity that arises from the lower-order perturbations. The imposition of this constraint ensures the vanishing of the perturbed Hamiltonian constraint as $t \to 0$ and also causes a term proportional to $\ln t$ in the (kth-order) perturbation of $\mathscr{H}_a - \beta_a \mathscr{H}_3$ to drop out in the limit as $t \to 0$. The remaining contributions to the kth-order constraints can be forced to vanish as $t \to 0$ by imposing equations of the form

$$\left(\frac{\partial}{\partial x^a}\left[\frac{2\sqrt{{}^{(2)}g}}{N}\, e^{4\phi} g^{ab} h_b^{(1),0}\right]\right)\Bigg|_{t=0} = \text{source},$$

$$\left\{\frac{1}{2}\frac{\hat{g}_{ac}}{\sqrt{{}^{(2)}\hat{g}}}\hat{\nabla}_b(\gamma^{bc}) - e^{2\phi}\hat{g}^{bd}(\beta_{a,b}-\beta_{b,a})h_d^{(1),0} - \sigma_{,a}\right\}\Big|_{t=0} = (\text{source})_a, \tag{3.40}$$

where

$$\sigma = (-2\lambda^{(1)} + h^{(1),1} + 2h^{(1),0} - \tfrac{1}{2}g^{cd}h_{cd}^{(1),0}) \tag{3.41}$$

and where $\hat{g}_{ab}$ and γ^{ab} are defined as in Eq. (3.7). As above, "source" stands for an analytic inhomogeneity arising from the lower-order perturbations. Equations (3.39) and (3.40) are inhomogeneous generalizations of Eqs. (3.5) and (3.6) and are necessary and sufficient for the vanishing of the k th-order constraints in the limit as $t\to 0$.

To show that Eqs. (3.39) and (3.40) are also sufficient for the vanishing of the perturbed constraints for $t>0$ we appeal to the evolution equations (3.2), which represent the (contracted) Bianchi identities for our system. Since Eqs. (3.2) are linear and homogeneous in $\{\mathscr{H}, \mathscr{H}_3, \mathscr{H}_a\}$, the k th-order perturbations of these equations yield equations of precisely the same form for the k th-order perturbed constraints (provided, of course, that the constraints up through order $k-1$ have already been imposed). In fact, the surviving (analytic) coefficients of (ln t) in the k th-order constraints must separately satisfy these same evolution equations since the contributions to the full equations proportional to (ln t) cannot be canceled [on any interval of the form $(0,t_0)$, for example] by the contributions lacking this factor. However, when Eqs. (3.39) and (3.40) are imposed the contributions proportional to (ln t) vanish in the limit as $t\to 0$. Equations (3.2) imply therefore the vanishing of these (analytic) coefficients $\forall t>0$ just as in the earlier example of purely first-order perturbations.

Thus the k th-order constraints reduce to purely analytic expressions (all the logarithmic terms and terms involving negative powers of t having been canceled out) that satisfy Eqs. (3.2) and that, moreover, vanish as $t\to 0$. Applying Eqs. (3.2) once again we thus find that the k th-order perturbed constraints vanish identically (i.e., $\forall t>0$ within the domain of existence of the background solution).

The only possible obstruction to implementing the above scheme for solving the constraints at k th order is the possibility that, in fact, Eqs. (3.39) and (3.40) may fail to have solutions. We shall now show that this eventuality never occurs unless the background space-time admits "nontrivial" Killing symmetries.

First of all, Eq. (3.39) is purely algebraic and can always be solved for (say) $(g^{ab}h_{ab}^{(1),1})|_{t=0}$ in terms of the remaining quantities. Thus we shall regard the trace of $(h_{ab}^{(1),1})|_{t=0}$ as fixed by Eq. (3.39). To solve the first of Eqs. (3.40) we decompose the vector field

$$V^a = ((2e^{4\phi}/N)g^{ab}h_b^{(1),0})|_{t=0} \tag{3.42}$$

into L^2-orthogonal summands of the form

$$V^a = V^{a^T} + g^{ab}\chi_{,b}, \tag{3.43}$$

where $\nabla_a V^{a^T} = 0$. This decomposition always exists and is unique on the compact Riemannian manifold (K,g_{ab}) by standard linear elliptic analysis (the "Fredholm alternative"). Substituting this decomposition into the first of Eqs. (3.40) leads to a Poisson equation for the function χ,

$$\frac{\partial}{\partial x^a}(\sqrt{{}^{(2)}g}g^{ab}\chi_{,b}) = \text{source}. \tag{3.44}$$

Again by standard linear elliptic theory this equation admits a global solution on K if and only if the source satisfies the integrability condition

$$\int_K (\text{source}) = 0. \tag{3.45}$$

However, from the form of the exact constraint $\mathscr{H}_3$ (i.e., the fact that it is a divergence) it is easy to see that the source term will inevitably have the form of a divergence of some vector density constructed from the lower-order perturbations. Therefore the integrability condition (3.45) will automatically be satisfied and thus the "longitudinal part" of V^a can always be chosen [via Eq. (3.44)] in such a way that the first of Eqs. (3.40) is satisfied.

Finally we must solve the second of Eqs. (3.40). We first decompose γ^{ab} (which represents the trace-free part of $h_{ab}^{(1),1}$) as in Eq. (3.8) and then substitute this decomposition into the constraint equation. The result is a second-order linear elliptic equation for the vector field $Y^a\,\partial/\partial x^a$ of the same form as Eq. (3.10) (but supplemented with the "source" term). This equation admits a solution (unique up to the addition of a conformal Killing field of $\hat{g}_{ab}$) if and only if the inhomogeneous term is L^2-orthogonal to every conformal Killing field of $\hat{g}_{ab}$. This integrability condition takes the form

$$\int_K \{\sqrt{{}^{(2)}g}e^{2\phi}Z^a g^{bd}(\beta_{a,b}-\beta_{b,a})h_d^{(1),0} + \sqrt{{}^{(2)}\hat{g}}Z^a\sigma_{,a} + \sqrt{{}^{(2)}\hat{g}}Z^a(\text{source})_a\}|_{t=0} = 0, \tag{3.46}$$

for every $Z^a\,\partial/\partial x^a$ satisfying Eq. (3.12). Reexpressing this somewhat by making use of the background regularity condition $(N/e^{2\phi})|_{t=0} = 1$ and (after integration by parts) the first of Eqs. (3.40), one finds the condition

$$0 = \int_K \{-(\sqrt{{}^{(2)}\hat{g}}Z^a)_{,a}\sigma + \sqrt{{}^{(2)}\hat{g}}Z^a(\text{source})_a - (\tfrac{1}{2}Z^a\beta_a)(\text{source}) - \tfrac{1}{2}(\mathscr{L}_{{}^{(2)}Z}\beta)_b V^b\sqrt{{}^{(2)}g}\}|_{t=0}, \tag{3.47}$$

where $V^a\,\partial/\partial x^a$ is defined as in Eq. (3.42). Decomposing $(\mathscr{L}_{{}^{(2)}Z}\beta)_a\,dx^a$ into L^2-orthogonal summands of the form [cf Eq. (3.43)]

$$(\mathscr{L}_{{}^{(2)}Z}\beta)_a = (\mathscr{L}_{{}^{(2)}Z}\beta)_a^T + \delta_{,a} \tag{3.48}$$

and substituting this expression into the condition (3.47) one finds, after some further manipulation, the equivalent condition

$$0 = \int_K \{-(\sqrt{{}^{(2)}\hat{g}}Z^a)_{,a}\sigma + \sqrt{{}^{(2)}\hat{g}}Z^a(\text{source})_a - (\tfrac{1}{2}Z^a\beta_a)(\text{source}) - \tfrac{1}{2}\sqrt{{}^{(2)}g}(V^b)^T(\mathscr{L}_{{}^{(2)}Z}\beta)_b^T + \tfrac{1}{2}\delta(\text{source})\}|_{t=0}. \tag{3.49}$$

At this point $(V^a)^T$ and σ are still completely arbitrary and may always be adjusted to satisfy the (finite number of) integrability conditions given by Eq. (3.49) unless their coefficients vanish identically. Thus we can always solve the kth-order constraint equations without obstruction unless both

$$(\sqrt{{}^{(2)}\hat{g}}Z^a)_{,a} = 0 \tag{3.50}$$

and

$$(\mathscr{L}_{{}^{(2)}Z}\beta)_a = \delta_{,a} \tag{3.51}$$

hold for some conformal Killing field ${}^{(2)}Z = Z^a\,\partial/\partial x^a$ of $\hat{g}_{ab}$. Equation (3.50) implies that ${}^{(2)}Z$ must be an actual Killing field of $\hat{g}_{ab}$. Identifying ${}^{(2)}Z$ with the vector field ${}^{(2)}\mathring{X}'$ in Eqs. (3.32), setting

$$\mathring{Y} = -2\mathscr{L}_{{}^{(2)}Z}\mathring{\phi},$$
$$\mathring{X}^{3'} = -\mathring{Y} - 2\delta, \tag{3.52}$$

and recalling that $\hat{g}_{ab} = e^{-2\phi}g_{ab}$ and that $(N/e^{2\phi})|_{t=0} = 1$, one finds that all of Eqs. (3.32) are satisfied and thus that $\{\mathring{Y}, Z^a, \mathring{X}^{3'}\}$ are initial data for an additional, "nontrivial" Killing field of ${}^{(4)}g$.

Thus the integrability conditions for the solution of the kth-order perturbed constraints are satisfiable (for arbitrary "sources") provided there exist no "nontrivial" Killing fields of ${}^{(4)}g$. If such Killing fields exist then the source terms appearing in the constraints must be subjected to further conditions of the form

$$\int_K \{\sqrt{{}^{(2)}\hat{g}}Z^a(\text{source})_a + \tfrac{1}{2}(\delta - Z^a\beta_a)(\text{source})\}|_{t=0} = 0, \tag{3.53}$$

for each $Z^a\,\partial/\partial x^a$ satisfying Eqs. (3.50) and (3.51). To simplify the analysis we may simply assume that the background metric ${}^{(4)}g$ has been chosen to have no "nontrivial" Killing symmetries (i.e., no Killing fields independent of $\partial/\partial x^{3'}$). In this case the integrability conditions (3.49) may always be satisfied with suitable restrictions upon $(V^a)^T$ and σ and the solutions of the perturbed constraints carried out to arbitrarily high order.

IV. CONCLUDING REMARKS

In subsequent work we plan to show how one can collect together the "dominant singular terms" from each order of perturbation theory and sum the resultant truncated series to determine what is presumably the asymptotic behavior of the perturbed Einstein space-times near their singular boundaries. A special case of this program has already been carried out for vacuum Gowdy metrics on $T^3\times R$ (which have two spacelike Killing fields rather than only one as in the problem treated here).[14,15] For the Gowdy problem we found that the asymptotic behavior of the perturbative solutions was governed by certain "geodesic loops" propagating in hyperbolic two-space and that the asymptotic behavior of (say) the Riemann curvature tensor could be evaluated in terms of computations based upon an associated family of approximate (and explicitly computable) "geodesic loop space-times."

The validity of this geodesic loop approximation for Gowdy space-times has been rigorously established for the special case of polarized Gowdy metrics (on $T^3\times R$, $S^3\times R$, and $S^2\times S^1\times R$) by Isenberg and the author[16] and an infinite-dimensional subfamily of unpolarized Gowdy metrics that display the indicated asymptotic behavior has been discovered by Mansfield.[17] An analogous family of vacuum solutions having only one spacelike Killing field was discussed by the author Ref. 18. In addition, Mansfield has extended the perturbation analysis to the Einstein–Maxwell equations for electrovacuum space-times of the Gowdy symmetry type on $T^3\times R$.[15] Chruściel and the author have made significant progress towards a proof of the validity of the geodesic loop approximation for general (i.e., unpolarized) Gowdy metrics but their analysis is not yet complete.

It seems likely that the analog of the geodesic loop approximation, and its implications for the asymptotic behavior of the perturbed generalized Taub–NUT space-times considered here, can be derived from the results of the present paper by the method suggested above. If so, then the asymptotic properties of an extremely large family of singular vacuum space-times (each having only one spacelike Killing field) could be determined by straightforward computations. A powerful check on such an approach would then be provided by the rigorously defined family of singular solutions discussed in Ref. 18 [just as Mansfield's family (cf. Ref. 17) provides a check on the perturbative approach to the Gowdy problem].

What is more, there is no reason to suppose that the methods of the present paper are limited to the study of symmetric perturbations. As a model problem for the study of completely general (nonsymmetric) perturbations of these same (generalized Taub–NUT) backgrounds, the author has recently derived the perturbative solutions (to all orders) for the nonlinear wave equation on such backgrounds.[19] He has shown that one can sum the leading-order terms of the full perturbation series for the nonlinear wave equation to derive what is presumably the asymptotic behavior of the general solution of this equation near a cosmological Cauchy horizon of the Taub–NUT type. He has also shown that the asymptotic solutions may be naturally classified into "Lagrangian submanifolds" of an associated asymptotic phase space. This last result is quite analogous to the classification of polarized Gowdy solutions into Lagrangian submanifolds discussed by Isenberg and the author in Ref. 16. A project to extend the results of the present paper to the study of completely general (nonsymmetric) perturbative solutions of Einstein's equations is currently underway.

A further exciting possibility, which was suggested to the author by Cosgrove, is that, at least for the Gowdy metrics problem, one may be able to sum higher-order contributions to the full perturbation series systematically by applying something like the "two-timing" method of conventional applied analysis.[20] Cosgrove has already sketched such a treatment of the closely related problem of studying singularities in the solutions of the stationary axisymmetric problem near an axis of symmetry. If this method generalizes to the perturbation problem treated in the present paper it promises to yield further significant insights into the nature of the singularities of the perturbative solutions described herein.

The present paper, and the references cited herein, have only touched the surface of the class of general relativistic problems amenable to the higher-order perturbation methods we have developed. It seems quite conceivable that the general solution of Einstein's equations near a cosmological singularity is open to study by a further development of these methods.

ACKNOWLEDGMENTS

I am grateful to the Université de Paris VI and the C.N.R.S. for hospitality and support during some of the preliminary work on this project in 1986. I am also grateful to the Centre for Mathematical Analysis of the Australian National University for hospitality while this work was completed in 1988.

This research was supported in part by NSF Grant No. PHY-8503072 to Yale University.

[1]V. Moncrief, Ann. Phys. (NY) **141**, 83 (1982).
[2]V. Moncrief, J. Geom. Phys. **1**, 107 (1984).
[3]V. Moncrief and J. Isenberg, Commun. Math. Phys. **89**, 387 (1983).
[4]J. Isenberg and V. Moncrief, J. Math. Phys. **26**, 1024 (1985).
[5]J. Arms, J. Marsden, and V. Moncrief, Ann. Phys. (NY) **144**, 81 (1982).
[6]J. Arms, J. Marsden, and V. Moncrief, Commun. Math. Phys. **78**, 455 (1981); see also A. Fischer, J. Marsden, and V. Moncrief, Ann. Inst. H. Poincaré **33**, 147 (1980), as well as other references cited in these papers.
[7]See, for example, C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 21.
[8]These equations appeared incorrectly in Eqs. (2.14) of Ref. 1. The argument given in that reference applies without modification to the corrected equations and yields the correctly stated results of the earlier paper.
[9]M. Berger and D. Ebin, J. Differ. Geom. **3**, 379 (1969).
[10]The decomposition occurring in Eq. (3.16) of Ref. 1 appeared incorrectly without the transverse-traceless summand. Reference 1 also had the constant k defined by Eq. (2.1) of this paper set equal to unity. Unless the range of the x^3 coordinate is allowed to vary, the choice $k = 1$ is a restriction upon the generality of the metric form.
[11]See, for example, Ref. 9 for a discussion of these methods in the context of splittings of symmetric tensor fields.
[12]A. Fischer and A. Tromba, Math. Ann. **267**, 311 (1984).
[13]A discussion of this point appears in Sec. III of Ref. 1.
[14]V. Moncrief, "Asymptotic behavior of Gowdy metrics on $T^3 \times R$," in preparation.
[15]P. Mansfield, paper in preparation.
[16]J. Isenberg and V. Moncrief, "Asymptotic behavior of the gravitational field and the nature of singularities in Gowdy spacetimes," unpublished.
[17]P. Mansfield, paper in preparation.
[18]V. Moncrief, Class. Quant. Gravit. **4**, 1555 (1987).
[19]V. Moncrief, J. Math. Phys. **30**, 1760 (1989).
[20]C. Cosgrove (private communication).

Spherically symmetric solutions in higher dimensions

K. D. Krori, P. Borgohain, and Kanika Das
Mathematical Physics Forum, Cotton College, Gauhati-781001, India

(Received 19 July 1988; accepted for publication 19 April 1989)

An exact general solution of Einstein's equations for spherically symmetric distribution of a perfect fluid in N dimensions is presented from which the whole class of spherically symmetric solutions may be obtained. As examples, some particular solutions obtainable from a general solution are presented.

I. INTRODUCTION

In view of the recent emergence of superstring theory as the most promising theory developed thus far, having the potential to lead us a step closer toward unification of four forces, studies in higher dimensions have obtained a new importance inspiring a host of workers to enter into this field of study. Already a number of important solutions of Einstein's equation in higher dimensions have been obtained. Yoshimura's[1] solutions of higher-dimensional Einstein's equations in a vacuum and the work of Koikawa and Yoshimura's[2] in the presence of matter and Koikawa's[3] solution and a Schwarzschild-like exterior solution[4] are a few of them. We have also worked out a solution of Einstein's equation in higher dimensions in the presence of matter.[5] In the present paper we present an exact general solution of Einstein's equations for spherically symmetric perfect fluids in N dimensions, from which the whole class of spherically symmetric solutions (both in N as well as four dimensions) may be obtained. We present some particular solutions in Sec. III. Our work is a higher-dimensional generalization of the work of Berger, Hojman, and Santamarina.[6]

II. FIELD EQUATIONS AND THEIR SOLUTION

We consider the line element in the form

$$ds^2 = g^2(x)dt^2 - dx^2 - r^2\, d\Omega^2, \tag{1}$$

where $d\Omega^2$ is the line element on a unit $(N-2)$ sphere.

Einstein's field equations with cosmological constants are

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = T_{\mu\nu}. \tag{2}$$

For a perfect fluid with spherical symmetry the energy-momentum tensor is

$$T_{\mu\nu} = (\rho + P)U_\mu U_\nu - Pg_{\mu\nu}, \tag{3}$$

where $U^\mu = (1/g)\delta^\mu{}_0$ is a unit vector with the flow lines tangent to it.

It may be readily seen that the change of variable

$$\overline{P} = P - \Lambda, \quad \bar{\rho} = \rho + \Lambda \tag{4}$$

transforms Eqs. (2) into an equivalent system with $\Lambda = 0$. Therefore twirls can now be dropped, keeping in mind that the $\Lambda \neq 0$ case is already included.

The field equations for the metric (1) are given by

$$(N-2)(r''/r) - [(N-2)(N-3)/2r^2](1 - r'^2) = -\rho, \tag{5}$$

$$(N-2)(g'r'/gr) - [(N-2)(N-3)/2r^2](1 - r'^2) = P, \tag{6}$$

$$\frac{g''}{g} + \frac{(N-3)g'r'}{gr} + (N-3)\frac{r''}{r} - \frac{(N-3)(N-4)}{2r^2}(1 - r'^2) = P, \tag{7}$$

and the equation of hydrostatic equilibrium is

$$(\rho + P)(g'/g) + P' = 0, \tag{8}$$

where the primes denote differentiations with respect to x.

Multiplying Eq. (5) by $r^{N-2}r'$ and integrating over x one obtains

$$r'^2 = 1 - 2m(r)/r^{N-3}, \tag{9}$$

where

$$\frac{dm}{dr} = \frac{\rho r^{N-2}}{(N-2)}. \tag{10}$$

Using Eqs. (8) and (9) in (6) we obtain

$$(r^{N-3} - 2m)\left[\frac{(N-2)(N-3)}{2r^{N-1}} - \frac{(N-2)}{r^{N-2}}\frac{dp}{dr}\frac{1}{(\rho + P)}\right] = P + \frac{(N-2)(N-3)}{2r^2}. \tag{11}$$

We now define a function G as

$$G = -\frac{(r^{N-3} - 2m)}{P + (N-2)(N-3)/2r^2}. \tag{12}$$

Equation (11), when expressed in terms of G, takes the form

$$r^3G\left[\frac{G(N-2)(N-3)}{2} - r^{N-1}\right]\frac{dp}{dr} + pr^3 \times\left[\frac{G(N-2)(N-3)}{2} + r^{N-1}\right]\left(\frac{dG}{dr} + \frac{2r^{N-2}}{N-2}\right) + \left[\frac{G(N-2)(N-3)}{2} + r^{N-1}\right]\left[(N-3)r^{N-1} + \frac{r(N-2)(N-3)}{2}\frac{dG}{dr} - G(N-2)(N-3)\right] = 0. \tag{13}$$

This equation can be integrated for $p(r)$ if $G(r)$ is a given function,

 0022-2488/89/102315-04$02.50

$$p(r)=\exp\int\frac{[r^{N-1}+G(N-2)(N-3)/2]}{G[r^{N-1}-G(N-2)(N-3)/2]}\left(\frac{dG}{dr}+\frac{2r^{N-2}}{N-2}\right)dr$$
$$\times\left\{K-\int\left[\frac{\{G(N-2)(N-3)/2+r^{N-1}\}\{(N-3)r^{N-1}+r(N-2)(N-3)/2(dG/dr)-G(N-2)(N-3)\}}{Gr^3\{G(N-2)(N-3)/2-r^{N-1}\}}\right.\right.$$
$$\left.\left.\times\exp\int\frac{\{r^{N-1}+G(N-2)(N-3)/2\}[dG/dr+2r^{N-2}/(N-2)]}{G\{G(N-2)(N-3)/2-r^{N-1}\}}dr\right]dr\right\},\qquad(14)$$

where K is an integration constant.

The function $\rho(r)$ can be obtained with the help of (10):

$$\rho(r)=\frac{(N-2)}{2r^{N-2}}\left[(N-3)r^{N-4}+G\frac{dp}{dr}+\frac{dG}{dr}p-\frac{(N-2)(N-3)G}{r^3}+\frac{(N-2)(N-3)}{2r^2}\frac{dG}{dr}\right],\qquad(15)$$

where $p(r)$ is given by (14).

The metric coefficient g can be found by direct integration of (8) and using (11) and (12),

$$g^2(x)=g_0^2\exp\left[-2\int\frac{dp/dr}{(\rho+P)}dr\right]=g_0^2\frac{1}{r^{N-3}}\exp\left[-\int\frac{2r^{N-2}}{(N-2)G}dr\right].\qquad(16)$$

To complete the integration we can recover the link between the metric coefficient r and the original variable x from Eq. (9),

$$x=\int\frac{dr}{\sqrt{1-2m/r^{N-3}}}.\qquad(17)$$

III. DERIVATION OF PARTICULAR SOLUTIONS

A. Schwarzschild-like exterior solution

To obtain the Schwarzschild-like vacuum solution,[4] on the basis of the above theory, we choose

$$G(r)=-[2r^2/(N-2)(N-3)](r^{N-3}-2M),\qquad(18)$$

where M is a constant and $K=0$.

Under such conditions we have

$$(N-3)r^{N-1}+\frac{r(N-2)(N-3)}{2}\frac{dG}{dr}-G(N-2)(N-3)=0\qquad(19)$$

and consequently

$$p(r)=0.\qquad(20)$$

Equations (14), (19), and (20) then lead to

$$\rho(r)=0.\qquad(21)$$

Comparing Eqs. (12) and (18) we obtain

$$m(r)=M.$$

Equation (17) can thus be written as

$$dx^2=\frac{dr^2}{1-2M/r^{N-3}}.\qquad(22)$$

Finally, from Eq. (16),

$$g^2(r)=g_0^2(1-2M/r^{N-3}).\qquad(23)$$

Equation (23) represents the Schwarzschild-like exterior solution in higher dimensions.

B. Interior Krori–Borgohain–Das solution

Now we derive an interior Schwarzschild-like solution in N dimensions recently obtained by us.[5] For this we choose G in the form

$$G(r)=-\frac{2r^{N-1}}{(N-2)(N-3)}\frac{[A_0\sqrt{1-r^2/R^2}-B_0(1-r^2/R^2)]}{[A_0\sqrt{1-r^2/R^2}-B_0\{1-(N-1)r^2/(N-3)R^2\}]},\qquad(24)$$

where

$$1/R^2=2\rho/(N-1)(N-2)\qquad(25)$$

and A_0, B_0, and R are constants.

Putting $K=0$ in the expressions (14) for $p(r)$, it is found after some calculation that

$$p(r)=\frac{(N-2)}{2R^2}\left[\frac{(N-1)B_0\sqrt{1-r^2/R^2}-(N-3)A_0}{A_0-B_0\sqrt{1-r^2/R^2}}\right].\qquad(26)$$

Also, Eq. (15) gives

$$\rho=(N-1)(N-2)/2R^2.\qquad(27)$$

Finally Eqs. (16) and (17) give

$$g^2(r)=g_0^2(A_0-B_0\sqrt{1-r^2/R^2})^2\qquad(28)$$

and

$$dx^2=\frac{dr^2}{1-r^2/R^2}.\qquad(29)$$

This has already been derived directly by Krori, Borgohain, and Das (KBD).[5]

C. An interior solution

We now consider a particular interior solution, which immediately leads to a simple equation of state. For this we choose G in the form

$$G=Ar^{N-1}.\qquad(30)$$

Taking the integration constant $K=0$, we obtain from (14) and (15), the expressions for pressure and density, respectively, as

$$p(r) = \frac{(N-3)[(A/2)(N-2)(N-3)+1]^2}{r^2[2A\{(A/2)(N-2)(N-3)-1\} - \{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}]} \tag{31}$$

and

$$\rho(r) = \frac{(N-2)(N-3)}{2r^2}\left\{\frac{A}{2}(N-2)(N-3)+1\right\} \times\left[1 + \frac{A\{(A/2)(N-2)(N-3)+1\}}{2A\{(A/2)(N-2)(N-3)-1\} - \{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}}\right]. \tag{32}$$

The pressure and density are positive when

$$A^2(1-N)(N-2)^2(N-3) > 4A(N-2)(N-1)+4 \tag{33}$$

and

$$1+\frac{A}{2}(N-2)(N-3) > \frac{-A[1+(A/2)(N-2)(N-3)]^2}{[2A\{(A/2)(N-2)(N-3)-1\} - \{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}]}. \tag{34}$$

The equation of state is given by $p = (\gamma - 1)\rho$, where

$$\gamma = \frac{2\{(A/2)(N-2)(N-3)+1\} + \frac{3}{2}A^2(N-2)^2(N-3) - A(N-2) - \{(A/2)(N-2)(N-3)+1\}\{A(N-1)(N-2)+2\}}{(N-2)[\frac{3}{2}A^2(N-2)(N-3) - A - \{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}]}. \tag{35}$$

It can be seen that for suitable values of the parameter A, the conditions (33)–(35) are satisfied and the solution describes a physical configuration. A possible range for values of A, i.e., $-1 \leqslant A \leqslant -0.34$ when $N=4$ (that is, the four-dimensional case). In this case, the values of $\gamma = 1.6$ for $A = -0.4$. Finally, Eqs. (16) and (17) give us

$$g^2(r) = g_0^2 r^{-(N-3)-2/A(N-1)} \tag{36}$$

and

$$dx^2 = \frac{dr^2}{1-2m/r^{N-3}}. \tag{37}$$

D. A composite solution

The solution obtained in Sec. III C is not free from singularity at the center and hence such a solution cannot form the core of a physical structure. Also the condition of vanishing of pressure at the boundary of configuration described by our solution in Sec. III C leads to a situation where pressure and density become zero for all values of r. Hence such a solution cannot also be used to describe the outermost layer of a composite structure. However, such a solution can form an intermediate layer of a composite system. As an illustration here we consider a composite sphere with a core of radius r_1, described by a solution given by us (KBD) in Sec. III B, an intermediate layer of internal and external radii r_1 and r_2, respectively, given by our solution in Sec. III C, and another outer layer of internal and external radii r_2 and a, respectively, described by a KBD-type solution.

Continuity of metric coefficients at $r = r_1$ gives us

$$r_1^{-(N-3)/2-1/A(N-1)} = C - D\sqrt{1-r_1^2/R_1^2} \tag{38}$$

and

$$r_1^{N-1} = 2mR_1^2. \tag{39}$$

Also, continuity of pressure at the same internal boundary gives

$$\frac{(N-3)[(A/2)(N-2)(N-3)+1]^2}{r_1^2[2A\{(A/2)(N-2)(N-3)-1\} - \{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}]} = \frac{(N-2)}{2R_1^2}\left[\frac{(N-1)D\sqrt{1-r_1^2/R_1^2} - (N-3)C}{C - D\sqrt{1-r_1^2/R_1^2}}\right]. \tag{40}$$

Equations (38)–(40) express the parameters C, D, and R_1, respectively, of the core, in terms of parameters of the intermediate shell.

Also at the boundary $r = r_2$, the continuity of metric coefficients and pressure gives us

$$r_2^{-(N-3)/2-1/A(N-1)} = E - F\sqrt{1-r_2^2/R_2^2} \tag{41}$$

and

$$r_2^{N-1} = 2mR_2^2. \tag{42}$$

Also,

$$\frac{(N-3)[(A/2)(N-2)(N-3)+1]^2}{r_2^2[2A\{(A/2)(N-2)(N-3)-1\}-\{(A/2)(N-2)(N-3)+1\}\{[A(N-1)(N-2)+2]/(N-2)\}]}$$
$$=\frac{(N-2)}{2R_2^2}\left[\frac{(N-1)F\sqrt{1-r_2^2/R_2^2}-(N-3)E}{E-F\sqrt{1-r_2^2/R_2^2}}\right], \quad (43)$$

where E, F, and R_2 are the counterparts of C, D, and R_1, respectively, in the outermost layer of the composite structure.

Again, at the outermost boundary of the composite structure, i.e., at $r = a$, the continuity of metric coefficients gives us

$$[E-F\sqrt{1-a^2/R_2^2}\,]^2 = 1-2M/a^{N-3} \quad (44)$$

and

$$a^{N-1} = 2MR_2^2. \quad (45)$$

The pressure at $r = a$ is zero. Hence we obtain

$$FD\sqrt{1-a^2/R_2^2} = [(N-3)/(N-1)]E. \quad (46)$$

The mass of the configuration is given by

$$M = a^{N-1}/2R_2^2. \quad (47)$$

ACKNOWLEDGMENTS

The authors express their profound gratitude to the Government of Assam, Dispur, for all facilities provided at Cotton College, Gauhati-781001, India, to carry out the investigations reported in this paper.

[1] M. Yoshimura, Phys. Rev. D **34**, 1021 (1986).
[2] T. Koikawa and M. Yoshimura, Prog. Theor. Phys. **75**, 977 (1986).
[3] T. Koikawa, Phys. Lett. A **117**, 279 (1986).
[4] R. C. Myers and M. J. Perry, Ann. Phys. (N.Y.) **172**, 304 (1986).
[5] K. D. Krori, P. Borgohain, and K. Das, Phys. Lett. A **132**, 321 (1988).
[6] S. Berger, R. Hojman, and J. Santamarina, J. Math. Phys. **28**, 2949 (1987).

Killing tensor conservation laws and their generators

Kjell Rosquist
Department of Physics, University of Stockholm, Vanadisvägen 9, S-113 46 Stockholm Sweden

(Received 18 October 1988; accepted for publication 7 June 1989)

The generators of Killing vector and tensor geodesic conservation laws are derived. It is shown that the generator of a Killing vector conservation law coincides with the Killing field itself. For Killing tensors the generators are not space-time vector fields but rather depend on the geodesic tangent vector and therefore lie in a jet space of the geodesic equations. By regarding the metric as a field on the one-jet space of the geodesic equations, the action of the Killing tensor generators on the metric can be defined in a natural way. It is found that the metric is not invariant under Killing tensor symmetries. This happens because the Killing tensor symmetries, unlike the Killing vector symmetries, are divergence symmetries.

I. INTRODUCTION

Killing tensors were for a long time regarded as rather mysterious objects. Associated with quadratic geodesic constants of the motion, a geometrical interpretation in terms of symmetries, like that for Killing vectors, was lacking until recently. Although some authors were aware that Killing tensor constants of the motion corresponded to symmetries of the geodesic equations (see, e.g., the papers on separation of variables by Benenti and Francaviglia[1] and Kalnins and Miller[2] and references therein), the first explicit identification of the symmetries corresponding to Killing tensors was made by Prince and Crampin[3,4] using a projective action formalism. In this paper we use the standard modern approach to Lie symmetries given in Olver[5] to analyze the geodesic equations. In that picture little more than elementary differential geometry is used, leading to a more transparent view of the geodesic symmetries. It then becomes obvious that the Killing vector symmetries are point symmetries of the geodesic equations while the Killing tensor symmetries are generalized symmetries (Lie–Bäcklund symmetries), i.e., they depend on the derivatives of the dependent variables in a nontrivial way. Furthermore, the Killing vector symmetries do not depend on the affine parameter. It is this property together with their point character that makes it possible to interpret them as space-time symmetries.

In the projective formalism of Prince and Crampin, the projective action of the Killing tensor symmetry on the metric vanishes. We show that there is a natural way to define an action of the Killing tensor symmetry on the metric. This follows since the Killing tensor symmetries, like the Killing vector symmetries, do not involve the affine parameter. Therefore the metric can be regarded as a field on the one-jet space of the geodesic equations, i.e., the space where the (prolonged) geodesic symmetries live. It then turns out that the action of the Killing tensor symmetry on the metric does not vanish, a result that at first sight would seem to contradict the result of Prince and Crampin. The basic reason for this action not to vanish is that the Killing tensor symmetry, unlike the Killing vector symmetry, is a divergence symmetry. However, the action vanishes "on shell," i.e., on solutions of the geodesic equations, a fact that can account for the result of Prince and Crampin.

II. THE GEODESIC EQUATIONS AND KILLING CONSTANTS OF THE MOTION

We shall use Lie's theory of symmetries of differential equations in the form given by Olver[5] (also see Rosquist[6] for a discussion in a general relativistic context). The geodesic equations constitute a system of ordinary differential equations given by

$$G^a \equiv x''^a + \Gamma^a{}_{bc} x'^b x'^c = 0, \quad x'^a \equiv \frac{dx^a}{d\lambda}, \tag{1}$$

where $U^a \equiv x'^a$ are the components of the geodesic tangent vector U. The independent variable is λ and the four (in four space-time dimensions) dependent variables are x^a. The spaces of independent and dependent variables are denoted by $\mathscr{X}$ and $\mathscr{U}$, respectively. Point symmetries involve coupled transformations of the independent and dependent variables, i.e., transformations of the product space $\mathscr{X} \times \mathscr{U}$. To handle transformations of the derivatives up to the nth order one needs the nth order *jet space*, $\mathscr{X} \times \mathscr{U} \times \mathscr{U}_1 \times \cdots \times \mathscr{U}_n$, where $\mathscr{U}_k$ is the one-dimensional space with coordinate $d^k x/d\lambda^k$. The general form of a point symmetry of the geodesic equations is

$$\mathbf{v} = \psi \frac{\partial}{\partial \lambda} + \phi^a \frac{\partial}{\partial x^a}, \tag{2}$$

where ψ and ϕ^a are functions of λ and x^a. In general, a symmetry $\mathbf{v}$ cannot be interpreted as a space-time vector field. Such an interpretation is only possible when $\psi = 0$ and ϕ^a is a function only of x^a. For generalized symmetries[5] ψ and ϕ^a also depend on the derivatives x'^a, x''^a, etc. Any symmetry is equivalent to its *evolutionary representative*, $\mathbf{v} = Q^a\, \partial/\partial x^a$, where

$$Q^a \equiv \phi^a - x'^a \psi \tag{3}$$

is the *characteristic* of the symmetry.

The geodesic equations can be derived from the Lagrangian

$$L = \tfrac{1}{2} g_{ab} x'^a x'^B. \tag{4}$$

The Euler operator is

$$\mathbf{E}_a = D_\lambda \frac{\partial}{\partial x'^a} - \frac{\partial}{\partial x^a}, \tag{5}$$

where D_λ is the *total derivative*[5] given by

$$D_\lambda = \frac{\partial}{\partial\lambda} + x'^a \frac{\partial}{\partial x^a} + x''^a \frac{\partial}{\partial x'^a} + \cdots . \qquad (6)$$

The sum is formally infinite but the total derivative is only applied to functions that depend on a finite number of derivatives, so only a finite number of terms are needed in any given situation. A short calculation yields the Euler–Lagrange equations

$$\mathbf{E}_a(L) = g_{ab} G^b. \qquad (7)$$

Now let ξ^a be a Killing vector field and ∇_U the covariant derivative along the geodesic. Then

$$\nabla_U(\xi_a U^a) = \xi_{a;b} U^b U^a + \xi_a U^a{}_{;b} U^b = 0. \qquad (8)$$

The two terms between the equality signs vanish separately. This is the usual relativistic calculation showing that $\xi_a U^a$ is a constant of the motion for the geodesic. The corresponding (but different) calculation using the total derivative is

$$\begin{aligned} D_\lambda(\xi_a x'^a) &= (D_\lambda \xi_a) x'^a + \xi_a D_\lambda x'^a = \xi_{a,b} x'^a x'^b + \xi_a x''^a \\ &= \Gamma^c{}_{ab} \xi_c x'^a x'^b + \xi_a x''^a = \xi_a G^a = \xi^a \mathbf{E}_a(L), \end{aligned} \qquad (9)$$

where we have used the Killing equation $\xi_{(a,b)} = \Gamma^c{}_{a,b}\xi_c$ and $G^a = g^{ab}\mathbf{E}_b(L)$ from Eq. (7). Note that the two terms after the first equality sign do not vanish separately. The characteristic form of a conservation law is $\operatorname{Div} P = Q \cdot \mathbf{E}(L)$, where P is the conserved quantity and $\operatorname{Div} = D_\lambda$ since $\mathscr{X}$ is one dimensional. Hence we conclude that ξ^a is the characteristic corresponding to the constant of the motion $\xi_a x'^a$. The symmetry generator is therefore $\mathbf{v} = \xi^a \partial/\partial x^a$ and coincides with the Killing field itself. Thus the Killing vector fields plays two roles. It generates both a space-time symmetry and a symmetry of the geodesic equations.

For a geodesic symmetry of the form $\mathbf{v} = \phi^a \partial/\partial x^a$, where ϕ^a does not depend on λ, it is possible to define a natural action on the metric by regarding the metric as a field on the one-jet space of the geodesic equations according to

$$ds^2 = g_{ab} x'^a x'^b \, d\lambda^2 = 2L \, d\lambda^2,$$

where $d\lambda^2$ is treated as a constant. Thus the metric coincides with the Lagrangian in this interpretation (up to a numerical factor). The action of a symmetry on the Lagrangian is given by $\operatorname{pr} \mathbf{v}(L)$, where $\operatorname{pr} \mathbf{v}$ is the (first) prolongation of $\mathbf{v}$,

$$\operatorname{pr} \mathbf{v} = \mathbf{v} + \phi^{(1)a} \frac{\partial}{\partial x'^a}, \qquad (10)$$

where $\phi^{(1)}$ is the first prolongation coefficient given by

$$\phi^{(1)a} = D_\lambda \phi^a = \phi^a{}_{,b} x'^b. \qquad (11)$$

The prolonged action on the Lagrangian by $\mathbf{v} = \phi^a \partial/\partial x^a$ is then calculated as

$$\begin{aligned} &\operatorname{pr} \mathbf{v}(L) \\ &= \tfrac{1}{2} \operatorname{pr} \mathbf{v}(g_{ab}) x'^a x'^b + g_{ab} x'^a \operatorname{pr} \mathbf{v}(x'^b) \\ &= \tfrac{1}{2}(\phi^c g_{ab,c} + 2 g_{ac} \phi^c{}_{,b}) x'^a x'^b = \tfrac{1}{2}(\pounds_{\mathbf{v}} g_{ab}) x'^a x'^b, \end{aligned} \qquad (12)$$

where $\pounds_{\mathbf{v}}$ is the Lie derivative with respect to $\mathbf{v}$. If ϕ^a are the components of a Killing vector field, then $\pounds_{\mathbf{v}} g_{ab} = 0$ leading to $\operatorname{pr} \mathbf{v}(L) = 0$. In general, the criterion for a symmetry $\mathbf{v} = Q^a \partial/\partial x^a$ to be a variational symmetry is $\operatorname{pr} \mathbf{v}(L) = \operatorname{Div} B$ for some function B on some (finite-order) jet space. Symmetries with $B \neq 0$ are *divergence symmetries.* Thus the Killing vector symmetry is a nondivergence symmetry. The action on the metric becomes

$$\operatorname{pr} \mathbf{v}(ds^2) = (\pounds_{\mathbf{v}} g_{ab}) dx^a \, dx^b, \qquad (13)$$

which shows that $\operatorname{pr} \mathbf{v}(ds^2) = 0$ for a Killing vector symmetry.

It is the nondivergence property that makes the Killing vector symmetries especially useful for the solution generating methods of general relativity. The approach there is to look for invariances of a "decoupled" part of the total Lagrangian (see, e.g., Kramer *et al*[7]). By restricting attention to nondivergence symmetries only the decoupled variables come into play.

The general criterion for a geodesic symmetry (variational or nonvariational) is that $\operatorname{pr} \mathbf{v}(G^a)$ vanishes on shell. Here $\operatorname{pr} \mathbf{v}$ stands for the second prolongation of $\mathbf{v}$. For a symmetry involving derivatives up to the first order, an equivalent condition is that

$$\operatorname{pr} \mathbf{v}(G^a) = (P^{(0)a}{}_b + P^{(1)a}{}_b D_\lambda) G^b$$

for some jet space functions $P^{(0)a}{}_b$ and $P^{(1)a}{}_b$. Prince and Crampin only allowed for a right-hand side proportional to G^a, which seems to be overly restrictive.

III. KILLING TENSOR SYMMETRIES

A second rank Killing tensor ξ_{ab} is symmetric and satisfies $\xi_{(ab;c)} = 0$ (Kramer *et al*[7]). Then $\xi_{ab} x'^a x'^b$ is a quadratic geodesic constant of the motion. To find its characteristic we apply the total derivative, as in Eq. (9), leading to

$$D_\lambda(\xi_{ab} x'^a x'^b) = \xi_a{}^b x'^a \mathbf{E}_b(L), \qquad (14)$$

where we have used the Killing tensor relation $\xi_{(ab,c)} = 2\Gamma^d{}_{(ab}\xi_{c)d}$. It follows that the characteristic is $Q^a = \xi_b{}^a x'^b$. Thus unlike the Killing vector case this symmetry depends on the derivative x'^a and therefore it cannot be interpreted as a space-time vector field. A symmetry is a point symmetry if the characteristic can be written as $Q^a = \phi^a - x'^a \psi$ for some functions ϕ and ψ of λ and x^a. A simple calculation shows that a Killing tensor symmetry is a point symmetry only in the trivial case when ξ_{ab} is proportional to the metric. In that case the symmetry is $\mathbf{v} = \partial/\partial\lambda$, reflecting the fact that the geodesic equations do not contain the affine parameter explicitly. The corresponding conserved quantity is the length of the tangent vector.

As before we can compute the action of the symmetry on the metric and on the Lagrangian. The first prolongation coefficient of $\mathbf{v} = \xi^a{}_b x'^b \partial/\partial x^a$ is given by

$$\phi^{(1)a} = D_\lambda(\xi^a{}_b x'^b) = \xi^a_{b,c} x'^b x'^c + \xi^a{}_b x''^b. \qquad (15)$$

The action of the symmetry on the Lagrangian then becomes

$$\operatorname{pr} \mathbf{v}(L) = \Gamma^d{}_{ab} \xi_{cd} x'^a x'^b x'^c + \xi_{ab} x'^a x''^b, \qquad (16)$$

where we have used the Killing tensor relation $\xi_{(ab,c)} = 2\Gamma^d_{(ab}\xi_{c)d}$. It follows that the symmetry is a divergence symmetry or in other words, the action of the Killing tensor symmetry on the Lagrangian (and the metric) does not vanish by contrast to the Killing vector case. From the general theory we know that pr $\mathbf{v}(L) = \text{Div } B = D_\lambda B$ for some function B. To find B we note that Eq. (16) can be written as

$$\text{pr } \mathbf{v}(L) = \xi^a{}_b x'^b \mathbf{E}_a(L) . \quad (17)$$

Therefore the symmetry action vanishes on-shell. It is this fact that lies behind the result of Prince and Crampin.[4] Comparison with (14) shows that the right-hand side of (17) can be written as Div B, where B is the conserved quantity, i.e., $B = \xi_{ab} x'^a x'^b$. This simple relation is a consequence of the quadratic nature of the constant of the motion.

IV. CONCLUDING REMARKS

One might ask whether the Killing vector symmetries are the only variational point symmetries of the form $\mathbf{v} = \xi^a(x)\partial/\partial x^a$. That this is indeed the case can be seen from the following argument. It follows from Eq. (1) that

$$\text{pr } \mathbf{v}(L) = (\pounds_\mathbf{v} g_{ab}) x'^a x'^b \equiv \bar{g}_{ab} x'^a x'^b \equiv \bar{L}.$$

Thus $\mathbf{v}$ is a variational symmetry if the new Langrangian $\bar{L}$ is a total divergence, i.e., $\bar{L} = D_\lambda B$ for some B. This is the case if the Euler-Langrange equations of $\bar{L}$ vanish identically.[5] But $\mathbf{E}_a(\bar{L}) = \bar{g}_{ab}(x''^b + \bar{\Gamma}^b{}_{cd} x'^c x'^d)$ so $\mathbf{E}_a(\bar{L})$ is identically zero only if $\bar{g}_{ab} = \pounds_\mathbf{v} g_{ab} = 0$, so $\mathbf{v}$ is indeed a Killing vector symmetry.

[1] S. Benenti and M. Francaviglia, *General Relativity and Gravitation* edited by A. Held (Plenum, New York, 1980).
[2] E. G. Kalnins and W. Miller, Jr., SIAM J. Math. Anal. **11**, 1011 (1980).
[3] G. E. Prince, and M. Crampin, Gen. Relativ. Gravit. **16**, 921 (1984).
[4] G. E. Prince, and M. Crampin, Gen. Relativ. Gravit. **16**, 1063 (1984).
[5] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).
[6] K. Rosquist, unpublished.
[7] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB Deutscher, Berlin, 1980).

Initial value problem for colliding gravitational plane waves. II

Isidore Hauser
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676
Frederick J. Ernst
Department of Mathematics and Computer Science and Department of Physics, Clarkson University, Potsdam, New York 13676

(Received 14 April 1989; accepted for publication 17 May 1989)

As a preliminary step in the development of a Hilbert problem (HP) approach to the initial value problem (IVP) for colliding gravitational plane waves with noncollinear polarizations, the IVP for colliding gravitational plane waves with *collinear* polarizations is reformulated in two different ways as an HP in a complex plane. The solutions of both forms of the HP are found and each of these agrees with the solution obtained by another method in the previous paper of this series [I. Hauser and F. J. Ernst, J. Math. Phys. **30**, 872 (1989)]. The conditions imposed on the initial data of the IVP by the vacuum field equations are discussed in detail. Anticipating the next paper of this series, the generalization of one form of the HP to noncollinear polarizations is briefly described.

I. INTRODUCTION

A. Objective

This is the second of a series[1] of papers on the initial value problem (IVP) for colliding gravitational plane waves, i.e., on the search for systematic methods of computing the outcoming scattered wave when the two incoming plane waves are prescribed. One method that shows some promise has been developed by the present authors. It involves replacing the usual formulation of the IVP in terms of a nonlinear partial differential equation, viz. the Ernst equation,[2] by an equivalent 2×2 matrix homogeneous Hilbert problem (HHP) in a complex plane. This HHP will be sketched in Sec. V and will be covered in full detail and geneality in a future paper.

The present paper will be chiefly devoted to a relatively simple one-dimensional Hilbert problem which is equivalent to the restriction of our matrix HHP to the case when the polarizations of the plane waves are collinear. This pursuit of the collinear case as opposed to an immediate exposition of the generally applicable matrix HHP has good reason. The point is that the key ideas of the matrix HHP will be nicely illustrated by the Hilbert problem (HP) for the collinear case and we shall thereby be able to introduce many concepts of value for subsequent papers without having to cope immediately with mathematical difficulties which beset the noncollinear case. These difficulties are illustrated by the fact that there is strong evidence that the general solution of the IVP for the noncollinear case is not expressible in a finite closed form. Moreover, the practical art of solving the matrix HHP for particular noncollinear cases is still in its infancy.

In contrast, the IVP for the collinear case requires only that one solve a Cauchy problem for a certain *linear* hyperbolic partial differential equation and it so happens that the general solution in a finite closed form is known. The first complete solution was obtained by Szekeres[3] by employing the Green's formula method of Riemann (not to be confused with the Riemann–Hilbert problem). Another form of the general solution for the collinear case was obtained by the present authors[1] by employing a classical method of linear superposition. In the present paper, yet another form of the general solution will be obtained by the radically different method of solving an HP and will be shown to be closely related in an instructive way to the solution in Ref. 1.

To enable us to describe the contents of this paper in greater detail, some specifics on the IVP for the collinear case, on several important concepts that will be used throughout the paper, and on the HP and its solution will now be given.

B. The IVP

The line element in the space-time region that is both covered by our chart[4] and occupied by the scattered wave is, in the collinear case,

$$ds^2 = \rho[e^{-2\psi}(dx^1)^2 + e^{2\psi}(dx^2)^2] - 2\rho^{-1/2}e^{2\Gamma}\,du\,dv, \tag{1.1}$$

where, ρ, ψ, and Γ depend only on the coordinates u and v over a simply connected planar domain

$$\text{IV}:=\{(u,v):\ 0\leqslant u<u_0, 0\leqslant v<v_0, 0<\rho(u,v)\}. \tag{1.2}$$

Each of the constants u_0 and v_0 is a real positive number or ∞. The ignorable coordinates x^1 and x^2 are scaled so that $\rho(0,0)=1$ and $\psi(0,0)=0$. The vacuum field equations imply that[5]

$$\rho(u,v)=\tfrac{1}{2}[s(v)-r(u)], \tag{1.3}$$

where $r(u)$ is a monotonic increasing function over $0\leqslant u<u_0$, $s(v)$ is a monotonic decreasing function over $0\leqslant v<v_0$, and

$$r(0)=-1,\quad s(0)=1,$$
$$-1<r_0\leqslant 1,\quad -1\leqslant s_0<1, \tag{1.4}$$

where

$$r_0:=\lim_{u\to u_0} r(u),\quad s_0:=\lim_{v\to v_0} s(v).$$

As was detailed in Ref. 1, the field $\Gamma(u,v)$ is simply ex-

 0022-2488/89/102322-15$02.50

pressed in terms of definite integrals[6] once $\psi(u,v)$ is known.

The key problem is to find the solution ψ of the hyperbolic field equation

$$2\rho\psi_{uv}+\rho_u\psi_v+\rho_v\psi_u=0, \quad (1.5)$$

where $\psi_u:=\partial\psi/\partial u$, etc., corresponding to the prescribed initial data

$$r(u),\quad s(v),\quad \psi_3(u):=\psi(u,0),\quad \psi_2(v):=\psi(0,v). \quad (1.6)$$

This is the essence of our IVP.

Some constraints on the initial data that are consistent with the existence and vanishing of the Ricci tensor are that the initial data functions (1.6) be of differentiability class $\mathbf{C}^1$ and satisfy $\psi_3(0)=\psi_2(0)=0$, Eqs. (1.4), and

$$\dot r(u)>0 \text{ if } 0<u,\quad \dot s(v)<0 \text{ if } 0<v, \quad (1.7)$$

where $\dot r(u):=dr(u)/du$, etc. *We shall follow the precedent of Ref. 1 and, except in a part of Sec. IV, impose no constraints on the initial data other than those just stated.* There are, however, additional constraints which will be detailed in Sec. IV and which are imposed by the requirements that the Ricci tensor exist and vanish. By ignoring these constraints, we are actually treating a broader class of metrics than those that strictly represent colliding gravitational plane waves. However, this does not hinder our ability to solve the IVP and even turns out to be useful, as we shall discuss in Sec. V.

C. Definitions of D_{IV}, γ, γ_3, γ_2, g_3, and g_2

In Ref. 1 we employed r and s to designate functions, as we have done above, and to designate coordinate variables.[7] We shall follow the same dual usage here and depend on context to distinguish one meaning from the other.

Note that the mapping $(u,v)\to(r(u),s(v))$ is one-to-one, bicontinuous, and maps IV as defined by Eqs. (1.2) and (1.3) onto

$$D_{IV}:=\{(r,s):\ -1\leqslant r<r_0, s_0<s\leqslant 1, r<s\}, \quad (1.8)$$

where r_0 and s_0 are defined by Eqs. (1.4). Henceforth, we shall almost everywhere employ D_{IV} instead of IV as a domain. The reasons for this are that D_{IV} is easier to visualize than IV and the use of D_{IV} leads to simpler expressions.

Definitions: We shall denote by γ, γ_3, and γ_2 those functions whose domains are D_{IV}, the interval $-1\leqslant r<r_0$, and the interval $s_0<s\leqslant 1$, respectively, and whose values $\gamma(r,s)$, $\gamma_3(r)$, and $\gamma_2(s)$ are given by

$$\begin{aligned}&\gamma(r(u),s(v)):=\psi(u,v),\\ &\gamma_3(r(u)):=\psi_3(u),\\ &\gamma_2(s(v)):=\psi_2(v).\end{aligned} \quad (1.9)$$

From Eqs. (1.4) and (1.6), note that

$$\gamma_3(r)=\gamma(r,1),\quad \gamma_2(s)=\gamma(-1,s). \quad (1.10)$$

Definitions: We shall denote by g_3 and g_2 those functions whose domains are the open intervals $-1<\sigma<r_0$ and $s_0<\sigma<1$, respectively, and whose values are given by

$$g_3(\sigma):=\int_{-1}^{\sigma}dr\,\frac{\dot\gamma_3(r)}{\sqrt{\sigma-r}},\quad g_2(\sigma):=-\int_{1}^{\sigma}ds\,\frac{\dot\gamma_2(s)}{\sqrt{s-\sigma}}, \quad (1.11)$$

where $\dot\gamma_3(r):=d\gamma_3(r)/dr$, etc.

We employ the Lebesgue definition of an integral as in Ref. 1, where the existences and some properties of g_j ($j=3,2$) were established.[8]

D. The spectral potential ϕ

The main ingredient of our HP is a function $\phi(r,s,\tau)$ of a complex (spectral) parameter τ as well as of r and s. The definition of ϕ in terms of a given γ will be given in Sec. II, where we shall also derive some pertinent properties of ϕ. One of these properties is that $\phi(r(u),s(v),\tau)$ satisfies the same hyperbolic equation (1.5) as $\psi(u,v)$. Another property is that ϕ is uniquely determined by γ. In particular, the initial values of ϕ are uniquely determined by the initial data as follows:

$$\begin{aligned}\phi_3(r,\tau)&:=\phi(r,1,\tau)\\ &=-\frac{\chi_3(r,\tau)}{\tau+1}\int_{-1}^{r}dr'\,\chi_3(r',\tau)\dot\gamma_3(r'),\\ \phi_2(s,\tau)&:=\phi(-1,s,\tau)\\ &=-\frac{\chi_2(s,\tau)}{\tau-1}\int_{1}^{s}ds'\,\chi_2(s',\tau)\dot\gamma_2(s'),\end{aligned} \quad (1.12)$$

where

$$\begin{aligned}\chi_3(r,\tau)&:=[(\tau+1)/(\tau-r)]^{1/2},\\ \chi_2(s,\tau)&:=[(\tau-1)/(\tau-s)]^{1/2},\end{aligned} \quad (1.13)$$

$\chi_3(-1,\tau)=\chi_2(1,\tau):=1$ and, for fixed $r\neq -1$ and $s\neq 1$, we employ those holomorphic branches of $\chi_3(r,\tau)$ and $\chi_2(s,\tau)$ that have the cuts $[-1,r]$ and $[s,1]$, respectively, on the real axis of the τ plane and satisfy $\chi_3(r,\infty)=\chi_2(s,\infty):=1$. The initial values of ϕ can also be determined by using

$$\begin{aligned}\phi_3(r,\tau)&=\frac{1}{\pi}\int_{-1}^{r}d\sigma\,\frac{g_3(\sigma)}{\sqrt{r-\sigma}(\sigma-\tau)},\\ \phi_2(s,\tau)&=\frac{1}{\pi}\int_{1}^{s}d\sigma\,\frac{g_2(\sigma)}{\sqrt{\sigma-s}(\sigma-\tau)}\end{aligned} \quad (1.14)$$

if g_3 and g_2 have already been found. Note that for fixed $r\neq -1$, $\phi_3(\tau)=\phi_3(r,\tau)$ is a holomorphic function of τ throughout $C-[-1,r]$, where C is the extended complex plane. Likewise, for fixed $s\neq 1$, $\phi_2(\tau)$ is holomorphic on $C-[s,1]$.

As regards other properties of ϕ that will be obtained in Sec. II, consider any fixed (r,s) in D_{IV}. Then the following statements hold for $\phi(\tau)=\phi(r,s,\tau)$[9]:

$$\begin{aligned}&\phi(\tau)\text{ is holomorphic on } C-([-1,r]\cup[s,1]),\\ &\phi(\tau)-\chi_2(\tau)\phi_3(\tau)\text{ is holomorphic on }[-1,r],\\ &\phi(\tau)-\chi_3(\tau)\phi_2(\tau)\text{ is holomorphic on }[s,1],\end{aligned} \quad (1.15)$$

and

$$\phi(\infty)=0,\quad [-\tau\phi(\tau)]_{\tau=\infty}=\gamma. \quad (1.16)$$

Statements (1.15) and (1.16) will constitute the HP for the collinear case.

Again consider any fixed (r,s) in $D_{\rm IV}$ and suppose $r \neq -1$ or $s \neq 1$ (or both). The remaining properties of $\phi(\tau)$ that will be derived in Sec. II concern its boundary values $\partial^{\pm}\phi(\sigma) = \partial^{\pm}\phi(r,s,\sigma)$, which are defined below for all σ on the union of the open intervals $-1<\sigma<r$ and $s<\sigma<1$.

Definitions: Let h_σ be any complex number such that $\operatorname{Im} h_\sigma > 0$ if $-1<\sigma<r$ and $\operatorname{Im} h_\sigma <0$ if $s<\sigma<1$. Then $\partial^{+}\phi(\sigma)$ and $\partial^{-}\phi(\sigma)$ are defined by[10]

$$\partial^{\pm}\phi(\sigma) := \lim_{h_\sigma \to 0} \phi(\sigma \pm h_\sigma). \tag{1.17}$$

In Sec. II we shall prove that $\partial^{\pm}\phi(\sigma)$ exist and satisfy

$$\partial^{+}\phi(\sigma) - \partial^{-}\phi(\sigma) = 2i\omega(\sigma)f(\sigma), \tag{1.18}$$

where

$$\omega(\sigma) := \sqrt{(1-\sigma^2)/(r-\sigma)(s-\sigma)} \tag{1.19}$$

and

$$f(\sigma) := \begin{cases} g_3(\sigma)/\sqrt{1+\sigma}, & \text{if} \quad -1<\sigma<r, \\ g_2(\sigma)/\sqrt{1-\sigma}, & \text{if} \quad s<\sigma<1. \end{cases} \tag{1.20}$$

Equation (1.18) will be the basis for an alternative useful form of the HP for the collinear case.

E. The HP and its solution

The HP and its solution will be covered in Sec. III. We shall give both here without providing the derivation of the solution and other proofs.

Assume that we have prescribed initial data and have used Eqs. (1.12) to compute ϕ_3 and ϕ_2. However, since we do not yet know γ, we cannot compute ϕ by using its definition, in terms of γ. So let us lay aside the definition of ϕ in terms of γ and seek an alternative definition, viz. one that regards ϕ as the solution of a certain HP.

For any fixed (r,s) in $D_{\rm IV}$, the HP[11] is to find a function $\phi(\tau)$ which satisfies statements (1.15) and the equation $\phi(\infty)=0$. In Sec. III we shall prove that the solution is unique and is

$$\phi(\tau) = -\frac{1}{2\pi i}\int_{\Gamma_3} d\tau' \frac{\chi_2(\tau')\phi_3(\tau')}{\tau'-\tau} - \frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \frac{\chi_3(\tau')\phi_2(\tau')}{\tau'-\tau}, \tag{1.21}$$

where Γ_3 and Γ_2 are any positively oriented (simple and smooth) contours such that

$$\begin{aligned} &[-1,r] \subset \Gamma_3^{+}, \quad [s,1] \subset \Gamma_3^{-}, \\ &[s,1] \subset \Gamma_2^{+}, \quad [-1,r] \subset \Gamma_2^{-}, \\ &\tau \in C - (\Gamma_3 \cup \Gamma_3^{+} \cup \Gamma_2 \cup \Gamma_2^{+}), \end{aligned} \tag{1.22}$$

and Γ_j^{+} and Γ_j^{-} are those open subsets of C which are bounded and unbounded, respectively, and have Γ_j as their common boundary. In descriptive terms, Γ_j^{+} and Γ_j^{-} are the open regions inside and outside Γ_j, respectively. Typical choices of Γ_3, Γ_2, and τ are shown in Fig. 1.

We define γ in terms of the above HP solution by the second of Eqs. (1.16), whereupon

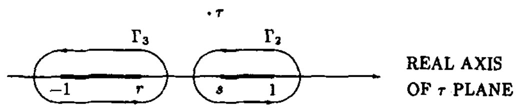


FIG. 1. Illustrative choices of Γ_3, Γ_2, and τ for the contour integrals in the solution of the HP adapted to (ϕ_3,ϕ_2).

$$\gamma = -\frac{1}{2\pi i}\int_{\Gamma_3} d\tau' \chi_2(\tau')\phi_3(\tau') - \frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \chi_3(\tau')\phi_2(\tau'). \tag{1.23}$$

In Sec. III we shall prove that Eq. (1.23) solves the IVP for the collinear case. Insertion of expressions (1.12) into Eq. (1.23) and interchange of the order of integration over r' and τ' yields a Green's function form of the solution obtained in Ref. 1 by other means.[12] The same process applied to Eq. (1.21) would yield a Green's function form for $\phi(\tau)$.

F. A second form of the HP and its solution

If one permits the contours Γ_3 and Γ_2 to contract until they "collapse onto" the cuts $[-1,r]$ and $[s,1]$ they enclose, then one obtains an expression for $\phi(\tau)$ in terms of definite integrals on $[-1,r]$ and $[s,1]$. Another way of arriving at this definite integral expression for $\phi(\tau)$ is by solving an alternative form of our HP, viz. an HP on

$$L(r,s) := \text{the pair of oriented arcs } -1 \text{ to } r \text{ and } 1 \text{ to } s \text{ on the real axis of } C. \tag{1.24}$$

Specifically, assume that g_3 and g_2 have been computed from the initial data by using (for example) Eqs. (1.11). Then *the HP on $L(r,s)$ is to find a $\phi(\tau)$ which is holomorphic on $C - L(r,s)$ such that $\phi(\infty)=0$, which has $\partial^{\pm}\phi(\sigma)$ existing such that Eq. (1.18) holds for all σ in the open intervals $-1<\sigma<r$ and $s<\sigma<1$ and which satisfies the following endpoint conditions:*

$$\phi(\tau) - \chi_2(\tau)\phi_3(\tau) \text{ is bounded as } \tau \to -1 \text{ and } \tau \to r, \tag{1.25}$$
$$\phi(\tau) - \chi_3(\tau)\phi_2(\tau) \text{ is bounded as } \tau \to 1 \text{ and } \tau \to s,$$

where ϕ_3 and ϕ_2 are given by Eqs. (1.14). The solution will be proven in Sec. III to be the line integral

$$\phi(\tau) = \frac{1}{\pi}\int_{L(r,s)} d\sigma \frac{\omega(\sigma)f(\sigma)}{\sigma-\tau}. \tag{1.26}$$

The second of Eqs. (1.16) then yields

$$\gamma = \frac{1}{\pi}\int_{L(r,s)} d\sigma\, \omega(\sigma)f(\sigma), \tag{1.27}$$

which is precisely the solution obtained in Ref. 1 by other means.

G. On g_3 and g_2

Section IV will cover properties of g_3 and g_2 which are important for calculations and perhaps for later extensions to the generalizations of g_3 and g_2 which occur in the noncollinear case. For example, we shall prove that g_j obeys a

Hölder condition of index $\frac{1}{2}$ on any closed subinterval of its (open interval) domain and we shall derive new integral expressions which facilitate computations of g_j. Also, we shall obtain useful implications concerning g_j for the members of a broad subclass of the class of all initial data sets (1.6) which satisfy all constraints imposed by the existence and vanishing of the Ricci tensor.

II. THE SPECTRAL POTENTIAL ϕ

A. Definitions of the space D_{IV}, the potential γ, and the duality operator *

There are two preliminary topics that we shall cover in Secs. II A and II B, respectively, before we define ϕ. The first topic is on the potential γ and its domain D_{IV} which were introduced in Sec. I C. However, we shall begin Sec. II without presupposing all the concepts and relations that were mentioned in Sec. I. The only concepts that we assume to be given at this time are the initial data functions $r(u)$, $\psi_3(u)$ and $s(v)$, $\psi_2(v)$.

Premises: The initial data functions $r(u)$, $\psi_3(u)$ and $s(v)$, $\psi_2(v)$ are C^1 over their domains $0\leqslant u < u_0$ and $0\leqslant v < v_0$, respectively, and satisfy $\psi_3(0) = \psi_2(0) = 0$ and Eqs. (1.7) and (1.4).

Definitions: The set IV is defined by Eqs. (1.2) and (1.3). The set D_{IV} is defined by Eqs. (1.4) and (1.8). The *topologies* of IV and D_{IV} will be the sets of all IV$\cap S$ and $D_{IV} \cap S$, respectively, such that S is any set in the usual topology of R^2. The *boundary* of D_{IV} will be

$$\partial D_{IV} := \{(r,s)\in D_{IV} : r = -1 \text{ or } s = 1\}.$$

Differentiation of any function is defined using only those sequences of points that lie in the domain of the function.

Definitions: We shall denote by Σ that one-to-one bicontinuous mapping of D_{IV} onto IV such that

$$\Sigma^{-1}(u,v) := (r(u),s(v)). \tag{2.1}$$

We assign that atlas to D_{IV} that consists of all charts $\Sigma' : D_{IV} \to R^2$ for which

$$(\Sigma' \circ \Sigma^{-1})(u,v) = (U(u),V(v)) =: (u',v'), \tag{2.2}$$

where U and V are any functions with domains $0\leqslant u < u_0$ and $0\leqslant v < v_0$, respectively, such that U and V are C^1 and have positive derivatives throughout their domains.

Note: The ordered pair that consists of D_{IV} and the atlas defined above is not a C^1 manifold since the atlas is not maximal. However, many of the concepts and results of manifold theory are clearly applicable to this structure and will be freely used in Sec. II.

Definition: We shall denote by γ any real-valued function whose domain is D_{IV} such that the function $\psi := \gamma\circ\Sigma^{-1}$ is C^1, has a continuous mixed second-order partial derivative, satisfies Eq. (1.5) throughout IV, and satisfies $\psi(u,0) = \psi_3(u)$ and $\psi(0,v) = \psi_2(v)$.

Definition: The symbol * will denote that duality operator on one-forms in IV such that

$$*du = du, \quad *dv = -dv \tag{2.3}$$

and will also denote a duality operator on one-forms in D_{IV} such that

$$*dr = dr, \quad *ds = -ds.$$

We shall rely on context to avoid confusion between the above two uses of the asterisk.

Note: It is easy to prove that for any one-forms ω_1 and ω_2 in IV (or in D_{IV})

$$**\omega_1 = \omega_1, \quad (*\omega_1)\omega_2 = -\omega_1(*\omega_2), \tag{2.4}$$

where $(*\omega_1)\omega_2 := (*\omega_1)\wedge\omega_2$. *We omit the wedges in all exterior products and exterior derivatives of differential forms.*

Note: Consider any one-form

$$\lambda(u,v) = du\,\alpha(u,v) + dv\,\beta(u,v)$$

which is defined and continuous on an open set M in IV. Here α and β need not be C^1 to construct a useful definition of $d\lambda$. It is sufficient to assume that α_v and β_u exist and are continuous, whereupon one way of defining the exterior derivative of λ is as follows:

$$d\lambda(u,v) := du\,dv[\beta_u(u,v) - \alpha_v(u,v)].$$

Note that the hyperbolic equation (1.5) is expressible as the two-form equation

$$d(\rho * d\psi) = 0. \tag{2.5}$$

Without entering into any specifics, we remark that one can define a two-dimensional tangent vector space and its dual at each point of D_{IV} in a manner similar to that used for differentiable manifolds. Then differential forms in D_{IV} can be introduced. We shall denote by Σ^* the pullback operator corresponding to the chart Σ. The key fact that we use below is that for each p-form μ in D_{IV} such that the domain of μ is a set M in the topology of D_{IV}, there is exactly one p-form λ in IV such that the domain of λ is $\Sigma(M)$ and $\mu = \Sigma^*\lambda$.

Definitions: The p-form μ is *continuous* if and only if λ is continuous and *$d\mu$ exists* if and only if $d\lambda$ exists and

$$d\mu := \Sigma^*(d\lambda).$$

Also, an alternative definition for the duality operation on any one-form μ is

$$*\mu := \Sigma^*(*\lambda).$$

It is easily proven that the above concepts are independent of the choice of the chart Σ. As examples of interest to us,

$$\Sigma^*\psi = \gamma, \quad \Sigma^*(d\psi) = d\gamma, \quad \Sigma^*(d*d\psi) = d*d\gamma$$

all exist and are continuous on the domain D_{IV}. Also, $\Sigma^*(d^2\psi) = d^2\gamma$ exists and vanishes on D_{IV}. Application of Σ^* to Eq. (2.5) yields

$$d(\rho*d\gamma) = 0, \text{ where } \rho(r,s) := \tfrac{1}{2}(s-r) \tag{2.6}$$

and where we note that ρ is assigned the dual role of denoting that function with domain IV for which $\rho(u,v) = \frac{1}{2}[s(v) - r(u)]$ and that function with domain D_{IV} for which $\rho(r,s) = \frac{1}{2}(s-r)$. The intended meaning of ρ will be clear from the context in which it appears.

Equation (2.6) requires a cautionary addendum. Our premises on $r(u)$ and $s(v)$ imply that *γ is continuous, that $\gamma_r := \partial\gamma/\partial r$ exists and is continuous at all (r,s) in D_{IV} such that $s\neq 1$, and that γ_{rs} exists and is continuous at all (r,s) in $D_{IV} - \partial D_{IV}$.* Consider, for example, γ_r. From Eqs. (1.4), (1.7), and the relation $\gamma(r(u),s(v)) = \psi(u,v)$,

$$\gamma_r(r(u),s(v)) = \psi_u(u,v)/\dot r(u)$$

if $u > 0$, i.e., if $r(u) > -1$. However, since $\dot{r}(0)$ may be zero, $\gamma_r(r(u),s(v))$ may not exist at $u=0$, i.e., at $r(0) = -1$. *In fact, as we demonstrated in Ref. 1, the vacuum field equations imply that $\dot{r}(0)=\dot{s}(0)=0$ and that $\gamma_r(r,s)$ is unbounded as $r\to -1$ and $\gamma_s(r,s)$ is unbounded as $s\to 1$.*[13]

Therefore, although $d\gamma$ exists and is continuous at all points of $D_{\rm IV}$, *one must note that the identity mapping of D_{IV} onto D_{IV} is not generally a chart in our* $\mathbf{C}^1$ *atlas* and that

$$d\gamma(r,s) = dr\,\gamma_r(r,s) + ds\,\gamma_s(r,s)$$

generally applies only to (r,s) in $D_{\rm IV} - \partial D_{\rm IV}$. The same reservation holds for

$$\tfrac{1}{2}\,dr\,ds[2(s-r)\gamma_{rs} - \gamma_s + \gamma_r] = 0, \tag{2.7}$$

which is the restriction of Eq. (2.6) to $D_{\rm IV} - \partial D_{\rm IV}$.

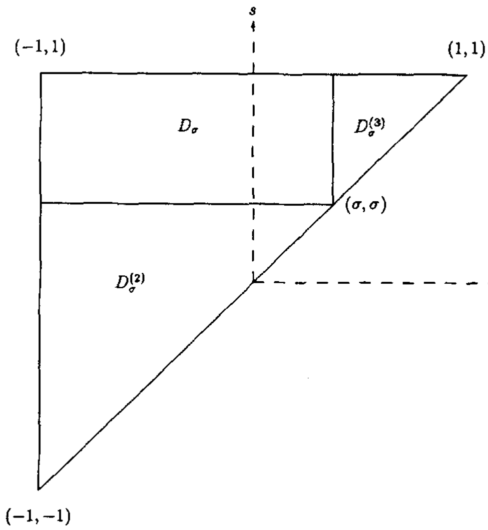


FIG. 2. The subregions D_σ, $D_0^{(3)}$, and $D_\sigma^{(2)}$ of $D_{\rm IV}$ when $r_0 = 1$, $s_0 = -1$, and $\tau = \sigma$, where $-1 < \sigma < 1$. The vertices of the triangular region $D_{\rm IV}$ are $(-1,1)$, $(1,1)$, and $(-1,-1)$.

B. Definitions of χ, D, $D_{(r,s)}$, D_τ, and $D_\sigma^{(j)}$

Our second preliminary topic before we define ϕ concerns a family of complex-valued solutions χ of

$$d(\rho * d\chi) = 0 \tag{2.8}$$

such that χ is a function of r times a function of s and $(r,s)\in D_{\rm IV}$. The reader can verify that a solution of this kind is given by

$$\chi(r,s,\tau) := \chi_3(r,\tau)\chi_2(s,\tau), \tag{2.9}$$

where τ is a complex separation parameter and can be any point in $C - \{-1,1\}$. For any fixed (r,s) in $D_{\rm IV}$, definitions (1.13) of χ_3 and χ_2 imply that the domain of $\chi(\tau) = \chi(r,s,\tau)$ in the τ plane is

$$D_{(r,s)} := C - ([-1,r] \cup [s,1]) \tag{2.10}$$

and that $\chi(\tau)$ is holomorphic on $D_{(r,s)}$. The function χ thus has the domain

$$D := \{(r,s,\tau) : (r,s)\in D_{\rm IV}, \tau\in D_{(r,s)}\}, \tag{2.11}$$

and for any fixed τ in $C - \{-1,1\}$ the domain of χ in the (r,s) plane (i.e., the τ section of D) is

$$D_\tau := \{(r,s)\in D_{\rm IV} : \tau\in D_{(r,s)}\}. \tag{2.12}$$

It is useful to have a mental image of D_τ. This is supplied by the following easily verifiable statements.

(i) The set $D_\tau = D_{\rm IV}$ if $\tau = \infty$, Im $\tau \neq 0$, or if τ is real and $|\tau| > 1$.

(ii) If $\tau = \sigma$ is real and $|\sigma| < 1$, then D_σ is the rectangular region consisting of all (r,s) in $D_{\rm IV}$ such that $-1 \leqslant r < \sigma < s \leqslant 1$. The set D_σ is illustrated in Fig. 2 for the case when $r_0 = 1$ and $s_0 = -1$ (which are the values chosen for r_0 and s_0 in almost every paper on colliding gravitational plane waves). Recall that $-1 < r_0 \leqslant 1$ and $-1 \leqslant s_0 < 1$ and that $-1 \leqslant r < r_0$ and $s_0 < s \leqslant 1$ are the ranges of r and s.

Note the triangular regions $D_\sigma^{(3)}$ and $D_\sigma^{(2)}$ which are also subsets of $D_{\rm IV}$ and appear in Fig. 2. The significances of these regions will now be explained. The boundary values $\partial^\pm \chi_j(\sigma)$ of the holomorphic functions $\chi_j(\tau)$ $(j = 3,2)$ are defined exactly as $\partial^\pm\phi(\sigma)$ were defined in Eq. (1.17). [Simply replace $\phi(\sigma + h_\sigma)$ by $\chi_j(\sigma + h_\sigma)$ in that definition.] From Eqs. (1.13), we obtain

$$\begin{aligned}
&\partial^\pm\chi_3(\sigma) = \mp i\sqrt{(1+\sigma)/(r-\sigma)} \text{ and } \partial^\pm\chi_2(\sigma) = \sqrt{(1-\sigma)/(s-\sigma)}, \text{ if } -1<\sigma<r,\\
&\partial^\pm\chi_3(\sigma) = \sqrt{(1+\sigma)/(\sigma-r)} \text{ and } \partial^\pm\chi_2(\sigma) = \mp i\sqrt{(1-\sigma)/(\sigma-s)}, \text{ if } s<\sigma<1,
\end{aligned} \tag{2.13}$$

where all square roots are positive and (r,s) is any fixed point in $D_{\rm IV}$ such that $r \neq -1$ and/or $s \neq 1$. (Recall that $r < s$ in $D_{\rm IV}$.) From Eqs. (2.9) and (2.13), we further obtain

$$\partial^\pm\chi(\sigma) = \mp i\omega(\sigma), \tag{2.14}$$

where $\omega(\sigma) > 0$ is defined by Eq. (1.19) for all σ such that $-1 < \sigma < r$ or $s < \sigma < 1$.

Now let us consider the domains of $\partial^\pm\chi(\sigma)$ in the (r,s) plane when σ is a fixed real number such that $-1 < \sigma < r_0$ or $s_0 < \sigma < 1$. One sees that $\partial^+\chi(\sigma)$ and $\partial^-\chi(\sigma)$ both have the domain $D_\sigma^{(3)} \cup D_\sigma^{(2)}$, where

$$\begin{aligned}
D_\sigma^{(3)} &:= \{(r,s)\in D_{\rm IV} : -1 < \sigma < r < s \leqslant 1\},\\
D_\sigma^{(2)} &:= \{(r,s)\in D_{\rm IV} : -1 \leqslant r < s < \sigma < 1\}.
\end{aligned} \tag{2.15}$$

As a final note concerning $\partial^\pm\chi(\sigma)$, observe from Eqs. (1.19) and (2.14) that $\partial^\pm\chi(\sigma)$ as well as $\chi(\tau)$ are annihilated by the operator $d\rho * d$.

Since $d(\rho * d\chi(\tau)) = 0$ and since D_τ is simply connected, there is a scalar field whose domain is D_τ and whose gradient is $\rho * d\chi(\tau)$. In fact, the reader can verify that

$$\rho * d\chi(\tau) = -d[(\tau - z)\chi(\tau)], \tag{2.16}$$

where

$$\rho(r,s) := \tfrac{1}{2}(s-r), \quad z(r,s) := \tfrac{1}{2}(s+r). \tag{2.17}$$

It is useful to note that in a neighborhood of $\tau = \infty$,

$$(\tau - z)\chi(\tau) = \tau + O(\tau^{-1}). \tag{2.18}$$

Note: Although the points $(r,s,\tau) = (r,s,\pm 1)$ are not in the domain D of χ, observe that $\chi(-1,s,\tau) = \chi_2(s,\tau)$ and is (for fixed s) holomorphic at $\tau = -1$ *in the sense that it has a holomorphic extension which covers* $\tau = -1$. In the same sense, $\chi(r,1,\tau) = \chi_3(r,\tau)$ is (for fixed r) holomorphic at $\tau = 1$ and $\chi(-1,1,\tau) = 1$ is holomorphic at $\tau = -1$ and $\tau = 1$. Similar remarks apply to the functions Ψ and ϕ that will be defined below.

C. Definitions of Ψ and ϕ

We shall be considering the complex-valued functions χ, Ψ, and ϕ which each have the domain D. In this subsection we shall let $\chi(\tau)$, $\Psi(\tau)$, and $\phi(\tau)$ denote those functions that each have the domain D_τ such that $\chi(\tau)(r,s) := \chi(r,s,\tau)$, $\Psi(\tau)(r,s) := \Psi(r,s,\tau)$, and $\phi(\tau)(r,s) := \phi(r,s,\tau)$. This notational device will not be used in the later parts of Sec. II.

Theorem: There exists a complex-valued function Ψ with domain D such that for any given τ in $C - \{-1,1\}$, $d\Psi(\tau)$ exists, is continuous[14] on D_τ, and satisfies

$$d\Psi(\tau) = (\tau^2 - 1)^{-1}\chi(\tau)(\tau - z - \rho *)d\gamma \tag{2.19}$$

and

$$\Psi(-1,1,\tau) = 0. \tag{2.20}$$

Proof: From Eqs. (2.4), (2.6), and (2.8),

$$d\,[\chi(\tau)(\rho * d\gamma) - \gamma(\rho * d\chi(\tau))] = 0$$

on D_τ. Substituting from Eq. (2.16) into the above, we obtain

$$d\,[\chi(\tau)(\rho * d\gamma) + \gamma d\{(\tau - z)\chi(\tau)\}]$$
$$= d\,[\chi(\tau)(\rho * d\gamma) - (\tau - z)\chi(\tau)d\gamma] = 0,$$

from which we deduce the existence of a complex-valued function F_τ with domain D_τ such that dF_τ exists, is continuous, and satisfies

$$dF_\tau = (\tau^2 - 1)^{-1}\chi(\tau)(\tau - z - \rho *)d\gamma. \tag{2.21}$$

From Eqs. (2.18) and (2.21), dF_τ is holomorphic in a neighborhood of $\tau = \infty$ and

$$dF_\tau = \frac{d\gamma}{\tau} + O(\tau^{-2}).$$

Therefore, we can fix the arbitrary additive function of τ in F_τ by specifying that $F_\tau(-1,1) = 0$. Now let Ψ denote that function whose domain is D and satisfies $\Psi(\tau) = F_\tau$, whereupon the theorem follows.

Corollary: The function is holomorphic in a neighborhood of $\tau = \infty$ and in this neighborhood

$$\Psi(\tau) = \gamma/\tau + O(\tau^{-2}). \tag{2.22}$$

Also, $d^2\Psi(\tau)$ exists and vanishes and $d * d\Psi(\tau)$ exists and is continuous (throughout D_τ).

Proof: The corollary follows from Eq. (2.19), Eq. (2.18), and the facts that $d\gamma$, $d\chi(\tau)$, and $d(*d\gamma)$ exist and are continuous, while $d^2\gamma$ and $d(\rho * d\gamma)$ exist and vanish throughout D_τ.

Definition:

$$\phi := -\chi\Psi. \tag{2.23}$$

Theorem: (i) The function ϕ is uniquely determined by γ and $\phi(\tau)$ is holomorphic in a neighborhood of $\tau = \infty$, where it satisfies Eqs. (1.16).

(ii) The differential $d\phi(\tau)$ exists and is continuous, $d^2\phi(\tau)$ exists and vanishes, and $d * d\phi(\tau)$ exists and is continuous.[14]

(iii) The function $\phi(\tau)$ satisfies

$$\phi(-1,1,\tau) = 0 \tag{2.24}$$

and

$$(\tau - z + \rho *)d\phi(\tau) - dz\,\phi(\tau) = -d\gamma. \tag{2.25}$$

Proof: Parts (i) and (ii) of the above theorem follow directly from definition (2.23) and the preceding theorem and corollary. Equation (2.24) is implied by Eq. (2.20).

Next, Eqs. (1.13), (2.4), (2.9), and (2.17) imply

$$(\tau^2 - 1)^{-1}(\tau - z + \rho *)(\tau - z - \rho *) = [\chi(\tau)]^{-2},$$

whereupon Eq. (2.19) is seen to be equivalent to

$$\chi(\tau)(\tau - z + \rho *)d\Psi(\tau) = d\gamma. \tag{2.26}$$

Furthermore, Eq. (2.16) is equivalent to

$$(\tau - z + \rho *)d\chi(\tau) = dz\,\chi(\tau). \tag{2.27}$$

Equation (2.25) now derives from Eqs. (2.23), (2.26), and (2.27). Q.E.D.

Corollary:

$$d(\rho * d\phi) = 0. \tag{2.28}$$

Proof: Take the exterior derivative of both sides of Eq. (2.25).

D. Holomorphy properties of ϕ

To compute $\phi(r,s,\tau)$ for τ in $C - \{-1,1\}$ and (r,s) in D_τ, we can integrate Eq. (2.19) along any segmentally smooth path which lies entirely in D_τ and which has $(-1,1)$ as its initial point and (r,s) as its final point. Then $\phi(r,s,\tau) = -\chi(r,s,\tau)\Psi(r,s,\tau)$. In particular, let us use the following two paths, which are each composed of a pair of straight line segments on or parallel to the axes of D_{IV}:

$$(-1,1) \to (r,1) \to (r,s), \quad (-1,1) \to (-1,s) \to (r,s).$$

The two integrations are straightforward and yield the two expressions

$$\phi(r,s,\tau) = \chi_2(s,\tau)\phi_3(r,\tau) + \xi_2(r,s,\tau),$$
$$\phi(r,s,\tau) = \chi_3(r,\tau)\phi_2(s,\tau) + \xi_3(r,s,\tau), \tag{2.29}$$

where ϕ_3 and ϕ_2 are given by Eqs. (1.12) and

$$\xi_2(r,s,\tau) := -\frac{\chi_2(s,\tau)}{\tau - 1}\int_1^s db\,\chi_2(b,\tau)\gamma_b(r,b),$$
$$\xi_3(r,s,\tau) := -\frac{\chi_3(r,\tau)}{\tau + 1}\int_{-1}^r da\,\chi_3(a,\tau)\gamma_a(a,s). \tag{2.30}$$

From Eqs. (2.30) the expressions for ϕ_3 and ϕ_2 in Eqs. (1.12), and the definitions of χ_3 and χ_2 in Eqs. (1.13), the following theorem is evident.

Theorem: (i)
$\phi_3(-1,\tau)=\phi_2(1,\tau)=\xi_3(-1,s,\tau)=\xi_2(r,1,\tau)=0.$

(ii) For fixed $r\neq -1$, $\phi_3(r,\tau)$ and $\xi_3(r,s,\tau)$ are holomorphic on the subset $C-[-1,r]$ of the τ plane. For fixed $s\neq 1$, $\phi_2(s,\tau)$ and $\xi_2(r,s,\tau)$ are holomorphic on $C-[s,1]$.

Corollary: Conditions (1.15) hold.

Proof: Use the preceding theorem and Eqs. (2.29).

It is a striking and easily proven statement that for given ϕ_3 and ϕ_2, the holomorphy conditions (1.15) taken together with the condition $\phi(r,s,\infty)=0$ uniquely determine ϕ and, therefore, uniquely determine γ. This will be the basis for our HP.

E. The boundary values $\partial^{\pm}\phi(\sigma)$

In this subsection we shall be using definitions (1.17) of $\partial^{\pm}\phi(\sigma)$. We shall also be using some theorems on generalized Abel transforms which we proved in Ref. 1.[8]

Let us begin by recalling the definitions:

$$\gamma_3(r):=\gamma(r,1),\quad \gamma_2(s):=\gamma(-1,s),$$
$$\phi_3(r,\tau):=\phi(r,1,\tau),\quad \phi_2(s,\tau):=\phi(-1,s,\tau). \tag{2.31}$$

From its definition, $\phi_3(r,\tau)$ is that integral of Eq. (2.25) along the path $s=1$ in D_τ such that $\phi_3(-1,\tau)=0$. Likewise, $\phi_2(s,\tau)$ is that integral of Eq. (2.25) along the path $r=-1$ such that $\phi_2(1,\tau)=0$. In other words, as one can see with the aid of Eqs. (2.3) and (2.17), defining equations for ϕ_3 and ϕ_2 are as follows:

$$(\tau-r)d\phi_3(\tau)-\tfrac{1}{2}dr\,\phi_3(\tau)=-d\gamma_3,\quad \phi_3(-1,\tau)=0,$$
$$(\tau-s)d\phi_2(\tau)-\tfrac{1}{2}ds\,\phi_2(\tau)=-d\gamma_2,\quad \phi_2(1,\tau)=0. \tag{2.32}$$

In fact, explicit expressions for the integrals of Eqs. (2.32) have already been given by Eqs. (1.12). Our current interests are the alternative expressions for ϕ_3 and ϕ_2 which are given by Eqs. (1.14).

Theorem: Equations (1.14) hold.

Proof: It is sufficient to prove the first of Eqs. (1.14), since the proof of the second equation is similar. For $-1<r<r_0$ and $\tau\in C-[-1,r]$ and for real σ such that $-1<r<\sigma<r_0$, the first of Eqs. (2.32) yields

$$d\left[\sqrt{\sigma-r}\,\phi_3(r,\tau)\right]=\sqrt{\sigma-r}\,d\phi_3(r,\tau)-\frac{dr\,\phi_3(r,\tau)}{2\sqrt{\sigma-r}}$$
$$=(\sigma-\tau)\frac{d\phi_3(r,\tau)}{\sqrt{\sigma-r}}-\frac{d\gamma_3(r)}{\sqrt{\sigma-r}}.$$

Upon integrating the above over r in the interval $[-1,\sigma]$, we obtain

$$\int_{-1}^{\sigma}dr\frac{\dot\phi_3(r,\tau)}{\sqrt{\sigma-r}}=\frac{g_3(\sigma)}{\sigma-\tau},\quad \text{for } \tau\in C-[-1,\sigma], \tag{2.33}$$

where $g_3(\sigma)$ is defined by Eq. (1.11).

Now we refer to the discussion of (generalized) Abel transforms[8] in Ref. 1. The Abel transform of $\dot\gamma_3(r)$ is $g_3(\sigma)$ as given by Eq. (1.11). Similarly, the Abel transform of $\dot\phi_3(r,\tau)$ is $g_3(\sigma)/(\sigma-\tau)$ as given by Eq. (2.33). A theorem in Ref. 1 on the inversion of the Abel transform (1.11) yields

$$\gamma_3(r)=\frac{1}{\pi}\int_{-1}^{r}d\sigma\frac{g_3(\sigma)}{\sqrt{r-\sigma}}.$$

The same theorem applied to Eq. (2.33) yields the first of Eqs. (1.14). Note that the integral in Eq. (1.14) exists since according to a theorem in Ref. 1, $g_3(\sigma)/\sqrt{r-\sigma}$ is an integrable function of σ over $[-1,r]$ and since $(\sigma-\tau)^{-1}$ is bounded and continuous on $[-1,r]$.

Lemma: For any given point (r,s) in D_{IV} such that $r\neq -1$ or $s\neq 1$, $\partial^{\pm}\phi_3(\sigma)$ and $\partial^{\pm}\phi_2(\sigma)$ exist and

$$\partial^{+}\phi_3(\sigma)-\partial^{-}\phi_3(\sigma)=\begin{cases}2ig_3(\sigma)/\sqrt{r-\sigma}, & \text{if } -1<\sigma<r,\\ 0, & \text{if } s<\sigma<1\end{cases} \tag{2.34}$$

and

$$\partial^{+}\phi_2(\sigma)-\partial^{-}\phi_2(\sigma)=\begin{cases}0, & \text{if } -1<\sigma<r,\\ 2ig_2(\sigma)/\sqrt{\sigma-s}, & \text{if } s<\sigma<1.\end{cases} \tag{2.35}$$

Proof: It is sufficient to prove Eq. (2.34). In Ref. 1 we proved that $g_3(\sigma)$ obeys a Hölder condition on every closed subinterval of its domain $-1<\sigma<r_0$. Therefore, for given $r>-1$, $g_3(\sigma)/\sqrt{r-\sigma}$ obeys a Hölder condition on every closed subinterval of the open interval $-1<\sigma<r$. Therefore, from a theorem of Plemelj,[15] the first of Eqs. (1.14) implies that $\partial^{\pm}\phi_3(\sigma)$ exist and satisfy Eq. (2.34) if $-1<\sigma<r$.

If $s<\sigma<1$, i.e., $-1<r<s<\sigma<1$, then $\phi_3(r,\tau)$ is holomorphic at $\tau=\sigma$. Hence, $\partial^{+}\phi_3(r,\sigma)=\partial^{-}\phi_3(r,\sigma)$ if $s<\sigma<1$. This completes the proof of Eq. (2.34).

Theorem: For any given point (r,s) in D_{IV} such that $r\neq -1$ or $s\neq 1$, $\partial^{\pm}\phi(\sigma)=\partial^{\pm}\phi(r,s,\sigma)$ exist and Eqs. (1.18)–(1.20) hold.

Proof: The theorem follows straightforward from the preceding lemma, Eqs. (2.13), Eqs. (2.29), and the theorem that follows Eqs. (2.30). Use the first of Eqs. (2.29) to prove the theorem for values of σ in the interval $-1<\sigma<r$ and use the second of Eqs. (2.29) for $s<\sigma<1$. We leave the details of the proof to the reader.

III. THE HP FOR THE COLLINEAR CASE

A. Definition of the HP adapted to (ϕ_3,ϕ_2)

We shall now set aside the definitions of γ and ϕ which were given in Sec. II and define them instead as the results of solving a certain HP. The equivalences of the definitions in Sec. II and the definitions that we shall give below will be proven in Sec. III C.

Assume that the initial data functions $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$ have been given and satisfy the premises stated after Eqs. (1.6). Also, assume that $\phi_3(r,\tau)$ and $\phi_2(s,\tau)$ have been computed from the given initial data by using Eqs. (1.12) or any other form of the integrals of Eqs. (2.32).

Definitions: The HP[11] *adapted to* (ϕ_3,ϕ_2) is the search for a function ϕ with domain D such that the following τ-plane conditions are satisfied by $\phi(r,s,\tau)$ for any given (r,s) in D_{IV} [where we suppress (r,s) below]:

$\phi(\tau) - \chi_2(\tau)\phi_3(\tau)$ is holomorphic on $C - [s,1]$,

$\phi(\tau) - \chi_3(\tau)\phi_2(\tau)$ is holomorphic on $C - [-1,r]$,

$\phi(\infty) = 0$.

The field γ is defined in terms of the solution ϕ by

$$\gamma := [-\tau\phi(\tau)]_{\tau=\infty} \tag{3.1}$$

By using the facts that $\chi_3(\tau)$ and $\phi_3(\tau)$ are holomorphic on $C - [-1,r]$ and $\chi_2(\tau)$ and $\phi_2(\tau)$ are holomorphic on $C - [s,1]$, one can see that the first pair of conditions and in the above definition is equivalent to the triad of conditions (1.15) that were used in the informal statement of the HP in Sec. I E.

B. Solution of the HP

Theorem: A solution ϕ of the HP exists, is unique, and has the following value for any given (r,s,τ) in D:

$$\phi(r,s,\tau) = {}_3\phi(r,s,\tau) + {}_2\phi(r,s,\tau), \tag{3.2}$$

where

$$\begin{aligned} {}_3\phi(r,s,\tau) &= -\frac{1}{2\pi i}\int_{\Gamma_3} d\tau' \frac{\chi_2(s,\tau')\phi_3(r,\tau')}{\tau'-\tau}, \\ {}_2\phi(r,s,\tau) &= -\frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \frac{\chi_3(r,\tau')\phi_2(s,\tau')}{\tau'-\tau}, \end{aligned} \tag{3.3}$$

and Γ_3, Γ_2, and τ satisfy conditions (1.22) as in Fig. 1. Furthermore,

$$\gamma(r,s) = {}_3\gamma(r,s) + {}_2\gamma(r,s), \tag{3.4}$$

where

$$\begin{aligned} {}_3\gamma(r,s) &:= -\frac{1}{2\pi i}\int_{\Gamma_3} d\tau' \chi_2(s,\tau')\phi_3(r,\tau'), \\ {}_2\gamma(r,s) &:= -\frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \chi_3(r,\tau')\phi_2(s,\tau'). \end{aligned} \tag{3.5}$$

Proof: The proof will be given in two parts. (i) Assume that a solution ϕ of the HP exists. Then the first two conditions of the definition in Sec. III A of the HP imply, for any given (r,s,τ) in D,

$$\frac{1}{2\pi i}\int_{\Gamma_3} d\tau' \frac{\phi(\tau') - \chi_2(\tau')\phi_3(\tau')}{\tau'-\tau} + \frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \frac{\phi(\tau') - \chi_3(\tau')\phi_2(\tau')}{\tau'-\tau} = 0,$$

where we are suppressing (r,s). The above equation is equivalent to

$$\frac{1}{2\pi i}\int_{\Gamma} d\tau' \frac{\phi(\tau')}{\tau'-\tau} + {}_3\phi(\tau) + {}_2\phi(\tau) = 0, \tag{3.6}$$

where ${}_3\phi$ and ${}_2\phi$ are defined by Eqs. (3.3) and Γ is any positively oriented contour such that

$$([-1,r]\cup[s,1])\subset\Gamma^+, \quad \tau\in\Gamma^-.$$

The third condition of the HP implies that the integral over Γ in Eq. (3.6) equals $-\phi(\tau)$, which yields Eq. (3.2). The application of Eq. (3.1) to Eqs. (3.2) and (3.3) then yields Eqs. (3.4) and (3.5). We have thus proven that if a solution ϕ of the HP exists, it is given by Eqs. (3.2) and (3.3) and the corresponding γ is given by Eqs. (3.4) and (3.5). This, of course, implies the uniqueness of the solution.

(ii) We shall next establish the existence of a solution by showing that ϕ *as defined by* Eqs. (3.2) and (3.3) satisfies all the conditions of the HP. The arguments (r,s) will be suppressed in the proof. Since $\chi_2(\infty)\phi_3(\infty) = 0$,

$$-\chi_2(\tau)\phi_3(\tau) = \frac{1}{2\pi i}\int_{\Gamma_3+\Gamma_2} d\tau' \frac{\chi_2(\tau')\phi_3(\tau')}{\tau'-\tau}.$$

Upon adding the two sides of the above equation to the corresponding sides of Eq. (3.2) and using Eqs. (3.3), we obtain

$$\phi(\tau) - \chi_2(\tau)\phi_3(\tau) = \xi_2(\tau),$$

where

$$\xi_2(\tau) := \frac{1}{2\pi i}\int_{\Gamma_2} d\tau' \frac{\chi_2(\tau')\phi_3(\tau') - \chi_3(\tau')\phi_2(\tau')}{\tau'-\tau}. \tag{3.7}$$

By the conventional method of deforming Γ_2 to accommodate any given τ in $C - [s,1]$ so that $\tau\in\Gamma_2^-$, $\xi_2(\tau)$ can be holomorphically extended to the domain $C - [s,1]$. Therefore, ϕ satisfies the first condition of the HP.

The second condition of the HP is proven by a similar method (which need not be detailed here) and the third condition follows directly from Eqs. (3.2) and (3.3). This completes our account of the proof.

Note: In Sec. II we proved that the ϕ and γ defined there satisfy Eq. (3.1) and all the conditions of the HP. Therefore, we could have used the known existence theorem on the solution of any hyperbolic equation such as Eq. (1.5) to prove the existence of a solution of the HP. However, we opted for the existence proof given above since it illustrates the complex plane methods which constitute an important part of our formalism.

Corollary: The solution ϕ of the HP and the corresponding γ satisfy

$$\begin{aligned} &\phi(r,1,\tau) = \phi_3(r,\tau), \quad \gamma(r,1) = \gamma_3(r), \\ &\phi(-1,s,\tau) = \phi_2(s,\tau), \quad \gamma(-1,s) = \gamma_2(s). \end{aligned} \tag{3.8}$$

Proof: Set $s = 1$ in Eqs. (3.3) and (3.5) and use the relations $\chi_2(1,\tau) = 1$, $\phi_2(2,\tau) = 0$, and

$$[-\tau\phi_3(r,\tau)]_{\tau=\infty} = \gamma_3(r),$$

which derives from Eq. (1.12) and the relations $\chi_3(r,\infty) = 1$ and $\gamma_3(-1) = 0$. Then set $r = -1$ in Eqs. (3.3) and (3.5) and use the relations $\chi_3(-1,\tau) = 1$, $\phi_3(-1,\tau) = 0$, and

$$[-\tau\phi_2(s,\tau)]_{\tau=\infty} = \gamma_2(s),$$

which derives from Eq. (1.12) and the relations $\chi_2(s,\infty) = 1$ and $\gamma_2(1) = 0$. We thus obtain

$$\begin{aligned} &{}_3\phi(r,1,\tau) = \phi_3(r,\tau), \quad {}_3\gamma(r,1) = \gamma_3(r), \\ &{}_2\phi(-1,s,\tau) = \phi_2(s,\tau), \quad {}_2\gamma(-1,s) = \gamma_2(s), \\ &{}_2\phi(r,1,\tau) = {}_2\gamma(r,1) = {}_3\phi(-1,s,\tau) = {}_3\gamma(-1,s) = 0. \end{aligned} \tag{3.9}$$

Substitution of (3.9) into Eqs. (3.2) and (3.4) yields Eqs. (3.8).

Corollary: For any given τ in $C - \{-1,1\}$, the functions [of (r,s)]

$$ {}_j\phi(\tau),\quad d\,[{}_j\phi(\tau)],\quad d*d\,[{}_j\phi(\tau)] $$

exist and are continuous on D_τ. The functions

$$ {}_j\gamma,\quad d({}_j\gamma),\quad d*d({}_j\gamma) $$

exist and are continuous on D_{IV} and $d^2({}_j\gamma)$ exists and vanishes on D_{IV}.

Proof: The proof employs Eqs. (3.3) and (3.5) as well as the definitions of continuity and exterior differentiation of differential forms in the space D_{IV} given in Sec. II A.

C. Relation of the HP to the IVP

We shall next give a theorem and two corollaries which establish that the solution of the HP solves the IVP. The z and ρ that appear in the following theorem are defined by Eqs. (2.17).

Theorem: The ${}_j\phi$ and ${}_j\gamma$ as defined by Eqs. (3.3) and (3.5) satisfy

$$ {}_j\phi(-1,1,\tau)=0 \tag{3.10} $$

and, on the domain D_τ,

$$ (\tau-z+\rho*)d\,[{}_j\phi(\tau)]-dz[{}_j\phi(\tau)] = -d\,[{}_j\gamma]. \tag{3.11} $$

Proof: Equation (3.10) is implied by Eqs. (3.8). As regards Eq. (3.11), it is sufficient to supply a proof for $j=3$. From the definition of $\chi_2(s,\tau)$ in Eq. (1.13),

$$ (\tau-s)d\chi_2(s,\tau)-\tfrac{1}{2}\,ds\,\chi_2(s,\tau)=0. $$

Upon combining the above equation with Eq. (2.32), we obtain, with the aid of Eq. (2.17) and the relations $*d\phi_3(r,\tau)=d\phi_3(r,\tau)$ and $*d\chi_2(s,\tau)=-d\chi_2(s,\tau)$,

$$ (\tau-z-\rho*)d[\chi_2(\tau)\phi_3(\tau)] - dz[\chi_2(\tau)\phi_3(\tau)] = -\chi_2(\tau)d\gamma_3. $$

Application of the above to ${}_3\phi(\tau)$ as given by Eq. (3.3) and the use of the expression for ${}_3\gamma$ in Eqs. (3.5) yields

$$ (\tau-z+\rho*)d\,[{}_3\phi(\tau)]-dz[{}_3\phi(\tau)] = -\frac{1}{2\pi i}\int_{\Gamma_3} d\tau'\,\frac{(\tau-\tau')d\,[\chi_2(\tau')\phi_3(\tau')]-\chi_2(\tau')d\gamma_3}{\tau'-\tau} = -d\,[{}_3\gamma]. \qquad \text{Q.E.D.} $$

Corollary: For $j=3$ and $j=2$,

$$ d\,[\rho*d({}_j\phi)]=0,\quad d\,[\rho*d({}_j\gamma)]=0. \tag{3.12} $$

Proof: Take the exterior derivative of both sides of Eq. (3.11) to obtain the first of Eqs. (3.12). Then apply the operator $\rho*$ to both sides of Eq. (3.11) and take the exterior derivative of the result. The second of Eqs. (3.12) is then obtained after using Eqs. (2.4) and the easily proven relation $*dz=-d\rho$.

Corollary: The solution ϕ of the HP and the corresponding γ satisfy Eqs. (2.6), (2.24), (2.25), and (2.28).

Proof: Sum Eqs. (3.10)–(3.12) over j and use Eqs. (3.2) and (3.4).

Let us see where we stand. In Sec. II A we defined γ as a function whose domain is D_{IV} such that $\psi(u,v):=\gamma(r(u),s(v))$ satisfies the hyperbolic Eq. (1.5) and the initial value conditions $\psi(u,0)=\psi_3(u)$ and $\psi(0,v)=\psi_2(v)$. In Sec. II C we defined ϕ in terms of the given γ by Eqs. (2.19), (2.20), and (2.23) or, equivalently, by Eqs. (2.24) and (2.25). Finally, in Secs. II C and II D we proved that this ϕ *satisfies Eq. (3.1) and all of the defining conditions of the HP.* [See part (i) of the second theorem in Sec. II C, the theorem and its corollary in Sec. II D, and the paragraph following the definition of the HP in Sec. III A.]

Conversely, Eqs. (3.8) and the last corollary that we proved above establish that the ϕ and γ defined by the HP satisfy all the defining equations of the ϕ and γ of Sec. II. *Thus the ϕ and γ defined by the HP are identical with the ϕ and γ which were defined in Sec. II. Furthermore, from the theorem in Sec. III B, the solution of the IVP for the collinear case exists, is unique, and is $\psi(u,v)=\gamma(r(u),s(v))$, where γ is given by Eq. (1.23).*

D. The HP adapted to (g_3, g_2)

Here we shall assume that g_3 and g_2 have been computed from the initial data by using Eqs. (1.11). It is not necessary to compute ϕ_3 and ϕ_2 in order to use the HP defined below. It is only necessary to know that ϕ_3 and ϕ_2 are given in terms of g_3 and g_2 by Eqs. (1.14). Equation (1.17) and Sec. II B should be consulted for the definition of the various symbols that appear in the following definition.

Definitions: The HP adapted to (g_3,g_2) is the search for a function ϕ with domain D such that the following τ-plane conditions are satisfied by $\phi(r,s,\tau)$ for any given (r,s) in D_{IV} [where we suppress (r,s) below]:

(i) The function $\phi(\tau)$ is holomorphic on $D_{(r,s)}$.

(ii) If $r\neq -1$ or $s\neq 1$ (or both), $\partial^{\pm}\phi(\sigma)$ exist and Eqs. (1.18) hold for all σ in the open intervals $-1<\sigma<r$ and $s<\sigma<1$.

(iii) The endpoint conditions (1.25) hold.

(iv) The value of $\phi(\infty)=0$.

Finally, γ is defined in terms of ϕ by Eq. (3.1).

Theorem: There is not more than one solution of the HP adapted to (g_3,g_2).

Proof: Let $\xi:=\phi'-\phi$, where ϕ and ϕ' are any solutions of the HP. For any given (r,s) in D_{IV}, condition (ii) of the HP implies that

$$ \partial^{\pm}\xi(\sigma)\ \text{exist},\quad \partial^{+}\xi(\sigma)=\partial^{-}\xi(\sigma) \tag{3.13} $$

for all σ such that $-1<\sigma<r$ or $s<\sigma<1$. According to a known theorem, a function $\xi(\tau)$ that satisfies condition (i) of the HP and Eq. (3.13) has an extension [which we shall also denote by $\xi(\tau)$] that is holomorphic throughout $C-\{-1,r,1,s\}$, i.e., holomorphic throughout C except perhaps for poles or isolated essential singularities at -1, r, 1, or s.

Next, note that

$$ \begin{aligned} \xi(\tau) &= [\phi'(\tau)-\chi_2(\tau)\phi_3(\tau)]-[\phi(\tau)-\chi_2(\tau)\phi_3(\tau)] \\ &= [\phi'(\tau)-\chi_3(\tau)\phi_2(\tau)]-[\phi(\tau)-\chi_3(\tau)\phi_2(\tau)]. \end{aligned} $$

Therefore, condition (iii) of the HP taken together with Eq. (3.13) implies that $\xi(\tau)$ remains bounded as $\tau\to -1$, r, 1, or s through any sequence of points in $C-\{-1,r,1,s\}$. Therefore, $\xi(\tau)$ has no singularities at -1, r, 1, and s and is, according to Liouville's theorem, constant in value throughout C. Hence $\phi'(\tau)=\phi(\tau)$ throughout $D_{(r,s)}$. Q.E.D.

Lemma: Alternative expressions for the functions ${}_3\phi$ and ${}_2\phi$ that were defined by Eqs. (3.3) are

$$
\begin{aligned}
{}_3\phi(r,s,\tau) &= \frac{1}{\pi}\int_{-1}^{r} d\sigma \frac{g_3(\sigma)\chi_2(s,\sigma)}{(\sigma-\tau)\sqrt{r-\sigma}}, \\
{}_2\phi(r,s,\tau) &= \frac{1}{\pi}\int_{1}^{s} d\sigma \frac{g_2(\sigma)\chi_3(r,\sigma)}{(\sigma-\tau)\sqrt{\sigma-s}}.
\end{aligned}
\tag{3.14}
$$

Proof: We shall only supply a proof of the first of Eqs. (3.14). Substitution from the first of Eqs. (1.14) into the first of Eqs. (3.3) yields

$$
{}_3\phi(r,s,\tau) = -\frac{1}{2\pi^2 i}\int_{\Gamma_3} d\tau' \int_{-1}^{r} d\sigma \frac{g_3(\sigma)\chi_2(s,\tau')}{(\sigma-\tau')(\tau'-\tau)\sqrt{r-\sigma}}, \tag{3.15}
$$

where we recall that $[-1,r]\subset\Gamma_3^+$ and $\tau\in\Gamma_3^+$. Now, $\chi_2(s,\tau)$ is holomorphic on $C-[s,1]$ and is, therefore, holomorphic on $\Gamma_3\cup\Gamma_3^+$. Hence, we obtain the first of Eqs. (3.14) after interchanging (as is permissible here) the order of integrations in Eq. (3.15) and applying the Cauchy theorem to the integral over τ'.

Theorem: The solution of the HP adapted to (g_3, g_2) exists and is equal to the solution of the HP adapted to (ϕ_3,ϕ_2). In other words, it exists and is given by Eqs. (3.2) and (3.3) or, equivalently, by Eqs. (3.2) and (3.14).

Proof: We prove the theorem by showing that ϕ *as defined by* Eqs. (3.2) and (3.14) satisfies conditions (i)–(iv) of the HP adapted to (g_3, g_2).

We proved in Sec. III B that ϕ as defined by Eqs. (3.2) and (3.3) satisfies the conditions of the HP adapted to (ϕ_3,ϕ_2). These conditions trivially imply conditions (i), (iii), and (iv) of the HP adapted to (g_3, g_2). Therefore, it remains only to prove that ϕ as defined by Eqs. (3.2) and (3.3) also satisfies condition (ii) of the HP adapted to (g_3, g_2).

In Ref. 1 we proved that $g_j(\sigma)$ obeys a Hölder condition on any given closed subinterval of its open interval domain. Now, for fixed (r,s) in $D_{\rm IV}$ such that $r\neq-1$ and $s\neq1$,

$$
\chi_2(s,\sigma)/\sqrt{r-\sigma}, \quad \chi_3(r,\sigma)/\sqrt{\sigma-s}
$$

are $\mathbf{C}^1$ functions of σ over any given closed subintervals of the open intervals $-1<\sigma<r$ and $s<\sigma<1$, respectively. Therefore, the products

$$
g_3\chi_2(s,\sigma)/\sqrt{r-\sigma}, \quad g_2\chi_3(r,\sigma)/\sqrt{\sigma-s}
$$

obey Hölder conditions on any given closed subintervals of the intervals $-1<\sigma<r$ and $s<\sigma<1$, respectively, whereupon a theorem of Plemelj declares that $\phi(\tau)$ as defined by Eqs. (3.2) and (3.14) has existing boundary values $\partial^{\pm}\phi(\sigma)$ which satisfy[15]

$$
\partial^+\phi(\sigma)-\partial^-\phi(\sigma) = \begin{cases} 2ig_3(\sigma)\chi_2(s,\sigma)/\sqrt{r-\sigma}, & \text{if } -1<\sigma<r, \\ 2ig_2(\sigma)\chi_3(r,\sigma)/\sqrt{\sigma-s}, & \text{if } s<\sigma<1. \end{cases}
$$

The above equation is equivalent to Eq. (1.18), as can be seen from Eqs. (1.19) and (1.20). Therefore, condition (ii) of the HP adapted to (g_3, g_2) is satisfied.

Corollary: Here ϕ and γ are also given by Eqs. (1.26) and (1.27).

Proof: Use Eqs. (1.19), (1.20), (3.2), and (3.14).

Note: The solution of the HP adapted to (g_3, g_2) could have been derived without using the information supplied by Sec. III B. Specifically, the same theorem of Plemelj[15] that we used to prove the preceding theorem directly tells us that Eq. (1.26) satisfies conditions (i), (ii), and (iv) of the HP adapted to (g_3, g_2). It remains only to prove that condition (iii) [i.e., the endpoint conditions (1.25)] is also satisfied by Eq. (1.26): This can be done by a direct method that makes no use of any results that we proved in Sec. III B. We leave the construction of this direct method to the interested reader.

IV. PROPERTIES OF THE INITIAL DATA FUNCTIONS AND OF g_3 AND g_2

A. Constraints on the initial data due to the vacuum condition

The only premises that have been granted so far concerning the initial data are those given after Eqs. (1.6) in Sec. I B. As we stated at the end of Sec. I B these premises do not include all constraints imposed on the initial data by the requirement that the Ricci tensor exist and vanish throughout the space-time.

The topic of constraints on the initial data was covered in Ref. 1 for the noncollinear as well as the collinear cases. We shall here summarize those conclusions in Ref. 1 that pertain to the collinear case. We shall assume that the metrical functions ρ, ψ, and Γ that occur in the line element (1.1) satisfy the following conditions throughout their common domain IV:

$$
\begin{aligned}
&\rho \text{ is } \mathbf{C}^2,\ \psi,\ \Gamma \text{ are } \mathbf{C}^1, \\
&\psi_{uv},\ \Gamma_{uv} \text{ exist and are continuous.}
\end{aligned}
\tag{4.1}
$$

When discussing the vacuum field equations in Ref. 1, we also assumed that ψ is $\mathbf{C}^2$. However, a reexamination of our work in Ref. 1 shows that the set of conditions (4.1) is sufficient for the existence of the Ricci tensor throughout the domain of the chart employed in Ref. 1 and is sufficient to deduce the conclusions given below.

Recall that the ignorable coordinates x^1, x^2 are scaled to make

$$
\rho(0,0)=1, \quad \psi(0,0)=0 \tag{4.2}
$$

and that the initial data functions are related to ρ and ψ as follows:

$$
\begin{aligned}
r(u) &= 1-2\rho(u,0), \quad \psi_3(u)=\psi(u,0), \\
s(v) &= 2\rho(0,v)-1, \quad \psi_2(v)=\psi(0,v).
\end{aligned}
\tag{4.3}
$$

The premises (4.1) and Eqs. (4.2) and (4.3) imply[16]

$$
\psi_3(u),\ \psi_2(v) \text{ are } \mathbf{C}^1, \quad \psi_3(0)=\psi_2(0)=0, \tag{4.4}
$$

and

$$
r(u),\ s(v) \text{ are } \mathbf{C}^2, \quad r(0)=-1, \quad s(0)=1. \tag{4.5}
$$

Furthermore, we proved in Ref. 1 that the vacuum field equations over that space-time region occupied by the two plane waves and the two null hypersurfaces that are the fronts of these waves prior to their collision imply

$$\dot r(0)=\dot s(0)=0,\quad \dot r(u)>0 \text{ if } 0<u<u_0,$$
$$\dot s(v)<0 \text{ if } 0<v<v_0, \tag{4.6}$$

and

$$\lim_{u\to 0}\{\ddot r(u)-2[1-r(u)][\dot\psi_3(u)]^2\}/\dot r(u) \text{ exists},$$
$$\lim_{v\to 0}\{\ddot s(v)+2[1+s(v)][\dot\psi_2(v)]^2\}/\dot s(v) \text{ exists}. \tag{4.7}$$

Conversely, if one assumes that the initial data functions are given and satisfy the constraining conditions (4.4)–(4.7), then the vacuum field equations over the space-time region that is occupied by the scattered wave have unique solutions ρ, ψ, and Γ which satisfy conditions (4.1)–(4.3). The solutions ρ and ψ are given by Eqs. (1.3) and (1.27), respectively, and the solution Γ is given in Ref. 1.[6]

Observe that conditions (4.4)–(4.7) imply that

$$\ddot r(0)=4[\dot\psi_3(0)]^2,\quad \ddot s(0)=-4[\dot\psi_2(0)]^2, \tag{4.8}$$

which imply in turn that $\ddot r(0)\geqslant 0$ and $\ddot s(0)\leqslant 0$. We shall refer to Eqs. (4.8) later.

Conditions (4.4)–(4.7) are formulated in terms of a particular choice of null coordinates u and v. However, if u and v are subjected to any $\mathbf{C}^2$ transformations

$$u\to u' \text{ with domain } 0\leqslant u<u_0 \text{ and range } 0\leqslant u'<u'_0,$$
$$v\to v' \text{ with domain } 0\leqslant v<v_0 \text{ and range } 0\leqslant v'<v'_0$$

such that du'/du and dv'/dv are positive, then conditions (4.4)–(4.7) remain true after replacing u, v, u_0, v_0, $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$ by u', v', u'_0, v'_0, $r'(u')$, $s'(v')$, $\psi'_3(u')$, and $\psi'_2(v')$, respectively, where (of course)

$$r'(u')=r(u),\quad s'(v')=s(v),$$
$$\psi'_3(u')=\psi_3(u),\quad \psi'_2(v')=\psi_2(v).$$

Moreover, the numbers r_0 and s_0 that were defined in Eqs. (1.4) and the functions γ_3 and γ_2 that were defined by Eqs. (1.9) are unchanged by the transformation $(u,v)\to(u',v')$. The above statements are easy to prove, as is the statement that the group generated by all such transformations is the group of all null coordinate transformations under which conditions (4.5) and (4.6) persist and r_0 and s_0 are unchanged. Application of any element of this group does not change γ_3 and γ_2 and leaves conditions (4.4)–(4.7) true if they are initially true.

The following theorem supplies another group of transformations under which conditions (4.4)–(4.7) persist.

Theorem: Let α_3 and α_2 be any real-valued functions with domains $-1\leqslant r<r_0$ and $s_0<s\leqslant 1$, respectively, such that α_3 and α_2 are $\mathbf{C}^1$ and $\alpha_3(-1)=\alpha_2(1)=0$. Then if statements (4.4)–(4.7) are true, they remain true after the substitutions

$$\psi_3(u)\to\psi_3(u)+\alpha_3(r(u)),\quad r(u)\to r(u),$$
$$\psi_2(v)\to\psi_2(v)+\alpha_2(s(v)),\quad s(v)\to s(v).$$

[Note: If $\ddot r(0)=0$, then one can replace the premise that α_3 is $\mathbf{C}^1$ by the lighter premise that $\dot\alpha_3(r)$ exists, is continuous over $-1<r<r_0$, and is bounded in at least one interval $-1<r<-1+\epsilon$ such that $\epsilon>0$. A like statement holds for α_2 if $\ddot s(0)=0$.]

Proof: Use the relations

$$\frac{d}{du}[\alpha_3(r(u))]=\dot r(u)\dot\alpha_3(r(u)),$$
$$\frac{d}{dv}[\alpha_2(s(v))]=\dot s(v)\dot\alpha_2(s(v)),$$

whereupon the proof is straightforward.

B. Special $r(u)$ and $s(v)$

In this subsection we shall focus attention on any choices of $r(u)$ and $s(v)$ that satisfy the following definition.

Definition: The functions $r(u)$ and $s(v)$ will be called *special* if they satisfy conditions (4.5) and (4.6), the conditions

$$\lim_{u\to 0}\left[\frac{1+r(u)}{\dot r(u)}\right]=\lim_{v\to 0}\left[\frac{1-s(v)}{-\dot s(v)}\right]=0, \tag{4.9}$$

and the conditions that

$$m_3(0):=\lim_{u\to 0}m_3(u),\quad m_2(0):=\lim_{v\to 0}m_2(v) \tag{4.10}$$

exist and have the values

$$m_3(0)\neq 0,\quad m_2(0)\neq 0, \tag{4.11}$$

where

$$m_3(u):=\ddot r(u)[1+r(u)]/\dot r(u)^2,$$
$$m_2(v):=-\ddot s(v)[1-s(v)]/\dot s(v)^2 \tag{4.12}$$

for $0<u<u_0$ and $0<v<v_0$, respectively.

Examples: The reader will discover by trial that any choices of $r(u)$ and $s(v)$ that satisfy conditions (4.5) and (4.6) and are deemed palatable for constructing specific colliding gravitational plane-wave solutions are *special* in the sense defined above. Examples of special $r(u)$ families are given by

$$r(u)=2u^n-1,\quad \text{where } n\geqslant 2,\ u_0:=1,$$
$$r(u)=\left[2\exp\left(-\frac{1}{u^n}\right)\right]-1,\quad \text{where } n>0, u_0:=\infty.$$

Similar examples that are $\mathbf{C}^2$, but are not $\mathbf{C}^\infty$, are easily constructed. A family of $r(u)$ choices that satisfy conditions (4.5) and (4.6), but are not special is given by

$$r(u)=2(n+1)(n+2)\int_0^u db\int_0^b da\, a^n\left(\sin\frac{1}{a}\right)^2-1,$$

where $n>0$ and $u_0:=1$.

Theorem: If $r(u)$ and $s(v)$ are special, then

$$0<m_j(0)\leqslant 1\ (j=3,2). \tag{4.13}$$

Furthermore, there exist real ϵ_3 and ϵ_2 such that $0<\epsilon_3<u_0$, $0<\epsilon_2<v_0$, and

$$\ddot r(u)>0,\quad \text{if } 0<u<\epsilon_3,\quad -\ddot s(v)>0,\quad \text{if } 0<v<\epsilon_2. \tag{4.14}$$

Proof: From Eqs. (4.9), (4.10), and (4.12), $m_3(u)$ and $m_2(v)$ are $\mathbf{C}^1$ over $0\leqslant u<u_0$ and $0\leqslant v<v_0$, respectively,

$$\frac{d}{du}\left[\frac{1+r(u)}{\dot{r}(u)}\right] = 1 - m_3(u),$$

$$\frac{d}{dv}\left[\frac{1-s(v)}{-\dot{s}(v)}\right] = 1 - m_2(v),$$

and the values of the above derivatives at $u = 0$ and $v = 0$ are given by

$$1 - m_3(0) = \lim_{u\to 0}\left[\frac{1+r(u)}{u\dot{r}(u)}\right],$$
$$1 - m_2(0) = \lim_{v\to 0}\left[\frac{1-s(v)}{-v\dot{s}(v)}\right]. \tag{4.15}$$

Conditions (4.5) and (4.6) imply that the expressions in brackets in Eqs. (4.15) are positive. Therefore, $1 \geqslant m_j(0)$.

Now, suppose $m_3(0) < 0$. Then there would exist real ϵ such that $0 < \epsilon < u_0$ and $m_3(u) < 0$ when $0 \leqslant u < \epsilon$. Definition (4.12) of $m_3(u)$ would then imply that $\ddot{r}(u) < 0$ when $0 < u < \epsilon$: This would contradict the conditions $\dot{r}(0) = 0$ and $\dot{r}(u) > 0$ when $u > 0$. Therefore, $m_3(0) \geqslant 0$. A similar proof yields $m_2(0) \geqslant 0$.

We have proven above that $0 \leqslant m_j(0) \leqslant 1$. Conclusions (4.13) of the theorem now follow from (4.11). The proof of conclusions (4.14) of the theorem is straightforward and left for the reader, as is the proof of the following corollary, which should be compared with Eqs. (4.8).

Corollary: If $r(u)$ and $s(v)$ are special, then $\ddot{r}(0) \geqslant 0$ and $-\ddot{s}(0) \leqslant 0$.

A stronger result is given by the following corollary.

Corollary: If $r(u)$ and $s(v)$ are special, then there exists at least one pair of real-valued functions $\psi_3(u)$ and $\psi_2(v)$ with domains $0 \leqslant u < u_0$ and $0 \leqslant v < v_0$, respectively, such that $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$ obey conditions (4.4)–(4.7).

Proof: Let ϵ_3 and ϵ_2 be defined by conclusions (4.14) of the preceding theorem. Let

$$\psi_3(u) := \begin{cases} \int_0^u da\,\sqrt{2\ddot{r}(a)/[1-r(a)]}, & \text{for } 0 \leqslant u \leqslant \epsilon_3, \\ (u - \epsilon_3 + 1)\psi_3(\epsilon_3), & \text{for } \epsilon_3 < u < u_0, \end{cases}$$

and similarly construct $\psi_2(v)$. The rest of the proof is straightforward.

Definitions: Let

$$p := \sqrt{(1+r)/2}, \quad q := \sqrt{(1-s)/2},$$
$$p_0 := \sqrt{(1+r_0)/2}, \quad q_0 := \sqrt{(1-s_0)/2}, \tag{4.16}$$
$$p(u) := \sqrt{[1+r(u)]/2}, \quad q(u) := \sqrt{[1-s(u)]/2}.$$

For any given ψ_3 and ψ_2, let β_3 and β_2 denote those functions whose domains are $0 \leqslant p < p_0$ and $0 \leqslant q < q_0$, respectively, and whose values are given by

$$\beta_3(p) := \gamma_3(r), \quad \beta_2(q) := \gamma_2(s). \tag{4.17}$$

Equivalently,

$$\beta_3(p(u)) := \psi_3(u), \quad \beta_2(q(v)) := \psi_2(v). \tag{4.18}$$

Observe that Eqs. (4.16) and (2.17) yield

$$\rho = 1 - p^2 - q^2, \quad z = p^2 - q^2. \tag{4.19}$$

Note that p and q are identical with u and v when $r(u) = 2u^2 - 1$ and $s(v) = 1 - 2v^2$, but this is not generally true here.

From definitions (4.17), $\beta_3(p)$ and $\beta_2(q)$ are continuous over $0 \leqslant p < p_0$ and $0 \leqslant q < q_0$, respectively. Moreover, $\dot{\beta}_3(p) := d\beta_3(p)/dp$ and $\dot{\beta}_2(q) := d\beta_2(q)/dq$ exist and are continuous over $0 < p < p_0$ and $0 < q < q_0$, respectively. However, for general $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$, the derivatives $\dot{\beta}_3(p)$ and $\dot{\beta}_2(q)$ may not be "well behaved" as $p \to 0$ and $q \to 0$.

Theorem: Suppose $r(u)$ and $s(v)$ are special and suppose $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$ satisfy conditions (4.4)–(4.7). Then $[\dot{\beta}_3(p)]^2$ and $[\dot{\beta}_2(q)]^2$ are continuous over $0 \leqslant p < p_0$ and $0 \leqslant q < q_0$, respectively, and

$$[\dot{\beta}_j(0)]^2 = 2m_j(0). \tag{4.20}$$

Equivalently, the set of limit points of $\dot{\beta}_j$ as its argument approaches the origin is

$$\{\sqrt{2m_j(0)}, -\sqrt{2m_j(0)}\}, \{\sqrt{2m_j(0)}\} \text{ or } \{-\sqrt{2m_j(0)}\}.$$

Proof: Consider, for example, $j = 3$. Observe that Eqs. (4.16) and (4.18) imply that

$$[1 + r(u)][\dot{\psi}_3(u)]^2/\dot{r}(u)^2 = \tfrac{1}{8}[\dot{\beta}_3(p(u))]^2.$$

Therefore, by multiplying the expression in curly brackets in the first of conditions (4.7) by $[1 + r(u)]/\dot{r}(u)^2$ and by using Eqs. (4.9) and (4.12), we obtain

$$m_3(u) - \tfrac{1}{4}[1 - r(u)][\dot{\beta}_3(p(u))]^2$$

and deduce that the above expression is continuous over $0 \leqslant u < u_0$ and vanishes at $u = 0$. The theorem for $j = 3$ then follows from the relation $r(0) = -1$. A similar proof is used for the case $j = 2$.

C. On g_3 and g_2

We next cover some properties of the functions g_3 and g_2 which were defined by Eqs. (1.11) and have played important roles both in this paper and in Ref. 1. The first theorem is general. It assumes nothing more than the premises on the initial data which are stated following Eqs. (1.6).

Theorem: The function $g_j(\sigma)$ obeys a Hölder condition[17] of index $\frac{1}{2}$ on any given closed subinterval of its (open interval) domain.[18]

Proof: We need to give a proof only for $j = 3$ since the proof for $j = 2$ is similar. Let $[a,b]$ be any closed subinterval of the domain $-1 < \sigma < r_0$ of $g_3(\sigma)$. Let

$$c := a - \tfrac{1}{2}(a + 1)$$

and let σ and σ' be any points on $[a,b]$ such that $\sigma' > \sigma$. Thus

$$-1 < c < a \leqslant \sigma < \sigma' \leqslant b < r_0. \tag{4.21}$$

From Eq. (1.11),

$$g_3(\sigma') - g_3(\sigma) = I_1(\sigma,\sigma') - I_2(\sigma,\sigma') - I_3(\sigma,\sigma'), \tag{4.22}$$

where

$$I_1(\sigma,\sigma') := \int_\sigma^{\sigma'} dr\,\frac{\dot{\gamma}_3(r)}{\sqrt{\sigma' - r}}, \tag{4.23}$$

$$I_2(\sigma,\sigma') := \int_c^{\sigma} dr\,\dot{\gamma}_3(r)\left[\frac{1}{\sqrt{\sigma - r}} - \frac{1}{\sqrt{\sigma' - r}}\right], \tag{4.24}$$

$$I_3(\sigma,\sigma') := \int_{-1}^{c} dr\,\dot{\gamma}_3(r)\left[\frac{1}{\sqrt{\sigma - r}} - \frac{1}{\sqrt{\sigma' - r}}\right]. \tag{4.25}$$

Since $\dot{\gamma}_3$ is continuous on the closed interval $[c,b]$,

$$M_1(a,b) := \max\{|\dot\gamma_3(r)|:\ c\leqslant r\leqslant b\}$$

exists and is finite and positive. From Eq. (4.23),

$$|I_1(\sigma,\sigma')|\leqslant 2M_1(a,b)\sqrt{\sigma'-\sigma}. \tag{4.26}$$

From Eqs. (4.21) and (4.24),

$$|I_2(\sigma,\sigma')|\leqslant 2M_1(a,b)\ [\sqrt{\sigma-c}+\sqrt{\sigma'-\sigma}-\sqrt{\sigma'-c}\ ]$$
$$\leqslant 2M_1(a,b)\sqrt{\sigma'-\sigma}. \tag{4.27}$$

For all non-negative real numbers x and y such that $x>y$,

$$\sqrt{x}-\sqrt{y}\leqslant\sqrt{x-y}.$$

Therefore, from Eq. (4.25),

$$|I_3(\sigma,\sigma')|\leqslant\sqrt{\sigma'-\sigma}\int_{-1}^{c} dr\frac{|\dot\gamma_3(r)|}{\sqrt{(\sigma-r)(\sigma'-r)}}$$
$$\leqslant\frac{2M_2(a)}{a+1}\sqrt{\sigma'-\sigma}, \tag{4.28}$$

where

$$M_2(a) := \int_{-1}^{c} dr|\dot\gamma_3(r)|$$

and where we have used the inequalities

$$\sigma-r\geqslant\tfrac{1}{2}(a+1)\sigma'-r>\tfrac{1}{2}(a+1).$$

[The integral $M_2(a)$ exists and is finite. Integrals of this type were discussed in Ref. 1.] From Eqs. (4.22) and (4.26)–(4.28), there exists a positive real number $M(a,b)$ such that

$$|g_3(\sigma')-g_3(\sigma)|\leqslant M(a,b)\sqrt{\sigma'-\sigma}. \qquad \text{Q.E.D.}$$

The premise in the following corollary contradicts conditions (4.6) and (4.7) since these vacuum conditions imply, as we proved in Ref. 1, that $\dot\gamma_3(r)$ and $\dot\gamma_2(s)$ are unbounded as $r\to-1$ and $s\to 1$, respectively. However, the theorem in Sec. IV A shows that functions that satisfy the premise of the corollary have at least one use. (In the theorem of Sec. IV A, these functions were denoted by α_j.)

Corollary: Suppose $\gamma_3(r)$ and $\gamma_2(s)$ are $\mathbf{C}^1$ over $-1\leqslant r<r_0$ and $s_0<s\leqslant 1$, respectively. Then $g_3(\sigma)$ and $g_2(\sigma)$ have continuous extensions onto $-1\leqslant\sigma<r_0$ and $s_0<\sigma\leqslant 1$, respectively, and these extensions obey Hölder conditions of index $\frac{1}{2}$ on any given closed subinterval of their respective domains. [If $\gamma_3(r)$ and $\gamma_2(s)$ are $\mathbf{C}^1$ over $-1\leqslant r\leqslant r_0$ and $s_0\leqslant s\leqslant 1$, respectively, then $g_3(\sigma)$ and $g_2(\sigma)$ have continuous extensions onto $-1\leqslant\sigma\leqslant r_0$ and $s_0\leqslant\sigma\leqslant 1$, respectively, and these extensions obey Hölder conditions of index $\frac{1}{2}$ on their respective domains.]

The proof of the above corollary is a modified form of the proof of the preceding theorem and is left for the reader.

Finally, we consider alternative forms of the integrals (1.11) for $g_3(\sigma)$ and $g_2(\sigma)$. Upon introducing the new variables of integration

$$x=\sqrt{(\sigma-r)/(1+\sigma)},\quad y=\sqrt{(s-\sigma)/(1-\sigma)},$$

one obtains

$$g_3(\sigma)=2\sqrt{1+\sigma}\int_0^1 dx\,\dot\gamma_3(\sigma-(1+\sigma)x^2),$$
$$g_2(\sigma)=2\sqrt{1-\sigma}\int_0^1 dy\,\dot\gamma_2(\sigma+(1-\sigma)y^2). \tag{4.29}$$

Let us next employ the variables p and q defined by Eqs. (4.16) and the functions $\beta_3(p)$ and $\beta_2(q)$ defined by Eqs. (4.17). Equations (1.11) then become

$$g_3(\sigma)=\frac{1}{\sqrt{2}}\int_0^{\pi/2} d\theta\,\dot\beta_3\left[\left(\frac{1+\sigma}{2}\right)^{1/2}\sin\theta\right],$$
$$g_2(\sigma)=-\frac{1}{\sqrt{2}}\int_0^{\pi/2} d\theta\,\dot\beta_2\left\{\left(\frac{1-\sigma}{2}\right)^{1/2}\sin\theta\right\}, \tag{4.30}$$

after introducing a new variable of integration θ. *Equations (4.30) are especially useful for computing g_3 and g_2 when $r(u)$ and $s(v)$ are special, as defined in Sec. IV B, and $r(u)$, $s(v)$, $\psi_3(u)$, and $\psi_2(v)$ satisfy conditions (4.4)–(4.7).* The reason for this can be seen from Eq. (4.20) and the theorem that contains this equation. Specifically, if the premises of the theorem hold, then the integrals in Eqs. (4.30) are proper Riemann integrals and

$$g_3(-1)=(\pi/2\sqrt{2})\dot\beta_3(0)$$
$$:=\text{limit point of } g_3(\sigma) \text{ as } \sigma\to-1,$$
$$g_2(1)=-(\pi/2\sqrt{2})\dot\beta_2(0)$$
$$:=\text{limit point of } g_2(\sigma) \text{ as } \sigma\to 1, \tag{4.31}$$

where $[\dot\beta_j(0)]^2$ is given by Eqs. (4.20) and (4.12).

V. PERSPECTIVES

A. The HHP adapted to ($\mathscr{P}_3$, $\mathscr{P}_2$)

If one no longer assumes that the polarizations of the plane waves are collinear, then the line element (1.1) in the scattered wave region is replaced by[4]

$$ds^2=\rho S_{ab}\,dx^a\,dx^b-2\rho^{-1/2}e^{2\Gamma}\,du\,dv, \tag{5.1}$$

where $a, b=1,2$; the symmetric matrix S is positive definite; $\det S=1$; and ρ, S, and Γ depend only on u and v. We select the ignorable coordinates x^1, x^2 so that $\rho(0,0)=1$ and $S(0,0)=I$. As in the collinear case, the solution of the vacuum field equation for $\rho(u,v)$ is given by Eq. (1.3) such that Eqs. (4.5) and (4.6) hold for the functions $r(u)$ and $s(v)$. Also, the solution for $\Gamma(u,v)$ is expressed in terms of definite integrals with known integrands *once $S(u,v)$ is known.*[6] The key problem is to find the solution S of the 2×2 matrix nonlinear hyperbolic equation[19]

$$d(\rho S\sigma_2{*}dS)=0,$$
$$(\sigma_1,\sigma_2,\sigma_3 := \text{the usual representation of the Pauli spin matrices}) \tag{5.2}$$

corresponding to the prescribed initial data

$$r(u),\ s(v),\ S(u,0),\ S(0,v). \tag{5.3}$$

In the collinear case, x^1 and x^2 can further be chosen so that $S(u,v)$ is diagonal at all (u,v) in IV, whereupon one defines ψ by

$$S=\exp(-2\sigma_3\psi).$$

Equation (5.2) then reduces to the linear equation (2.5) for the function ψ.

The basic ingredient of our HHP for the general case is a 2×2 matrix spectral potential $\mathscr{P}(r,s,\tau)$ which has the same domain D as the spectral potential $\phi(r,s,\tau)$ defined in Sec. II C. In fact, for the collinear case, $\mathscr{P}(r,s,\tau)$ is given by[20]

$$\mathscr{P}_{\text{coll}} = e^{-\sigma_3\gamma}\mathscr{P}_0^K e^{(\tau\sigma_3 + i\sigma_1)\Psi}, \tag{5.4}$$

where

$$\mathscr{P}_0^K = \tfrac{1}{2}(I - \sigma_2)\chi_3 + \tfrac{1}{2}(I + \sigma_2)\chi_2 \tag{5.5}$$

and where we recall that $\phi := -\chi\Psi$, as discussed in Sec. II C. The definition of $\mathscr{P}$ for the general case and the derivations of Eq. (5.4) and of various properties of $\mathscr{P}$ will be given in a future paper of this series. Here we shall merely summarize those few properties of $\mathscr{P}$ that we need to describe the HHP.

We start by noting that the initial values

$$\mathscr{P}_3(r,\tau) := \mathscr{P}(r,1,\tau), \quad \mathscr{P}_2(s,\tau) := \mathscr{P}(-1,s,\tau) \tag{5.6}$$

are computed from the initial data functions (5.3) by integrating two separate 2×2 matrix ordinary differential equations

$$d\mathscr{P}_3 + \Gamma_3\sigma_2\mathscr{P}_3 = 0, \quad d\mathscr{P}_2 + \Gamma_2\sigma_2\mathscr{P}_2 = 0 \tag{5.7}$$

which are linear, homogeneous, and of the first order. The matrix coefficients Γ_j in Eqs. (5.7) are constructed in a simple way from the initial data functions and their differentials. Here $\mathscr{P}_j$ has exactly the same domain and holomorphy properties in the τ plane as the function ϕ_j, which is given by Eqs. (1.12) and is discussed in Sec. II D. For example, in the τ plane,

$$\begin{aligned}&\mathscr{P}_3(r,\tau) \text{ is holomorphic on } C - [-1,r],\\ &\mathscr{P}_2(s,\tau) \text{ is holomorphic on } C - [s,1].\end{aligned} \tag{5.8}$$

Also,

$$\mathscr{P}_3(-1,\tau) = \mathscr{P}_2(1,\tau) = I \tag{5.9}$$

throughout C.

As regards $\mathscr{P}(r,s,\tau) = \mathscr{P}(\tau)$, we have the following properties in the τ plane for any given (r,s) in D_{IV}:

$$\begin{aligned}&\mathscr{P}(\tau)[\mathscr{P}_3(\tau)]^{-1} \text{ is holomorphic on } C - [s,1],\\ &\mathscr{P}(\tau)[\mathscr{P}_2(\tau)]^{-1} \text{ is holomorphic on } C - [-1,r].\end{aligned} \tag{5.10}$$

Moreover,

$$\mathscr{P}(\infty) = I \tag{5.11}$$

and

$$\begin{aligned}&I - \rho(u,v)S(u,v)\\ &\quad = \text{Re}\{2\tau[I - \mathscr{P}(r(u),s(v),\tau)]\sigma_2\}_{\tau=\infty},\end{aligned} \tag{5.12}$$

Definition: The HHP *adapted to* $(\mathscr{P}_3, \mathscr{P}_2)$ is the search for a 2×2 matrix function $\mathscr{P}$ with domain D such that for any given (r, s) in D_{IV}, conditions (5.10) and (5.11) hold. Once $\mathscr{P}$ is found, then S is given by Eq. (5.12).

Our next paper will cover the above HHP in detail. There is a close parallel here with the HP adapted to (ϕ_3, ϕ_2) defined in Sec. III A. In fact, as we shall demonstrate in the next paper, the HP adapted to (ϕ_3, ϕ_2) is equivalent to the HHP adapted to $(\mathscr{P}_3,\mathscr{P}_2)$ when the polarizations of the plane waves are collinear. There is another form of the HHP that similarly generalizes the HP adapted to (g_3,g_2) as discused in Sec. III D, but we shall reserve that topic for a later paper.

We stress that there is no closed form of the solution of the HHP for arbitrary $(\mathscr{P}_3, \mathscr{P}_2)$. However, we have constructed a linear integral equation of the Cauchy type which will be given in our next paper and which is equivalent to the HHP adapted to $(\mathscr{P}_3, \mathscr{P}_2)$. We are hopeful of eventually reducing the IVP in the noncollinear case to that of solving a linear Fredholm equation of the second kind. Integral equations of this type should be effective for discovering properties of colliding gravitational plane waves.

The HHP adapted to $(\mathscr{P}_3,\mathscr{P}_2)$ is also a useful starting point in the study of methods of generating new colliding gravitational plane-wave solutions from already known solutions by employing a group of transformations similar to one that the present authors[21] used to construct stationary axisymmetric gravitational fields and that Ernst *et al.*[22] used to construct colliding gravitational plane waves. We shall be developing this theme in subsequent papers.

B. On solutions that violate the colliding wave conditions

In Secs. I–III, we did not employ all the constraints on the initial data which are imposed by the vacuum field equations. Specifically, we ignored the conditions $\dot{r}(0) = \dot{s}(0) = 0$ and the existence statements (4.7). In Ref. 1 we derived a generalization of (4.7) which is applicable to the noncollinear as well as the collinear case.[23] The two existence statements in this generalization of (4.7), taken together with the conditions that $r(u)$ and $s(v)$ are $\mathbf{C}^2$ and satisfy $\dot{r}(0) = \dot{s}(0) = 0$, will be called (as in Ref. 1) *the colliding wave conditions.*

In the next paper of this series we shall continue our policy of not excluding initial data functions which fail to satisfy all or some of the colliding wave conditions. There are two good reasons for this policy. One is that spectral potentials corresponding to initial data functions that violate colliding wave conditions often enter directly into the construction of families of spectral potentials which are consistent with the colliding wave conditions. For example, consider the spectral potential (5.4) for the general collinear case. The matrix function $\mathscr{P}_0^K$, which is given by Eq. (5.5) and appears as a function in Eq. (5.4), is the spectral potential corresponding to the Kasner metric of index 0. For this metric, $S = I$, i.e., $\psi = 0$ throughout region IV, which implies $\psi_3 = 0$ and $\psi_2 = 0$. The reader can easily verify that the existence conditions (4.7) are not satisfied if $\psi_3 = 0$, $\psi_2 = 0$, and $\dot{r}(0) = \dot{s}(0) = 0$.

The second reason for covering initial data functions that do not satisfy all or some of the colliding wave conditions is that we shall have occasion to consider families of initial data functions, say

$$r(u),\ s(v),\ S^{(n)}(u,0),\ S^{(n)}(0,v),$$

where n is a real parameter set; where the colliding wave conditions are satisfied only for a certain value n_0 of the parameter set; where the corresponding spectral potential $\mathscr{P}^{(n)}$ is a uniformly continuous function of n (with respect to any compact subset of D); and where it is straightforward to determine $\mathscr{P}^{(n)}$ when $n \neq n_0$, but difficult to do so when $n = n_0$. One can then find $\mathscr{P}^{(n_0)}$ by first computing $\mathscr{P}^{(n)}$

for $n \neq n_0$ and then letting $n \to n_0$. We shall have the opportunity to witness this theme in future papers.

ACKNOWLEDGMENT

This research was supported in part by Grant No. PHY-8605958 from the National Science Foundation.

[1]The first paper of this series is I. Hauser and F. J. Ernst, J. Math. Phys. **30**, 872 (1989).

[2]F. J. Ernst, Phys. Rev. **167**, 1176 (1968); **168**, 1415 (1968); J. Math. Phys. **15**, 1409 (1974). Also, see Sec. II F in Ref. 1.

[3]P. Szekeres, J. Math. Phys. **13**, 286 (1972). The corresponding solution when a Maxwell field is included was obtained by B. C. Xanthopoulos, J. Math. Phys. **27**, 2129 (1986).

[4]For a full description of the chart and line element, see Secs. II A and II B of Ref. 1.

[5]See Secs. II E and II H of Ref. 1.

[6]Equations (2.23), (2.31), (2.40), (2.41), and (2.53) of Ref. 1.

[7]This is discussed in Sec. II M of Ref. 1.

[8]Section III of Ref. 1.

[9]A function of τ is said to be *holomorphic at* $\tau_0 \in C$ if the function is holomorphic on at least one open neighborhood of τ_0. The function is said to be *holomorphic on* $S \subset C$ if it is holomorphic at every point of S.

[10]Some pertinent properties of $\partial^{\pm} \phi$ are derived by N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, Holland, 1953), Chap. 2, Sec. 14. This reference uses the conventional notations $\phi^{\pm}$ in place of our $\partial^{\pm} \phi$.

[11]In Sec. III we shall employ the more explicit term "HP adapted to (ϕ_3,ϕ_2)." This will help avoid confusion between our HP and certain problems that differ from ours (although there is a family connection) and are called HP's in Ref. 10. Incidentally, Secs. 34 and 39 in Chap. 5 of Ref. 10 contain enlightening remarks on the uses of the terms "Riemann problem," " Hilbert problem," and "Riemann–Hilbert problem."

[12]Sections IV D–IV G of Ref. 1.

[13]Section II G and Eqs. (2.23), (2.39), and (2.53) of Ref. 1.

[14]For each p-form μ in $D_{\rm IV}$, the definitions of $d\mu$ and the continuity of $d\mu$ are given in Sec. II A.

[15]See, e.g., Secs. 17 and 18 in Chap. 2 of Ref. 10.

[16]Sections II G–II I of Ref. 1.

[17]This is called a "Lipschitz condition" by some authors. A detailed discussion of the Hölder condition is given in Chap. 1 of Ref. 10.

[18]In Sec. III of Ref. 1, we proved the weaker statement that g_j obeys a Hölder condition of index $\frac{1}{4}$ on every closed subinterval of its domain. We also proved that g_j obeys a Hölder condition of index $(1 + 3\nu)/4$ on every closed subinterval of its domain if $\dot{\gamma}_j$ obeys a Hölder condition of index ν $(0 < \nu \leqslant 1)$ on every closed subinterval of its domain.

[19]This is equivalent to the Ernst equation for the complex-valued function $E := (1 + iS_{12})/S_{22}$. See, e.g., Sec. II F in Ref. 1.

[20]In some previous papers by the present authors, P was used to denote any member of a gauge of spectral potentials which differs from the present gauge and should not be confused with it.

[21]I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980); "The Riemann–Hilbert approach to the axial Einstein equations," in *Solitons in General Relativity*, edited by H. C. Morris and R. Dodd (Plenum, New York, to be published). Also, see I. Hauser "On the homogeneous Hilbert problem for effecting Kinnersley–Chitre transformations," in Lecture Notes in Physics, Vol. 205, *Solutions of Einstein's Equations: Techniques and Results*, edited by C. Hoenselaers and W. Dietz (Springer, Berlin, 1984), pp. 128–175.

[22]F. J. Ernst, A. García-Diaz, and I. Hauser, J. Math. Phys. **29**, 681 (1988).

[23]Section II I of Ref. 1.

The theorem of Greenberg and Robinson for two-dimensional quantum field theories

Klaus Baumann
Institut für Theoretische Physik, Bunsenstrasse 9, D-3400 Göttingen, Federal Republic of Germany

(Received 15 November 1988; accepted for publication 7 June 1989)

In two space-time dimensions there are quantum fields φ obeying $\Box\varphi = 0$, which nevertheless have nonvanishing higher truncated n-point functions. Such fields show up if one wants to adapt the theorem of Greenberg and Robinson to $1 + 1$ space-time dimensions. Using the Jost–Lehmann–Dyson representation, it is shown that if either (a) $\tilde{\varphi}(p) = 0$ for spacelike momenta or (b) $\widetilde{W}_{\varphi^+\varphi}(p)$ decreases at least like $\exp(-p^2)$, then the field φ is the sum of two local fields A and B, where A is a generalized free field and B satisfies $\Box B = 0$.

I. INTRODUCTION

Let $\varphi(x)$ be a relativistic quantum field in $n + 1$ space-time dimensions. If $n \geqslant 2$ and if we assume $\tilde{\varphi}(p) \equiv 0$ for $p^2 < 0$ (spacelike p), then $\varphi(x)$ is necessarily a generalized free field, i.e.,

$$[\varphi_i(x),\varphi_j(y)]_- = (\Omega,[\varphi_i(x),\varphi_j(y)]_-\Omega),$$

for Bose fields, respectively,

$$[\varphi_i(x),\varphi_j(y)]_+ = (\Omega,[\varphi_i(x),\varphi_j(y)]_+\Omega),$$

for Fermi fields.

This theorem proven by Greenberg[1] and Robinson[2] remains true in $1 + 1$ space-time dimensions, only if we exclude fields of mass zero. The reason is well known: In two dimensions there are fields obeying $\Box\varphi_i(x) = 0$ that are not free fields in the above sense. The best known example are the Wick products of $j_+ = j_0 + j_1$, resp. $j_- = j_0 - j_1$, where $j_\mu(x)$ is a free conserved current (see Ref. 3). In the following we shall show that the two-dimensional version of the theorem holds true up to a field $B(x)$, which fulfills $\Box B(x) = 0$.

II. RESULTS

Theorem: Let $\varphi_i(x)$ be the components of a Wightman field in $1 + 1$ space-time dimensions.

(a) If $\tilde{\varphi}_i(p) = 0$ for $-M_2^2 < p^2 < -M_1^2 \leqslant 0$ or (b) if $\widetilde{W}_{\varphi_i^+\varphi_i}(p)\exp(\alpha p^2) \in \mathscr{S}'(\mathbb{R}^2)$, $\alpha > 0$ for all components φ_i, then φ can be written as $\varphi(x) = A(x) + B(x)$ and we have (i) $A(x)$ is a generalized free field, i.e.,

$$[A_i(x),A_j(y)]_{(\mp)} = (\Omega,[A_i(x),A_j(y)]_{(\mp)}\Omega)$$

(ii) $B(x) = 0$,

(iii) $[A_i(x),B_j(y)]_{(\mp)} = (\Omega,[A_i(x),B_j(y)]_{(\mp)}\Omega)$.

Remarks:

(1) Of course $B(x)$ can be a free field too. (2) For the proof we do not have to assume that φ transforms covariantly under the Lorentz group. Therefore the theorem is valid also for fields φ with infinitely many components. But if φ transforms finitely covariantly then it suffices to assume $\tilde{\varphi}(p) = 0$ for two open sets contained in the right resp. left spacelike cone. (3) From our previous work[3] we expect that the assumption (b) can be weakened to

(b′) $\widetilde{W}_{\varphi_i^+\varphi_i}(p)\exp(\alpha\sqrt{p^2}) \in \mathscr{S}'(\mathbb{R}^2)$, $\alpha > 0$,

but we did not succeed to prove it. Nevertheless the assumption (b) is weaker than the original assumption made by Greenberg and Borchers—namely, $\widetilde{W}_{\varphi^+\varphi}(p) = 0$ for $p^2 > M^2$.

III. PROOF OF THE THEOREM

Depending on the statistics of the field φ either the commutator $[\varphi_i(x),\varphi_j(y)]_-$ (Bose field) or the anticommutator $[\varphi_i(x),\varphi_j(y)]_+$ (Fermi field) vanishes for spacelike distances because of locality. The proof of the theorem will be given in four lemmas. In Lemma 1 we analyze the support of $[\tilde{\varphi}(p_1),\tilde{\varphi}(p_2)]\Omega$. Lemma 2 characterizes the commutator. In Lemma 3 we show that the support of $\tilde{\varphi}$ is restricted to $\overline{V_+ \cup V_-}$. This fact will be used in Lemma 4 to decompose $\tilde{\varphi}$ according to the theorem.

For an arbitrary vector Ψ we consider

$$\widetilde{F}(p,q) := (\Psi,[\tilde{\varphi}_i(p/2 - q),\tilde{\varphi}_j(p/2 + q)]\Omega).$$

From the spectrum condition we know $\widetilde{F}(p,q) = 0$ if $p \notin \overline{V}_+$. As a first step we show the following lemma.

Lemma 1: For all $\Psi_\perp = (1 - P_\Omega)\Psi$ we have

$$\operatorname{supp} \widetilde{F} \subseteq \{p_0 = p_1 \geqslant 0\} \times \{q_0 = q_1\} \cup \{p_0 = -p_1 \geqslant 0\} \times \{q_0 = -q_1\}$$

i.e., $\widetilde{F}(p,q) \equiv 0$ unless $p \in L_+ = \{p; p^2 = 0, p_0 > 0\}$ and q parallel to p.

The proofs for the cases (a) and (b) differ; however, in both cases we use the Jost–Lehmann–Dyson representation.

Proof: Case (a): $\widetilde{F}(p,q)$ is the Fourier transform of $(\Psi[\varphi(x - \xi/2),\varphi(x + \xi/2)]\Omega)$. For fixed p the distribution

$$G_p(\xi) := \int \frac{d^2q}{2\pi}\, e^{iq\xi}\widetilde{F}(p,q)$$

vanishes if $\xi^2 < 0$ and we can use the JLD representation (see Ref. 4).

$$\widetilde{G}_p(q,\sigma) := \int \frac{d^2\xi}{2\pi} G_p(\xi)\cos(\sigma\sqrt{\xi^2})e^{-iq\xi}$$

is a solution of the wave equation $(\partial_{q_0}^2 - \partial_{q_1}^2 - \partial_\sigma^2)\widetilde{G}_p(q,\sigma) = 0$, symmetric in σ. For $\sigma = 0$ we have $\widetilde{G}_p(q,0) = \widetilde{F}(p,q)$ and $(\partial_\sigma \widetilde{G}_p)(q,0) = 0$. From our assumption (a) and for $p \in V_+$ it follows that $\widetilde{G}_p(q,\sigma) \equiv 0$ and therefore $\widetilde{F}(p,q) \equiv 0$ as has already been shown by

 0022-2488/89/102337-03$02.50

Greenberg[1] and Robinson.[2] For $p \in L_+$ we get by the same methods

$$\operatorname{supp} \tilde{G}_p(\cdot,\sigma) \subseteq \{\lambda p : \lambda \in \mathbb{R}\} \cup p/2 + \bar{V}_+ \cup -p/2 - \bar{V}_+$$

for all σ. The Cauchy data of $\tilde{G}_p(q,\sigma)$ in the plane $q_0 = 0$ are concentrated on the line $q_1 = 0$, i.e.,

$$\tilde{G}_p(0,q_1,\sigma) = \sum_{k=0}^{N} \delta^{(k)}(q_1) f_k(\sigma),$$

$$(\partial_{q_0} \tilde{G}_p)(0,q_1,\sigma) = \sum_{k=0}^{N} \delta^{(k)}(q_1) g_k(\sigma).$$

If we express $\tilde{G}_p(q,0)$ by its Cauchy data we realize immediately that $\tilde{G}_p(q,0) \equiv \tilde{F}(p,q)$ as a function of q is finite covariant with respect to Lorentz transformations. Therefore $\operatorname{supp} \tilde{F}(p,\cdot)$ is given by (i) $\{q_0 = q_1\}$ if $p_0 = p_1 > 0$, (ii) $\{q_0 = -q_1\}$ if $p_0 = -p_1 > 0$.

Case (b): In this case we use the same strategy as explained in our previous paper.[3] Therefore we shall be very sketchy and refer the reader to Ref. 3 for the details.

$\tilde{F}(p,q)$ can be written as

$$\tilde{F}(p,q) = (\Psi_\perp, \tilde{\varphi}_i(p/2-q)\tilde{\varphi}_j(p/2+q)\Omega)$$
$$\overline{(+)}(\Psi_\perp, \tilde{\varphi}_i(p/2+q)\tilde{\varphi}_j(p/2-q)\Omega)$$
$$= \tilde{F}_+(p,q)\ \overline{(+)}\ \tilde{F}_-(p,q)$$

and for the supports with respect to q we have for fixed $p \in \bar{V}_+$:

$$\operatorname{supp} \tilde{F}_+(p,\cdot) \subseteq -p/2 + \bar{V}_+,$$
$$\operatorname{supp} \tilde{F}_-(p,\cdot) \subseteq p/2 - \bar{V}_+.$$

From assumption (b) $\tilde{W}_{\varphi_i^+\varphi_i}(p)\exp(\alpha p^2) \in \mathscr{S}'(\mathbb{R}^2)$ for some $\alpha > 0$ and because of the Cauchy–Schwarz inequality we have

$$\tilde{F}_+(p,q)\exp[(\alpha/2)(p/2+q)^2] \in \mathscr{S}'(\mathbb{R}^4),$$

resp.

$$\tilde{F}_-(p,q)\exp[(\alpha/2)(p/2-q)^2] \in \mathscr{S}'(\mathbb{R}^4).$$

Therefore $\tilde{F}(p,q)\cosh(\sigma\sqrt{q^2})$ exists as a distribution in $\mathscr{S}'(\mathbb{R}^4) \times \mathscr{D}'(\mathbb{R})$ with respect to (p,q,σ). For fixed p the Fourier transform with respect to q

$$G_p(\sigma,\xi) = \int \frac{d^2q}{2\pi} \tilde{F}(p,q)\cosh(\sigma\sqrt{q^2})e^{iq\xi}$$

is a solution of the wave equation $(\partial_\sigma^2 + \partial_{\xi_0}^2 - \partial_{\xi_1}^2)G_p(\sigma,\xi) = 0$, symmetric in σ. Because of locality we get for $\sigma = 0$ $G_p(0,\xi) = 0$ if $\xi^2 < 0$. Asgeirsson's lemma implies $G_p(\sigma,\xi) = 0$ if $\xi^2 < -\sigma^2$ and because for fixed p and σ the support of $\tilde{F}(p,q)\cosh(\sigma\sqrt{q^2})$ is contained in $-p/2 + \bar{V}_+ \cup p/2 - \bar{V}_+$ we can use Araki's extension[5] of JLD to show $G_p(\sigma,\xi)$ for all $\xi^2 < 0$. From our assumption (b) it follows that $\tilde{F}(p,q)\exp((\alpha/4)q^2)$ defines a tempered distribution and can be written as

$$\tilde{F}(p,q)\exp\left(\frac{\alpha}{4}q^2\right) = \frac{1}{\sqrt{\pi\alpha}} \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2}{\alpha}\right)\tilde{F}(p,q)\cosh(\sigma\sqrt{q^2})d\sigma.$$

Therefore its Fourier transform

$$H_p(\xi) = \int \frac{d^2q}{2\pi} \tilde{F}(p,q)\exp\left(\frac{\alpha}{4}q^2\right)e^{i\xi q} = \frac{1}{\sqrt{\pi\alpha}} \int G_p(\sigma,\xi)\exp\left(-\frac{\sigma^2}{\alpha}\right)d\sigma$$

satisfies locality, i.e., $H_p(\xi) = 0$ if $\xi^2 < 0$.

Furthermore as long as $p \in V_+$,

$$\tilde{F}(p,q)\exp((\alpha/4)q^2)\exp(\pm\gamma q_0) \in \mathscr{S}'(\mathbb{R}^4)$$

for some $\gamma > 0$ and therefore $H_p(\xi)$ vanishes because it is analytic in ξ_0.

Now we consider $p \in L_+$, i.e., $p_0 = p_1 > 0$ or $p_0 = -p_1 > 0$. For $p_0 = p_1 > 0$ we have

$$\tilde{F}(p,q)\exp((\alpha/4)q^2)\exp(\pm\gamma q_-) \in \mathscr{S}'(\mathbb{R}^4)$$

for some $\gamma > 0$, where $q_- = q_0 - q_1$, $q_+ = q_0 + q_1$ and therefore $H_p(\xi)$ is analytic in ξ_+ because $q\xi = \frac{1}{2}\{q_+\xi_- + q_-\xi_+\}$. Because of locality this implies $H_p(\xi) = 0$ as long as $\xi_- \neq 0$ and we have the representation (for fixed p)

$$H_p(\xi) = \sum_{k=0}^{N} \delta^{(k)}(\xi_-) f_k(\xi_+),$$

therefore $\tilde{F}(p,q)\exp((\alpha/4)q^2)$ is a polynomial in q_+ and because of the support properties we have $\operatorname{supp} \tilde{F}(p,\cdot) \subseteq \{q_- = 0\}$ if $p_- = 0$. Starting from $p_0 = -p_1 > 0$ we get $\operatorname{supp} \tilde{F}(p,q) \subseteq \{q_+ = 0\}$ if $p_+ = 0$. This proves Lemma 1.

Lemma 2: For the commutator $[\varphi_i(x-\xi/2), \varphi_j(x+\xi/2)]$ we have (a) $\Box_x[\varphi(x-\xi/2), \varphi(x+\xi/2)] = 0$, (b) $[\varphi(x-\xi/2), \varphi(x+\xi/2)] = \Delta(\xi) + C_+(x_+,\xi) + C_-(x_-,\xi)$, where C_+, C_- are the nontrivial parts, and Δ is the VEV of the commutator. (c) For the Fourier transforms $\tilde{C}_\pm(p,q)$ we have $\tilde{C}_+(p,q) = \delta(p_+)\hat{C}_+(p_-,q)$, $\operatorname{supp} \hat{C}_+(p_-,\cdot) \subseteq \{q_+ = 0\}$ and $\tilde{C}_-(p,q) = \delta(p_-)\hat{C}_-(p_+,q)$, $\operatorname{supp} \hat{C}_-(p_+,\cdot) \subseteq \{q_- = 0\}$.

Proof: (a) $[\tilde{\varphi}_i(p/2-q), \tilde{\varphi}_j(p/2+q)]\Omega \equiv 0$ unless $p^2 = 0$ and because of positivity this implies $p^2[\tilde{\varphi}_i(p/2-q), \tilde{\varphi}_j(p/2+q)]\Omega \equiv 0$. This proves (a) because of the Reeh–Schlieder theorem. (b) From $\Box_x = \partial_{x_0}^2 - \partial_{x_1}^2 = \frac{1}{4}\partial_{x_+}\partial_{x_-}$ it follows

$$[\varphi(x-\xi/2), \varphi(x+\xi/2)] = (\Omega[\varphi(x-\xi/2), \varphi(x+\xi/2)]\Omega) + C_+(x_+,\xi) + C_-(x_-,\xi) \text{ with } (\Omega, C_\pm(x_\pm,\xi)\Omega) \equiv 0.$$

This proves (b), and (c) is an immediate consequence of (b) and Lemma 1.

Lemma 3: (a) For $f \in \mathscr{S}(\mathbb{R}^2)$ and for all momenta r with $r^2 \neq 0$ we have

$$[\varphi(f), \tilde{\varphi}(r)] = (\Omega, [\varphi(f), \tilde{\varphi}(r)]\Omega).$$

(b) $\operatorname{supp} \tilde{\varphi} \subseteq \overline{V_+ \cup V_-}$.

Proof: (a) Because of Lemma 2,

$$[\tilde{\varphi}(p/2-q), \tilde{\varphi}(p/2+q)] - (\Omega[\tilde{\varphi}(p/2-q), \tilde{\varphi}(p/2+q)]\Omega) = \delta(p_+)\hat{C}_+(p_-,q) + \delta(p_-)\hat{C}_-(p_+,q) \equiv 0$$

unless $p_+ = 0$ and $q_+ = 0$ or $p_- = 0$ and $q_- = 0$. But if $p_+ = 0$, $q_+ = 0$ or $p_- = 0$, $q_- = 0$ then $r^2 = (p/2 + q)^2 = 0$. This proves (a). (b) Take $f, g \in \mathscr{S}(\mathbb{R}^2)$ and $\operatorname{supp} \tilde{g} \subseteq \{p^2 < 0\}$ then $[\varphi(f), \varphi(g)] = 0$ because $(\Omega, [\varphi(f), \varphi(g)]\Omega)$ vanishes. Therefore $\varphi(g)\varphi(f_N) \cdots \varphi(f_1)\Omega \equiv 0$ because of the spectrum condition. This proves (b).

Remark: The support properties of $\tilde{\varphi}$ suggest to write φ as a sum of two fields A and B such that $\varphi(x) = A_\epsilon(x) + B_\epsilon(x)$ and $\widetilde{B}_\epsilon(p) = 0$ if $p^2 > \frac{3}{2}\epsilon$, respectively, $\widetilde{A}_\epsilon(p) = 0$ if $p^2 < \epsilon/2$. Normally such a decomposition will destroy locality, but because of the commutation relations given in Lemma 3(a) locality will be preserved.

With the help of the above three lemmas we can finally prove the theorem namely by use of the following lemma.

Lemma 4: The field $\varphi(x)$ can be written as $\varphi = A + B$ with (i) $A(x)$ is a generalized free field, i.e.,

$$[A_i(x), A_j(y)] = (\Omega, [A_i(x), A_j(y)]\Omega).$$

(ii) $B(x) = 0$,

(iii) $[A_i(x), B_j(y)] = (\Omega, [A_i(x), B_j(y)]\Omega)$.

Proof: As indicated above we write in momentum space

$$\tilde{\varphi}(p) = \chi_\epsilon(p^2)\tilde{\varphi}(p) + [1 - \chi_\epsilon(p^2)]\tilde{\varphi}(p) = \widetilde{A}_\epsilon(p) + \widetilde{B}_\epsilon(p),$$

where $0 \leqslant \chi_\epsilon(s) = \chi(s/\epsilon) \in C^\infty$, $0 \leqslant \chi'(s) \in \mathscr{D}([\frac{1}{2}, \frac{3}{2}])$ and

$$\chi(s) = \begin{cases} 0, & \text{if } s < \frac{1}{2} \\ 1, & \phantom{\text{if }} s > \frac{3}{2} \end{cases}.$$

Because of Lemma 3 we have the following commutation relations:

$$[A_\epsilon(x), A_\epsilon(y)] = (\Omega, [A_\epsilon(x), A_\epsilon(y)]\Omega),$$
$$[A_\epsilon(x), B_\epsilon(y)] = (\Omega, [A_\epsilon(x), B_\epsilon(y)]\Omega),$$

and

$$[B_\epsilon(x), B_\epsilon(y)] - (\Omega, [B_\epsilon(x), B_\epsilon(y)]\Omega) = C_+((x+y)/2, x-y) + C_-((x+y)/2, x-y)$$

independent of ϵ and given by

$$= [\varphi(x), \varphi(y)] - (\Omega, [\varphi(x), \varphi(y)]\Omega).$$

Therefore A_ϵ and B_ϵ define fields which fulfill locality. As ϵ goes zero the limits of the various commutators exist and therefore the limiting fields $A(x)$ and $B(x)$ exist too. Because of positivity $\Box B(x)\Omega$ must be zero. This proves Lemma 4.

[1] O. W. Greenberg, "Heisenberg fields which vanish on domains of momentum space," J. Math. Phys. 3, 859 (1962).

[2] D. W. Robinson, "Support of a field in momentum space," Helv. Phys. Acta **35**, 403 (1962).

[3] K. Baumann, "On the two-point functions of interacting Wightman fields," J. Math. Phys. **27**, 828 (1986).

[4] A. S. Wightman, "Analytic functions of several complex variables, in *Relations de Dispersion et Particules Élémentaires*, edited by C. De Witt and R. Omnès (Hermann, Paris, 1960).

[5] H. Araki, "A generalization of Borchers theorem," Helv. Phys. Acta **36**, 132 (1963).

Hyperfunction quantum field theory: Basic structural results

Erwin Brüning
Department of Physics, Princeton University, Princeton, New Jersey 08544
Shigeaki Nagamachi
Technical College, Tokushima University, Tokushima 770, Japan

(Received 23 February 1989; accepted for publication 5 April 1989)

The choice of the class E' of generalized functions on space-time in which to formulate general relativistic quantum field theory (QFT) is discussed. A first step is to isolate a set of conditions on E' that allows a formulation of QFT in otherwise the same way as the original proposal by Wightman [Ark. Fys. **28**, 129 (1965)], where E' is the class of tempered distributions. It is stressed that the formulation of QFT in which E' equals the class of Fourier hyperfunctions on space-time meets the following requirements: (A) Fourier hyperfunctions generalize tempered distributions thus allowing more singular fields as suggested by concrete models; (B) Fourier hyperfunction quantum fields are localizable both in space-time and in energy-momentum space thus allowing the physically indispensable standard interpretation of Poincaré covariance, local commutativity, and localization of energy-momentum spectrum; and (C) in Fourier hyperfunction quantum field theory almost all the basic structural results of "standard" QFT (existence of a PCT operator, spin-statistics theorems, existence of a scattering operator, etc.) hold. Finally, a short introduction to that part of Fourier hyperfunction theory needed in this context is given.

I. INTRODUCTION

A. Some motivation

A formulation of general relativistic quantum field theory (QFT) always has to start with a decision about the choice of the test-function space. For well known reasons the traditional choice for the test-function space E is

$$E = \mathscr{S}(\mathbb{R}^4,V) \simeq \mathscr{S}(\mathbb{R}^4) \otimes V, \tag{1.1}$$

where V is a finite-dimensional vector space and $\mathscr{S}(\mathbb{R}^4)$ is the Schwartz space of all C^∞ functions on space-time $\mathbb{R}^4$ that decay together with all their derivatives faster than any $(\text{polynomial})^{-1}$ (see Refs. 1–3).

Since the early days of QFT, for various reasons, there has been some discussion on this choice in the literature. Later we will discuss some of these proposals. The main reason for considering other test-function spaces are indications coming from model constructions that one has to admit (A) that there are more singular than tempered fields, respectively, stronger growth properties of the fields in energy-momentum space. This requirement is fulfilled by a test-function space E if the elements of E are "smoother" in coordinate space and decay more rapidly in energy-momentum space than those in $\mathscr{S}(\mathbb{R}^4)$. If one has a choice for the test-function space E that meets requirements (A) one usually gets into trouble with (B) an unambiguous and clear motion of localization in coordinate and momentum space, and accordingly not much is then known about (C) the permanence of the basic structural results known in QFT for tempered fields (more details follow later).

In this paper we want to show that there is a test-function space E that satisfies all three requirements (A)–(C). This test-function space

$$E = \underset{\sim}{\mathscr{O}}(\mathbb{D}^4,V) \tag{1.2}$$

is defined and described in Sec. II. Elements of its topological dual E' are called "Fourier hyperfunctions."

The suggestion to use a Fourier hyperfunction in quantum field theory has been made by Nagamachi and Mugibayashi in a series of papers.[4–7] This first suggestion is, at least for nonexperts in (Fourier) hyperfunctions, not always very transparent and clear, thus hiding in part its main achievements.

Accordingly one goal of this paper is to give a short but clear and complete introduction to QFT in terms of Fourier hyperfunctions. In particular, we present a more transparent (for nonexperts in hyperfunctions) account of the highly nontrivial fact that QFT in terms of Fourier hyperfunctions can deal very well in a "good physical understanding" with the localization problem in coordinate and momentum space [point (B) above] though the underlying space of test functions contains no elements of compact support, neither in coordinate nor in momentum space.

An important hint in favor of QFT in terms of Fourier hyperfunctions comes from the construction of concrete models. This has been discussed in more detail by Wightman.[8]

B. Quantum fields and their dependence on the test-function space

We begin by recalling the defining assumptions of general quantum field theory. For reasons that will become evident later we present here a variation of the set of assumptions proposed by Gårding and Wightman. In order to stress our point of view that the choice of the space of test functions is at one's disposal according to the problems at hand we start by isolating a list of conditions on a space E of functions

 0022-2488/89/102340-20$02.50

on space-time in order that E be "admissible" as a test-function space of a relativistic quantum field theory.

(H_0) *The test-function space E*:

(a) The test-function space E is a locally convex topological vector space of functions on space-time $\mathbb{R}^4$.

(i) E admits the Fourier transformation $\mathscr{F}$ as an isomorphism of topological vector spaces.

(ii) For continuous linear functionals on E and on $\widetilde{E} = \mathscr{F}E$ the notion of support is available.

(b) On E and on $\widetilde{E}$ continuous involutions $f \to f^*$ are defined satisfying $(\mathscr{F}f)^* = \mathscr{F}(f^*)$, for all $f \in E$.

(c) The vector space E has a Z_2 grading and is accordingly decomposed into subspaces of "even" and "odd" elements:

$$E = E_0 \oplus E_1.$$

(d) The universal covering group $G = \mathrm{iSL}(2,\mathbb{C})$ of the Poincaré group acts on E by continuous linear maps α_g: $E \to E$, $g \in G$, such that, for all $f \in E$ and all $g \in G$,

(i) $\alpha_g(f)^* = \alpha_g(f^*)$,

(ii) $g \to \alpha_g(f)$ is a differentiable map $G \to E$,

(iii) α_g preserves the grading.

(H_1) *Fields over E or fields with test-function space E*: A field A over such a vector space E with state space $\mathscr{H}$, domain $\mathscr{D}$, and cyclic unit vector Φ_0 is specified in the following way.

(a) The state space is a (separable) complex Hilbert space $\mathscr{H}$.

(b) The domain $\mathscr{D}$ is a dense subspace of $\mathscr{H}$ containing the cyclic vector Φ_0.

(c) The field A is a linear map form E into the algebra $L(\mathscr{D},\mathscr{D})$ of linear operators $\mathscr{D} \to \mathscr{D}$ such that the following conditions hold.

(i) For all $\Phi,\Psi \in \mathscr{D}$, $f \to (\Phi, A(f)\Psi)$ is a continuous linear map $E \to \mathbb{C}$.

(ii) For each $f \in E$, the adjoint operator $A(f)^*$ of the densely defined operator $A(f)$ in $\mathscr{H}$ is an extension of $A(f^*)$:

$$A(f^*) \subset A(f)^*.$$

(iii) The linear span

$$\mathscr{D}_0 = \text{lin span}\{\Phi_0, A(f_{j_1}) \cdots A(f_{j_n})\Phi_0 \,|\, f_{j_i} \in E,\; n = 1,2,\ldots\}$$

is dense in $\mathscr{H}$.

(H_2) *Poincaré covariance*: A field $(A,\mathscr{H},\mathscr{D},\Phi_0)$ over E is said to be Poincaré covariant if and only if there is a unitary continuous representation U of $G = \mathrm{iSL}(2,\mathbb{C})$ on the Hilbert space $\mathscr{H}$ such that, for all $g \in G$ and all $f \in E$,

$$U(g)\mathscr{D} = \mathscr{D},$$

$$U(g)A(f)U(g)^* = A(\alpha_g f).$$

(H_3) *Energy-momentum spectrum* Σ: The energy-momentum spectrum Σ of the theory equals the spectrum $\sigma(P)$ of the infinitesimal generator $P = (P^0,P^1,P^2,P^3)$ of the time–space translations in the representation U, i.e.,

$$U(a,\mathbf{1}) = e^{ia\cdot P}, \quad a \in \mathbb{R}^4.$$

It is contained in the closed "forward light cone"

$$\overline{V}_+ = \{(q^0,\mathbf{q}) \in \mathbb{R}^4 \,|\, q^0 \geqslant |\mathbf{q}|, \quad \mathbf{q} \in \mathbb{R}^3\}$$

and contains the origin, i.e., $0 \in \Sigma \subset \overline{V}_+$.

(H_4) *Locality* (*local commutativity*): The restrictions A_α of A to E_α, $\alpha = 0,1$, satisfy, for all $\alpha, \beta \in \{0,1\}$,

$$\operatorname{supp}\langle A_\alpha, A_\beta\rangle \subset K,$$

where

$$K = \{(x,y) \in \mathbb{R}^4 \times \mathbb{R}^4 \,|\, y - x \in \overline{V} = \overline{V}_+ \cup (-\overline{V}_+)\}$$

and where $\langle A_\alpha, A_\beta\rangle$: $E_\alpha \times E_\beta \to L(\mathscr{D},\mathscr{D})$ is defined by

$$\langle A_\alpha, A_\beta\rangle(f_\alpha, f_\beta) = A_\alpha(f_\alpha)A_\beta(f_\beta) - \sigma_{\alpha\beta} A_\beta(f_\beta)A_\alpha(f_\alpha),$$

with $\sigma_{\alpha\beta} \in \{1, -1\}$, usually $\sigma_{\alpha\beta} = (-1)^{\alpha\beta}$.

(H_5) *Uniqueness of the vacuum state*: The subspace $\mathscr{H}_0$ of all translation invariant states in $\mathscr{H}$, i.e.,

$$\mathscr{H}_0 = \{\Psi \in \mathscr{H} \,|\, U(a,\mathbf{1})\Psi = \Psi, \text{ for all } a \in \mathbb{R}^4\},$$

is one dimensional and is generated by the cyclic unit vector Φ_0:

$$\mathscr{H}_0 = C\Phi_0.$$

Conditions (H_0)–(H_5) characterize a *relativistic quantum field A over E*.

Remark 1.1:

(a) For well known reasons the original choice for the test-function space E was $E = \mathscr{S}(\mathbb{R}^4,V)$, where V is some finite-dimensional vector space depending on the "type of fields" under discussion. Here the type of a field is specified by its transformation properties with respect to G, i.e., by α_g, $g \in G$. Clearly this test-function space satisfies condition (H_0).

(b) Notice that by an appropriate choice of the test-function space E (together with the action α of G on E and the involution $*$) the general case of a finite number of scalar, vector, tensor, and/or spinor fields as well as the case of non-Hermitian fields is covered by our formulation.

(c) Sometimes the spectral assumption (H_3) is strengthened by the requirement that the point $p = 0$ be isolated in Σ.

(d) The realization of the locality condition (H_4) [and the spectral condition (H_3)] depends on the test-function space E. If E contains functions on space-time with compact support this is understood in the obvious way. Otherwise an appropriate interpretation of this condition has to be given.

In any case, (H_4) says that the bilinear functional

$$(f,g) \to (\Psi, \langle A_\alpha, A_\beta\rangle(f,g)\Phi)$$

on $E \times E$ has its "support" in K for any $\Psi,\Phi \in \mathscr{D}$.

The value of $\sigma_{\alpha\beta} \in \{1, -1\}$ has to be specified according to the type of fields in agreement with the "spin and statistics" theorem.

(e) It is mainly part (ii) of condition (a) in the characterization (H_0) of an "admissible" test-function space that prevents an easy and/or obvious choice of E besides the traditional one [(1.1)].

Clearly one was well aware already at the beginning of general QFT that the choice of the underlying space of test functions is not only a technical assumption but also has implications of physical relevance.

(1) The allowed growth properties for a field and its

singular behavior depend on the test-function space [i.e., point (A)].

(2) Accordingly the class of interactions that can be controlled depends on the test-function space (distinction between "renormalizable" and "nonrenormalizable" interactions).

(3) The concrete realization of the locality and spectral condition depends on the test-function space.

For further details on points (1) and (2) we refer the reader to Refs. 8 and 9. Point (3) will be discussed in considerable detail in a later section. The localization problem in connection with the choice of the test-function space is also discussed in Sec. 15.5 of Ref. 3.

As a last but important point we want to recall that for the usual choice (1.1) of the test-function space E there are still no "nontrivial" models of relativistic quantum fields on physical space-time.

These are some reasons for considering QFT over test-function spaces other than the traditional one. Further reasons are presented in Refs. 8 and 9. Accordingly several attempts have been made in this direction, which we want to review briefly. Before doing this, however, we want to stress that any interesting modification of the test-function space E should still allow us to deduce all the structural results of QFT or at least most of them in order to meet requirement (C).

These structural results of QFT we have in mind here are (1) the existence of a PCT operator, (2) the connection between spin and statistics, (3) the existence of a scattering operator, and some further important but more technical results: (4) the cluster property, (5) analyticity results, (6) the global nature of local commutativity, (7) the general form of the two-point function, (8) the Borchers class of a field, (9) the Jost–Schroer theorem, (10) Euclidean reformulation, and (11) dispersion relations. The proofs of these results as given in the literature[1–3] usually seem to rely on the assumption of "temperedness" in an essential way. Nevertheless it is possible to prove some of these results also for various test-function spaces other than $E = \mathscr{S}(\mathbb{R}^4, V)$ as our review will show.

For the test-function space (1.2) for Fourier hyperfunctions we will prove the results (1), (2), and (4)–(9). The remaining points (3), (10), and (11) will be discussed in the last section.

The main sources of difficulties in proving these statements are (i) that there are no test functions of compact support and (ii) that continuous linear functionals on $\underset{\sim}{\mathscr{O}}(\mathbb{D}^4, V)$ we have a "support at infinity."

C. A short review

In 1967, Jaffe[10] seems to have been the first to consider the choice of test-function spaces for relativistic quantum fields systematically. In order to be able to realize the locality condition in the traditional way he determined a class of function spaces E_J on energy-momentum space $\mathbb{R}^4$ such that (i) $\mathscr{D}(\mathbb{R}^4) \subset E_J \subset \mathscr{S}(\mathbb{R}^4)$, and (ii) $\mathscr{F} E_J$, i.e., the space of Fourier transforms of elements in E_J, contains (enough) functions of compact support.

Somewhat later (1969) Iofa and Fainberg[11] proposed using a test-function space $E_I = E_I(\mathbb{R}^4)$ of entire functions that are polynomially decreasing in any strip $|\mathrm{Im}\, z_j| < \delta$, $\delta > 0$. Since such a space does not contain any function on space-time of compact support the localization of the fields is not possible in the usual sense. Accordingly they are called *nonlocalizable fields*. Clearly the locality condition (H_4) also cannot be formulated in a natural way for such fields. Nevertheless several structural properties [(4) and (5)] could be proved and some others [(1) and (2)] were indicated in such a theory.

In 1971, Constantinescu observed[2] that localizability of the fields in the above sense and locality of the fields according to (H_4) are different notions. He explained this on the level of two-point functions. However, this is not sufficient for the locality of the whole theory. Constantinescu proposed an inductive limit space $E_C(\mathbb{R}^4_p)$ of C^∞ functions on energy-momentum space such that $\mathscr{D}(\mathbb{R}^4) \subset E_C \subset \mathscr{S}(\mathbb{R}^4)$ and $\mathscr{F} E_C$ consists of test functions f holomorphic in some strip $|\mathrm{Im}\, z_j| < \delta$, $\delta = \delta(f) > 0$. He proves some structural properties of QFT and discusses some others.

Finally there is a series of papers by Lücke concerning the choice of the test-function space and the corresponding realization of the locality condition (H_4) as well as the structural properties (1)–(10). A recent source of information about this and further references is Ref. 13.

Lücke proposes to take the Gel'fand spaces $\mathscr{S}^s(\mathbb{R}^4)$, $0 \leqslant s < \infty$, on space-time, defined and studied in Chap. IV of Ref. 14 as test-function spaces for QFT. If $s > 1$, then $\mathscr{S}^s(\mathbb{R}^4)$ contains enough C^∞ functions of compact support; hence localization in the usual sense is possible and thus the usual realization of the locality condition (H_4). If, however, $s \leqslant 1$, then the space $\mathscr{S}^s(\mathbb{R}^4)$ consists of holomorphic functions (entire functions for $0 \leqslant s < 1$) and hence localizability is lost for such test functions. In this case fields are again called nonlocalizable fields.

The locality condition (H_4) is accordingly replaced by the assumption that the fields are "essentially local" which means that "sufficiently many" matrix elements of the (anti-) commutator of the field operators $[A(x_1), A(x_2)]_\pm$ are locally continuous on K with respect to $\mathscr{S}^s(\mathbb{R}^4)$.[15] Since permutation symmetry of the Wightman functions can be proved for essentially local fields,[16] some of the structural properties follow also for this class of fields over $\mathscr{S}^s(\mathbb{R}^4)$, $s \leqslant 1$.[13]

However, on one side it can be shown that there is no clear and unambiguous notion of support for $F \in \mathscr{S}^s(\mathbb{R}^4)'$, $0 \leqslant s < 1$.[13] On the other side we think it to be important that in QFT a sensitive mathematical formulation of the locality condition (H_4) has to realize the idea that the (anti-) commutator of the field operators has its "support" only inside K. Therefore we think that our point (B) above is really indispensable and accordingly explain this point for the test-function space (1.2) in some detail. In particular, we will explain that the notion of support for Fourier hyperfunctions (Sec. II D) used in its realization is a straightforward generalization of the notion of support for distributions and thus provides a genuine realization of the locality condition.

II. FOURIER HYPERFUNCTIONS

A. Introduction

This section introduces the notions and explains the results from Fourier hyperfunction theory that we are going to use. For proofs we clearly have to refer to the literature.[6,17,18]

Recall that the main original motivation for introducing distributions came from the theory of linear partial differential operators with constant coefficients.[19] Similarly hyperfunctions have proved to be an appropriate frame for the theory of linear partial differential operators with real analytic coefficients and those that have "regular singularities."[20]

Hyperfunctions are finite sums of boundary values of certain analytic functions.[21] Thus hyperfunctions generalize distributions. They admit the same basic operations (differentiation, integration, and convolution) as distributions. Just as distributions do, they have a "good notion" of *localization* (which agrees with the known localization properties of distribution if applied to them).

In contrast to distributions, hyperfunctions admit a canonical definition of a product (at least in the simplest case) and this may turn out to be of great importance in applications to QFT.

However, in general, hyperfunctions do not admit a canonical definition of Fourier transform as an isomorphism. Some "growth restrictions at infinity" are needed for this. A way to achieve this is to compactify the underlying space $\mathbb{R}^n$. The *radial compactification* $\mathbb{D}^n$ of $\mathbb{R}^n$ has proved to be very useful here. It is defined in a natural way as follows: Let S_∞^{n-1} be the $(n-1)$-dimensional sphere at infinity, which is homeomorphic to the unit sphere $S^{n-1} = \{x\in\mathbb{R}^n \mid |x| = 1\}$ by the mapping $x \to x_\infty$, where the point $x_\infty \in S_\infty^{n-1}$ lies on the ray connecting the origin with the point $x\in S^{n-1}$. The set $\mathbb{R}^n \cup S_\infty^{n-1}$ equipped with its "natural topology" (a fundamental system of the neighborhood of x_∞ is given by all open cones of arbitrary vertex generated by an arbitrary open neighborhoods of x_∞ in S_∞^{n-1}) is denoted by $\mathbb{D}^n$.[17]

A natural extension of hyperfunctions on $\mathbb{R}^n$ to $\mathbb{D}^n$ leads to Fourier hyperfunctions. It turns out that Fourier hyperfunctions have all the properties of hyperfunctions and, in addition, the Fourier transformation is an isomorphism for them.

The realization of the idea that hyperfunctions on $\mathbb{R}^n$ (and on $\mathbb{D}^n$) are finite sums of boundary values of analytic functions is immediate for the case of one variable ($n = 1$). For $n \geqslant 2$ variables, however, we have to meet serious complications as a result of the considerably more complicated theory of analytic functions of more than one variable. New phenomena of analytic continuation cause the main complication in introducing an appropriate notion of boundary value in the $n \geqslant 2$ variable case. The first approach for a "good" notion of boundary values of holomorphic functions of several variables is due to Sato.[22] He introduced this notion by considering sheaves of germs of analytic functions and their relative cohomology.[23] Later it was realized how to introduce a hyperfunction without using cohomology theory.[24–27] According to Sec. I it is clear in QFT we prefer this second approach adapted to Fourier hyperfunctions in which Fourier hyperfunctions are defined as continuous linear functionals on some space of functions.

B. The test-function space for Fourier hyperfunctions

The spaces of test functions we are going to use later are

$$E = \underset{\sim}{\mathscr{O}}(\mathbb{D}^n, V) \simeq \underset{\sim}{\mathscr{O}}(\mathbb{D}^n) \otimes V, \tag{2.1}$$

with some finite-dimensional vector space V and an inductive limit space of functions on $\mathbb{D}^n$ for $n = 4$, which we describe now. By reasons that will become obvious soon we introduce and study the spaces

$$\underset{\sim}{\mathscr{O}}(K), \quad K \subset \mathbb{D}^n \text{ closed}. \tag{2.2}$$

Let $\{U_m, \ m\in\mathbb{N}\}$ be a fundamental sequence of neighborhoods of K in $\mathbb{Q}^n = \mathbb{D}^n + i\mathbb{R}^n$. Then let $\mathscr{O}_c^m(U_m)$ be the Banach space of functions f analytic on $U_m \cap \mathbb{C}^n$ and continuous on $\overline{U}_m \cap \mathbb{C}^n$ such that

$$\|f\|_m = \sup_{z\in U_m \cap \mathbb{C}^n} |f(z)| e^{|z|/m} \tag{2.3}$$

is finite. The space $\underset{\sim}{\mathscr{O}}(K)$ is now defined as the inductive limit of the Banach spaces $\mathscr{O}_c^m(U_m)$:

$$\underset{\sim}{\mathscr{O}}(K) = \operatorname*{ind\,lim}_{m\to\infty} \mathscr{O}_c^m(U_m). \tag{2.4}$$

The following propositions collect some properties of the space $\underset{\sim}{\mathscr{O}}(K)$ of a *rapidly decreasing analytic function* on K that are used in QFT.

Proposition 2.1: (a) $\underset{\sim}{\mathscr{O}}(K)$ is a DFS space (a dual Fréchet–Schwartz space).

(b) $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ is nuclear and barreled.

(c) $\otimes^n \underset{\sim}{\mathscr{O}}(\mathbb{D})$ is dense in $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ and $\hat{\otimes}^n \underset{\sim}{\mathscr{O}}(\mathbb{D}) = \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$.

And, as a consequence, we have the following proposition.

Proposition 2.2: Let M be a separately continuous n-linear form on $\underset{\sim}{\mathscr{O}}(\mathbb{D}^m)^n = \underset{\sim}{\mathscr{O}}(\mathbb{D}^m) \times \cdots \times \underset{\sim}{\mathscr{O}}(\mathbb{D}^m)$. Then the following conditions hold.

(a) M: $\mathscr{O}(\mathbb{D}^m)^n \to \mathbb{C}$ is jointly continuous.

(b) There is a unique continuous linear form F on $\underset{\sim}{\mathscr{O}}(\mathbb{D}^{mn})$ such that, for all

$f_i \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^m)$,

$M(f_1,\dots,f_n) = F(f_1 \otimes \cdots \otimes f_n)$.

For the proofs of these results we refer to Refs. 4 and 5. Note that the "kernel theorem" for this space [part (b) of Proposition 2.2] can be proved in an elementary way by using the explicit characterization of the topological dual $\underset{\sim}{\mathscr{O}}(\mathbb{D}^{mn})$, obtained in Sec. II C.

The following proposition is concerned with the Fourier transformation.

Proposition 2.3: The Fourier transformation $\mathscr{F}$ is well defined on $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ by

$$(\mathscr{F}f)(p) = (2\pi)^{-n/2} \int e^{ip\cdot x} f(x)\,dx, \tag{2.5}$$

where $\mathscr{F}$ is an isomorphism of the topological vector space $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ with inverse

$$(\overline{\mathscr{F}}f)(x) = (2\pi)^{-n/2} \int e^{-ip\cdot x} \tilde{f}(p)\,dp. \tag{2.6}$$

Proof: Clearly it suffices to consider the case $n = 1$. Here $f \in \mathcal{O}(\mathbb{D})$ means $f \in \mathcal{O}_c^m(U_m) \equiv \mathcal{O}_m$ for some m, where $U_m = \mathbb{D} + i(-1/m, 1/m)$. Then for all $k = p + iq \in \mathbb{C} \cap U_{m+1}$ we obtain, from (2.5),

$$|(\mathcal{F}f)(k)| \leqslant C_m \|f\|_m.$$

Now, if b is real, $|b| \leqslant 1/2m$, then the function f_b defined by $f_b(z) = f(z + ib)$, belongs to $\mathcal{O}_{2m}$ and satisfies

$$\|f_b\|_{2m} \leqslant C_m \|f\|_m, \quad b \in [-1/2m, 1/2m].$$

By analyticity, decay properties, and Cauchy's theorem one proves

$$(\mathcal{F}f_b)(k) = e^{-ik \cdot (ib)} (\mathcal{F}f)(k),$$

so that the above estimates imply, for all $|b| \leqslant 1/2m$,

$$|e^{kb} (\mathcal{F}f)(k)| \leqslant C_m \|f\|_m;$$

hence $\mathcal{F}f \in \mathcal{O}_{2m+1}$ and

$$\|\mathcal{F}f\|_{2m+1} \leqslant C_m \|f\|_m.$$

This proves $\mathcal{F}: \underset{\sim}{\mathcal{O}}(\mathbb{D}^n) \to \underset{\sim}{\mathcal{O}}(\mathbb{D}^n)$ to be a well defined continuous linear map.

Clearly the map $\overline{\mathcal{F}}$ has the same properties. And as usual one proves that $\overline{\mathcal{F}}\mathcal{F}$ is the identity on $\mathcal{O}(\mathbb{D}^n)$. Hence $\mathcal{F}$ is an isomorphism.

C. Fourier hyperfunctions

In our approach a *Fourier hyperfunction on* $\mathbb{D}^n$ is by definition an element of the topological dual $\underset{\sim}{\mathcal{O}}(\mathbb{D}^n)'$ of the space $\underset{\sim}{\mathcal{O}}(\mathbb{D}^n)$. In order to give an explicit characterization of $\underset{\sim}{\mathcal{O}}(\mathbb{D}^n)'$ and to relate this notion of Fourier hyperfunctions with its heuristic definition in Sec. II A as a finite sum of boundary values of holomorphic functions let us introduce the *sheaf* $\widetilde{\mathcal{O}}$ *of slowly increasing holomorphic functions on* $\mathbb{Q}^n$.

For an open subset $\Omega \subset \mathbb{Q}^n$ denote by $\widetilde{\mathcal{O}}(\Omega)$ the set of all analytic functions F on $\Omega \cap \mathbb{C}^n$ such that, for every $\epsilon > 0$ and every compact set $K \subset \Omega$,

$$\|F\|_{K,\epsilon} = \sup_{z \in K \cap \mathbb{C}^n} e^{-\epsilon|z|} |F(z)| \tag{2.7}$$

is finite. Here $\widetilde{\mathcal{O}}(\Omega)$ is called the ($\mathbb{C}$-vector) space of slowly increasing analytic functions on Ω.

If Ω' is another open set contained in Ω we obviously have a well defined restriction map

$$\rho_{\Omega'\Omega}: \quad \widetilde{\mathcal{O}}(\Omega) \to \widetilde{\mathcal{O}}(\Omega'), \quad \rho_{\Omega'\Omega}(F) = F_{|\Omega'}, \quad F \in \widetilde{\mathcal{O}}(\Omega), \tag{2.8}$$

such that

$$\rho_{\Omega''\Omega'} \cdot \rho_{\Omega'\Omega} = \rho_{\Omega''\Omega} \quad \text{and} \quad \rho_{\Omega\Omega} = \mathrm{id},$$

for all open sets $\Omega'' \subset \Omega' \subset \Omega$.

Thus with these restriction maps $\{\widetilde{\mathcal{O}}(\Omega) | \Omega \subset \mathbb{Q}^n \text{ open}\}$ is a presheaf on $\mathbb{Q}^n$. This presheaf actually is a sheaf since furthermore the following localization properties are satisfied.

(L_1) If an open set Ω is covered by open sets Ω_α, $\Omega = \cup_\alpha \Omega_\alpha$, and if all the restrictions $F_{|\Omega_\alpha}$ of a function $F \in \widetilde{\mathcal{O}}(\Omega)$ vanish then the function F itself vanishes.

(L_2) If any collection $\{\Omega_\alpha\}$ of open sets in $\mathbb{Q}^n$ is given together with a collection of functions $F_\alpha \in \widetilde{\mathcal{O}}(\Omega_\alpha)$ satisfying

$$F_{\alpha|\Omega_\alpha \cap \Omega_\beta} = F_{\beta|\Omega_\alpha \cap \Omega_\beta},$$

for all α and β, then there exists a function $F \in \widetilde{\mathcal{O}}(\cup_\alpha \Omega_\alpha)$ such that $F_{|\Omega_\alpha} = F_\alpha$, for all α.

For $j = 1, \ldots, n$, let us introduce the open subsets $W_j = \{z \in \mathbb{Q}^n | \operatorname{Im} z_j \neq 0\}$. The intersection

$$W = \bigcap_{j=1}^n W_j$$

of all these sets consists of 2^n open connected components separated by the "real points." Then

$$\hat{W}_k = \bigcap_{\substack{j=1 \\ j \neq k}}^n W_j$$

includes the real points in the k th variable.

In an obvious way we can consider

$$\sum_{k=1}^n \widetilde{\mathcal{O}}(\hat{W}_k)$$

as a subspace of $\widetilde{\mathcal{O}}(W)$. Thus the factor space

$$\mathcal{R} = \mathcal{R}(\mathbb{D}^n) = \widetilde{\mathcal{O}}(W) \Big/ \Big(\sum_{k=1}^n \widetilde{\mathcal{O}}(\hat{W}_k) \Big) \tag{2.9}$$

consists of equivalence classes $[F]$ of functions $f \in \widetilde{\mathcal{O}}(W)$ where two functions F and F' define the same class if and only if

$$F' - F = \sum_{k=1}^n F_k, \quad F_k \in \widetilde{\mathcal{O}}(\hat{W}_k). \tag{2.10}$$

The topological dual of $\underset{\sim}{\mathcal{O}}(\mathbb{D}^n)$ is now characterized in terms of this factor space as follows.

Proposition 2.4: The topological dual of $\underset{\sim}{\mathcal{O}}(\mathbb{D}^n)$ and the factor space (2.9) are isomorphic:

$$\mathcal{R}(\mathbb{D}^n) \simeq \underset{\sim}{\mathcal{O}}(\mathbb{D}^n)'.$$

This isomorphism and its inverse are given explicitly by the following formulas: For $\mu \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^n)'$, define a function $\hat{\mu}$ on W by

$$\hat{\mu}(z) = \mu(h_z), \quad h_z(t) = \prod_{j=1}^n \frac{e^{-(t_j - z_j)^2}}{2\pi i(t_j - z_j)}, \quad z \in W; \tag{2.11}$$

then $\hat{\mu} \in \widetilde{\mathcal{O}}(W)$ and thus $[\hat{\mu}] \in \mathcal{R}$. Conversely every equivalence class $[F] \in \mathcal{R}$ defines an element $\mu_{[F]} \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^n)'$ by

$$\mu_{[F]}(f) = \int_{\Gamma_1} \cdots \int_{\Gamma_n} F(z_1, \ldots, z_n) f(z_1, \ldots, z_n) \times dz_1 \cdots dz_n \equiv \int_{\Gamma_1 \cdots \Gamma_n} F(z) f(z) dz, \tag{2.12}$$

where $F \in \widetilde{\mathcal{O}}(W)$ is any representative of $[F]$ and where the paths $\Gamma_1, \ldots, \Gamma_n$ are chosen according to $f \in \underset{\sim}{\mathcal{O}}_c^m(U_m)$ for some m such that

$$\Gamma_1 \times \cdots \Gamma_n \subset U_m \cap W \cap \mathbb{C}^n, \text{ for instance,}$$

$$\Gamma_j = \Gamma_j^+ + \Gamma_j^-,$$

$$\Gamma_j^\pm = \{z_j | z_j = \pm x_j \pm i\delta_m, -\infty < x_j < \infty\}$$

with sufficiently small $\delta_m > 0$.

Proof[4]*:* The first part clearly relies on properties of the collection of functions h_z, $z \in W$. Those that are relevant here are contained in the following elementary lemma.

Lemma 2.5: (a) For every $z \in W \cap \mathbb{C}^n$, there is $m_0 = m_0(z)$ such that h_z belongs to $\mathscr{O}_c^m(U_m)$ and

$$\|h_z\|_m \leqslant \text{const}\, e^{|z|/m}/\delta(z), \quad \delta(z) = \text{dist}(z, \mathbb{R}^n),$$

for all $m \geqslant m_0$.

(b) For every $z^0 \in W \cap \mathbb{C}^n$ there is a polycircle

$$\mathscr{P} = \{z = (z_1,\ldots, z_n) \mid |z_j - z_j^0| < r_j, \quad j = 1,\ldots, n\}$$

around z^0 such that $\overline{\mathscr{P}} \subset W$ and there are functions $\Delta_j: \mathscr{P} \to \underset{\sim}{\mathscr{O}}(\mathbb{D}^n), j = 1,\ldots, n$ such that, for all $z \in \mathscr{P}$,

$$h_z - h_{z^0} = \sum_{j=1}^n (z_j - z_j^0)\Delta_j(z),$$

with

$$\Delta_j(z) \to \Delta_j(z^0) \quad \text{in} \quad \underset{\sim}{\mathscr{O}}(\mathbb{D}^n) \quad \text{for } z \to z^0.$$

Now take any $\mu \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$. Part (a) of the lemma implies immediately that $z \to \mu(h_z)$ is a well defined function on $W \cap \mathbb{C}^n$ and, according to part (b), this function has complex derivatives; hence $\hat{\mu}$ is analytic on $W \cap \mathbb{C}^n$.

Suppose a compact subset $K \subset W$ and a number $\epsilon > 0$ to be given. Then $\delta = \text{dist}(K \cap \mathbb{C}^n, \mathbb{R}^n) > 0$. There is $m_0 = m_0(K)$ such that $\epsilon > 1/m_0$ and $\text{dist}(U_m \cap \mathbb{C}^n, K \cap \mathbb{C}^n) \leqslant \delta/2$, for all $m \geqslant m_0$. Then, for fixed $m \geqslant m_0$, the collection of functions h_z, $z \in K \cap \mathbb{C}^n$, belongs to $\mathscr{O}_c^m(U_m)$ and the estimate of part (a) yields

$$\|\hat{\mu}\|_{K,\epsilon} \leqslant C_m \sup_{z \in K \cap \mathbb{C}^n} \|h_z\| e^{-\epsilon|z|} < \infty;$$

hence $\hat{\mu}$ is slowly increasing and therefore $\hat{\mu} \in \widetilde{\mathscr{O}}(W)$.

The growth restriction (2.7) for a function $F \in \widetilde{\mathscr{O}}(W)$ implies that the integral in Eq. (2.12) is well defined for all $f \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$; more precisely, for every $m \in \mathbb{N}$, there is $C_m = C_m(F)$ such that, for all $f \in \mathscr{O}_c^m(U_m)$, this integral is bounded in absolute value by

$$C_m \|f\|_m.$$

A function $F_k \in \widetilde{\mathscr{O}}(\widehat{W}_k)$ is, in particular, analytic in $z_k \in \mathbb{C}$. Hence if we rewrite the integral in (2.12) in the form

$$\int_{\Gamma_1} \cdots \left(\int_{\Gamma_k} F_k(z_1,\ldots, z_k,\ldots, z_n) f(z_1,\ldots, z_k,\ldots, z_n)\, dz_k \right) dz_1 \cdots dz_{k-1}\, dz_{k+1} \cdots dz_n,$$

we see that this integral vanishes according to Cauchy's theorem and the growth restriction on F and f.

Therefore all elements F' in the equivalence class $[F]$ of $F \in \widetilde{\mathscr{O}}(W)$ define the same continuous linear functional on $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$, that is, by (2.12), $\mathscr{R}(\mathbb{D}^n)$ is mapped linearly into $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$.

Another application of Cauchy's theorem together with the growth restrictions on F and f shows, by appropriate choice of the integration path: If $\mu_{[F]}(f) = 0$, for all $f \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$, then

$$F \in \sum_{k=1}^n \widetilde{\mathscr{O}}(\widehat{W}_k),$$

i.e., $[F] = 0$. Hence the mapping (2.12) $\mathscr{R}(\mathbb{D}^n) \to \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ is injective.

Since $h_z(\cdot)$ is a modified Cauchy kernel with appropriate decay properties, one knows, for all $f \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ in a suitable complex neighborhood of $\mathbb{D}^n$,

$$\int_{\Gamma_1 \times \cdots \times \Gamma_m} f(z) h_z\, dz = f(\cdot).$$

If $\mu \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ is applied to this equation one deduces

$$\mu_{[\hat{\mu}]}(f) = \mu(f);$$

hence the mapping (2.12) is an inverse of the mapping (2.11) and the proposition follows.

Via the isomorphism of Proposition 2.4 the heuristic definition of a Fourier hyperfunction as a finite sum of boundary values of slowly increasing holomorphic functions is easily given a precise meaning: The 2^n connected components of W can be described as

$$W(\alpha_1,\ldots, \alpha_n) = \{z \in \mathbb{Q}^n | \alpha_j \operatorname{Im} z_j > 0, \quad j = 1,\ldots, n\},$$
$$\alpha_j \in \{1, -1\}.$$

Now define, for $\alpha = (\alpha_1,\ldots, \alpha_n) \in \{1, -1\}^n$,

$$\sigma(\alpha) = \prod_{j=1}^n \alpha_j,$$

and then, for $F \in \widetilde{\mathscr{O}}(W)$,

$$F_\alpha = \sigma(\alpha) F \text{ on } W(\alpha) \text{ and } F_\alpha = 0 \text{ elsewhere.}$$

The *boundary value* of $F \in \widetilde{\mathscr{O}}(W)$ with respect to the cone $W(\alpha)$ is then defined by

$$\delta_{W(\alpha)}(F) = [F_\alpha]. \tag{2.13}$$

Clearly it follows that

$$[F] = \sum_{\alpha \in \{1, -1\}^n} \sigma(\alpha) \delta_{W(\alpha)}(F), \tag{2.14}$$

and hence, by Proposition 2.4, μ is the corresponding finite sum of boundary values of $\hat{\mu} \in \widetilde{\mathscr{O}}(W)$.

Next we establish the traditional view of boundary values as limits [in $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$] of slowly increasing functions. To this end we define, for $0 < \delta_j < 1/m$, a path $\Gamma_j(\alpha_j, \delta_j)$, $\alpha_j \in \{1, -1\}$, in the complex z_j plane by

$$\Gamma_j(\alpha_j, \delta_j) = \{z = \alpha_j x_j + i\alpha_j \delta_j | -\infty < x_j < \infty\},$$

so that

$$\Gamma(\alpha, \delta) = \Gamma_1(\alpha_1, \delta_1) \times \cdots \times \Gamma_n(\alpha_n, \delta_n)$$
$$\subset U_m \cap W(\alpha).$$

Then, for all $f \in \mathscr{O}_c^m(U_m)$ and all $F \in \widetilde{\mathscr{O}}(W)$,

$$\int_{\Gamma(\alpha,\delta)} F(z) f(z)\, dz = I(\alpha, \delta)$$

is independent of δ, $0 < \delta_j < 1/m$, and thus equals the limit $\delta_j \to 0, j = 1,\ldots, n$, of this integral denoted by

$$\langle F_\alpha(x_1 + i\alpha_1 0,\ldots, x_n + i\alpha_n 0),$$
$$f(x_1 + i\alpha_1 0,\ldots, x_n + i\alpha_n 0)\rangle,$$

where the duality $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$, $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ is used. If we sum over all $\alpha \in \{1, -1\}^n$, we obtain

$$\sum_{\alpha\in\{1,-1\}^n} I(\alpha,\delta) = \int_{\Gamma(\delta)} F(z)f(z)dz.$$

Thus, by Proposition 2.4, every $\mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ is a finite sum of boundary values of slowly increasing functions.

D. Support of Fourier hyperfunctions

If $K \subset \mathbb{D}^n$ is any closed subset, then relations (2.3) and (2.4) easily imply that $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ is contained in $\underset{\sim}{\mathscr{O}}(K)$. For a Fourier hyperfunction μ on $\mathbb{D}^n$, denote by $C(\mu)$ the class of all closed subsets $K \subset \mathbb{D}^n$ such that there is a continuous extension μ_K of μ to $\underset{\sim}{\mathscr{O}}(K)$:

$$\mu_K\in\underset{\sim}{\mathscr{O}}(K)', \quad \mu_{K|\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)} = \mu. \tag{2.15}$$

Any such subset $K\in C(\mu)$ is called a "carrier" of the Fourier hyperfunction μ. In contrast to a general analytic functional a Fourier hyperfunction μ has a smallest carrier, called the *support* of μ:

$$\operatorname{supp}\mu = \bigcap_{K\in C(\mu)} K. \tag{2.16}$$

This definition really works since one can prove[6] the following proposition.

Proposition 2.6: If $K_1,K_2\in C(\mu)$, then $K_1 \cap K_2\in C(\mu)$.

This result is by no means trivial. We give some hints. Having $K_j\in C(\mu)$ means that there are $\mu_j = \mu_{K_j}\in\underset{\sim}{\mathscr{O}}(K_j)'$ satisfying (2.15). Also, we have to define an extension to $\underset{\sim}{\mathscr{O}}(K_1 \cap K_2)$. Given $f\in\underset{\sim}{\mathscr{O}}(K_1 \cap K_2)$ there exist by the Mittag–Leffler theorem for rapidly decreasing functions[6] $f_j \in\underset{\sim}{\mathscr{O}}(K_j)$ such that

$$f = f_1 - f_2 \quad \text{on} \quad K_1 \cap K_2. \tag{2.17}$$

Now define a function μ:

$$\mu(f) = \mu_1(f_1) - \mu_2(f_2). \tag{2.18}$$

The right-hand side of Eq. (2.18) is independent of the special choice of the decomposition (2.17). Hence μ is well defined, and obviously μ is linear. One can prove continuity of μ by some general arguments.[6]

According to definition (2.16) the topological dual of $\underset{\sim}{\mathscr{O}}(K)$, $K \subset \mathbb{D}^n$ closed, is the set of Fourier hyperfunctions on $\mathbb{D}^n$ with support contained in K. With this interpretation in mind the space of *Fourier hyperfunctions on an open subset* V of $\mathbb{D}^n$ is naturally defined as the factor space of the space of all Fourier hyperfunctions on $\mathbb{D}^n$ with respect to the subspace of those Fourier hyperfunctions having support in the complement $V^c = \mathbb{D}^n - V$ of V:

$$\mathscr{R}(V) = \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'/\underset{\sim}{\mathscr{O}}(V^c)'. \tag{2.19}$$

It is known from the following proposition[17] that $\mathscr{R}(V)$ is isomorphic to

$$\mathscr{R}(V) \simeq \underset{\sim}{\mathscr{O}}(\overline{V})'/\underset{\sim}{\mathscr{O}}(\overline{V} - V)'.$$

Proposition 2.7: Let $K = \cup_{i=1}^p K_i$ be the union of p compact sets in $\mathbb{D}^n$. Suppose $\mu\in\underset{\sim}{\mathscr{O}}(K)'$; then there are $\mu_i\in\underset{\sim}{\mathscr{O}}(K_i)'$ such that $\mu = \Sigma_{i=1}^p \mu_i$.

Proof: Since the mapping $\underset{\sim}{\mathscr{O}}(K) \to \Pi_{i=p}^p \underset{\sim}{\mathscr{O}}(K_i)$, namely, $f \to \{f_{|K_i}\}_{i=1}^p$, is injective and of closed range, the mapping $\Pi_{i=1}^p \underset{\sim}{\mathscr{O}}(K_i)' \to \underset{\sim}{\mathscr{O}}(K)'$, namely, $\{\mu_i\}_{i=1}^p \to \Sigma_{i=1}^p \mu_i$, is accordingly surjective.

If V and W are open subsets in $\mathbb{D}^n$ with W contained in V, then we have $\underset{\sim}{\mathscr{O}}(V^c)' \subset \underset{\sim}{\mathscr{O}}(W^c)'$ and thus we get a restriction map

$$\begin{aligned}\rho_{WV}\colon \mathscr{R}(V) &= \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'/\underset{\sim}{\mathscr{O}}(V^c)' \to \mathscr{R}(W)\\ &= \underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'/\underset{\sim}{\mathscr{O}}(W^c)'\end{aligned} \tag{2.20}$$

in a canonical way.

For our purposes it turns out to be important to have the following result.

Theorem 2.8: The assignment of the factor spaces $\mathscr{R}(V)$ with open subsets V of $\mathbb{D}^n$ according to (2.19) together with the canonical restriction maps $\rho_{WV}\colon \mathscr{R}(V) \to \mathscr{R}(W)$ for open subsets $W \subset V$ according to (2.20) is a flabby sheaf on $\mathbb{D}^n$, called the *sheaf* $\mathscr{R}$ *of Fourier hyperfunctions on* $\mathbb{D}^n$.

This means in particular that Fourier hyperfunctions have the following localization properties (L_1) and (L_2):

(L_1) if $V = \bigcup_\alpha V_\alpha$, $\quad V_\alpha \subset \mathbb{D}^n$ open, $\quad \mu\in\mathscr{R}(V)$,

then $\mu_{|V_\alpha} = 0$, for all V_α, implies $\mu = 0$; (2.21)

(L_2) if $V_\alpha \subset \mathbb{D}^n$ open and

if $\mu_\alpha\in\mathscr{R}(V_\alpha)$ satisfies $\mu_{\alpha|V_\alpha \cap V_\beta} = \mu_{\beta|V_\alpha \cap V_\beta}$

then there is $\mu\in\mathscr{R}(V)$, such that $\mu_{|V_\alpha} = \mu_\alpha$. (2.22)

Furthermore, since $\mathscr{R}$ is flabby, any Fourier hyperfunction μ_V on any open subset $V \subset \mathbb{D}^n$ is the restriction of a Fourier hyperfunction $\mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ to V: $\mu_V = \mu_{|V}$.

Because of the localization properties of a sheaf the notion of *support of a Fourier hyperfunction* $\mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ can also be defined as the smallest closed subset $K \subset \mathbb{D}^n$ such that $\mu_{|K^c} = 0$. From (2.19) and (2.20) it is obvious that this notion of support agrees with the notion introduced previously in (2.15) and (2.16).

Proof of Theorem 2.8: First we assume $V = \cup_{\alpha\in I} V_\alpha$, $\mu\in\mathscr{R}(V)$. Let $\bar\mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ be a representative of μ. Then

$$\rho_{V_\alpha \mathbb{D}^n}(\bar\mu) = \rho_{V_\alpha V}\cdot\rho_{V\mathbb{D}^n}(\bar\mu) = \rho_{V_\alpha V}(\mu) = 0$$

implies $\operatorname{supp}\bar\mu \cap V_\alpha = \varnothing$, for all $\alpha\in I$, and hence $\operatorname{supp}\bar\mu \cap V = \varnothing$, which implies $\mu = 0$. Thus (L_1) is proved.

To prove (L_2) we begin with the case of just two open sets V_1 and V_2. Let $\bar\mu_\alpha\in\underset{\sim}{\mathscr{O}}(\overline{V}_\alpha)'$ be representatives of μ_α, for $\alpha = 1,2$. The support of $\mu_1 - \mu_2$ is contained in

$$(\overline{V}_1\cup\overline{V}_2) - (\overline{V}_1\cap\overline{V}_2) = (\overline{V}_1^c\cap\overline{V}_2)\cup(\overline{V}_1\cap\overline{V}_2^c);$$

thus Proposition 2.7 gives a decomposition

$$\bar\mu_1 - \bar\mu_2 = \bar\nu_1 - \bar\nu_2,$$

$$\bar\nu_1\in\underset{\sim}{\mathscr{O}}(\overline{V}_1^c\cap\overline{V}_2)', \quad \bar\nu_2\in\underset{\sim}{\mathscr{O}}(\overline{V}_1\cap\overline{V}_2^c)'.$$

Let

$$\mu = \bar\mu_1 - \bar\nu_1 = \bar\mu_2 - \bar\nu_2\in\underset{\sim}{\mathscr{O}}(\overline{V}_1\cap\overline{V}_2)'.$$

Then we have $\mu_{|V_\alpha} = \mu_\alpha$ because $\operatorname{supp}(\mu - \mu_\alpha)\cap V_\alpha = \varnothing$.

In the general case (L_2) is proved by using some topological argument (see Theorem 4.17 of Ref. 6).

The existence of the representative $\bar\mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'$ of $\mu\in\mathscr{R}(V)$ implies the flabbiness of $\mathscr{R}$.

Remark 2.1: The restriction of a Fourier hyperfunction to $\mathbb{R}^n$ gives a hyperfunction. Since the sheaf $\mathscr{R}$ of a Fourier hyperfunction is flabby any hyperfunction on $\mathbb{R}^n$ can be extended to $\mathbb{D}^n$ as a Fourier hyperfunction. This extension, however, is not unique since there are Fourier hyperfunctions with support at "infinity" (support$\subset S_\infty^{n-1}$).[28]

Remark 2.2: For any $f\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ we have $f_{|\mathbb{R}^n}\in\mathscr{S}(\mathbb{R}^n)$ by definition, and this injection of $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$ into $\mathscr{S}(\mathbb{R}^n)$ is continuous. Hence Fourier hyperfunctions generalize tempered distributions. Thus for a tempered distribution we have defined the notion of support in the sense of Fourier hyperfunctions also. We now show that this notion of support agrees with that in the sense of distributions.

Suppose that the support of T in the sense of distributions is contained in some closed set $K\subset\mathbb{R}^n$. Let $\{U_m\}$ be a fundamental system of neighborhoods of $\overline{K}$ in $\mathbb{Q}^n$, the closure of K in $\mathbb{D}^n$; then for any U_m there exists a C^∞- function χ such that supp $\chi\subset U_m\cap\mathbb{R}^n$ and $\chi=1$ on K. Since $\chi\cdot f\in\mathscr{S}(\mathbb{R}^n)$ for $f\in\mathscr{O}_c^m(U_m)$ and $T\cdot\chi(\varnothing)=T(\varnothing)$, for a $\varnothing\in\mathscr{S}(\mathbb{R}^n)$, T defines an element of $\underset{\sim}{\mathscr{O}}(\overline{K})'$. Hence the support of T in the sense of Fourier hyperfunctions is contained in $\overline{K}$.

Now consider the tempered distribution T as a Fourier hyperfunction and suppose that this Fourier hyperfunction has its support in a closed set $\overline{K}$ of $\mathbb{D}^n$, that is, $T\in\underset{\sim}{\mathscr{O}}(\overline{K})'$. Let $\varnothing\in C_0^\infty(\mathbb{R}^n)$ with supp $\varnothing\subset K^c$; then $\varnothing\in\underset{\sim}{\mathscr{O}}(\overline{K})$ and $T(\varnothing)=0$. This shows that the support of T in the sense of distributions is contained in K.

Remark 2.3: Let $n=n_1+n_2$. Note that $\mathbb{D}^n\neq\mathbb{D}^{n_1}\times\mathbb{D}^{n_2}$, but

$$\mathbb{Q}^n\cap\mathbb{R}^n=\mathbb{R}^n=(\mathbb{Q}^{n_1}\times\mathbb{Q}^{n_2})\cap\mathbb{R}^n$$

and

$$\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)=\underset{\sim}{\mathscr{O}}(\mathbb{D}^{n_1}\times\mathbb{D}^{n_2}).$$

Let K_j $(j=1,2)$ be closed sets in $\mathbb{R}^{n_j}$, $\overline{K}_j$ be the closure of K_j in $\mathbb{D}^{n_j}$, and $\overline{K_1\times K_2}$ be the closure of $K_1\times K_2$ in $\mathbb{D}^n$. Then we have

$$\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)\subset\underset{\sim}{\mathscr{O}}(\overline{K_1\times K_2})\subset\underset{\sim}{\mathscr{O}}(\overline{K}_1\times\overline{K}_2).$$

Thus we have

$$\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'\supset\underset{\sim}{\mathscr{O}}(\overline{K_1\times K_2})'\supset\underset{\sim}{\mathscr{O}}(\overline{K}_1\times\overline{K}_2)',$$

i.e., the elements of $\underset{\sim}{\mathscr{O}}(\overline{K}_1\times\overline{K}_2)'$ can be considered to be Fourier hyperfunctions.

E. Fourier, Fourier–Laplace transformation, and edge of the wedge theorem for (Fourier) hyperfunctions

According to Proposition 2.3 the Fourier transformation is an isomorphism of $\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)$. Hence by duality we have the Fourier transform as an isomorphism for Fourier hyperfunctions:

$$(\mathscr{F}\mu)(f)=\mu(\mathscr{F}f),\quad \text{for all } f\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n),\quad \mu\in\underset{\sim}{\mathscr{O}}(\mathbb{D}^n)'. \tag{2.23}$$

As in distribution theory, if appropriate support properties are available, the Fourier transformation has an extension to complex arguments to yield the "Fourier–Laplace transformation."

Proposition 2.9 (Paley–Wiener theorem for Fourier hyperfunctions)*:* Let Γ be a closed and strictly convex cone in $\mathbb{R}^n$ with its vertex at the origin such that $\Gamma\subset\{x\in\mathbb{R}^n|x\cdot e>0\}\cup\{0\}$ for some unit vector $e\in\mathbb{R}^n$. Let $\mathring{\Gamma}=\{\xi\,|x\cdot\xi>0,\text{ for all } x\in\Gamma\}$ be the polar set of Γ and let μ be a Fourier hyperfunction on $\mathbb{D}^n$.

(a) If supp $\mu\subset\overline{\Gamma}$ = closure of Γ in $\mathbb{D}^n$, then the *Fourier–Laplace transform* of μ,

$$(\mathscr{L}\mu)(\zeta)=\mu(e_\zeta),\quad e_\zeta=(2\pi)^{-n/2}e^{i\zeta\cdot z}, \tag{2.24}$$

is well defined for $\zeta\in\mathbb{R}^n+i\mathring{\Gamma}$ and is a holomorphic function of its argument satisfying the following growth condition: For every relatively compact open cone $\Gamma_0\subset\mathring{\Gamma}$ and every $0<\epsilon<\epsilon_0(\Gamma_0)$ there is a constant $C=C(\epsilon,\Gamma_0)$ such that, for all $\zeta\in\mathbb{R}^n+i\Gamma_0$,

$$|(\mathscr{L}\mu)(\zeta)|\leqslant Ce^{\epsilon|\mathrm{Re}\,\zeta|+\chi_\epsilon(\mathrm{Im}\,\zeta)}, \tag{2.25}$$

where

$$\chi_\epsilon(\eta)=\sup_{x\in\Gamma-\epsilon e}\{\epsilon|x|-x\cdot\eta\}.$$

(b) Conversely if a holomorphic function F on $\mathbb{R}^n+i\mathring{\Gamma}$ satisfies the above growth condition then it is the Fourier–Laplace transform of a Fourier hyperfunction μ on $\mathbb{D}^n$ with support in $\overline{\Gamma}$.

This is proved in Ref. 17. And as for distributions there is an immediate connection with the Fourier transformation.

Corollary 2.10: If $\mu\in\underset{\sim}{\mathscr{O}}(\overline{\Gamma})'$, then the Fourier–Laplace transform $\mathscr{L}\mu$ is holomorphic in $\mathbb{R}^n+i\mathring{\Gamma}$ and its boundary value with respect to the open cone $\mathring{\Gamma}$ equals the Fourier transform of μ:

$$\delta_{\mathring{\Gamma}}(\mathscr{L}\mu)=\mathscr{F}\mu. \tag{2.26}$$

The proof is obvious from Proposition 2.9 and the definitions.

Finally we will use the *edge of the wedge theorem for hyperfunctions* as proved in Ref. 29 which generalizes Epstein's version of this result for distributions.

Proposition 2.11: Let Γ_1 and Γ_2 be two open convex cones in $\mathbb{R}^n$. For any open set U in $\mathbb{R}^n$ and its complex neighborhood V there exists a complex neighborhood W of U such that $W\subset V$ and the following holds: If the boundary values $\delta_{\Gamma_1}(F_1)$ and $\delta_{\Gamma_2}(F_2)$ of two functions F_j holomorphic in $V\cap T(\Gamma_j)$, $T(\Gamma_j)=\mathbb{R}^n+i\Gamma_j$, $j=1,2$, agree on U in the sense of hyperfunctions then there exists a function F holomorphic in $W\cap T(\mathrm{ch}(\Gamma_1\cup\Gamma_2))$ such that

$$F=F_j\quad\text{on}\quad W\cap T(\Gamma_j),\quad j=1,2,$$

where ch A denotes the convex hull of a subset A in $\mathbb{R}^n$.

An immediate consequence is the following corollary.

Corollary 3.11: Let Γ be some open convex cone in $\mathbb{R}^n$ and F some holomorphic function on the tube $T(\Gamma)=\mathbb{R}^n+i\Gamma$. If the boundary value $\delta_\Gamma(F)$ of F in the sense of hyperfunctions vanishes in some open nonempty subset $U\subset\mathbb{R}^n$ then the function F itself vanishes.

III. QFT IN TERMS OF FOURIER HYPERFUNCTIONS

A. The test-function space

A QFT over a test-function space

$$E=\underset{\sim}{\mathscr{O}}(\mathbb{D}^4,V),\quad \dim V<\infty, \tag{3.1}$$

as introduced in Sec. II, is called a *Fourier hyperfunction quantum field theory* (FHQFT). If a QFT is formulated over a test-function space F such that the space $\underset{\sim}{\mathscr{O}}(\mathbb{D}^4,V)$ is densely and continuously embedded into F we call such a theory a *special FHQFT* since then continuous linear functionals on F are special Fourier hyperfunctions.

However, according to Sec. I, a function space E has to meet several requirements in order to be "admissible" as a test-function space of a QFT. So we show here that $E = \underset{\sim}{\mathcal{O}}(\mathbb{D}^4, V)$ indeed is "admissible." For convenience we do this explicitly only for $E = \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$, e.g., dim $V = 1$.

Since

$$\underset{\sim}{\mathcal{O}}(\mathbb{D}^4) = \operatorname*{ind\,lim}_{m\to\infty} \underset{\sim}{\mathcal{O}}{}_c^m(U_m)$$

with a fundamental sequence of neighborhoods U_m of $\mathbb{D}^4$ in $\mathbb{Q}^4 = \mathbb{D}^4 + i\mathbb{R}^4$ such that U^m is invariant under complex conjugation $z \to \bar{z}$, a continuous involution $f \to f^*$ on $\underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ is well defined by

$$f^*(z) = \overline{f(\bar{z})}, \tag{3.2}$$

since on each $\underset{\sim}{\mathcal{O}}{}_c^m(U_m)$ we have $\|f^*\|_m = \|f\|_m$.

The action of the group $G = \text{iSL}(2,\mathbb{C})$ on $\underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ is defined as usual by

$$(\alpha_{(a,A)} f)(z) = f(\Lambda(a)^{-1}(z-a)), \tag{3.3}$$

where $a \in \mathbb{R}^4$, $A \in \text{SL}(2,\mathbb{C})$, $\Lambda(A) \in L^{\uparrow}_{+}$, and thus

$$\alpha_g(f^*) = \alpha_g(f)^* \tag{3.4}$$

follows easily. The differentiability of the map $G \to \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$,

$$g \to \alpha_g(f),$$

for each fixed $f \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$, is proved in Sec. V where it is actually used to prove the cluster property.

According to Sec. II the function space $\underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ admits the Fourier transformation as an isomorphism. Furthermore, for elements in $\underset{\sim}{\mathcal{O}}(\mathbb{D}^{4n})'$, $n = 1,2,\ldots$, a "good" notion of support is available, expressing the intuitive meaning of support in this mathematical frame. Hence we get an adequate formulation of the locality condition (H_4) if the notion of support is understood in the sense of Fourier hyperfunctions:

$$\text{Supp}_{\text{HF}} \langle A_\alpha, A_\beta \rangle \subset \overline{K} = \text{closure of } K \text{ in } \mathbb{D}^8. \tag{3.5}$$

With these specifications of the test-function space a *scalar relativistic quantum field in terms of Fourier hyperfunctions* is a field over

$$E = \{\underset{\sim}{\mathcal{O}}(\mathbb{D}^4), {}^*, \alpha_g, \quad g \in G \equiv \text{iSL}(2,C)\}$$

satisfying (H_1)–(H_5).

B. HFQFT in terms of its n-point functions

In this subsection we briefly recall the description of a field in terms of the sequence of its n-point functions[1] and indicate, where necessary, the differences with respect to the "standard" approach as a result of the particularities of the test-function space $E = \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$.

Given a scalar field $\tilde{A}$ over $E = \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ satisfying (H_0) and (H_1) we can consider the sequence of separately continuous n-linear functionals on $E^n = E \times \cdots \times E$ (n times) defined by

$$(f_1, \ldots f_n) \to (\Phi_0, A(f_1) \cdots A(f_n)\Phi_0),$$

where Φ_0 denotes the cyclic unit vector. By Proposition 2.2 these functionals uniquely determine Fourier hyperfunctions $\mathscr{W}_n = \mathscr{W}_n^A \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^{4n})'$ such that

$$\mathscr{W}_n(f_1 \otimes \cdots \otimes f_n) = (\Phi_0, A(f_1) \cdots A(f_n)\Phi_0), \tag{3.6}$$

for all $f_j \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ and all $n = 1,2,\ldots$.

The sequence

$$\mathscr{W} = \mathscr{W}^A = \{1, \mathscr{W}_1^A, \mathscr{W}_2^A, \mathscr{W}_3^A, \ldots\} \tag{3.7}$$

of these *n-point functions* $\mathscr{W}_n^A$ of the field A is a *state* on the complete tensor algebra,

$$\underline{E} = \bigoplus_{n=0}^{\infty} E(n) \quad \text{(locally convex direct sum)},$$

$$E(0) = C, \quad E(n) = \underset{\sim}{\mathcal{O}}(\mathbb{D}^{4n}) = \hat{\otimes}^n \underset{\sim}{\mathcal{O}}(\mathbb{D}^4), \quad n \geqslant 1, \tag{3.8}$$

that is a continuous linear functional on $\underline{E}$, which is normalized according to

$$\mathscr{W}(\underline{1}) = 1, \quad \underline{1} = \{1,0,0,\ldots\}$$

and non-negative according to

$$\mathscr{W}(\underline{f}^* \cdot \underline{f}) \geqslant 0, \quad \text{for all } \underline{f} \in \underline{E}, \tag{3.9}$$

where the involution $*$ on $\underline{E}$ is given by canonical extension of the involution on E and the product is the usual product of tensor algebras. Conversely according to the well known reconstruction theorem[1–3] such a state $\mathscr{W}$ on $\underline{E}$ determines uniquely up to unitary equivalence a field A over $E = \underset{\sim}{\mathcal{O}}(\mathbb{D}^4)$ satisfying (H_1) and (3.6).

If the field A is covariant in the sense of condition (H_2) then the associated state $\mathscr{W} = \mathscr{W}^A$ on $\underline{E}$ is easily seen to be invariant under the action

$$\underline{\alpha}_g = \bigoplus_{n=0}^{\infty} \alpha_g^{\otimes n}, \quad \alpha_g^{\otimes 0} = 1, \tag{3.10}$$

of $G = \text{iSL}(2,\mathbb{C})$ on $\underline{E}$.

Conversely if a state $\mathscr{W}$ on $\underline{E}$ is invariant under the action (3.10) of G it determines as above a field A over E and a continuous unitary representation U of G satisfying (H_1) and (H_2).

Next we translate the locality condition (H_4) into properties of the n-point functions $\mathscr{W}_n$. This condition says

$$\text{supp}_{\text{HF}}(\Phi, [A(\cdot), A(\cdot)]_\sigma \psi) \subset \overline{K}, \quad \text{for all } \Phi, \psi \in \mathscr{D}_0,$$

where $[A(f), A(g)]_\sigma = A(f)A(g) - \sigma A(g)A(f)$, for all $f, g \in E$, $\sigma = (-1)^\alpha$, $\alpha = 0$ or 1, that is, we assume $A = A_\alpha$ in (H_4). Hence if we introduce, for $0 \leqslant j \leqslant n$ and $n = 0,1,2,\ldots$, Fourier hyperfunctions $\mathscr{W}_{n,j} \in \underset{\sim}{\mathcal{O}}(\mathbb{D}^{4(n+2)})'$ by

$$\begin{aligned} \mathscr{W}_{nj}&(x_1,\ldots,x_j,x,y,x_{j+1},\ldots,x_n) \\ &= \mathscr{W}_{n+2}(x_1,\ldots,x_j,x,y,x_{j+1},\ldots,x_n) \\ &\quad - \sigma \mathscr{W}_{n+2}(x_1,\ldots,x_j,y,x,x_{j+1},\ldots,x_n), \end{aligned} \tag{3.11}$$

we easily see that by definition of $\mathscr{D}_0$ the locality condition (H_4) is equivalent to

$$\text{supp}_{\text{HF}}\ \mathscr{W}_{n,j} \subset \overline{K}_{n,j} \tag{3.12}$$

or $\mathscr{W}_{n,j} \in \underset{\sim}{\mathcal{O}}(\overline{K}_{n,j})'$, for all $0 \leqslant j \leqslant n$ and all $n = 0,1,2,\ldots$, where $\overline{K}_{n,j}$ is the closure of $K_{n,j} = \mathbb{R}^{4j} \times K \times \mathbb{R}^{4(n-j)}$ in $\mathbb{D}^{4(n+2)}$.

In Sec. V the "cluster property" is proved to be equivalent to condition (H_5) (uniqueness of the vacuum state Φ_0).

Thus we are left with expressing the spectral condition (H_3) in terms of properties of the n-point functions. However, in HFQFT this is considerably more complicated than for the tempered field since also in energy-momentum space there are no test functions of compact support. Hence this point needs some additional arguments.

Suppose that (H_0)–(H_2) are satisfied. Then according to Proposition 2.2 there are continuous linear maps

$$\Phi_n: E(n) \to \mathscr{H}, \quad E(n) = \underset{\sim}{\mathscr{O}}(\mathbb{D}^{4n})$$

satisfying, for all $f_j \in E(1) = \underset{\sim}{\mathscr{O}}(\mathbb{D}^4)$,

$$\Phi_n(f_1 \otimes \cdots \otimes f_n) = A(f_1) \cdots A(f_n)\Phi_0, \quad n = 1,2,\ldots . \tag{3.13}$$

The covariance of the field under $G = \mathrm{iSL}(2,\mathbb{C})$ implies in particular the following transformation law for these maps Φ_n under translations:

$$\begin{aligned} &U(a)\Phi_n(f_n) = \Phi_n(f_{n,a}), \\ &f_{n,a}(x_1,\ldots,x_n) = f_n(x_1 - a,\ldots,x_n - a). \end{aligned} \tag{3.14}$$

The consequences of these transformation properties on the Fourier hyperfunctions Φ_n with values in the Hilbert space $\mathscr{H}$ are most conveniently analyzed if in its Fourier transform $\widetilde{\Phi}_n$ the following variables are introduced:

$$\begin{aligned} &(q_1,\ldots,q_n) = \chi_n^{-1}(p_1,\ldots,p_n), \quad q_k = \sum_{j=k}^{n} p_j, \\ &\widetilde{Z}_n = \Phi_n \cdot \chi_n . \end{aligned} \tag{3.15}$$

Denote by P the generator of the translations $U(a)$, i.e.,

$$U(a) = e^{iaP} = \int e^{iak} E(dk). \tag{3.16}$$

The spectrum $\Sigma = \sigma(P)$ of the operator P is given by the support of the projection-valued measure E:

$$\Sigma = \sigma(P) = \operatorname{supp} E. \tag{3.17}$$

For any continuous bounded function h we know

$$h(P) = \int h(k)E(dk) = \int da\, \tilde{h}(a)\, U(a) \tag{3.18}$$

to be a bounded operator.

The transformation property (3.14) can now be expressed in the following way: For every $f \in E(1) \equiv \underset{\sim}{\mathscr{O}}(\mathbb{D}^4)$, every $g \in E(n-1) = \underset{\sim}{\mathscr{O}}(\mathbb{D}^{4(n-1)})$, and every function $\tilde{h}$ in the multiplicator space of $\tilde{E}(1)$, one has

$$h(P)\widetilde{Z}_n(f \otimes g) = \widetilde{Z}_n(h \cdot f \otimes g). \tag{3.19}$$

This equation can be used to extend $\widetilde{Z}_n$ in its first argument f. For every $m \in \mathbb{N}$, define functions ρ_m and ψ_m by

$$\rho_m(q) = \prod_{i=0}^{3} \cosh\left(\frac{q_i}{m}\right) \quad \text{and} \quad \psi_m(q) = \rho_m(q)^{-1}. \tag{3.20}$$

It follows, for $m = 1,2,\ldots$,

$$\psi_m \in \underset{\sim}{\mathscr{O}}(\mathbb{D}^4), \quad \text{and} \quad |\rho_m(q)| \leqslant Ce^{|q|/m}.$$

So we can rewrite Eq. (3.19) as

$$\widetilde{Z}_n(f \otimes g) = (\rho_m \cdot f)(P)\widetilde{Z}_n(\psi_m \otimes g), \tag{3.21}$$

and thus $f \to \widetilde{Z}_n(f \otimes g)$ can be extended to all those f for which $(\rho_m \cdot f)(P)$ is a bounded operator on $\mathscr{H}$, i.e., for which

$$\sup_{q \in \Sigma} |\rho_m(q) f(q)| \leqslant C\, |f|_{m,\Sigma} \equiv C \sup_{q \in \Sigma} e^{|q|/m} |f(q)| < \infty$$

is finite. This is in particular the case for all continuous functions f of compact support K in $\mathbb{R}^4$:

$$|f|_{m,\Sigma} \leqslant C_{K \cap \Sigma} |f|_{\infty, K \cap \Sigma} = C_{K \cap \Sigma} \sup_{q \in K \cap \Sigma} |f(q)| < \infty .$$

Hence we have proved the first part of the following proposition.

Proposition 3.1: (a) The vector-valued Fourier hyperfunctions $\widetilde{Z}_n$ of Eq. (3.15) can be extended to continuous linear maps

$$C_0(\mathbb{R}^4) \times \mathscr{O}(\mathbb{D}^{4(n-1)}) \to \mathscr{H},$$

that is, $\widetilde{Z}_n$ is a Radon measure in q_1 and a Fourier hyperfunction in

$$q_2,\ldots,q_n, \quad n = 2,3,\ldots .$$

(b) For every $g \in E(n-1)$ the measure $h \to \widetilde{Z}_n(h \otimes g)$ is slowly increasing and has its support $\Sigma_n(g)$ contained in Σ.

(c) The energy-momentum spectrum Σ of the theory is given by

$$\Sigma = \mathrm{cl}\Big(\{0\} \cup \bigcup_{n=1}^{\infty} \Sigma_n\Big), \quad \Sigma_n = \bigcup_{g \in E(n-1)} \Sigma_n(g), \tag{3.22}$$

where $\mathrm{cl}(A)$ denotes the closure of A in $\mathbb{R}^n$.

To complete the proof note that by Eq. (3.21) $Z_n(h \otimes g)$ extends to all functions

$$h \in F = \operatorname*{ind\,lim}_{m \to \infty} F_m,$$

where F_m is the Banach space of continuous functions on $\mathbb{R}^n$ such that $|h|_{m,\Sigma}$ is finite. Hence this measure is slowly increasing and has its support Σ_n contained in Σ.

Finally part (c) follows from the fact that

$$\{\Phi_0\}, \ \{\widetilde{Z}_n(f \otimes g) \mid f \in E(1), \ g \in E(n-1), \ n = 1,2,\ldots\}$$

is a total set of vectors in the representation space $\mathscr{H}$ of the unitary representation U.

The connection of the vector-valued Fourier hyperfunctions $\widetilde{Z}_n$ with the n-point functions of the theory is described by the following proposition.

Proposition 3.3: (a) Define Fourier hyperfunctions $\widetilde{W}_{n-1}$, $n = 2,3,\ldots$, by

$$\begin{aligned} &\widetilde{W}_{n-1}(f) = (\Phi_0, \widetilde{Z}_n(\psi_m \otimes f)), \\ &f \in E(n-1) = \mathscr{O}(\mathbb{D}^{4(n-1)}), \end{aligned} \tag{3.23}$$

where $m \in \mathbb{N}$ is arbitrary; then the Fourier transform $\widetilde{\mathscr{W}}_n$ of the n-point function $\mathscr{W}_n$ satisfies

$$\widetilde{\mathscr{W}}_n \cdot \chi_n(q_1,\ldots,q_n) = \delta(q_1)\widetilde{W}_{n-1}(q_2,\ldots,q_n). \tag{3.24}$$

(b) These Fourier hyperfunctions $\widetilde{W}_{n-1}$ allow the following decompositions:

$$\begin{aligned} &(\widetilde{Z}_j(f_j \otimes \cdots \otimes f_1), \widetilde{Z}_{j+1}(f_{j+1} \otimes \cdots \otimes f_n)) \\ &\quad = \widetilde{W}_{n-1}(\bar{f}_1 \otimes \cdots \otimes \bar{f}_{j-1} \otimes \bar{f}_j \cdot f_{j+1} \otimes f_{j+2} \otimes \cdots \otimes f_n), \end{aligned} \tag{3.25}$$

for all $f_i \in \mathscr{O}(\mathbb{D}^4)$, $1 \leqslant j \leqslant n-1$, $n = 2,3,\ldots$.

Proof: (a) Define $\widetilde{W}_{n-1}$ by Eq. (3.23) with $m = 1$. Then for arbitrary $m \in \mathbb{N}$ we use Eq. (3.21) to get

$$\widetilde{Z}_n(\psi_1 \otimes f) = (\rho_m \cdot \psi_1)(P)\widetilde{Z}_n(\psi_m \otimes f),$$

and thus, since the cyclic unit vector Φ_0 is translation invariant and

$(\rho_m \psi_1)(0) = 1,$

$$(\Phi_0, \widetilde{Z}_n(\psi_1 \otimes f)) = (\rho_m \psi_1)(0)(\Phi_0, \widetilde{Z}_n(\psi_m \otimes f)) = (\Phi_0, \widetilde{Z}_n(\psi_m \otimes f)).$$

Hence the definition (3.23) is independent of $m \in \mathbb{N}$.

Similarly we have, for all $f_i \in \mathscr{O}(\mathbb{D}^4)$, according to Eqs. (3.6), (3.15), and (3.21),

$$\begin{aligned} \mathscr{W}_n \cdot \chi_n(f_1 \otimes \cdots \otimes f_n) &= (\Phi_0, \widetilde{Z}_n(f_1 \otimes \cdots \otimes f_n)) \\ &= (\rho_m \cdot f_1)(0)(\Phi_0, \widetilde{Z}_n(\psi_m \otimes \cdots \otimes f_n)) \\ &= f_1(0)\widetilde{W}_{n-1}(f_2 \otimes \cdots \otimes f_n), \end{aligned}$$

and this proves Eq. (3.24). Finally part (b) follows by straightforward calculations directly from the definitions.

Remark 3.1: Together with Proposition 3.1, Eq. (3.25) says that the Fourier hyperfunctions $\widetilde{W}_{n-1}$ can always be considered in one of its variables as slowly increasing Radon measure with support in Σ. In particular we have, for all $h \in E(1)$, $g \in E(n-1)$, $n = 1,2,\ldots,$

$$\widetilde{W}_{2n-1}(g^* \otimes \bar{h} \cdot h \otimes g) = (\widetilde{Z}_n(h \otimes g), \widetilde{Z}_n(h \otimes g)), \quad (3.26)$$

exhibiting positivity properties of these measures.

Finally we derive support properties of the Fourier hyperfunctions $\widetilde{W}_{n-1}$, $n = 2,3,\ldots$, in all variables.

Denote by $\overline{\Sigma}$ the closure of Σ in $\mathbb{D}^4$ and introduce for $j = 1,\ldots,n-1$ the closed set

$$U_j = \{(q_1,\ldots,q_{n-1}) \in [\mathbb{D}^4]^{(n-1)} | q_j \in \overline{\Sigma}\}.$$

Then Eq. (3.25) and Proposition 3.1 imply

$$\operatorname{supp} \widetilde{W}_{n-1} \subset \overline{U}_j, \quad (3.27)$$

that is,

$$\widetilde{W}_{n-1|V_j} = 0, \quad (3.27')$$

for $V_j = \overline{U}_j^c = [\mathbb{D}^4]^{(n-1)} - \overline{U}_j$ and $j = 1,2,\ldots,n-1$.

The localization property (L_1) of the sheaf of Fourier hyperfunctions on energy-momentum space implies

$$\widetilde{W}_{n-1|\cup_{j=1}^{n-1} V_j} = 0$$

or

$$\operatorname{supp} \widetilde{W}_{n-1} \subset \left(\bigcup_{j=1}^{n-1} V_j\right)^c = \bigcap_{j=1}^{n-1} \overline{U}_j = \overline{\Sigma}^{n-1}. \quad (3.28)$$

Here we consider that $\widetilde{W}_{n-1}$ is defined on $[\mathbb{D}^4]^{n-1}$. Since $\mathscr{O}(\overline{\Sigma}^{n-1}) \supset \mathscr{O}(\overline{\Sigma^{n-1}})$, we have $\operatorname{supp} \widetilde{W}_{n-1} \subset \overline{\Sigma^{n-1}}$, if we consider that it is defined on $\mathbb{D}^{4(n-1)}$ (see Remark 2.3). By Proposition 3.2 this proves the following corollary.

Corollary 3.3: In a relativistic quantum field theory over $E = \mathscr{O}(\mathbb{D}^4)$ [only (H_1)–(H_3) have to be assumed] the Fourier transform $\widetilde{\mathscr{W}}_n$ of the n-point function $\mathscr{W}_n$ has its support contained in the closure of

$$\left\{(p_1,\ldots,p_n) \in \mathbb{R}^{4n} \,\middle|\, \sum_{j=1}^n p_j = 0, \left(\sum_{j=2}^n p_j, \sum_{j=3}^n p_j, \ldots, p_{n-1} + p_n, p_n\right) \in \Sigma^{(n-1)}\right\}$$

in $\mathbb{D}^{4n}$.

Therefore also in HFQFT a field A can be characterized in the usual way in terms of its n-point functions $\mathscr{W}_n = \mathscr{W}_n^A$ if the relevant support conditions (in coordinate and energy-momentum space) are interpreted in the sense of hyperfunctions. From (2.26) we have $\operatorname{supp} \widetilde{Z}_n \subset \overline{\Sigma}^n$ or $\operatorname{supp} \widetilde{Z}_n \subset \overline{\Sigma^n}$. The support properties of the Fourier transforms $\widetilde{\mathscr{W}}_n$ of the n-point function $\mathscr{W}_n$ together with the Paley–Wiener theorem for Fourier hyperfunctions (Proposition 2.8) allow us to derive the basic analyticity properties of the Wightman functions as easily as for tempered fields.[1-3]

Theorem 3.4: The n-point functions $\mathscr{W}_n$ of a relativistic quantum field over $E = \mathscr{O}(\mathbb{D}^4)$ [only (H_1)–(H_3) have to be assumed] are boundary values of $L^+(\mathbb{C})$-invariant holomorphic functions $\widehat{\mathscr{W}}_n$.

(a)
$$\widehat{\mathscr{W}}_{n+1}(z_0,z_1,\ldots,z_n) = \widehat{W}_n(z_1 - z_0, z_2 - z_1, \ldots, z_n - z_{n-1}), \quad (3.29)$$

where $\widehat{W}_n$ is holomorphic and $L^+(\mathbb{C})$ invariant on the extended tube

$$\mathscr{T}'_n = \bigcup_{A \in L^+(\mathbb{C})} A\mathscr{T}^+_{n'} \quad \text{and} \quad \mathscr{T}^+_n = T(V^n_+)$$

is the forward tube.

(b) The restriction of $\widehat{W}_n$ to $\mathscr{T}^+_n$ is the Fourier–Laplace transform of the Fourier hyperfunction $\widetilde{W}$ defined in Proposition 3.2. As an identity for Fourier hyperfunctions we have, for fixed $y_j \in V_+$,

$$\mathscr{W}_{n+1}(x_0,x_1,\ldots,x_1) = \lim_{\epsilon \to +0} \widehat{W}_n(x_1 - x_0 + i\epsilon y_1, \ldots, x_n - x_{n-1} + i\epsilon y_n). \quad (3.30)$$

C. Characterization of locality, existence of PCT operator, and global nature of local commutativity

The locality condition (3.11) and (3.12) says that the n-point functions

$$\mathscr{W}_{n+2}(x_1,\ldots,x,y,x_{j+1},\ldots,x_n)$$

and

$$-(-1)^{\alpha\beta}\mathscr{W}_{n+2}(x_1,\ldots,x,y,x_{j+1},\ldots,x_n)$$

agree as Fourier hyperfunctions in particular on the subset

$$I_{n+2} = \{(x_1,\ldots,x,y,x_{j+1},\ldots,x_n) \in \mathbb{R}^{4(n+2)} | (x_i - x_j)^2 < 0 \ (i \neq j), (x-y)^2 < 0, \ (x_j - x)^2 < 0, \ (x_j - y)^2 < 0, \ j = 1,\ldots,n\},$$

which is open in $\mathbb{R}^{4(n+2)}$.

By Jost's characterization of the real points of the extended tube $\mathscr{T}'_{n+1}$ (see Ref. 2) it follows from Theorem 3.4 that I_{n+2} consists of real points of analyticity of the associated Wightman functions $\mathscr{W}_{n+2}(x_1,\ldots,x,y,x_{j+1},\ldots,x_n)$ and $\mathscr{W}_{n+2}(x_1,\ldots,y,x,x_{j+1},\ldots,x_n)$.

However, if two analytic functions agree, in the sense of Fourier hyperfunctions, on an open set of real points of analyticity they do so as analytic functions (Corollary 2.11). This implies now that we can argue as in the case of tempered fields and arrive at the following theorem.

Theorem 3.5: Consider the Wightman functions $\widehat{\mathscr{W}}_n$ of a relativistic quantum field over $E=\tilde{\mathscr{O}}(\mathbb{D}^4)$ [satisfying only (H_1)–(H_3)] as given by Theorem 3.4. Then the locality condition (H_4) holds if and only if the $\mathscr{W}_n$ are analytic in

$$S_n^\pi=\{(z_1,...,z_n)|(z_{\pi(2)}-z_{\pi(1)},...,z_{\pi(n)}-z_{\pi(n-1)})\in\mathscr{T}'_{n-1},\quad \text{for some permutation } \pi \text{ of } (1,...,n)\},$$

and are permutation symmetric there:

$$\widehat{W}_n(z_1,...,z_n)=\widehat{\mathscr{W}}_n(z_{\pi(1)},...,z_{\pi(n)}),\quad n=2,3,... .$$

Remark 3.2: Without giving further details it should be clear from the above discussion on analyticity results that the PCT theorem[1–3] continues to hold in HFQFT.

Later we will have to use the following technical result that relies in an essential way on the analyticity properties of the Wightman functions.

Proposition 3.6: If $A=(A_1,...,A_M)$ is a relativistic quantum field over $E=\tilde{\mathscr{O}}(\mathbb{D}^4,V)$, $\dim V=M$, then $A_{j_0}(f)\Phi_0=0$, for all $f\in\tilde{\mathscr{O}}(\mathbb{D}^4)$ and some $j_0\in\{1,...,M\}$, implies $A_{j_0}=0$ (Φ_0 denotes the cyclic vacuum vector).

Proof: Since A is supposed to be local, the components $\{A_j\}$— of A are local relative to each other, that is,

$$\operatorname{supp}[A_i(\cdot),A_j(\cdot)]_{\sigma_{ij}}\subset\overline{K},\quad i,j=1,...,M.$$

At all points $(x_1,...,x,x_{k+1},...,x_n)$ such that

$$(x_2-x_1,...,x-x_k,x_{k+1}-x,...,x_n-x_{n-1})=(\xi_1,...,\xi_n)$$

is a Jost point we have, by repeated application of the locality condition as an identity for Fourier hyperfunctions,

$$W_n(\xi_1,...,\xi_n)=(\Phi_0,A_{j_1}(x_1)\cdots A_{j_k}(x_k)A_{j_0}(x)A_{j_{k+1}}(x_{k+1})\cdots A_{j_n}(x_n)\Phi_0)$$
$$=\pm(\Phi_0,A_{j_1}(x_1)\cdots A_{j_n}(x_n)A_{j_0}(x)\Phi_0)=0.$$

Thus Theorem 3.4 implies that the Wightman functions $\widehat{W}_n\in\tilde{\mathscr{O}}(\mathscr{T}'_n)$ vanish on the open subset J_n of $\mathbb{R}^{4n}$. Hence by Corollary 2.11 $\widetilde{W}_n$ vanishes identically. Therefore again by Theorem 3.4 the boundary value

$$(\Phi_0,A_{j_1}(x_1)\cdots A_{j_k}(x_k)A_{j_0}(x)A_{j_{k+1}}(x_{k+1})\cdots A_{j_n}(x_n)\Phi_0)$$

vanishes identically. And this holds for all $j_i\in\{1,...,M\}$, all $1\leqslant k<n$, and all $n=1,2,...$.

Thus $A_{j_0}(f)$ vanishes for all $f\in\tilde{\mathscr{O}}(\mathbb{D}^4)$ on the minimal domain $\mathscr{D}_0$, and we are done, since by cyclicity of Φ_0 the minimal domain is dense in the Hilbert space.

Remark 3.3: The proof of the "global nature of local commutativity" (Chap. 4.1 of Ref. 1) for tempered fields relies on analyticity properties of the Wightman functions and on arguments about analytic completion for special tube domains. The basic analyticity properties are provided by Theorem 3.4. The proofs of Theorem 3.5, Proposition 3.6, and Theorem 6.1 show that also in HFQFT the appropriate tools are available to imitate the proof given for tempered fields. Hence we conclude the following.

Theorem 3.7: Let A be a relativistic quantum field over $E=\tilde{\mathscr{O}}(\mathbb{D}^4,V)$, $\dim V<\infty$, satisfying (H_0)–(H_5) but the locality condition (H_4) only in the weaker form

$$\operatorname{supp}\langle A_\alpha,A_\beta\rangle\subset\overline{M},$$

with some closed subset $M\subset\mathbb{R}^4\times\mathbb{R}^4$ satisfying

$$K\subset M\quad\text{and}\quad M^c\neq\varnothing.$$

Then A satisfies $\operatorname{supp}\langle A_\alpha,A_\beta\rangle\subset\overline{K}$, i.e., A satisfies the locality condition (H_4).

IV. CLUSTER PROPERTY

The proof of the cluster property as given by Jost and Hepp[30] applies whenever the minimal domain $\mathscr{D}_0$ of the field, spanned by

$$\{\Phi_0,A(f_{j_1})\cdots A(f_{j_n})\Phi_0|f_{j_i}\in E,\ n=1,2,...\},$$

is invariant under the infinitesimal generators of $G=\mathrm{iSL}(2,\mathbb{C})$ in the given representation U [see (H_2)]. By the definition of the minimal domain and the action of $U(g)$, $g\in G$, on it this follows immediately from the invariance of the underlying space E of test functions under the infinitesimal generators of the action α of G on E according to (H_0). We give an explicit proof of this latter invariance for the test-function space $E=\tilde{\mathscr{O}}(\mathbb{D}^4)$ of rapidly decreasing holomorphic functions on $\mathbb{D}^4$ and prepare it by a sequence of lemmas. The technical details are given only for the more complicated case of the subgroup $\mathrm{SL}(2,\mathbb{C})$ of G. The corresponding proof for the subgroup of translations is left as an exercise.

Let $t\to\Lambda_t$ be a function on $\mathbb{R}$ with values in the space of $n\times n$ matrices such that

$$\begin{aligned}&\text{(i)}\quad \Lambda_0=I=\text{identity matrix},\\&\text{(ii)}\quad \Lambda_t=I+t\Sigma+o(t),\quad \text{for } |t|\to0,\\&\qquad\text{with some } n\times n \text{ matrix } \Sigma,\\&\qquad\text{i.e., } t\to\Lambda_t \text{ is differentiable at } t=0.\end{aligned}\tag{4.1}$$

For $x\in\mathbb{R}^n$ we introduce

$$y_t=\Lambda_t x-x\quad\text{and}\quad y=\Sigma x\tag{4.2}$$

and get immediately, with some constant $C\in\mathbb{R}_+$,

$$|y_t-ty|=o(t)|x|,\quad |y_t|\leqslant|t|C|x|.\tag{4.3}$$

Now let U_m be a neighborhood of $\mathbb{D}^n$, e.g.,

$$U_m=\{x+iy\in\mathbb{Q}^n|\ |\mathrm{Im}\,y|<1/m\}.$$

Take some fixed $f \in \mathscr{O}_c^m = \mathscr{O}_c^m(U_m)$, and apply Taylor's theorem for fixed x and $|t| \to 0$:

$$f(\Lambda_t x) - f(x) = \sum_{j=1}^{n} y_t^j \partial_j f(x) + \sum_{j,k=1}^{n} y_t^j y_t^k (\partial_j \partial_k f)(x + \theta_t y_t), \tag{4.4}$$

where y_t^j is the jth component of y_t and $\theta_t = \theta_t(x)$ some real number between 0 and 1. The terms on the right-hand side of this equation will be controlled by some lemmas.

Lemma 4.1: For $f \in \mathscr{O}_c^m$ and $\beta = (\beta_1, ..., \beta_n) \in \mathbb{Z}_+^n$, we have $\partial_\beta f \in \mathscr{O}_c^{m'}$ for all $m' > m$.

Proof: For $m' > m$ there is $\delta > 0$ such that the polycircles

$$C(z) = C_1 \times \cdots \times C_n, \quad C_j = \{\zeta_j \mid |\zeta_j - z_j| = \delta\},$$
$$|\mathrm{Im}\, z| \leqslant 1/m',$$

are contained in $\{\zeta \mid |\mathrm{Im}\, \zeta| < 1/m\}$. Hence, for $\zeta \in C(z)$ and $m' > m$, we know

$$|z|/m' - |\zeta|/m \leqslant \delta/m'.$$

By Cauchy's integral formula, $\partial_\beta f(z)$ is easily dominated according to

$$|\partial_\beta f(z)| = (2\pi)^{-n} \left| \int_{C(z)} f(\zeta) \prod_{j=1}^{n} (\zeta_1 - z_j)^{-1-\beta} j d\zeta_j \right| \leqslant K_1 \|f\|_m e^{-|\zeta|/m},$$

and thus

$$\|\partial_\beta f\|_{m'} = \sup_{|\mathrm{Im}\, z| \leq 1/m'} e^{|z|/m'} |\partial_\beta f(z)| \leq K_1 \|f\|_m d e^{\delta/m'}.$$

Lemma 4.2: For fixed $g \in \mathscr{O}_c^m$ and all $m' > n.$,

$(1/t)\, y_t^j g$ converges for $|t| \to 0$ in $\mathscr{O}_c^{m'}$ to $y^j g$.

Proof: By (4.3), given $\epsilon > 0$, there is $\delta > 0$ such that, for all x and all $0 < |t| < \delta$,

$$|y_t - ty| \leqslant \epsilon |t|\, |x|.$$

Hence, for all $m' > m$ and all $0 < |t| < \delta$, we have, for $j = 1, ..., n$,

$$\|(1/t)\, y_t^j g - y^j g\|_{m'} = \sup_{|\mathrm{Im}\, x| \leqslant 1/m'} |(1/t) y_t^j - y^j| g(x)| e^{|x|/m'} \leqslant \sup_{|\mathrm{Im}\, x| \leqslant 1/m'} \epsilon |x|\, \|g\|_m e^{|x|(1/m' - 1/m)} = \epsilon \|g\|_m C_m.$$

This implies the statement of the lemma.

Lemma 4.3: For fixed $g \in \mathscr{O}_c^m$ and all $m' > m$,

$(1/t) y_t^j y_t^k g(\cdot + \theta_t(\cdot) y_t) \to 0$ in $\mathscr{O}_c^{m'}$ as $|t| \to 0$.

Proof: Choose $\delta = (1 - m/m')/C$, where the positive constant C is given by (4.3). Then we have by (4.3), for all $0 < |t| < \delta$,

$$|\theta_t(x) y_t| \leqslant |y_t| \leqslant |x|,$$

and thus

$$|x + \theta_t y_t| \geqslant |x|(1 - |t| C).$$

Furthermore, for all $0 < |t| < \delta/2$,

$$\frac{1}{m'} - \frac{1 - |t| C}{m} < \frac{1}{2}\left(\frac{1}{m'} - \frac{1}{m}\right) < 0$$

is known. This implies, for $0 < |t| < \delta/2$,

$$|(1/t) y_t^j y_t^k g(x + \theta_t(\cdot) y_t)| \leqslant C^2 |t|\, |x|^2 \|g\|_m e^{-|x + \theta_t(x) y_t|/m} \leqslant |t| C^2 \|g\|_m |x|^2 e^{-|x|(1 - C|t|)/m}$$

and

$$\|(1/t) y_t^j y_t^k g(\cdot + \theta_t(\cdot) y_t)\|_{m'} \leqslant |t| C^2 \|g\|_m \sup_{|\mathrm{Im} x| \leqslant 1/m} |x|^2 e^{|x|[1/m' - (1 - C|t|)/m]} \leqslant |t|\, \|g\|_m C_m,$$

for all $0 < |t| < \delta/2$, follows easily.

Proposition 4.4: The space $\mathscr{O}(\mathbb{D}^n)$ is invariant under the infinitesimal generators of the induced action of (4.1) on $\mathscr{O}(\mathbb{D}^n)$, that is, for fixed $f \in \mathscr{O}(\mathbb{D}^n)$, one has

$$\lim_{|t| \to 0} \{f(\Lambda_t \cdot) - f(\cdot)\} = \hat{\Sigma} f \quad \text{in } \mathscr{O}(\mathbb{D}^n), \tag{4.5}$$

where

$$(\hat{\Sigma} f)(x) = \sum_{j=1}^{n} (\Sigma x)^j (\partial_j f)(x).$$

Proof: Since $\mathscr{O}(\mathbb{D}^n)$ is the inductive limit of the Banach spaces $\mathscr{O}_c^m = \mathscr{O}_c^m(\mathbb{D}^n)$ for $m \to \infty$ it suffices to show that, for fixed $f \in \mathscr{O}_c^m$, there is some $m' > m$ such that the above limit relation holds in $\mathscr{O}_c^{m'}$. According to Eq. (4.4),

$$(1/t)\{f(\Lambda_t x) - f(x)\} - (\hat{\Sigma} f)(x) = \sum_{i=1}^{n} \left(\frac{1}{t} y_t^j - y^j\right) \partial_j f(x) + \sum_{j,k=1}^{n} \frac{1}{t} y_t^j y_t^k (\partial_j \partial_k f)(x + \theta_t y_t).$$

Lemmas 4.1 and 4.2 imply that the first term of the right-hand side tends to zero in $\mathscr{O}_c^{m'}$ for $|t| \to 0$ for all $m' > m + 1$. Similarly the second term converges to zero in $\mathscr{O}_c^{m'}$ for $|t| \to 0$ for all $m' > m + 2$ by Lemmas 4.1 and 4.3. This proves (4.5).

Corollary 4.5: A test-function space of the form $E = \mathscr{O}(\mathbb{D}^4, V)$ with action α of $G = \mathrm{iSL}(2,\mathbb{C})$ on E specified by Eq. (3.3) is invariant under the infinitesimal generators of this action; i.e., E is invariant under the generators of the translations and the generators of the Lorentz transformations on E.

Proof: If $t \to A_t$ is a one-parameter subgroup of the Lie group $\mathrm{SL}(2,\mathbb{C})$, we take, in Proposition 4.4,

$$\Lambda_t = \Lambda(A_{-t}), \quad t \in \mathbb{R},$$

where Λ is the canonical homomorphism from $\mathrm{SL}(2,\mathbb{C})$ onto $L_+^\uparrow$. Since $t \to S(A_{-t})$ is easily seen to be differentiable (compare Sec. III A), Proposition 4.4 implies that

$$\lim_{|t| \to 0} (1/t)\{\alpha_{(0,A_t)} f - f\} = \lim_{|t| \to 0} (1/t)\{S(A_{-t}) f(\Lambda_t \cdot) - f(\cdot)\}$$

exists in E and thus proves the invariance of this test-function space under the generators of the "Lorentz transformations" on E. The case of translations is even simpler. Consider the translation group in direction

$$e\in\mathbb{R}^4,\ |e|=1:\quad \alpha_{(te,1)}f(x)=f(x-te),\quad t\in\mathbb{R}.$$

If we identify $y_t=-te$ and $y=-e$ we have, instead of (4.3),

$$y_t-ty=0 \quad\text{and}\quad |y_t|=|t|, \tag{4.3'}$$

and thus the proof of Proposition 4.4 simplifies considerably. Finally this implies the invariance of E under the generators of the translations on E.

Theorem 4.6: In a relativistic quantum field theory over a test-function space $E=\mathscr{O}(\mathbb{D}^4,V)$, where the point $p=0$ is isolated in the energy momentum spectrum Σ, the following identity holds for arbitrary but fixed a, $a^2<0$ [only (H_0)–(H_3) are assumed]:

$$\operatorname*{w-lim}_{\lambda\to\infty} U(\lambda a,1)=Q_0\equiv E(\{0\}), \tag{4.6}$$

where Q_0 is the projection operator onto the subspace of translation invariant states.

Hence the theory has a unique vacuum state [i.e., condition (H_5) holds] if and only if the *cluster property*

$$\mathscr{W}(\underline{g}\cdot\underline{\alpha}_{\lambda a}\underline{f})\to\mathscr{W}(\underline{g})\mathscr{W}(\underline{f}),\quad\text{for }\lambda\to\infty,$$
$$\text{for all }\underline{g},\ \underline{f}\in\underline{E}, \tag{4.7}$$

is satisfied.

Proof: Corollary 4.5 assures that all assumptions for the "Jost–Hepp proof" of this statement[30] are satisfied. Thus we are finished.

V. CONNECTION BETWEEN SPIN AND STATISTICS

There is a set of results in QFT usually referred to as the spin-statistics theorem by which the form of the commutation relation for the field (used in the formulation of the locality condition) is related to the type of field (spinor or tensor).[1–3]

The main results in this respect are the theorem of Burgoyne, Lüders, and Zumino on one side and the theorem of Dell'Antonio on the other side. The proof of the result mentioned first relies on properties of the Lorentz group and its representations and on analyticity properties of the Wightman functions [analyticity and $L^+(\mathbb{C})$ covariance in the extended tubes, existence of Jost points, and the fact that $-\mathbf{1}_4\in L^+(\mathbb{C})$]. Since these properties are also available in HFQFT (see Sec. III) the theorem of Burgoyne, Lüders, and Zumino still holds in HFQFT. Dell'Antonio's theorem reads in its HFQFT version as follows:

Theroem 5.1: If a relativistic quantum field $A=(A_1,\dots,A_M)$ over $E=\mathscr{O}(\mathbb{D}^4,V)$, dim $V=M$, satisfies

$$\operatorname{supp}[A_k(x),A_j^*(y)]_-\subset\overline{K} \tag{5.1}$$

and

$$\operatorname{supp}[A_k(x),A_j(y)]_+\subset\overline{K}, \tag{5.2}$$

then either $A_j=0$ or $A_k=0$.

The same conclusion holds if in (5.1) and (5.2) the signs $+$ and $-$ are exchanged.

Compared to the situation in "standard" QFT this result is considerably harder to prove in HFQFT.

The starting point for a proof is the following elementary identity which holds for all test functions $f,g\in\mathscr{O}(\mathbb{D}^4)$ and all $\lambda\geqslant 0$:

$$\begin{aligned}&\|A_j(f)A_k(g_\lambda)\Phi_0\|^2\\&\quad+(A_k(g)^*A_k(g)\Phi_0,U(\lambda a)A_j(f)^*A_j(f)\Phi_0)\\&=(A_k(g_\lambda)\Phi_0,A_j(f)^*[A_j(f),A_k(g_\lambda)]_+\Phi_0)\\&\quad+(A_k(g_\lambda)\Phi_0,[A_k(g_\lambda),A_j(f)^*]_-A_j(f)\Phi_0)\\&\equiv \mathrm{I}_\lambda+\mathrm{II}_\lambda,\end{aligned} \tag{5.3}$$

where

$$U(\lambda a)A_j(g)U(\lambda a)^{-1}=A_j(g_\lambda),\quad g_\lambda=g_{\lambda a},$$

with some spacelike vector $a=(0,\mathbf{a})$, $\mathbf{a}^2=1$.

If test functions f and g of compact support were available one could choose λ sufficiently large so that the functions f and g_λ would have spacelike separated supports. Then the assumptions (5.1) and (5.2) would easily imply that the right-hand side of Eq. (5.3) vanishes, and by the cluster property one would easily conclude the proof. In our case of HFQFT the control over the rhs of Eq. (5.3) needs considerably more preparation relating geometrical facts about a product of Minkowski spaces with the topology of the underlying test-function space $\mathscr{O}(\mathbb{D}^4)$ as well as the precise formulation of the "locality conditions" (5.1) and (5.2).

Denote by $\overline{V}_+$ the closed forward light cone $\{\xi^3,\xi)\in\mathbb{R}^4|\xi^0\geqslant|\xi|\}$ and by $V=\overline{V}_+\cup\overline{V}_-$, $\overline{V}_-=-\overline{V}_+$ the closed light cone.

The set K of (5.1) and (5.2) decomposes into $\overline{K}=\overline{K}^4\cup\overline{K}^-$, where

$$K^\pm=\{z=(x,y)\in\mathbb{R}^4\times\mathbb{R}^4|x-y\in\overline{V}_\pm\}.$$

The following lemma establishes some facts about the separation of the set K from "spacelike" points. It is proved in the Appendix.

Lemma 5.2: For $a=(0,\mathbf{a})\in\mathbb{R}^4$, $\mathbf{a}^2=1$, denote $\hat{a}=(0,a)\in\mathbb{R}^4\times\mathbb{R}^4-K$. Then the following hold.

(a) The $\operatorname{dist}(K,\lambda\hat{a})=\lambda/2$, $\lambda\geqslant 0$, is attained at

$$\lambda a^\pm=\lambda((\mp 1,\mathbf{a}),(\pm 1,3\mathbf{a}))\in K.$$

(b) Denote by $e^\pm=a^\pm-\hat{a}$; then $z\cdot e^+\geqslant 0$, for all $z\in K^+$, and $z\cdot e^-\geqslant 0$, for all $z\in K^-$. Hence, for all $z\in K^\pm$,

$$|z-\lambda\hat{a}|\geqslant|z-\lambda a^\pm|.$$

(c) Given $\delta_0>0$, define $\lambda_0=\lambda_0(\delta_0)$ by

$$\lambda_0=[16|a^\pm|/(2|a^\pm|-1)]\delta_0,\quad |a^\pm|=\sqrt{3}/2;$$

then, for all $\lambda\geqslant\lambda_0$ and all $z\in K$,

$$|z-\lambda\hat{a}|\geqslant\epsilon_0|z|+\delta_0+\lambda/4,$$

where $\epsilon_0^{-1}=8|a^\pm|=4\sqrt{3}$.

Lemma 5.3: There exist a positive constant L_0 and a symmetric neighborhood $U=-U$ of $\overline{K}$ (the closure of K in $\mathbb{D}^8$) in $\mathbb{D}^8+i\mathbb{R}^8$ such that

$$\operatorname{dist}(U\cap\mathbb{C}^8,\lambda\hat{a})\geqslant\lambda/4,\quad\text{for all }\lambda\geqslant L_0.$$

Proof: Let $B(z,r)$ be the open ball of radius r and center $z\in\mathbb{R}^8$. For $0<\epsilon$, $0<\delta$, introduce the real neighborhood

$$K_{\epsilon,\delta}=\bigcup_{z\in K}B(z,\epsilon|z|+\delta)$$

of K and then the complex neighborhood

$$W_{\epsilon,\delta} = \{z\in\mathbb{C}^8 | \operatorname{Re} z\in K_{\epsilon,\delta}, \quad |\operatorname{Im} z| < \delta\}.$$

If we choose now $0<\epsilon\leqslant\epsilon_0$ and $\delta_0\leqslant 2\delta$ and apply part (c) of Lemma 5.2 we see easily that $|z-\lambda\hat{a}|\geqslant\lambda/4$ holds for all $z\in W_{\epsilon,\delta}$ and all $\lambda\geqslant\lambda_0(\delta_0)$.

By definition of $\mathbb{D}^8$ there exists a neighborhood U of $\overline{K}$ in $\mathbb{D}^8 + i\mathbb{R}^8$ such that $U\cup\mathbb{C}^8 = W_{\epsilon,\delta}$ holds.

Lemma 5.4: For any $f\in\mathscr{Q}(\mathbb{D}^8)$, define f_λ for $\lambda\geqslant 0$ by

$$f_\lambda(z) = f(z+\lambda\hat{a}), \quad z\in\mathbb{D}^8,$$

where $\hat{a}$ is defined in Lemma 5.2. Then

$$f_\lambda\to 0, \quad \text{for} \quad \lambda\to\infty \quad \text{in} \quad \mathscr{Q}(\overline{K}).$$

Proof: The space $\mathscr{Q}(\overline{K})$ is defined to be the inductive limit of a sequence of Banach spaces $\mathscr{O}_c^m(U_m)$, $m\in\mathbb{N}$, where $\{U_m | m\in\mathbb{N}\}$ is a fundamental system of neighborhoods of $\overline{K}$ in $\mathbb{Q}^8 = \mathbb{D}^8 + i\mathbb{R}^8$; hence we have to show that, for some m,

$$f_\lambda\to 0 \quad \text{for} \quad \lambda\to\infty \quad \text{in} \quad \mathscr{O}_c^m(U_m).$$

Since $f\in\mathscr{Q}(\mathbb{D}^8)$ there exist positive numbers δ and C such that

$$|f(z)|\leqslant Ce^{-\delta|z|} \quad \text{in} \quad \{z|\ |\operatorname{Im} z|\leqslant\delta\}.$$

With this $\delta>0$, do the construction of Lemma 5.3 to obtain a neighborhood U of $\overline{K}$ in $\mathbb{Q}^8$ such that

$$U\cap\mathbb{C}^8\subset\{z|\ |\operatorname{Im} z|<\delta\},$$

and, for sufficiently large λ,

$$\operatorname{dist}(U\cap\mathbb{C}^8,\lambda\hat{a})\geqslant\lambda/4.$$

Then there is $M\in\mathbb{N}$ such that, for all $m\geqslant M$,

$$U_m\subset U \quad \text{and} \quad m\delta\geqslant 4.$$

Next observe that

$$z-\lambda\hat{a}\in U\cap\mathbb{C}^8 \quad \text{implies} \quad |z|\geqslant\lambda/4,$$

since then $\zeta=\lambda\hat{a}-z\in U\cap\mathbb{C}^8$ and $z=\lambda\hat{a}-\zeta$; thus, by Lemma 5.3,

$$|z|\geqslant\operatorname{dist}(U\cap\mathbb{C}^8,\lambda\hat{a})\geqslant\lambda/4,$$

if $\lambda\geqslant L_0$.

Now fix $m\geqslant M$ and choose $\lambda\geqslant L_0$. Then the following chain of inequalities holds:

$$\|f_\lambda\|_m = \sup_{z\in U_m\cap\mathbb{C}^8} |f_\lambda(z)|e^{|z|/m}$$

$$\leqslant C \sup_{z-\lambda\hat{a}\in U_m\cap\mathbb{C}^8} e^{-\delta|z|}e^{|z-\lambda\hat{a}|/m}$$

$$\leqslant C \sup_{|z|\geqslant\lambda/4} e^{-(\delta-1/m)|z|+\lambda/m}\leqslant Ce^{-\delta\lambda/8}.$$

Thus we conclude the proof of Lemma 5.4.

Proof of Theorem 5.1: Since $\|A_k(g_\lambda)\Phi_0\| = \|A_k(g)\Phi_0\|$ is known, the first term I_λ in Eq. (5.3) is dominated by

$$|\mathrm{I}_\lambda|\leqslant\|A_k(g)\Phi_0\|\ \|A_j(f)^*[A_j(f),A_k(g_\lambda)]_+\Phi_0\|.$$

Assumption (5.2) means that for any $\Phi,\Psi\in\mathscr{D}$ the functional

$$h_1\times h_2\to(A_j(f)\Phi,[A_k(h_1),A_j(h_2)]_+\Psi)$$

belongs to $\mathscr{Q}(\overline{K})'$. Since $\mathscr{Q}(\overline{K})$ is barreled it follows that

$$h_1\times h_2\to\|A_j(f)^*[A_k(h_1),A_j(h_2)]_+\Psi\|$$

$$= \sup_{\Phi\in\mathscr{D},\|\Phi\|=1}|(A_j(f)\Phi,[A_k(h_1),A_j(h_2)]_+\Psi)|$$

is a continuous seminorm on this space. By Lemma 5.4 we know $f\times g_\lambda\to 0$ for $\lambda\to\infty$ in $\mathscr{Q}(\overline{K})$, for every $f,g\in\mathscr{Q}(\mathbb{D}^4)$. This implies $\mathrm{I}_\lambda\to 0$ for $\lambda\to\infty$. Similarly assumption (5.1), barreledness of $\mathscr{Q}(\overline{K})$, and Lemma 5.4 are used to conclude $\mathrm{II}_\lambda\to 0$ for $\lambda\to\infty$.

Thus the right-hand side of Eq. (5.3) has a vanishing limit for $\lambda\to\infty$. By the cluster property (Theorem 4.6) the second term on the left-hand side of Eq. (5.3) has the limit

$$\|A_k(g)\Phi_0\|^2\|A_j(f)\Phi_0\|^2,$$

for $\lambda\to\infty$. Hence $\|A_j(f)A_k(g_\lambda)\Phi_0\|^2$ also has a limit for $\lambda\to\infty$ and this limit is

$$-\|A_k(g)\Phi_0\|^2\|A_j(f)\Phi_0\|^2.$$

We conclude $\|A_k(g)\Phi_0\|^2\|A_j(f)\Phi_0\|^2=0$ and obtain either

$$A_k(g)\Phi_0=0 \quad \text{or} \quad A_j(f)\Phi_0=0, \quad \text{for all} \quad g,\ f\in\mathscr{Q}(\mathbb{D}^4).$$

If, for instance, $A_j(\cdot)\Phi_0=0$, then, by Proposition 3.6, $A_j=0$ follows.

VI. SOME CHARACTERIZATIONS OF TRIVIAL QUANTUM FIELDS

A. Characterization of generalized free fields

Our first result here assures us that the well known characterization of generalized free fields in terms of commutator properties for the field operators still holds in HFQFT.

Theorem 6.1: For a relativistic quantum field A over $E=\mathscr{Q}(\mathbb{D}^4)$ with cyclic vacuum vector Φ_0 the following conditions are equivalent:

$$[A(f),A(g)]\subset\mathbb{W}_2(f\otimes g-g\otimes f)\mathbf{1}, \quad \text{for all } f,g\in E, \tag{6.1}$$

$$[A(f),A(g)]\Phi_0=\mathbb{W}_2(f\otimes g-g\otimes f)\Phi_0, \quad \text{for all } f,g\in E. \tag{6.2}$$

Proof: (a) We only have to show that condition (6.2) implies (6.1). And by cyclicity of the vacuum state this follows from

$$\begin{aligned}&\mathbb{W}_{n+m+2}(f_1\otimes\cdots\otimes f_n\otimes f\otimes g\otimes g_1\otimes\cdots\otimes g_m)\\&\quad-\mathbb{W}_{n+m+2}(f_1\otimes\cdots\otimes f_n\otimes g\otimes f\otimes g_1\otimes\cdots\otimes g_m)\\&\quad=\mathbb{W}_2([f,g])\mathbb{W}_{n+m}(f_1\otimes\cdots\otimes f_n\otimes g_1\otimes\cdots\otimes g_m),\end{aligned} \tag{6.3}$$

for all $f,g,f_j,g_j\in E$ and $n,m=0,1,2,\ldots$, with $[f,g]=f\otimes g-g\otimes f$. By Theorem 3.4 the hyperfunctions $\mathbb{W}_N$, $N=2,3,\ldots$, are boundary values of analytic functions $\widehat{\mathbb{W}}_N$. So we study the analytic functions $\mathbb{W}^\pm$ on tubes $T(\Gamma_\pm)$ defined by

$$\mathbb{W}^+(\underline{z}_n,z,w,\underline{w}_m)=\widehat{\mathbb{W}}_{n+m+2}(\underline{z}_n,z,w,\underline{w}_m)-\widehat{\mathbb{W}}_2(z,w)\widehat{\mathbb{W}}_{n+m}(\underline{z}_n,\underline{w}_m),$$

$$\begin{aligned}\Gamma_+=\{(\xi_n,\xi,\eta,\eta_m)|&\xi_{j+1}-\xi_j\in V_+,\quad j\leqslant n-1,\\&\xi-\xi_n\in V_+,\quad \eta-\xi\in V_+,\quad \eta_1-\eta\in V_+,\\&\eta_{j+1}-\eta_j\in V_+,\quad j\leqslant m-1\}\end{aligned} \tag{6.4}$$

and

$$\hat{\mathbb{W}}^-(\underline{z}_n,z,w,\underline{w}_m) = \hat{\mathbb{W}}_{n+m+2}(\underline{z}_n,w,z,\underline{w}_m) - \hat{\mathbb{W}}_2(w,z)\hat{\mathbb{W}}_{n+m}(\underline{z}_n,\underline{w}_m),$$

$$\Gamma_- = \{(\underline{\xi}_n,\xi,\eta,\underline{\eta}_m) | \xi_{j+1}-\xi_j \in V_+, \ j \leqslant n-1, \ -\eta+\xi \in V_+, \ \eta - \xi_n \in V_+, \ \eta_1 - \xi \in V_+, \ \eta_{j+1}-\eta_j \in V_+, \ j \leqslant m-1\}. \quad (6.5)$$

(b) By Theorem 3.4 these analytic functions $\hat{\mathbb{W}}^\pm$ have boundary values $\mathbb{W}^+ = \delta_{\Gamma_+}\hat{\mathbb{W}}^+$ (resp. $\mathbb{W}^- = \delta_{\Gamma_-}\hat{\mathbb{W}}^-$) in the sense of hyperfunctions. By locality (H_4) and our assumption (6.2) these boundary values agree on the open subset U of $\mathbb{R}^{4(n+m+2)}$,

$$U = \{(\underline{x}_n,x,y,\underline{y}_m) | (x-y_j)^2 < 0, \ (y-y_j)^2 < 0, \ j=1,\dots,m\}.$$

Hence by the edge of the wedge theorem for hyperfunctions (Proposition 2.10) there exist a complex neighborhood V of U and a function

$$\hat{\mathbb{W}} \in \tilde{\mathcal{O}}(V \cap T(\mathrm{ch}(\Gamma_+ \cup \Gamma_-))$$

such that (6.6)

$$\hat{\mathbb{W}} = \hat{\mathbb{W}}^\pm \quad \text{on} \quad V \cap T(\Gamma_\pm).$$

(c) Observe now that the complex cone

$$\Gamma_0 = \{(\underline{\xi}_n,\xi,\eta,\underline{\eta}_m) | \xi_{j+1}-\xi_j \in V_+, \ j \leqslant n-1, \ \times \xi - \xi_n \in V_+, \ \eta = \xi, \ \eta_1 - \eta \in V_+, \ \times \eta_{j+1} - \eta_j \in V_+, \ j \leqslant m-1\} \quad (6.7)$$

is contained in the convex hull $\mathrm{ch}(\Gamma_+ \cup \Gamma_-)$ of $\Gamma_+ \cup \Gamma_-$.

To prove this, suppose $(\underline{\xi}_n,\xi,\eta,\underline{\eta}_m) \in \Gamma_0$ to be given. Then there are $\zeta_j \in V_+$ such that $\zeta_1+\zeta_3-\zeta_2 \in V_+$,

$$\xi - \xi_n = \zeta_1 + \tfrac{1}{2}\zeta_3 \quad \text{and} \quad \eta_1 - \eta = \zeta_2 + \tfrac{1}{2}\zeta_3.$$

Write $\underline{\xi}_n = (\xi_1, \underline{\xi}'_n)$ and define

$$\underline{\zeta}^+ = (\underline{\xi}_n^+,\xi^+,\eta^+,\eta_1^+,\underline{\eta}_m'^+) \equiv (\underline{\xi}_n,\xi_n+\zeta_1,\zeta_1+\zeta_3+\xi_n,\zeta_1+\zeta_2+\zeta_3+\xi_n,\underline{\eta}'_m)$$

and

$$\underline{\zeta}^- = (\underline{\xi}_n^-,\xi^-,\eta^-,\eta_1^-,\underline{\eta}_m'^-) \equiv (\underline{\xi}_n,\xi+\tfrac{1}{2}\xi_3,\eta+\tfrac{1}{2}\xi_3-\xi_2,\eta_1,\underline{\eta}'_m).$$

Then $\underline{\zeta}^\pm \in \Gamma_\pm$ and

$$(\underline{\xi}_n,\xi,\eta,\underline{\eta}_m) = (\underline{\zeta}^+ + \underline{\zeta}^-)/2 \in \mathrm{ch}(\Gamma_+ \cup \Gamma_-).$$

(d) By (6.6) and (6.7) we conclude that $\hat{\mathbb{W}}^+$ and $\hat{\mathbb{W}}^-$ are analytically continued with respect to the variables (z,w) to $\mathrm{Im}(z-w)=0$. Therefore $\mathbb{W}^\pm(\underline{x}_n,x,y,\underline{y}_m)$ can be considered to be hyperfunctions in the variables $\underline{x}_n, \underline{y}_m$ with real analytic parameters x,y. And these hyperfunctions are boundary values of functions $\hat{\mathbb{W}}^\pm(x,y;\underline{z}_n,\underline{w}_m)$ analytic in the tube $T(\Gamma)$,

$$\Gamma = \{(\underline{\xi}_n,\underline{\eta}_m) | \xi_{j+1}-\xi_j \in V_+, \ j \leqslant n-1, \ \xi_n \in V_+, \ \eta_1 \in V_+, \ \eta_{j+1}-\eta_j \in V_+, \ j \leqslant m-1\},$$

hence

$$\hat{\mathbb{W}}(x,y,\underline{z}_n,\underline{w}_m) = \hat{\mathbb{W}}^+(x,y;\underline{z}_n,\underline{w}_m) - \hat{\mathbb{W}}^-(x,y;\underline{z}_n,\underline{w}_m)$$

is analytic in $T(\Gamma)$.

(e) If U_1 is some open bounded nonempty set in $\mathbb{R}^{2\cdot 4}$ there exists an open set U_2 in $\mathbb{R}^{4m}$ such that all $(x,y,\underline{y}_m) \in U_1 \times U_2$ satisfy

$$(x-y_j)^2 < 0 \quad \text{and} \quad (y-y_j)^2 < 0, \quad \text{for} \quad j=1,\dots,m.$$

For all $(x,y) \in U_1$ consider the boundary values

$$\mathbb{W}(x,y;\underline{x}_n,\underline{y}_m) = \delta_\Gamma \hat{\mathbb{W}}(x,y;\underline{z}_n,\underline{w}_m)$$

in $\mathbb{R}^{4n} \times U_2$ in the sense of hyperfunctions. As we have shown above in (b) all the boundary values vanish ($\mathbb{R}^{4n} \times U_1 \times U_2 \subset U$). Hence by Corollary 2.11 all the analytic functions $\hat{\mathbb{W}}(x,y;\dots)$ on $T(\Gamma)$, $x,y \in U_1$, vanish. Since $\hat{\mathscr{W}}(x,y;\dots)$ is real analytic in x, y, this function vanishes identically.

By definitions (6.4) and (6.5) this proves Eq. (6.3) and thus we have the theorem.

A relativistic quantum field A over $\mathcal{Q}(\mathbb{D}^4)$ that satisfies condition (6.1) or (6.2) is called a *generalized free field* over $\mathcal{Q}(\mathbb{D}^4)$. Such fields have been studied in some detail by Roberts.[31] Clearly as in the case of tempered fields relation (6.1) determines easily all n-point functions of the theory. The relevant formulas are given in the following corollary.

Corollary 6.2: If a relativistic quantum field A over $E = \mathcal{Q}(\mathbb{D}^4)$ has a vanishing one-point function $\mathbb{W}_1$ and satisfies (6.1) or (6.2) its n-point functions are

$$\mathbb{W}_{2n+1} = 0, \quad \text{for } n = 0,1,2,\dots, \qquad \mathbb{W}_{2n} = \mathbb{W}_{2n}^0 \equiv \circledS^n \mathbb{W}_2, \quad \text{for } n = 1,2,\dots, \quad (6.8)$$

where $\mathbb{W}_{2n}^0$ are recursively defined by

$$\mathbb{W}^0_{2(n+1)}(f_1 \otimes \cdots \otimes f_{2n+2}) = \sum_{j=2}^{2n+2} \mathbb{W}_2^0(f_1 \otimes f_j)\mathbb{W}_{2n}^0(f_1 \otimes \cdots \otimes \hat{f}_j \otimes \cdots \otimes f_{2n+2}). \quad (6.9)$$

B. Jost–Schroer theorem

According to Theorem 6.1 and Corollary 6.2 the four-point function can be used to decide whether a field over $E = \mathcal{Q}(\mathbb{D}^4)$ is a generalized free field or not. In the case of tempered fields Jost and Schroer[32] have observed that this result can be used to determine a scalar relativistic quantum field completely if its two-point function is known to have a special form (that of a free scalar field). For hyperfunction quantum fields this characterization continues to hold if we add a technical assumption on the support of the four-point function.

Theorem 6.3 (Jost–Schroer theorem for HFQFT): If the two-point function $\mathbb{W}_2$ of a relativistic quantum field A over $E = \mathcal{Q}(\mathbb{D}^4)$ with cyclic vacuum Φ_0 equals that of a free field of mass $m > 0$ and if the four-point function has no "pathological support" (see Remark 6.1) in energy momentum space, then A is a free field of mass m.

Remark 6.1: According to the results of Sec. III B,

$$\tilde{g} \to \tilde{W}_3(\tilde{h} \otimes \tilde{f}\tilde{f}_1 \otimes \tilde{g}) \quad (6.10)$$

is a well defined Fourier hyperfunction with support in $\bar{\Sigma}$ (the closure of Σ in $\mathbb{D}^4$) for arbitrary $\tilde{h} \in \mathcal{Q}(\mathbb{D}^4)$ and $\tilde{f}\tilde{f}_1 \in \mathscr{D}(\mathbb{R}^4)$. Then we say that $\tilde{\mathbb{W}}_4$ has no "pathological support" if the support of the Fourier hyperfunction (6.10) has no connected component contained in S^3_∞.

Remark 6.2: This technical assumption is actually known to be satisfied in some cases. If the theory is formulated over a certain slightly bigger test-function space $E_1 \supset E$ or if the four-point function has a continuous extension to this space then this support property can be shown to hold. A concrete example of such a space E_1 is described in Ref. 7.

Proof of Theorem 6.3: (a) By assumption we have, for all $f \in E$,

$$\|A((\Box + m^2)f)\Phi_0\|^2 = c\langle \delta^+_m(p), (-p^2+m^2)^2 |\tilde{f}(p)|^2\rangle = 0.$$

Hence the field B defined by $B = (\Box + m^2)A$ satisfies $B(f)\Phi_0$, for all $f \in E$. Since the local field B clearly is relatively local to the field A Proposition 3.6 implies $B = 0$, that is the field A solves the linear differential equation

$$(\Box + m^2)A(x)\Phi = 0, \quad \text{for} \quad \Phi \in \mathscr{D},$$

and this implies, for the Fourier transform $\tilde{A}\Phi$ of $A\Phi$,

$$\operatorname{supp} \tilde{A}\Phi \subset \overline{H}_m, \tag{6.11}$$

where H_m is the mass hyperboloid

$$H_m = H^+_m \cup H^-_m, \quad H^\pm_m = \{(p_0, \mathbf{p}) | p_0 = \pm\sqrt{\mathbf{p}^2 + m^2}\},$$

and $\overline{H}_m$ denotes its closure in $\mathbb{D}^4$.

For such a field one obtains more refined support properties for the Fourier hyperfunction $\tilde{Z}_2$ introduced by Eqs. (3.13) and (3.15):

$$\operatorname{supp} \tilde{Z}_2 \subset \overline{\left(\bigcup_{p\in\Sigma} \{p\} \times T^+(p)\right)},$$

$$T^+(p) = \overline{H^+_m} \cap \overline{(p - H_m)} = [H^+_m \cap (p - H_m)] \cup \overline{H^+_m} \cap S^3_\infty$$

by Proposition 3.2 and Corollary 3.3.

(b) In order to complete the proof it suffices, according to (6.11) and Theorem 6.1, to show Eq. (6.2). To this end we study the Fourier hyperfunction $[A(x_1), A(x_2)]\Phi_0$ in the coordinates

$$x = (x_1 + x_2)/2, \quad \xi = x_2 - x_1, \quad \Psi^+(x,\xi) = A(x_1)A(x_2)\Phi_0, \tag{6.12}$$

$$\Psi^-(x,\xi) = A(x_2)A(x_1)\Phi_0 = \Psi^+(x, -\xi).$$

The Fourier transform of Ψ^+ satisfies

$$\tilde{\Psi}^+(p,q) = \tilde{Z}_2(p,(q+p)/2),$$

$$\tilde{\Psi}^+(p,(q-p)/2) = \tilde{Z}_2(p,q), \tag{6.13}$$

and it follows, for $\tilde{\Psi} = \tilde{\Psi}^+ - \tilde{\Psi}^-$,

$$\sup \tilde{\Psi} \subset \left(\bigcup_{p\in\Sigma} \{p\} \times S^+(p)\right),$$

$$S(p) = S^+(p) \cup S^-(p), \quad S^-(p) = -S^+(p),$$

$$S^+(p) = \overline{(-p/2 + H^+_m)} \cap \overline{(p/2 - H_m)} = [(-p/2 + H^+_m) \cap (p/2 - H_m)] \times \cup \overline{H^+_m} \cap S^3_\infty. \tag{6.14}$$

Elementary geometry shows

(i) $S^+(0) = \overline{H}^+_m$,

(ii) if $B \subset \Sigma$ is compact in $\mathbb{R}^4$ and $0 \notin B$, then $\left(\bigcup_{p\in B} \{p\} \times S(p)\right) \subset B \times (\hat{B} \cup \overline{H}_m \cap S^3_\infty)$ with some compact set $\hat{B} \subset \mathbb{R}^4$. (6.15)

(c) In the same way as in Proposition 3.1 we can show that $\tilde{\Psi}(p,q)$ is a Radon measure in p and a Fourier hyperfunction in q. Thus we may choose $\tilde{f} \in \mathscr{D}(\mathbb{R}^4)$, $\operatorname{supp} \tilde{f} = B$ with $0 \notin B \subset \Sigma$ and know by Proposition 3.1 that $\tilde{g} \to \tilde{\Psi}(\tilde{f} \otimes \tilde{g})$ is a well defined Fourier hyperfunction, $\tilde{g} \in \mathscr{O}(\mathbb{D}^4)$. Relations (6.12) show

$$\tilde{\Psi}(\tilde{f} \otimes \tilde{g}) = \langle [A(x_1), A(x_2)]\Phi_0, f((x_1 + x_2)/2) g(x_2 - x_1)\rangle, \tag{6.16}$$

therefore by locality the Fourier transform of $\tilde{\Psi}(\tilde{f}, q)$ vanishes for $\xi^2 < 0$. According to statement (6.15) the support of the Fourier hyperfunction $\tilde{\Psi}(\tilde{f}, q)$ is contained in $\hat{B} \cup \overline{H}_m \cap S^3_\infty$. Now apply Eq. (3.26) for $\tilde{f} \in \mathscr{D}(\mathbb{R}^4)$ and $\tilde{f}_1, \tilde{g}, \tilde{g}_1 \in \mathscr{O}(\mathbb{D}^4)$ to get

$$(\tilde{Z}_2(\tilde{f}_1 \otimes \tilde{g}_1), \quad \tilde{\Psi}(\tilde{f} \otimes \tilde{g})) = \tilde{W}_3(\bar{\tilde{g}}_1 \otimes \bar{\tilde{f}}_1 \tilde{f} \otimes (\tilde{g} - \hat{\tilde{g}})),$$

where $\hat{\tilde{g}}(q) = \tilde{g}(-q)$. This is a Fourier hyperfunction with respect to $\tilde{g}$ with support in $\hat{B} \cup \overline{H}_m \cap S^3_\infty$ and its support is contained in $\hat{B}$ by Remark 6.1. Since $\tilde{f}_1, \tilde{g}_1 \in \mathscr{O}(\mathbb{D}^4)$ are arbitrary, the support of $\tilde{\Psi}(\tilde{f}, q)$ is contained in $\hat{B}$. Hence its Fourier Laplace transform with respect to q is an entire analytic function of ζ that vanishes on the open subset $\xi^2 < 0$. Thus $\tilde{\Psi}(\tilde{f}, q)$ vanishes and we deduce by choice of $\tilde{f}$

$$\operatorname{supp} \Psi \subset \{0\} \times S(0) = \{0\} \times \overline{H}_m. \tag{6.17}$$

(d) Denote by χ_ϵ the characteristic function of a ball of radius $\epsilon > 0$ and center $p = 0$. Proposition 3.1 and relation (6.17) imply that $\chi_\epsilon \tilde{\Psi}(p,q)$ is well defined and that, for all $\epsilon > 0$,

$$\chi_\epsilon \tilde{\Psi} = \tilde{\Psi}$$

holds. Thus we get for

$$\chi_\epsilon(P)[A(f), A(g)]\Phi_0 = [A(f), A(g)]\Phi_0,$$

$$[A(f), A(g)]\Phi_0 = \langle \tilde{\Psi}(p,q), \tilde{f}(p/2 - q)\tilde{g}(p/2 + q)\rangle.$$

But by uniqueness of the vacuum state (H_5) we know that

$$\chi_\epsilon(P) = \int \chi_\epsilon(p)\, dE(p)$$

converges strongly for $\epsilon \to 0$ to the projection operator $|\Phi_0\rangle\langle\Phi_0|$. This then proves Eq. (6.2).

C. Borchers classes

As with the result about the existence of a PCT operator for a QFT in terms of Fourier hyperfunctions we only indicate in this subsection that also for quantum fields over $E = \mathscr{O}(\mathbb{D}^4, V)$ the concept of the "Borchers class" of some field is available since the possibility for this concept relies exclusively on analyticity properties of the n-point functions and the existence of a PCT operator. These analyticity properties are provided by Theorem 3.4, and the techniques of the proof are very similar to those explained in detail in Secs. III C and VI A.

However, as expected, compared to the case of tempered fields the Borchers class of a field in HFQFT is considerably bigger. In order to see this recall that for tempered fields the Borchers class of a free massive field consists of all Wick polynomials including derivatives of that field.[33]

In Ref. 34 it has been shown that all power series

$$B(x) = \sum_{n=0}^{\infty} c_n :A^n:\frac{(x)}{n!}, \quad \lim_{n\to\infty} [\,|c_n|^2/n!\,]^{1/n} = 0, \tag{6.18}$$

define a relativistic quantum field over $E = \mathscr{O}(\mathbb{D}^4)$ and that the associated sequence of Wick polynomials

$$B_N(x) = \sum_{n=0}^{N} c_n :A^n:\frac{(x)}{n!}, \quad N = 1,2,..., \tag{6.19}$$

converges in the relevant topology to $B(x)$. Since all the $B_N(x)$ are known to be relatively local with respect to A it follows that B, too, is relatively local with respect to A.

Hence all entire function of A as described in (6.18) belong to the Borchers class of A.

VII. CONCLUSIONS

In order to give a comprehensive picture about QFT in terms of Fourier hyperfunctions we discuss here the status of the remaining points of the basic structural results of QFT mentioned in Sec. I. The existence of a scattering operator [point (3)] in HFQFT has been proved in Ref. 7; however, it has been proved only for a special class of Fourier hyperfunctions, that is, for a somewhat larger test-function space than $\mathscr{O}(\mathbb{D}^4)$. Though this is already quite a satisfactory result it would be preferable to have a scattering operator also for fields over $\mathscr{O}(\mathbb{D}^4)$. This point is under consideration.

In general form of the two-point function [point (7)] can also be determined in HFQFT. The result is the obvious generalization of the form given by Källen and Lehmann. If we combine the information provided by Propositions 3.1 and 3.2 with Eq. (3.26) for $n = 2$ we get immediately that the two-point function of a scalar field over $\mathscr{O}(\mathbb{D}^4)$ has the following general form:

$$\mathbb{W}_2(f\otimes g) = \int \tilde{t}(dp)\tilde{f}(-p)\tilde{g}(p),$$

with some $L^{\uparrow}_{+}$-invariant positive Radon measure t with support Σ, which is slowly increasing in the sense of Proposition 3.1(b). The structures of such measures are known[35]:

$$t(dp) = c\delta(p)d^4p + \int_0^\infty \rho(d\kappa)\delta_\kappa^+(p)d^4p, \quad c\geqslant 0,$$

with some positive slowly increasing measure ρ on $(0,\infty)$ that is not necessarily polynomially bounded as for tempered fields.

The possibility of a Euclidean reformulation [point (10)] of relativistic QFT in terms of Fourier hyperfunctions has been indicated to exist by Nagamachi and Mugibayashi in Ref. 5 shortly after Osterwalder and Schrader's solution of this problem in terms of distribution. At the price of introducing an even smaller test-function space $\tilde{\mathscr{O}}(\mathbb{D}^4) \subset \mathscr{O}(\mathbb{D}^4)$, Nagamachi and Mugibayashi[6] could actually prove a complete symmetry between the Euclidean and relativistic formulation of "HFQFT" without any additional growth restrictions as in the distributional setting.

However, the space $\tilde{\mathscr{O}}(\mathbb{D}^4)$ has some disadvantages as a space of test functions for QFT. So one might reconsider this problem for the test-function space $\mathscr{O}(\mathbb{D}^4)$. Admitting eventually similar additional growth restrictions as in the distributional setting the proof of equivalence between the Euclidean and relativistic formulation of HFQFT seems to be possible.

We have not tried to prove dispersion relations, which is quite an involved matter. However, we expect that it is possible to prove the necessary analyticity properties for the 2–2-particle scattering amplitude but not the necessary growth restrictions in order to be able to write a dispersion relation with a finite number of subtractions.

Finally we sum up the main points of this paper and give an outlook for further applications of HFQFT.

Since there is no *a priori* choice for the test-function space in relativistic quantum field theory we have isolated conditions on a space E of functions on space-time in order that E be "admissible" as the test-function space of a relativistic QFT [condition (H_0) in Sec. I]. As our short review shows it has been known since the early days of general QFT and has emerged more clearly later by considering model constructions that the traditional choice $E = \mathscr{S}(\mathbb{R}^4, V)$ has to be modified for various important reasons. And accordingly several attempts have been made in the past to generalize the notion of a "tempered relativistic quantum field." Most of these suggestions have considerable difficulties with an appropriate notion of localization in coordinate and/or momentum space. Though it might not have been so clear from the beginning, the only suggestion that has a precise notion of localization in both spaces has been that of Nagamachi and Mugibayashi.[4]

In this paper we have stressed the point of view that a sensible generalization of the notion of a tempered quantum field should not only have these localization properties but should also allow us to derive (hopefully) all the basic structural results of QFT known for tempered fields.

And accordingly in this paper we have presented a short introduction to QFT in terms of Fourier hyperfunctions and have shown that indeed most of the structural results of QFT continue to hold in this more general approach. We mention some further results of HFQFT that we think to be important for future applications.

The existence of entire functions of a free massive field A, for instance,

$$:e^{igA(x)}:, \quad g\in\mathbb{R},$$

can be used in the construction of concrete models.

For instance, the transformation

$$A(x) \to :e^{igA(x)}:$$

can be used for an easy "decoupling" of the interaction of the "derivative coupling model" and thus to obtain a solution of this model.[34] We expect that a renormalization theory based on Fourier hyperfunctions would admit a clearer and more powerful notion of "renormalizable interactions" than in the traditional approach based on (tempered) distributions. An example has been treated in Ref. 34.

One important reason for the great success of Euclidean methods in the construction of models in lower-dimensional space-time clearly is the fact that these methods allow us to take into account in a natural and powerful way the relevant positivity condition. Since the test-function space $E = \mathscr{Q}(\mathbb{D}^n, V)$ of HFQFT is a nuclear DFS space its topological dual E' has a well developed theory of Radon probability measures on it as for the standard case $\mathscr{S}'$ or $\mathscr{D}'$. This might turn out to be important for the construction of HFQFT models with nontrivial interactions according to the "functional integral point of view."

Thus we think that our paper clearly shows that the test-function space (1.2) of Fourier hyperfunctions provides quite a comprehensive realization of the requirements (A)–(C) of the Introduction. In any case this approach is much more natural with respect to the realization of the localization problems [requirement (B)] and is considerably more powerful in the realization of the structural results of QFT [requirement (C)] than any other approach. Furthermore as indicated above there are convincing prospects of further successful applications in model constructions.

ACKNOWLEDGMENTS

One of us (S.N.) wishes to thank BiBoS Research Center (Bielefeld) for the financial support and the hospitality that enabled our cooperation.

E.B. gratefully acknowledges financial support by the Max Kade Foundation (New York), and with great pleasure he thanks the Department of Physics, Princeton University (in particular, Professor A. S. Wightman) for their kind hospitality.

Finally we would like to thank Professor A. S. Wightman for pointing out Ref. 31.

APPENDIX: PROOF OF LEMMA 5.2

(a) For $y \in V^c = \mathbb{R}^4 - V$ the distance to the light cone V is easily calculated to be

$$\operatorname{dist}(V, y) = (|\mathbf{y}| - |y_0|)/\sqrt{2} \tag{A1}$$

and it is attained at a point $e(y)$ of the boundary ∂V

$$e(y) = (|y_0| + |\mathbf{y}|)(\operatorname{sgn} y_0, \hat{\mathbf{y}})/2, \quad \hat{\mathbf{y}} = \mathbf{y}/|\mathbf{y}|. \tag{A2}$$

The points of K are parametrized by

$$K = \{z = (y - \xi, y) \,|\, y \in \mathbb{R}^4, \quad \xi \in V\}.$$

We calculate

$$\operatorname{dist}(K, \lambda\hat{a}) = \inf_{z \in K} |z - \lambda\hat{a}|$$

in two steps using (A1): The first step is simply

$$\inf_{\xi \in V, y \in V^c} |(y - \xi, y) - \lambda\hat{a}| = \inf_{y \in V} |y - \lambda a| = \lambda/2, \tag{A3}$$

and for the second we note

$$\inf_{\xi \in V,\ y \in V^c} |(y - \xi, y) - \lambda\hat{a}|^2$$

$$= \inf_{y \in V^c} \{\inf_{\xi \in V} [|y - \xi|^2 + |y - \lambda a|^2]\}.$$

For $\xi = e(y)$ this equals, according to (A1) and (A2),

$$\inf_{y \in V^c} \{(|\mathbf{y}| - |y_0|)^2/2 + |y - \lambda a|^2\}$$

$$= \inf_{|\mathbf{y}| > |y_0|} \{(|\mathbf{y}| - |y_0|)^2/2 + y_0^2 + (|\mathbf{y}| - \lambda)^2\}.$$

The last infimum is attained at

$$|y_0| = \lambda/4 \quad \text{and} \quad |\mathbf{y}| = 3\lambda/4$$

and equals $\lambda^2/4$; hence

$$\inf_{\xi \in V} \inf_{y \in V^c} |(y - \xi, y) - \lambda\hat{a}| = \lambda/2 \tag{A4}$$

and this is attained at

$$y_\lambda^\pm = \lambda(\pm 1, 3\mathbf{a})/4, \quad \xi = e(y_\lambda^\pm),$$

or at $\lambda a^\pm = (y_\lambda^\pm - e(y_\lambda^\pm), y_\lambda^\pm)$, i.e.,

$$\lambda a^\pm = \lambda(\mp 1, 3\mathbf{a}), (\pm 1, 3\mathbf{a})) \tag{A5}$$

and we conclude

$$\operatorname{dist}(K, \lambda\hat{a}) = |\lambda a^\pm - \lambda\hat{a}| = \lambda|e^\pm| = \lambda/2,$$

where

$$e^\pm = a^\pm - \hat{a} = (-\alpha^\pm, \alpha^\pm), \quad \alpha^\pm = (\pm 1, -\mathbf{a})/4. \tag{A6}$$

This proves part (a).

(b) For $z \in K^\pm$, that is, $\xi = y - x \in V^\pm$, we get

$$z \cdot e^\pm = \xi \cdot \alpha^\pm = (\pm \xi^0 - \xi \cdot \mathbf{a})/4 \geqslant (\pm \xi^0 - |\xi|)/4;$$

hence

$$\begin{aligned} &z \cdot e^+ \geqslant 0, \quad \text{for } z \in K^+, \\ &z \cdot e^- \geqslant 0, \quad \text{for } z \in K^-. \end{aligned} \tag{A7}$$

It follows that

$$\begin{aligned} &|z - \lambda\hat{a}| \geqslant |z - \lambda a^+|, \quad \text{for } z \in K^+, \\ &|z - \lambda\hat{a}| \geqslant |z - \lambda a^-|, \quad \text{for } z \in K^-. \end{aligned} \tag{A8}$$

Thus (b) follows.

(c) In order to prove part (c) we distinguish two cases. If $|z| < (\lambda/4 - \delta)\epsilon_0^{-1}$, then, by part (a),

$$\lambda/4 + \delta + \epsilon_0|z| < \lambda/4 + \delta + \lambda/4 - \delta = \lambda/2$$

$$= \operatorname{dist}(K, \lambda\hat{a}) \leqslant |z - \lambda\hat{a}|,$$

if z also belongs to K.

If, however, $|z| \leqslant (\lambda/4 - \delta)8|a^\pm|$, $z \in K$, we use (A8) to obtain

$$|z - \lambda\hat{a}| \geqslant |z - \lambda a^\pm| \geqslant \big| |z| - \lambda|a^\pm| \big| \geqslant |z| - \lambda|a^\mp|$$

$$\geqslant \epsilon_0|z| + \delta + \lambda/4 + \lambda(2|a^\pm| - 1)/2 - 8|a^\pm|\delta$$

$$\geqslant \epsilon_0|z| + \delta + \lambda/4,$$

since $\lambda \geqslant \lambda_0(\delta)$ is equivalent to $\lambda(2|a^\pm| - 1)/2 - 8|a^\pm|\delta \geqslant 0$.

[1]R. F. Streater and W. A. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[2]R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, RI, 1965).
[3]N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, London, 1975).
[4]S. Nagamachi and N. Mugibayashi, Commun. Math. Phys. **46**, 119 (1976).
[5]S. Nagamachi and N. Mugibayashi, Commun. Math. Phys. **49**, 257 (1976).
[6]S. Nagamachi and N. Mugibayashi, Publ. RIMS Kyoto Univ. **12** Suppl., 309 (1977).
[7]S. Nagamachi and N. Mugibayashi, Rep. Math. Phys. **16**, 181 (1979).
[8]A. S. Wightman, Math. Anal. Appl. B **7**, 769 (1981).
[9]A. S. Wightman, Phys. Scr. **24**, 813 (1981).

[10]A. Jaffe, Phys. Rev. **158**, 1454 (1961).
[11]M. Z. Iofa and V. Ya. Fainberg, Sov. Phys. JETP **29**, 880 (1969).
[12]F. Constantinescu, J. Math. Phys. **12**, 293 (1971).
[13]W. Lücke, in *Proceedings of the XIII International Conference on Differential Geometric Methods in Theoretical Physics*, Shumen, Bulgaria, 1984, edited by H. D. Doebner and T. D. Palev (World Scientific, Singapore, 1986), pp. 163–169.
[14]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 2.
[15]J. Bümmerstede and W. Lücke, J. Math. Phys. **16**, 1203 (1975).
[16]W. Lücke, J. Math. Phys. **17**, 1515 (1976).
[17]T. Kawai, J. Fac. Sci. Univ. Tokyo IA **17**, 467 (1970).
[18]S. Nagamachi, Publ. RIMS Kyoto Univ. **17**, 25 (1981).
[19]L. Hörmander, *Linear Partial Differential Operators* (Springer, Berlin, 1963).
[20]M. Sato, T. Kawai, and M. Kashiwara, *Lecture Notes in Mathematics*, Vol. 287 (Springer, Berlin, 1973), pp. 264–529.
[21]H. Schlichtkrull, *Hyperfunctions and Harmonic Analysis on Symmetric Space* (Birkhäuser, Boston, 1984).
[22]M. Sato, J. Fac. Sci. Univ. Tokyo, Sect. I **8**, 387 (1959).
[23]G. Bredon, *Sheaf Theory* (McGraw–Hill, New York, 1967).
[24]P. Schapira, *Lecture Notes on Mathematics*, Vol. 126 (Springer, Berlin, 1970).
[25]L. Hörmander, *The Analysis of Linear Partial Differential Operators I* (Springer, Berlin, 1983).
[26]T. Matsuzawa, Nagoya Math. J. **108**, 67 (1987).
[27]Y. Ito, J. Math. Tokushima Univ. **15**, 1 (1981).
[28]M. Morimoto and K. Yoshino, Proc. Jpn. Acad. Ser. A **56**, 357 (1980).
[29]M. Morimoto, *Lecture Notes in Mathematics*, Vol. 287 (Springer, Berlin, 1973), pp. 41–81.
[30]R. Jost and K. Hepp, Helv. Phys. Acta **35**, 34 (1962).
[31]D. Roberts, Princeton University thesis, 1983 (unpublished).
[32]R. Jost, *Lextures on Field Theory and the Many-body Problem*, edited by E. R. Caianiello (Academic, New York, 1961), pp. 127–145.
[33]H. Epstein, Nuovo Cimento **27**, 886 (1963).
[34]S. Nagamachi and N. Mugibayashi, J. Math. Phys. **27**, 832 (1986).
[35]M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1975), Vol. 2.

Killing–Yano tensors and variable separation in Kerr geometry

E. G. Kalnins
Mathematics Department, University of Waikato, Hamilton, New Zealand
W. Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
G. C. Williams
Mathematics Department, University of Waikato, Hamilton, New Zealand

(Received 1 June 1988; accepted for publication 29 March 1989)

A complete analysis of the free-field massless spin-s equations ($s = 0, \frac{1}{2}, 1$) in Kerr geometry is given. It is shown that in each case the separation constants occurring in the solutions obtained from a potential function can be characterized in an invariant way. This invariant characterization is given in terms of the Killing–Yano tensor admitted by Kerr geometry.

I. INTRODUCTION AND MATHEMATICAL PRELIMINARIES

A complete understanding of the characterization of solutions of spin-s free-field equations in Kerr geometry has yet to be achieved. Interest in these equations originated with the investigations of Teukolsky,[1] who showed that in the Newman–Penrose[2] formalism separable solutions were possible for certain Maxwell and Weyl scalars in Kerr geometry.[3] (Kerr geometry is the space-time geometry of the gravitational background due to a rotating black hole.)

Chandrasekhar[4] has shown that these results can be extended to the Dirac equation. These results have been further extended[5,6] and shown to hold for more general classes of space-time. In the original work of Carter[7] it was established that the Hamilton–Jacobi and Schrödinger equations admitted a solution for the Kerr geometry via standard separation of variables techniques. Because of this property, Kerr space-time admits a quadratic constant of the motion in addition to the already known two Killing vector fields. However, the key property at the heart of the solution of the equations for spin-s $(0,\frac{1}{2},1)$ is the existence of a Killing–Yano tensor.[8] The role played by such a tensor for the solutions of the Dirac equation has been explained in Refs. 9 and 10. In this paper we indicate how this characterization works for massless particles with spins 0, $\frac{1}{2}$, and 1 and massive particles with spins-0, $\frac{1}{2}$. In so doing we clarify the role of the Killing–Yano tensor. The results for spin-1 are new and the treatment of spins-0, $\frac{1}{2}$, while not new, is presented in a unified way.

Once this work is extended we expect to better understand the methods by which a theory of "variable separation" can be constructed for general spin-s equations. Earlier work by the authors,[11] although not incorrect, did not succeed in giving an intrinsic characterization of the separation parameters appearing in the solution of Maxwell's equations. What was in fact achieved in Ref. 11 was a characterization of a particular choice of gauge. The contents of the present paper are arranged as follows. In Sec. I we outline the conventions and notations used, together with the relevant definitions and properties of Killing–Yano tensors. In Secs. II and III we deal with the zero-mass equations of spin-0, $\frac{1}{2}$, and 1, respectively.

In this paper we consistently use the spinor notation of Penrose and Rindler.[12] In addition, we employ the null tetrad formalism as described by Chandrasekhar.[4] Specifically, we restrict ourselves to the Kinnersley null tetrad of vectors with the components

$$\begin{aligned} l^i &= (1/\Delta)(r^2 + a^2,\Delta,0,a),\\ n^i &= (1/2\rho^2)(r^2 + a^2, -\Delta,0,a),\\ m^i &= (1/\sqrt{2}\bar{\rho})(ia\sin\theta,0,1,i\csc\theta),\\ \bar{m}^i &= (1/\sqrt{2}\bar{\rho}^*)(-ia\sin\theta,0,1,-i\csc\theta), \end{aligned} \tag{1.1}$$

where

$$\Delta = r^2 + a^2 - 2Mr,\ \rho^2 = r^2 + a^2\cos^2\theta,$$
$$\bar{\rho} = (r + ia\cos\theta).$$

The Kerr solution of the Einstein equations has the line element

$$\begin{aligned} ds^2 &= \left(1 - \frac{2Mr}{\rho^2}\right)dt^2 - \frac{\rho^2}{\Delta}dr^2 - \rho^2\,d\theta^2\\ &\quad + \frac{4aMr\sin^2\theta}{\rho^2}\,dt\,d\phi\\ &\quad - \left((r^2 + a^2) + \frac{2a^2Mr\sin^2\theta}{\rho^2}\right)\sin^2\theta\,d\phi^2. \end{aligned} \tag{1.2}$$

A Killing–Yano tensor $K_{AA'BB'}$ is a (skew symmetric) tensor satisfying

$$\nabla_{(CC'}K_{AA')BB'} = 0,\ K_{AA'BB'} + K_{BB'AA'} = 0. \tag{1.3}$$

The Killing–Yano tensor can also be equivalently represented in terms of the pair of symmetric Killing spinors $K_{AB'}$, $\bar{K}_{A'B'}$ via

$$K_{AA'BB'} = \tfrac{1}{2}(\epsilon_{A'B'}K_{AB} + \epsilon_{AB}\bar{K}_{A'B'}). \tag{1.4}$$

Conditions (1.3) are then equivalent to

$$\begin{aligned} &\nabla_{(AA'}K_{BC)} = 0,\ \nabla_{A(A'}\bar{K}_{B'C')} = 0,\\ &\nabla_{BA'}K_A{}^B + \nabla_{AB'}\bar{K}_{A'}{}^{B'} = 0. \end{aligned} \tag{1.5}$$

 0022-2488/89/102360-06$02.50

We have the following result: In Kerr space-time the equations for a Killing–Yano tensor have only one solution. The nonzero components of this tensor in the null tetrad formalism using the Kinnersley tetrad are

$$K_{00'11'} = ia\cos\theta,\ K_{01'10'} = r. \tag{1.6}$$

The Klein–Gordon equation for a spin-0 free-field is

$$\Box\phi = (\nabla_{AA'}\nabla^{AA'})\phi = m^2\phi. \tag{1.7}$$

In Newman–Penrose notation (1.7) has the form

$$\begin{aligned}\Box\phi = [&(D-\rho-\rho^*)\Delta + (\Delta-\gamma-\gamma^*+\mu+\mu^*)D \\ &-(\delta^*-\alpha+\beta^*-\tau^*+\pi)\delta \\ &-(\delta+\beta-\alpha^*-\tau+\pi^*)\delta^*]\phi = m^2\phi.\end{aligned} \tag{1.8}$$

In terms of the coordinates used to describe the line element (1.2) this equation reads

$$\begin{aligned}\Box\phi = (-1/2\rho^2)\{&\Delta(\mathscr{D}_1\mathscr{D}_0^+ + \mathscr{D}_1^+\mathscr{D}_0) \\ &+ (\mathscr{L}_1\mathscr{L}_0^+ + \mathscr{L}_1^+\mathscr{L}_0)\}\phi = m^2\phi.\end{aligned} \tag{1.9}$$

Equation (1.9) admits a separable solution

$$\phi = R_0(r)S_0(\theta)e^{im\phi + i\sigma t}, \tag{1.10}$$

where the separation equations are

$$\begin{aligned}&[\Delta(\mathscr{D}_1\mathscr{D}_0^+ + \mathscr{D}_1^+\mathscr{D}_0) + 2m^2r^2 + \lambda]R_0 = 0, \\ &[\mathscr{L}_1\mathscr{L}_0^+ + \mathscr{L}_0^+\mathscr{L}_0 + 2m^2a^2\cos^2\theta - \lambda]S_0 = 0.\end{aligned} \tag{1.11}$$

The directional derivatives in expression (1.11) are defined by

$$\begin{aligned}\mathscr{D}_n &= \partial_r + iK/\Delta + 2n(r-M)/\Delta, \\ \mathscr{D}_n^+ &= \partial_r - iK/\Delta + 2n(r-M)/\Delta, \\ \mathscr{L}_n &= \partial_\theta + Q + n\cot\theta, \\ \mathscr{L}_n^+ &= \partial_\theta - Q + n\cot\theta,\end{aligned} \tag{1.12}$$

where $K = (r^2+a^2)\sigma + am$ and $Q = a\sigma\sin\theta + m\csc\theta$.

From the theory of separation of variables for the Klein–Gordon equation it follows that there exists a second-order symmetry operator U such that

$$U\phi = \lambda\phi \tag{1.13}$$

for a separable solution ϕ. (We say that U is a symmetry operator if it commutes with $\Box$: $[\Box, U] = 0$.)

In terms of the Killing–Yano tensor,

$$\begin{aligned}U &= (K^{AA'BB'}\nabla_{BB'})(K_{AA'}{}^{CC'}\nabla_{CC'}) - K^{AA'BB'}M_{BB'}\nabla_{AA'} \\ &= (1/2\rho^2)[a^2\cos^2\theta[\Delta(\mathscr{D}_1\mathscr{D}_0^+ + \mathscr{D}_1^+\mathscr{D}_0)] \\ &\quad - r^2[\mathscr{L}_1\mathscr{L}_0^+ + \mathscr{L}_1^+\mathscr{L}_0]],\end{aligned} \tag{1.14}$$

where

$$M_{AB'} = \tfrac{1}{3}\nabla^{BA'}K_{BB'AA'}.$$

We also note here that the symmetric tensor

$$\mathscr{K}^{AA'CC'} = K^{AA'}{}_{BB'}K^{BB'CC'} \tag{1.15}$$

is a second-order Killing tensor satisfying the Killing equation

$$\nabla_{(AA'}\mathscr{K}_{BB'CC')} = 0. \tag{1.16}$$

This fact is crucial in the separability of the corresponding Hamilton–Jacobi and Schrödinger equations.

II. THE DIRAC EQUATION

In spinor notation the Dirac equation has the form

$$\nabla_{AX'}\chi^{X'} = (im_e\sqrt{2})\phi_A,\ \nabla_{AX'}\phi^A = -(im_e/\sqrt{2})\chi_{X'}. \tag{2.1}$$

Equations (2.1), when written in Newman–Penrose notation, are

$$\begin{aligned}&(D-\rho^*)\chi_{1'} - (\delta+\pi^*-\alpha^*)\chi_{0'} = (im_e/\sqrt{2})\phi_0, \\ &(\delta^*+\beta^*-\tau^*)\chi_{1'} - (\Delta+\mu^*-\gamma^*)\chi_{0'} = (im_e/\sqrt{2})\phi_1, \\ &(D-\rho)\phi_1 - (\delta^*+\pi-\alpha)\phi_0 = -(im_e/\sqrt{2})\chi_{0'}, \\ &(\delta+\beta-\tau)\phi_1 - (\Delta+\mu-\gamma)\phi_0 = -(im_e/\sqrt{2})\chi_{1'}.\end{aligned} \tag{2.2}$$

Chandrasekhar[4] found solutions of the form

$$\begin{aligned}\phi_1 &= (1/\bar\rho^*)R_{-1/2}S_{-1/2}e^{i\sigma t + im\phi}, \\ \phi_0 &= -R_{1/2}S_{1/2}e^{i\sigma t + im\phi}, \\ \chi_{1'} &= -(1/\bar\rho)R_{-1/2}S_{1/2}e^{i\sigma t + im\phi}, \\ \chi_{0'} &= -R_{1/2}S_{-1/2}e^{i\sigma t + im\phi}.\end{aligned} \tag{2.3}$$

The second-order separation equations are

$$\begin{aligned}&\{\Delta\mathscr{D}_{1/2}^+\mathscr{D}_0 - [im_e/(\lambda + im_e r)]\Delta\mathscr{D}_0 \\ &\quad - (\lambda^2 + m_e^2r^2)\}R_{-1/2} = 0, \\ &\{\Delta\mathscr{D}_{1/2}\mathscr{D}_0^+ + [im_e/(\lambda - im_e r)]\Delta\mathscr{D}_0^+ \\ &\quad - (\lambda^2 + m_e^2r^2)\}\Delta^{1/2}R_{1/2} = 0, \\ &\{\mathscr{L}_{1/2}\mathscr{L}_{1/2}^+ + [am_e\sin\theta/(\lambda + am_e\cos\theta)]\mathscr{L}_{1/2}^+ \\ &\quad + (\lambda^2 - a^2m_e^2\cos^2\theta)\}S_{-1/2} = 0, \\ &\{\mathscr{L}_{1/2}^+\mathscr{L}_{1/2} - [am_e\sin\theta/(\lambda - am_e\cos\theta)]\mathscr{L}_{1/2} \\ &\quad + (\lambda^2 - a^2m_e^2\cos^2\theta)\}S_{1/2} = 0.\end{aligned} \tag{2.4}$$

The separated solutions satisfy the eigenvalue equations

$$\begin{aligned}L_{AA'}\chi^{A'} &= (K_{AA'}{}^{BB'}\nabla_{BB'} - M_{AA'})\chi^{A'} = (\lambda/\sqrt{2})\phi_A, \\ N_{AA'}\phi^A &= (K_{AA'}{}^{BB'}\nabla_{BB'} + M_{AA'})\phi^A = (\lambda/\sqrt{2})\chi_{A'}.\end{aligned} \tag{2.5}$$

From Eqs. (2.5) follows the conditions

$$\begin{aligned}&[\nabla_{AA'}L^{AX'} + N_{AA'}\nabla^{AX'}]\chi_{X'} = 0, \\ &[\nabla^{CA'}N_{AA'} + L^{CA'}\nabla_{AA'}]\phi^A = 0.\end{aligned} \tag{2.6}$$

From (2.5) we can construct the operator

$$\Lambda = \begin{bmatrix} 0 & L_A^{A'} \\ N_{A'}^A & 0 \end{bmatrix} \tag{2.7}$$

acting on the Dirac spinors

$$\begin{bmatrix}\phi_A \\ \chi_{A'}\end{bmatrix}.$$

The operator (2.7) anticommutes with the Dirac Hamiltonian

$$H = \begin{bmatrix} (im_e/\sqrt{2})\epsilon_B{}^A & -\nabla_B{}^{A'} \\ \nabla^A{}_{B'} & -(im_e/\sqrt{2})\epsilon_{B'}{}^{A'} \end{bmatrix}. \tag{2.8}$$

The proof of relations (2.6) is instructive; we now prove the first of these relations. Consider the operator

$$Q_{A'C'} = N_{AA'}\nabla^A{}_{C'} - \nabla^A{}_{A'}L_{AC'} \tag{2.9}$$

using

$$\nabla_{AA'}K_{BC} = \epsilon_{AB}M_{CA'} + \epsilon_{AC}M_{BA'},$$
$$\nabla_{AA'}\bar{K}_{B'C'} = -\epsilon_{A'B'}M_{AC'} - \epsilon_{A'C'}M_{AB'}. \tag{2.10}$$

We find that

$$Q_{A'C'} = \tfrac{1}{2}[K_A{}^B\nabla_{BA'}\nabla^A{}_{C'} + \bar{K}_{A'}{}^{B'}\nabla_{AB'}\nabla^A{}_{C'} - K_A{}^B\nabla^A{}_{A'}\nabla_{BC'} - \bar{K}_{C'}{}^{B'}\nabla^A{}_{A'}\nabla_{AB'} + (\epsilon_{A'C'}M^{AB'} + \epsilon_{A'}{}^{B'}M^A{}_{C'})\nabla_{AB'} + 3M^B{}_{A'}\nabla_{BC'}] + M_{AA'}\nabla^A{}_{C'} + M_{AC'}\nabla^A{}_{A'} + (\nabla^A{}_{A'}M_{AC'}). \tag{2.11}$$

Noting that

$$K_A{}^B\nabla_{BA'}\nabla^A{}_{C'} = K_A{}^B\nabla^A{}_{A'}\nabla_{BC'} \tag{2.12}$$

since K_{AB} is symmetric,

$$\begin{aligned}\bar{K}_{A'}{}^{B'}\nabla_{AB'}\nabla^A{}_{C'} - \bar{K}_{C'}{}^{B'}\nabla^A{}_{A'}\nabla_{AB'} &= \tfrac{1}{2}\bar{K}_{A'}{}^{B'}(\nabla_{AB'}\nabla^A{}_{C'} + \nabla^A{}_{C'}\nabla_{AB'} + [\nabla_{AB'},\nabla^A{}_{C'}]) \\ &\quad - \tfrac{1}{2}\bar{K}_{C'}{}^{B'}(\nabla^A{}_{A'}\nabla_{AB'} + \nabla_{AB'}\nabla^A{}_{A'} + [\nabla^A{}_{A'},\nabla_{AB'}]) \\ &= \tfrac{1}{2}\epsilon_{A'C'}\bar{K}_{D'}{}^{B'}(\nabla_{AB'}\nabla^{AD'} + \nabla^{AD'}\nabla_{AB'}) \\ &\quad + \tfrac{1}{2}\bar{K}_{A'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{C'}] + \tfrac{1}{2}\bar{K}_{C'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{A'}] \\ &= \tfrac{1}{2}\bar{K}_{A'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{C'}] + \tfrac{1}{2}\bar{K}_{C'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{A'}],\end{aligned} \tag{2.13}$$

$$\epsilon_{A'C'}M^{AB'}\nabla_{AB'} = M^A{}_{C'}\nabla_{AA'} - M^A{}_{A'}\nabla_{AC'}, \tag{2.14}$$

we can write

$$Q_{A'C'} = \tfrac{1}{4}\bar{K}_{A'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{C'}] + \tfrac{1}{4}\bar{K}_{C'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{A'}] + (\nabla^A{}_{A'}M_{AC'}). \tag{2.15}$$

Now consider

$$\begin{aligned}\nabla_{AA'}M_{BB'} &= \tfrac{1}{3}\nabla_{AA'}\nabla_{CB'}K^C{}_B \\ &= \tfrac{1}{3}(\nabla_{CB'}\nabla_{AA'} + [\nabla_{AA'},\nabla_{CB'}])K^C{}_B \\ &= \tfrac{1}{3}(\nabla_{AB'}M_{BA'} + \epsilon_{AB}\nabla_{CB'}M^C{}_{A'} + \epsilon_{A'B'}\Psi_{ABCD}K^{CD}),\end{aligned} \tag{2.16}$$

from which the following results can be obtained:

$$\nabla_{(A(A'}M_{B)B')} = 0, \quad \nabla_{AA'}M^{AA'} = 0,$$
$$\nabla_{(AA'}M_{B)}{}^{A'} = \tfrac{1}{2}\Psi_{ABCD}K^{CD} = W_{AB}, \quad \text{defining } W_{AB}. \tag{2.17}$$

Note that we can also write $\nabla_{AA'}M_{BB'} = -\tfrac{1}{3}\nabla_{AA'}\nabla_{BC'}\bar{K}^{C'}{}_{B'}$ and proceed in a similar manner as before to obtain the additional result

$$\nabla_{A(A'}M^A{}_{B')} = -\tfrac{1}{2}\bar{\Psi}_{A'B'C'D'}\bar{K}^{C'D'} = -\bar{W}_{A'B'}, \quad \text{defining } \bar{W}_{A'B'}. \tag{2.18}$$

Now since (by reducing to symmetric spinors) we can write for any $T_{ABA'B'}$,

$$T_{ABA'B'} = T_{(AB)(A'B')} + \tfrac{1}{2}\epsilon_{A'B'}T_{(AB)K'}{}^{K'} + \tfrac{1}{2}\epsilon_{AB}T_K{}^K{}_{(A'B')} + \tfrac{1}{4}\epsilon_{AB}\epsilon_{A'B'}T_K{}^K{}_{K'}{}^{K'} \tag{2.19}$$

it follows that

$$\nabla_{AA'}M_{BB'} = \tfrac{1}{2}\epsilon_{A'B'}W_{AB} - \tfrac{1}{2}\epsilon_{AB}\bar{W}_{A'B'}. \tag{2.20}$$

We also note in passing that $\nabla_{AA'}M_{BB'}$ is a skew-symmetric tensor, i. e., M_a satisfies

$$\nabla_{(b}M_{a)} = 0, \tag{2.21}$$

i.e., M_a is a Killing vector.

Returning to the operator $Q_{A'C'}$, we can now write

$$Q_{A'C'} = \tfrac{1}{4}\bar{K}_{A'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{C'}] + \tfrac{1}{4}\bar{K}_{C'}{}^{B'}[\nabla_{AB'},\nabla^A{}_{A'}] + \bar{W}_{A'C'}, \tag{2.22}$$

from which its action on a spinor $\phi^{C'}$ is as follows:

$$Q_{A'C'}\phi^{C'} = \tfrac{1}{4}\bar{K}_{A'}{}^{B'}\epsilon_A{}^A\bar{\Psi}_{B'C'}{}^{C'}{}_{M'}\phi^{M'} + \tfrac{1}{4}\bar{K}_{C'}{}^{B'}\epsilon_A{}^A\bar{\Psi}_{B'A'}{}^{C'}{}_{M'}\phi^{M'} + \tfrac{1}{2}\bar{\Psi}_{A'C'L'M'}\bar{K}^{L'M'}\phi^{C'} = 0. \tag{2.23}$$

Thus

$$N_{AA'}\nabla^A{}_{C'}\phi^{C'} = \nabla^A{}_{A'}L_{AC'}\phi^{C'} \tag{2.24}$$

If we consider only the neutrino equation $\nabla^{AA'}\phi_A = 0$ for the case in which $m_e = 0$ (massless spin-$\tfrac{1}{2}$), then the separation constant λ^2 stems from the eigenvalue equation

$$L_B{}^{A'}N^A{}_{A'}\phi_A = (\lambda^2/2)\phi_B. \tag{2.25}$$

III. THE MAXWELL EQUATIONS

For the case of the Maxwell equations corresponding to mass-zero spin-1 the characterization of separation parameters in terms of the components of the Killing–Yano tensor can also be achieved. Maxwell's equations are commonly formulated in terms of the skew-symmetric energy momentum tensor $F_{AA'BB'}$, which satisfies

$$\nabla_{AA'}F_{BB'CC'} + \nabla_{CC'}F_{AA'BB'} + \nabla_{BB'}F_{CC'AA'} = 0,$$
$$\nabla^{AA'}F_{AA'BB'} = 0, \quad F_{AA'BB'} + F_{BB'AA'} = 0. \tag{3.1}$$

As with the case of Killing–Yano tensor, $F_{AA'BB'}$ can be realized via the symmetric spinors ϕ_{AB}, $\phi_{A'B'}$ according to

$$F_{AA'BB'} = \epsilon_{AB}\bar{\phi}_{A'B'} + \epsilon_{A'B'}\phi_{AB}. \tag{3.2}$$

In terms of these symmetric spinors, Maxwell's equations have the form

$$\nabla^A{}_{A'}\phi_{AB} = 0, \tag{3.3a}$$
$$\nabla_A{}^{C'}\bar{\phi}_{C'B'} = 0. \tag{3.3b}$$

In Ref. 4 Chandrasekhar has obtained explicit solutions for these equations: viz.

$$\begin{aligned}&(D - 2\rho)\phi_{01} - (\delta^* + \pi - 2\alpha)\phi_{00} = 0, \\ &(D - \rho)\phi_{11} - (\delta^* + 2\pi)\phi_{01} = 0, \\ &(\delta - 2\tau)\phi_{01} - (\Delta + \mu - 2\gamma)\phi_{00} = 0, \\ &(\delta - \tau + 2\beta)\phi_{11} - (\Delta + 2\mu)\phi_{01} = 0.\end{aligned} \tag{3.4}$$

From the crucial observation that

$$\bar{\rho}(\delta - 2\tau)(D - 2\rho) = (D - 2\rho)\bar{\rho}(\delta - 2\tau), \tag{3.5}$$

Teukolsky[1] deduced that if $\phi_{00} = \psi_{00}e^{i\sigma t + im\phi}$, then the function ψ_{00} satisfies

$$[\Delta\mathscr{D}_1\mathscr{D}_1^+ + \mathscr{L}_0^+\mathscr{L}_1 - 2i\sigma\bar{\rho}]\psi_{00} = 0. \tag{3.6}$$

This function admits a separable solution $\psi_{00} = R_1S_1$, where the separation equations are

$$(\Delta\mathscr{D}_1\mathscr{D}_1^+ - 2i\sigma r - \lambda)R_1 = 0,$$
$$(\mathscr{L}_0^+\mathscr{L}_1 + 2a\sigma\cos\theta + \lambda)S_1 = 0. \tag{3.7}$$

If Eqs. (3.4) are analyzed further and we write

$$\phi_{11} = (2(\bar\rho^*)^2)^{-1}\psi_{11}e^{i\sigma t + im\phi},$$

we find that the function ψ_{11} satisfies

$$[\Delta\mathscr{D}_0^+\mathscr{D}_0 + \mathscr{L}_0\mathscr{L}_1^+ + 2i\sigma\bar\rho]\psi_{11} = 0, \tag{3.8}$$

which admits a separable solution $\psi_{11} = R_{-1}S_{-1}$ with the separation equations

$$\begin{aligned}&[\Delta\mathscr{D}_0^+\mathscr{D}_0 + 2i\sigma r + \lambda]R_{-1} = 0,\\ &[\mathscr{L}_0\mathscr{L}_1^+ - 2a\sigma\cos\theta - \lambda]S_{-1} = 0.\end{aligned} \tag{3.9}$$

Equations (3.7) and (3.9) were first derived by Teukolsky.[1] The functions $R_{\pm 1}$, $S_{\pm 1}$ are called Teukolsky functions by Chandrasekhar. If instead of $R_{\pm 1}$ we choose the function $P_{-1} = R_{-1}$, $P_{+1} = \Delta R_{+1}$, then the functions exhibit interesting properties, which are summarized in the Appendix. Chandrasekhar proceeded further and showed that ϕ_{01} can be written in the form

$$\begin{aligned}\phi_{01} = {}&(1/\sqrt{2}\bar\rho^*\mathscr{C})[\mathscr{D}_0\mathscr{L}_1 - (1/\bar\rho^*)\\ &\times(\mathscr{L}_1 + ia\sin\theta\mathscr{D}_0)]P_{-1}S_{+1},\end{aligned} \tag{3.10}$$

where $\mathscr{C}$ is as in (A1).

We now seek the invariant characterization of the parameters λ and $\mathscr{C}$. To determine this we draw on the results of Cohen and Kegeles,[12] who showed how to obtain solutions of (3.3) via the use of a Debye potential $\bar P^{X'Y'}$ and a gauge degree of freedom $G_B^{\ W'}$. If these functions satisfy

$$\nabla^{A(M'}\nabla_{AW'}\bar P^{N')W'} = 2\nabla^{A(M'}G_A^{\ N')}, \tag{3.11}$$

then

$$\phi_{AB} = \nabla_{(AW'}\nabla_{B)X'}\bar P^{W'X'} - 2\nabla_{(AW'}G_{B)}^{\ \ W'} \tag{3.12}$$

is a solution of (3.3). More specifically, if one chooses $G_A^{\ W'} = -U_{AA'}\bar P^{A'W'}$, where

$$U_{00'} = \rho^*_1,\ U_{10'} = \tau^*_1,\ U_{01'} = -\pi^*,\ U_{11'} = -\mu^*,$$

then $\bar P^{A'W'}$ satisfies the decoupled equation

$$\begin{aligned}&[(1-p/2)(\Delta + \mu - \gamma + (1-2p)\gamma^* - p\mu^*)(D + (1-p)\rho^*)\\ &\quad - (1-p/2)(\delta + \beta - \tau + (1-2p)\alpha^* - p\pi^*)(\delta^* + 2(1-p)\beta^* + (1-p)\tau^*)\\ &\quad - (p/2)(\delta^* + \pi - \alpha + (3-2p)\beta^* + (2-p)\tau^*)(\delta + 2(1-p)\alpha^* - (p-1)\pi^*)\\ &\quad + (p/2)(D - \rho + (2-p)\rho^*)(\Delta + 2(1-p)\gamma^* - (p-1)\mu^*)]\bar P^{A'W'} = 0,\end{aligned} \tag{3.13}$$

where p is the number of "ones" appearing in the indices of $\bar P^{A'W'}$.

The choice of $G_A^{\ W'}$ made above is particularly interesting since it yields three equivalent representations for the same function, viz.

(i) $p = 0$, $\bar P^{0'0'} = P_{-1}S_1 e^{i\sigma t + im\phi}$:

$$\begin{aligned}&\phi_{00} = \mathscr{D}_0\mathscr{D}_0\bar P^{0'0'},\\ &\phi_{01} = (1/\sqrt{2}\bar\rho^*)[\mathscr{D}_0\mathscr{L}_1 - (1/\bar\rho^*)\\ &\qquad\times(\mathscr{L}_1 + ia\sin\theta\mathscr{D}_0)]\bar P^{0'0'},\\ &\phi_{11} = [1/2(\bar\rho^*)^2]\mathscr{L}_0\mathscr{L}_1\bar P^{0'0'}.\end{aligned} \tag{3.14}$$

(ii) $p = 2$, $\bar P^{1'1'} = (\bar\rho)^2\Delta^{-1}P_{+1}S_{-1}e^{i\sigma t + im\phi}$:

$$\begin{aligned}&\phi_{00} = \mathscr{L}_0^+\mathscr{L}_1^+\bar P^{1'1'},\\ &\phi_{01} = -(\Delta/\sqrt{2}\bar\rho^*)[\mathscr{D}_1^+\mathscr{L}_1^+ - (1/\bar\rho^*)\\ &\qquad\times(\mathscr{L}_1^+ + ia\sin\theta\mathscr{D}_1^+)]\bar P^{1'1'},\\ &\phi_{11} = [\Delta^2/2(\bar\rho^*)^2]\mathscr{D}_1^+\mathscr{D}_1^+\bar P^{1'1'}.\end{aligned} \tag{3.15}$$

From the identities given in the Appendix it is straightforward to establish that (3.14) and (3.15) are representations of the same functions ϕ_{AB}.

(iii) $p = 1$; in this case $\bar P^{0'1'}$ satisfies

$$\left[\Delta\left(\mathscr{D}_1^+ - \frac{1}{\bar\rho}\right)\left(\mathscr{D}_0 + \frac{1}{\bar\rho}\right) + \left(\mathscr{L}_1^+ + \frac{ia\sin\theta}{\rho}\right)\times\left(\mathscr{L}_0 - \frac{ia\sin\theta}{\bar\rho}\right)\right]\frac{1}{\bar\rho}\bar P^{0'1'} = 0.$$

An examination of this equation shows that $\bar P^{0'1'}$ satisfies the same equation as $\bar\rho\bar\phi_{0'1'}$. Hence a solution may be taken to be

$$\bar P^{0'1'} = (1/\sqrt{2}\mathscr{C})[\mathscr{D}_0\mathscr{L}_1^+ - (1/\rho)(\mathscr{L}_1^+ - ia\sin\theta\mathscr{D}_0)]\times P_{-1}S_{-1}e^{i\sigma t + im\phi}.$$

With this choice the components of ϕ_{AB} can be written as

$$\begin{aligned}&\phi_{00} = (\sqrt{2}/\bar\rho)\mathscr{D}_0\mathscr{L}_0^+\bar P^{0'1'},\\ &\phi_{01} = \frac{-1}{4\rho^2}\Big[\Delta\left(\mathscr{D}_1 - \frac{2}{\bar\rho^*}\right)\mathscr{D}_0^+ + \Delta\left(\mathscr{D}_1^+ - \frac{2}{\bar\rho^*}\right)\mathscr{D}_0\\ &\qquad + \left(\mathscr{L}_1^+ - \frac{ia\sin\theta}{\bar\rho}\right)\mathscr{L}_0\Big]\\ &\qquad + \left(\mathscr{L}_1^+ \frac{ia\sin\theta}{\bar\rho}\right)\mathscr{L}_0^+\Big]\bar P^{0'1'},\\ &\phi_{11} = (\Delta/\sqrt{2}\rho^2\bar\rho^*)\mathscr{D}_0^+\mathscr{L}_0\bar P^{0'1'}.\end{aligned} \tag{3.16}$$

Again, using the identities in the Appendix it can be verified that expression (3.16) for ϕ_{AB} is identical to those given when $p = 0$ or 2.

These representations (and the corresponding ones for $\phi_{A'B'}$) are invaluable for the proof of our principal result.

Theorem: If the functions ϕ_{AB} are the solutions of $\nabla^A{}_{A'}\phi_{AB} = 0$ as represented by, say, (3.14), then the parameters λ and $\mathscr{C}$ are intrinsically defined via the relations

$$\begin{aligned}&C^{AB}{}_{A'B'}\phi_{AB}\\ &\quad = (K^A{}_{(A'}{}^{CC'}K^B{}_{B')}{}^{DD'}\nabla_{CC'}\nabla_{DD'}\\ &\qquad + 4M^A{}_{(A'}K^B{}_{B')}{}^{DD'}\nabla_{DD'} + 2M^A{}_{(A'}M^B{}_{B')})\phi_{AB}\\ &\quad = \tfrac{1}{2}\mathscr{C}\phi_{A'B'},\end{aligned} \tag{3.17}$$

$$\Lambda_{(A}{}^{B}\phi_{C)B} = (K_A{}^{A'EE'}\nabla_{EE'} - M_A{}^{A'}) \times (K^B{}_{A'}{}^{DD'}\nabla_{DD'} + 2M^B{}_{A'})\phi_{BC} = (\lambda/2)\phi_{AC}. \quad (3.18)$$

The proof of relations (3.17) and (3.18) is (in principle) straightforward. The use of the algebraic computing language MACSYMA has been particularly useful in this respect. For relations (3.17) the given result can simply be verified using identities (A4) of the Appendix.

Relation (3.18) is somewhat more difficult. For the cases when $A = C = 0$ or 1 the result is relatively straightforward to establish. However, the result when $A = 0$, $C = 1$ requires extensive computation. In particular, the verification of the identity

$$\frac{a^2\cos^2\theta\Delta}{4\rho^2\bar{\rho}^*}(\mathscr{D}_1\mathscr{D}_0^+ + \mathscr{D}_1^+\mathscr{D}_0)\bar{\rho}^*\phi_{10} - \frac{r^2}{4\rho^2\bar{\rho}^*}(\mathscr{L}_1\mathscr{L}_0^+\mathscr{L}_1^+\mathscr{L}_0)\bar{\rho}^*\phi_{10} + \frac{Ma^2\cos^2\phi}{4(\bar{\rho}^*)^3}\phi_{10} - \frac{iar\cos\theta(\bar{\rho}^*+5\bar{\rho})}{4\rho^2\bar{\rho}^*}\phi_{10} - \frac{a^2}{(\bar{\rho}^*)^2}\phi_{10} - \frac{ia\sin\theta}{2(\bar{\rho}^*)^3}\phi_{00} + \frac{ia\sin\theta}{2(\bar{\rho}^*)}\phi_{11} = \frac{\lambda}{2}\phi_{01} \quad (3.19)$$

is nontrivial.

It can also be verified that the following holds.

(i) If ϕ_{AB} is a solution of (3.3a), then so is $\phi'_{AB} = \Lambda_{(A}{}^{C}\phi_{B)C}$.

In fact,

$$\nabla^{CC'}\Lambda_{(C}{}^{A}\phi_{B)A} = [\Lambda_B{}^A\epsilon_{A'}{}^{C'} - (K_B{}^{C'EE'}\nabla_{EE'} - M_B{}^{C'})M^A{}_{A'} + M_B{}^{C'}(K^A{}_{A'}{}^{DD'}\nabla_{DD'} + 2M^A{}_{A'}) - 2M_B{}^{C'}M^A{}_{A'} + \tfrac{1}{2}\nabla_B{}^{C'}K^A{}_{A'}{}^{DD'}M_{DD'}]\nabla^{CA'}\phi_{AC}. \quad (3.20)$$

(ii) If $\phi_{A'B'}$ is a solution of (3.3b), then $\phi'_{AB} = \mathrm{C}^{A'B'}{}_{AB}\phi_{A'B'}$ is a solution of (3.3a). Moreover,

$$\phi'_{C'D'} = \mathrm{C}_{C'D'}{}^{AB}\mathrm{C}^{A'B'}{}_{AB}\phi_{A'B'} = \mathbb{C}_{C'D'}{}^{A'B'}\phi_{A'B'}$$

is a solution of (3.3b) if $\phi_{A'B'}$ is a solution as well.

(iii) If ϕ_{AB} is a solution of (3.3a) then

$$\Lambda_C{}^F\mathbb{C}_{FD}{}^{KL}\phi_{KL} + \Lambda_D{}^F\mathbb{C}_{FC}{}^{KL}\phi_{KL} = \mathbb{C}_{CD}{}^{AB}[\Lambda_A{}^E\phi_{EB} + \Lambda_B{}^E\phi_{EA}].$$

The operator $\mathrm{C}_{AB}^{A'B'}$ is essentially the operator introduced by Torres del Castillo.[13]

We take the opportunity here to give a more complete discussion of the vector potential $A_{BB'}$ which gives rise to the corresponding $F_{AA'BB'}$:

$$F_{CC'BB'} = \nabla_{CC'}A_{BB'} - \nabla_{BB'}A_{CC'}. \quad (3.21)$$

As is well known, the choice of vector potential is not unique. A derivative of a gauge function can always be added according to $A_{CC'} \to A_{CC'} + \nabla_{CC'}\phi$. As in Ref. 11, we choose the gauge in which the components $A_{CC'}$ are divergenceless; then these functions satisfy

$$\Box A_{CC'} = (\nabla^{BB'}\nabla_{BB'})A_{CC'} = 0, \quad \nabla^{BB'}A_{BB'} = 0. \quad (3.22)$$

There are two independent solutions for the above equation which correspond to the same $F_{AA'BB'}$. These solutions are the analogs of electric and magnetic multipoles,[14]

$$A_{00'} = [P_{+1}(\mathscr{L}_1 S_{+1} - \mathscr{L}_1^+ S_{-1})\Delta^{-1}]e^{i\sigma t + im\phi},$$
$$A_{11'} = [P_{-1}(\mathscr{L}_1^+ S_{-1} - \mathscr{L}_1 S_{+1})(2\rho^2)^{-1}]e^{i\sigma t + im\phi},$$
$$A_{01'} = -(\mathscr{D}_0^+ P_{+1} + \mathscr{D}_0 P_{-1})S_{+1}(\sqrt{2}\bar{\rho})^{-1}e^{i\sigma t + im\phi},$$
$$A_{10'} = (\mathscr{D}_0^+ P_{+1} + \mathscr{D}_0 P_{-1})S_{-1}(\sqrt{2}\bar{\rho}^*)^{-1}e^{i\sigma t + im\phi}; \quad (3.23a)$$

$$A_{00'} = [P_{+1}(ia\cos\theta\mathscr{L}_1^+ + ia\sin\theta)S_{-1}\Delta^{-1}]e^{i\sigma t + im\phi},$$
$$A_{11'} = [P_{-1}(ia\cos\theta\mathscr{L}_1 + ia\sin\theta)S_{+1}(2\rho^2)^{-1}]e^{i\sigma t + im\phi},$$
$$A_{01'} = -(r\mathscr{D}_0 - 1)P_{-1}S_{+1}e^{i\sigma t + im\phi},$$
$$A_{10'} = -(r\mathscr{D}_0^+ - 1)P_{+1}S_{-1}e^{i\sigma t + im\phi}. \quad (3.23b)$$

Indeed, the (3.23a) corresponds to electric multipoles and (3.23b) corresponds to magnetic multipoles. In establishing (3.23a) use was made of the identity

$$\mathscr{C}[(\mathscr{L}_1 - ia\sin\theta\mathscr{D}_0^+)P_{+1}S_{+1} + (\mathscr{L}_1^+ - ia\sin\theta\mathscr{D}_0)P_{-1}S_{+1}] = i\sigma\bar{\rho}[\bar{\rho}^*\mathscr{D}_0\mathscr{L}_1 - (\mathscr{L}_1 + ia\,\theta\mathscr{D}_0)]P_{-1}S_{+}1.$$

It should be noted that the method of Cohen and Kegeles[12] also gives expressions for the vector potential. More specifically, the vector

$$A_{CC'} = (\nabla_{CE'}\bar{P}^{E'}{}_{C'} - 2G_{CC'}) + \text{complex conjugate}$$

is such that $A_{CC'}$ is a solution of the Maxwell equations. However, the choice of functions $\bar{P}^{X'Y'}$ and $G_A^{Y'}$ given previously does not lead to solutions in the divergence-free gauge.

IV. CONCLUSION

In this paper we have explicitly shown how the separation parameters that occur for spin-$s = 0, \frac{1}{2}, 1$ equations can be intrinsically characterized in terms of covariant operators whose coefficients can be written in terms of the Killing–Yano tensor and its covariant derivatives. In Minkowski space we subsequently show that these characterizations and their natural generalizations hold true for any s. There are well-known difficulties with the generalizations of equations of type (3.3).[15] In this respect it is our intention to examine the nature of the intrinsic operator characterization of the functions $A_{CC'}$ and their generalizations for higher spin. All these results provide a nontrivial example of solutions to spin-s equations. Ideally, a suitable theory of such solutions to this type of equation would enable us to derive the existence of such solutions from intrinsic properties only.

ACKNOWLEDGMENT

One of the authors (WM) would like to acknowledge partial support by the National Science Foundation under Grant DMS 86-00372.

APPENDIX: SUMMARY OF THE CHANDRASEKHAR[4] RESULTS

Chandrasekhar,[4] in his treatment of electromagnetic waves in Kerr geometry, has thoroughly developed the properties of the Teukolsky[1] functions. We summarize his results in the following theorems.

Theorem A1: For a suitable choice of the relative normalization of the functions $P_{\pm 1}$ it is possible to arrange that

$$\Delta\mathscr{D}_0\mathscr{D}_0 P_{-1} = \mathscr{C} P_{+1}, \quad \Delta\mathscr{D}_0^+\mathscr{D}_0^+ P_{+1} = \mathscr{C} P_{-1}, \tag{A1}$$

where

$$\mathscr{C}^2 = \lambda^2 - 4(a^2\sigma^2 + am\sigma).$$

Theorem A2: If the functions $S_{\pm 1}$ are normalized to unity,[4] then it is possible to arrange that

$$\mathscr{L}_0\mathscr{L}_1 S_{+1} = \mathscr{C} S_{-1}, \quad \mathscr{L}_0^+\mathscr{L}_1^+ S_{-1} = \mathscr{C} S_{+1}, \tag{A2}$$

with $\mathscr{C}$ as in Theorem A1.

Corollary: The derivatives of the functions $P_{\pm 1}$ and $S_{\pm 1}$ can again be expressed as combinations of the same functions:

$$\begin{aligned} &\mathscr{D}_0^+ P_{+1} = (-i/2K)[(\lambda + 2i\sigma r)P_{+1} - \mathscr{C} P_{-1}], \quad \mathscr{D}_0 P_{-1} = (i/2K)[(\lambda - 2i\sigma r)P_{-1} - \mathscr{C} P_{+1}], \\ &\mathscr{L}_1^+ S_{-1} = (-1/2Q)[(\lambda - 2a\sigma\cos\theta)S_{-1} + \mathscr{C} S_{+1}], \quad \mathscr{L}_1 S_{+1} = (1/2Q)[(\lambda + 2a\sigma\cos\theta)S_{+1} + \mathscr{C} S_{-1}]. \end{aligned} \tag{A3}$$

In addition to identities (A3) the following relations are instrumental in the establishment of (3.17):

$$\begin{aligned} &\mathscr{D}_0^+\mathscr{L}_0^+ = (\mathscr{C}/\Delta)\bar{\psi}_2 + \mathscr{L}_0^+(1/\bar{\rho}^*)[-\psi_1 + ia\sin\theta(1/\Delta)\psi_2], \\ &\mathscr{D}_0\mathscr{L}_0\psi_1 = \mathscr{C}\bar{\psi}_0 - \mathscr{L}_0(1/\bar{\rho}^*)(\psi_1 + ia\sin\theta\psi_0), \\ &\mathscr{L}_1^+\mathscr{L}_0\psi_1 = \mathscr{C}\bar{\psi}_1 + (1/\bar{\rho})(\mathscr{L}_1^+ - ia\sin\theta\mathscr{D}_0)\psi_2 - \mathscr{L}_1^+(1/\bar{\rho}^*)(\psi_2 + ia\sin\theta\psi_1), \\ &\Delta\mathscr{D}_1\mathscr{D}_0^+\psi_1 = -\mathscr{C}\bar{\psi}_1 - (1/\bar{\rho})(\mathscr{L}_1^+ - ia\sin\theta\mathscr{D}_0)\psi_2 - \mathscr{D}_0(1/\bar{\rho}^*)(\Delta\psi_1 - ia\sin\theta\psi_2), \\ &\Delta\mathscr{D}_1^+\mathscr{D}_0\psi_1 = -\mathscr{C}\bar{\psi}_1 + (1/\bar{\rho})(\mathscr{L}_1 - ia\sin\theta\mathscr{D}_0^+)\Delta\psi_0 - \mathscr{D}_0^+(\Delta/\bar{\rho}^*)(\psi_1 + ia\sin\theta\psi_0), \\ &\mathscr{L}_1\mathscr{L}_0^+\psi_1 = \mathscr{C}\bar{\psi}_1 - (1/\bar{\rho})(\mathscr{L}_1 - ia\sin\theta\mathscr{D}_0^+)\Delta\psi_0 + \mathscr{L}_1(1/\bar{\rho}^*)(\Delta\psi_0 - ia\sin\theta\psi_1), \\ &\Delta\mathscr{D}_1^+\mathscr{L}_1\psi_0 = -\mathscr{C}\bar{\psi}_1 + (1/\bar{\rho})(\mathscr{L}_1 - ia\sin\theta\mathscr{D}_0^+)\Delta\psi_0, \\ &\mathscr{D}_0\mathscr{L}_1^+\psi_2 = \mathscr{C}\bar{\psi}_1 + (1/\bar{\rho})(\mathscr{L}_1^+ - ia\sin\theta\mathscr{D}_0)\psi_2, \end{aligned} \tag{A4}$$

where

$$\psi_0 = 2\phi_{00}, \quad \psi = \sqrt{2}\bar{\rho}^*\phi_{01}, \quad \psi_2 = 2(\bar{\rho}^*)^2\phi_{11}.$$

[1] S. A. Teukolsky, "Rotating black holes. Separable wave equations for gravitational and electromagnetic perturbation," Phys. Rev. Lett. **29**, 1114 (1972).

[2] E. T. Newman and R. Penrose, "An approach to gravitational radiation by a method of spin coefficients," J. Math. Phys. **3**, 566 (1962).

[3] R. P. Kerr, "Gravitational field of spinning mass as an example of algebraically special metrics," Phys. Rev. Lett. **11**, 237 (1963).

[4] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., Oxford, 1984).

[5] N. Kamran and R. G. McLenaghan, "Separation of variables for higher spin zero mass field equations on type D vacuum backgrounds with cosmological constant, to be published in *Ivor Robinson Festschrift*, edited by W. Rindler and A. Trautman.

[6] R. Güven, "The solution of Dirac's equation in a class of type D backgrounds," Proc. R. Soc. London Ser. A **356**, 465 (1977).

[7] B. Carter, "Hamilton–Jacobi and Schrödinger separable solutions of Einstein's equations," Commun. Math. Phys. **10**, 280 (1968).

[8] K. Yano and S. Bochner, *Curvature and Betti Numbers*, Annals of Mathematics Studies #32 (Princeton U.P., Princeton).

[9] R. G. McLenaghan and P. Spindel, "Quantum numbers of Dirac fields in curved space-time," Phys. Rev. D **20**, 409 (1979).

[10] B. Carter and R. G. McLenaghan, "Generalized total angular momentum for the Dirac equation in curved space-time," Phys. Rev. D **19**, 1093 (1979).

[11] E. G. Kalnins, W. Miller, Jr., and G. C. Williams, "Electromagnetic waves in Kerr geometry," Proc. R. Soc. London Ser. A **408**, 23 (1986).

[12] R. Penrose and W. Rindler, *Spinors and Space-time. Vol. 1. Two-spinor Calculus and Relativistic Fields* (Cambridge U.P., London, 1984).

[13] J. M. Cohen and L. S. Kegeles, "Constructive procedure for perturbations in space-times," Phys. Rev. D **19**, 1641 (1979).

[14] G. F. Torres del Castillo, "Killing spinors and massless spinor fields," Proc. R. Soc. London Ser. A **400**, 119 (1985).

[15] M. E. Rose, *Multipole Fields* (Wiley, New York, 1955).

Topological aspects of a fermion and the chiral anomaly

Ashim Roy and Pratul Bandyopadhyay
Indian Statistical Institute, Calcutta-700035, India

(Received 22 February 1989; accepted for publication 26 April 1989)

It is here shown that the chiral anomaly is related to the topological properties of a fermion. The quantization procedure of a relativistic particle requires that the particle be an extended one, and to quantize a Fermi field, it is necessary to introduce an anisotropic feature in the internal space of the particle so that it gives rise to two internal helicities corresponding to a particle and an antiparticle. This specific quantum geometry of a Dirac particle gives rise to the solitonic feature as envisaged by Skyrme and the Skyrme term appears as an effect of quantization. When in the Lagrangian formulation the effect of this topological property is taken into account, it is found that the anomaly vanishes.

I. INTRODUCTION

In recent times, the old idea of the topological origin of the baryon number proposed by Skyrme[1] and Finkelstein and Rubinstein[2] has been revived. These authors put forward the idea that conserved quantum numbers arise as a consequence of the topological properties and that particles that carry conserved quantum numbers are built up from classical fields of nontrivial topology. In this picture baryons appear as solitons, commonly known as skyrmions. In a recent paper[3] it has been shown that the Skyrme term, which is necessary for the stability of a soliton, may appear as a consequence of the anisotropic feature of the internal space-time, where we have assumed that there is a fixed axis corresponding to a "direction vector" and this property of internal space-time helps us to have a consistent quantization of a Fermi field. In this scheme all fermions appear as solitons and the Skyrme term may be considered as an effect of quantization.

It may be added that Sternberg[4] has studied in detail the operation of charge conjugation and has argued that geometrically charged conjugation is induced by the Hodge star operator acting on a twistor space. It has been pointed out elsewhere[5] that the geometrical formulation of conformal inversion, which is induced by a charge conjugation acting on a spinor, in effect, corresponds to the inversion of the internal helicity for a spinor. This internal helicity may be taken to correspond to a fixed direction vector in the internal space of a massive spinor or a direction vector (vortex line) attached to the space-time point of a massless or massive spinor in a composite system of hadrons. The Hodge star operation in twistor space eventually inverts the orientation of the direction vector. In view of this, the internal helicity may be taken to represent the fermion number and can be taken to be of topological origin.

Jackiw[6] first pointed out the significance of topological effects in gauge field theories and its relationship with anomalies in quantum field theory. In a very elegant way he has shown how anomalies arise due to quantum mechanical symmetry breaking. Alvarez-Gaume and Ginsparg[7] studied non-Abelian anomalies from topological considerations. In this paper we shall show that the topological aspect of the stochastic quantization procedure of a Fermi field, where a direction vector is attached to a space-time point corresponding to the anisotropic feature of the internal space giving rise to the fermion number, helps us to find out the origin of the chiral anomaly in quantum field theory. This anomaly is avoided when we take into account this quantum geometry to study interactions involving gauge fields.

II. CONFORMAL GEOMETRY, TWISTOR SPACE, AND TOPOLOGICAL ASPECTS OF A FERMION

It is well known that the wave function of the form $\Psi(X_\mu, Y_\mu)$, where Y_μ is an attached vector that extends the Lorentz group SO(3,1) to the de Sitter group SO(4,1). Now in the stochastic quantization procedure for a fermion, it has been shown that a massive fermion is characterized by a fixed direction vector in the internal space that helps us to derive the fermionic propagator in Minkowski space from the two-point correlation of the stochastic fields $\varphi(Z_\mu) = \varphi(X_\mu) + i\varphi(Y_\mu)$, where the coordinate is given by $Z_\mu = X_\mu + iY_\mu$ in a complex manifold.[8] This indicates that the internal space of a massive fermion is disconnected in nature. This disconnectedness of the internal space gives rise to an internal helicity of the particle that corresponds to the fermion number. This follows from the fact that since the group structure is now given by SO(4,1), the irreducible representations of SO(4), the maximal compact subgroup of SO(4,1), are characterized by two numbers (k,n), where k is an integer or half-integer and n is a natural number. These two numbers are related to the eigenvalues of the Casimir operators by

$$\begin{aligned} \tfrac{1}{2} S^{\alpha\beta} S_{\alpha\beta} &= k^2 + (|k| + n)^2 - 1, \\ \tfrac{1}{8} \epsilon^{\alpha\beta\gamma\delta} S_{\alpha\beta} S_{\gamma\delta} &= k(|k| + n), \end{aligned} \tag{1}$$

where $S_{\alpha\beta}$, $\alpha,\beta = 1,2,3,4$, are the generators of the group. Barut and Bohm[9] have shown that the representations of SO(4,1) given by $S = \frac{1}{2}$ and $k = \pm\frac{1}{2}$ can be fully extended to two inequivalent representations of the conformal group SO(4,2). In fact these values actually correspond to the eigenvalues of the operator $K = \frac{1}{2}(a^+a - b^+b)$ in the oscillator representation of the $SO(3)_1 \times SO(3)_2$ basis of SO(4). The value of k as well as its signature is an SO(4,2) invariant. The representation $(s = 0, k = 0)$ in the conformal interpretation of SO(4,2) describes a massless spin-0 particle. The

 0022-2488/89/102366-07$02.50

representation $s=\frac{1}{2}, k= \pm\frac{1}{2}$ describes the helicity state of a massless spinor. Now for a massive particle, the conformal invariance breaks down and $k= \pm\frac{1}{2}$ cannot be represented as helicity states in the conventional sense, but represents an "internal helicity" or orientation so that the mutually opposite orientations are equivalent to particle and antiparticle states.

Since these representations can be fully extended to the conformal group SO(4,2), we can now deal with eight-component conformal spinors. The simplest conformally covariant spinor field equation postulated as an O(4,2) covariant equation in a pseudo-Euclidean manifold $M^{4,2}$ is of the form

$$\left(\Gamma_a \frac{\partial}{\partial \eta_a} + m\right)\xi(\eta) = 0, \quad a = 0,1,2,3,5,6, \tag{2}$$

where the elements of the Clifford algebra Γ_a are the basis unit vectors of $M^{4,2}$, m is a constant matrix, and $\xi(\eta)$ is an eight-component spinor field. Cartan[10] has shown that in the fundamental representation where the unit vectors are represented by 8×8 matrices of the form

$$\Gamma_a = \begin{vmatrix} 0 & \Xi \\ H & 0 \end{vmatrix}, \tag{3}$$

the conformal spinors ξ are of the form

$$\xi = \begin{vmatrix} \phi_1 \\ \phi_2 \end{vmatrix}, \tag{4}$$

where ϕ_1 and ϕ_2 are Cartan semispinors. The characteristic property of these spinors is that for any reflection ϕ_1 and ϕ_2 interchange. In this basis, Eq. (2) becomes equivalent to the coupled equations in the Minkowski space

$$\begin{aligned} i\partial\!\!\!/\phi_1 &= m\phi_2, \\ i\partial\!\!\!/\phi_2 &= m\phi_1. \end{aligned} \tag{5}$$

However it is also possible to obtain from Eq. (2) a pair of standard Dirac equations in Minkowski space. To this end, we have to work with a unitary transformation C_1 given by

$$C_1 = \begin{vmatrix} L & R \\ R & L \end{vmatrix}, \tag{6}$$

where $L=\frac{1}{2}(1+\gamma_5)$, $R=\frac{1}{2}(1-\gamma_5)$ with

$$\gamma_5 = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}.$$

With this, we have

$$C_1\xi = \xi^D = \begin{vmatrix} \psi_1 \\ \psi_2 \end{vmatrix} \tag{7}$$

and

$$C_1^{-1}\Gamma_\mu C_1 = \Gamma_\mu{}^D = \begin{vmatrix} \gamma_\mu & 0 \\ 0 & \gamma_\mu \end{vmatrix}.$$

This suggests that Eq. (2) is equivalent in Minkowski space to the pair of standard Dirac equations

$$\begin{aligned} (i\partial\!\!\!/ + m)\psi_1 &= 0, \\ (i\partial\!\!\!/ + m)\psi_2 &= 0. \end{aligned} \tag{8}$$

It is to be noted that space or time reflection interchanges φ_1 and φ_2 and transforms ψ_1 and ψ_2 into themselves; conformal reflection interchanges both $\varphi_1 \leftrightarrow \varphi_2$ and $\psi_1 \leftrightarrow \psi_2$. It should be added that ψ_1 and ψ_2 may represent physical free massive fermions whereas φ_1 and φ_2 do not unless they are massless since they obey coupled equations. However, in the case $m\neq 0$, if we define φ_1 and φ_2 such that they represent two different "internal helicity" states given by $k=+\frac{1}{2}$ and $-\frac{1}{2}$, i.e., $\varphi_1=\psi(k=\frac{1}{2})$ and $\varphi_2=\psi(k=-\frac{1}{2})$, Eqs. (5) can be reduced to a single equation with two internal degrees of freedom when the linear combination of $\psi(k=+\frac{1}{2})$ and $\psi(k=-\frac{1}{2})$ represents an eigenstate. Now, to retain the four-component nature of the spinor in Minkowski space, these two internal degrees of freedom should be associated with particle–antiparticle states. Evidently this property of φ_1 and φ_2 satisfies the criteria that space, time, or conformal reflection transforms one into the other. This follows from the facts that (a) the parity operator changes the sign of k; (b) the time reversal operator T changes the orientation of the internal helicity and hence changes the sign of k; (c) as φ_1 and φ_2 are related here to particle–antiparticle states, conformal reflection changes one into the other. Thus each member of the doublet of massive spinors having the internal helicity $k=+\frac{1}{2}$ and $-\frac{1}{2}$, corresponding to particle and antiparticle states, represents a Cartan semispinor.

To have a geometrical interpretation of the doublet of Cartan semispinors it may be noted that it is possible to regard the components of the semispinor as the homogeneous coordinates of a point in three-dimensional projective space whereas those of another semispinor are regarded as the homogeneous coordinates of a plane in P^3 (Ref. 11). Moreover, a point-plane correspondence exists in P^3 that reflects the conjugation relation of semispinors. On the other hand, according to the analysis of Penrose,[12] there also exists a 1-1 correspondence between twistors of valence $\binom{1}{0}$ $\binom{0}{1}$ and a point plane in P^3. Thus the semispinors into which an eight-component spinor splits in the Cartan basis are identical to Penrose twistors. This reflects the analysis of Sternberg[4] that charge conjugation corresponds to Hodgestar operation in twistor space.

This analysis along with the fact that the anticommutation relation of the eight-component conformal spinors gives rise to supersymmetry algebra[13] suggests that we can introduce a spinor structure at each space-time point so that we have additional degrees of freedom to our space-time manifold E parametrized by $(x_\mu,\theta,\bar\theta)$, where $\theta=\binom{\theta_1}{\theta_2}$ is a two-component spinor. This effectively corresponds to a superspace. Indeed, the additional degrees of freedom θ, $\bar\theta$ in the space-time structure may be related here with the internal helicity given by the values $k=+\frac{1}{2}$ and $-\frac{1}{2}$ in the representation space of $SO(4)=SO(3)_1\otimes SO(3)_2$. To this end, we choose the chiral coordinates in the superspace as

$$Z^\mu = x^\mu + (i/2)\lambda^\mu_\alpha\theta^\alpha \quad (\alpha=1,2), \tag{9}$$

where we identify the coordinate in the complex manifold $Z_\mu = X_\mu + iY_\mu$ with $Y^\mu=\frac{1}{2}\lambda^\mu_\alpha\theta^\alpha$. We now replace the chiral coordinates by the matrices

$$Z^{AA'} = X^{AA'} + (i/2)\lambda^{AA'}_\alpha\theta^\alpha, \tag{10}$$

where $\lambda^{AA'}_\alpha(\alpha=1,2)\in SL(2,c)$. With these relations the twistor equation is now modified as

$$\bar Z_a Z^a + \lambda^{AA'}_\alpha\theta^\alpha\bar\pi_A\pi_{A'} = 0, \tag{11}$$

where $\bar\pi_A$ $(\pi_{A'})$ is the spinorial variable corresponding to the four-momentum variable p^μ, the conjugate of X^μ, and is given by in the matrix representation

$$p^{AA'} = \bar{\pi}^A \pi^{A'} \tag{12}$$

and

$$Z^a = (\omega^A, \pi_{A'}), \quad \bar{Z}_a = (\bar{\pi}_A, \bar{\omega}^{A'}), \tag{13}$$

with

$$\omega^A = i(X^{AA'} + (i/2)\lambda_\alpha^{AA'}\theta^\alpha)\pi_{A'}. \tag{14}$$

Equation (11) now involves the helicity operator

$$S = -\lambda_\alpha^{AA'}\theta^\alpha \pi_A \pi_{A'}. \tag{15}$$

It may be noted that in the complex manifold, we have taken the matrix representation of P_μ, the conjugate of X_μ in the complex coordinate $Z_\mu = X_\mu + iY_\mu$, as $p^{AA'} = \bar{\pi}^A \pi^{A'}$ implying $P_\mu^2 = 0$ and so the particle will attain its mass due to the nonvanishing characteristic of the quantity Y_μ^2. In the null plane where $Y_\mu^2 = 0$, we can write the chiral coordinates as follows:

$$Z^{AA'} = X^{AA'} + (i/2)\bar{\theta}^A \theta^{A'}, \tag{16}$$

where the coordinate Y^μ is replaced by $Y^{AA'} = \frac{1}{2}\bar{\theta}^A\theta^{A'}$. In this case, the helicity operator is given by

$$S = -2Y^{AA'}\bar{\pi}_A \pi_{A'} = -\bar{\theta}^A \theta^{A'}\bar{\pi}_A \pi_{A'} = \bar{\epsilon}\epsilon, \tag{17}$$

with $\epsilon = i\theta^{A'}\pi_{A'}$, $\bar{\epsilon} = -i\bar{\theta}^A\bar{\pi}_A$. In this case, following Shirafuji[14] we can apply the canonical quantization procedure where $i\,\bar{z}_a$ and $i\,\bar{\epsilon}$ are canonically conjugate to Z^a and ϵ, respectively, and we can postulate the canonical commutation and anticommutation relations given by

$$[Z^\alpha, \bar{Z}_\beta] = \delta^\alpha{}_\beta, \tag{18}$$

$$\{\epsilon^i, \bar{\epsilon}_j\} = \delta^i{}_j. \tag{19}$$

Symmetrizing $\bar{Z}_a$ and Z^a and antisymmetrizing $\bar{\epsilon}$ and ϵ we require that the state vectors should satisfy

$$(\{\bar{Z}_a, z^a\} + [\bar{\epsilon}, \epsilon])|\psi\rangle = 0. \tag{20}$$

From this we find

$$(\bar{S} + \tfrac{1}{2}\bar{\epsilon}\epsilon - \tfrac{1}{4})|\psi\rangle = 0, \tag{21}$$

where

$$\bar{S} = \tfrac{1}{4}\{\bar{Z}_a, Z^a\}. \tag{22}$$

Now defining the operators

$$S_i^a = \bar{\epsilon}_i Z^a, \quad S_a^i = \bar{Z}_a \epsilon^i, \tag{23}$$

we have the commutation relations

$$[\bar{S}, S_i^a] = -\tfrac{1}{2}S_i^a, \qquad [\bar{S}, S_a^i] = \tfrac{1}{2}\bar{S}_a^i, \tag{24}$$

which indicates that $\bar{S}_a^i$ and S_i^a are the helicity raising and lowering operators, respectively. The state with the internal helicity $+\frac{1}{2}$ is the vacuum state of the fermion operator

$$\epsilon|S = +\tfrac{1}{2}\rangle = 0. \tag{25}$$

Similarly, the state with the internal helicity $-\frac{1}{2}$ is the vacuum state of the fermion operator

$$\bar{\epsilon}|S = -\tfrac{1}{2}\rangle = 0. \tag{26}$$

In case of a massive spinor, we can define a negative definite plane D^- where for the coordinate $Z = X + iY$, Y belongs to the interior of the forward light cone ($Y \gg 0$) and as such represents the upper half-plane with the condition $\det Y > 0$ and $\frac{1}{2}\operatorname{Tr} Y > 0$. The positive definite plane D^+ is given by the set of all coordinates Z with Y in the interior of the backward light cone ($Y \ll 0$). The map $Z \to Z^*$ sends a negative definite plane to a positive definite plane. The space M of null space ($\det Y = 0$) is the Shilov boundary so that a function holomorphic in $D^-(D^+)$ is determined by its boundary values. Thus if we consider that any function $\varphi(Z) = \varphi(X) + i\phi(Y)$ is holomorphic in the whole domain, we note that the helicity $+\frac{1}{2}(-\frac{1}{2})$ given by the operator $i\theta^A\pi_A(-i\bar{\theta}^A\bar{\pi}_A)$ in the null plane may be taken to be the limiting value of the "internal helicity" in the upper (lower) half-plane. This indicates that in the massive spinor case, we can consider that the helicity given by

$$S = -\lambda_\alpha^{AA'}\theta^\alpha \bar{\pi}_A \pi_{A'} \tag{27}$$

represents the internal helicity $+\frac{1}{2}$ where we have $Y \gg 0$. Since the map $Z \to Z^*$ transforms a negative definite plane to a positive definite plane, we will have an opposite internal helicity $-\frac{1}{2}$ with the coordinate $Z^\mu = X_\mu - iY_\mu$ replaced by the matrices $Z^{AA'} = X^{AA'} - (i/2)\lambda_\alpha^{AA'}\bar{\theta}\alpha$ having $\frac{1}{2}\operatorname{Tr} Y < 0$. In the null plane we will have the condition $Y^{AA'} = \frac{1}{2}\bar{\theta}^A\theta^{A'}$ so that we can have the simultaneous existence of two helicities $+\frac{1}{2}$ and $-\frac{1}{2}$ corresponding to the spin projections on the z axis for a massless spinor. In this way, we can relate the spinorial variables θ and $\bar{\theta}$ in the superspace given by the coordinate $(X_\mu, \theta, \bar{\theta})$ with the internal helicity of a massive spinor. Evidently, this corresponds to the values $k = +\frac{1}{2}$ and $-\frac{1}{2}$ in the representation space of $SO(4) = SO(3)_1 \otimes SO(3)_2$ in the de Sitter space.

Now we want to point out that when the extension of a particle is given by the coordinate $(X_\mu, \theta, \bar{\theta})$, we can have a gauge field theoretic description of this extension when the corresponding gauge fields have the group structure $SL(2,c)$. Indeed, the metric tensor $g_{\mu\nu}^{AA'}(X,\theta,\bar{\theta}) = g_{\mu\nu}(x)\bar{\theta}^A\theta^{A'}$ can be taken to be described by the $SL(2,c)$ gauge fields in Minkowski space-time with the gauge field strength tensor given by

$$F_{\mu\nu} = \partial_\mu B_\gamma - \partial_\gamma B_\mu + [B_\mu, \beta_\nu], \tag{28}$$

where B_μ is the matrix-valued potential and belongs to $SL(2,c)$ (Ref. 3). The asymptotic zero curvature condition then implies $F_{\mu\nu} = 0$ so that we can write the non-Abelian gauge field as

$$B_\mu = U^{-1}\partial_\mu U, \quad \text{where} \quad U \in SL(2,c).$$

With the substitution, we note that the corresponding Lagrangian is given by

$$L = M^2 \operatorname{Tr}(\partial_\mu U'\partial_\mu U) + \operatorname{Tr}[\partial_\mu UU^+, \partial_\nu UU^+]^2, \tag{29}$$

where M is a suitable constant having the dimension of mass.

Thus we find that the quantization of a Fermi field considering an anisotropy in the internal space leading to an internal helicity description corresponds to the realization of a nonlinear sigma model—where the Skyrme term in the Lagrangian ($L_{\text{Skyrme}} = \operatorname{Tr}[\partial_\mu UU^+, \partial_\nu UU^+]^2$) automatically arises for stabilizing the soliton. Thus in this picture, fermions appear as solitons and the fermion number is found to have a topological origin. Indeed, for the Hermitian representation, we can take the group manifold as SU(2) and this leads to a mapping from the space three-sphere S^3 to the group space S^3 [$SU(2) = S^3$] and the corresponding winding number is given by

$$q = \frac{1}{24\pi^2}\int_{S^3} ds_\mu\, \epsilon^{\mu\nu\alpha\beta}\,\mathrm{Tr}[U^{-1}\partial_\nu U U^{-1}\partial_\lambda U U^{-1}\partial_\beta U]. \tag{30}$$

Evidently q is a topological index and represents the fermion number.

III. TOPOLOGICAL ASPECTS OF A FERMION AND THE CONSERVED CURRENT

The above analysis can be used to link up the topological origin of fermion number with the internal helicity. Then the wave function for a particle and an antiparticle is implicitly represented as $\psi(x,\theta)$ and $\psi(X,\bar\theta)$, $\theta,\bar\theta$ indicating the internal helicity $+\frac{1}{2}$ and $-\frac{1}{2}$, respectively, and the metric tensor is given by $g_{\mu\nu}(X,\theta,\bar\theta)$. That is, spinor structures are introduced to each space-time point and we have a superspace. This geometry effectively gives rise to the SL(2,c) gauge fields (as the spinor-affine connection) having the field strength

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu + [B_\mu, B_\nu],$$

where B_μ is the matrix-valued potential. In superspace a given covariant tensor $F_{\mu\nu}$ does not have contravariant components $F^{\mu\nu}$. Therefore, following Carmeli and Malin[15] we choose the simplest Lagrangian density which is invariant under SL(2,c) transformations

$$\alpha = -\tfrac{1}{4}\,\mathrm{Tr}(\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}) \tag{31}$$

where $\epsilon^{\alpha\beta\gamma\delta}$ is the completely antisymmetric tensor density in four dimensions with $\epsilon^{0123} = 1$. Applying the usual procedure of variational calculus, we get the field equations

$$\partial_\delta(\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}) - [B_\delta, \epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}] = 0. \tag{32}$$

Taking the infinitesimal generators of the group SL(2,c) as

$$g_1 = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix},\quad g_2 = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix},\quad g_3 = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \tag{33}$$

we can write

$$\begin{aligned} B_\mu &= b^a_\mu g^a = \mathbf{b}_\mu\cdot\mathbf{g},\\ F_{\mu\nu} &= F^a_{\mu\nu}g^a = \mathbf{f}_{\mu\nu}\cdot\mathbf{g}\quad (a = 1,2,3). \end{aligned} \tag{34}$$

Evidently in this space, these SL(2,c) gauge fields will appear as background fields.

Thus to describe a matter field in this geometry, the Lagrangian will be modified by the introduction of this SL(2,c) invariant Lagrangian density (31). Hence for a massless spinor field we write for the Lagrangian

$$L = -\bar\psi\gamma^\mu D_\mu\psi - \tfrac{1}{4}\,\mathrm{Tr}\,\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}, \tag{35}$$

where D_μ is the SL(2,c) gauge covariant derivatives defined by

$$D_\mu = \partial_\mu - igB_\mu,$$

where g is some coupling strength. It is to be observed that by the introduction of the SL(2,c) gauge field Lagrangian in (35), we are effectively taking into account the effect of the extension of the fermionic particle giving rise to the internal helicity in terms of the gauge fields.[16] That is, writing the space-time coordinate and the four-momentum variables as

$$\begin{aligned} Q_m &= q_\mu + i\hat{Q}_\mu,\\ P_\mu &= p_\mu + i\hat{p}_\mu, \end{aligned} \tag{36}$$

where $q_\mu(p_\mu)$ corresponds to the mean position (momentum) relating to the external space-time and $\hat{Q}_\mu(\hat{P}_n)$ corresponds to the internal stochastic extension, we can write, following Brooke and Prugovecki,[17] the following representation of Q_μ/ω and P_μ/ω, $\omega = \hbar/lmc$ (l being a fundamental length) acting on functions defined in phase space:

$$\begin{aligned} \frac{Q_\mu}{\omega} &= -i\left(\frac{\partial}{\partial p_\mu} + \phi_\mu\right),\\ \frac{P_\mu}{\omega} &= i\left(\frac{\partial}{\partial q_\mu} + \psi_\mu\right), \end{aligned} \tag{37}$$

where ϕ_μ and ψ_μ are matrix-valued functions. Thus identifying ϕ_μ with the SL(2,c) gauge field B_μ, we note that this spatial extension will give rise to a Lagrangian density given by (31) in addition to the point-particle spinorial Lagrangian density $\bar\psi\gamma_\mu\partial_\mu\psi$. Besides we can conceive of a coupling with this backgroundfield with the spinor and this leads to Eq. (35) for the effective Lagrangian of the spinorial matter field.

From this, we can now construct a conserved current corresponding to this Lagrangian and we get (neglecting the coupling with the gauge field)

$$\mathbf{j}^\mu = \bar\psi\gamma^\mu\psi + \epsilon^{\mu\nu\alpha\beta}\mathbf{b}_\nu\times\mathbf{f}_{\alpha\beta} = \mathbf{j}^\mu_x + \mathbf{j}^\mu_\theta. \tag{38}$$

Indeed from the properties of SL(2,c) generators we find from (32) that

$$\epsilon^{\mu\nu\alpha\beta}(\partial_\nu\mathbf{f}_{\alpha\beta} - \mathbf{b}_\nu\times\mathbf{f}_{\alpha\beta}) = 0.$$

This suggests that

$$\mathbf{j}^\mu_\theta = \epsilon^{\mu\nu\alpha\beta}\mathbf{b}_\nu\times\mathbf{f}_{\alpha\beta} = \epsilon^{\mu\nu\alpha\beta}\partial_\nu\mathbf{f}_{\alpha\beta}. \tag{39}$$

Then using the antisymmetric property of the Levi-Civita tensor density $\epsilon^{\mu\nu\alpha\beta}$ we get

$$\partial_\mu\mathbf{j}^\mu_\theta = \epsilon^{\mu\nu\alpha\beta}\partial_\mu\,\partial_\nu\,\mathbf{f}_{\alpha\beta} = 0. \tag{40}$$

Now noting that for spinor field, the vector current density is conserved, we finally have

$$\partial_\mu\mathbf{j}^\mu = \partial_\mu(\mathbf{j}^\mu_x + \mathbf{j}^\mu_\theta) = 0. \tag{41}$$

However, in the Lagrangian (35), if we split the Dirac massless spinor in chiral forms and identify the internal helicity $(+\frac{1}{2})$ $(-\frac{1}{2})$ with left (right) chirality corresponding to θ and $\bar\theta$, we can write

$$\begin{aligned} \bar\psi\gamma_\mu D_\mu\psi &= \bar\psi\gamma_\mu\partial_\mu\psi - ig\bar\psi\gamma_\mu B^a_\mu g^a\psi\\ &= \bar\psi\gamma_\mu\partial_\mu\psi - (ig/2)\{\bar\psi_R\gamma_\mu B^1_\mu\psi_R - \bar\psi_R\gamma_\mu B^2_\mu\psi_R\\ &\quad + \bar\psi_L\gamma_\mu B^2_\mu\psi_L + \bar\psi_L\gamma_\mu B^3_\mu\psi_L\}. \end{aligned} \tag{42}$$

Then the three SL(2,c) gauge field equations give rise to the following three conservations laws,

$$\begin{aligned} &\partial_\mu[\tfrac{1}{2}(-ig\bar\psi_R\gamma_\mu\psi_R) + j^1_\mu] = 0,\\ &\partial_\mu[\tfrac{1}{2}(-ig\bar\psi_L\gamma_\mu\psi_L + ig\bar\psi_R\gamma_\mu\psi_R) + j^2_\mu] = 0,\\ &\partial_\mu[\tfrac{1}{2}(-ig\bar\psi_L\gamma_\mu\psi_L) + j^3_\mu] = 0. \end{aligned} \tag{43}$$

These three equations represent a consistent set of equations if we choose

$$j_\mu^1 = -j_\mu^2/2, \quad j_\mu^3 = j_\mu^2/2,$$

which evidently guarantees the vector current conservation. Then we can write

$$\begin{aligned} &\partial_\mu(\bar\psi_R\gamma_\mu\psi_R + j_\mu^2) = 0, \\ &\partial_\mu(\bar\psi_L\gamma_\mu\psi_L - j_\mu^2) = 0. \end{aligned} \tag{44}$$

From these, we find

$$\partial_\mu(\bar\psi\gamma_\mu\gamma_5\psi) = \partial_\mu j_\mu^5 = -2\partial_\mu j_\mu^2. \tag{45}$$

Thus the anomaly is expressed here in terms of the second $SL(2,c)$ component of the gauge field current j_μ^2. However, since in this formalism the chiral currents are modified by the introduction of j_μ^2, we note from Eq. (44) that the anomaly vanishes.

From these equations, two separately conserved charges emerge, viz.,

$$\begin{aligned} \widetilde{Q}_L &= \int \psi_L^+ \psi_2\, d^3x - \int j_0^2\, d^3x, \\ \widetilde{Q}_R &= \int \psi_R^+ \psi_R\, d^3x + \int j_0^2\, d^3x. \end{aligned} \tag{46}$$

The charge corresponding to the gauge field part is

$$q = \int j_0^2\, d^3x = \int_{\text{surface}} \epsilon^{ijk}\, da_i\, F_{jk}^2 \quad (i,j,k = 1,2,3). \tag{47}$$

Visualizing F_{jk}^2 to be the magnetic fieldlike components for the vector potential B_i^2, we see that $(i = 1,2,3)$ is actually associated with the magnetic pole strength for the corresponding field distribution.

The term $\epsilon^{\alpha\beta\gamma\delta}\,\mathrm{Tr}\, F_{\alpha\beta}F_{\gamma\delta}$ in the Lagrangian can be actually expressed as a four-divergence of the form $\partial_\mu\Omega^\mu$, where

$$\Omega^\mu = -(1/16\pi^2)\epsilon^{\mu\alpha\beta\gamma}\,\mathrm{Tr}\left[\tfrac{1}{2}B_\alpha F_{\beta\gamma} - \tfrac{2}{3}(B_\alpha B_\beta B_\gamma)\right]. \tag{48}$$

We recognize that the gauge field Lagrangian is related to the Pontryagin density

$$P = -(1/16\pi^2)\,\mathrm{Tr}^*F_{\mu\nu}F^{\mu\nu} = \partial_\mu\Omega^\mu \tag{49}$$

and Ω^μ is the corresponding Chern–Simons secondary characteristic class. The Pontryagin index

$$q = \int P d^4x \tag{50}$$

is then a topological invariant. If we consider Euclidean four-dimensional space-time, then the above integral may be reduced to a three-surface integral where the three-surface is topologically equivalent to S^3. Now it is noted that we must have $F_{\alpha\beta} = 0$ at all spatial and temporal infinity points so that the action $S = \int L\, d^4x$ gives rise to a finite energy gauge field configuration. Then the gauge potentials tend to a pure gauge at large distances in all four directions, i.e., we have

$$B_\mu \underset{x_\mu\to\alpha}{\rightarrow} U^{-1}\partial_\mu U. \tag{51}$$

This then helps us to write

$$q = \frac{1}{24\pi^2}\int_{S^3} ds_\mu\, \epsilon^{\mu\alpha\beta\gamma} \times \mathrm{Tr}\left[U^{-1}\partial_\alpha U U^{-1}\partial_\beta U U^{-1}\partial_\gamma U\right]. \tag{52}$$

We observe that this is nothing but the fermion number as discussed in the previous section. In four-dimensional space-time, if we assume B_μ to go faster than $1/r$, B_μ being zero at negative infinity of the time coordinate but tends to a pure gauge at positive infinity of the time coordinate, we can write

$$q = \int d\mathbf{r}\, \Omega^0 \Big|_{x^0=\alpha}. \tag{53}$$

From this, it appears that the axial vector current is now modified as

$$\tilde{j}_\mu^5 = j_\mu^5 + 2\hbar\Omega_\mu \tag{54}$$

and though $\partial_\mu j_\mu^5 \neq 0$, we have $\partial_\mu \tilde{j}_\mu^5 = 0$. That means, when the topological properties of a fermion related to the origin of fermion number is taken into account, we are not confronted with the chiral anomaly. The origin of the chiral anomaly is thus found to be due to the naive form of the point particle current without any topological structure, which turns out to be essential for the quantization of a Fermi field.

IV. FERMIONS AND THE INTERACTION WITH AN EXTERNAL ABELIAN GAUGE FIELD

The chiral description of the matter field in terms of the spinorial variables $\theta,\bar\theta$ in the metric tensor $g_{\mu\nu}(x,\theta,\bar\theta)$ giving rise to the $SL(2,c)$ gauge field currents necessitates the introduction of a disconnected gauge group for the external Abelian field interacting with the matter field in a chiral symmetric way. In the case where the external Abelian gauge field is the electromagnetic field, the Lagrangian density is given by

$$\begin{aligned} L = &-\bar\psi\gamma_\mu D_\mu\psi - \tfrac{1}{4}\,\mathrm{Tr}(\epsilon^{\alpha\beta\gamma\delta}\widetilde{F}_{\alpha\beta}\widetilde{F}_{\gamma\delta}) \\ &- \tfrac{1}{4}\,\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) + \mathrm{Tr}(j_\mu A^\mu). \end{aligned} \tag{55}$$

Here D_μ is the $SL(2,c)$ gauge covariant derivative and, considering the order of $\psi - B_\mu$ coupling to be negligible compared to the matter current electromagnetic field coupling, we can replace it by ∂_μ,

$$\widetilde{F}_{\alpha\beta} = \partial_\alpha B_\beta - \partial_\beta B_\alpha + [B_\alpha, B_\beta], \quad B_\alpha \in SL(2,c)$$

and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, A_μ being the electromagnetic gauge potential and j_μ is the matter current matrix given by

$$j_\mu = \begin{bmatrix} \bar\psi_R\gamma_\mu\psi_R + j_\mu^2 & 0 \\ 0 & \bar\psi_L\gamma_\mu\psi_L - j_\mu^2 \end{bmatrix}, \tag{56}$$

where j_μ^2 is the second component of the $SL(2,c)$ gauge field current as discussed in the previous section. It is evident that this matrix structure of j_μ exhibiting the chiral form suggests that for A_μ we should take the disconnected gauge group $U_{1L}\times U_{1R} = U_1 x\{1,d\}$ where d is the orientation reversing operation. Evidently in such an interaction the field strength and current are not gauge invariant but only gauge covariant, each changing sign under d. This is similar to the non-Abelian theories where field strengths and currents are only gauge covariant even under gauge transformations connected to the identity. The internal symmetry group here is $O(2)$ which is given by the relation

$$O(2) = SO(2)\times\{1,d\} = U_1\times\{1,d\}, \quad (57)$$

where d is the orientation reversing operator. Indeed, we can take

$$A_\mu = \begin{bmatrix} A_{\mu+} & 0 \\ 0 & A_{\mu-} \end{bmatrix}. \quad (58)$$

Kiskis[18] has studied the interactions having disconnected gauge group. Following Kiskis, we can think of a large system of observers each responsible for a small open region U_i of the connected space-time manifold M. Let us consider that all the frames in U_i have the same orientation. Physically this means that the space is simply connected and the observer can give an unambiguous definition of positive charge everywhere. This suggests that we can introduce the connection (gauge field) in the Lagrangian

$$L^i = L_g^{(i)} + L_M^{(i)}, \quad (59)$$

where i identifies quantities associated with the region U_i,$L_M^{(i)}$ is the matter field Lagrangian, and $L_g^{(i)}$ is the kinetic energy term for the connection. The gauge symmetry of the $L_g^{(i)}$ is given by

$$A \to g^{-1}(\partial + A)g, \quad (60)$$

with g a smooth map

$$g = U_i \to O(2) \quad (61)$$

which may lie in either component of O(2). A transformation that reverses the orientation at each point can be written as

$$g = dg_0,$$
$$g_o = U_i \to SO(2), \quad (62)$$
$$d = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This gives

$$A \to g_0^{-1}(\partial - A)g_0. \quad (63)$$

We see that it is a combination of charge conjugation and orientation preserving gauge rotation. Evidently in this formalism the chiral currents interact with the gauge field in a disconnected form. Indeed, writing

$$A_\mu = \begin{bmatrix} A_{\mu+} & 0 \\ 0 & A_{\mu-} \end{bmatrix}$$

we find the interaction term is given by

$$\begin{bmatrix} (\bar\psi_L \gamma_\mu \psi_L - j_\mu^2)A_{\mu+} & 0 \\ 0 & (\bar\psi_R \gamma_\mu \psi_R + j_\mu^2)A_{\mu-} \end{bmatrix}. \quad (64)$$

Evidently there is no term like $A_{\mu+} A_{\mu-}$ in the Lagrangian.

As Kiskis[18] has discussed, in the overlap region

$$U_{ij} = U_i \cap U_j$$

there are two observers studying the same physical system where each observer has set up his own basis in the internal symmetry space over U_{ij}. The relation between these bases is a gauge transformation

$$g_{ij}: U_{ij} \to O(2),$$

where the map lies in either component of O(2). That is, observers i and j may have opposite charge convention. If they have opposite convention about charge, they will have opposite convention about field. In fact, if we designate *a priori* what is a particle and what is an antiparticle, the left and right directions can be determined by any parity violating interaction. On the other hand, if we designate what is left and what is right the particle–antiparticle designation remains fixed. Thus any parallel transport from a region U_i to U_j of the manifold will be such that either the orientation remains the same and the observer will see the same charge or the orientation is opposite when by reversing the orientation of U_j the observer will see the same charge. Thus any path from any region U_i to U_j will be such that either this will give the same orientation for U_j or it is opposite when reversing the orientation, the observer will identify a left-handed or a right-handed particle.

V. DISCUSSION

We have shown above that the chiral anomaly is connected with the topological properties of a fermion. Indeed, the topological property of a fermion gives rise to the fermion number which is always conserved and helps us to treat fermions as solitons. The Skyrme term here arises just as an effect of quantization of a fermion[3] and is related to the quantum geometry of a relativistic particle. The relativistic generalization of a quantum particle necessitates the particle to be an extended one and to attain the fermionic property, we need to introduce an anisotropic feature in the internal space of the particle so that it gives rise to two internal helicities corresponding to a particle and an antiparticle. This specific quantum geometry of a Dirac particle gives rise to the solitonic feature as envisaged by Skyrme[1] as well as by Finkelstein and Rubinstein.[2] When in the Lagrangian formulation the effect of this topological property is taken into account, we find that the anomaly vanishes.

This analysis suggests that the origin of anomaly lies in the fact that fermions are conventionally treated as localized point particles devoid of any specific geometrical and topological feature. But when this topology is taken into account anomaly vanishes implying that when we study quantum mechanical symmetry breaking, we must take into account the geometrical features involved in the quantization procedure. That is, quantum mechanical effects have their origin in quantum geometry and need to be studied in this perspective.

[1]T. H. R. Skyrme, Proc. R. Soc. London Ser. A **260**, 127 (1961). Nucl. Phys. **31**, 556 (1962).

[2]D. Finkelstein, J. Math. Phys. **7**, 1218 (1966); D. Finkelstein and J. Rubinstein, *ibid.* 1762 (1968).

[3]P. Bandyopadhyay and K. Hajra, J. Math. Phys. **28**, 711 (1987).

[4]S. Sternberg, Comm. Math. Phys. **109**, 649 (1987).

[5]P. Bandyopadhyay, "Holomorphic Quantum Mechanics, Conformal Reflection, and the Internal Symmetry of Hadrons," to be published in Int. J. Mod. Phys.

[6]R. Jackiw, *Topological Investigations of Quantized Gauge Theories* (Relativity, Groups and Topology, Les Houches 1983), edited by B. S. Dewitt and R. Stora (North-Holland, Amsterdam, 1984).

[7]L. Alvarez-Gaume and P. Ginsparg, Nucl. Phys. B **243**, 449 (1984).

[8]P. Bandyopadhyay and K. Hajra, "Stochastic Quantization in Minkowski Space" (submitted for publication).

[9]A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970).

[10]E. Cartan, *The Theory of Spinors* (Paris, 1966).
[11]M. Daniel, Phys. Lett. B **65**, 246 (1976).
[12]R. Penrose, Int. J. Theor. Phys. **1**, 61 (1968).
[13]A. Haag, J. Lopusanski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975); M. Daniel and C. N. Ktorides, Nucl. Phys. B **115**, 313 (1976).
[14]T. Shirafuji, Prog. Theor. Phys. **70**, 18 (1983).
[15]S. Malin and M. Carmeli, Ann. Phys. **103**, 208 (1977).
[16]P. Bandyopadhyay and K. Hajra, "Stochastic Quantization, Localizability of a Relativistic Particle and Quantum Geometry" (submitted for publication).
[17]J. A. Brooke and E. Prugovecki, Nuovo Cimento A **89**, 126 (1985).
[18]J. Kiskis, Phys. Rev. **17**, 3196 (1978).

Dirac equation in external vector fields: New exact solutions

German V. Shishkin and Victor M. Villalba[a)]
Department of Theoretical Physics, Byelorussian State University, Minsk, 220080, Union of Soviet Socialist Republics

(Received 17 May 1988; accepted for publication 19 April 1989)

New exact solutions are searched for on the basis of the method of separation of variables proposed in earlier work by the present authors [J. Math. Phys. 30, 2132 (1989)]. The essence of this method consists of constructing first-order matrix differential operators that define the dependence of the Dirac bispinor on the related variables, but commutation of such operators with the operator of the equation or between them is not assumed.The classical problems are considered as possibilities, namely, electrons in the field of plane monochromatic electromagnetic waves (Volkov's problem) and electrons in the Coulomb field (hydrogen atom). Then "plane" external electromagnetic fields are considered for which some new exact solutions are obtained in terms of special functions. Four new exact solutions of the Dirac equation in the fields with axial cylindrical symmetry are also shown, and lastly one "free" solution with exotic geometry is demonstrated, namely, "free" parabolic cylindrical spinor waves.

I. INTRODUCTION

We have proved rigorous theorems about necessary and sufficient conditions on the external vector fields that allow us to have partial or complete separation of variables in the Dirac equation in Ref. 1 (see Theorems 1–6 for Cartesian coordinates and Theorems 7 and 8 for the general orthogonal curvilinear coordinates).

Taking into account the conditions of complete separation of variables in the Dirac equation in Cartesian coordinates, we can see that all the possible potentials that allow such a separation contain additively in the components the dependences on the corresponding variables, i.e.,

$$A_k = \widetilde{A}_k(x^k) + B_k(x^m, x^n, x^l), \quad k \neq m \neq n \neq l. \tag{1.1}$$

Therefore it is possible to simplify according to

$$\Psi = \widetilde{\Psi} \exp\left\{ i \sum \int \widetilde{A}_k(x^k) dx^k \right\}, \quad k = i,j,m,n. \tag{1.2}$$

As a result of transformation (1.2), the bispinor Ψ satisfies the Dirac equation in the form

$$\{\gamma^i(\partial_i - iA_i) + \gamma^j(\partial_j - iA_j) + \gamma^m(\partial_m - iA_m) + \gamma^n(\partial_n - iA_n) + m_0\}\Psi = 0, \tag{1.3}$$

where the Lorentz condition on the vector potential is fulfilled automatically:

$$\partial_m A^m = 0 \tag{1.4}$$

(summation on m takes place). Here and thereafter we use the nomenclature of Ref. 1.

Simplifications analogous to (1.2) are possible also in the case of curvilinear coordinates.

Note that the conditions of separation of variables in the Dirac equation according to Ref. 1 require the components of the vector potential to be sums of functions of separable variables. However, the physical fields may have structure other than as sums of functions of separable variables; for example, as a field of a plane monochromatic electromagnetic wave:

$$A_x = Ae^{i\omega(z-t)}. \tag{1.5}$$

Here the wave propagates along the Z axis with transversal polarization. In such cases we can often reduce the Dirac equation to the form (1.3) by means of transition to the new variables. However, such transition is connected with a mixing of space and time variables and as a result the matrices multiplying the corresponding derivatives may not have definite Hermitian form and, moreover, may be degenerate. In the case (1.5) such a situation takes place for the well known variables

$$u = z - t, \quad v = z + t. \tag{1.6}$$

The degenerate matrices must be handled with special care. In other words, after reducing the Dirac equation to the form (1.3) we can use the method of separation of variables proposed in Ref. 1.

In the case (1.5) the Dirac equation takes the form

$$\{\gamma^1(\partial_x - iA(u)) + \gamma^2 \partial_y + \gamma^u \partial_u + \gamma^v \partial_v + m_0\}\Psi = 0, \tag{1.7}$$

where

$$\gamma^u = \gamma^3 - \gamma^4, \quad \gamma^v = \gamma^3 + \gamma^4. \tag{1.8}$$

Here the matrices γ^u and γ^v are degenerate.

Separating successively x,y, and v according to Theorem 3 of Ref. 1 we have

$$\hat{K}_x = -i\partial_x, \quad \hat{K}_y = -i\partial_y, \quad \hat{K}_v = -i\partial_v, \tag{1.9}$$

$$\hat{K}_u = \gamma^1(k_x + A(u)) + \gamma^2 k_y - i\gamma^4 \partial_u + \gamma^v k_v - im_0. \tag{1.10}$$

Choosing the representation

$$\gamma^1 = \begin{pmatrix} 0 & \sigma^1 \\ \sigma^1 & 0 \end{pmatrix}, \quad \gamma^2 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma^3 = \begin{pmatrix} 0 & \sigma^3 \\ \sigma^3 & 0 \end{pmatrix}, \quad \gamma^4 = i\begin{pmatrix} 0 & \sigma^2 \\ \sigma^2 & 0 \end{pmatrix}, \tag{1.11}$$

we obtain the exact solution

[a)] Permanent address: Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), Apdo 21817, Caracas 1020-A, Venezuela.

$$\Psi=\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix},\quad \Psi_2=\frac{ik_y+m_0}{k}\Psi_1,\quad \Psi_1=\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}, \tag{1.12}$$

$$k^2=k_y^2+m_0^2, \tag{1.13}$$

$$\varphi_1=\exp\left\{-\frac{i}{4}\int\{(k_x+A(u))^2+k^2\}du\right\},$$

$$\varphi_2=\frac{-i}{2k_x}\{k-i(k_x+A(u))\}\varphi_1. \tag{1.14}$$

This is the well-known Volkov solution.[2,3]

Our nomenclature, however, is closer to Refs. 4 and 5.

Taking the Dirac equation in spherical coordinates in the presence of a Coulomb field in a diagonal gauge tetrad we have

$$\left\{\gamma^1\partial_r+\frac{\gamma^2}{r}\partial_\theta+\frac{\gamma^3}{r\sin\theta}\partial_\varphi+\gamma^4\partial_t-i\gamma^4\frac{\alpha}{r}+m_0\right\}\Psi=0. \tag{1.15}$$

According to Theorem 8 of Ref. 1, separating t, θ, and φ successively we have

$$\hat{K}_t=-i\partial_t,\quad \hat{K}_\varphi=-i\partial_\varphi,$$
$$\hat{K}_\theta=-i\left\{\gamma^2\partial_\theta+\frac{\gamma^3}{\sin\theta}m\right\}\gamma^1\gamma^4, \tag{1.16}$$
$$\hat{K}_r=\left\{-i\gamma^1\partial_r-\frac{k}{r}\gamma^1\gamma^4-\gamma^4\left(E-\frac{\alpha}{r}\right)-im_0\right\}.$$

Here m is an eigenvalue of the operator $\hat{K}_\varphi$ and k is an eigenvalue of the operator $\hat{K}_\theta$.

Taking the matrices γ^1 and γ^4 in the form

$$\gamma^1=\begin{pmatrix}0 & -I\\ I & 0\end{pmatrix},\quad \gamma^4=i\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix}, \tag{1.17}$$

we can have from (1.16) the standard radial equations of a hydrogenlike atom

$$\left(\frac{\partial}{\partial r}+\frac{k}{r}\right)\chi-\left(E-\frac{\alpha}{r}+m_0\right)\varphi=0,$$
$$\left(-\frac{\partial}{\partial r}+\frac{k}{r}\right)\varphi+\left(E-\frac{\alpha}{r}-m_0\right)\chi=0,\qquad \tilde{\Psi}=\begin{pmatrix}\varphi\\ \chi\end{pmatrix}. \tag{1.18}$$

Gathering the operators $\hat{K}_\varphi$ and $\hat{K}_\theta$ after the unitary transformation that is the inverse of that of (3.25) of Ref. 1, we obtain the operator (expressed in the Cartesian gauge)

$$\hat{K}_{\theta\varphi}=\left\{\gamma_1\gamma_4\gamma_2\frac{\partial}{i\,\partial\theta}+\gamma_1\gamma_4\gamma_3\sin^{-1}\theta\frac{\partial}{i\,\partial\varphi}+\frac{\gamma_4}{i}\right\}, \tag{1.19}$$

connected in the standard way with the momentum of Dirac's particle:

$$\hat{K}^2_{\theta\varphi}=\left(\hat{\mathbf{L}}+\frac{1}{2}\boldsymbol{\sigma}\right)^2+\frac{1}{4}=\hat{\mathbf{J}}^2+\frac{1}{4}=\{\boldsymbol{\sigma}\hat{\mathbf{L}})+1\}^2. \tag{1.20}$$

Here $\hat{\mathbf{L}}$ is the orbital momentum operator and $\frac{1}{2}\boldsymbol{\sigma}$ is the spin momentum.

Thus we have classical results for the classical problems in our scheme of separation of variables proposed in Ref. 1.

II. SOLUTIONS OF THE DIRAC EQUATION IN CARTESIAN COORDINATES

We assume that when necessary the transformation (1.2) is fulfilled and, according to Theorem 6 of Ref. 1, the components of the vector potential satisfying the Lorentz condition and allowing the complete separation of variables take the form

$$A_i=0,\quad A_j=A_j(x^i),\quad A_m=0,\quad A_n=A_n(x^m), \tag{2.1}$$

or the trivial case when the components of vector potential depend only on one variable. The separation is most simple if first we separate the variables in pairs. Then instead of (1.3) we have

$$\{\gamma^1\partial_3+\gamma^2(E-A_4)\}\gamma^3\gamma^4\tilde{\Psi}=k\tilde{\Psi}, \tag{2.2}$$
$$\{\gamma^1\partial_1+\gamma^2(\partial_2-iA_2)+m_0\}\gamma^1\gamma^2\tilde{\Psi}=ik\tilde{\Psi}, \tag{2.3}$$

where

$$\tilde{\Psi}=\hat{\zeta}\Psi=\tfrac{1}{2}(1+i\gamma^2\gamma^4)(1+\gamma^1\gamma^3)\Psi. \tag{2.4}$$

In contrast to Ref. 1 throughout the paper, we take the fourth variable to be imaginary.

In the standard representation of the Dirac matrices,

$$\gamma^k=\begin{pmatrix}0 & \sigma^k\\ \sigma^k & 0\end{pmatrix},\quad k=1,2,3,\quad \gamma^4=i\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix} \tag{2.5}$$

(here and after γ^4 is anti-Hermitian), instead of (2.2) and (2.3) we can write

$$k\xi_1=\{i\partial_3+(E-A_4)\}\xi_2, \tag{2.6}$$
$$\{i\partial_3-(E-A_4)\}\xi_1=-k\xi_1, \tag{2.7}$$
$$i(m_0-k)A+(-\partial_1+(k_2+A_2))B=0, \tag{2.8}$$
$$i(m_0+k)B+(\partial_1+(k_2+A_2))A=0, \tag{2.9}$$

where k is an eigenvalue of the operators of Eqs. (2.2) and (2.3).

Then we have for the structure of a bispinor

$$\tilde{\Psi}=\begin{pmatrix}A(\xi_1-\xi_2)\\ B(\xi_1+\xi_2)\\ iA(\xi_1+\xi_2)\\ -iB(\xi_1-\xi_2)\end{pmatrix}. \tag{2.10}$$

Note that the potentials

$$A_l(x^k),\quad l\neq k,\quad A_i(x^j),\quad i\neq j, \tag{2.11}$$

in our nomenclature correspond to the parallel electric and magnetic fields ($\vec{E}\,\|\vec{H}$), respectively.

Let us consider situations where Eqs. (2.6)–(2.9) allow exact solutions in terms of special functions.

A. $A_4=cz$, $c=$const

After introducing a new variable y,

$$y=E-cz \tag{2.12}$$

(here E is the energy of the state), Eqs. (2.6) and (2.7) take the form

$$-k\xi_1+(-ic\,\partial_y+y)\xi_2=0, \tag{2.13}$$
$$-(ic\,\partial_y+y)\xi_1+k\xi_2=0. \tag{2.14}$$

After the change,

$$x=\sqrt{2/c}(E-cz)=\sqrt{2/c}y, \tag{2.15}$$

and from (2.13) and (2.14) we have the equation of the parabolic cylinder,

$$\{\partial^2_{xx} - (k^2 + ic)/2c + \tfrac{1}{4}x^2\}\xi_1 = 0, \tag{2.16}$$

for which exact solutions in terms of degenerate hypergeometric functions are well known:

$$\xi_1 = a_1 e^{-ix^2/4} M(-ia/2 + \tfrac{1}{4}, \tfrac{1}{2}, ix^2/2) + b_1 x e^{-ix^2/4} M(-ia/2 + \tfrac{3}{4}, \tfrac{3}{2}, ix^2/2). \tag{2.17}$$

Noting that Eqs. (2.13) and (2.14) are mutually conjugated we have

$$\xi^2 = a_2 e^{ix^2/4} M(ia/2 + \tfrac{1}{4}, \tfrac{1}{2}, -ix^2/2) + b_2 x e^{ix^2/4} M(ia/2 + \tfrac{3}{4}, \tfrac{3}{2}, -ix^2/2). \tag{2.18}$$

Taking into account the relation

$$M(\alpha,\beta,x) = e^x(\beta - \alpha,\beta, -x), \tag{2.19}$$

instead of (2.18) we can write

$$\xi_2^* = a_2^* e^{-ix^2/4} M(ia/2 + \tfrac{1}{4}, \tfrac{1}{2}, +ix^2/2) + b_2^* x e^{-ix2/4} M(ia/2 + \tfrac{3}{4}, \tfrac{3}{2}, +ix^2/2). \tag{2.20}$$

Substituting (2.17) and (2.20) in (2.13) we find the connections between coefficients to be

$$a_2 = i\sqrt{2c}/k, \quad b_1 = a_1^*, \quad b_2 = ik/\sqrt{2c}, \quad a_2 = b_1^*, \tag{2.21}$$

The determination of the functions A and B is trivial. Thus the problem is solved exactly.

B. $A_2 = cx$, $c = \text{const}$

Introducing the variable

$$z = k_2 + cx, \tag{2.22}$$

instead of Eqs. (2.8) and (2.9) we have

$$i(m_0 - k)A + (-c\,\partial_z + z)B = 0, \tag{2.23}$$

$$i(m_0 + k)B + (c\,\partial_z + z)A = 0. \tag{2.24}$$

As a result of the substitution

$$z = \sqrt{c/2}y, \tag{2.25}$$

we again have the equations of the parabolic cylinder,

$$\{\partial^2_{yy} - \tfrac{1}{4}y^2 + (k^2 - m_0^2 + c)/2c\}A = 0, \tag{2.26}$$

$$\{\partial^2_{yy} - \tfrac{1}{4}y^2 + (k^2 - m_0^2 - c)/2c\}B = 0, \tag{2.27}$$

for which exact solutions are known:

$$A = a_1 e^{-y^2/4} M(a/2 + \tfrac{1}{4}, \tfrac{1}{2}, y^2/2) + a_1 y e^{-y^2/4} M(a/2 + \tfrac{3}{4}, \tfrac{1}{2}, y^2/2), \tag{2.28}$$

$$B = b_1 e^{-y^2/4} M(a/2 + \tfrac{3}{4}, \tfrac{1}{2}, y^2/2) + b_2 y e^{-y^2/4} M(a/2 + \tfrac{5}{4}, \tfrac{3}{2}, y^2/2). \tag{2.29}$$

If

$$(k^2 - m_0^2)/4c = -n/2, \quad n = 1,2,3,..., \tag{2.30}$$

the degenerate hypergeometric functions reduce to the Hermite polynomials, and the quantization of energy takes place:

$$E = (m_0^2 + k_z^2 + 2nc)^{1/2}. \tag{2.31}$$

Returning to Eqs. (2.26) and (2.27), we have the relations between the coefficients of the solutions:

$$a_1 = -i\sqrt{2c}b_2/(m_0 - k), \quad b_1 = \{i\sqrt{2c}/(m_0 + k)\}a_2. \tag{2.32}$$

It is trivial to solve Eqs. (1.6) and (1.7) in this case. The problem is solved exactly.

C. $A_4 = \beta e^{\eta z}$, $\beta = \text{const}$, $\eta = \text{const}$

Equations (2.6) and (2.7) for this potential after the change

$$\mu = e^{\eta z} \tag{2.33}$$

take the form

$$-k\xi_1 + (i\eta\mu\partial_\mu + (E - \beta\mu))\xi_2 = 0, \tag{2.34}$$

$$(i\eta\mu\,\partial_\mu - (E - \beta\mu))\xi_1 + k\xi_2 = 0, \tag{2.35}$$

and therefore

$$\left\{\frac{\partial^2}{\partial\mu^2} + \frac{1}{\mu}\frac{\partial}{\partial\mu}\frac{1}{\mu^2}\left(-\frac{i\mu\beta}{\eta} + \frac{1}{\eta^2}((E - \beta\mu)^2 - k^2)\right)\right\}\xi_1 = 0. \tag{2.36}$$

The change

$$\xi_1 = \exp(-i\beta\mu/\eta)\mu^{ik_z/\eta}Y_1 \tag{2.37}$$

leads to the equation of degenerate hypergeometric functions

$$\zeta Y_1'' + Y_1'(2q - 1 - \zeta) + (iE/\eta - q)Y_1 = 0, \tag{2.38}$$

where

$$\zeta = 2i\beta\mu/\eta, \tag{2.39}$$

$$q = ik_z/\eta. \tag{2.40}$$

Therefore

$$\xi_1 = e^{-i\beta\mu/\eta}\left\{a_1\mu^{-ik_z/\eta}M\left(-\frac{ik_z}{\eta} - \frac{iE}{\eta} + 1, -\frac{2ik_z}{\eta} + L, \frac{2i\beta\mu}{\eta}\right) + a_2\mu^{ik_z/\eta}M\left(\frac{ik_z}{\eta} - \frac{iE}{\eta + 1}, \frac{2ik_z}{\eta} + L, \frac{2i\beta\mu}{\eta}\right)\right\}, \tag{2.41}$$

$$\xi_2 = e^{-i\beta\mu/\eta}\left\{b_1\mu^{ik_z/\eta}M\left(\frac{ik_z}{\eta} - \frac{iE}{\eta}, \frac{2ik_z}{\eta} + L, \frac{2i\beta\mu}{\eta}\right) + b_2\mu^{-ik_z/\eta}M\left(-\frac{ik_z}{\eta} - \frac{iE}{\eta}, -\frac{2ik_z}{\eta} + L, \frac{2i\beta\mu}{\eta}\right)\right\}. \tag{2.42}$$

According (2.35) and (2.36) the coefficients are connected by the relation

$$a_1^* = b_1, \quad a_2^* = b_2, \quad a_1/a_2 = i(E + k_z)/k. \tag{2.43}$$

The problem is solved exactly.

In the case of free motion $[\beta \to 0, \quad M(\alpha,\beta,0) = 1]$, if β is not an integer number then we have superposition of two free waves running in opposite directions.

D. $A_2 = \beta e^{\eta x}, \beta = \text{const}, \eta = \text{const}$

After the change

$$\mu = e^{\eta x}, \tag{2.44}$$

instead of (2.8) and (2.9) we have

$$i(m - k)A + (-\eta\mu\, \partial_\mu + (k_2 + \beta\mu))B = 0, \tag{2.45}$$

$$i(m + k)B + (\eta\mu\, \partial_\mu + (k_2 + \beta\mu))A = 0. \tag{2.46}$$

Thus again we have the exact solution

$$A = e^{-\beta\mu/\eta}\Big\{a_1\mu^{ik_1/\eta}M\Big(\frac{k_2 + ik_1}{\eta}, \frac{2ik_1}{\eta} + L, \frac{2\beta\mu}{\eta}\Big) + a_2\mu^{-ik_1/\eta}M\Big(\frac{k_2 - ik_1}{\eta}, -\frac{2ik_1}{\eta} + L, \frac{2\beta\mu}{\eta}\Big)\Big\} \tag{2.47}$$

$$B = e^{-\beta\mu/\eta}\Big\{b_1\mu^{ik_1/\eta}M\Big(\frac{k_2 + ik_1}{\eta} + 1, \frac{2ik_1}{\eta} + 1, \frac{2\beta\mu}{\eta}\Big) + b_2\mu^{-ik_1/\eta}\mu\Big(\frac{k_2 - ik_1}{\eta} + 1, -\frac{2ik_1}{2} + 1, \frac{2\beta\mu}{\eta}\Big)\Big\}. \tag{2.48}$$

The coefficients are related as follows:

$$b_1 = ia_1(k_2 + ik_1)/(m_0 + k), \quad b_2 = ia_2(k_2 - ik_1)/(m_0 + k). \tag{2.49}$$

E. $A_4 = l/z, l = \text{const}$

Now we have instead of (2.6) and (2.7)

$$-k\xi_1 + (i\partial_z + (E - l/z))\xi_2 = 0, \tag{2.50}$$

$$(i\,\partial_z - (E - l/z))\xi_1 + k\xi_2 = 0. \tag{2.51}$$

Passing to the second-order equations we have the well-known Whittaker ones:

$$\{\partial^2_{zz} + il/z^2 + (E - l/z)^2 - k^2\}\xi_1 = 0, \tag{2.52}$$

$$\{\partial^2_{zz} - il/z^2 + (E - l/z)^2 - k^2\}\xi_2 = 0. \tag{2.53}$$

The corresponding solutions have the form

$$\xi_1 = e^{-y/2}\{y^{1-il}a_1M(1 - il - \tilde{k}, 2 - 2il, y) + a_2M(il - \tilde{k}, 2il, y)y^{il}\}, \tag{2.54}$$

$$\xi_2 = e^{-y/2}\{y^{1-il}b_1M(1 + il - \tilde{k}, 2 + 2il, y) + b_2M(-il - \tilde{k}, -2il, y)y^{il}\}, \tag{2.55}$$

where

$$y = 2\sqrt{k^2 - E^2}z, \tag{2.56}$$

$$\tilde{k} = -lE/\sqrt{k^2 - E^2}. \tag{2.57}$$

The connections between the coefficients may be deduced from substitution of (2.55) and (2.56) into (2.51) and (2.52):

$$b_1 = -ia_2k/[2\sqrt{k^2 - E^2}(1 + 2il)], \quad b_2 = -[2\sqrt{k^2 - E^2}(1 - 2il)]/ik. \tag{2.58}$$

The behavior of solutions in the asymptotic region when $z \to \infty$ is given by

$$\xi_1 = e^{-y/2}y^{1-il}c_1 + e^{-y/2}y^{il}c_2, \tag{2.59}$$

$$\xi_2 = e^{-y/2}y^{1+il}d_1 + e^{-y/2}y^{-il}d_2. \tag{2.60}$$

F. $A_2 = l/z, l = \text{const}$

Equations (2.8) and (2.9) with such potential take the form

$$i(m - k)A + (-\partial_x + (k_2 + l/x))B = 0, \tag{2.61}$$

$$i(m + k)B + (\partial_x + (k_2 + l/x))A = 0. \tag{2.62}$$

Analogously to those presented previously, these equations may be solved exactly; namely, we have

$$A = e^{-y/2}\{a_1y^{l+1}M(1 + l + 2lk_2/\lambda, 2 + 2l, y) + a_2y^{-l}M(-l + 2lk_2/\lambda, 1 - 2l, y)\}, \tag{2.63}$$

$$B = e^{-y/2}\{a_1\{i(m_0 + k)/2\sqrt{k_2^2 + m_0^2 - k^2}(2l + 1)\}^{-1} \times M(l + 2lk_2/\lambda, 2ly)y^l + a_1\{(m_0 - k)/2i\sqrt{k_2^2 + m_0^2 - k^2}(1 - 2l)\}^{-1} \times M(1 - l + 2lk_2/\lambda, 2 - 2l, y)y^{1-l}\}. \tag{2.64}$$

III. SOLUTION OF THE DIRAC EQUATION IN CYLINDRICAL COORDINATES

The Dirac equation in general cylindrical coordinates

$$x = e^\mu \cos\theta, \quad y = e^\mu \sin\theta, \quad z, \quad t \tag{3.1}$$

in the presence of vector fields in a diagonal tetrad gauge takes the form

$$\Big\{\gamma^1\Big(\frac{\partial_\mu}{e^\mu} - iA_\mu\Big) + \gamma^2\Big(\frac{\partial_\theta}{e^\mu} - iA_\theta\Big) + \gamma^3(\partial_z - iA_z) + \gamma^4(\partial_t - iA_4) + m_0\Big\}\Psi = 0. \tag{3.2}$$

In the case where the vector potential has only one component, i.e., $A_\mu = A_\theta = A_z = 0$, $A_4 = A_4(\mu)$, Eq. (3.2) allows a separation of variables such as

$$\hat{K}_{12} = -\gamma^4\gamma^2\,\partial_\mu/e^\mu + \gamma^4\gamma^1\,\partial_\theta/e^\mu + i\gamma^1\gamma^2(A_4 + E), \tag{3.3}$$

$$\hat{K}_z = -\gamma^1\gamma^2\gamma^3\gamma^4\,\partial_z + \gamma^1\gamma^2\gamma^4m_0, \tag{3.4}$$

$$[\hat{K}_{12}, \hat{K}_z] = 0, \quad (\hat{K}_{12} + \hat{K}_z)\tilde{\Psi} = 0. \tag{3.5}$$

Here it is taken into account that the operator of Eq. (2.2) and the operator of energy $-i\,\partial_t$ are commute.

In order to solve Eq. (3.2) and to find the explicit form of the bispinor let us transform from standard representation of the Dirac matrices to some "almost standard" representation through the transformation

$$\gamma \to \hat{S}^{-1}\gamma\hat{S}, \quad \hat{S} = (1/\sqrt{2})(1 + i\gamma^3\gamma^4), \tag{3.6}$$

which corresponds to the transitions $\gamma^3 \to i\gamma^4$ and $\gamma^4 \to i\gamma^3$.

The operator $\hat{K}_z$ leads to the equations

$$-(\partial_z + m_0)\chi = k\varphi,$$
$$(\partial_z - m_0)\varphi = k\chi, \tag{3.7}$$

$$\tilde{\Psi} = \begin{pmatrix} \tilde{\varphi} \\ \tilde{\chi} \end{pmatrix}, \tag{3.8}$$

where

$$\tilde{\Psi} = S^{-1}\Psi = \Psi_0 \exp\{i(k_z z - Et)\}, \tag{3.9}$$

$$\Psi_0 = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}. \tag{3.10}$$

Thus we have

$$k^2 = k_z^2 + m_0^2, \tag{3.11}$$

where k is an eigenvalue number of the operator $\hat{K}_z$.

Equation (3.3) leads to the system

$$\{(\sigma^1/e^\mu)\partial_\mu + i\sigma^2 m/e^\mu - \sigma^3(A_4 + E)\}\varphi = k\varphi, \tag{3.12}$$

$$\{(\sigma^1/e^\mu)\partial_\mu + i\sigma^2 m/e^\mu - \sigma^3(A_4 + E)\}\chi = k\chi, \tag{3.13}$$

where m is an eigenvalue of the operator $-i\,\partial_\theta$. The number m takes the values permitted by the corresponding boundary conditions in the Cartesian tetrad gauge.

The transformation $\hat{S}(\theta)$ that achieves the transition from the diagonal tetrad gauge to the Cartesian one has the form

$$\hat{S}(\theta) = e^{-\mu/2}\exp(-(\theta/2)\gamma^1\gamma^2), \tag{3.14}$$

i.e.,

$$\Psi_{\text{Cart}} = \hat{S}^{-1}(\theta)\Psi_{\text{Diag}}. \tag{3.15}$$

According to (3.14), $\hat{S}(\theta + 2\pi) = -\hat{S}(\theta)$ and therefore

$$\Psi_{\text{Diag}}(\theta + 2\pi) = -\Psi_{\text{Diag}}(\theta). \tag{3.16}$$

So we have $m = n + \frac{1}{2}$, where n is an integer number.

Equation (3.12) leads to the system

$$\{e^{-\mu}(\partial_\mu + m)\}\varphi_2 = \{k + (A_4 + E)\}\varphi_1, \tag{3.17}$$

$$\{e^{-\mu}(\partial_\mu - m)\}\varphi_1 = \{k - (A_4 + E)\}\varphi_2, \tag{3.18}$$

where

$$\varphi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}. \tag{3.19}$$

The system of equations (3.17) and (3.18) allows exact solutions in the case of the potential $A_4 = l/e^\mu$.

A. $l=0$, free case

In this case, instead of (3.17) and (3.18) we have

$$\rho^2\frac{\partial^2\varphi_1}{\partial\rho^2} - \{m(m-1) + (k^2 - E^2)\rho^2\}\varphi_1 = 0, \tag{3.20}$$

$$\rho^2\frac{\partial^2\varphi_2}{\partial\rho^2} - \{m(m+1) + (k^2 - E^2)/\rho^2\}\varphi_2 = 0, \tag{3.21}$$

where a new variable $e^\mu = \rho$ is introduced.

The solutions of (3.20) and (3.21) are well known:

$$\varphi_1 = c_1\sqrt{\rho}J_{m-1/2}(i\sqrt{k^2 - E^2}\rho), \tag{3.22}$$

$$\varphi_2 = c_2\sqrt{\rho}J_{m+1/2}(i\sqrt{k^2 - E^2}\rho). \tag{3.23}$$

After substitution of (3.22) and (3.23) in (3.17) and (3.18) we have the relation between the coefficients

$$c_1 = ic_2\sqrt{(k-E)/(E+K)}. \tag{3.24}$$

In the Cartesian gauge the radial solutions have the form

$$\varphi_1 \sim J_{m-1/2}(i\sqrt{k^2 - E^2}\rho), \tag{3.25}$$

$$\varphi_2 \sim J_{m+1/2}(i\sqrt{k^2 - E^2}\rho). \tag{3.26}$$

For the asymptotic behavior (large ρ) we have the well known cylindrical wave

$$J_{m+1/2}(i\sqrt{k^2 - E^2}\rho) \to \sqrt{\frac{2}{\pi}\frac{1}{\sqrt{k^2E^2}\rho}}\cos\left\{\sqrt{E^2 - k^2}\rho - \frac{\pi}{2}\left(m + \frac{1}{2}\right) - \frac{\pi}{4}\right\}. \tag{3.27}$$

B. $A_4 = l/e^\mu$, $l = \text{const}$

Taking into account the change

$$e^\mu = \rho, \tag{3.28}$$

we have instead of (3.12)

$$(\partial_\rho + m/\rho)\varphi_2 - (E + k + l/p)\varphi_1 = 0, \tag{3.29}$$

$$(\partial_\rho - m/\rho)\varphi_1 + (E - k + l/\rho)\varphi_2 = 0. \tag{3.30}$$

We find the solutions of (3.29) and (3.30) in the form

$$\varphi_1 = \sqrt{1 - E/k}\,e^{-\lambda\rho}\{F_1(2\lambda\rho) - F_2(2\lambda\rho)\}, \tag{3.31}$$

$$\varphi_2 = \sqrt{1 - E/k}\,e^{-\lambda\rho}\{F_1(2\lambda\rho) + F_2(2\lambda\rho)\}, \tag{3.32}$$

where

$$\lambda = \sqrt{k^2 - E^2}. \tag{3.33}$$

The substitution of these expressions in (3.16) and (3.17) leads us to the equations

$$\left(\partial_{\tilde\rho} + \frac{lE}{\lambda\tilde\rho} - 1\right)F_1 + \left(\frac{lk}{\lambda\tilde\rho} + \frac{m}{\tilde\rho}\right)F_2 = 0, \tag{3.34}$$

$$\left(\partial_{\tilde\rho} - \frac{lE}{\lambda\tilde\rho}\right)F_2 - \left(\frac{lk}{\lambda\tilde\rho} - \frac{m}{\tilde\rho}\right)F_1 = 0, \tag{3.35}$$

where $\tilde\rho = 2\lambda\rho$.

Let us consider the behavior of the solutions of these equations when $\tilde\rho \to 0$. Then it is convenient to write

$$F_1 = a_1\tilde\rho^\nu, \quad F_2 = a_2\tilde\rho^\nu. \tag{3.36}$$

Here a_1, a_2, and ν are constants connected according to (3.34) and (3.35) by the relations

$$(\nu + lE/\lambda)a_1 + (m + lk/\lambda)a_2 = 0, \tag{3.37}$$

$$(-m + lk/\lambda)a_1 + (-\nu + lE/\lambda)a_2 = 0, \tag{3.38}$$

from which we have

$$\nu = \sqrt{m^2 - l^2}. \tag{3.39}$$

Introducing the function

$$F_2 = \tilde\rho^\nu G_2(\tilde\rho), \tag{3.40}$$

we can see that the function G_2 satisfies the equation

$$\tilde\rho G_2'' + (2\nu + 1 - \tilde\rho)G_2' + (lE/\lambda - \nu)G_2 = 0. \tag{3.41}$$

Then

$$G_2 = cM(\nu - lE/\lambda, 2\nu + 1, \tilde\rho). \tag{3.42}$$

Taking into account Eqs. (3.34) and (3.35) we have finally

$$F_1 = -c[(\nu\lambda - lE)/(m\lambda - lk)]\tilde{\rho}^\nu \times M(\nu + 1 - lE/\lambda, 2\nu + 1, \tilde{\rho}), \quad (3.43)$$

$$F_2 = c\tilde{\rho}^\nu M(\nu - lE/\lambda, 2\nu + 1, \tilde{\rho}). \quad (3.44)$$

Formulas (3.42)–(3.44) define the functions of the variable $\tilde{\rho}$ for $E<k$ and $E>k$.

Note that in the case $E<k$ the energy spectrum will be discrete. In order to find the possible values of energy it is necessary to use the condition of finiteness of the functions F_1 and F_2 when $\rho \to \infty$.

We have the following behavior of the hypergeometric functions for $\rho \gg 1$ (see Ref. 6):

$$M(a,b,\rho) \to e^{-i\pi a}\frac{\Gamma(b)}{\Gamma(b-a)}\rho^{-a} + \frac{\Gamma(b)}{\Gamma(a)}\rho^{a-b}e^{\rho}, \quad (3.45)$$

from which we can see that $M(a,b,\rho)$ does not contain the exponentially increasing term if

$$1/\Gamma(a) = 0. \quad (3.46)$$

Application of this condition to the functions F_1 and F_2 gives

$$1/\Gamma(\nu - lE/\lambda) = 0. \quad (3.47)$$

Since the poles of the gamma function are negative integer numbers and zero, we have at last

$$\nu - lE/\lambda = -n. \quad (3.48)$$

So the possible values of the energy are

$$E_n = k/\{[l/(n+\nu)]^2 + 1\}^{1/2}. \quad (3.49)$$

The solution of the Dirac equation in the diagonal tetrad gauge has the structure

$$\Psi_{\text{Diag}} = \begin{pmatrix} \dfrac{k}{ik_z + m} & \begin{pmatrix} \varphi_1 \\ -\varphi_2 \end{pmatrix} \\ & \begin{pmatrix} \varphi_1 \\ -\varphi_2 \end{pmatrix} \end{pmatrix} \times \exp\{i(k_z z + m\varphi - Et)\}. \quad (3.50)$$

Now we shall investigate the possibilities for exact solutions of the Dirac equation (3.2) if we separate the variables first with pair separation of μ,θ from z,t, i.e.,

$$\left\{\gamma^\mu\left(\frac{\partial_\mu}{e^\mu} - iA_\mu\right) + \gamma^\theta\left(\frac{\partial_\theta}{e^\mu} - iA_\theta\right)\right\}\gamma^z\gamma^4\tilde{\Psi} = -ik\tilde{\Psi}, \quad (3.51)$$

$$\{\gamma^z(\partial_z - iA_z) + \gamma^4(\partial_t - iA_4)\}\gamma^z\gamma^4\tilde{\Psi} = ik\tilde{\Psi}. \quad (3.52)$$

Taking into account the standard representation of the Dirac matrices (2.5) we can rewite Eq. (3.51) in the form

$$\left\{i\sigma^2\left(\frac{\partial_\mu}{e^\mu} - iA_\mu\right) - i\sigma^1\left(\frac{\partial_\theta}{e^\mu} - iA_\theta\right) + k\right\}\Psi_L = 0, \quad (3.53)$$

$$\left\{-i\sigma^2\left(\frac{\partial_\mu}{e^\mu} - iA_\mu\right) + i\sigma^1\left(\frac{\partial_\theta}{e^\mu} - iA_\theta\right) + k\right\}\Psi_2 = 0, \quad (3.54)$$

$$\tilde{\Psi} = \begin{pmatrix}\Psi_1 \\ \Psi_2\end{pmatrix}, \quad (3.55)$$

which leads to

$$\Psi_1 = C(z,t)\sigma^3\Psi_2. \quad (3.56)$$

Using the explicit form of the Pauli matrices,

$$\sigma^1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad \sigma^2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma^3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad (3.57)$$

we have from (3.53)

$$k\varphi_1 - i\{(\partial_\theta/e^\mu - iA_\theta) + i(\partial_\mu/e^\mu - iA_\mu)\}\varphi_2 = 0, \quad (3.58)$$

$$k\varphi_2 + i\{i(\partial_\mu/e^\mu - iA_\mu) - (\partial_\theta/e^\mu - iA_\theta)\}\varphi_1 = 0. \quad (3.59)$$

The corresponding equivalent second-order equations take place:

$$\{d^2_{\rho\rho} - (k_\theta + 1)k_\theta/\rho^2 + d_\rho A_\theta - A^2_\theta - 2k_\theta A_\theta/\rho + k^2\}\varphi_1 = 0, \quad (3.60)$$

$$\{d^2_{\rho\rho} - (k_\theta - 1)k_\theta/\rho^2 + d_\rho A_\theta - A^2_\theta - 2k_\theta A_\theta/\rho + k^2\}\varphi_2 = 0. \quad (3.61)$$

These equations admit exact solutions for the potentials $A_\theta = Cr$ and $A_\theta = C$, where C is constant.

C. $A_\theta = Cr$, $C = \text{const}$

This case corresponds to a constant magnetic field

$$\mathbf{B} = 2C\hat{e}_z. \quad (3.62)$$

After the change

$$\mu = \gamma\rho^2, \quad (3.63)$$

(3.60) and (3.61) lead to

$$\left\{\mu\,\partial^2_{\mu\mu} + \frac{1}{2}\partial_\mu - \frac{(k_\theta+1)k_\theta}{4\mu} - \frac{C^2\mu}{4\gamma^2} + (k^2 - 2k_\theta C + C)\right\}\varphi_1 = 0, \quad (3.64)$$

$$\left\{\mu\,\partial^2_{\mu\mu} + \frac{1}{2}\partial_\mu - \frac{(k_\theta-1)k_\theta}{4\mu} - \frac{C^2\mu}{4\gamma^2} + (k^2 - 2k_\theta C - C)\right\}\varphi_2 = 0. \quad (3.65)$$

Introducing the new unknown functions

$$\varphi_{1,2} = e^{\alpha\mu}\mu^\beta f_{1,2}, \quad (3.66)$$

where α and β are constants, we find for f_1 and f_2 the differential equations

$$\left\{\mu\,\partial^2_{\mu\mu} + \left(\mu\left(2\alpha + \frac{\beta}{\mu}\right) + \frac{1}{2}\right)\partial_\mu + \mu\alpha^2 + 2\alpha\beta + \frac{(\beta^2-\beta)}{\mu} + \frac{\alpha}{2} + \frac{\beta}{2\mu} + \frac{k^2 - 2k_\theta C + C}{4\gamma} - \frac{(k_\theta+1)k_\theta}{4\mu} - \frac{C^2}{4\gamma^2}\mu\right\}f_1 = 0, \quad (3.67)$$

$$\left\{\mu\,\partial^2_{\mu} + \left(\mu\left(2\alpha + \frac{\beta}{\mu}\right) + \frac{1}{2}\right)\partial_\mu + \mu\alpha^2 + 2\alpha\beta + \frac{(\beta^2-\beta)}{\mu} + \frac{\alpha}{2} + \frac{\beta}{2\mu} + \frac{k^2 - 2k_\theta C - C}{4\gamma} - \frac{(k_\theta-1)k_\theta}{4\mu} - \frac{C^2}{4\gamma^2}\mu\right\}f_2 = 0. \quad (3.68)$$

These equations may be simplified if we set

$$\dot{\alpha}^2 - C^2/4\gamma^2 = 0, \quad (3.69)$$

$$\beta(\beta - \tfrac{1}{2}) - (k_\theta \pm 1)k_\theta/4 = 0, \quad (3.70)$$

i.e.,

$$\beta = (k_\theta + 1)/2 \quad \text{or} \quad k_\theta/2, \quad (3.71)$$

and for

$$C = \gamma, \quad \alpha = \tfrac{1}{2}, \quad (3.72)$$

Eqs. (3.67) and (3.68) take the form

$$\{\mu\, \partial^2_{\mu\mu} + (k_\theta + \tfrac{3}{2} - \mu)\partial_\mu + (k^2/4\gamma - 1)\} f_1 = 0, \quad (3.73)$$

$$\{\mu\, \partial^2_{\mu\mu} + (k_\theta + \tfrac{1}{2} - \mu)\partial_\mu + k^2/4\gamma\} f_2 = 0. \quad (3.74)$$

These equations admit polynomial solutions under the condition

$$k^2/4\gamma = n. \quad (3.75)$$

Then, instead of (3.73) and (3.74) we have

$$\{\mu\, \partial^2_{\mu\mu} + (m + 1 - \mu)\partial_\mu + (n-1)\} f_1 = 0, \quad (3.76)$$

$$\{\mu\, \partial^2_{\mu\mu} + (m - \mu)\partial_\mu + n\} f_2 = 0, \quad (3.77)$$

where

$$m = k_\theta + \tfrac{1}{2}. \quad (3.78)$$

The solutions of these equations are the Laguerre polynomials[6]

$$f_1 = c_1 L^{m}_{n-1}, \quad (3.79)$$

$$f_2 = c_2 L^{m-1}_{n-1}, \quad (3.80)$$

where c_1 and c_2 are constant coefficients.

Substituting (3.79) and (3.80) in (3.76) and (3.77) we have the relation between the coefficients

$$c_1/c_2 = 2\gamma^{1/2}/k. \quad (3.81)$$

After a unitary transformation that is the inverse of (3.14), we finally have the expressions of φ_1 and φ_2 in the Cartesian tetrad gauge:

$$\varphi_{1\,\text{Cart}} = e^{-(1/2)\mu}\mu^{m/2}(2\gamma^{1/2}/k) L^{m}_{n-1}, \quad (3.82)$$

$$\varphi_{2\,\text{Cart}} = e^{-(1/2)\mu}\mu^{(m-1)/2} L^{m-1}_{n-1}. \quad (3.83)$$

D. $A_\theta = C_\theta = \text{const}$

Now Eqs. (3.60) and (3.61) may be reduced to the form

$$\{\rho^2 \partial^2_{\rho\rho} - (k_\theta - 1)k_\theta - (C^2_\theta - k^2)\rho^2 - 2k_\theta C_\theta \rho\}\varphi_1 = 0, \quad (3.84)$$

$$\{\rho^2 \partial^2_{\rho\rho} - (k_\theta + 1)k_\theta - (C^2_\theta - k^2)\rho^2 - 2k_\theta C_\theta \rho\}\varphi_2 = 0. \quad (3.85)$$

After the change

$$z = \alpha\rho, \quad (3.86)$$

we have

$$\{\partial^2_{zz} + (-\tfrac{1}{4} + k/z + (\tfrac{1}{4} - (k_\theta + \tfrac{1}{2})^2)/z^2)\}\varphi_1 = 0, \quad (3.87)$$

$$\{\partial^2_{zz} + (-\tfrac{1}{4} + k/z + (\tfrac{1}{4} + (k_\theta - \tfrac{1}{2})^2)/z^2)\}\varphi_2 = 0, \quad (3.88)$$

where

$$\alpha = 2\sqrt{C^2_\theta - k^2}, \quad k = -2k_\theta C_\theta/\alpha; \quad \mu_{1,2} = \pm(k_\theta + \tfrac{1}{2}). \quad (3.89)$$

The solutions of these Whittaker equations are the functions

$$M_{k,\mu}(z) = e^{-z/2} z^{1/2+\mu} M(\tfrac{1}{2} + \mu - k, 1 + 2\mu, z). \quad (3.90)$$

Because of (3.87) and (3.88) the parameter μ takes two values and therefore we formally have two functions of type (3.90) as the solutions but one of them is singular at $z = 0$.

Note that Eqs. (3.60) and (3.61) are very simple with the potential $A_\theta = C_\theta$,

$$k\varphi_1 + (\partial_\rho - (k_\theta/\rho + C_\theta))\varphi_2 = 0, \quad (3.91)$$

$$k\varphi_2 = (\partial_\rho + (k_\theta/\rho + C_\theta))\varphi_1 = 0. \quad (3.92)$$

According to (3.90) we finally have

$$\varphi_1 = a e^{-z/2} z^{1/2+\mu_1} M(\tfrac{1}{2} + \mu_1 - k, 1 + 2\mu_1, z) + b e^{-z/2} z^{1/2-\mu_1} M(\tfrac{1}{2} - \mu_1 - k, 1 - 2\mu_1, z), \quad (3.93)$$

$$\varphi_2 = c e^{-z/2} z^{1/2+\mu_2} M(\tfrac{1}{2} + \mu_2 - k, 1 + 2\mu_2, z) + d e^{-z/2} z^{1/2-\mu_2} M(\tfrac{1}{2} - \mu_2 - k, 1 - 2\mu_2, z), \quad (3.94)$$

where

$$\mu_1 = k_\theta + \tfrac{1}{2}, \quad \mu_2 = k_\theta - \tfrac{1}{2}. \quad (3.95)$$

Substituting these solutions into (3.91) and (3.92) we have the relation between the coefficients

$$a/c = k/\alpha(2k_\theta + 1), \quad a/b = k/\alpha(2k_\theta - 1). \quad (3.96)$$

So instead of (3.93) and (3.94) we have

$$\varphi_1 = a e^{-z/2} z^{k_\theta+1} M(k_\theta + 1 - k, 2k_\theta + 2, z) + b e^{-z/2} z^{-k_\theta} M(-k_\theta - k, -k_\theta, z), \quad (3.97)$$

$$\varphi_2 = (a\alpha/k)(2k_\theta + 1) e^{-z/2} z^{k_\theta} M(k_\theta - k, 2k_\theta, z) + (bk/\alpha)(2k_\theta - 1)^{-1} e^{-z/2} z^{1-k_\theta} M \times (1 - k_\theta - k, 2 - 2k_\theta, z). \quad (3.98)$$

Note that Eqs. (3.51) and (3.52) are investigated in Ref. 7 with an application to quaternions. But the author only looks into some of the possibilities we consider and he does not deduce the relations between the coefficients of the solutions.

IV. SOLUTION OF THE DIRAC EQUATION IN PARABOLIC CYLINDRICAL COORDINATES IN THE ABSENCE OF FIELDS

Here we consider the most complex case of separation of variables in the Dirac equation in the search for an exact solution when Lame's functions depend on two variables even in the diagonal tetrad gauge. The general approach for such cases is proposed in Ref. 1. We consider here only the case of free motion, i.e., we investigate the "free" parabolic cylindrical spinor waves.

The free solutions of wave equations, i.e., in the absence of fields, play an important role both for understanding the nature of the spreading of the waves of different geometric types in space and for the construction of a theory of interactions. It is well known that the latest aspect is most elegantly achieved by free plane waves interacting locally. It is difficult to accept as truly free waves the well known cylindrical, spherical, and other types of waves because the amplitudes of such waves change with distance. We can, however, talk about a natural damping of divergent spherical or cylindrical waves. But how do we accept the "supernatural" amplification of convergent waves? What is the powerful source that generates them in the infinite distance and forces them to converge along fixed congruences? In this sense if we accept the primitive correspondence "free particle–free wave" and take into account that the free particle moves with constant impulse and energy, i.e, with constant direction and amplitude in the wave sense, we can accept as the only possible truly free wave the de Broglie's wave, which has won respect for its constant "love of freedom."

The free plane, cylindrical, and spherical waves are well known in the literature.[8,9] Here remembering about the conditional freedom of geometrically complex waves we present the "free" parabolic cylindrical wave, i.e., the free solution of the Dirac equation in the corresponding coordinates.

The Dirac equation in the parabolic cylindrical coordinates

$$x=(\mu^2-\nu^2)/2,\quad y=\mu\nu,\quad z,\quad t \tag{4.1}$$

takes the form

$$\{(\gamma^\mu/h)\partial_\mu+(\gamma^\nu/h)\partial_\nu+\gamma^z\partial_z+\gamma^4\partial_t+m_0\}\widetilde{\Psi}=0, \tag{4.2}$$

$$h=(\mu^2+\nu^2)^{1/2}. \tag{4.3}$$

As the single Lame's function in (4.1) depends on two variables μ and ν, it is natural first to fulfill the separations in pairs (μ,ν from z,t):

$$\{\gamma^\mu\partial_\mu+\gamma^\nu\partial_\nu)\gamma^z\gamma^4+ihk\}\Phi=0, \tag{4.4}$$

$$\{\gamma^4\partial_z+\gamma^z\partial_t+m_0\gamma^z\gamma^4-ik\}\Phi=0, \tag{4.5}$$

where k is the constant of separation.

Writing

$$\Phi=\begin{pmatrix}\varphi\\ \chi\end{pmatrix},\quad \varphi=\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix},\quad \chi=\begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix}, \tag{4.6}$$

we can find the structure of the bispinor Φ and its dependence on the variables z and t:

$$\Phi=\begin{pmatrix}\varphi\\ \{i\sigma^3(E-m_0)/(k_z-ik)\}\varphi\end{pmatrix}e^{i(k_z z-Et)}. \tag{4.7}$$

Here Φ is a new unknown bispinor connected with Ψ by the relation

$$\Phi=\gamma^z\gamma^4\widetilde{\Psi}, \tag{4.8}$$

E is the energy of the state, k_z is the z component of the wave vector, and φ depends only on variables μ and ν.

Using the standard representation of the Dirac matrices (2.5) we have from (4.4)

$$(\sigma^2\partial_\mu-\sigma^1\partial_\nu+kh)\varphi=0, \tag{4.9}$$

$$(-\sigma^2\partial_\mu+\sigma^1\partial_\nu+kh)\chi=0. \tag{4.10}$$

Because of the structure (4.7), Eqs. (4.9) and (4.10) are equivalent and therefore we only consider Eq. (4.9).

The presence of the expression $h=(\mu^2+\nu^2)^{1/2}$ in Eq. (4.9) does not allow us to separate variables immediately in this equation. However, after a similarity transformation by means of the operator

$$\hat{S}=(1/\sqrt{2})\{(h+\mu)^{1/2}+i\sigma^3(h-\mu)^{1/2}\} \tag{4.11}$$

(see Ref. 1) we have the more simple equation

$$\{(\sigma^2\partial_\mu+ik\mu)-i(\sigma^2\partial_\nu-ik\nu)\}\zeta=0,\quad \varphi=\hat{S}\zeta. \tag{4.12}$$

Introducing a new two-component spinor

$$\zeta=\{-i(\sigma^2\partial_\nu-ik\nu)-\sigma^3(\sigma^2\partial_\mu+ik\mu)\}\xi, \tag{4.13}$$

we have the second-order matrix differential equation

$$\{(\partial^2_{\mu\mu}+i\sigma^2k+k^2\mu^2)+(\partial^2_{\nu\nu}-i\sigma^2k+k^2\nu^2)\}\xi=0, \tag{4.14}$$

which allows the separation of variables

$$(\partial^2_{\mu\mu}+i\sigma^2k+k^2\mu^2)\xi=\lambda\xi, \tag{4.15}$$

$$(\partial^2_{\nu\nu}-i\sigma^2k+k^2\nu^2)\xi=-\lambda\xi, \tag{4.16}$$

where λ is the separation constant.

Now we make a transition to a new representation of Pauli matrices in order that $\sigma^2\to\sigma^3$. It is easy to see that such a transition may be achieved by means of the corresponding unitary transformation

$$\hat{U}=(1/\sqrt{2})(1+i\sigma^1),\quad \hat{U}^{-1}\sigma^2U=\sigma^3. \tag{4.17}$$

Then we have instead of (4.15) and (4.16)

$$(\partial^2_{\mu\mu}+i\sigma^3k+k^2\mu^2-\lambda)\eta=0, \tag{4.18}$$

$$(\partial^2_{\nu\nu}-i\sigma^3k+k^2\nu^2+\lambda)\eta=0. \tag{4.19}$$

Because of the separability of the variables μ and ν, we can now write

$$\eta_1=P(\mu)Q(\nu),\quad \eta_2=R(\mu)S(\nu), \tag{4.20}$$

where the unknown functions satisfy the equations

$$(\partial^2_{\mu\mu}+ik-\lambda+k^2\mu^2)P(\mu)=0, \tag{4.21}$$

$$(\partial^2_{\mu\mu}-ik-\lambda+k^2\mu^2)R(\mu)=0, \tag{4.22}$$

$$(\partial^2_{\nu\nu}-ik+\lambda+k^2\nu^2)Q(\nu)=0, \tag{4.23}$$

$$(\partial^2_{\nu\nu}+ik+\lambda+k^2\nu^2)S(\nu)=0. \tag{4.24}$$

The standard solutions of these equations of the parabolic cylinder are well known.[6]

Taking into account the transformation (4.17) we have a spinor structure

$$\xi=\frac{1}{\sqrt{2}}(1-i\sigma^1)\begin{pmatrix}PQ\\ RS\end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix}PQ-iRS\\ -iPQ+RS\end{pmatrix}. \tag{4.25}$$

Analogously in correspondence with (4.11) and (4.13) we have

$$\begin{aligned}\varphi_1=\{(h+\mu)^{1/2}+i\sigma^3(h-\mu)^{1/2}\}\{&i(Q_\nu-ikQ\nu)P\\ &+i(ikR_\mu-R\mu)S-(S_\nu+ikS\nu)R\\ &-(P_\mu+ikP\mu)Q,\end{aligned} \tag{4.26}$$

$$\begin{aligned}\varphi_2=\{(h+\mu)^{1/2}+i\sigma^3(h-\mu)^{1/2}\}\{&(Q_\nu-ikQ\nu)P\\ &+(ikR\mu-R_\mu)S-i(S_\nu+ikS\nu)R\\ &-i(P_\mu+ikP\mu)Q,\end{aligned} \tag{4.27}$$

Finally with (4.7) and (4.8) we obtain the exact solution of Eq. (4.2):

$$\tilde{\Psi} = \begin{pmatrix} i\dfrac{E-m}{k_z - ik}\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix} \\ i\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}\end{pmatrix} e^{i(k_z z - Et)}. \tag{4.28}$$

Regarding the discussion of the exact solution (4.28), i.e., the parabolic cylindrical wave, note that the space dependence of this wave on the x and y variables is determined by the equation on the μ and ν variables (4.4) separated from z and t (4.5). It allows us to affirm that the vector potential of the kind $A_4(z)$ or $A_z(t)$ may be included in the solution (4.28) in a trivial way.

For example, in the case $A_4(z)$, instead of (4.5) we have

$$\{\gamma^4\,\partial_z + \gamma^z(\partial_t - iA_4(z)) + m_0\gamma^z\gamma^4\}\Phi = ik\Phi. \tag{4.29}$$

Taking into account the stationarity of the state after the similarity transformation

$$\{H\}\to\hat{S}^{-1}\{H\}\hat{S},\quad \eta = \hat{S}\Phi = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}, \tag{4.30}$$

by means of the operator[1]

$$S = \tfrac{1}{2}(1 - \gamma^\mu\gamma^z)(1 + i\gamma^\nu\gamma^4), \tag{4.31}$$

we have the equations

$$\begin{aligned}&\{\sigma^1\,\partial_z - \sigma^2(E + A_4)\}\eta_2 - (i\sigma^3 m_0 - k)\eta_1 = 0,\\ &\{\sigma^1\,\partial_z - \sigma^2(E + A_4)\}\eta_1 - (i\sigma^3 m_0 - k)\eta_2 = 0.\end{aligned} \tag{4.32}$$

According to the structure of these equations,

$$\eta_1 = \eta_2\begin{pmatrix}a(z)\\ b(z)\end{pmatrix}, \tag{4.33}$$

where the functions a and b in the standard representation of the Pauli matrices satisfy

$$\begin{aligned}&\{\partial_z + i(E + A_4)\}b(z) - (im_0 - k)a(z) = 0,\\ &\{\partial_z - i(E + A_4)\}a(z) + (im_0 + k)b(z) = 0.\end{aligned} \tag{4.34}$$

By (4.30) and (4.31) we have

$$\Phi = \begin{pmatrix}\varphi_1(\mu,\nu)(a+b)\\ \varphi_2(\mu,\nu)(a+b)\\ \varphi_3(\mu,\nu)(a-b)\\ \varphi_4(\mu,\nu)(a-b)\end{pmatrix} e^{-iEt}, \tag{4.35}$$

where φ_1, φ_2, φ_3, and φ_4 are components of the bispinor solution of Eq. (4.4).

Analogously, the vector potential $A_z(t)$ may be introduced.

Finally we consider the asymptotic behavior of the solution (4.28) to $x\to\infty$, $y\to\infty$. Because of (4.1) the conditions $x\to\infty$, $y\to\infty$ are equivalent to the requirement $\mu\to\infty$ for any ν (or inversely $\nu\to\infty$ for any μ).

According to the relation[6]

$$M(a,c,z) \sim \frac{\Gamma(c)}{\Gamma(a)}\,e^z z^{(a-c)} + \frac{\Gamma(c)}{\Gamma(c-a)}\,z^{-a}e^{i\pi a}, \tag{4.36}$$

we find

$$\varphi \sim e^{ik\rho}\begin{pmatrix}\mu^{i\lambda/2k}A + e^{ik\mu^2}\mu^{-i\lambda/2k}B\\ \mu^{i\lambda/2k}C + e^{ik\mu^2}\mu^{-i\lambda/2k}D\end{pmatrix}, \tag{4.37}$$

where the functions A, B, C, and D depend only on ν and

$$\rho = (x^2 + y^2)^{1/2} = (\mu^2 + \nu^2)/2. \tag{4.38}$$

Noting the explicit connection of x,y with μ,ν (or, equivalently, with ρ), we have

$$\varphi \sim \begin{pmatrix}(x+\rho)^{i\lambda/4k}Ae^{-ik\rho} + e^{ikx}(x+\rho)^{-i\lambda/4k}B\\ (x+\rho)^{i\lambda/4k}Ce^{-ik\rho} + e^{ikx}(x+\rho)^{-i\lambda/4k}D\end{pmatrix}, \tag{4.39}$$

for large x and y (or, equivalently, for large ρ). Then our solution takes the form of the superposition of free plane and cylindrical waves with amplitude modulation.

V. DISCUSSION

The possibilities for the method of separation of variables in the Dirac equation in the presence of vector fields in the search for exact solutions are not exhausted by the solutions considered here. They only demonstrate the usefulness of our method.[1] Even in the statement of the problem we have other possibilities for exact solutions. These are the solutions in the same geometries and for the same external fields but with other orders of separation of variables. Then we have the natural problem, namely, the problem of the unique solution (Cauchy problem). Note that we have not considered the boundary conditions for our solutions. Only after the introduction of boundary conditions can we select the physical solutions. Note, too, that as the Dirac matrices have the special structure and each Dirac 4×4 matrix consists of the 2×2 matrices, each component of the Dirac bispinor finally is determined by the second-order differential equation for each variable, whose solution contains two linearly independent functions. So the number of mathematically different solutions will be determined by the number of combinations of different solutions of the equations on the separated variables. The physical solution will be determined by criteria of finiteness, square integrability, boundary conditions, and others.

In any case, we have demonstrated the possibilities for the search for exact solutions of the Dirac equation by means of our method proposed in Ref. 1 for all types of variables, namely, in the "plane" Cartesian coordinates, in the curvilinear cylindrical coordinates, and finally in the exotic parabolic cylindrical coordinates where Lame's function does not separate its variables in the diagonal tetrad gauge.

[1] G. V. Shishkin and V. M. Villalba, J. Math. Phys. **30**, 2132 (1989).
[2] D. M. Volkov, Z. Phys. **94**, 3 (1936).
[3] D. M. Volkov, Sov. Phys. JETP **7**, 1286 (1937).
[4] F. I. Fedorov, Sov. Phys. JETP **35**, 495 (1958).
[5] F. I. Fedorov, Dokl. Acad. Nauk USSR **174**, 334 (1967) (in Russian).
[6] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[7] A. Hautot, J. Math. Phys. **13**, 710 (1972).
[8] A. S. Davydov, *Quantum Mechanics* (Pergamon, Oxford, 1965).
[9] A. I. Akheizer and V. B. Berestetskii, *Quantum Electrodynamics* (Science, Moscow, 1969), in Russian.

The equivariant inverse problem and the uniqueness of the Yang–Mills equations

M. C. López and R. J. Noriega
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina
C. G. Schifini
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina and CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

(Received 23 August 1988; accepted for publication 3 May 1989)

The equivariant inverse problem for Yang–Mills-type Euler–Lagrange expressions is solved in the affirmative. This leads to a proof of the uniqueness of the Yang–Mills equations.

I. INTRODUCTION

Let $P(M,G,\pi)$ be a principal fiber bundle with base space M, total space P, and structural group G. Let $n=\dim M$, $r=\dim G$. For α,β non-negative integers we define $V:=T^\alpha_\beta(LG)$, the space of α contravariant, β covariant tensors on the Lie algebra LG of G and $\rho: G \to \mathrm{GL}(V)$ by

$$\rho:=\mathrm{Ad}\otimes\cdots\otimes\mathrm{Ad}\otimes\widetilde{\mathrm{Ad}}\otimes\cdots\otimes\widetilde{\mathrm{Ad}}, \tag{1}$$

where

$$\widetilde{\mathrm{Ad}}(a)(\eta)(Xe):=\eta(\mathrm{Ad}(a)(Xe)), \tag{2}$$

and Ad is the adjoint representation of G. Let z be the local chart around e in G given by exp.

A *gauge field* is a connection form w on P. If U is an open set in M then a *gauge* is a pair (U,σ) where $\sigma: U \to P$ is a smooth section of π. For a gauge (U,σ) let $\omega_\sigma:=\sigma^*\omega$. Then ω_σ is an LG-valued one-form defined on U. If (x,V) is a local chart in M such that $U\cap V\neq\varnothing$ then

$$\omega_\sigma=(A^\alpha_i\,dx^i)e_\alpha$$

(latin letters run from 1 to n, greek letters run from 1 to r and we use the summation convention). The A^α_i are called the *gauge potentials* of ω associated to (U,σ), (x,V), and e_α.

If (U,σ) and (U',σ') are two gauges such that $\sigma(m)=\sigma'(m)$ for some m in $U\cap U'$, then there is a smooth function $\psi: U\cap U' \to G$ such that $\sigma\cdot\psi=\sigma'$ in $U\cap U'$. It is well known that

$$A'^\alpha_i=\mathrm{Ad}^\alpha_\beta\circ\psi^{-1}A^\beta_i+l^\alpha_\beta\circ\psi\frac{\partial\psi^\beta}{\partial x^i}, \tag{3}$$

where $\psi^\beta:=z^\beta\circ\psi$; $l^\alpha_\beta\,dz^\beta$ are the left invariant one-forms generated by the dual basis of e_α, and $\mathrm{Ad}(a)e_\alpha=\mathrm{Ad}^\beta_\alpha(a)e_\beta$.

We say that T is a *gauge tensor field of type* $(V;r,s,w)$ if it gives for every gauge (U,σ) a V-valued relative tensor field T_σ of type (r,s,w) defined over U. We say that T is a gauge tensor field of type $(\rho;r,s,w)$ if furthermore

$$T_{\sigma'}=\rho(\psi^{-1})T_\sigma \text{ in } U\cap U', \tag{4}$$

where ρ is given by (1).

The coefficients of the curvature form, defined as

$$F^\alpha_{ij}:=A^\alpha_{j,i}-A^\alpha_{i,j}+C^\alpha_{\beta\gamma}A^\beta_iA^\gamma_j, \tag{5}$$

where $C^\alpha_{\beta\gamma}$ are the structure constants associated to e_α, are the components of a gauge tensor F of type $(\mathrm{Ad};0,2,0)$.

If we have a Lorentz metric g_{ij} on M, the gauge covariant derivative of F is defined as

$$F^\alpha_{ij\|h}:=F^\alpha_{ij,h}-F^\alpha_{kj}\Gamma^k_{ih}-F^\alpha_{ik}\Gamma^k_{jh}+F_{ij}C^\alpha_{\beta\gamma}A^\beta_h, \tag{6}$$

where Γ^k_{jh} are the Christoffel symbols associated to g_{ij}.

The Yang–Mills equations are

$$B_{\alpha\beta}F^{\alpha ij}{}_{\|j}=J^i_\beta, \tag{7}$$

where $B_{\alpha\beta}$ are the coefficients of an Ad G-invariant bilinear symmetric form on LG. They can be obtained through a variational principle as follows. If $L=L(g_{ij};A^\alpha_i;A^\alpha_{i,j})$, then through a variation of A^α_i one obtains the Euler–Lagrange equations

$$E^i_\alpha(L)=J^i_\alpha, \tag{8}$$

where

$$E^i_\alpha(L)=\frac{\partial L}{\partial A^\alpha_i}-\frac{\partial}{\partial x^j}\left(\frac{\partial L}{\partial A^\alpha_{i,j}}\right). \tag{9}$$

To allow for possible interaction with a gravitational field, it is of interest to define the energy momentum tensor

$$T^{ij}=E^{ij}(L)=\frac{\partial L}{\partial g_{ij}}. \tag{10}$$

If one chooses L as

$$L=B_{\alpha\beta}F^{\alpha ij}F^\beta{}_{ij}, \tag{11}$$

then Eq. (8) becomes (7) and Eq. (10) becomes

$$T^{ij}=B_{\alpha\beta}(F^{\alpha i}{}_kF^{\beta jk}-\tfrac14 g^{ij}F^{\alpha hk}F^\beta_{hk}). \tag{12}$$

The left-hand sides of (7) and (12) are gauge tensors; the same is true for the Lagrangian (11). However, while the gauge invariance (i.e., being a gauge tensor) is mandatory for the field equations, it is not so for the Lagrangian, because in general it has no physical meaning. The equivariant inverse problem[1] for this particular case could be stated as follows. If L is a function of the type

$$L=L(g_{ij};F^\alpha_{ij}), \tag{13}$$

such that $E^i_\alpha(L)$ and $E^{ij}(L)$ are gauge tensors, is there a gauge invariant Lagrangian $\hat L$ such that $E^i_\alpha(L)=E^i_\alpha(\hat L)$ and $E^{ij}(L)=E^{ij}(\hat L)$? We will prove that this is the case, a

 0022-2488/89/102382-06$02.50

fact that restricts severely the form of the function given by (13). Using this restriction, we will prove that Yang–Mills equations are the only possibility, thereby proving the uniqueness of these equations. It is worthwhile to note that the F_{ij} satisfy identically the following[2]:

$$*F^{\alpha ij}{}_{\|j} = 0. \qquad (14)$$

This problem was treated by Horndeski.[3] He added the following strong hypothesis about the form of the Lagrangian in flat space-time: that it reduces to $4g^{1/2}\,\gamma_{\alpha\beta}\,F^{\beta ij}{}_{\|j}$, where $\gamma_{\alpha\beta}$ are the components of a symmetric Ad G invariant bilinear form on LG. This allows him to find the general form of the Lagrangian. We do not assume such a hypothesis.

II. THE EQUIVARIANT INVERSE PROBLEM

We will work in a four-dimensional space-time, and *we will only assume in this section that $E^{ij}(L)$ is a gauge tensor.* This leads to Theorem 1. Let us denote $L^{ij} = E^{ij}(L)$ and let us suppose that $r = \dim G = 3$. If we consider H^{ihjk} given by

$$\sqrt{g}\,H^{ihjk} = L^{ij}g^{hk} - L^{ik}g^{hj} - L^{hj}g^{ik} + L^{hk}g^{ij}, \qquad (15)$$

then it is clear that H^{ihjk} is skew-symmetric in i,h and skew-symmetric in j,k. Then, by a previous result[4] we can write

$$H^{ihjk} = a'_{\alpha\beta}\,F^{\alpha ih}\,F^{\beta jk} + b'_{\alpha\beta}\,F^{\alpha ih}\,*F^{\beta jk} + c_{\alpha\beta}\,*F^{\alpha ih}\,F^{\beta jk} + d_{\alpha\beta}\,*F^{\alpha ih}\,*F^{\beta jk}, \qquad (16)$$

where the coefficients are gauge invariant scalar densities and $1 \leqslant \alpha, \beta \leqslant 3$. Since $H^{jhik} = H^{ikjh}$, it follows at once that

$$a'_{\alpha\beta} = a'_{\beta\alpha}, \quad b'_{\alpha\beta} = c_{\beta\alpha}, \quad d_{\alpha\beta} = d_{\beta\alpha}. \qquad (17)$$

Replacing in (16) and multiplying by g_{hk} it follows that

$$(1/\sqrt{g})L^{ij} = a'g^{ij} + 2a_{\alpha\beta}\,F^{\alpha i}{}_k\,F^{\beta kj} + b'_{\alpha\beta}(F^{\alpha i}{}_k{*F}^{\beta jk} + F^{\alpha j}{}_k\,*F^{\beta ik}), \qquad (18)$$

taking account that

$$*F^{\alpha i}{}_k\,*F^{\beta jk} = -\tfrac{1}{2}g^{ij}\phi^{\alpha\beta} + F^{\alpha j}{}_k\,F^{\beta ik}, \qquad (19)$$

where

$$\phi^{\alpha\beta} = F^{\alpha ij}\,F^{\beta}{}_{ij}. \qquad (20)$$

Decomposing $b'_{\alpha\beta}$ in its symmetrical and skew-symmetrical parts, and taking account that

$$F^{\alpha j}{}_k\,*F^{\beta ik} + F^{\beta j}{}_k\,*F^{\alpha ik} = \tfrac{1}{2}g^{ij}\,\psi^{\alpha\beta}, \qquad (21)$$

where

$$\psi^{\alpha\beta} = F^{\alpha ij}\,*F^{\beta}{}_{ij}, \qquad (22)$$

then it follows that

$$L^{ij} = \sqrt{g}\,\{ag^{ij} + a_{\alpha\beta}\,T^{\alpha\beta ij} + b_{\alpha\beta}\,S^{\alpha\beta ij}\}. \qquad (23)$$

We have proved the following lemma.

Lemma 1: If $L^{ij} = L^{ij}(g_{hk}; F^{\alpha}_{hk})$ is a tensorial concomitant symmetric in i,j, then L^{ij} has the form (23), where $a_{\alpha\beta} = a_{\beta\alpha}$, $b_{\alpha\beta} = -b_{\beta\alpha}$, and a, $a_{\alpha\beta}$, and $b_{\alpha\beta}$ are gauge invariant scalars. Also,

$$T^{\alpha\beta ij} = F^{\alpha i}{}_s\,F^{\beta sj} + F^{\alpha j}{}_s\,F^{\beta si}, \qquad (24)$$

$$S^{\alpha\beta ij} = F^{\alpha i}{}_s\,*F^{\beta sj} + F^{\alpha j}{}_s\,*F^{\beta si}. \qquad (25)$$

By the way, since (23) is valid for any symmetric tensor L^{ij}, it follows that the ten tensors g^{ij}, $T^{\alpha\beta ij}$, and $S^{\alpha\beta ij}$ are linearly independent $(1 \leqslant \alpha, \beta \leqslant 3)$.

By a previous result[5] it is known that

$$a = a(\phi^{\alpha\beta}; \psi^{\alpha\beta}), \quad a_{\alpha\beta} = a_{\alpha\beta}(\phi^{\mu\nu}; \psi^{\mu\nu}), \quad b_{\alpha\beta} = b_{\alpha\beta}(\phi^{\mu\nu}; \psi^{\mu\nu}) \qquad (26)$$

in a dense subset of the set of concomitance variables (i.e., as long as $F^1, F^2, F^3, *F^1, *F^2, *F^3$ is a linearly independent set; see Refs. 4 and 5), and where $\phi^{\alpha\beta}$ and $\psi^{\alpha\beta}$ are given by (20) and (22).

Since $L^{ij} = \partial L/\partial g_{ij}$, then $L^{ij;hk} = L^{hk;ij}$. In view of (26), this can be written, after some reductions as

$$\left(a_{\alpha\beta} - 2\frac{\partial a}{\partial\phi^{\alpha\beta}} - \frac{\partial a_{\alpha\beta}}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}\right)T^{\alpha\beta ij}g^{hk} + \left(-a_{\alpha\beta} + 2\frac{\partial a}{\partial\phi^{\alpha\beta}} + \frac{\partial a_{\alpha\beta}}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}\right)T^{\alpha\beta hk}g^{ij}$$
$$+ 2\left(\frac{\partial a_{\alpha\beta}}{\partial\phi^{\mu\nu}} - \frac{\partial a_{\mu\nu}}{\partial\phi^{\alpha\beta}}\right)T^{\alpha\beta ij}T^{\mu\nu hk} + 2\frac{\partial b_{\alpha\beta}}{\partial\phi^{\mu\nu}}S^{\alpha\beta ij}T^{\mu\nu hk} - \frac{\partial b_{\alpha\beta}}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}S^{\alpha\beta ij}g^{hk}$$
$$- 2\frac{\partial b_{\alpha\beta}}{\partial\phi^{\mu\nu}}S^{\alpha\beta hk}T^{\mu\nu ij} + \frac{\partial b_{\alpha\beta}}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}S^{\alpha\beta hk}g^{ij} + b_{\alpha\beta}\,g^{ih}(F^{\alpha j}{}_s\,*F^{\beta sk} - F^{\alpha k}{}_s\,*F^{\beta sj})$$
$$+ b_{\alpha\beta}\,g^{jh}(F^{\alpha i}{}_s\,*F^{\beta sk} - F^{\alpha k}{}_s\,*F^{\beta si}) + b_{\alpha\beta}\,g^{ik}(F^{\alpha j}{}_s\,*F^{\beta sh} - F^{\alpha h}{}_s\,*F^{\beta sj}) + b_{\alpha\beta}\,g^{jk}(F^{\alpha i}{}_s\,*F^{\beta sh} - F^{\alpha h}{}_s\,*F^{\beta si}) = 0. \qquad (27)$$

Multiplying (27) by g_{jk} and F^{γ}_{ih} for an arbitrary F^{γ}_{ih} we obtain

$$-8C_{\alpha\beta\mu\nu}\,\phi^{\alpha\beta\nu\mu\gamma} + 8b_{\alpha\beta\mu\nu}(-\psi^{\nu\mu\gamma\alpha\beta} + \psi^{\gamma\mu\nu\alpha\beta}) - 12b_{\alpha\beta}\,\psi^{\gamma\alpha\beta} = 0, \qquad (28)$$

where

$$C_{\alpha\beta\mu\nu} = \partial a_{\alpha\beta}/\partial\phi^{\mu\nu} - \partial a_{\mu\nu}/\partial\phi^{\alpha\beta},$$
$$b_{\alpha\beta\mu\nu} = \partial b_{\alpha\beta}/\partial\phi^{\mu\nu},$$
$$\phi^{\alpha\beta\nu\mu\gamma} = F^{\alpha i}{}_j\,F^{\beta j}{}_h\,F^{\nu h}{}_k\,F^{\mu k}{}_s\,F^{\gamma s}{}_i,$$
$$\psi^{\nu\mu\gamma\alpha\beta} = F^{\nu i}{}_j\,F^{\mu j}{}_h\,F^{\gamma h}{}_k\,F^{\alpha k}{}_s\,*F^{\beta s}{}_i,$$

and

$$\psi^{\gamma\alpha\beta} = F^{\gamma i}{}_j\,F^{\alpha j}{}_h\,*F^{\beta h}{}_i.$$

It is very easy to prove that $\psi^{\gamma\alpha\beta}$ is skew-symmetric in all of its greek indices. Also, from the definition we have

$$\psi^{\nu\mu\gamma\alpha\beta} = -\psi^{\alpha\gamma\mu\nu\beta}. \qquad (29)$$

From (21) it follows easily that

$$\psi^{\nu\mu\gamma\alpha\beta} = -\psi^{\nu\mu\gamma\beta\alpha} - \tfrac{1}{2}\phi^{\nu\mu\gamma}\psi^{\alpha\beta}, \tag{30}$$

and

$$\psi^{\nu\mu\gamma\alpha\beta} = -\psi^{\beta\mu\gamma\alpha\nu} - \tfrac{1}{2}\phi^{\mu\gamma\alpha}\psi^{\nu\beta}. \tag{31}$$

From (30) and (31) it follows that

$$\psi^{\nu\mu\gamma\alpha\beta} = -\psi^{\alpha\mu\gamma\nu\beta} - \tfrac{1}{2}\phi^{\mu\gamma\nu}\psi^{\alpha\beta} + \tfrac{1}{2}\phi^{\beta\mu\gamma}\psi^{\nu\alpha} - \tfrac{1}{2}\phi^{\mu\gamma\alpha}\psi^{\nu\beta}, \tag{32}$$

and from (7) in Ref. 4 we know that (9) in Ref. 4 is true. Written out in full and using (30) we have

$$\psi^{\gamma\alpha\beta\mu\nu} = \psi^{\beta\mu\gamma\alpha\nu} + \tfrac{1}{2}\phi^{\beta\mu\gamma}\psi^{\alpha\nu} - \tfrac{1}{8}\psi^{\beta\mu\gamma}\phi^{\alpha\nu} - \tfrac{1}{2}\psi^{\beta\mu\nu}\phi^{\gamma\alpha} - \tfrac{1}{2}\phi^{\beta\mu\alpha}\psi^{\gamma\nu}, \tag{33}$$

where $\phi^{\beta\mu\gamma} = F^{\beta i}{}_{j}\, F^{\mu j}{}_{h}\, F^{\gamma h}{}_{i}$ is skew-symmetric in all of its indices.

It is straightforward to prove that

$$\begin{aligned} {}^{*}F^{\alpha ij}F^{\beta hk} = \tfrac{1}{2}\{ & \psi^{\alpha\beta}(g^{ih}g^{jk} - g^{ik}g^{jh}) \\ & + 2F^{\alpha mk}\,{}^{*}F^{\beta j}{}_{m}\,g^{ih} - 2F^{\alpha mh}\,{}^{*}F^{\beta j}{}_{m}\,g^{ik} \\ & + 2F^{\alpha km}\,{}^{*}F^{\beta i}{}_{m}\,g^{jh} - 2F^{\alpha hm}\,{}^{*}F^{\beta i}{}_{m}\,g^{jk} \\ & + 2F^{\alpha hk}\,{}^{*}F^{\beta ij}\}. \end{aligned} \tag{34}$$

Multiplying (34) by $F^{\gamma}_{hk}\,F^{\mu l}{}_{i}\,F^{\nu}{}_{lj}$ we deduce

$$\psi^{\nu\mu\gamma\alpha\beta} = \psi^{\mu\nu\gamma\alpha\beta} + \tfrac{1}{2}\psi^{\nu\mu\alpha}\phi^{\beta\gamma} - \tfrac{1}{2}\phi^{\nu\mu\gamma}\psi^{\alpha\beta} - \tfrac{1}{2}\psi^{\nu\mu\beta}\phi^{\alpha\gamma}. \tag{35}$$

Using (31), (29), (31) again, and (35) it follows that

$$\begin{aligned} & b_{\alpha\beta\mu\nu}(-\psi^{\nu\mu\gamma\alpha\beta} + \psi^{\gamma\mu\nu\alpha\beta}) \\ & = b_{\alpha\beta\mu\nu}(\psi^{\mu\gamma\beta}\phi^{\alpha\nu} + \phi^{\gamma\mu\beta}\psi^{\alpha\nu}). \end{aligned} \tag{36}$$

Due to the skew-symmetry of $\psi^{\mu\gamma\beta}$ and $\phi^{\mu\gamma\beta}$ it follows that the only terms in (36) that could be different from zero are those that have

$$\begin{aligned} & \mu\neq\gamma\neq\beta\neq\mu, \quad \mu\neq\gamma\neq\alpha\neq\mu, \\ & \nu\neq\gamma\neq\beta\neq\nu, \quad \nu\neq\gamma\neq\alpha\neq\nu. \end{aligned} \tag{37}$$

Being $r=3$ we deduce from (37) that it must be $\alpha=\beta$, and so all terms in (36) are null because of the skew-symmetry of $b_{\alpha\beta}$. Then (28) reads

$$8C_{\alpha\beta\mu\nu}\,\phi^{\alpha\beta\nu\mu\gamma} + 12b_{\alpha\beta}\,\psi^{\gamma\alpha\beta} = 0. \tag{38}$$

Since $F^{\gamma}{}_{ih}$ was arbitrary, we can differentiate (38) with respect to $F^{\gamma}{}_{ih}$ and then multiply by ${}^{*}F^{\gamma}{}_{ih}$ to obtain

$$-8C_{\alpha\beta\mu\nu}\,\psi^{\alpha\beta\nu\mu\gamma} + 12b_{\alpha\beta}\,\phi^{\gamma\alpha\beta} = 0. \tag{39}$$

Using (33), (30), (31), (30) again, (35), and (29) it follows that

$$C_{\alpha\beta\mu\nu}\,\psi^{\alpha\beta\nu\mu\gamma} = \tfrac{1}{2}C_{\alpha\beta\mu\nu}\,\psi^{\alpha\mu\gamma}\phi^{\nu\beta} + \tfrac{1}{2}C_{\alpha\beta\mu\nu}\,\phi^{\mu\alpha\gamma}\psi^{\beta\nu}, \tag{40}$$

and so, repeating the argument that led to (37) we deduce that (39) reads

$$\begin{aligned} -8C_{\alpha\beta\mu\nu}\,\psi^{\alpha\beta\nu\mu\gamma} = & -16(C_{1122}\,\psi^{1122\gamma} \\ & + C_{1133}\,\psi^{1133\gamma} + C_{2233}\,\psi^{2233\gamma}). \end{aligned} \tag{41}$$

Similarly, using (7) of Ref. 4 and the identity $\phi^{\alpha\beta\nu\mu\gamma} = -\phi^{\mu\nu\beta\alpha\gamma}$ we obtain

$$\begin{aligned} 8C_{\alpha\beta\mu\nu}\,\phi^{\alpha\beta\nu\mu\gamma} = 16(& C_{1122}\,\phi^{1122\gamma} + C_{1133}\,\phi^{1133\gamma} \\ & + C_{2233}\,\phi^{2233\gamma}). \end{aligned} \tag{42}$$

Taking $\gamma=3$ in (38) and (39), and taking account of (41) and (42) we have the 2×2 system,

$$-16C_{1122}\,\phi^{11223} - 24b_{12}\,\psi^{123} = 0, \tag{43}$$

$$-16C_{1122}\,\psi^{11223} + 24b_{12}\,\phi^{123} = 0. \tag{44}$$

By working out an example we see that the determinant of (43) and (44) is not identically zero and so being a polynomial, it is non-null in an open dense set. Then $b_{12}=0$ everywhere. Similarly, taking $\gamma=2$, and $\gamma=1$, we obtain $b_{13}=b_{23}=0$.

Now, from (27) and from the independence of tensorial products of independent tensors, we deduce

$$a_{\alpha\beta} - 2\frac{\partial a}{\partial\phi^{\alpha\beta}} - \frac{\partial a_{\alpha\beta}}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu} = 0, \tag{45}$$

and

$$\frac{\partial a_{\alpha\beta}}{\partial\phi^{\mu\nu}} - \frac{\partial a_{\mu\nu}}{\partial\phi^{\alpha\beta}} = 0. \tag{46}$$

From (46) we deduce that there is a function $f = f(\phi^{\alpha\beta};\psi^{\alpha\beta})$ such that

$$a_{\alpha\beta} = \frac{\partial f}{\partial\phi^{\alpha\beta}}. \tag{47}$$

From (47), we can write (45) as

$$\frac{\partial}{\partial\phi^{\alpha\beta}}\left(f - 2a - \frac{\partial f}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}\right) = 0, \tag{48}$$

and then we obtain

$$a = \frac{1}{2}\left(f - \frac{\partial f}{\partial\psi^{\mu\nu}}\,\psi^{\mu\nu}\right) + h(\psi^{\alpha\beta}), \tag{49}$$

for some function h. Substituting everything in (23) we have

$$L^{ij} = L_0^{ij} + \sqrt{g}\,h(\psi^{\alpha\beta})g^{ij}, \tag{50}$$

where $L_0 = \sqrt{g}\,f$ is a scalar density.

It is important to note that within the domain of the variables $\phi^{\alpha\beta}$, $\psi^{\alpha\beta}$, it is included that $\psi^{\alpha\beta}=0$ for all α,β, even if we take account of our hypothesis that $F^1, F^2, F^3, {}^{*}F^1, {}^{*}F^2, {}^{*}F^3$ are linearly independent. In fact, it is enough to consider $(g_{ij}) = \mathrm{diag}(-1,1,1,1)$ and $F^1_{ij} = \delta^1_i\cdot\delta^2_j - \delta^2_i\cdot\delta^1_j$, $F^2_{ij} = \delta^1_i\cdot\delta^3_j - \delta^3_i\cdot\delta^1_j$, and $F^3_{ij} = \delta^1_i\cdot\delta^4_j - \delta^4_i\cdot\delta^1_j$.

Now, let k be the function

$$k(\psi^{\alpha\beta}) = h(\psi^{\alpha\beta}) - h(0) - \frac{\partial h}{\partial\psi^{\alpha\beta}}(0)\cdot\psi^{\alpha\beta}, \tag{51}$$

where h is given in (49) and (50).

Since $k(0)=0$ and $(\partial k/\partial\psi^{\alpha\beta})(0)=0$, it is known[6] that

$$k(\psi^{\alpha\beta}) = a_{\mu\nu\epsilon\theta}(\psi^{\alpha\beta})\cdot\psi^{\mu\nu}\psi^{\epsilon\theta}.$$

Then, $(1/\lambda^2)k(\lambda\psi^{\alpha\beta}) = a_{\mu\nu\epsilon\theta}(\lambda\psi^{\alpha\beta})\cdot\psi^{\mu\nu}\psi^{\epsilon\theta}$.

Let v be defined as

$$v = v(\psi^{\alpha\beta}) = -\int_0^1 \frac{1}{\lambda^2}\,k(\lambda\psi^{\alpha\beta})\,d\lambda.$$

Then a straightforward computation proves that

$$k = v - \frac{\partial v}{\partial \psi^{\alpha\beta}} \psi^{\alpha\beta},$$

and so

$$L^{ij} = L_0^{ij} + (2\sqrt{g}\, v)^{ij} + (2\sqrt{g}\, h(0))^{ij}$$
$$+ \sqrt{g}\, g^{ij} \lambda_{\alpha\beta} \psi^{\alpha\beta}$$
$$= L_1^{ij} + \sqrt{g}\, g^{ij} \lambda_{\alpha\beta}\, \psi^{\alpha\beta}$$
$$= L_1^{ij} + ((\ln g)\lambda_{\alpha\beta}\, \tilde{\psi}^{\alpha\beta})^{ij}. \qquad (52)$$

Then we have proved the following lemma.

Lemma 2: If $L^{ij} = L^{ij}(g_{hk}; F^{\alpha}_{hk})$ is a symmetric tensorial concomitant such that $L^{ij} = E^{ij}(L) = \partial L/\partial g_{ij}$ for a function $L = L(g_{hk}; F^{\alpha}_{hk})$ (not necessarily a scalar), and if the Lie group is three-dimensional (i.e., $\alpha = 1,2,3$), then L^{ij} has the form (52), where L_1 is a scalar density, $\tilde{\psi}^{\alpha\beta} = \sqrt{g}\, \psi^{\alpha\beta}$, and $\lambda_{\alpha\beta}$ are real numbers.

The result (52) is proved for $r = \dim G = 3$. Let us suppose now that $r \geqslant 3$, and let AS(4) be

$$\mathrm{AS}(4) = \{A \in R^{4\times 4}: A = -A^t\}.$$

Let g_{ij} be a symmetric matrix in $R^{4\times4}$ of signature (1,3) and let F^1, F^2, F^3 be elements of AS(4) such that $F^1, F^2, F^3, {}^*F^1, {}^*F^2, {}^*F^3$, be linearly independent. If we denote

$$A' = \{a_\alpha(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij})\, F^{\alpha}_{hk}$$
$$+ b_\alpha(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij})\, {}^*F^{\alpha}_{hk}:$$
$$1 \leqslant \alpha \leqslant 3,\ a_\alpha \text{ and } b_\alpha \text{ scalar concomitants}\},$$

it is obvious that $A' \subset \mathrm{AS}(4)$. Since $\dim \mathrm{AS}(4) = 6$ and the real numbers are examples of scalar concomitants, then it follows easily that $\mathrm{AS}(4) \subset A'$, and so $\mathrm{AS}(4) = A'$.

As a consequence, if $(F^4_{ij}) \in \mathrm{AS}(4)$ is a fixed but arbitrary skew-symmetric matrix, then there are scalar concomitants $a_\alpha(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij})$ and $b_\alpha(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij})$ $(1 \leqslant \alpha \leqslant 3)$ such that

$$F^4_{ij} = a_\alpha\, F^{\alpha}_{ij} + b_\alpha\, {}^*F^{\alpha}_{ij}. \qquad (53)$$

From (40) it follows at once that

$$\phi^{\beta 4} = a_\alpha\, \phi^{\alpha\beta} + b_\alpha\, \psi^{\alpha\beta} \quad (1 \leqslant \beta \leqslant 3), \qquad (54)$$

$$\phi^{44} = (a_\alpha a_\beta - b_\alpha b_\beta)\phi^{\alpha\beta} + 2a_\alpha b_\beta\, \psi^{\alpha\beta}, \qquad (55)$$

$$\psi^{\beta 4} = a_\alpha\, \psi^{\alpha\beta} - b_\alpha\, \phi^{\alpha\beta} \quad (1 \leqslant \beta \leqslant 3), \qquad (56)$$

and

$$\psi^{44} = (a_\alpha a_\beta - b_\alpha b_\beta)\psi^{\alpha\beta} - 2a_\alpha b_\beta\, \phi^{\alpha\beta}. \qquad (57)$$

If $a = a(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij})$ is a scalar concomitant, then, for F^4 fixed, we can write, in a dense subset of the set of concomitance variables

$$a = a_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta}) \quad (1 \leqslant \alpha, \beta \leqslant 3), \qquad (58)$$

as a consequence of (54)–(57) and Ref. 5. It is clear that

$$a_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta}) = a_{\bar{F}^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta}), \qquad (59)$$

where $\bar{F}^4$ is F^4 when computed in any other coordinate system.

The relation (59) is valid for the scalar a appearing in (23) as well as for the scalars $a_{\alpha\beta}$ and $b_{\alpha\beta}$ in (23). Then, we can repeat the proof leading to (50), and so to (52). Thus (52) is valid for all $r \geqslant 3$.

We deduce from (52) that

$$L = L_1 + (\ln g)\lambda_{\alpha\beta}\, \tilde{\psi}^{\alpha\beta} + T(F^{\alpha}_{ij})$$
$$= L_1 + (\ln g)t(\tilde{\psi}^{\alpha\beta}) + T(F^{\alpha}_{ij}) \quad (1 \leqslant \alpha, \beta \leqslant 3). \quad (60)$$

If now we vary F^4 then it follows from (60) that

$$L(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij})$$
$$= \sqrt{g}\, f_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta}) + \sqrt{g}(\ln g)t_{F^4}(\tilde{\psi}^{\alpha\beta})$$
$$+ T_{F^4}(F^{\alpha}_{ij}) \quad (1 \leqslant \alpha, \beta \leqslant 3). \qquad (61)$$

But we can choose $f_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta})$ such that

$$f_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta}) = f_{\bar{F}^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta})$$

because f is chosen to satisfy

$$\frac{\partial f}{\partial \phi^{\alpha\beta}} = a_{\alpha\beta_{F^4}} = a_{\alpha\beta_{\bar{F}^4}},$$

then $\sqrt{g}\, f_{F^4}(\phi^{\alpha\beta}; \psi^{\alpha\beta})$ is a scalar density concomitant of $g_{ij}, F^1_{ij}, F^2_{ij}, F^3_{ij}, F^4_{ij}$. Since

$$h = a - \frac{1}{2}\left(f - \frac{\partial f}{\partial \psi^{\mu\nu}}\, \psi^{\mu\nu}\right),$$

then $h_{F^4} = h_{\bar{F}^4}$ and so it defines a scalar concomitant of $g_{ij}, F^1_{ij}, F^2_{ij}, F^3_{ij}, F^4_{ij}$.

Finally,

$$t_{F^4}(\tilde{\psi}^{\alpha\beta}) = \frac{\partial h}{\partial \psi^{\alpha\beta}}(0)\cdot \tilde{\psi}^{\alpha\beta},$$

and so

$$t_{\bar{F}^4}(\bar{\tilde{\psi}}^{\alpha\beta}) = B\cdot t_{F^4}(\bar{\tilde{\psi}}^{\alpha\beta}),$$

where $B = \det(\partial x^i/\partial \bar{x}^j)$. Then it defines a scalar density $t = t(F^1_{ij}; F^2_{uj}; F^3_{ij}; F^4_{ij})$. Let us find its form.

From the invariance identities[7] we have

$$t^{ij;hk}_{\alpha\beta}(\lambda F^1_{ij}; \lambda F^2_{ij}; \lambda F^3_{ij}; \lambda F^4_{ij})$$
$$= t^{ij;hk}_{\alpha\beta}(F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij}). \qquad (62)$$

Since $\tilde{\psi}^{\alpha\beta}$ can be zero, we can take $\lim_{\lambda\to 0}$ in (62) to obtain

$$t^{ij;hk}_{\alpha\beta}(F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij})$$
$$= t^{ij;hk}_{\alpha\beta}(0;0;0;0) = l_{\alpha\beta}\epsilon^{ijhk}, \qquad (63)$$

where the result follows from Ref. 8 ($l_{\alpha\beta}$ are real numbers). Integrating (63) we obtain

$$t = l_{\alpha\beta}\tilde{\psi}_{\alpha\beta} = l_{\alpha\beta}\epsilon^{ijhk}F^{\alpha}_{ij}F^{\beta}_{hk} \quad (1 \leqslant \alpha,\beta \leqslant 4). \qquad (64)$$

From (61) and the following considerations we deduce

$$L(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij})$$
$$= L_1(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij}; F^4_{ij})$$
$$+ (\ln g)l_{\alpha\beta}\tilde{\psi}^{\alpha\beta} + T(F^{\alpha}_{ij}) \quad (1 \leqslant \alpha,\beta \leqslant 4).$$

Similarly it follows that

$$L(g_{ij}; F^{\alpha}_{ij}) = L_1(g_{ij}; F^{\alpha}_{ij}) + (\ln g)l_{\alpha\beta}\tilde{\psi}^{\alpha\beta}$$
$$+ T(F^{\alpha}_{ij}) \quad (1 \leqslant \alpha,\beta \leqslant r). \qquad (65)$$

We have proved the following theorem.

Theorem 1: If $L(g_{ij}; F^{\alpha}_{ij})$ is such that $E^{ij}(L)$ is a tensorial density, then L is given by (65), where L_1 is a scalar density and $l_{\alpha\beta}$ are real numbers.

Remark: It is not difficult to prove that $(\ln g) l_{\alpha\beta}\tilde{\psi}^{\alpha\beta}$ is not equivalent (it does not have the same Euler–Lagrange expressions E^{ij}) to any scalar density.

Let us denote $L_2 = (\ln g) l_{\alpha\beta}\tilde{\psi}^{\alpha\beta}$. Since $E^i_\alpha(L)$ is a gauge invariant tensorial density, then

$$E^i_\alpha(L) = L^{ij;hk}_{\alpha\beta} F^\beta_{hk\|j}.$$

Differentiating this last equation with respect to $A^\beta_{h,kj}$, it follows that $L^{ij;hk}_{\alpha\beta} + L^{ik;hj}_{\alpha\beta}$ is tensorial density. But from (65) we deduce

$$L^{ij;hk}_{\alpha\beta} + L^{ik;hj}_{\alpha\beta} = L_1{}^{ij;hk}_{\alpha\beta} + L_1{}^{ik;hj}_{\alpha\beta} + L_2{}^{ij;hk}_{\alpha\beta} + L_2{}^{ik;hj}_{\alpha\beta} + T^{ij;hk}_{\alpha\beta} + T^{ik;hj}_{\alpha\beta}. \quad (66)$$

Now, $L_2{}^{ij;hk}_{\alpha\beta} + L_2{}^{ik;hj}_{\alpha\beta} = 0$, as we can see from the definition of L_2. Also $L_1{}^{ij;hk}_{\alpha\beta} + L_1{}^{ik;hj}_{\alpha\beta}$ is a tensorial density because L_1 is a scalar density. Then, from (66), $T^{ij;hk}_{\alpha\beta} + T^{ik;hj}_{\alpha\beta}$ is a tensorial density.

To find the contribution of T to (66) we consider, for each α, β between 1 and r, the scalar density

$$C_{\alpha\beta} = (T^{ij;hk}_{\alpha\beta} + T^{ik;hj}_{\alpha\beta}) F^{r+1}_{ij} F^{r+2}_{hk}, \quad (67)$$

where F^{r+1} and F^{r+2} are arbitrary and independent skew-symmetric tensors. We will prove later that

$$C_{\alpha\beta} = C_{\alpha\beta\gamma\delta}\tilde{\psi}^{\gamma\delta}, \quad (68)$$

where $1 \leqslant \gamma, \delta \leqslant r+2$, and $C_{\alpha\beta\gamma\delta}$ are real numbers.

Differentiating successively (68) with respect to F^{r+1}_{ts} and F^{r+2}_{lm} we obtain

$$4T^{ts;lm}_{\alpha\beta} + T^{tm;ls}_{\alpha\beta} - T^{tl;ms}_{\alpha\beta} - T^{sm;lt}_{\alpha\beta} + T^{sl;mt}_{\alpha\beta} = C_{\alpha\beta,r+1,r+2}\epsilon^{tslm}. \quad (69)$$

Changing m with t in (69) and adding the corresponding equalities, we have

$$5T^{ts;lm}_{\alpha\beta} + 5T^{ms;lt}_{\alpha\beta} - T^{tl;ms}_{\alpha\beta} - T^{ml;ts}_{\alpha\beta} = 0. \quad (70)$$

Changing l with s in (70), and comparing (70) with the resulting identity, it follows that

$$T^{ts;lm}_{\alpha\beta} = -T^{ms;lt}_{\alpha\beta},$$

and so the contribution of T to (66) is null.

It remains to prove (68). In order to achieve this, we proceed as we did to derive (65) from (52) to write

$$C_{\alpha\beta} = \sqrt{g}\, G_{\alpha\beta}(\phi^{\mu\nu};\psi^{\mu\nu}) \quad (r \leqslant \mu,\nu \leqslant r+2). \quad (71)$$

Differentiating (71) with respect to g_{ij} we have

$$0 = \frac{\partial C_{\alpha\beta}}{\partial g} = \frac{1}{2}\sqrt{g}g^{ij}\left(G_{\alpha\beta} - \frac{\partial G_{\alpha\beta}}{\partial\psi^{\mu\nu}}\psi^{\mu\nu}\right) + \frac{\partial G_{\alpha\beta}}{\partial\phi^{\mu\nu}} T^{\mu\nu ij} \quad (r \leqslant \mu,\nu \leqslant r+2). \quad (72)$$

Since g^{ij} and $T^{\mu\nu ij}$ $(r \leqslant \mu,\nu \leqslant r+2)$ are linearly independent in a dense subset, it follows from (72) that

$$G_{\alpha\beta} = \frac{\partial G_{\alpha\beta}}{\partial\psi^{\mu\nu}}\psi^{\mu\nu}, \quad \frac{\partial G_{\alpha\beta}}{\partial\phi^{\mu\nu}} = 0,$$

and so

$$C_{\alpha\beta} = \sqrt{g}\, G_{\alpha\beta}(\psi^{\mu\nu}), \quad (73)$$

and

$$G_{\alpha\beta}(\lambda\psi^{\mu\nu}) = \lambda G_{\alpha\beta}(\psi^{\mu\nu}). \quad (74)$$

Differentiating (74) with respect to λ and making $\lambda = 0$ it follows that

$$G_{\alpha\beta}(\psi^{\mu\nu}) = \frac{\partial G_{\alpha\beta}}{\partial\psi^{\mu\nu}}(0)\cdot\psi^{\mu\nu} = C_{\alpha\beta\mu\nu}\psi^{\mu\nu},$$

which, together with (73), proves (68).

From the identity following (70) we deduce easily that $T^{ij;hk}_{\alpha\beta}$ is skew-symmetric in all of its latin indices. Then

$$T^{ij;hk}_{\alpha\beta} = d_{\alpha\beta}\epsilon^{ijhk}. \quad (75)$$

Differentiating (75) with respect to F^γ_{rs} we have

$$T^{ij;hk;rs}_{\alpha\beta\ \gamma} = \frac{\partial d_{\alpha\beta}}{\partial F^\gamma_{rs}}\epsilon^{ijhk}. \quad (76)$$

But the left-hand side of (76) is skew-symmetric in all of its latin indices. Since we are working in a four-dimensional space, then it is null, and so the $d_{\alpha\beta}$ in (75) are real numbers.

Integrating (75) gives

$$T = d_{\alpha\beta}\tilde{\psi}^{\alpha\beta} + K^{ij}_\alpha F^\alpha_{ij} + K. \quad (77)$$

Let $\hat{K} = K^{ij}_\alpha F^\alpha_{ij} + K$. We know that $E^i_\alpha(L)$ is gauge invariant. Using the replacement theorem of Horndeski[9] and taking account that $E^i_\alpha(K^{hj}_\beta F^\beta_{hj}) = 2K^{ij}_\beta C^\beta_{\theta\alpha}A^\theta_j$, we have

$$E^i_\alpha(L)(g_{hk};0;-\tfrac{1}{2}F^\gamma_{hk}) = E^i_\alpha(\hat{L})(g_{hk};0;-\tfrac{1}{2}F^\gamma_{hk}) + E^i_\alpha(L_2)(g_{hk};0;-\tfrac{1}{2}F^\gamma_{hk}) \quad (78)$$

and so $E^i_\alpha(L_2)$ is a tensorial density. We will prove that it is zero. In fact

$$E^i_\gamma(L_2) = E^i_\gamma((\ln g) l_{\alpha\beta}\tilde{\psi}^{\alpha\beta}) = (\ln g) E^i_\gamma(l_{\alpha\beta}\tilde{\psi}^{\alpha\beta}) + l_{\alpha\beta}\frac{\partial\tilde{\psi}^{\alpha\beta}}{\partial A^\gamma_{i,j}} g^{hk}g_{hk,j}. \quad (79)$$

By making a change of coordinates such that $(g_{ij}) = \mathrm{diag}(-1,1,1,1)$, $g_{ij,h} = 0$, and $\det(\partial x^i/\partial\bar{x}^j) = 1$, it follows from (79) that $E^i_\alpha(L_2) = 0$. Since $E^i_\alpha(L_2)$ is a scalar density, then it is null in every coordinate system. Now, taking a coordinate system where $g = 1$ and $g^{hk}g_{hk,j} \neq 0$, it follows from (79) and the vanishing of $E^i_\alpha(L_2)$ that

$$l_{\alpha\beta}\frac{\partial\tilde{\psi}^{\alpha\beta}}{\partial A^\gamma_{i,j}} = 0. \quad (80)$$

Differentiating (80) with respect to $A^\theta_{h,k}$ we have

$$l_{\alpha\beta}\epsilon^{ijhk} = 0,$$

from where it follows $l_{\alpha\beta} = 0$. Then

$$L_2 = 0. \quad (81)$$

Using (64) we deduce that $\hat{L}$ is a scalar density defined everywhere (the tensorial character follows from the continuity of the invariance identities). Besides

$$E^i_\alpha(K^{hj}_\beta F^\beta_{hj}) = 2K^{ij}_\beta C^\beta_{\theta\alpha}A^\theta_j$$

is a tensorial density. It follows easily[8] that it is zero. We have proved then

$$L = \hat{L} + \hat{K}, \quad (82)$$

where $E^i(\hat{K}) = 0$, $E^{ij}(\hat{K}) = 0$. Then it follows

$$E^{ij}(L) = E^{ij}(\hat{L}), \quad E^i(L) = E^i(\hat{L}). \tag{83}$$

But, from (83), $\hat{L}$ is a scalar density whose Euler–Lagrange expressions are gauge invariant. We deduce[10] that $\hat{L}$ can be replaced by a gauge invariant scalar density without modifying the Euler–Lagrange expressions. This solves in affirmative the equivariant inverse problem for L, and so we have the following theorem.

Theorem 2: If $L(g_{ij};F^{\alpha}_{ij})$ is such that $E^{ij}(L)$ and $E^i_{\alpha}(L)$ are gauge invariant tensorial densities, then there is a gauge invariant scalar density $\hat{L}$ such that

$$E^{ij}(L) = E^{ij}(\hat{L}), \quad E^i_{\alpha}(L) = E^i_{\alpha}(\hat{L}).$$

III. THE UNIQUENESS OF THE YANG–MILLS EQUATIONS

Let $\hat{L}$ be defined as in Theorem 2. By similarity with (7) and (14), we could claim the field equations to be

$$\hat{L}^{ij}_{\alpha\|j} = J^i_{\alpha}, \tag{84}$$

$$*\hat{L}^{ij}_{\alpha\|j} = 0. \tag{85}$$

Now, Eqs. (85) corresponding to the Yang–Mills internal equations do not depend on the charge and current distribution, and so they should be satisfied identically.

Let $H^{ij}_{\alpha} = *\hat{L}^{ij}_{\alpha}$, so that $H^{ij}_{\alpha\|j} = 0$.

Since H^{ij}_{α} is gauge invariant, this can be written as

$$\frac{\partial H^{ij}_{\alpha}}{\partial F^{\beta}_{hk}} F^{\beta}_{hk\|j} = 0. \tag{86}$$

Differentiating (86) with respect to $A^{\beta}_{k,hj}$ we have

$$H^{ij;hk}_{\alpha\beta} + H^{ih;jk}_{\alpha\ \beta} = 0, \tag{87}$$

so that it follows easily that $H^{ij;hk}_{\alpha\beta}$ is skew-symmetric in all of its latin indices, and so

$$H^{ij;hk}_{\alpha\beta} = \mu_{\alpha\beta}\eta^{ijhk}. \tag{88}$$

Differentiating (88) with respect to F^{γ}_{rs} we have

$$H^{ij;hk;rs}_{\alpha\beta\ \gamma} = \frac{\partial\mu_{\alpha\beta}}{\partial F^{\gamma}_{rs}}\eta^{ijhk}. \tag{89}$$

But

$$H^{ij;hk;rs}_{\alpha\beta\ \gamma} = H^{ij;rs;hk}_{\alpha\beta\gamma} = -H^{ir;js;hk}_{\alpha\ \beta\gamma},$$

and so $H^{ij;hk;rs}_{\alpha\beta\ \gamma}$ is skew-symmetric in all of its latin indices. Since we are working in a four-dimensional space, then $H^{ij;hk;rs}_{\alpha\beta\ \gamma} = 0$ and so $\partial\mu_{\alpha\beta}/\partial F^{\gamma}_{rs} = 0$, which means that $\mu_{\alpha\beta}$ is a scalar concomitant of g_{ij} alone. Then[11] $\mu_{\alpha\beta}/\sqrt{g}$ is a constant for each α,β. From (88) we have, integrating

$$H^{ij}_{\alpha} = \mu_{\alpha\beta} *F^{\beta ij} + \lambda^{ij}_{\alpha}(g_{ij}).$$

But $\lambda^{ij}_{\alpha} = 0$ from Ref. 11, so that $H^{ij}_{\alpha} = \mu_{\alpha\beta} *F^{\beta ij}$, and so

$$*\hat{L}^{ij}_{\alpha} = \mu_{\alpha\beta} *F^{\beta ij}. \tag{90}$$

Multiplying (90) by η_{ijhk} and then multiplying by $g^{hr}g^{ks}$ it follows that

$$\hat{L}^{rs}_{\alpha} = \mu_{\alpha\beta}F^{\beta rs},$$

and so, integrating

$$\hat{L} = \mu_{\alpha\beta}\phi^{\alpha\beta} + C\sqrt{g},$$

in which case $\hat{L}^{ij}_{\alpha\|j} = \mu_{\alpha\beta}F^{\beta ij}{}_{\|j}$, which means that $E^i_{\alpha}(\hat{L}) = J^i_{\alpha}$ are the usual Yang–Mills equations.

ACKNOWLEDGMENT

We gratefully acknowledge the referee for very helpful suggestions.

[1]I. M. Anderson, Ann. Math. **120**, 329 (1984).

[2]We denote $*F^{\alpha ij} = \eta^{ijhk}F^{\alpha}_{hk}$, where $\eta^{ijhk} = (\sqrt{g})^{-1}\epsilon^{ijhk}$, where ϵ^{ijhk} are the Levi-Civita permutation symbols. In Eq. (7), J^i_{α} is the charge-current vector.

[3]G. W. Horndeski, Arch. Rat. Mech. Anal. **75**, 229 (1981).

[4]R. J. Noriega and C. G. Schifini, Gen. Relativ. Gravit. **20**, 337 (1988).

[5]R. J. Noriega and C. G. Schifini, J. Math. Phys. **30**, 617 (1989).

[6]M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965).

[7]H. Rund, Abh. Math. Sem. Univ. Hamburg **29**, 243 (1966).

[8]R. J. Noriega, Rev. Un. Mat. Argentina **31**, 149 (1984).

[9]G. W. Horndeski, Utilitas Math. **19**, 215 (1981).

[10]M. C. Calvo, M. C. López, R. J. Noriega, and C. G. Schifini, "Gauge invariance of Euler–Lagrange expressions in Einstein–Yang–Mills field theories," submitted to Gen. Relativ. Gravit.

[11]D. Lovelock, Arch. Rat. Mech. Anal. **33**, 45 (1969).

Tadpole graph in covariant closed string field theory

Guillermo Zemba and Barton Zwiebach
Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 7 February 1989; accepted for publication 19 April 1989)

A one-loop tadpole graph in covariant closed string field theory must generate all inequivalent tori with one puncture. In order to determine the region of integration in moduli space we map the tadpole to a two-sheeted sphere and use the analyticity of the modular parameter with respect to the complex propagator parameter. It is shown that the naive closed string extension of the Witten open string vertex fails to give the correct modular region for this one-loop amplitude.

I. INTRODUCTION AND SUMMARY

In light cone field theory closed strings interact via a three-point vertex that is the naive extension of the open string three-point vertex.[1] In covariant field theory open strings interact via a three-point vertex suggested by Witten.[2] Its naive extension, however, did not appear to be a compelling candidate for a closed string vertex.[2] Nevertheless it has been studied somewhat.[3] It was realized that it would not generate completely the modular region for the four-point scattering amplitude[4] and therefore it would have to be supplemented by an elementary four-point closed string interaction. There are even concrete proposals for such an interaction.[5]

There is, however, a very simple test that a covariant closed string vertex must pass. When two of the legs of the vertex are contracted with a propagator to give a one-loop tadpole one must generate all inequivalent tori with one puncture. This criterion was advocated in Ref. 6 as the basis for a search for a satisfactory closed string vertex and it was shown that it rules out all vertices based on $SL(2,C)$ functions (projective type vertices). For the case of the naive extension of Witten's vertex, which is not projective, qualitative arguments were put forth that it would give the wrong answer for the one-loop tadpole.[7]

In this paper we discuss the evaluation of one-loop tadpoles. Our motivation for studying this problem in detail is based on the following. The tadpole graph cannot be treated with light-cone type methods because it corresponds to a surface with one puncture, while light-cone graphs correspond to surfaces with at least two punctures. Therefore this is a somewhat novel type of problem, which we have found can be studied quite effectively using the automorphic function of the level-two congruence subgroup $\Gamma(2)$ of the modular group Γ. Given that projective vertices are ruled out[6] as candidates for closed string vertices, we wished to learn how to deal with nonprojective vertices and this vertex was a prime candidate. As in Ref. 6, we concentrate on finding the region of integration for the modular parameter τ describing the torus. The modular parameter τ is a function of the complex parameter $T = t + i\theta$ (t is the length and θ the twist angle) describing the closed string propagator and knowledge of this function is sufficient to determine if one is getting the correct modular region. Finally, we wished to give quantitative confirmation of the argument of Ref. 7. Indeed, the naive closed string vertex fails badly in giving the correct modular region for the one-loop tadpole and it does not seem to have any special modular properties. This implies that used in conjunction with the usual closed string propagator, for which $0 \leqslant t < \infty$ and $-\pi \leqslant \theta < \pi$, it cannot lead to a completely satisfactory field theory. Perhaps it is possible to use our methods and result to understand possible modifications for the closed string propagator.[8]

In Sec. II we first discuss issues related to analytic behavior. In a one-loop tadpole, two legs of a given three-point vertex are joined by a propagator specified by the length parameter t and the twist angle θ. The modular parameter τ of this torus, for a fixed three-point vertex, can only depend on t and θ. We argue that τ is an analytic function of the complex parameter $T = t + i\theta$. This fact simplifies enormously the work of finding the region of integration in moduli space, since it is sufficient to work with rectangular tori ($\theta = 0$) to find $\tau(t)$ and extend this analytically letting $t \to T$. Such type of arguments could not be made in light-cone diagrams because the propagator parameters are not independent. Namely, between two interaction points, the length of all tubes must be equal, so that this constraint requires the nonanalytic equation $\mathrm{Re}\, T_1 = \mathrm{Re}\, T_2 = \cdots$. In covariant field theory there are no such constraints and one has analytic dependence of the modular parameters on the propagator parameters, which play the role of complex coordinates for moduli space.

We then turn to review the representation of a torus as two sheets joined across two branch cuts, one extending from 0 to 1 and the other from a point x up to ∞. This standard mathematical presentation has been used in recent string theory works.[9] Any square torus ($-i\tau$ real and positive) can be mapped by a Schwarz–Christoffel transformation into two sheets with cuts running from 0 to 1, and from x, real and greater than one, to ∞. We verify that for a general torus the same type of Schwarz–Christoffel map works, but x becomes complex. The only complication that arises is that if one wished to present the fundamental region of the torus as a parallelogram, the cuts between 0 and 1 and x to ∞ would be curved. This is an illustration of the idea, which will be used in Sec. III, that general tori can be dealt with using mapping functions that are an analytic continuation of the ones used for rectangular tori.

In the two-sheeted construction x is the modular parameter. For the general case of a torus with a modular param-

 0022-2488/89/102388-09$02.50

eter τ, x depends on τ, in fact $x(\tau) = \lambda^{-1}(\tau)$, where $\lambda(\tau)$ is Picard's automorphic function of the modular subgroup $\Gamma(2)$.[10,11] We found the two-sheeted presentation of the torus especially useful, since it was possible to cut the tadpole graph in two pieces and relate via a Schwarz–Christoffel transformation each piece to one sheet. From the relation between x and τ one can see what is the region in the x plane that corresponds to the fundamental region of the modular group (Fig. 4). Special limits for $\tau(x)$, necessary for Sec. III, are given for the cases $x \to 1$ and $x \to \infty$.

In Sec. III we concentrate on the naive closed string vertex and give the Schwarz–Christoffel map of the tadpole (ρ plane) into the two-sheeted sphere presentation (z plane). The map is interesting in that $\rho = \int^z \phi$, where ϕ is not a meromorphic differential on the torus. It could not have been a meromorphic differential, since given that it should map into a tadpole, which has one external leg, it would have to have a single first-order pole with a nonvanishing residue corresponding to the length parameter. But this is impossible for meromorphic differentials, since the sum of their residues has to equal zero.[12] The differential we have is a multivalued one, it has two zeroes of order $+\frac{1}{2}$ and a simple pole (like a square root of a Weierstrass function). We argue that it can be used consistently to perform the conformal mapping.

Our aim is to find $x(T)$. While, in general, this is very hard to do explicitly, it can be done both for $T \to 0$, which corresponds to $x \to 1$, and $T \to \infty$, which corresponds to $x \to \infty$. Comparing with the limits given in Sec. II for $\tau(x)$, one is able to relate T and τ. We find that for $T \to 0$ one has that $\tau \sim T$. This implies that the neighborhood of $\tau = 0$ is covered (see Fig. 10) and since this region contains an infinite number of copies of the modular region we find that the naive closed string vertex fails to give the correct region of moduli space.

In Sec. IV we give some conclusions and discuss how our arguments can be applied to simplify the discussion of four-point closed string scattering.

II. ANALYTICITY AND TORI AS TWO-SHEETED SPHERES

A. Analytic behavior of the modular parameter

We first wish to argue that the parameters that define the propagator, namely, t and θ form a natural complex parameter $T = t + i\theta$ and that the modular parameter, which in our case of interest is τ, is an analytic function of T, namely, we have $\tau(T)$. A similar situation should hold for higher genus diagrams in covariant closed string field theory, each propagator parameter is an unconstrained complex variable, and the modular parameters must be analytic functions of those parameters.

The process of evaluating a tadpole graph can be set in a formal way,[6] following the formalism of Ref. 13. As discussed in Sec. V of Ref. 6 for any closed string vertex, one has two local coordinates, z_1 and z_2, corresponding to the two legs of the tadpole that are to be contracted, and two functions $h_1(z_1)$ and $h_2(z_2)$ that define mappings from the local coordinates to the z plane, where the strings interact. Figure 1 shows the local coordinates and the images of the unit

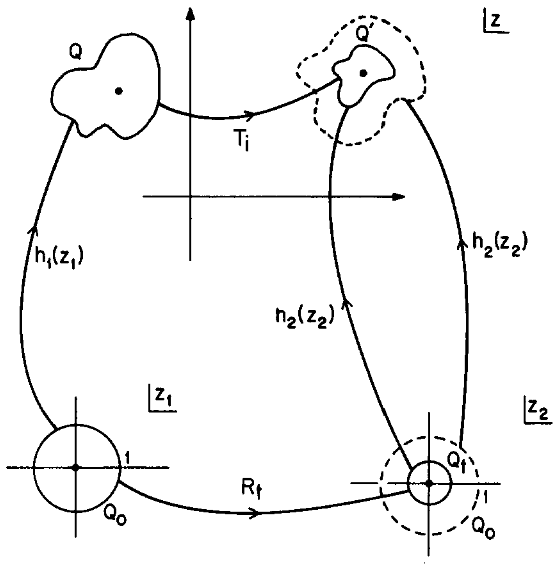


FIG. 1. Formulation of the general problem of evaluating a tadpole graph. The process of identification between Q and Q' is indicated. The z_1 and z_2 variables are the local coordinates.

circles Q_0 under the functions h. A particular torus is built gluing Q with Q', which is the image of the circle Q_t in the plane z_2. The circle Q_t itself is obtained from Q_0, via the transformation $R_t(z) = t/z$. Here t is a complex parameter. It seems fairly clear that the holomorphic differential on the surface will depend on t but not on $\bar{t}$, which does not appear anywhere. [A mathematical proof of this would probably use the analytic dependence on t of the identifying transformation $T_i = h_2 \circ R_t \circ (h_1)^{-1}$ in the z plane.] It then follows that the modular parameter will only depend on t. But t is nothing else than the complex propagator parameter in disguise, since the annulus between Q_0 and Q_t in the z_2 plane corresponds to the tube that joins the two legs of the vertex. The mapping $\ln z_2$ takes the annulus into a cylinder of parameter $T = \text{length} + i\,\text{twist} = -\ln t = -\ln|t| - i\arg(t)$. Since the modular parameter depends only on t, it will depend only on T and not on $\bar{T}$.

B. Tori as two-sheeted spheres

The mapping that takes a torus into two sheets glued across two cuts is just a Schwarz–Christoffel map. Since this standard presentation for tori is one of the main tools for our work we discuss it next, and illustrate how for tori that involve a twist angle, the same mapping function works if one lets some parameters become complex. The mapping function is nothing else than the integral of the holomorphic differential on the torus. Consider the rectangular torus indicated in Fig. 2. The short sides of length π are glued to each other, and so are the sides of length T. The torus is cut along the line AA' into two pieces, which can be glued together following the pattern indicated by the signs to again form the torus [Fig. 2(b)]. The left piece is mapped into the z plane via

$$\rho(z) = N \int_0^z \frac{dz'}{\sqrt{z'(z'-1)(z'-x)}}. \qquad (2.1)$$

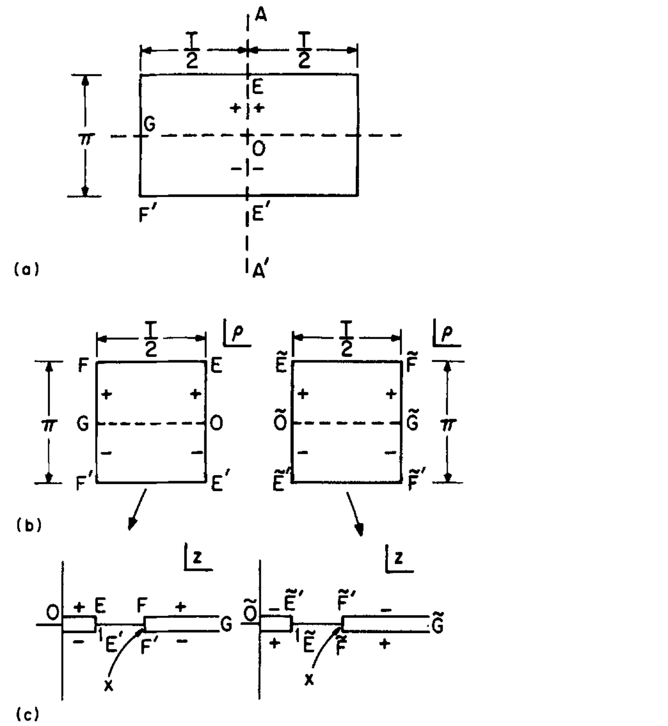


FIG. 2. Construction of a square torus (a) as a two-sheeted sphere (c). Here x is the modular parameter.

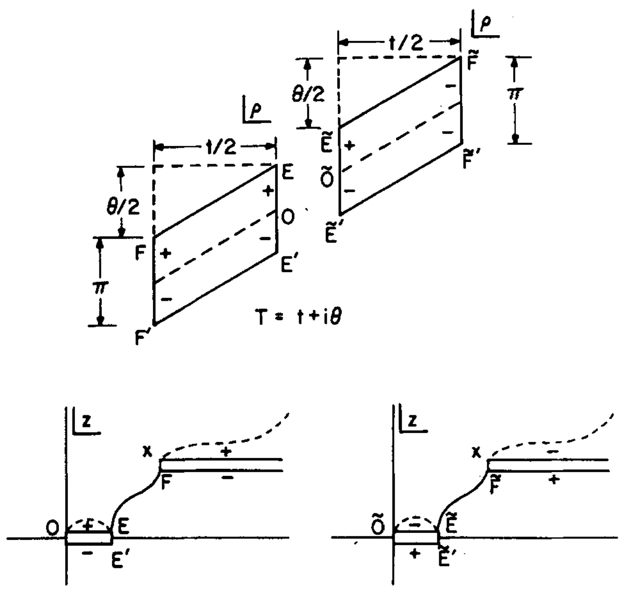


FIG. 3. Construction of a general torus using two sheets with cuts.

Note that the sides that have to be glued to the other half go into the slits from 0 to 1, and from a parameter x to ∞. The horizontal sides FE and $F'E'$, which are identified in Fig. 2(b), are effectively glued in Fig. 2(c). The powers of $(-\frac{1}{2})$ in the above equation are there because around $z_i = 0,1,x$ one must have $\rho \approx (z - z_i)^{1/2}$. The normalization constant N is fixed requiring $\rho(E) = i\pi/2$, and T depends only on x via the relation $T/2 = \rho(E) - \rho(F)$. The other side of the diagram will be mapped similarly, as shown in the figure, but note that the pattern of signs is automatically inverted as it should, in order to be possible to glue the two surfaces.[12] This is so because the diagram to the right of Fig. 2(b) has to be rotated by an angle of 180° around $\tilde{O}$ in order to get a configuration equivalent to the one on the left and therefore be able to use the same mapping function for both.

In order to work with tori that have a twist angle one just lets T and x become complex, there is no need to change the mapping function. We illustrate this in Fig. 3. There a complex torus with modular parameter $\tau = T/2\pi = (t + i\theta)/2\pi$ is cut again in half. One may think that the Schwarz–Christoffel map that takes one of the pieces into the z plane may even need irrational powers to smooth the corners for arbitrary θ, but this is wrong. The fact is that in the ρ plane the total internal angle at point E, which is identified with E', is still π and therefore comparing with the situation in the z plane where the angle is 2π one sees that the power of $(-\frac{1}{2})$ is still needed. The same holds for point F. Thus the mapping function is the same, the normalization condition is the same, and the only thing that changes is the fact that the relation between T and x (which is the same) requires x complex, given that T is complex. It should also be noted that the mapping function would not take the straight slits indicated in Fig. 3 into the straight lines joining E and E' or F and F'. It would produce an equivalent copy of the half torus with curved but identical lines joining E and E', and F and F'. This is perfectly acceptable. Equally acceptable curved slits shown in dotted lines in the figure could be found that would give straight lines in the ρ plane. It is interesting to realize how the identification procedure between the slits works in both cases. For the case displayed in Fig. 3, in which we have curved slits in the z plane, the argument is the same as in the case of the square torus. For the case in which we have straight slits in the z plane, we notice that $\rho(z)$ follows a curve that is symmetric around O between points E and E' when z goes around the slit and therefore the previous argument can also be applied to this case.

Any torus can therefore be mapped conformally into two sheets (Riemann spheres) glued across slits cut from 0 to 1 and from a given point x to ∞. In this rather convenient presentation for tori the modular parameter τ for the torus can be calculated as

$$\tau = \frac{\omega_2}{\omega_1}, \quad \omega_2 = \int_1^x \frac{dz}{y}, \qquad \omega_1 = \int_0^1 \frac{dz}{y}, \quad y^2 = z(z-1)(z-x). \tag{2.2}$$

Here dz/y is a holomorphic differential in the two-sheeted surface [in fact, the integrand in (2.1)] and the ω's are the integrals of this differential over the cycles of the torus. These integrals are just complete elliptic integrals and τ, which just depends on x, can therefore be written as

$$\tau(x) = i[K(k')/K(k)], \tag{2.3}$$

where $K(k)$ is the complete elliptic integral of the first kind (see the Appendix for notation), $k^2 = x^{-1}$, and $k' = \sqrt{1-k^2}$ is the complementary modulus.

Two simple observations can be made: (i) in the above construction four points were singled out, namely, 0, 1, x, and ∞; the branches can run between any pair of points and one still gets conformally the same torus; (ii) for a given value of x there are five other values of x that represent the same torus, the six values of x are obtained when one considers the six $SL(2,C)$ transformations that generate the permutations of the points 0, 1, and ∞. These six transformations form a finite group, the Λ group or group of anharmonic ratios[10] relevant to the problem of the four-point scattering in closed string field theory.[14]

The function $x(\tau)$ is in fact just $1/\lambda(\tau)$, where λ is Picard's function. Under modular transformations one has

$$x(\tau+1) = 1 - x(\tau), \quad x(-1/\tau) = x(\tau)/[x(\tau)-1]. \tag{2.4}$$

The function $x(\tau)$ is invariant or automorphic under some important transformations of the modular group. One has that

$$x[(a\tau+2b)/(2c\tau+d)] = x(\tau), \quad ad-4bc=1, \tag{2.5}$$

where a, b, c, and d are integers. These transformations form a subgroup of the modular group denoted $\Gamma(2)$, where $\Gamma(N)$ denotes the level N principal congruence subgroup of the modular group Γ:

$$\Gamma(N) \equiv \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \,|\, a,d \equiv 1 \bmod N;\ b,c \equiv 0 \bmod N \right\}. \tag{2.6}$$

Note that $\Gamma(N)$ is, in fact, a normal subgroup of Γ. This implies that the quotient $\Gamma/\Gamma(2)$ is itself a group. This group is just the Λ group, of six elements. Therefore $\Gamma(2)$ has index six in Λ and consequently its fundamental region $\mathscr{F}_2$ [Fig. 4(a)] contains six copies of the fundamental region of the modular group $\mathscr{F}_0$. Here $\mathscr{F}_2$ is the region of the upper τ plane bounded by the vertical lines $\tau = \pm 1$ and by two circles of radius $\frac{1}{2}$ centered at $\pm\frac{1}{2}$. The subgroup $\Gamma(2)$ is generated by the transformations

$$\tau \to \tau+2, \quad \tau \to \tau/(1-2\tau), \tag{2.7}$$

and the effect of the generating transformations on the boundaries of $\mathscr{F}_2$ is indicated in Fig. 4(a). The function $\lambda(\tau)$ can be expressed in terms of theta functions

$$\lambda(\tau) \equiv 1/x(\tau) = \theta_1^4(0|\tau)/\theta_3^4(0|\tau). \tag{2.8}$$

Most important for our purposes is that $\lambda(\tau)$ [and as a consequence $x(\tau)$] maps $\mathscr{F}_2$ onto the Riemann sphere and so is the $N=2$ analog of the modular invariant $J(\tau)$. In Fig. 4(b) we indicate how each of the six copies of $\mathscr{F}_0$ that make $\mathscr{F}_2$ is mapped under $x(\tau)$.

For the study of the tadpole diagram we need to know the behavior of $x(\tau)$ for special situations. It is possible to obtain asymptotic expressions for τ for some regions of x. First consider the case when $x \to 1$. Inspection of Fig. 4(b) shows that $\tau \to 0$. In this case $k \to 1$ and it is possible to expand both integrals using Eqs. (A3) and (A4). The result for τ is

$$\tau(x) = -\frac{i\pi}{\ln[(x-1)/16]} + O\left(\frac{x-1}{\ln(x-1)}\right) \quad (x \to 1). \tag{2.9}$$

Note that the above function indeed maps a neighborhood of $x=1$ into the zero angle wedge at $\tau = 0$, as shown in Figs. 4(a) and 4(b).

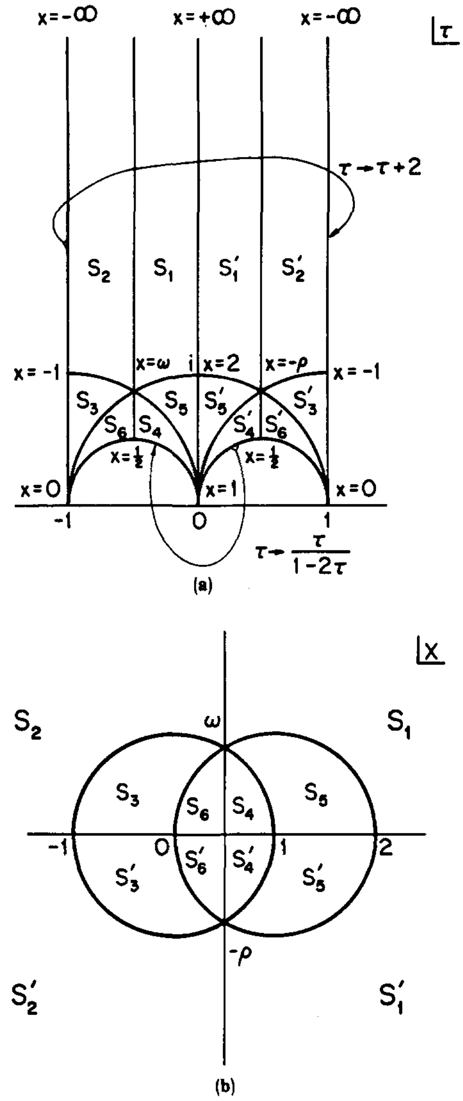


FIG. 4. (a) Fundamental region $\mathscr{F}_2$ of $\Gamma(2)$ represented in the τ plane showing the identifying transformations. (b) Image of $\mathscr{F}_2$ under $x(\tau) = \lambda^{-1}(\tau)$. Here $\omega = e^{i(\pi/3)}$, and $\rho = e^{i(2\pi/3)}$.

The other limit of interest arises when $x \to \infty$. In this case $\tau \to \infty$ and one finds the following expression:

$$\tau(x) = (i/\pi)[\ln(16x) - 1/2x + O(x^{-2}\ln x)] \quad (x \to \infty). \tag{2.10}$$

III. THE CONFORMAL MAP

The aim of this section is to provide the tools for the determination of the region in moduli space covered by the one-loop tadpole diagram using the bilocal, naive extension to closed strings of the vertex that Witten proposed for open string field theory [Fig. 5(a)]. In this vertex, three cylinders representing free propagating closed strings join in a symmetrical way. The total angle at each interacting point is 3π. The one-loop tadpole graph we wish to study is shown in Fig. 5(b). There is one incoming string (C) that splits into two strings at the interaction points B,D and a closed string propagator pairing these two strings.

To find the region of integration in moduli space, we

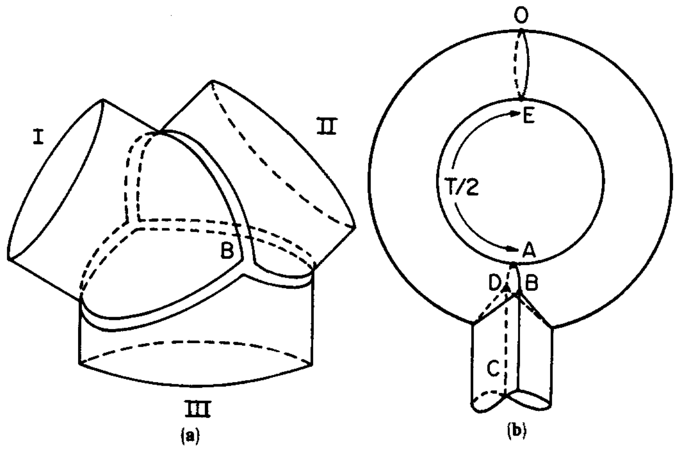


FIG. 5. (a) Diagram representing the naive extension of Witten's vertex to closed string field theory. Here I, II, and III represent the three interacting closed strings. Note that B is one of the two interaction points. The total angle at any interaction point is 3π. (b) The one-loop tadpole graph. The intermediate loop parameter is called T. The marked points are in correspondence with the ones in Figs. 6–9. The diagram is cut along the $CBADC$ and EO lines.

perform a conformal mapping from the Riemann surface for the tadpole [Fig. 5(b)], which is conformally equivalent to a torus with one puncture and to a double-sheeted complex plane with two cuts. The conformal mapping is carried out in three stages. First, we cut the original diagram along a line that contains the two interaction points of the vertex and slices the external propagator along a diameter [line $CBADC$ in Fig. 5(b)]. We also cut along the curve EO in the loop portion of the diagram and we end up with two pieces. For the moment, we assume there is no twist in the propagator. Each one of the two components of the diagram can be represented as the portion of the complex plane shown in Fig. 6. Segments EA and $E'A'$ are identified in this construction. The other segments of the boundary are glued with the corresponding ones of the symmetrical piece representing the other half of the diagram. The upper half w plane can be mapped to this region by a Schwarz–Christoffel transformation

$$\rho(w) = 2N \int_0^w dw \frac{(w^2-\beta^2)^{1/2}}{(w^2-1)^{1/2}(w^2-\alpha^2)^{1/2}}. \quad (3.1)$$

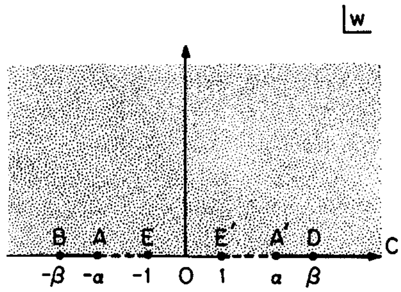


FIG. 7. The region of the complex w plane that is mapped by (3.1) to the region of Fig. 6. Because of the symmetry and overall normalization freedom it depends on the position of two points α and β, as shown.

Here N is a normalization factor. The mirror symmetry of the diagram around the real axis in the ρ plane (Fig. 6) is realized as a reflection symmetry around the imaginary axis in the upper w plane (Fig. 7). By virtue of this symmetry, the origin in the w plane is mapped onto the origin in the ρ plane and the mapping depends only on two real numbers, α and β. Points ± 1 (E',E) are mapped onto points $\rho = \mp i(\pi/2)$. Points $\pm\alpha$ (A',A) are mapped onto points $\rho = \mp i(\pi/2) + t/2$, where t is the Schwinger parameter of the loop propagator. Since the internal angles at E, E', A, and A' are $\pi/2$, the exponent for the factors $(w \pm 1)$ and $(w \pm \alpha)$ is $(-\frac{1}{2})$. Points $\pm\beta$ (D,B) are mapped onto points $\mp i(\pi/4) + t/2$. The internal angles at B and D are $3\pi/2$ and therefore the exponent for the factors $(w \pm \beta)$ is $(+\frac{1}{2})$. The external state (C) is mapped to infinity.

In the next stage, we map the upper half w plane to the whole complex z plane using the transformation $z = w^2$. It is useful to define $x = \alpha^2$ and $y = \beta^2$. After performing the same set of operations for each half of the original diagram we end up with two sheets, each one with two cuts on it: one between 0 and 1 and the other from a point x up to infinity (Fig. 8). From this picture, it is clear that y plays the role of an auxiliary variable and that x is the modular parameter. The segments EA and $E'A'$ that had to be identified in the w plane are effectively glued in the z plane across the slit that goes from 1 to x. In the final stage, opposite sides of the cuts of the two sheets are identified in order to recover the original Riemann surface. The final expression for the mapping is

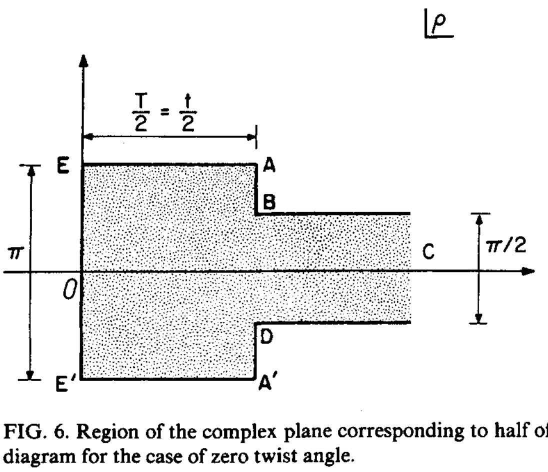


FIG. 6. Region of the complex plane corresponding to half of the tadpole diagram for the case of zero twist angle.

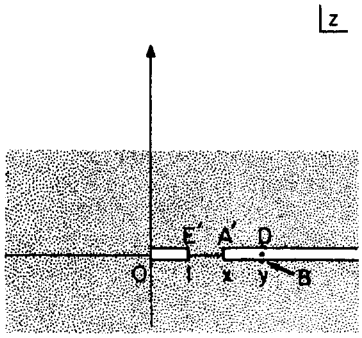


FIG. 8. The region of the complex z plane that is mapped by (3.2) to the region of Fig. 6. Here $x = \alpha^2$ and $y = \beta^2$ are real numbers for the case of zero twist angle.

then given by

$$\rho(z) = N \int_0^z dz' \left(\frac{(z'-y)}{z'(z'-1)(z'-x)} \right)^{1/2}. \qquad (3.2)$$

As we already pointed out in the Introduction, the differential in (3.2) is not meromorphic. To evaluate the integral we have to specify the sign assignments given to the object that is being integrated in each sheet. On the first sheet (Fig. 8), we define the integrand of (3.2) to have opposite signs across a slit that goes from 0 to 1 and across a slit from x to y. For points beyond y we assign the same sign to points immediately above and below the cut in the surface. Signs are completely reversed on the second sheet. It is possible to check that with this definition the object that is being integrated in (3.2) is well defined over the double-sheeted plane.

As it is customary, we normalize all closed string widths to π. The conditions that we will impose on this map are the following: (a) let the strip width between E and E' be π,

$$\rho(E) - \rho(E') = i\pi; \qquad (3.3)$$

(b) let the length of segment AE be equal to the length of segment $A'E'$ and both should equal $t/2$, where t is the internal propagator parameter,

$$\rho(A') - \rho(E') = \rho(A) - \rho(E) = t/2; \qquad (3.4)$$

and (c) let the strip width between B and D be $\pi/2$,

$$\rho(B) - \rho(D) = i\pi/2. \qquad (3.5)$$

The overall normalization constant N can be fixed by considering the limit $z \to \infty$. In this case the integral becomes a logarithmic one and application of condition (c) gives $N = \frac{1}{4}$. Conditions (a) and (b) leave us with two elliptic integrals:

$$2\pi = \int_0^1 dz \left(\frac{(y-z)}{z(1-z)(x-z)} \right)^{1/2}, \qquad (3.6)$$

$$t = \frac{1}{2} \int_1^x dz \left(\frac{(y-z)}{z(z-1)(x-z)} \right)^{1/2}. \qquad (3.7)$$

The first one determines y as a function of x and the second one gives the intermediate time t as a function of x and y. Using both conditions we can find t as a function of x alone. For the case in which the intermediate loop propagator contains a twist we consider it as being performed before gluing the two matching circumferences of the loop. Half of the total amount of the twist is assigned to each component piece. We may still use the previous technique, provided we extend t to a complex parameter $T = t + i\theta$, in which t is interpreted as before and θ is the twist angle that runs between $-\pi/2$ and $\pi/2$ (Fig. 9). We rely on the analyticity of (3.6) and (3.7) to conclude that these expressions can be extended to the case of complex T, provided x and y become complex variables. For later convenience, we notice that (3.6) and (3.7) can be rewritten as (Ref. 15)

$$2\pi = [2y/\sqrt{x(y-1)}]\,\Pi(\alpha_2^2, k_2), \qquad (3.8)$$

$$T = [1/\sqrt{x(y-1)}][-\Pi(\alpha_1^2, k_1) + yK(k_1)]. \qquad (3.9)$$

Here $K(k)$ and $\Pi(\alpha^2, k)$ are complete elliptic integrals of the first and third kind, respectively. The parameters k (modulus) and α are given by

$$k_1^2 = (x-1)y/(y-1)x, \quad k_2^2 = 1 - k_1^2,$$
$$\alpha_1^2 = (x-1)/x, \quad \alpha_2^2 = 1/(1-y). \qquad (3.10)$$

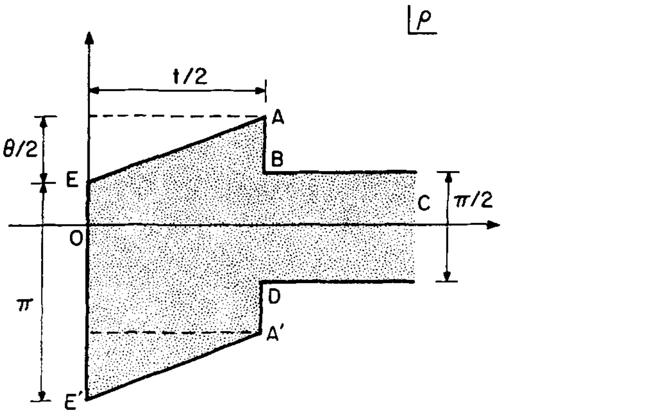


FIG. 9. Region of the complex plane corresponding to half of the tadpole diagram for the case of complex $T = t + i\theta$.

It is worthwhile to realize[16] that there is an expected symmetry in expressions (3.6) and (3.7). In order to see this, consider again Fig. 1. It is clear that the answer for the tadpole amplitude should be independent of an arbitrary rotation of the circles Q_0 or Q_t around the origin by any integer multiple of 2π. According to the interpretation of the propagator parameter $T = t + i\theta$, given in Sec. II, a full rotation of the circles results, in our conventions, in a shift of T to $T \pm i\pi$. Under this operation, we should get the same torus up to a disconnected diffeomorphism. It can be verified that when $x \to 1 - x$ and $y \to 1 - y$, (3.6) is invariant and (3.7) shifts according to $T \to T \pm i\pi$ [this requires a short calculation using the invariance of (3.6)]. But $x \to 1 - x$ corresponds to $\tau \to 1 \pm \tau$ [Eq. (2.4)] and indeed we recover the same torus up to one of the generators of the modular group.

The region of integration in moduli space: Our strategy for determining the region of moduli space covered by the diagram consists in finding the limit curve in the τ plane, namely, the boundary of the region in moduli space. To do so, we consider the boundary of the region $\mathscr{R}_T$ in the T plane, defined as the semi-infinite rectangle that satisfies $\mathrm{Re}(T) \geqslant 0$ and $-\pi/2 \leqslant \mathrm{Im}(T) \leqslant \pi/2$. This region represents the standard free propagator. Our calculation of $x(T)$ implies that we know $\tau(T)$. The limit curve is the image of the boundary of $\mathscr{R}_T$ in the τ plane under this map. Knowledge of this curve is enough to determine if the diagram reproduces or not the correct integration region in moduli space. In order to gain some understanding of the mapping, it is useful to consider some special limits in which analytic forms can be found for the previous expressions.

First we would like to consider the limiting case defined by $x \to \infty$ and $y \to \infty$. For simplicity we consider the case of real x and y and then we extend the results by analyticity. We look for an approximation to the integral (3.6) for the case in which both x and y are real and large. Since $x < y$, if $x \to \infty$ we get the condition $y = 4x + O(1)$, in order that (3.6) be true to lowest order in x^{-1}. It is possible to obtain the following asymptotic relations, as it is shown in the Appendix:

$$y = 4x - \tfrac{3}{2} - \tfrac{39}{128}(1/x) + O(x^{-2}), \qquad (3.11)$$

$$T(x) = \ln x + \tfrac{3}{2}\ln(\tfrac{16}{3}) + O(x^{-1}). \qquad (3.12)$$

Using (2.10), we see that in this limit $T(x)$ and $\tau(x)$ are similar and to leading order we have a linear dependence,

$$T(x) = -i\pi\tau(x) \quad (x\to\infty), \qquad (3.13a)$$

which implies

$$\tau = (i/\pi)T \quad (T\to\infty). \qquad (3.13b)$$

Subsequent terms do not follow this relation. This is the expected result since the region that we are considering corresponds to the case of both $\mathrm{Re}(T)$ and $\mathrm{Im}(\tau)$ large. In this region, the mapping that relates T and τ is the mapping between two strips and therefore it should be of the form (3.13b). From a physical point of view, it is clear that in this limit the length of the intermediate tube is very large and so T is essentially the modular parameter independently of the details of the gluing at the end points of the tube.

The second limit we will consider is given by $x\to 1$, which corresponds to small intermediate times and is characterized by $\alpha_1\to 0$ and $k_1\to 0$. Note that if $x\to 1$ then $y\to 1$, in order that (3.6) remains infinite. By using the asymptotic expansion for $\Pi(\alpha^2,k)$, given in the Appendix, and setting $x = 1+\epsilon$, where ϵ is small, (3.8) gives

$$\epsilon = 16\frac{(y-1)}{y}\exp\left[-\frac{2}{\sqrt{y-1}}\left(\pi - \tan^{-1}\frac{1}{\sqrt{y-1}}\right) + O\left(\frac{\epsilon}{(y-1)^2}\right)\right] + O(\epsilon^2). \qquad (3.14)$$

In the limit $\epsilon\to 0$, we can invert the above relation to get

$$y(x) = 1 + \pi^2/\ln^2(x-1) + O(\ln^{-3}(x-1)). \qquad (3.15)$$

The time integral can be evaluated in a similar fashion, using expansions for the elliptic integrals given in the Appendix. [We notice that $\alpha_1^2 = O(\epsilon)$ while $k_1^2 = O(\epsilon\ln^2\epsilon)$.] The result is

$$T = \frac{\pi}{2}\sqrt{y-1} + \epsilon\frac{\pi}{4}\left(\frac{1}{2}\frac{2-y}{\sqrt{y-1}} - 1\right) + O(\epsilon^2\ln^4\epsilon). \qquad (3.16)$$

Note that T can also be written in terms of x by replacing (3.15) in (3.16):

$$T(x) = -(\pi^2/2)[1/\ln(x-1)] + O((x-1)\ln(x-1)). \qquad (3.17)$$

Recalling (2.9) we observe that in this limit there is also a linear relation between the leading-order terms of $T(x)$ and $\tau(x)$, given by

$$T(x) = -i(\pi/2)\tau(x) \quad (x\to 1) \qquad (3.18a)$$

and therefore

$$\tau = i(2/\pi)T \quad (T\to 0). \qquad (3.18b)$$

This last property implies that the modular region will be covered an infinite number of times. This can be seen from the fact that as $T\to 0$, the part of the neighborhood of $T=0$ that is in $\mathscr{R}_T$ will be mapped under (3.18b) into a similar neighborhood around $\tau = 0$, as shown in the bottom portion of Fig. 10. Such a neighborhood contains an infinite number of copies of the modular region. We see that the vertex fails in providing a finite number of coverings because it does not reproduce the familiar wedge around $\tau = 0$ shown in Fig. 4(a).

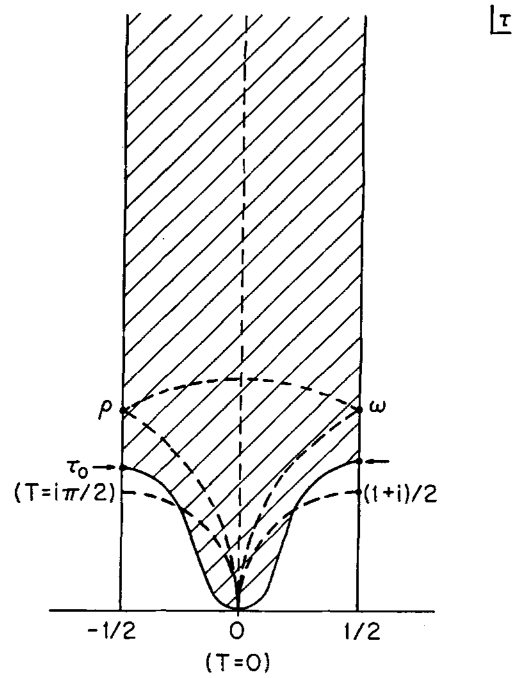


FIG. 10. Qualitative picture of the modular region covered by the one-loop tadpole graph.

For completeness, we now discuss briefly how to obtain a qualitative picture of the limit curve in the τ plane. The interesting part of this curve is the image of $T = i\theta$ in the τ plane. We wish to find the value τ_0 of the modular parameter for $T = i(\pi/2)$ (see Fig. 10). For x of the form $x = \frac{1}{2} + i\,\mathrm{Im}(x)$, y is of the form $y = \frac{1}{2} + i\,\mathrm{Im}(y)$, as can be verified using Eq. (3.6), and τ is of the form $\tau = -\frac{1}{2} + i\,\mathrm{Im}(\tau)$ (Fig. 4). Equation (3.7) and the values of x and y imply that T is of the form $T = i(\pi/2) + \mathrm{Im}(T)$. The actual value of $\mathrm{Im}(T)$ for a given x can be found by solving Eqs. (3.6) and (3.7) numerically. We find that $\mathrm{Im}(x) = 0.103$ and $\mathrm{Im}(y) = 1.382$ correspond to $\mathrm{Im}(T) = 0$ and therefore to $T = i(\pi/2)$. The corresponding value of τ_0 lies on the segment that separates region S_4 from S_6 with $\frac{1}{2} < \mathrm{Im}(\tau_0) < \sqrt{3}/2$ (Fig. 4). Given (3.18b), we know T for small values of τ, and therefore a continuous line starting at $\tau = 0$ and ending at $\tau = \tau_0$ must join all points corresponding to purely instantaneous twists $T = i\theta$ (Fig. 10). We observe that the curve does not appear to have any special modular properties and that the modular region is covered an infinite number of times.

IV. CONCLUSIONS

The method developed here for the determination of the limit curve in moduli space is quite general and may be applied to test other types of vertices. It is based on the representation of tori as two-sheeted spheres and the analyticity of the modular parameter with respect to the closed string propagator parameter.

As this paper has further confirmed, the naive closed string vertex is not a satisfactory vertex for covariant closed string field theory. It has been shown that it fails in giving the correct region of integration in moduli space for the one-loop tadpole. The reason for this failure is the infinite covering of the modular region that arises for very short intermediate times. This behavior can be traced back to the fact that in this vertex the interacting strings overlap too much. A condition that must be fulfilled by a given vertex, in order to achieve a finite number of coverings, is that it should be able to map the neighborhood of $T=0$, which is in $\mathscr{R}_T$ to the familiar wedge around $\tau=0$ in the modular plane. In fact, it is possible to obtain such behavior, as it is shown in Ref. 17. In view of the failure of the naive vertex in producing the correct answer for the tadpole problem, one may still consider some unlikely ways out. One possibility, suggested in Ref. 18, is that the quantization of covarient closed string field theory may not be straightforward and the tree-level Feynman rules need to be modified at the loop level. Another possibility is that even though this vertex fails in providing the right answer for surfaces with one puncture, it may still be capable of producing the correct answer for surfaces with two punctures or more.

Our method of analysis of closed string amplitudes by considering their behavior for real values of the closed string propagator parameter can be applied to the problem of four particle scattering in covariant closed string field theory. In fact, the open string results of Ref. 19 could be used as follows to verify that the closed string extension does not work.[4] Such results define a function $\lambda(t)$, which gives the $SL(2,C)$ invariant cross ratio as a function of the intermediate open string propagator parameter. For closed strings, one continues analytically, letting $t\to T$. As argued repeatedly in Ref. 17 we, in fact, know what $\lambda(T)$ would be. In order to get the complete region of integration $\lambda(T)$ must equal $\hat{\lambda}(T)$, where $\hat{\lambda}$ is a map from $\mathscr{R}_T$ into two copies of the fundamental region of the Λ group.[14] It is possible to decide whether or not $\lambda=\hat{\lambda}$ by considering their behavior on the real axis.

ACKNOWLEDGMENTS

We wish to thank J. I. Latorre for many useful discussions. Thanks are also due to M. Bershadsky and A. Eskin for raising interesting equations.

This work was supported in part by funds provided by the U.S. Department of Energy under Contract No. DE-AC02-76ER03069.

APPENDIX

1. Notation and definitions

Here we give the definition of the complete elliptic integrals according to Ref. 15. The complete elliptic integral of the first kind is defined as

$$K(k)\equiv\int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}=\int_0^{\pi/2}\frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}, \tag{A1}$$

whereas the complete elliptic integral of the third kind is defined as

$$\Pi(\alpha^2,k)\equiv\int_0^1\frac{dt}{(1-\alpha^2t^2)\sqrt{(1-t^2)(1-k^2t^2)}} =\int_0^{\pi/2}\frac{d\theta}{(1-\alpha^2\sin^2\theta)\sqrt{1-k^2\sin^2\theta}}. \tag{A2}$$

Here $\alpha^2\neq 1$. In both cases k is any complex number known as the modulus.

2. Useful expansions for the elliptic integrals

The following expansions for the elliptic integrals of the first and third kind are quoted from Ref. 15. For the elliptic integrals of the first kind,

$$K(k)=\ln\left(\frac{4}{k'}\right)+\frac{1}{4}\left[\ln\left(\frac{4}{k'}\right)-1\right]k'^2 +\frac{9}{64}\left[\ln\left(\frac{4}{k'}\right)-\frac{7}{6}\right]k'^4+\cdots \quad (k'\to 0), \tag{A3}$$

$$K(k)=(\pi/2)\left[1+\tfrac{1}{4}k^2+\tfrac{9}{64}k^4+\cdots\right] \quad (k\to 0). \tag{A4}$$

For the elliptic integrals of the third kind,

$$\Pi(\alpha^2,k)=(\pi/2)(1+\tfrac{1}{2}\alpha^2+\tfrac{1}{4}k^2+\cdots) \quad (k,\alpha\to 0), \tag{A5}$$

$$\Pi(\alpha^2,k)=[1/(1-\alpha^2)][\ln(4/k') +\sqrt{-\alpha^2}\tan^{-1}(\sqrt{-\alpha^2})]+O(k'^2). \tag{A6}$$

The last expression is valid for $k^2<1$ and $-\alpha^2>0$. In all cases $k'=\sqrt{1-k^2}$ is the complementary modulus.

3. Calculation of asymptotic expressions of relevant elliptic integrals

We look for an approximation to the integral (3.6) for the case in which both x and y are real and large,

$$\int_0^1 dz\left(\frac{(y-z)}{z(1-z)(x-z)}\right)^{1/2}=2\pi. \tag{A7}$$

If $x\to\infty$, we get the condition $y=4x+O(1)$, in order that (A7) be true to lowest order in x^{-1}. Next we set

$$y=4x+\xi(x),$$

where

$$\xi(x)=a_0+a_1/x+a_2/x^2+\cdots.$$

Expanding the integrand, we find

$$\left(\frac{y-z}{x-z}\right)^{1/2}=2\left\{1+\frac{1}{8x}(a_0+3z)+\frac{1}{8x^2}\times\left[\left(a_1-\frac{a_0^2}{16}\right)+\frac{5}{8}a_0z+\frac{39}{16}z^2\right]\right\}+O(x^{-3}).$$

Replacing in (A7) we get

$$2\pi+(1/x)I_1+(1/x^2)I_2+O(x^{-3})=2\pi,$$

where

$$I_1=\frac{1}{4}\int_0^1\frac{dz}{\sqrt{z(1-z)}}(a_0+3z),$$

$$I_2 = \frac{1}{4}\int_0^1 \frac{dz}{\sqrt{z(1-z)}}\left[\left(a_1 - \frac{a_0^2}{16}\right) + \frac{5}{8}a_0 z + \frac{39}{16}z^2\right].$$

By requiring $I_1 = 0$, $I_2 = 0$, etc., we get

$$y = 4x - \tfrac{3}{2} - \tfrac{39}{128}(1/x) + O(x^{-2}). \quad \text{(A8)}$$

In order to evaluate the time integral (3.7) it is convenient to define

$$t(x,u) = \frac{1}{2}\int_1^x dz\left(\frac{(u-z)}{z(z-1)(x-z)}\right)^{1/2}, \quad \text{(A9)}$$

so that

$$t(x,y) = t(x,x) + \int_x^y du\,\frac{dt(x,u)}{du}. \quad \text{(A10)}$$

Here

$$t(x,x) = \frac{1}{2}\int_1^x \frac{dz}{\sqrt{z(z-1)}} = \frac{1}{2}\ln 4x + O(x^{-1}), \quad \text{(A11)}$$

$$\frac{dt(x,u)}{du} = \frac{1}{4}\int_1^x \frac{dz}{\sqrt{z(z-1)(x-z)(u-z)}} = \frac{1}{2\sqrt{x(u-1)}}K(k), \quad \text{(A12)}$$

where

$$k^2 = (x-1)u/(u-1)x. \quad \text{(A13)}$$

We consider $x<u<y$. When x and y are large $k^2 \to 1$ and $k'^2 \to 0$, so that we can use (A3) in (A12) and replace this result for the integrand in (A10). Carrying out the integration in (A10) we obtain the desired result (3.12).

[1]M. Kaku and K. Kikkawa, Phys. Rev. D **10**, 110 (1974); M. Kaku, *ibid.* **10**, 3943 (1974); E. Cremmer and J.-L. Gervais, Nucl. Phys. B **76**, 209 (1974).
[2]E. Witten, Nucl. Phys. B **268**, 253 (1986).
[3]J. Lykken and S. Raby, Nucl. Phys. B **278**, 256 (1986); L. J. Romans, Phys. Lett. B **194**, 499 (1987).
[4]M. Peskin (unpublished); J. Lykken and S. Raby (unpublished).
[5]M. Kaku and J. Lykken, Phys. Rev. D **38**, 3067 (1988).
[6]B. Zwiebach, Ann. Phys. **186**, 111 (1988).
[7]S. B. Giddings and E. Martinec, Nucl. Phys. B **278**, 91 (1986).
[8]R. Woodard, Phys. Lett. B **213**, 144 (1988); C. B. Thorn, UFTP-88-7, Lectures given on superstrings at Spring School, Trieste, Italy, April 1988; M. Awada, Phys. Lett. B **215**, 642 (1988).
[9]M. Bershadsky and A. Radul, Int. J. Mod. Phys. A **2**, 165 (1987); L. Dixon, D. Friedan, E. Martinec, and S. Shenker, Nucl. Phys. B **282**, 13 (1987); D. Gross and P. Mende, *ibid.* **303**, 407 (1988).
[10]L. R. Ford, *Automorphic Functions* (Chelsea, New York, 1951).
[11]Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[12]G. Springer, *Introduction to Riemann Surfaces* (Chelsea, New York, 1957).
[13]A. LeClair, M. E. Peskin, and C. Preitschopf, Nucl. Phys. B **317**, 411, 464 (1989).
[14]B. Zwiebach, Phys. Lett. B **213**, 25 (1988).
[15]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, New York, 1971), 2nd ed.
[16]We thank M. Peskin for suggesting for us to check this property in our expressions.
[17]B. Zwiebach, Nucl. Phys. B **317**, 147 (1989).
[18]H. Hata, Phys. Lett. B **217**, 438 (1989).
[19]S. B. Giddings, Nucl. Phys. B **278**, 242 (1986).

Relations for Clebsch–Gordan and Racah coefficients in $su_q(2)$ and Yang–Baxter equations

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Komaba, Tokyo, 153, Japan

(Received 5 January 1989; accepted for publication 24 May 1989)

Relations are exploited among Clebsch–Gordan (CG) and Racah coefficients in the algebra $su_q(2)$, known as a deformation of su(2). These are used to show that the Yang–Baxter (YB) relation for the IRF (interaction round a face) model results from one of the symmetry relations for the 9-*j* symbol specific to $su_q(2)$, and that in an asymptotic limit this YB relation becomes the YB relation for the two-dimensional vertex model. The Racah coefficient, which has a particularly simple dependence on *q*, is efficiently used such that an asymptotic limit of the Racah coefficient is the CG coefficient and another limit gives the factorized *S* matrix of the vertex model.

I. INTRODUCTION

The quantum theory of angular momentum, i.e., the algebra su(2), has been widely used in many branches of physics.[1–5] In recent years, extensive studies on Yang–Baxter (YB) relations[6,7] (sufficient conditions to yield commuting transfer matrices specific to exactly solvable models) have inspired generalization of Lie groups, called quantum deformations of the Lie groups.[8–11] In particular, the generalized algebra of su(2), called the algebra $su_q(2)$,[8–13] has become of great importance in studies of the braid group,[14,15] six-vertex algebra,[16] etc.

The Clebsch–Gordan (CG) and the Racah coefficients, which play central roles in the theory of angular momentum, were generalized in terms of *q* analogs by Askey and Wilson.[17] Recently, these *q* analogs have been identified as the CG and Racah coefficients of the algebra $su_q(2)$. Kirillov and Reshetikhin[13] have discussed relations among these coefficients and the YB relations. The present author[18] has pointed out that a known relation among Racah coefficients, which describes one of the symmetry relations for the 9-*j* (or, generally, the 12-*j*) symbol, is a kind of YB relation for the IRF (interaction round a face) model[7] even in the framework of su(2).

The aim of this paper is to present some new relations among the CG and the Racah coefficients of the algebra $su_q(2)$ and to obtain YB relations in terms of these coefficients. One of our devices lies in expressing the CG coefficient as an asymptotic form of the Racah coefficient: The corresponding relationship in su(2) was given by Biedenharn.[19] This provides efficient manipulation that circumvents the very involved expression of the CG coefficient in $su_q(2)$. We show that one of the symmetry relations specific to a kind of 9-*j* symbol, which is described in terms of three Racah coefficients, produces the YB relation for the IRF model. In an asymptotic limit, the YB relation becomes that for the two-dimensional vertex model[7] which concerns the factorized S (R) matrix of the process

$$j_1m_1 + j_2m_2 \to j_1m_1' + j_2m_2'. \tag{1.1}$$

In the algebra $su_q(2)$, many of the quantities are functions in the indeterminate *q* (Refs. 8–17). While the *q* dependence of the CG coefficient is very complicated,[12,13] the expression of the Racah coefficient is easily transcribed into the corresponding expression of the algebra su(2) and vice versa. In our formalism, this advantage of the Racah coefficient is utilized as much as possible.

We use the 6-*j* symbol in place of the Racah coefficient: The distinction between these lies in a sign factor. It is partly because the 6-*j* symbol, as well as its modified form, is used in Ref. 13 and partly because we often use symmetries of the 6-*j* symbol.

In Sec. II, a brief review is given of the algebra $su_q(2)$, the CG coefficient, and the 6-*j* symbol (the Racah coefficient). Sections III–V constitute discussions on the CG coefficient, the 6-*j* symbol, and various new relations among them. In Sec. III, the CG coefficient is described as an asymptotic limit of the 6-*j* symbol. Symmetry relations for the CG coefficient are then deduced from those for the 6-*j* symbol. In Sec. IV, another asymptotic limit of the 6-*j* symbol is presented which plays a decisive role in discussing YB relations. Section V is devoted to various relations among CG coefficients and/or 6-*j* symbols. It includes two relations, each of which describes the 6-*j* symbol as a weighted linear sum of CG coefficients. In Sec. VI, a kind of 9-*j* symbol is defined. It has the same kinds of symmetries as the 9-*j* symbol of su(2). In Sec. VII, the operator (matrix) R is defined as a quantity to express the overlap of a pair of coupled bases, one specified by *q* and the other by $1/q$. Relations among R, CG coefficients, and 6-*j* symbols are discussed. In Sec. VIII, it is shown that one of the symmetry relations for the 9-*j* symbol is rewritten as the YB relation for the IRF model and that its asymptotic limit gives the YB relation for the vertex model. The operator R is interpreted as the factorized S matrix of the vertex model. Concluding remarks are given in Sec. IX.

II. PRELIMINARIES ON $su_q(2)$, THE CG COEFFICIENT, AND THE 6-*j* SYMBOL

This section is devoted to a brief review of the algebra $su_q(2)$. In particular, algebraic expressions for the CG coefficient and of the 6-*j* symbol, discussed by Kirillov and Reshetikhin,[13] are summarized. For recent studies on q analogs in group representations, see works by Milne.[20,21]

The algebra $su_q(2)$ is generated by the relation[8–13]

$$[X^{\pm}, H/2] = \mp X^{\pm}, \tag{2.1}$$

and

$$[X^+, X^-] = \frac{q^{H/2} - q^{-H/2}}{q^{1/2} - q^{-1/2}}, \tag{2.2}$$

where q is an indeterminate defined as a positive number. In the case of $q \to 1$, the operators $X^{\pm}$ and $H/2$ become $J^{\pm}$ and J_z of the algebra su(2) (Refs. 1–5), respectively. The generators H, $X^{\pm}$ acting on the bases $|jm\rangle$ yield

$$X^{\pm}|jm\rangle = ([j \mp m][j \pm m + 1])^{1/2}|jm \pm 1\rangle, \tag{2.3}$$

and

$$(H/2)|jm\rangle = m|jm\rangle. \tag{2.4}$$

For a real number n we define $[n]$ by

$$[n] = (q^{n/2} - q^{-n/2})/(q^{1/2} - q^{-1/2}). \tag{2.5}$$

Our definitions of q and $[n]$ are the same as those given in Ref. 13: q and $[n]$ here correspond to q^{-2} and (n) of Ref. 12, respectively. The product $[1][2]\cdots[n]$ is denoted as $[n]!$. We postulate that $[0]! = 1$ and $[n]! = 0$ if $n < 0$. Notice that

$[n]$ is invariant under replacement of q by $1/q$. (2.6)

The symbol $[n]$ with a natural number n is rewritten as

$$[n] = q^{(n-1)/2} + q^{(n-3)/2} + \cdots + q^{-(n-1)/2}, \quad n \geqslant 1. \tag{2.7}$$

Some miscellaneous relations for $[n]$ are

$$q^{-n/2}[n'] + [n]q^{n'/2} = [n + n'], \tag{2.8}$$

$$\begin{aligned}[2j+1][2j'+1] &= [2|j-j'|+1] + [2(|j-j'|+1)+1] \\ &\quad + \cdots + [2(j+j')+1],\end{aligned} \tag{2.9}$$

and

$$[n][n'+n''] - [n+n''][n'] = [n-n'][n'']. \tag{2.10}$$

The comultiplication (coproduct) Δ_q, which acts on the product of two bases, $v_i \otimes v_k$, is defined such that

$$\Delta_q(X^{\pm}) = X_i^{\pm} \otimes q^{H_k/4} + q^{-H_i/4} \otimes X_k^{\pm}, \tag{2.11a}$$

and

$$\Delta_q(H) = H_i \otimes I_k + I_i \otimes H_k, \tag{2.11b}$$

where I denotes the identity. The expression (2.11a) is a generalization of

$$J^{\pm} = J_i^{\pm} \otimes I_k + I_i \otimes J_k^{\pm}, \tag{2.12}$$

defined in the algebra su(2). Notice that the right-hand side (rhs) of (2.11b) is independent of q. In the case $q \to 1$, most of the expressions throughout the paper approach those of su(2). The coupled basis $|(j_i j_k)jm\rangle_q$ is defined such that the comultiplication $\Delta_q(X^{\pm})$ acting on it yields (2.3). The algebra on $su_q(2)$ is a Hopf algebra.[8–13] This means that the antipode and counit on the generators $X^{\pm}$ and H are properly defined. This specific feature is not referred to explicitly in the following discussions.

The CG coefficient is defined in the usual way as the expansion coefficient between the set of uncoupled bases and the set of coupled bases. It is expressed as[13]

$$\begin{aligned}(j_1 j_2 m_1 m_2|jm)_q &= \Delta(j_1 j_2 j) q^{\{x(j_1)+x(j_2)-x(j)+2(j_1 j_2 + j_1 m_2 - j_2 m_1)\}/4}\{[j_1+m_1]![j_1-m_1]![j_2+m_2]! \\ &\quad \times [j_2-m_2]![j+m]![j-m]![2j+1]\}^{1/2} \sum_z (-1)^z q^{-z(j_1+j_2+j+1)/2} \\ &\quad \times \{[z]![j_1+j_2-j-z]![j_1-m_1-z]![j_2+m_2-z]![j-j_2+m_1+z]![j-j_1-m_2+z]!\}^{-1}.\end{aligned} \tag{2.13}$$

The sum over z is taken such that none of the factorials could have a negative argument. We have defined

$$\Delta(abc) = \left(\frac{[-a+b+c]![a-b+c]![a+b-c]!}{[a+b+c+1]!}\right)^{1/2}, \tag{2.14}$$

and

$$x(a) = a(a+1). \tag{2.15}$$

The symbol $\Delta(abc)$ is invariant under replacement of q by $1/q$, as is shown from (2.6). [The symbol Δ is used also to denote comultiplication (2.11), which, however, will not give rise to serious confusion.] The symbol $x(a)$ is the one denoted as $c(a)$ in Ref. 13. The expression (2.13) is the q analog of the formula by van der Waerden and by Racah.[1] It is normalized so as to be unity in the case of $j = m = j_1 + j_2$.

For later convenience and for checking expressions in this paper, we give (2.13) in a few special cases:

$$(jjm-m|00) = (-1)^{j-m} q^{m/2}/\sqrt{[2j+1]}, \tag{2.16}$$

$$\begin{aligned}(j_1 j_2 m_1 m_2|jj)_q &= (-1)^{j_1-m_1} q^{\{-x(j_1)+x(j_2)-x(j)+2m_1(j+1)\}/4} \\ &\quad \times \left(\frac{[2j+1]![j_1+m_1]![j_2+m_2]![j_1+j_2-j]!}{[j_1-m_1]![j_2-m_2]![j_1-j_2+j]![-j_1+j_2+j]![j_1+j_2+j+1]!}\right)^{1/2}\end{aligned} \tag{2.17}$$

The 6-j symbol is defined in the conventional way as the transformation coefficient between different coupling schemes of three angular momenta. It is expressed in terms of CG coefficients by

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = (-1)^{a+b+c+d}(edm_{12}m_3|cm)_q{}^{-1} \sum_{m_1} (abm_1m_2|em_{12})_q$$
$$\times (bdm_2m_3|fm_{23})_q (afm_1m_{23}|cm)_q/\sqrt{[2e+1][2f+1]}. \tag{2.18}$$

The 6-j symbol multiplied by $(-1)^{a+b+c+d}$ gives the Racah coefficient, $W(abcd;ef)$. Kirillov and Reshetikhin[13] deduced the following expression of the 6-j symbol, using (2.13) and (2.18):

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = \Delta(abe)\Delta(acf)\Delta(cde)\Delta(dbf) \sum_z (-1)^z[z+1]!\{[z-a-b-e]![z-a-c-f]![z-b-d-f]!$$
$$\times [z-d-c-e]![a+b+c+d-z]![a+d+e+f-z]![b+c+e+f-z]!\}^{-1}. \tag{2.19}$$

The rhs of (2.19) is simpler than that of (2.13), the expression for the CG coefficient, since (2.19) is described only in terms of symbols [ ]!. From (2.6) or (2.19), we see that the 6-j symbol is invariant under the replacement of q by $1/q$: Because of this, we abbreviate the suffix q or $1/q$ of the 6-j symbol. Furthermore, expression (2.19) is transcribed into a known expression for the 6-j symbol of su(2) by means of replacement of all the symbols $[n]!$ by $n!$. This transcription cannot be applied to the expression of the CG coefficient, (2.13), in general. This transcription is not available even with such symmetric CG coefficients as, for example, $(j_1j_200|j0)_q$.

Various symmetries of the 6-j symbol that are known in the algebra of su(2) (Refs. 3 and 5) hold also in the algebra of $su_q(2)$. That is, the 6-j symbol of this paper is invariant under any permutation of columns and also under an interchange of upper and lower arguments in each of any two of its columns. Further, Regge's symmetry holds.[22]

The CG coefficient and the 6-j symbols in the algebra of $su_q(2)$ fulfills the same forms of orthogonality relations as those[3,5] in the algebra su(2). For example, it holds that

$$\sum_j (j_1j_2m_1m_2|jm)_q (j_1j_2m_1'm_2'|jm)_q = \delta(m_1,m_1')\delta(m_2,m_2'). \tag{2.20}$$

III. THE CG COEFFICIENT AS AN ASYMPTOTIC LIMIT OF THE 6-j SYMBOL

Here, we show that the CG coefficient can be expressed as an asymptotic limit of the 6-j symbol: The corresponding relation in su(2) is well known.[3,19] The relation provides a way to circumvent uses of the involved expression (2.13).

We prepare the following expression for studies of various expressions in asymptotic limits:

$$\frac{[S+a]!}{[S+b]!} \text{ with a very large } S$$
$$\approx \begin{cases} [S]^{a-b}q^{\{x(a)-x(b)\}/4}, & \text{if } q\geqslant 1 \quad (3.1a) \\ [S]^{a-b}q^{-\{x(a)-x(b)\}/4}, & \text{if } q<1. \quad (3.1b) \end{cases}$$

The factor $[S]$ on the rhs is asymptotically $q^{S/2}/(q^{1/2}-q^{-1/2})$ in the case of $q>1$ and $q^{-S/2}/(q^{-1/2}-q^{1/2})$ in the case of $q<1$.

In order to relate (2.19) with (2.13), let us replace labels of (2.19) such that $c\to S+c$, $d\to S+d$, and $f\to S+f$. We replace further the label z as $z\to 2S-z+a+b+c+d$ and $z\to 2S+z+c+d+e$, respectively, for $q\geqslant 1$ and $q<1$. We subsequently make $S\to\infty$ in (2.19), using (3.1a) and (3.1b). We compare the resultant expression with (2.13) to obtain

$$\lim_{S\to\infty} (-1)^{2S}\begin{Bmatrix} a & b & e \\ S+d & S+c & S+f \end{Bmatrix}\sqrt{[2S+1]}$$
$$= \begin{cases} (-1)^{a+b+c+d}q^{-f/2}(a,b,f-c,d-f|e,-c+d)_q/\sqrt{[2e+1]}, & q\geqslant 1, \quad (3.2a) \\ (-1)^{c+d+e}q^{f/2}(a,b,c-f,f-d|e,c-d)_q/\sqrt{[2e+1]}, & q<1. \quad (3.2b) \end{cases}$$

Let us show that the CG coefficient has the symmetry,

$$(j_1j_2m_1m_2|jm)_q = (-1)^{j_1+j_2-j}(j_1j_2-m_1-m_2|j-m)_{1/q}. \tag{3.3}$$

Proof: We prove it separately for the cases of $q>1$, $q<1$, and $q=1$. In the case of $q>1$, we replace q in (3.2b) by $1/q$, and compare with (3.2a). After changing notations such that $a=j_1$, $b=j_2$, $e=j$, $f-c=m_1$, $d-f=m_2$, and $d-c=m$, we get (3.3). To prove (3.3) in the case of $q<1$ we replace q in (3.2a) by $1/q$ to compare with (3.2b). Putting $a=j_1$, $c-f=m_1$, $f-d=m_2$, etc., we get (3.3). The relation (3.3) with $q=1$ is well known[3] in the algebra su(2).

Various symmetry relations for the CG coefficient other than (3.3) can be deduced from (3.2a) and (3.2b) being combined with symmetry relations for the 6-j symbol. Among them, we have

$$(j_1j_2m_1m_2|jm)_q = (-1)^{j_1+j_2-j}(j_2j_1m_2m_1|jm)_{1/q} \tag{3.4}$$

$$= (-1)^{j_2 - j - m_1} q^{m_1/2} \left(\frac{[2j+1]}{[2j_2+1]} \right)^{1/2} (j_1 j - m_1 m | j_2 m_2)_q \tag{3.5}$$

$$= \{ (j_1 + j_2 + m)/2, (j_1 + j_2 - m)/2, (j_1 - j_2 + m_1 - m_2)/2, (j_1 - j_2 - m_1 + m_2)/2 | j, j_1 - j_2 \}_q . \tag{3.6}$$

The last relation, which expresses Regge's symmetry[23] of the CG coefficient, is linked to Regge's symmetry[22] of the 6-j symbol. Kirillov and Reshetikhin[13] deduced (3.4) and (3.5), using (2.13): A misprint in their expression is corrected in (3.5). The combination of (3.3–3.6) generates q analogs of all the known symmetry relations for the CG coefficient of su(2).

It is possible to define the q analog of the 3-j symbol by the rhs of (3.2). We should multiply, however, an extra factor $q^{\pm (c+d+f)/6}$, where $\pm$ means $+$ or $-$ according to $q \geqslant 1$ or $q < 1$, to unify (3.2a) and (3.2b). Symmetries of the 3-j symbol, rewritings of (3.3)–(3.6), are linked to the symmetries of the 6-j symbol.

In a previous paper,[24] the author presented a new expression of the 6-j symbol in the algebra su(2). Transcription of it in the q analog is done with the replacement of every factorial $n!$ by its q analog $[n]!$ so as to yield

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = (-1)^{c+d+e} (bae)(cde) / \{ (caf)(bdf) \}$$

$$\times \sum_z \frac{(-1)^{a-z} [a+z]! [c+f-z]! [b-c+d+z]!}{[a-z]! [b-e+z]! [b+e+z+1]! [-c+f+z]! [-b+c+d-z]!} , \tag{3.7}$$

where

$$(abc) = \left(\frac{[a+b-c]! [a-b+c]! [a+b+c+1]!}{[-a+b+c]!} \right)^{1/2} ; \tag{3.8}$$

it is equivalent to (2.19), though these are apparently different from each other.

There is another expression of the CG coefficient, which is apparently different from (2.13):

$$(j_1 j_2 m_1 m_2 | jm)_q = (-1)^{j_1 - m_1} q^{\{ -x(j_1) + x(j_2) - x(j) \}/4 + m_1 (m+1)/2}$$

$$\times \left(\frac{[j_1 - m_1]! [j_2 - m_2]! [j+m]! [j-m]! [j_1 + j_2 - j]! [2j+1]}{[j_1 + m_1]! [j_2 + m_2]! [j_1 - j_2 + j]! [-j_1 + j_2 + j]! [j_1 + j_2 + j + 1]!} \right)^{1/2}$$

$$\times \sum_z (-1)^z q^{z(j+m+1)/2} \frac{[j_1 + m_1 + z]! [j_2 + j - m_1 - z]!}{[z]! [j - m - z]! [j_1 - m_1 - z]! [j_2 - j + m_1 + z]!} . \tag{3.9}$$

Proof: We prove separately the cases $q > 1$, $q < 1$, and $q = 1$. In the case of $q > 1$, we replace the labels in (3.7) as $a \to b$, $b \to S + f$, $c \to e$, $d \to S + c$, $e \to S + d$, $f \to a$, and $z \to -z + c + e - f$, and make $S \to \infty$ using (3.2a). We subsequently change notations of labels such that $a = j_1$, $b = j_2$, $e = j$, $f - c = m_1$, $d - f = m_2$, $d - c = m$. We compare the resultant expression with (3.2), and get (3.9). In the case of $q < 1$, we first put $q = 1/q'$ in (3.9) so as to be $q' < 1$. Next, we use (3.4) to rewrite the left-hand side (lhs) as $(-1)^{j_1 + j_2 - j} (j_2 j_1 m_2 m_1 | jm)_{q'}$. We subsequently interchange labels as $j_1 \leftrightarrow j_2$ and $m_1 \leftrightarrow m_2$ together with the replacement of z by $j - m - z'$. After replacing q' and z' by q and z, respectively, we get (3.9) with $q < 1$. In the case of $q = 1$, the expression (3.9) becomes the so-called Racah's first form.[3]

The expression (3.9) is to (3.8) what (2.13) is to (2.19).

IV. ANOTHER ASYMPTOTIC LIMIT OF THE 6-j SYMBOL

In this section, we give another asymptotic limit of the 6-symbol.

Let us replace labels of (2.19) such that $b \to S + b$, $c \to S + c$, $e \to S + e$, $f \to S + f$, and $z \to 2S + z'$ so as to make $S \to \infty$. It is shown that the label z' does not exceed the minimum of $a + b + c + d$ and $a + d + e + f$ and that the term of the largest order in S is specified with the largest possible value of z', irrespective of q. Using these, we transform (2.19) separately for the cases of $q \geqslant 1$ and $q < 1$. We get the following result, valid for a positive number q:

$$\lim_{S \to \infty} (-1)^{2S+a+b+c+d} \begin{Bmatrix} a & S+b & S+e \\ d & S+c & S+f \end{Bmatrix} \frac{[2S+e+f+1]!}{[2S+b+c]!} = \begin{cases} T(a,d,f-c,b-f,b-e,e-c), & b+c \leqslant e+f, \\ 0 & b+c > e+f. \end{cases} \tag{4.1}$$

We have defined T by

$$T(a,d,a',d',a'',d'') = \frac{1}{[a'-a'']!} \left(\frac{[a+a']! [a-a'']! [d-d']! [d+d'']!}{[a-a']! [a+a'']! [d+d']! [d-d'']!} \right)^{1/2} \tag{4.2}$$

V. RELATIONS AMONG CG COEFFICIENTS AND/OR 6-j SYMBOLS

Here, we give some relations among CG coefficients and/or 6-j symbols. Some new devices are developed to deduce them.

Let us put in (2.18) $m_3 = d$ and $m = c$. We make use of (2.17) to express the first and the last two CG coefficients on the rhs of (2.18) in terms of q-analog factorials. It then follows that

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = q^{\{-x(a)-x(b)+x(e)-2x(c)+2x(f)\}/4} \frac{([c+d-e]![c+d+e+1]!)^{1/2}}{(caf)(dbf)}$$
$$\times\sum_m (-1)^{e+d+f+m} q^{m(c+d+1)/2} \frac{[c+f-m]!}{[-c+f+m]!}\left(\frac{[a+m]![b-c+d+m]!}{[a-m]![b+c-d-m]!}\right)^{1/2}$$
$$\times (a,b,m,c-d-m|e,c-d)_q/\sqrt{[2e+1]}. \qquad (5.1)$$

We have another expression similar in form to (5.1):

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} = (-1)^{a+b+c+d} q^{\{-x(a)-x(b)+x(e)-2x(c)+2x(f)\}/4} \frac{(caf)(dbf)}{\{[c+d-e]![c+d+e+1]!\}^{1/2}}$$
$$\times\sum_m \frac{q^{-m(c+d+1)/2}}{[c-f+m]![c+f+m+1]!}\left(\frac{[a+m]![b+c-d+m]!}{[a-m]![b-c+d-m]!}\right)^{1/2}$$
$$\times (a,b,m,-c+d-m|e,-c+d)_q/\sqrt{[2e+1]}. \qquad (5.2)$$

In the algebra su(2), the present author[24] obtained (5.2) with $q=1$.

The relation (5.1) is termwise invariant under the replacement

$$a\to s-d,\quad b\to s-c,\quad c\to s-b,$$
$$d\to s-a,\quad m\to m+(a-b-c+d)/2. \qquad (5.3)$$

Similarly, (5.2) is termwise invariant under the replacement

$$a\to s-c,\quad b\to s-d,\quad c\to s-a,$$
$$d\to s-b,\quad m\to m+(a-b+c-d)/2. \qquad (5.4)$$

These invariances result from the Regge's symmetries of the 6-j symbol and of the CG coefficient, (3.6).

The following relation holds among CG coefficients:

$$(j_1j_2m_1m_2|jm)_q = (-1)^{j_1+j_2-j} q^{-\{x(j_1)+x(j_2)-x(j)\}/2} \sum_{m_2'>m_2} (-1)^{m_1-m_1'} q^{\{m_1-m_1'-(m_1+m_1')(m_2+m_2')\}/4}$$
$$\times (q^{1/2}-q^{-1/2})^{m_1-m_1'} (j_1j_2-m_1'-m_2'|j-m)_q T(j_1,j_2,m_1,m_2,m_1',m_2'). \qquad (5.5)$$

Proof: We consider the case of $q\neq 1$, since (5.5) with $q=1$ gives a trivial equality. Let us replace labels of (5.1) such that $c\to S+c$, $d\to S+d$, and $f\to S+f$ to make $S\to\infty$. To its rhs and lhs we apply (3.1) and (3.2), respectively: The cases of $q>1$ and $q<1$ are to be treated separately. We replace labels such that $a\to j_1$, $b\to j_2$, and $c\to j$. Further, we put $f-c\to m_1$, $d-f\to m_2$, $d-c\to m$, $m\to -m_1'$, $m+d-c\to m_2'$ in the case of $q>1$, and put $c-f\to m_1$, $f-d\to m_2$, $c-d\to m$, $m\to m_1'$, $c-d-m\to m_2'$ in the case of $q<1$. It then follows (5.5).

We transform (5.5), using (2.20) and (3.3), to get

$$\sum_j q^{\{x(j_1)+x(j_2)-x(j)\}/2} (j_1j_2m_1m_2|jm)_q (j_1j_2m_1'm_2'|jm)_{1/q}$$
$$= \begin{cases} (-1)^{m_1-m_1'} q^{\{m_1-m_1'-(m_1+m_1')(m_2+m_2')\}/4} (q^{1/2}-q^{-1/2})^{m_1-m_1'} T(j_1,j_2,m_1,m_2,m_1',m_2'), m_1\geqslant m_1', \\ 0, m_1<m_1' \end{cases}. \qquad (5.6)$$

The combination of (4.1) with (5.6) yields

$$\lim_{S\to\infty} (-1)^{2S+a+b+c+d} \begin{Bmatrix} a & S+b & S+e \\ d & S+c & S+f \end{Bmatrix} \frac{[2S+e+f+1]!}{[2S+b+c]!}$$
$$= (-1)^{b+c-e-f} q^{\{x(b)+x(c)-x(e)-x(f)\}/4} q^{-(bc-ef)/2} (q^{1/2}-q^{-1/2})^{b+c-e-f} \sum_j q^{\{x(a)+x(d)-x(j)\}/2}$$
$$\times (a,d,f-c,b-f|j,b-c)_q (a,d,b-e,e-c|j,b-c)_{1/q}. \qquad (5.7)$$

Corresponding to Racah's sum rule in the algebra su(2), we have the relation

$$\sum_g (-1)^{e+f+g}[2g+1] \begin{Bmatrix} a & d & g \\ c & b & e \end{Bmatrix} \begin{Bmatrix} a & d & g \\ b & c & f \end{Bmatrix} q^{-\{x(e)+x(f)+x(g)\}/2} = \begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} q^{-\{x(a)+x(b)+x(c)+x(d)\}/2}. \qquad (5.8)$$

Proof: Let us show that the lhs is transformed into the rhs. In the first 6-j symbol on the lhs, we substitute (5.1) after labels b, c, d, e, and f in (5.1) are replaced by d, b, c, g, and e, respectively. In a similar way, we substitute (5.2) in the second 6-j symbol on the lhs of (5.8), after labels b, d, and e in (5.2) are replaced by d, b, and g, respectively. The sum over g on the lhs of (5.8) is then carried out by using the orthogonality relation for the CG coefficient, (2.20). We compare the resultant expression with the expression (3.7), and obtain the rhs of (5.8).

The relation (5.7) can be deduced also from an asymptotic limit of (5.8). To show it, we replace b, c, e, and f in (5.8) by $S+b$, $S+c$, $S+e$, and $S+f$, respectively, and make $S\to\infty$. The cases $q\geqslant 1$ and $q<1$ are treated separately. We apply (3.2) to two 6-j symbols on the lhs and obtain (5.7).

The relation (5.8) in the case of $f=0$, $a=d$, and $b=c$ gives

$$\sum_g [2g+1]\begin{Bmatrix} a & b & e\\ a & b & g\end{Bmatrix}_q q^{-\{x(e)+x(g)\}/2} = (-1)^{2a+2b}q^{-x(a)-x(b)}. \qquad (5.9)$$

From the combination of (5.1) and (5.2), where q in (5.2) is replaced by $1/q$, we can deduce the orthogonality relation for the 6-j symbol. This in turn gives a check of (5.1) and (5.2).

The Biedenharn–Elliott (BE) rule in the algebra su(2) holds also in su$_q$(2):

$$\sum_z (-1)^{b+f+z}[2z+1]\begin{Bmatrix} a & c & h\\ k & g & z\end{Bmatrix}\begin{Bmatrix} a & g & z\\ f & b & e\end{Bmatrix}\begin{Bmatrix} c & z & k\\ f & d & b\end{Bmatrix} = (-1)^{a+c+k+g+d+e+h}\begin{Bmatrix} a & c & h\\ d & e & b\end{Bmatrix}\begin{Bmatrix} h & g & k\\ f & d & e\end{Bmatrix}. \qquad (5.10)$$

Here, the labels of the 6-j symbols are arranged in a way suitable for the deduction of (7.12) given later.

The simplest way to deduce (5.10) is to make use of the transformation from one specific coupling scheme of four angular momenta to another specific one, the way well known in the algebra su(2) (Refs. 3–5). There is no additional q-dependent factor in (5.10), since transposition of angular momenta is not involved in the transformation of the coupling schemes of concern.

VI. A KIND OF 9-j SYMBOL AND ITS SYMMETRIES

Here, a kind of 9-j symbol is presented which has specific symmetries similar to the 9-j symbol of su(2).

In the algebra su(2), the 9-j symbol is defined as the transformation coefficient between different coupling schemes of four angular momenta.[3–5] Two defining expressions, equivalent to each other, of the 9-j symbol are widely known, one in terms of six CG coefficients and the other in terms of three 6-j symbols.[3–5] As is discussed on p. 458 of Ref. 4, the transformation of the coupling schemes consists of three pure recouplings and two transpositions.

In the algebra su$_q$(2), the situation is slightly different. The transformation coefficient is really expressed in terms of six CG coefficients. However, two of the coefficients, whose labels are arranged in the standard form, are specified by the indeterminate $1/q$, while the other four are by q. It is due to the presence of transpositions involved in the recoupling of four angular momenta. Contrary to the case of su(2), the transformation coefficient is not expressed as a sum over products of three 6-j symbols.

Here, we define an analog of the 9-j symbol by

$$\begin{Bmatrix} a & b & e\\ c & d & f\\ h & k & g\end{Bmatrix}_q = \sum_z (-1)^{2z}q^{-\{x(z)+x(h)+x(d)+x(e)\}/2}[2z+1] \times\begin{Bmatrix} a & c & h\\ k & g & z\end{Bmatrix}\begin{Bmatrix} b & d & k\\ c & z & f\end{Bmatrix}\begin{Bmatrix} e & f & g\\ z & a & b\end{Bmatrix}. \qquad (6.1)$$

This 9-j is not invariant under interchange of q and $1/q$, contrary to the 6-j symbol.

Significance of the symbol defined by (6.1) lies in the following symmetry relation, valid for every $q>0$,

$$\begin{Bmatrix} a & b & e\\ c & d & f\\ h & k & g\end{Bmatrix}_q = (-1)^A q^{-x'(A)/2}\begin{Bmatrix} a & e & b\\ c & f & d\\ h & g & k\end{Bmatrix}_{1/q}, \qquad (6.2)$$

in which A and $x'(A)$ are defined by

$$A=\sum_{i=1}^{9} a_i = a+b+c+d+e+f+g+h+k, \qquad (6.3)$$

and

$$x'(A)=\sum_{i=1}^{9} x(a_i) = x(a)+x(b)+\cdots+x(k). \qquad (6.4)$$

Proof: We rewrite the first 6-j symbol on the rhs of (6.1) in terms of two 6-j symbols by using the Racah's sum rule (5.8) after labels in (5.8) are replaced such that $b\to c$, $c\to g$, $d\to k$, $e\to h$, and $f\to z$ together with $q\to 1/q$. Subsequently, we use the BE rule (5.10) for taking the sum over z on the resultant rhs of (6.1). From this we get the rhs of (6.2).

The 9-j symbol satsifies the same kind of relation as (6.2) under interchange of the first two columns: The proof is essentially the same as that given to (6.2). Further, the 9-j symbol is unchanged under the interchange of rows and columns (matrix transposition).[3–5] From these it is shown that the 9-j symbol (6.1) has the same symmetries as that of su(2).

VII. THE MATRIX R EXPRESSED IN TERMS OF A 6-j SYMBOL

The operator (matrix) R is defined here as a quantity to express a degree of overlap of a pair of coupled bases, one specified by q and the other by $1/q$: The operator R is interpreted in the next section as the factorized S matrix that describes the process (1.1).

We define the operator R by

$$R^{j_1j_2}=\sum_{jm} |(j_1j_2)jm\rangle_q F(j_1j_2j)_q\langle (j_2j_1)jm|, \qquad (7.1)$$

in which

$$F(j_1j_2j) = (-1)^{j_1+j_2-j}q^{-\{x(j_1)+x(j_2)-x(j)\}/2}. \quad (7.2)$$

The inverse of R is shown to be

$$(R^{j_1j_2})^{-1} = \sum_{jm} |(j_2j_1)jm\rangle_q F(j_1j_2j)^{-1}{}_q\langle(j_1j_2)jm|. \quad (7.3)$$

The R matrix in the case of $q=1$ implies simply the unit operator.

It is shown from (7.1) that $\Delta_{1/q}(X^{\pm})R = R\Delta_q(X^{\pm})$. The operator R, defined by (7.1), corresponds to the operator R of Ref. 13.

The matrix element of R is expressed as

$$(R^{j_1j_2})^{m_1',m_2'}_{m_1,m_2} = \langle(j_1,m_1')(j_2,m_2')|R^{j_1j_2}|(j_1,m_1)(j_2,m_2)\rangle \quad (7.4)$$

$$= \sum_j F(j_1j_2j)(j_1j_2\,m_1'm_2'|jm)_q \times(j_2j_1m_2m_1|jm)_q \quad (7.5)$$

$$= \sum_j (-1)^{j_1+j_2-j}F(j_1j_2j) \times(j_1j_2\,m_1'm_2'|jm)_q(j_1j_2m_1m_2|jm)_{1/q}. \quad (7.6)$$

In the last step, (3.4) has been used. Similarly, we express the matrix element of R^{-1} in terms of CG coefficients. We compare the result with the lhs of (5.6) to find that

$$((R^{j_1j_2})^{-1})^{m_1',m_2'}_{m_1,m_2} = \text{the lhs of (5.6)}. \quad (7.7)$$

As can easily be shown, the operators $(R^{j_1j_2})^{-1}$ and $R^{j_1j_2}$ are interchanged if the indeterminate q is replaced by its inverse $1/q$.

We combine (5.6) with (7.6), using (7.7), to get

$$(R^{(j_1j_2)})^{m_1',m_2'}_{m_1,m_2} = \begin{cases} q^{\{-m_1+m_1'+(m_1+m_1')(m_2+m_2')\}/4}(q^{1/2}-q^{-1/2})^{m_1-m_1'}T(j_1,j_2,m_1,m_2,m_1',m_2'), & m_1\geqslant m_1', \\ 0, \quad m_1<m_1'. \end{cases} \quad (7.8)$$

As a special case, it follows that

$$(R^{(j_1j_2)})^{m_1,m_2}_{m_1,m_2} = q^{m_1m_2}. \quad (7.9)$$

The expression of R, (7.8), differs significantly from that of Ref. 13 as to q-dependent factors.

We modify the rhs of (7.6) using symmetries of the CG coefficient, (3.3) and (3.4), to get the following.

> The matrix element of R, (7.4), is invariant under replacement of labels (m_1, m_2, m_1', m_2') by $(-m_1', -m_2', -m_1, -m_2)$ and under replacement of $(j_1,j_2, m_1, m_2, m_1', m_2')$ by $(j_2,j_1, m_2', m_1', m_2, m_1)$. (7.10)

> The expression (7.8) actually fulfulls these symmetries. Combination of (4.1) with (7.8) gives

$$\lim_{S\to\infty} (-1)^{2S+a+b+c+d}\begin{Bmatrix} a & S+b & S+e \\ d & S+c & S+f\end{Bmatrix} \times\frac{[2S+e+f+1]!}{[2S+b+c]!} = (R^{ad})^{b-e,e-c}_{f-c,b-f}q^{-\{x(b)+x(c)-x(e)-x(f)\}/4} \times q^{(bc-ef)/2}(q^{1/2}-q^{-1/2})^{b+c-e-f}. \quad (7.11)$$

We have described the relationship between R, T, the CG coefficient, and the 6-j symbol in (4.1), (5.6), (5.7), (7.6), (7.8), and (7.11).

The BE rule (5.10) has the following asymptotic limit:

$$\sum_z (R^{af})^{g-z,z-b}_{e-b,g-e}(R^{cf})^{z-k,k-d}_{b-d,z-b} \times(a,c,g-z,z-k\,|h,g-k)_q = (R^{hf})^{g-k,k-d}_{e-d,g-e}(a,c,e-b,b-d\,|h,e-d)_q. \quad (7.12)$$

Proof: In (5.10) we replace labels as $z\to S+z$, $k\to S+k$, $g\to S+g$, $b\to S+b$, $e\to S+e$, $d\to S+d$ and make $S\to\infty$. The cases of $q>1$, $q<1$, and $q=1$ are treated separately. Here, we give the proof only for the case of $q<1$, for simplicity. We use (3.2b) for the first 6-j symbol on each side of (5.10), and use (7.11) for the other three 6-j symbols so as to express the $S\to\infty$ limit of (5.10) in terms of R and CG coefficients. Notice that $[2S+2z+1]$, for example, is of the form $q^{-(S+z+1/2)}/(q^{-1/2}-q^{1/2})$. After rewriting expressions by the use of (3.1b), we obtain (7.12).

VIII. YANG–BAXTER RELATIONS IN TERMS OF 6-j SYMBOLS

It is shown here that one of symmetry relations for the 9-j symbol generates YB relations for the IRF model and for the vertex model.

Invariance of the 9-j symbol (6.1) under the third column being put onto the lhs of the first column is expressed in terms of the 6-j symbol as

$$\sum_z (-1)^{2z}[2z+1]\begin{Bmatrix} a & g & z \\ k & c & h\end{Bmatrix}\begin{Bmatrix} a & e & b \\ f & z & g\end{Bmatrix}\begin{Bmatrix} k & b & d \\ f & c & z\end{Bmatrix}q^{-\{x(z)+x(h)+x(e)+x(d)\}/2} = \sum_z (-1)^{2z}[2z+1]\begin{Bmatrix} k & e & z \\ f & h & g\end{Bmatrix}\begin{Bmatrix} a & z & d \\ f & c & h\end{Bmatrix}\begin{Bmatrix} a & e & b \\ k & d & z\end{Bmatrix}q^{-\{x(z)+x(g)+x(c)+x(b)\}/2}. \quad (8.1)$$

The lhs is a reexpression of the rhs of (6.1). We have used symmetries of the 6-j symbol to arrange labels suitable for obtaining (8.5) given later. Let us define w and w' (the symbol w stands for a weight[7] and not for the Racah coefficient) by

$$w(h,c,z,g;a,k) = (-1)^{a+k+h+z} q^{\{x(c)+x(g)-x(h)-x(z)\}/2} \times \sqrt[4]{[2h+1][2c+1][2z+1][2g+1]} \times \begin{Bmatrix} a & g & z \\ k & c & h \end{Bmatrix}, \quad (8.2)$$

and

$$w'(h,c,z,g;a,k) = \sqrt[4]{[2h+1][2z+1]/\{[2c+1][2g+1]\}} \times w(h,c,z,g;a,k). \quad (8.3)$$

We express (8.1) in terms of w and w' as

$$\sum_z w(h,c,z,g;a,k)w(g,z,b,e;a,f)w'(z,c,d,b;k,f) = \sum_z w'(g,h,z,e;k,f)w(h,c,d,z;a,f)w(z,d,b,e;a,k). \quad (8.4)$$

In this relation, it is allowed to replace simultaneously all the functions w by the corresponding w'. The functional relation (8.4) is the YB relation for the IRF model exploited by Baxter[7]: For the prescription to construct commuting transfer matrices from w and w', see Ref. 7. In Ref. 18, the present author discussed (8.4) with $q=1$.

In (8.1) we put $z\to S+z$, $b\to S+b$, $c\to S+c$, $d\to S+d$, $e\to S+e$, $g\to S+g$, $h\to S+h$ and make $S\to\infty$. Subsequently, we use (7.11) to describe the asymptotic limit of (8.1) in terms of R. The cases of $q<1$, $q>1$, and $q=1$ are to be treated separately. It then follows that

$$\sum_z (R^{ak})^{g-z,z-c}_{h-c,g-h}(R^{af})^{e-b,b-z}_{g-z,e-g}(R^{kf})^{b-d,d-c}_{z-c,b-z} = \sum_z (R^{kf})^{e-z,z-h}_{g-h,e-g}(R^{af})^{z-d,d-c}_{h-c,z-h}(R^{ak})^{e-b,b-d}_{z-d,e-z}. \quad (8.5)$$

In the case of $q\leqslant 1$, each side is equal to

$$(-1)^{b-d+h-g} q^{-(b+d+h+g)/2}(q^{1/2}-q^{-1/2})^{-2} \times \lim_{S\to\infty} q^{x(S+c)+x(S+e)-2S-1} \times \begin{Bmatrix} S+b & S+e & a \\ S+d & f & S+c \\ k & S+g & S+h \end{Bmatrix}_q. \quad (8.6)$$

If $q>1$, expression (8.5) with q being replaced by $1/q$ gives each side of (8.5).

The specific form of (8.5) is known as the YB relation for the vertex model.[6,7] The matrix element of R describes two-particle scattering process (1.1). Beside q the labels a, k, and f act as parameters to specify $w^{(\prime)}$ and R. However, none of them could satisfy the additivity of spectral parameters.[7]

We compare (8.4) and (8.5) to find the correspondence

$$w^{(\prime)}(h,c,z,g;a,k) \leftrightarrow (R^{ak})^{g-z,z-c}_{h-c,g-h}. \quad (8.7)$$

This is in agreement with the so-called Wu–Kadanoff–Wegner transformation.[25–27] In the present formalism, the correspondence (8.7) is based on (7.11), and is linked originally to the correspondence of the 6-j symbol and the CG coefficient (3.2).

Using symmetries of the 6-j symbol, it is shown that

$$w^{(\prime)}(h,c,z,g;a,k) = w^{(\prime)}(z,g,h,c;a,k) = w^{(\prime)}(z,c,h,g;k,a). \quad (8.8)$$

The symmetries correspond to the symmetries of R, (7.10), which is compatible with (8.7).

IX. CONCLUDING REMARKS

We have discussed relations among CG coefficients and 6-j symbols (Racah coefficients, in the same sense) of the algebra $su_q(2)$, and have shown that one of symmetry relations for the 9-j symbol, described in terms of three 6-j symbols, generates YB relations for the IRF model and for the vertex model.

As shown in (3.2), the 6-j symbol becomes the CG coefficient in an asymptotic limit. This 6-j symbol becomes the matrix element of R in another asymptotic limit, as shown in (7.11). Manipulation of the algebra in $su_q(2)$ is remarkably simplified by virtue of these relations among R, the CG coefficient, and the 6-j symbol, together with the fact that the 6-j symbol in $su_q(2)$ has simple dependence on q in comparison with CG coefficient.

Relations such as (3.5), (3.9), (5.8) (5.9), and (7.8) were already given in Ref. 13. However, serious confusion seems to exist in each of their expressions, especially as to q-dependent factors. The expression of R we get in (7.8) differs from that of Ref. 13.

It is hoped that many of the present results will hold in various "quantum" groups, so far as the CG coefficient and the 6-j symbol are properly defined. It is easy to extend the present discussion to the algebra of $su_q(1,1)$ (Ref. 28), although we do not discuss it here.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor T. Kotera for guidance on recent topics of solvable models, and to Professor M. Wadati, Dr. A. Kuniba, and Mr. T. Deguchi for stimulating discussions on solvable models.

[1]G. Racah, Phys. Rev. **61**, 186 (1942); **62**, 438 (1942); **63**, 367 (1943).
[2]E. P. Wigner, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), p. 87.
[3]L. C. Biedenharn and J. D. Louck, *Encyclopedia of Mathematics and Its Applications*, edited by G.-C. Rota (Addison–Wesley, Reading, MA, 1981), Vol. 8.
[4]L. C. Biedenharn and J. D. Louck, *Encyclopedia of Mathematics and Its Applications*, edited by G.-C. Rota (Addison–Wesley, Reading, MA, 1981), Vol. 9.
[5]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963); A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[6]C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[7]R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).

[8]V. G. Drinfeld, Sov. Math. Dokl. **32**, 254 (1985).
[9]M. Jimbo, Lett. Math. Phys. **10**, 63 (1985).
[10]E. K. Sklyanin, Usp. Math. Nauk. **40**, 214 (1985).
[11]M. Rosso, Commun. Math. Phys. **117**, 581 (1988).
[12]V. Pasquier, Commun. Math. Phys. **118**, 355 (1988).
[13]A. N. Kirillov and N. Yu. Reshetikhin, "Representations of the algebra $U_q(2)$, q-orthogonal polynominals and invariants of links," preprint, 1988.
[14]V. F. R. Jones, Bull. Amer. Math. Soc. **12**, 103 (1985).
[15]Y. Akutsu, T. Deguchi, and M. Wadati, J. Phys. Soc. Jpn. **56**, 3464 (1987).
[16]H. M. Babujian and A. M. Tsevelik, Nucl. Phys. B **265**, 24 (1986).
[17]R. Askey and J. Wilson, SIAM J. Math. Anal. **10**, 1008 (1979).
[18]M. Nomura, J. Phys. Soc. Jpn. **57**, 3653 (1988).
[19]L. C. Biedenharn, J. Math. Phys. (M.I.T.) **31**, 287 (1953).
[20]S. C. Milne, Adv. Math. **57**, 71 (1985).
[21]S. C. Milne, Adv. Math. **58**, 1 (1985).
[22]T. Regge, Nuovo Cimento **11**, 116 (1959).
[23]T. Regge, Nuovo Cimento **10**, 544 (1958).
[24]M. Nomura, J. Phys. Soc. Jpn. **58**, 2655 (1989).
[25]F. Y. Wu, Phys. Rev. B **4**, 2312 (1971).
[26]L. P. Kadanoff and F. J. Wegner, Phys. Rev. B **4**, 3989 (1971).
[27]K. Sogo, Y. Akutsu, and T. Abe, Prog. Theor. Phys. **70**, 739 (1983).
[28]H. Ui, Ann. Phys. (New York) **49**, 69 (1968).

Vectorlike coherent states with noncompact stability group

L. C. Papaloucas
Institute of Mathematics, University of Athens, 106 79 Athens, Greece
J. Rembieliński and W. Tybor
Institute of Physics, University of Łódź, Nowotki 149/153, 90-236 Łódź, Poland

(Received 6 September 1988; accepted for publication 26 April 1989)

An investigation is made of coherent states that differ from the usual ones in two ways: (a) they are connected with the coset space G/H, where the stability subgroup H may be *noncompact*; and (b) the notion of an H-invariant ray is replaced by the more general notion of an H-invariant subspace. A general framework is given for vectorlike coherent states with the help of the nonlinear realization technique as well as with the rigged Hilbert space theory. The vectorlike coherent states are found for the Poincaré group and the hyperbolic coherent states are found for the SU(1,1) group.

I. INTRODUCTION

The coherent states technique is a powerful method widely applied to various branches of modern physics and mathematical physics as well as to mathematics. It was originated from the pioneering papers by Glauber[1] and Klauder.[2] The concept was generalized by Radcliffe,[3] Perelomov,[4] Thompson,[5] and Gilmore and co-workers[6,7] and next extensively applied and developed by a number of authors. An exhaustive review of the theory and applications of coherent states as well as a reference list is given in the excellent books by Klauder and Skagerstam,[8] Perelomov,[9] and Hecht.[10] We also should mention some works dealing with different systems of overcomplete states: the continuous representations of quantum states by Barut and Girardello,[11] Skagerstam,[12] and Nieto.[13]

In this paper we investigate the coherent states understood in the spirit of the Perelomov definition but with some generalization. In order to explain this fact let us recall the definition of Perelomov: "The system of states $|\Psi_g\rangle = U(g)|\Psi_0\rangle$, where g are elements of Lie group G, $U(G)$ is an unitary representation of G in the Hilbert space $\mathscr{H}$ and $|\Psi_0\rangle$ is a fixed vector in $\mathscr{H}$, is called the coherent states system." Thus a coherent state $|\Psi_g\rangle$ is determined by a point $x(g)$ of the coset space G/H, where H is the stability subgroup of the ray $|\Psi_0\rangle$. Now, because that vector lies in the proper Hilbert space, the stability group H is thus compact.[14] As we will see below, the compactness of H is a very convenient assumption for technical reasons. Although this last requirement increases our difficulties, it is highly desirable since it drastically extends the applicability of the coherent states method.

The aim of our paper is to propose a procedure for the case of the noncompact stability group. An additional generalization of the Perelomov definition we have adopted is the replacement of the notion of the stable ray $|\Psi_0\rangle$ by the stable (H-invariant) subspace of $\mathscr{H}$ (see also Refs. 10 and 15–24). The plan of this paper is as follows: In Sec. II we study problems connected with the noncompactness of H and we give a framework for dealing with such a case. Section III is devoted to a description of the vectorlike coherent states in terms of the language of nonlinear realizations. In Sec. IV we investigate the vectorlike coherent states for the Poincaré group with the Lorentz stability group. In Sec. V we find the hyperbolic coherent states for the SU(1,1) group. In Sec. VI we briefly discuss the results.

II. THE EXTENSION PROBLEM FOR REPRESENTATIONS OF *G*

To explain some characteristic features of the coherent states with a noncompact stability group let us begin with an elementary but very pedagogical example of the special nilpotent group the (Heisenberg–Weyl[2,5] group). The Lie algebra of this group is generated by $\hat{q}$, $\hat{p}$, and $\hat{e}$: $[\hat{q},\hat{p}] = i\hat{e}$, $[\hat{e},\hat{q}] = [\hat{e},\hat{p}] = 0$. The unitary representation of the Heisenberg–Weyl group, labeled by a real number λ, is given via the Stone[26] and von Neumann[27] theorems by

$$U_\lambda(x,p,\phi) = \exp[i(\lambda\phi I + p\hat{q} - x\hat{p})].$$

Here $\hat{q}$ and $\hat{p}$ are unbounded self-adjoint with domains dense in the underlying Hilbert space $\mathscr{H}$, while I is the identity and $\hat{e} = \lambda I$. For simplicity we choose $\lambda = 1$, i.e., $\hat{e} = I$. Note that the group space is the product $S^1 \times \mathbb{R}^2$.

Now, the "classical" set of coherent states is obtained by choosing the compact phase group $\exp(i\phi)$ as the stability group, by fixing its eigenvector, and finally by the action of the operator $U(x,p,\phi)$ on this eigenvector. As a result one obtains a set of states parametrized by the points of the coset space $S^1\times\mathbb{R}^2/S^1 \simeq \mathbb{R}$, i.e., by points of the plane (x,p).

Now, let us look for another choice of the stability subgroup, namely, the subgroup generated by $\hat{q}$ and $\hat{e}$: $\exp[i(\phi I + p\hat{q})]$. The eigenvectors of this group do not belong to the Hilbert space $\mathscr{H}$, but rather to the space of tempered distributions S' [the space of continuous linear functionals over some nuclear subspace S of $\mathscr{H}$ (see Ref. 28)]. They are simply the position operator eigenstates: $\hat{q}|q\rangle = q|q\rangle$. Therefore we extend the unitary representation of the Heisenberg–Weyl group from $\mathscr{H}$ to S' with help of the duality of S and S'. Finally, by the action of the extended operator U on $|q\rangle$ we obtain the set of "coherent states" $|q + x\rangle$, q fixed:

$$\exp[i(\phi I + p\hat{q} - x\hat{p})]|q\rangle = \exp[i(\phi + xp/2 + pq)]|q + x\rangle;$$

we have $q = 0$ without loss of generality. As the result we

 0022-2488/89/102406-05$02.50

obtain again the eigenstates of $\hat{q}$; they correspond to the points of the coset space $S^1 \times \mathbb{R}^2/S^1 \times \mathbb{R}^1 \simeq \mathbb{R}^1$ and lie outside the Hilbert space $\mathscr{H}$.

Some remarks follow from the above example.

(i) The noncompactness of the stability group H causes the nonexistence of H-invariant irreducible subspaces in $\mathscr{H}$.[14]

(ii) We have to extend the underlying space $\mathscr{H}$ appropriately as well as the representation $U(G)$ in order to find the H-invariant subspace.

(iii) By the group action on an H-invariant subspace we obtain the set of generalized coherent states lying outside the initial representation space $\mathscr{H}$.

We are going to show that such a construction can be done for many groups. As above, G is a Lie group, H its proper subgroup, and $U(G)$ is a unitary representation of G acting in the Hilbert space $\mathscr{H}$. Both G and H may be noncompact. Let us denote by $D \subset \mathscr{H}$ a set of differentiable (or analytic) vectors of the representation $U(G)$. For several cases, the set D is nuclear so D', the set of all continuous linear functionals on D, is a natural and minimal domain for working with algebraic infinitesimal methods, especially if operators with a continuous spectra should be diagonalized. This is strongly related to the problem of subduction $U(G)\downarrow H$ with a noncompact H as well as with reduction of $U(H)$ to the block-diagonal form. Now, having the Gel'fand-like triple[28] $D \subset \mathscr{H} \subset D'$ we can extend by duality the representation $U(G)$ to $U'(G)$ acting in D'. As was shown by Nagel and Lindbland,[14] an unusual property of $U'(G)$ is that the subduced representation $U'(H) = U'(G)\downarrow H$, with a noncompact H, contains, in general, a number of "redundant" unitary as well as nonunitary representations of H which do not appear in $U(H) = U(G)\downarrow H$. Only for compact H do the representations $U'(H)$ and $U(H)$ always have the same representation content.[14] In this latter case, H-invariant subspaces always belong to $\mathscr{H}$, so there is no reason to extend the representation space to D'; this is just the source of the simplification in the case of coherent states with a compact stability group. In the noncompact case we should find first a nuclear set D for $U(G)$. In Ref. 14 it was shown that D is nuclear at least for unitary irreducible representations of a large class of groups; namely, for semisimple G with finite center, semidirect product $A \times K$, where A is Abelian while K is compact, nilpotent G, and Poincaré-like groups (with $m^2 > 0$).

III. THE VECTORLIKE COHERENT STATES

This section is devoted to giving a general definition of the vectorlike coherent states and to studying their basic properties. To do this we use the technique of the nonlinear realizations of Lie groups of Coleman *et al.*[29] and Salam and Strathdee.[30] They define the (left) group action on the quotient space G/H using the decomposition of $g = g_{G/H}h$, where h belongs to the subgroup H, while $g_{G/H}$ belongs to the part of G corresponding to G/H. In the following we will denote $g_{G/H}$ by $\xi(x)$, where the x are coordinates in G/H. The relation

$$g\xi(x) = \xi(x')h(g,x) \tag{1}$$

defines the nonlinear transformation law for x, that is, a transitive realization of G on G/H. With the help of Eq. (1) it is possible to introduce a variety of realizations of G in the associated bundles $(G/H,V)$, where V is a representation space for the subgroup H, by the following transformation law[29,30]:

$$\begin{aligned} g\xi(x) &= \xi(x')h(g,x), \\ \phi' &= D(h(g,x))\phi, \end{aligned} \tag{2}$$

where $\phi \in V$ and $D(H)$ is a linear representation of H acting in V. Note that contrary to the action of G on G/H the corresponding action of G on V is not transitive; H acts on a orbit in V generated from a fixed vector, say, ϕ^0. Furthermore, if the stability group of ϕ^0 is the subgroup H^0 of H, we conclude that the manifold $(G/H,V)$ splits into the sum of the manifolds $(G/H,H/H^0)$ with a transitive group action. Now, let us define the "boost" from a fixed point (x^0,ϕ^0) of the orbit $(G/H,H/H^0)$ to an arbitrary one (x,ϕ) by

$$\begin{aligned} g_{x\phi}\xi(x^0) &= \xi(x)h_{x\phi}, \\ D(h_{x\phi})\phi^0 &= \phi. \end{aligned} \tag{3}$$

Here $h_{x\phi} = h(g_{x\phi},x^0)$ [see Eq. (2)]. The boost $g_{x\phi}$ is defined by Eq. (3) up to an arbitrary element of H^0.

Let us assume that a unitary irreducible representation $U(G)$ of G has in $\mathscr{H}$ the set D of differentiable (or analytic) vectors which is dense in $\mathscr{H}$ and nuclear. So we can extend U to U' acting in D' dual to D. Next, let a set of vectors $|x^0,\phi^0,k\rangle$ form the basis in a H^0-invariant linear submanifold of D'. So, under action of $U'(H^0)$,

$$U'(h^0)|x^0,\phi^0,k\rangle = R_k{}^j(h^0)|x^0,\phi^0,j\rangle. \tag{4}$$

Here $h^0 \in H^0$ and $R(H^0)$ is a linear representation of H^0. We define

$$|x,\phi,k\rangle := U'(g_{x\phi})|x^0,\phi^0,k\rangle. \tag{5}$$

Using Eqs. (3)–(5) we obtain the transformation law for the above vectors. First, let us note that the group element $h^0(g,x,\phi)$, defined by

$$h^0(g,x,\phi) = g_{x'\phi'}^{-1}gg_{x\phi} = h_{x'\phi'}^{-1}h(g,x)h_{x\phi}, \tag{6}$$

belongs to H^0. So by the use of the famous Wigner trick we have

$$U'(g)|x,\phi,k\rangle = R_k{}^j[h^0(g,x,\phi)]|x',D(h_{x\phi})\phi(x),j\rangle. \tag{7}$$

Here x' is given by Eq. (1). Now we are ready to give the definition of the generalized coherent states.

Definition: Let $M \subset D'$ be a linear manifold invariant under the action of the group $H^0 \subset H \subset G$, i.e.,

$$U'(H^0)|x^0,\phi^0,k\rangle = R_k{}^j(H^0)|x^0,\phi^0,j\rangle,$$

where $|x^0,\phi^0,k\rangle$ forms a basis in M while $R(H^0)$ is a linear representation of H^0, but not the necessary unitary one. The set of vectors

$$|x,\phi,k\rangle = U'(g_{x\phi})|x^0,\phi^0,k\rangle,$$

where x runs over G/H, forms the system of so-called vectorlike coherent states of the type $[G \supset H \supset H^0, U(G), D(H), R(H^0)]$. Note that for compact $H^0 = H$ and for a one-dimensional $M \subset \mathscr{H}$ this definition coincides with the Perelomov one.

Some remarks are in order. First, the system of vectorlike coherent states is not always overcomplete in D'. A typical example was given in Sec. II: the set $\{|x\rangle\}$ of the position operator eigenstates does not form an overcomplete basis in S'. However, at least for semisimple H^0, the system of vectorlike coherent states is the complete one. The operator

$$E=\int_{G/H} d\mu(x)\int_{H/H^0} d\mu(\phi)\sum_{k,j} \Phi^{kj}|x,\phi,k\rangle\langle x,\phi,j|, \quad (8)$$

where $d\mu(x)$ and $d\mu(\phi)$ are the left invariant measures on G/H and H/H^0, respectively, is an invariant operator if there exists a Φ such that

$$R^\dagger(H^0)\Phi R(H^0)=\Phi. \quad (9)$$

This holds at least for semisimple H^0's. Consequently, irreducibility of U implies, via Schur's lemma, that $E\simeq I$, so completeness holds.

In general, if there exists a Φ satisfying condition (9), the distribution

$$C_{kj}(x,\phi;y,\psi):=\langle x,\phi,k|y,\psi,j\rangle$$

behaves like a generalized reproducing kernel in the space of the distributions $\langle\Psi|x,\phi,k\rangle$.

IV. VECTORLIKE COHERENT STATES FOR THE POINCARÉ GROUP

In this section, we are going to construct a less trivial example, namely, we choose $G=P$ (the Poincaré group), while $H=H^0=L$ (the Lorentz group). In this case the quotient space P/L is simply the Minkowski space-time. Denoting by $g=(\Lambda,a)$ a Poincaré group element, where the Λ are the pseudorotations and the a are the translations, we see that under the identification $\xi(x)=(I,x)$ Eq. (1) takes the form

$$(\Lambda,a)(I,x)=(\Lambda,\Lambda x+a),$$

i.e., we obtain the standard Poincaré group action on the Minkowski space:

$$x'=\Lambda x+a. \quad (10)$$

Now, to specify the coherent states, let us choose the representation $D(L)$ as the trivial one, i.e., $D(L)=I$ while the representation $R(L)$ is assumed to be finite dimensional. The latter one is denoted below as $D(\Lambda)$. Finally, the irreducible unitary representation U of P is taken as $U^{m,s}$, where the mass square $m^2>0$ and the spin s is fixed integer or half-integer. So we have deal with coherent states of the type $[P\supset L,U^{m,s},D(L)]$.

The representation $U^{m,s}$ can be uniquely extended[31] from the Hilbert space $\mathscr{H}$ to the space of tempered distributions S' where it takes the famous Wigner form (we have omitted the superscript prime at U)

$$U(\Lambda,a)|p,\sigma;m,s\rangle = \exp(iap)\sum_{\tau=-s}^{s}\mathscr{D}^s(R_{\Lambda p})_\sigma{}^\tau|\Lambda p,\tau;m,s\rangle. \quad (11)$$

Here $ap=a^\nu p_\nu$, where p_ν is the four-momentum, and $\mathscr{D}^s$ denotes a representation of the SU(2) group labeled by the spin value s, while $R_{\Lambda p}$ is the Wigner rotation.

Now, let us look for an L-invariant subspace M of the S'. To do this let us expand the base vectors $|x^0,k\rangle$, $k=1,2,...,\dim M$, of M on the complete set of the vectors $|p,\sigma;m,s\rangle$:

$$|x^0,k\rangle=\int d\mu(p)\sum_{\sigma=-s}^{s} u_k(p,\sigma)|p,\sigma;m,s\rangle. \quad (12)$$

Here $d\mu(p)=\theta(p)\delta(p^2-m^2)d^4p$. Because of the assumption

$$U(\Lambda,0)|x^0,k\rangle=D_k{}^j(\Lambda)|x^0,j\rangle, \quad (13)$$

we obtain from Eqs. (11) and (12) the consistency condition for the expansion coefficients $u(p)=[u_k(p,\sigma)]$:

$$u(\Lambda p)=D(\Lambda^{-1})u(p)\mathscr{D}^s(R_{\Lambda p}). \quad (14)$$

The set of the vectorlike coherent states is now obtained from a vector $|x^0,k\rangle$ by the action of the unitary operator $U(I,x)$:

$$|x,k\rangle=U(I,x)|x^0,k\rangle = \int d\mu(p)\sum_{\sigma=-s}^{s}\exp(ixp)\times u_k(p,\sigma)|p,\sigma;m,s\rangle. \quad (15)$$

Obviously we can choose $x^0=0$. It is a matter of direct verification that the vectorlike coherent states (15) are simply obtainable from the vacuum state $|0\rangle$ by action of a local field operator, say, $\phi_k(x)$. The coefficients $u_k(p,\sigma)$, satisfying the consistency condition (14), play the role of amplitudes in the Fourier expansion of $\phi_k(x)$. Consequently, the scalar product of two coherent states gives in this case the two-point Wightman function

$$\langle\bar{\phi}_k(x)|\phi_j(y)\rangle = (2\pi)^3\bar{u}(-i\partial_\nu)_k u(-i\partial_\nu)_j\Delta_+(x-y;m), \quad (16)$$

where $\bar{u}$ denotes the Dirac conjugation of u and

$$\Delta_+(x)=(2\pi)^{-3}\int d\mu(p)\exp(-ixp).$$

Concluding, the vectorlike coherent states for the Poincaré group describe a free quantum motion of a relativistic particle with mass m and spin s.

V. HYPERBOLIC COHERENT STATES FOR SU(1,1)

As was claimed in Sec. II, for unitary representations of semisimple groups with a finite center, the set of differentiable vectors is dense and nuclear in the Hilbert space $\mathscr{H}$. The simplest example is given by the special pseudounitary group SU(1,1). Now, we construct for this group the set of coherent states of the type $(\mathrm{SU}(1,1)\supset\mathrm{SO}(1,1), U^{\Phi,E_0}, D[\mathrm{SO}(1,1)])$, namely, we choose $G=\mathrm{SU}(1,1)$, $H=H^0=\mathrm{SO}(1,1)$, $D(H)=I$, $R(H^0)=D_\lambda[\mathrm{SO}(1,1)]$, and $U=U^{\Phi,E_0}$. Here SO(1,1) is the hyperbolic subgroup of SU(1,1) while D_λ is a one-dimensional representation of SO(1,1) labeled by the real number λ. The U^{Φ,E_0} denotes the unitary representation of SU(1,1) fixed by the real numbers Φ and E_0.[32] According to the standard procedure[32] the base vectors in the representation space of SU(1,1) can be chosen as the monomials of two complex variables ζ_ϵ, $\epsilon=1,2$:

$$|\Phi,m\rangle=N_m(\zeta_1\zeta_2)^\Phi\left(\frac{\zeta_1}{\zeta_2}\right)^{E_0+m} \quad (17)$$

Here m is the eigenvalue of the generator of the SO(2) subgroup while the normalization factor $N_m = 1$ for the principal series of representations and $N_m = [\Gamma(m+E_0-\Phi)/\Gamma(m+E_0+\Phi+1)]^{1/2}$ for the other ones.[32] Under the global SU(1,1) transformation, ζ_ϵ behaves as a spinor:

$$\begin{pmatrix}\zeta_1\\ \zeta_2\end{pmatrix}' = \begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix}\begin{pmatrix}\zeta_1\\ \zeta_2\end{pmatrix} \tag{18}$$

with $|\alpha|^2 - |\beta|^2 = 1$. The hyperbolic subgroup SO(1,1) is generated by the differential operator $K = \frac{1}{2}(\xi_2\partial_1 - \zeta_1\partial_2)$, where $\partial_\epsilon = \partial/\partial\zeta_\epsilon$.

Taking into account the homogeneity degree of vectors in an irreducible representation space, we obtain

$$|\Phi,\lambda\rangle = C_\lambda(\zeta_1 + i\zeta_2)^{\Phi+i\lambda}(\zeta_1 - i\zeta_2)^{\Phi-i\lambda} \tag{19}$$

as the solution of the eigenvalue equation $K|\Phi,\lambda\rangle = \lambda|\Phi,\lambda\rangle, \lambda\in\mathbb{R}$. Here C_λ is a normalization factor. Now, we can follow the usual procedure to construct the coherent states. We use the following parametrization of the cosets SU(1,1)/SO(1,1):

$$\mathrm{SU}(1,1)/\mathrm{SO}(1,1) \sim \left\{\begin{pmatrix} z & \omega\\ \omega & \bar z\end{pmatrix}\right\}, \tag{20}$$

where $\omega\in\mathbb{R}$, $z\in\mathbb{C}$, and $|z|^2 - \omega^2 = 1$. Let us begin with the construction of the hyperbolic coherent states $|\Phi,\lambda;z,\omega\rangle$ for the discrete series D^+ of the unitary representations of SU(1,1). In this case, $E_0 = -\Phi > 0$ and $m = 0,1,2,\ldots$. Using Eqs. (18) and (19), we obtain

$$\begin{aligned}|\Phi,\lambda;z,\omega\rangle &= U'(z,\omega)|\Phi,\lambda\rangle\\ &= C_\lambda i^{2i\lambda}(z+i\omega)^{\Phi+i\lambda}(z-i\omega)^{\Phi-i\lambda}\\ &\quad\times\zeta_2^{2\Phi}[1 - i\zeta_1(\bar z - i\omega)/\zeta_2(z-i\omega)]^{\Phi+i\lambda}\\ &\quad\times[1 + i\zeta_1(\bar z + i\omega)/\zeta_2(z-i\omega)]^{\Phi-i\lambda}.\end{aligned} \tag{21}$$

The right-hand side can be expanded with help of the formula[33]

$$(1-s)^{p-q}(1-s+sz)^{-p} = \sum_{n=o}^{\infty} s^n \binom{q-n-1}{n} F(-n,p,q;z),$$

where F is the hypergeometric function. As a result we obtain

$$\begin{aligned}|\Phi,\lambda;z,\omega\rangle &= C_\lambda e^{-\pi\lambda}\Gamma(-2\Phi)^{-1}(z+i\omega)^{\Phi+i\lambda}(z-i\omega)^{\Phi-i\lambda}\sum_{m=0} i^m(\bar z - i\omega)^m(z-i\omega)^{-m}\Gamma(m-2\Phi)^{1/2}\Gamma(m+1)^{-1/2}\\ &\quad\times F(-m,-\Phi+i\lambda,-2\Phi;2[1-i\omega(z+\bar z)]^{-1})|\Phi,m\rangle.\end{aligned} \tag{22}$$

The scalar product of two coherent states reads

$$\langle\Phi,\lambda';z',\omega'|\Phi,\lambda,;z,\omega\rangle = \bar C_{\lambda'}C_\lambda\Gamma(-2\Phi)w^{i(\lambda-\lambda')}x^{\Phi+i\lambda'}y^{\Phi-i\lambda}F(-\Phi-i\lambda',-\Phi+i\lambda,-2\Phi;4(xy)^{-1}), \tag{23}$$

where

$$w = z\bar z' - \bar z z' + i\omega(z'+\bar z') - i\omega'(z+\bar z),$$
$$x = z\bar z' + \bar z z' - 2\omega\omega' - i\omega(z'-\bar z') + i\omega'(z-\bar z),$$
$$y = z\bar z' + \bar z z' - 2\omega\omega' + i\omega(z'-\bar z') - i\omega'(z-\bar z).$$

An analogous procedure for the principal and supplementary series gives

$$\begin{aligned}|\Phi,\lambda;z,\omega\rangle &= C_\lambda i^{\Phi+E_0}[1+i\omega(z+\bar z)]^{\Phi}\left[-\frac{z+i\omega}{z-i\omega}\right]^{i\lambda}\left[\frac{\bar z+i\omega}{z-i\omega}\right]^{E_0}\sum_{\substack{n=0\\m=0}}^{\infty} i^{n-m}\left[\frac{\bar z-i\omega}{z+i\omega}\right]^n\left[\frac{\bar z+i\omega}{z-i\omega}\right]^m\\ &\quad\times\binom{E_0-\Phi+n-1}{n}\binom{\Phi+E_0}{m}N_{n-m}^{-1}F(-n,E_0+i\lambda,E_0-\Phi;2[1-i\omega(z+\bar z)]^{-1})|\Phi,n-m\rangle.\end{aligned} \tag{24}$$

Here[32] for the principal series $\Phi = -\frac{1}{2}+i\chi$, $\chi\in\mathbb{R}_+$, $-\frac{1}{2}<E_0<\frac{1}{2}$, $m = 0, \pm 1,\ldots$, and for the supplementary series $\Phi\in\mathbb{R}$, $-\frac{1}{2}<E_0<\frac{1}{2}$, $|\Phi+\frac{1}{2}|<\frac{1}{2}-|E_0|$, $|m=0, \pm 1, \pm 2,\ldots$.

To our knowledge, the hyperbolic coherent states for SU(1,1) were not investigated in the literature [the elliptic coherent states with the stability group SO(2) are exhaustively discussed in Ref. 9; they were first introduced by Perelomov[4] and Gilmore[34]].

VI. DISCUSSION

As follows from the specific examples we have discussed in Secs. II, IV, and V, the properties of the introduced coherent states strongly depend on the defining groups. This holds because the class of noncompact groups is too rich to give universal properties for the corresponding coherent states. A typical example is given in Sec. II: The position operator eigenvectors $|x\rangle$ are orthonormal and form a complete (not overcomplete) set in S', contrary to the coherent states with a compact stability group. As a consequence, no representation of operators by symbols exists.

A somewhat different situation is the case of the vector-like coherent states for the Poincaré group (Sec. IV). These states are normalized to matrix elements of the projector on the positive energy subspace (with a fixed mass) of the Hilbert space of square-integrable functions with respect to the measure d^4x. Furthermore, this sytem is also complete rather than overcomplete.

Finally, the hyperbolic coherent states of the type SU(1,1)/SO(1,1) (Sec. V) form a nonorthogonal and overcomplete set of vectors [compare Eq. (23)].

Summarizing, some properties universal in the case of compact groups do not necessarily hold for noncompact stability groups. Some possible relations between the group type and properties of the vectorlike coherent states are under investigation.

Now, let us discuss a possible physical application of the introduced states. To do this, let us consider first a simple extension of the example given in Sec. II, i.e., the coherent states for a special solvable group N generated by $\hat{q}$, $\hat{p}$, $\hat{e}$, and $\hat{H}$. The generator $\hat{H}$, interpreted as the harmonic oscillator Hamiltonian, satisfies $[\hat{H},\hat{q}] = -i\hat{p}$, $[\hat{H},\hat{p}] = i\hat{q}$, and $[\hat{H},\hat{e}] = 0$.

We choose the noncompact group generated by $\hat{q}$ and $\hat{e}$ as the stability group H. Thus the stable vectors of H are the position operator eigenvectors $|x\rangle$, i.e., they lie out of the underlying Hilbert space of the unitary representation of N. The set of coherent states has the form

$$|t,x+y\rangle = \exp[it\hat{H}]\exp[iy\hat{p}]|x\rangle$$
$$= \exp[it\hat{H}]|x+y\rangle \equiv |t,q\rangle,$$

i.e., they describe the time evolution of $|q\rangle$ in the Schrödinger picture. Consequently the scalar product of two coherent states $\langle q',t'|q,t\rangle$ is simply the propagation function for the quantum oscillator between two space-time points (q',t') and (q,t).

A quite similar situation arises in the case of the vectorlike coherent states for the Poincaré group (Sec. IV): The coherent states defined by Eq. (15) are simply the relativistic-covariant wave functions of free quantum particles with spin in the coordinate representation. Therefore, their scalar product (16) gives a two-point Wightman function of the corresponding free quantum field theory.

The above examples suggest that in terms of vectorlike coherent states it is possible to describe the quantum motion of particles in homogeneous space-times $M = G/H$. A typical example is the de Sitter space-time SO(4,1)/SO(3,1) and the anti-de Sitter space-time SO(3,2)/SO(3,1) and their supersymmetric counterparts.[35] In that spirit the generalized coherent states can be applied to the Kaluza–Klein cosmologies—our investigations in this direction are in progress.

ACKNOWLEDGMENT

Two of the authors (J. R. and W. T.) were supported by CPBP 01.03 of the Polish Ministry of National Education.

[1]R. J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[2]J. R. Klauder, J. Math. Phys. **4**, 1055 (1963); **6**, 177 (1964).
[3]J. M. Radcliffe, J. Phys. A **4**, 313 (1971).
[4]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[5]B. V. Thompson, J. Phys. A **5**, 1455 (1972).
[6]F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[7]R. Gilmore, Ann. Phys. (NY) **74**, 391 (1972).
[8]J. R. Klauder and B. S. Skagerstam, *Coherent States Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).
[9]A. M. Perelomov, *Generalised Coherent States and Their Applications* (Springer, Berlin, 1986).
[10]K. T. Hecht, *The Vector Coherent State Method and its Application to Problems of Higher Symmetries* (Springer, Berlin, 1986).
[11]A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[12]B. S. Skagerstam, J. Phys. A **18**, 1 (1985).
[13]M. M. Nieto, *Coherent States with Classical Motion; from an Analytic Method Complementary to Group Theory*, in *Group Theoretical Methods in Physics*, Vol. 2, edited by M. A. Markov (Nauka, Moscow, 1983), p. 174.
[14]In the general case the subduction problem $U(H) = U(G)\downarrow H$ corresponds to an eigenvalue problem for Casimir operators of the stability subgroup H (and eventually for other elements from the center of the enveloping algebra of the Lie algebra of H). For noncompact H some of these operators have continuous spectrum so they cannot be diagonalized in the underlying Hilbert space of the considered unitary representation of G. See B. Nagel, in *Studies in Mathematical Physics*, edited by A. O. Barut, (Reidel, Dordrecht, 1973); G. Lindbland and B. Nagel, Ann. Inst. H. Poincaré **13**, 27 (1970).
[15]E. Gutkin, Funk. Anal. Pril. **9**(3), 89 (1975).
6
[16]D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
[17]D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A **17**, L399 (1984).
[18]D. J. Rowe, B. G. Wybourne, and P. H. Butler, J. Phys. A **18**, 939 (1985).
[19]D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).
[20]J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984).
[21]C. Quesne, J. Math. Phys. **27**, 428, 869 (1986).
[22]O. Castaños, E. Chacón, and M. Moshinsky, J. Math. Phys. **25**, 2107 (1984).
[23]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985).
[24]O. Castaños, P. Kramer, and M. Moshinsky, J. Math. Phys. **27**, 924 (1986).
[25]H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, Leipzig, 1928); P. Cartier, in *Proceedings of the Symposium on Pure Mathematics*, Vol. 9 (Am. Math. Soc. Providence, RI, 1966), p. 361.
[26]M. Stone, Proc. Natl. Acad. Sci. USA **16**, 172 (1930).
[27]J. von Neumann, Math. Ann. **104**, 570 (1931).
[28]L. Schwartz, *Theorie des distributions* (Hermann, Paris, 1957); I. M. Gel'fand and A. G. Kostyushenko, Dokl. Akad. Nauk. USSR **103**, 349 (1955).
[29]S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969); G. Callan, S. Coleman, J. Wess, and B. Zumino, *ibid.* **177**, 2247 (1969).
[30]A. Salam and J. Strathdee, Phys. Rev. **184**, 1750 (1969).
[31]H. Joos, Fortschr. Phys. **10**, 65 (1962).
[32]A. O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965).
[33]H. Bateman and A. Erdelyi, *Higher Transcendental Functions*, Vol. I (McGraw-Hill, New York, 1953).
[34]R. Gilmore, J. Math. Phys. **15**, 2090 (1974).
[35]J. Rembielinski and W. Tybor, Acta Phys. Pol. B **15**, 611 (1984); P. Kosinski, J. Rembielinski, and W. Tybor, *ibid.* **16**, 827 (1985).

Uniqueness problems in formal nonlinear representations of the Poincaré group in 2+1 dimensions

G. Rideau[a)]
Université Paris VII and C.N.R.S. UPR 177, Paris, France

(Received 26 April 1989; accepted for publication 31 May 1989)

It is shown that the cocycles of extension of one massless representation of the Poincaré group $\mathscr{P}_3$ in 2 + 1 dimensions by the tensor product of n massless representations are coboundaries when the space of the representation is the space of the C^∞ functions on $\mathbb{R}^2$ rapidly decreasing at infinity and at the origin. As a direct consequence, the equivalence classes of formal nonlinear representations of $\mathscr{P}_3$ with an irreducible physical representation as linear term are isomorphic to the classes of extension of the linear term by its symmetrical tensor product of order 2.

I. INTRODUCTION

It has been shown in a previous paper that the Poincaré group $\mathscr{P}_3$ in 2 + 1 dimensions has nonlinear formal representations as defined in the general theory of Ref. 1. These representations started with massless representations as the initial linear term and were indexed by distribution in $\mathscr{D}'(]0,1[)$ characterizing the classes of extension of the initial representation by its symmetrical tensor product.

Essentially the proof lies in recurrently exhibiting a particular solution verifying some conditions of support of the nth obstructive equation, but we completely discarded the possible contributions of the solutions of the homogeneous equations. These solutions are nothing but the cocycles of extension of the initial linear representation by its symmetrical tensor power of order n. Their contribution will only be effective if they are nontrivial cocycles. For, in the opposite case, we can get rid off of them, at least formally, by a sequence of formal nonlinear transforms, each of which modifies a given term of the formal expansion by a well-determined coboundary (see Appendix B). Since we are only concerned with classes of equivalent representations, we could then speak of the uniqueness of the representations in Ref. 2.

The purpose of this paper is precisely to prove that this situation is the actual one, at least if we conveniently redefine the representations spaces. In so far as the spaces $\mathscr{S}_\eta$ introduced below densely contain the spaces $\mathscr{D}_\eta$ of Ref. 2, we do not find too high the price we have paid.

Strictly speaking, we should again prove the results of Sec. III in Ref. 2. But the reader will convince himself that the cohomological considerations of Ref. 2 are meaningful for the new spaces provided we index the equivalence classes by distributions in $\mathscr{D}'([0,1])$.

Let us begin now by recalling the basic factors about $\mathscr{P}_3$, its massless irreducible representations, and the cohomology of extensions concerned with it.

Let M_3 be the three-dimensional Minkowski space, $\hat{M}_3$ its dual, C_+ the future cone without the origin in $\hat{M}_3$, and ω the point of C_+ with coordinates $(1/2, 0, -1/2)$. With any $k\in C_+$, we associate $\Lambda_k \in \mathrm{SL}(2,\mathbb{R})$ such that

$$k = \Lambda_k^{-1}\cdot\omega, \qquad (1.1)$$

where Λ_k is given by

$$\Lambda_k = \begin{vmatrix} \exp t & 0 \\ 0 & \exp(-t) \end{vmatrix} \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix},$$
$$t\in\mathbb{R}, \quad -\pi/2 < \varphi \leqslant \pi/2. \qquad (1.2)$$

The Minkowskian coordinates of k are written in terms of (t,φ):

$$k_0 = \tfrac{1}{2}\exp(-2t), \quad k_1 = \tfrac{1}{2}\exp(-2t)\sin 2\varphi,$$
$$k_2 = -\tfrac{1}{2}\exp(-2t)\cos 2\varphi. \qquad (1.3)$$

In the following we mark the point k of C_+ by $\mathbf{k} = (k_1 k_2)$ or the parameters (t,φ) indifferently.

We define linear topological spaces $\mathscr{S}_\eta$, $\eta = 0,1$: $f(t,\varphi)\in\mathscr{S}_\eta$ if:

(i) $f(t,\varphi)$ is C^∞ in t,φ;

(ii) $\hat{f}(t,\varphi) = (\mathrm{sgn}\,\varphi)^\eta f(t,\varphi - \pi/2\,\mathrm{sgn}\,\varphi)$ is C^∞ in t,φ;

(iii) for any integer r we have

$$\sup_{\substack{t\in\mathbb{R} \\ -\pi/2\leqslant\varphi\leqslant\pi/2 \\ p+q\leqslant r}} (\mathrm{ch}\, 2t)^r \left| \frac{\partial^{p+q} f}{\partial t^p \partial\varphi^q}(t,\varphi) \right| < \infty;$$

(iv) a sequence $f_n(t,\varphi)$ converges to zero in $\mathscr{S}_\eta$ if, for any integer r

$$\lim_{n\to\infty} \sup_{\substack{t\in\mathbb{R} \\ -\pi/2\leqslant\varphi\leqslant\pi/2 \\ p+q\leqslant r}} (\mathrm{ch}\, 2t)^r \left| \frac{\partial^{p+q} f_n}{\partial t^p \partial\varphi^q}(t,\varphi) \right| = 0.$$

$\mathscr{S}_0$ can be identified with the space of C^∞ functions in $\mathbb{R}^2$ rapidly decreasing at infinity and at the origin.

$\mathscr{S}_1$ is isomorphic to $\mathscr{S}_0$: if $f\in\mathscr{S}_1$, then $\exp(i\varphi)f\in\mathscr{S}_0$. Consequently, if $T\in\mathscr{S}_1'$, $\exp(-i\varphi)T\in\mathscr{S}_0'\cdot\mathscr{S}_0(\mathscr{S}_1)$ is obviously nuclear.

$\mathscr{S}_\eta$ are $K(M_p)$ spaces, with properties (P) and (N) as defined in Ref. 3.

Let V_η be the representation of $\mathscr{P}_3$ in $\mathscr{S}_\eta$ defined by

$$f(\mathbf{k}) \to \exp i\langle a,k\rangle \epsilon^\eta(\Lambda,k) f(\Lambda^{-1}\mathbf{k}),$$
$$a\in\mathbb{R}^3 \quad \Lambda\in\mathrm{SL}(2,\mathbb{R}) \qquad (1.4)$$

where $\langle a,k\rangle$ is the Minkowskian scalar product and

$$\epsilon(\Lambda,k) = \mathrm{sgn}(\delta - \beta\, tg\varphi), \quad \Lambda = \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} \in \mathrm{SL}(2,\mathbb{R}). \qquad (1.5)$$

[a)] Postal address: Université Paris VII-L.P.T.M., Tour Centrale-3ème étage, 2, place Jussieu, 75251 Paris Cedex 05, France.

 0022-2488/89/102411-08$02.50

Here, V_η is the restriction to $\mathscr{S}_\eta$ of the unitary massless representation of $\mathscr{P}_3$ with helicity $\eta/2$.

We are looking for mappings $Z(g) \in \mathscr{L}(\mathscr{S}_{\eta_1} \otimes \cdots \otimes \mathscr{S}_{\eta_n}, \mathscr{S}_\eta)$, $g \in \mathscr{P}_3$ satisfying the following cohomological equation:

$$Z(gg') = Z(g) + V_\eta(g) Z(g') V^{-1}_{\eta_1 \cdots \eta_n}(g), \quad g,g' \in \mathscr{P}_3, \tag{1.6}$$

where $V_{\eta_1 \cdots \eta_n}(g)$ stands for $V_{\eta_1}(g) \otimes \cdots \otimes V_{\eta_n}(g)$. Similarly we will use $\mathscr{S}_{\eta_1 \cdots \eta_n}$ instead of $\mathscr{S}_{\eta_1} \otimes \cdots \otimes \mathscr{S}_{\eta_n}$. Here, $Z(g)$ defines an extension of $V_\eta(g)$ by $V_{\eta_1 \cdots \eta_n}(g)$. We are interested only in nontrivial extensions and we will identify two solutions of (1.6) when their difference is a coboundary, i.e., a particular solution of (1.6) written:

$$Z(g) = A - V_\eta(g) A V^{-1}_{\eta_1 \cdots \eta_n}(g), \quad A \in \mathscr{L}(\mathscr{S}_{\eta_1 \cdots \eta_n}, \mathscr{S}_\eta).$$

Accordingly, we do not restrict the generality in assuming $Z(g)$ equal to zero on SO(2), the compact subgroup of SL(2,$\mathbb{R}$).

Lemma 1.1: Under this condition, the solutions of (1.6) are identically zero when $\eta + \eta_1 + \cdots + \eta_n = 0$, mod 2.

Proof: $(0,-I) \in \mathrm{SO}(2)$ and $(0,-I) g (0,-I) = g$. A repeated use of (1.6) gives

$$Z(g) = V_\eta(0,-I) Z(g) V_{\eta_1 \cdots \eta_n}(0,-I) = (-1)^{\eta + \eta_1 + \cdots + \eta_n} Z(g).$$

In the following, we shall find it convenient to associate with $Z(g)$ the distribution $Z_{\mathbf{k}}(g) \in \mathscr{S}'_{\eta_1 \cdots \eta_n}$, defined by

$$(Z_{\mathbf{k}}(g), f) = (Z(g) f)(\mathbf{k}), \quad f \in \mathscr{S}_{\eta_1 \cdots \eta_n}. \tag{1.7}$$

The cohomological equation (1.6) now reads

$$Z_{\mathbf{k}}((a,\Lambda)(a',\Lambda')) = Z_{\mathbf{k}}(a,\Lambda) + \exp i\langle a,k\rangle \epsilon^\eta(\Lambda,k) \times Z_{\Lambda^{-1}\mathbf{k}}(a',\Lambda') V^{-1}_{\eta_1 \cdots \eta_n}(a,\Lambda). \tag{1.8}$$

Let Γ_ω be the stabilizer of ω in SL(2,$\mathbb{R}$); then $h_k(\Lambda) = \Lambda_k \Lambda \Lambda^{-1}_{\Lambda^{-1}k}$ belongs to Γ_ω, and we have

$$(0,\Lambda_k)(a,\Lambda) = (\Lambda_k a, h_k(\Lambda))(0,\Lambda_{\Lambda^{-1}k}).$$

Applying (1.8) to this identity, we get the following expression of $Z_{\mathbf{k}}(a,\Lambda)$:

$$\begin{aligned} Z_{\mathbf{k}}(a,\Lambda) &= Z_\omega(\Lambda_k a, h_k(\Lambda)) V_{\eta_1 \cdots \eta_n}(0,\Lambda_k) \\ &\quad - Z_\omega(0,\Lambda_k) V_{\eta_1 \cdots \eta_n}(0,\Lambda_k) \\ &\quad + \exp i\langle a,k\rangle \epsilon^\eta(\Lambda,k) Z_\omega(0,\Lambda_{\Lambda^{-1}k}) \\ &\quad \times V_{\eta_1 \cdots \eta_n}(0,\Lambda_{\Lambda^{-1}k}) V^{-1}_{\eta_1 \cdots \eta_n}(a,\Lambda). \end{aligned} \tag{1.9}$$

This formula brings back the solution of (1.6) to the determination of $Z_\omega(a,\Lambda)$ for $a \in \mathbb{R}^3$, $\Lambda \in \Gamma_\omega$. According to Lemma 1.1, we are concerned only with $\eta + \eta_1 + \cdots + \eta_n = 0$, mod 2.

II. TRIVIALITY OF $Z(a,I)$

Lemma 2.1: There exists $T \in \mathscr{S}_{\eta_1 \cdots \eta_n}$ such that

$$Z_\omega(a,I) = T - \exp i\langle a,\omega\rangle T V^{-1}_{\eta_1 \cdots \eta_n}(a,I). \tag{2.1}$$

Proof: It is sufficient to prove (2.1) for $\eta_i = 0$, $i = 1,\dots,n$. Indeed $Z_\omega(a,I) \in \mathscr{S}'_{\eta_1 \cdots \eta_n}$ and therefore $Z_\omega(a,I) \exp(-i\Sigma_1^n \eta_i \varphi_i) \in \mathscr{S}'_{0 \cdots 0}$. Furthermore, $V_{\eta_1 \cdots \eta_n}(a,I)$ does not depend on the η_i's and commutes with the multiplication by $\exp(-i\Sigma_1^n \eta_i \varphi_i)$.

Let us introduce

$$\zeta_\mu = \frac{\partial}{\partial a_\mu} Z_\omega(a,I)|_{a=0}, \quad \mu = 0,1,2.$$

The abelianness of the translations and Eq. (1.6) imply

$$\left(\omega_\mu - \sum_1^n k_{i,\mu}\right)\zeta_\nu = \left(\omega_\nu - \sum_1^n k_{i,\nu}\right)\zeta_\mu, \quad \mu,\nu = 0,1,2. \tag{2.2}$$

If (2.1) is true, we have

$$\left(\omega_\mu - \sum_1^n k_{i,\mu}\right) T = i\zeta_\mu. \tag{2.3}$$

Let us multiply the equation for $\mu = 0$ by a function f written as

$$f = P_0 + \Sigma |k_i| P_i + \sum_{i<j} |k_i||k_j| P_{ij} + \cdots + \prod_{i=1}^n |k_i| R,$$

where $P_0, P_i, P_{ij}, \dots, R$ are polynomials in the components of $\mathbf{k}_1 \cdots \mathbf{k}_n$, determined by requiring that

$$f\left(1 - \sum_1^n |k_i|\right)$$

is also a polynomial in the components of $\mathbf{k}_1,\dots,\mathbf{k}_n$. It is easy to see that there exists such an f: the various polynomials verify a homogeneous linear system which contains one equation less than the number of unknowns.

The proof given in Ref. 4 of the divisibility of distributions by polynomials can be extended to the distributions of $\mathscr{S}'_{0 \cdots 0}$ with minor modifications. Therefore we can proceed essentially as in Ref. 5 for proving (2.1).

Proposition 2.1: Each equivalence class of solutions of (1.6) contains a cocycle equal to zero on the semidirect product $\mathbb{R}^3 \cdot \mathrm{SO}(2)$.

Proof: We get from (1.9)

$$Z_{\mathbf{k}}(a,I) = A_{\mathbf{k}} - \exp i\langle a,k\rangle A_{\mathbf{k}} V^{-1}_{\eta_1 \cdots \eta_n}(a,I),$$

$$A_{\mathbf{k}} = (T - Z_\omega(0,\Lambda_k)) V_{\eta_1 \cdots \eta_n}(0,\Lambda_k). \tag{2.4}$$

We shall prove that $(A_{\mathbf{k}}, f)$ is in $\mathscr{S}_\eta$ for any $f \in \mathscr{S}_{\eta_1 \cdots \eta_n}$. This will be done in several steps, which adjust to the present topological spaces the method of Ref. 5.

(1) $(A_{\mathbf{k}}, f)$ is C^∞ in t,φ. For, on the one hand,

$$Z_\omega(0,\Lambda_k) = Z_\omega\left(0, \begin{vmatrix} \exp t & 0 \\ 0 & \exp(-t) \end{vmatrix}\right),$$

and on the other hand, the elements of $\mathscr{S}_{\eta_1 \cdots \eta_n}$ are, by construction, differentiable vectors of the representation $V_{\eta_1 \cdots \eta_n}$.

(2) $\widehat{(A_{\mathbf{k}}, f)}$ is C^∞ in t,φ. Indeed,

$$V_{\eta_1 \cdots \eta_n}\left(0, \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix}\right)$$

is replaced by

$$\begin{aligned} &(\operatorname{sgn}\varphi)^\eta V_{\eta_1 \cdots \eta_n}\left(0, \operatorname{sgn}\varphi \begin{vmatrix} \sin\varphi & -\cos\varphi \\ \cos\varphi & \sin\varphi \end{vmatrix}\right) \\ &\quad = V_{\eta_1 \cdots \eta_n}\left(0, \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix}\right)(\operatorname{sgn}\varphi)^\eta \\ &\quad \times V_{\eta_1 \cdots \eta_n}\left(0, \begin{vmatrix} 0 & -\operatorname{sgn}\varphi \\ \operatorname{sgn}\varphi & 0 \end{vmatrix}\right). \end{aligned}$$

By the action of

$$V_{\eta_1\cdots\eta_n}\left(0,\begin{vmatrix} 0 & -\operatorname{sgn}\varphi \\ \operatorname{sgn}\varphi & 0 \end{vmatrix}\right),$$

each f in $\mathscr{S}_{\eta_1\cdots\eta_n}$ becomes $(\operatorname{sgn}\varphi)^{\eta_1\cdots\eta_n}f=(\operatorname{sgn}\varphi)^{\eta}\hat{f}$. Therefore

$$(A_{\mathbf{k}},f)=(A_{\mathbf{k}},\hat{f})$$

and we conclude as above.

(3) Let X be an infinitesimal generator of $\mathscr{P}_3$. Then, for each semi-norm q on $\mathscr{S}_\eta$, we can find a semi-norm p such that

$$q\left(\frac{d^r}{du^r}V_\eta(\exp uX)f\right)\leqslant p(f)\exp c_r^1|u|\,,\quad f\in\mathscr{S}_\eta$$

for some positive constant c_r^1. This is true for $r=0$ (cf. Ref. 5). We prove it for any r recurrently from the identity:

$$\frac{d^r}{du^r}V_\eta(\exp uX)f=\frac{d^{r-1}}{du^{r-1}}V_\eta(\exp uX)dV_\eta(X)f,$$

where $dV_\eta(X)$ is a continuous operator on $\mathscr{S}_\eta$.

Similarly, for each seminorm q on $\mathscr{S}_{\eta_1\cdots\eta_n}$, we can find a seminorm p such that

$$q\left(\frac{d^r}{du^r}V_{\eta_1\cdots\eta_n}(\exp uX)f\right)\leqslant p(f)\exp C_r^n|u|,\quad f\in\mathscr{S}_{\eta_1\cdots\eta_n}$$

for some positive constant C_r^n.

(4) Let ζ_X be $(d/du)Z(\exp uX)|_{u=0}$. From the cohomological equation, we get

$$\frac{d^r}{du^r}Z(\exp uX)=V_\eta(\exp uX)\zeta_X^{(r)}V^{-1}_{\eta_1\cdots\eta_n}(\exp uX)\,,$$

where $\zeta_X^{(r)}$ is a continuous mapping of $\mathscr{S}_{\eta_1\cdots\eta_n}$ into $\mathscr{S}_\eta$ defined recurrently by

$$\zeta_X^{(r)}=dV_\eta(X)\zeta_X^{(r-1)}dV_{\eta_1\cdots\eta_n}(X)\,,\quad \zeta_X^{(1)}=\zeta_X\,. \tag{2.5}$$

Proceeding as in Ref. 5, we prove that for each seminorm p in $\mathscr{S}_\eta$, we can find a seminorm q in $\mathscr{S}_{\eta_1\cdots\eta_n}$ such that

$$p\left(\frac{d^r}{du^r}Z(\exp uX)f\right)<q(f)\exp c_r|u|$$

for some positive constant c_r.

(5) Let X be the infinitesimal generator of

$$\begin{vmatrix}\exp t & 0\\ 0 & \exp(-t)\end{vmatrix}$$

and choose $p(f)=|f\omega)|$. With $t=-\frac{1}{2}\log 2|\mathbf{k}|$, we get from the above inequalities

$$\left|\left(\frac{d^r}{dt^r}\left[Z_\omega\left(0,\begin{vmatrix}\exp t & 0\\ 0 & \exp(-t)\end{vmatrix}\right)\times V_{\eta_1\cdots\eta_n}\left(0,\begin{vmatrix}\exp t & 0\\ 0 & \exp(-t)\end{vmatrix}\right)\right],f\right)\right|\leqslant q_r(f)\exp\gamma_r\left|\log|\mathbf{k}|\right|$$

for some semi-norm q_r in $\mathscr{S}_{\eta_1\cdots\eta_n}$ and some positive constant γ_r. As $T\in\mathscr{S}'_{\eta_1\cdots\eta_n}$, we have similarly

$$\left|\left(T\frac{d^r}{dt^r}V_{\eta_1\cdots\eta_n}\left(0,\begin{vmatrix}\exp t & 0\\ 0 & \exp(-t)\end{vmatrix}\right),f\right)\right|\leqslant q_r(f)\exp\gamma_r\left|\log|\mathbf{k}|\right|\,.$$

Using the estimate of (3) when X is the infinitesimal generator of the compact subgroup SO(2), we deduce that for each seminorm p on $\mathscr{S}_{\eta_1\cdots\eta_n}$, we can find a seminorm q_s (on $\mathscr{S}_{\eta_1\cdots\eta_n}$) such that

$$p\left(\frac{d^s}{d\varphi^s}V_{\eta_1\cdots\eta_n}\left(0,\begin{vmatrix}\cos\varphi & \sin\varphi\\ -\sin\varphi & \cos\varphi\end{vmatrix}\right),f\right)\leqslant\alpha_s q_s(f)$$

for some positive constant α_s.

Piecing together all these estimates, we get finally

$$\left|\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}}\,,f\right)\right|\leqslant q_{r,s}(f)\exp\gamma_{r,s}\left|\log|\mathbf{k}|\right| \tag{2.6}$$

for some positive constant $\gamma_{r,s}$.

(6) Let P_0 be the generator of the time translations. From (2.4) we get

$$i|\mathbf{k}|A_{\mathbf{k}}-A_{\mathbf{k}}\,dV_{\eta_1\cdots\eta_n}(P_0)=\zeta_{\mathbf{k}}(P_0)\,.$$

This implies, with $|\mathbf{k}|=\frac{1}{2}\exp(-2t)$ that

$$\frac{i}{2}\exp(-2t)\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}}-\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}}\,dV_{\eta_1\cdots\eta_n}(P_0)=\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}\zeta_{\mathbf{k}}(P_0)-\frac{1}{2}\exp(-2t)\sum_{q=1}^{r}C_q^r(-2)^q\frac{\partial^{r-q,s}}{\partial t^{r-q}\partial\varphi^s}A_{\mathbf{k}}\,. \tag{2.7}$$

Let us assume it has been already shown that for all $r'<r$, $((\partial^{r'+s}/\partial t^{r'}\partial\varphi^s)A_{\mathbf{k}},f)$ goes to zero faster than any power of $1/|\mathbf{k}|$ when $|\mathbf{k}|\to\infty$ or any power of $|\mathbf{k}|$ when $|\mathbf{k}|$ goes to zero. We will prove the same is true for $((\partial^{r+s}/\partial t^r\partial\varphi^s)A_{\mathbf{k}},f)$. Let us denote by $\xi_{\mathbf{k}}^{(r,s)}$ the right-hand side of (2.7). By successive iterations, we get on the one hand,

$$\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}},f\right)=\sum_{p=0}^{m-1}(i|\mathbf{k}|)^{p-m}(\xi_{\mathbf{k}}^{(r,s)}\,dV_{\eta_1\cdots\eta_n}(P_0)^{m-p-1},f)+\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}}\,dV^m_{\eta_1\cdots\eta_n}(P_0),f\right)(i|\mathbf{k}|)^{-m} \tag{2.8}$$

and, on the other hand,

$$\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}},f\right)=(i|\mathbf{k}|)^m\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}A_{\mathbf{k}}\,dV^{-m}_{\eta_1\cdots\eta_n}(P_0),f\right)-\sum_0^{m-1}(i|\mathbf{k}|)^p(\xi_{\mathbf{k}}^{r,s}\,dV^{-p-1}_{\eta_1\cdots\eta_n}(P_0),f) \tag{2.9}$$

for any integer m.

By definition of $\zeta_{\mathbf{k}}(P_0)$,

$$\left(\frac{\partial^{r+s}}{\partial t^r\partial\varphi^s}\zeta_{\mathbf{k}}(P_0),f\right)\in\mathscr{S}_\eta.$$

The mappings

$$f \to dV^{\pm p}_{\eta_1 \cdots \eta_n}(P_0) f, \quad f \in \mathscr{S}_{\eta_1 \cdots \eta_n}$$

are continuous mappings of $\mathscr{S}_{\eta_1 \cdots \eta_n}$ into itself. According to (2.6) $((\partial^{r+s}/\partial t^r \partial\varphi^s) A_{\mathbf{k}}, f)$ behaves at most as $|\mathbf{k}|^{\gamma_{r,s}}$ when $|\mathbf{k}| \to \infty$ and $|\mathbf{k}|^{-\gamma_{r,s}}$ for $|\mathbf{k}| \to 0$. Combining all that with the recurrence hypothesis, we get the conclusion we are looking for.

(7) The continuity of the mapping $f \to (A_{\mathbf{k}}, f)$ derives from (2.8) and (2.9). Indeed, we have the estimates:

$$|\mathbf{k}|^p \left| \left(\frac{\partial^{r+s}}{\partial t^r \partial \varphi^s} A_{\mathbf{k}}, f \right) \right|$$
$$\leqslant |\mathbf{k}|^{p+\gamma-m} q_m(f) + \sum_{p=0}^{m-1} |\mathbf{k}|^{l-m+p}$$
$$\times |(\xi_{\mathbf{k}}^{(r,s)} dV^{m-p-1}_{\eta_1 \cdots \eta_n}(P_0), f)|, \quad |\mathbf{k}| \gg 1,$$
$$|\mathbf{k}|^{-p} \left| \left(\frac{\partial^{r+s}}{\partial t^r \partial \varphi^s} A_{\mathbf{k}}, f \right) \right|$$
$$\leqslant |\mathbf{k}|^{m-p+\gamma} q_m(f) + \sum_{l=0}^{m-1} |\mathbf{k}|^{l-p}$$
$$\times |(\xi_{\mathbf{k}}^{(r,s)} dV^{-p-1}_{\eta_1 \cdots \eta_n}(P_0), f)|, \quad |\mathbf{k}| \ll 1,$$

where $\xi_{\mathbf{k}}^{r,s}$ contains only the derivatives $(\partial^{r'+s} A_{\mathbf{k}}/\partial t^{r'} \partial\varphi^s, f)$ for $r' < r$. Therefore we can get the proof by recurrence.

Finally, we have still to prove the cancellation on SO(2) of the coboundary generated by $A_{\mathbf{k}}$, which is equivalent to our statement: after substration of this coboundary from $Z_{\mathbf{k}}(a,\Lambda)$ we get an equivalent cocycle identically zero on $\mathbb{R}^3 \cdot \mathrm{SO}(2)$. Now we have, for

$$u = \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix}:$$

$$\epsilon^\eta(k,u) A_{\mathbf{u}^{-1}\cdot\mathbf{k}} V^{-1}_{\eta_1 \cdots \eta_n}(0,u)$$
$$= (\mathrm{sgn}\cos(\varphi+\theta))^\eta \left(T - Z_\omega \left(0, \begin{vmatrix} \exp t & 0 \\ 0 & \exp(-t) \end{vmatrix} \right) \right.$$
$$\left. \times V_{\eta_1 \cdots \eta_n} \left(0, \begin{vmatrix} \exp t & 0 \\ 0 & \exp(-t) \end{vmatrix} \right) \right)$$
$$\times V_{\eta_1 \cdots \eta_n} \left(0, \begin{vmatrix} \cos(\omega_\varphi - \theta) & \sin(\omega_\varphi - \theta) \\ -\sin(\omega_\varphi - \theta) & \cos(\omega_\varphi - \theta) \end{vmatrix} \right), \tag{2.10}$$

where ω_φ is given by

$\omega_\varphi = \varphi + \theta, \quad -\varphi - \pi/2 < \theta < \pi/2 - \varphi, \ (\cos(\varphi+\theta) > 0),$

$\omega_\varphi = \varphi + \theta - \pi, \quad \pi > \theta > \pi/2 - \varphi, \ (\cos(\varphi+\theta) < 0),$

$\omega_\varphi = \varphi + \theta + \pi, \quad -\pi < \theta < -\varphi - \pi/2, \ (\cos(\varphi+\theta) < 0).$

Therefore, the last term written in (2.10) is equal to

$$V_{\eta_1 \cdots \eta_n} \left(0, \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix} \right)$$

for $|\varphi + \theta| < \pi/2$ and to

$$(-1)^{\eta_1 + \cdots + \eta_n} V_{\eta_1 \cdots \eta_n} \left(0, \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix} \right)$$

for $|\varphi + \theta| > \pi/2$. But in this case the factor $(-1)^{\eta_1 + \cdots + \eta_n}$ is compensated by the factor $(\mathrm{sgn}\cos(\varphi+\theta))^\eta$. Thus we have

$$\epsilon^\eta(k,u) A_{\mathbf{u}^{-1}\cdot\mathbf{k}} V^{-1}_{\eta_1 \cdots \eta_n}(0,u) = A_{\mathbf{k}}.$$

From now on, we restrict ourselves to the consideration of the only cocycles identically zero on $\mathbb{R}^3 \cdot \mathrm{SO}(2)$. According to what we have just proved, the right-hand side of (1.9) is then reduced to its first term.

III. DETERMINATION OF $Z_\omega(0,h), h \in \Gamma_\omega$

We recall the two basic assumptions on the cocycle $Z(a,\Lambda)$: (i) $Z(a,\Lambda)$ is identically zero on $\mathbb{R}^3 \cdot \mathrm{SO}(2)$; (ii) $\eta + \eta_1 + \cdots + \eta_n = 0$, mod 2, for otherwize $Z(a,\Lambda)$ is identically zero. Accordingly, we will have for $a \in \mathbb{R}^3$, $h = \epsilon \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix}$, $\epsilon = \pm 1$, $x \in \mathbb{R}$:

$$Z_\omega(a,h) = Z_\omega \left(0, \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix} \right).$$

From the cohomological equation, we get

$$\frac{d}{dx} \left(Z_\omega \left(0, \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix} \right), f \right)$$
$$= \left(S, V_{\eta_1 \cdots \eta_n} \left(0, \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix} \right) f \right), \quad f \in \mathscr{S}_{\eta_1 \cdots \eta_n},$$

where the distribution S is given by

$$S(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \frac{d}{dx} Z_\omega \left(0, \begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix}; \mathbf{k}_1,\ldots,\mathbf{k}_n \right) \Big|_{x=0}$$

and verifies the identity

$$\left(1 - \exp i \left(a, \omega - \sum_1^n k_i \right) \right) S(\mathbf{k}_1,\ldots,\mathbf{k}_n) = 0. \tag{3.1}$$

The cocycle $Z_\omega(a,h)$ will be a coboundary if we can find $T \in \mathscr{S}'_{\eta_1,\ldots,\eta_n}$ such that T verifies (3.1) and

$$(T, dV_{\eta_1,\ldots,\eta_n}(X) f) = (S, f), \tag{3.2}$$

where X is the infinitesimal generator of

$$\begin{vmatrix} 1 & x \\ 0 & 1 \end{vmatrix}.$$

Following (3.1), the supports of S and T are contained in the manifold $\varphi_i = 0$, $i = 1,\ldots,n$. Thus, it is not restrictive to assume that the test functions are identically zero (with all their derivatives) outside some definite interval J of variations of each φ_i, strictly contained in $]-\pi/2, +\pi/2[$. Then we perform the following change of variables (with absorption of the Jacobian into the distributions):

$$x_i = \exp(-t_i)\sin\varphi_i, \quad r_i = \exp(-t_i)\cos\varphi_i,$$
$$i = 1,\ldots,n, \quad \varphi_i \in J \subset]-\pi/2, +\pi/2[, \quad t_i \in \mathbb{R}$$

and we replace each $\mathscr{S}_{\eta_i}$ by $\mathscr{S}^J_{\eta_i}$, with the following definition:

$$f(r_i, x_i) \in \mathscr{S}^J_{\eta_i}$$

when (i) $f(r_i,x_i)$ is C^∞ in r_i, x_i, and identically zero for $|x_i/r_i| > C_J$

$$\text{(ii)}\quad \sup_{\substack{r_i,x_i\\ u+v<p}} (r_i^2+x_i^2+(r_i^2+x_i^2)^{-1})^p \left|\frac{\partial^{u+v}f}{\partial x_i^u\,\partial r_i^v}\right| < \infty. \tag{3.3}$$

Here, $\mathscr{S}^J_{\eta_i}$ is also a $K(M_p)$ space with properties (P) and (N). In the following we use the shortened notation $\mathscr{S}^J_{\eta_1\cdots\eta_n}$ for the tensor product $\mathscr{S}^J_{\eta_1}\otimes\cdots\otimes\mathscr{S}^J_{\eta_n}$.

In the new variables, (3.1) and (3.2) now read

$$\sum_1^n x_i\frac{\partial T}{\partial r_i}=S,\quad T,S\in(\mathscr{S}^J_{\eta_1\cdots\eta_n})', \tag{3.4}$$

where T is bound to verify, as S itself:

$$\left(1-\sum_1^n r_i^2\right)T=0,\quad \left(\sum_1^n x_i^2\right)T=0,$$
$$\left(\sum_1^n x_i r_i\right)T=0. \tag{3.5}$$

We introduce spherical coordinates:

$$r_n=R\cos\theta_{n-1},$$
$$r_i=R\cos\theta_{i-1}\sin\theta_i\cdots\sin\theta_{n-1},$$
$$i=2,\dots,n-1,\quad 0\leqslant\theta_i\leqslant\pi/2$$
$$r_1=R\sin\theta_1\cdots\sin\theta_{n-1}$$

and we put

$$T=\delta(1-R)t,\quad S=\delta(1-R)s.$$

Then we have from (3.5)

$$\left(\sum_1^n x_i^2\right)t=0,\quad (x_n\cos\theta_{n-1}+y_{n-1}\sin\theta_{n-1})t=0, \tag{3.6}$$

and similarly for s. As for (3.4), it now reads

$$\left[(-x_n\sin\theta_{n-1}+y_{n-1}\cos\theta_{n-1})\frac{\partial}{\partial\theta_{n-1}}+\sum_1^{n-2}\left(\frac{y_i}{\sin\theta_{i+1}\cdots\sin\theta_{n-1}}\right)\frac{\partial}{\partial\theta_i}\right]t=s, \tag{3.7}$$

where the y_i, $i=1,\dots,n-1$, are orthogonal linear combinations of $x_1,\dots,x_{n-1}$:

$$y_1=x_1\cos\theta_1-x_2\sin\theta_1,$$
$$y_i=x_1\sin\theta_1\cdots\sin\theta_{i-1}\cos\theta_i-\sum_2^{i-1}x_j\cos\theta_{j-1}\times\sin\theta_j\cdots\sin\theta_{i-1}\cos\theta_i+x_i\cos\theta_{i-1}\cos\theta_i-x_{i+1}\sin\theta_i,$$
$$y_{n-1}=x_1\sin\theta_1\cdots\sin\theta_{n-2}+\sum_2^{n-2}x_j\cos\theta_{j-1}\times\sin\theta_j\cdots\sin\theta_{n-2}+x_{n-1}\cos\theta_{n-2}. \tag{3.8}$$

We denote by S_n the set of equations (3.6), (3.7). We prove now the following proposition.

Proposition 3.1: For any given s verifying (3.6), we can find a solution t of (3.7) verifying (3.6).

Proof: According to the second equation (3.5) and following the general structure of distributions over $K(M_p)$ spaces with properties (P) and (N) (Ref. 3), we have the finite expansion:

$$t=\sum t_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})\delta^{(p_1)}(x_1)\cdots\delta^{(p_n)}(x_n), \tag{3.9}$$

where the "coefficients" $t_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})$ are distributions in $\mathscr{D}'([0,\pi/2]^{n-1})$. Indeed, following (3.3) $t_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})$ is defined on a space of C^∞ functions on $[0,\pi/2]^{n-1}$ going to zero faster than any power of $\theta_i,\pi/2-\theta_i$ as θ_i goes to zero or $\pi/2$. But this space is easily identified to the space of C^∞ functions on $\mathbb{R}^{n-1}$ identically zero outside the hypercube $[0,\pi/2]^{n-1}$. Similarly, we have

$$s=\sum s_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})\theta^{(p_1)}(x_1)\cdots\delta^{(p_n)}(x_n),$$
$$s_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})\in\mathscr{D}'([0,\pi/2]^{n-1}). \tag{3.10}$$

Substituting (3.9) and (3.10) in (3.6) and (3.7) we get a system equivalent to S_n and relating the unknowns $t_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})$ to the data $s_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})$.

The key of the proof of our proposition will be provided by the following recurrence hypothesis.

Recurrence hypothesis: The proposition is true for all systems S_p, with $p=3,\dots,n-1$.

We prove it now for $p=n$.

(1) Let N be such that $s_{p_1\cdots p_n}(\theta_1,\dots,\theta_{n-1})$ is zero when $p_n>N$. We claim there exists a solution of S_n with the same property.

It will be convenient to write (3.9), (3.10) in the shortened form:

$$t=\sum_{p\leqslant N}t_p(x_1,\dots,x_{n-1};\theta_1,\dots,\theta_{n-1})\delta^{(p)}(x_n),$$
$$s=\sum_{p\leqslant N}s_p(x_1,\dots,x_{n-1};\theta_1,\dots,\theta_{n-1})\delta^{(p)}(x_n). \tag{3.11}$$

We get from (3.6)

$$(p+1)(p+2)t_{p+2}+\left(\sum_1^{n-1}x_i^2\right)t_p=0,$$
$$-(p+1)\cos\theta_{n-1}t_{p+1}+\sin\theta_{n-1}y_{n-1}t_p=0, \tag{3.12}$$

and from (3.7) after using the last equation:

$$(p+1)\left(\frac{\partial}{\partial\theta_{n-1}}-\cot\theta_{n-1}\right)t_{p+1}+\left(y_{n-2}\frac{\partial}{\partial\theta_{n-2}}+\sum_1^{n-3}\frac{y_i}{\sin\theta_{i+1}\dots\sin\theta_{n-2}}\frac{\partial}{\partial\theta_i}\right)\times t_p=\sin\theta_{n-1}s_p. \tag{3.13}$$

Accordingly, we get for t_N

$$\left(\sum_1^{n-1}x_i^2\right)t_N=0,\quad y_{n-1}t_N=0,$$
$$\left(y_{n-2}\frac{\partial}{\partial\theta_{n-2}}+\sum_1^{n-3}\frac{y_i}{\sin\theta_{i+1}\dots\sin\theta_{n-2}}\frac{\partial}{\partial\theta_i}\right)t_N=\sin\theta_{n-1}s_N. \tag{3.14}$$

As s_p verifies (3.12) also, we have necessarily

$$\left(\sum_1^{n-1}x_i^2\right)s_N=0,\quad y_{n-1}s_N=0.$$

Going back to (3.8), we conclude that s_N, t_N satisfy the same set of equations s,t up to the replacement of n by $n-1$.

According to the recurrence hypothesis, we can find t_N.

(2) The last equation (3.12) gives

$$t_{N-k} = \frac{N!}{(N-k!)} \left(\frac{\cos\theta_{n-1}}{\sin\theta_{n-1}}\right)^k \times \tilde{t}_{N,k} + t^0_{N-k}, \quad k=1,...,N, \tag{3.15}$$

where $\tilde{t}_{N,k}$ is a distribution verifying

$$y^k_{n-1}\tilde{t}_{N,k} = t_N. \tag{3.16}$$

We need the following lemma.

Lemma 3.1: The distributions $\tilde{t}_{N,k}$ in (3.15) can be defined in such a way that the distributions t^0_p verify (3.12).

Proof: The distributions t^0_p verify (3.12) if we have

$$y_{n-1}\tilde{t}_{N,k} = \tilde{t}_{N,k-1}, \tag{3.17}$$

$$\tilde{t}_{N,k-2} + \left(\sum_1^{n-1} x_i^2\right)\frac{\cos^2\theta_{n-1}}{\sin^2\theta_{n-1}}\tilde{t}_{N,k} = 0. \tag{3.18}$$

With $t_N = \delta(y_{n-1})u_N$, we take $\tilde{t}_{N,k}$ in the form:

$$\tilde{t}_{N,k} = (-1)^k/k!\delta^{(k)}(y_{n-1})u_N + (-1)^k \times \sum_1^{[k/2]} v_l\delta^{(k-2l)}(y_{n-1})/(k-2l)!; \tag{3.19}$$

(3.17) is obviously satisfied. As for (3.18), we remark that

$$\sum_1^{n-1} x_i^2 = \sum_1^{n-1} y_i^2$$

so that by (3.17) we get

$$\tilde{t}_{N,k-2} + \cos^2\theta_{n-1}\left(\sum_1^{n-2} y_i^2\right)\tilde{t}_{N,k} = 0. \tag{3.20}$$

Substituting (3.19) in (3.20) and taking into account

$$\left(\sum_1^{n-2} y_i^2\right)u_N = 0$$

we obtain

$$\left(\sum_1^{n-2} y_i^2\right)v_1 + u_N = 0,$$
$$\left(\sum_1^{n-2} y_i^2\right)v_{l+1} + v_l = 0, \quad l=1,..., \left[\frac{N}{2}\right]. \tag{3.21}$$

But the distributions u_N and v_l are finite linear combination of derivatives of $\delta(x_1)\cdots\delta(x_{n-1})$, so that the division by the polynomial $\Sigma_1^{n-2}y_i^2$ is meaningful. Therefore, (3.21) can be solved in term of u_N and the resulting $\tilde{t}_{N,k}$ verify (3.17) and (3.18).

Let us go back to the main proof. If we substitute (3.15) in (3.13), we get the same system for t^0_p as for t_p with the only following differences: (i) $t^0_N = 0$ by construction; (ii) the right-hand sides $\sin\theta_{n-1}s_p$ are replaced by distributions W_p such that $W_p = 0$ for $p \geqslant N-1$. Therefore the existence of a solution t of S_n for some N will result recurrently of the existence of a solution for $N=0$. But this is already contained in the recurrence hypothesis.

Finally it remains to prove the proposition for $n=3$.

We introduce, as above, the spherical coordinates R, θ_1, θ_2 and we put

$$z_3 = x_1 \sin\theta_1 \sin\theta_2 + x_2 \sin\theta_2 \cos\theta_1 + x_3\cos\theta_2,$$

$$z = x_1\cos\theta_1 - x_2\sin\theta_1 + i\cos\theta_2(x_1\sin\theta_1 + x_2\cos\theta_1) - ix_3\sin\theta_2. \tag{3.22}$$

Then (3.6) is written

$$(z_3^2 + z\bar{z})t = 0, \quad z_3 t = 0 \tag{3.23}$$

and similarly for s. As for (3.7), it becomes

$$z\left(\frac{\partial}{\partial\theta_1} - i\sin\theta_2\frac{\partial}{\partial\theta_2}\right)t + \bar{z}\left(\frac{\partial}{\partial\theta_1} + i\sin\theta_2\frac{\partial}{\partial\theta_2}\right)t = 2\sin\theta_2 s. \tag{3.24}$$

According to (3.23), we have the following finite expansions for t and s:

$$t = \delta(z_3)(t_0\delta(z)) + \sum_{p\geqslant 1}(t_p^+\delta^{(p,0)}(z) + t_p^-\delta^{(0,p)}(z)),$$
$$s = \delta(z_3)(s_0\delta(z)) + \sum_{p\geqslant 1}(s_p^+\delta^{(p,0)}(z) + s_p^-\delta^{(0,p)}(z)), \tag{3.25}$$

where $t_p^\pm, s_p^\pm$ belong to $\mathscr{D}'([0,\pi/2]^2)$ and $\delta^{(p,0)},\delta^{(0,p)}$ are the derivatives of $\delta(z) = \delta(z+\bar{z}/2)\delta(z-\bar{z}/2i)$ of order p in z and $\bar{z}$, respectively. Taking into account (3.23), the dependence of z, z_3 on θ_1, θ_2 and the relations

$$z\delta^{(p,q)}(z) = -p\delta^{(p-1,q)}(z), \quad \bar{z}\delta^{(p,q)}(z) = -q\delta^{(p,q-1)}(z)$$

we get after substitution of (3.25) into (3.24) and the replacement of $t_p^\pm$ by $v_p^\pm = \sin^p\theta_2 t_p^\pm$ and of $s_p^\pm$ by $u_p^\pm = \sin^p\theta_2 s_p^\pm$:

$$-\frac{\partial v^+_{p+1}}{\partial\theta_1} + i\sin\theta_2\frac{\partial v^+_{p+1}}{\partial\theta_2} = \frac{2\sin^2\theta_2}{(p+1)}u_p^+,$$
$$-\frac{\partial v^-_{p+1}}{\partial\theta_1} - i\sin\theta_2\frac{\partial v^-_{p+1}}{\partial\theta_2} = \frac{2\sin^2\theta_2}{(p+1)}u_p^-, \quad p\geqslant 1, \tag{3.26}$$

and for $p=0$

$$\frac{\partial v_1^+}{\partial\theta_1} - i\sin\theta_2\frac{\partial v_1^+}{\partial\theta_1} + \frac{\partial v_1^-}{\partial\theta_1} + i\sin\theta_2\frac{\partial v_1^-}{\partial\theta_2} = -2\sin^2\theta_2 u_0. \tag{3.27}$$

According to Appendix A, each Eq. (3.26) has at least one solution in $\mathscr{D}'([0,\pi/2]^2)$. As for (3.27) it can be solved by quadrature if we impose the supplementary condition

$$\frac{\partial}{\partial\theta_2}(v_1^+ - v_1^-) = 0.$$

Thus we conclude the proof of Proposition 3.1. As a mere corollary of it, we can state the essential result of this paper.

Theorem 1: The extension of a massless representation of the Poincaré group in 2 + 1 dimensions with helicity $\eta/2$, $\eta = 0,1$ defined in the space $\mathscr{S}_\eta$ isomorphic to the space of functions on $\mathbb{R}^2$ with rapid decrease at infinity and at the origin by the tensor product of n massless representations with helicity $\eta_i/2$, $i, = 1,...,\eta$ defined in $\mathscr{S}_{\eta_i}$, is always trivial for $n \geqslant 3$.

Proof: According to (1.9) and Proposition (3.1), we have

$$Z_{\mathbf{k}}(a,\Lambda) = TV_{\eta_1\cdots\eta_n}(0,\Lambda_k) - \epsilon^{\eta}(\Lambda,k) \times TV_{\eta_1\cdots\eta_n}(0,\Lambda_{\Lambda^{-1}k})V^{-1}_{\eta_1\cdots\eta_n}(0,\Lambda)$$

where T in $\mathscr{S}'_{\eta_1\cdots\eta_n}$ verifies (3.1).

Using the structure of T as displayed above, we show that

$$TV_{\eta_1\cdots\eta_n}(0,\Lambda_k)\in\mathscr{L}(\mathscr{S}_{\eta_1\cdots\eta_n};\mathscr{S}_{\eta}).$$

IV. APPLICATIONS AND CONCLUSIVE REMARKS

As pointed out in the introduction, the theorem above has the following immediate consequence.

Theorem 2: The equivalence classes of the formal nonlinear representations of $\mathscr{P}_3$ with irreducible physical representation as linear term are isomorphic to the classes of extension of this linear term by its symmetrical tensor product.

Proof: For the massive representations, we know from Ref. 5 that any nonlinear formal representation is nonlinearly equivalent to the representation itself. Similarly for the massless representation with helicity 1/2 according to the results in Ref. 2 and the present paper. So we have only to discuss the case of the massless representation with helicity zero.

We recall that a formal nonlinear representation is given by the following expansion:

$$f(\mathbf{k}) \xrightarrow{(a,\Lambda)} V_0(a,\Lambda)f(\mathbf{k})$$
$$(Z^2_{\mathbf{k}}(a,\Lambda),V_0(a,\Lambda)f\otimes V_0(a,\Lambda)f)$$
$$+\sum_{n>2}(Z^n_{\mathbf{k}}(a,\Lambda),\overbrace{V_0(a,\Lambda)f\otimes\cdots\otimes V_0(a,\Lambda)f}^{n\text{-terms}}).$$

Let us suppose we have built two representations with the same cocycle $Z^2_{\mathbf{k}}(a,\Lambda)$. Then the difference of the third terms of the expansion is a solution of the homogeneous equations, i.e., a cocycle of extension of $V_0(a,\Lambda)$ by its symmetrical third power. We know by Theorem 1 that this cocycle is a coboundary, so that we can transform the second formal representation in an equivalent representation identical to the first representation as for the first three terms of the expansion (see Appendix B). Then we know the difference of the fourth term is a coboundary and another nonlinear transform makes it equal to zero, etc. The infinite product of nonlinear transforms we need to achieve the identification is convergent in the space of formal series, for the nth term in the product does not change the $(n-1)$ first terms of the expansion: it defines a formal nonlinear transform which identifies the two original representations. This technique of proof has already been used in Ref. 6.

Remark 1: The set of nonlinear formal representations of $\mathscr{P}_3$ we have just obtained appears to be a very restricted set. But is would take a long time to exhaust all the possibilities: we can take as a linear term a direct sum of irreducible representations or of indecomposable representations. Owing to the nonlinearity, the resulting nonlinear representations cannot be deduced from the nonlinear representations built for each component separately.

Remark 2: The theorem cannot be extended to the spaces $\mathscr{D}_{\eta}$ of Ref. 2. The point is that the dual of $\mathscr{D}_{\eta_1\cdots\eta_n}$ contains distributions of infinite order. Consequently expansion (3.9) is no longer finite but must satisfy the rather strong condition of local finiteness. Let us consider the case $n=3$. We can proceed as in the third part of the proof of Proposition 3.1, so that we have finally to solve the system (3.26), (3.27) with $u^{\pm}_p, v^{\pm}_p$ in $\mathscr{D}'(]0,\pi/2[^2)$. Let us take as data:

$$u^+_p = \alpha_p\delta(p+\omega_1)\delta(\theta_1-\pi/4),\quad p\geqslant 1,\ \alpha_p\in\mathbb{C},$$
$$u_0 = 0,\quad u^-_p = 0,\quad p\geqslant 1,$$

in the variables $\theta_1,\omega_1 = \log tg\theta/2$ (see Appendix A). The requirement of local finiteness is obviously verified. Then we have, with $\zeta = \theta_1 + i\omega_1$:

$$v^+_{p+1} = -\frac{4\alpha_p}{\pi(p+1)}(e^p+e^{-p})^{-2}\left(\bar{\zeta}-\frac{\pi}{4}-ip\right)^{-1} + W^+_p(\bar{\zeta}),\quad p\geqslant 1,$$
$$v^-_{p+1} = W^-_p(\bar{\zeta}),\quad p\geqslant 1,$$

where $W^{\pm}_p$ are antiholomorphic function in the strip $B =]0,\pi/2[\times\mathbb{R}_-$.

The local finiteness of expansion (3.9) means that for a given compact K in B, the restriction of $v^{\pm}_p$ to K is identically zero for p sufficiently large. Therefore, we have on K for such p:

$$W^+_p = (4\alpha_p/\pi(p+1))(e^p+e^{-p})^{-2}(\bar{\zeta}-\pi/4-ip)^{-1}.$$

The uniqueness of the analytical extension contradicts the antiholomorphy of W^+_p on B. Thus although we can solve (3.26), (3.27), the solution must be rejected because it does not verify the local finiteness condition. Thus we have a counter-example to Proposition 3.1.

ACKNOWLEDGMENTS

We are indebted to Dr. J. Bertrand and Professor B. Gaveau for their interest in this work and many constructive discussions.

APPENDIX A: SOLUTION OF (3.26)

First we replace the test functions f by $\sin\theta_2 f$ and the variable θ_2 by ω_1:

$$\omega_1 = \log\tan\theta_2\quad \omega_1\in\mathbb{R}_-.$$

The space $\mathscr{D}([0,\pi/2]^2)$ is mapped onto the space $\mathscr{S}_1(B)$ of C^∞ functions on $\mathbb{R}^2$, identically zero outside the strip $B=[0,\pi/2]\times\mathbb{R}_-$ and going to zero at infinity on the lines $\theta_1 = \text{Cte}$ with all their derivatives faster that any power of $\exp\omega_1$. The topology of $\mathscr{S}_1(B)$ is defined by the family of semi-norms:

$$\|f\|_p = \sup_{\substack{\omega_1,\theta_1\in B\\ u+v\leqslant p}} \exp(-p\omega_1)\left|\frac{\partial^{u+v}f}{\partial\theta_1^u\,\partial\omega_1^v}(\theta_1,\omega_1)\right|.\quad\text{(A1)}$$

Then we have to prove that, for given u^+_p in $\mathscr{S}'_1(B)$, the following equation:

$$\frac{\partial v_{p+1}^+}{\partial \zeta} = -\frac{4}{p+1}(\exp \omega_1 + \exp(-\omega_1))^{-2} \times u_p^+, \quad \zeta = \theta_1 + i\omega_1$$

has at least a solution v_{p+1}^+ in $\mathscr{S}_1'(B)$.

According to Cor. 4, chap. 18 in Ref. 7, it is equivalent to prove that if $h_n = \partial f_n/\partial \zeta$ goes to zero in $\mathscr{S}_1(B)$, then f_n goes to zero in $\mathscr{S}_1(B)$.

Following Ref. 8 we have

$$f_n = \frac{1}{\pi}\int_{\mathbb{C}} \frac{h_n(\zeta',\bar\zeta')}{\bar\zeta - \bar\zeta'} d\zeta'\, d\bar\zeta' + \varphi_n(\bar\zeta),$$

where $\varphi_n(\bar\zeta)$ is some antiholomorphic function. But by definition, $h_n(\zeta',\bar\zeta')$ is orthogonal to any antiholomorphic function on B, in particular to $1/(\bar\zeta - \bar\zeta')$ when $\zeta \notin B$. Therefore $\varphi_n(\bar\zeta)$ is identically zero outside B and the uniqueness of the analytical extension implies $\varphi_n(\bar\zeta) = 0$. Thus we can write

$$f_n = \frac{1}{\pi}\int_{\mathbb{C}} \frac{h_n(\zeta',\bar\zeta')}{\bar\zeta - \bar\zeta'} d\zeta' d\bar\zeta' = -\frac{1}{\pi}\int_{\mathbb{C}} \frac{h_n(\zeta+\eta, \bar\zeta+\bar\eta)}{\bar\eta} d\eta\, d\bar\eta.$$

Accordingly, we obtain with $\eta = \rho \exp(i\alpha)$:

$$\exp(-p\omega_1)\left|\frac{\partial^{u+v} f_n}{\partial\theta_1^u \partial\omega_1^v}\right| \leqslant \frac{1}{\pi}\int_{-\pi/2}^{0} d\alpha \int_0^{\pi/2\cos\alpha} \rho\, d\rho \exp(\rho \sin\alpha) \times \exp\left(\frac{-p(\zeta+\eta-\bar\zeta-\bar\eta)}{2i}\right) \times \left|\frac{\partial^{u+v} h_n}{\partial\theta_1^u\, \partial\omega_1}(\zeta+\eta, \bar\zeta+\bar\eta)\right|.$$

But we have

$$\frac{1}{\pi}\int_{-\pi/2}^{0} d\alpha \int_0^{\pi/2\cos 2} \rho\, d\rho \exp(p\rho \sin\alpha) = \frac{1}{2p},$$

so that we get

$$\|f_n\|_p \leqslant 1/2p\|h_n\|_p$$

and f_n goes to zero in $\mathscr{S}_1(B)$ with h_n.

APPENDIX B: DEVICE USED IN PROOF OF THEOREM 2

For the sake of completeness, we give here the general device used in the proof of theorem 2 (see Ref. 6).

Let G be a group, $V(g)$ a representation of G in some linear topological space E. We consider the nonlinear formal representation of G given by the expansion:

$$f \to V(g)f - \sum_2^\infty Z^n(g)V(g)f \overset{n\text{-times}}{\hat\otimes \cdots \hat\otimes} V(g)f, \quad g \in G,\ f \in E, \tag{B1}$$

where as usual $\hat\otimes$ denotes the projective tensor product and $Z^n(g)$ is a linear mapping of $E \overset{n\text{-times}}{\hat\otimes \cdots \hat\otimes} E$ into E.

Let A be a mapping of $E \overset{p\text{-times}}{\hat\otimes \cdots \hat\otimes} E$ into E. We consider the nonlinear mapping of E into E given by

$$f \to \varphi = f + Af \overset{p\text{-times}}{\hat\otimes \cdots \hat\otimes} f \tag{B2}$$

This nonlinear mapping has a formal inverse:

$$f = \varphi - A\varphi \overset{p\text{-times}}{\hat\otimes \cdots \hat\otimes} \varphi + \cdots. \tag{B3}$$

Substituting φ to f in (B1) by (B3) and transforming the resulting expansion by (B2), we get a realization of the formal representation in term of φ. In this realization, the $(p-1)$ first terms are unchanged and the pth term is written

$$Z^p(g) + A - V(g)AV(g^{-1}) \overset{p\text{-times}}{\hat\otimes \cdots \hat\otimes} V(g^{-1})$$

as it is shown easily by direct calculation.

Using successive transforms as (B2), we can drive out all the coboundaries in (B1).

[1]M. Flato, G. Pinczon, and J. Simon, Ann. Sci. Ec. Norm. Sup. **10**, 405 (1977).
[2]J. Bertrand and G. Rideau, J. Math. Phys. **28**, 1972 (1987).
[3]I. M. Gelfand and G. E. Shilow, *Generalized Functions* (Academic, New York, 1968), Vol. 2.
[4]L. Hörmander, Ark. Mat. **3**, 555 (1958).
[5]E. Taflin: J. Math. Phys. **25**, 765 (1984).
[6]M. Flato and J. Simon, Lett. Math. Phys. **2**, 155 (1977).
[7]F. Treves, *Topological Vector Spaces, Distributions and Kernel* (Academic, New York, 1967).
[8]L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1950, 1951).

Microscopic collective nuclear models with horizontal mixture

G. F. Filippov and A. L. Blokhin
Institute for Theoretical Physics, Kiev-130, 252130, Union of Soviet Socialist Republics

(Received 28 June 1988; accepted for publication 10 May 1989)

By accounting for the difference in the spin and isospin projection among nucleons, the Sp(6,R) dynamical symmetry of the nuclear collective model is amplified in parallel with a horizontal extension of the instrinic shell structure. The generating kernels for the extended symplectic model basis functions are constructed in the form of a shortened coherent state. Analytical expressions for the microscopic Hamiltonian matrix elements, including central and tensor interactions, are obtained between the coherent states indicated.

I. INTRODUCTION

The problem of calculating the Hamilton operator matrix elements with respect to the basis functions of symplectic and unitary group irreducible representations (irreps) is of permanent importance in the theory of systems with a finite particle number.[1–14] The field of corresponding physical applications covers collective motion studies,[4–10] effective Hamiltonian construction,[1,8,15–18] nuclear and atomic spectroscopy,[6,8,9,15,19] and other branches. The most complicated part of the problem indicated is to determine the potential energy operator matrix elements; reaching this stage several researchers reject the indispensible microscopic approach in favor of the phenomenological Hamiltonians in the polynomial form of the Lie algebra generators.[1,4,8,10,12,16,17] The present paper realizes a program of constructing the generating kernel for the matrix elements of two-body interaction operators (central and tensor forces) between the symplectic nuclear model states admitting various (both regular and irregular) occupation of the valence nucleon shell. As one can note, the composing of intrinsic subspace by means of several shell occupations essentially amplifies the model basis compared with the conventional Sp(6,R) model.[4] Really, the complementarity between the intrinsic and collective motion[2,3] requires the whole space of states to be made up of different Sp(6,R) irreps. Following Park *et al.*,[12] we suppose that making use of the basis states of a single Sp(6,R) configuration is no more successful even for the s-d shell, so the exit through the light nuclei region with realistic Hamiltonians needs mixed representation calculations. The assumption is based on essential physical arguments. First, the dominant SU(3) irrep spans up to 80% of the full shell basis in the beginning of the 2s-1d shell. This result of Akiyama *et al.*[20] with a microscopic Hamiltonian affirmed the significance of intrinsic configuration mixing for light and medium nuclei. The analogous conclusion was stated by Draayer *et al.*[21] who had diagonalized a semiempirical Hamiltonian for ^{20}Ne in the space of mixed Sp(6,R) representations. The experimental data for the systems of 18–20 nucleons[22] also testify that the quantity of collective bands in real spectra notably exceeds the one irrep predictions. Then, the intrinsic SU(3) symmetry is characteristic only for the nuclei with rotational spectra. The intrinsic symmetry of the coupled rotor–vibrator-type corresponds to the higher dimension unitary groups and hence the space of states for such nuclei occurs to be reducible with respect to the SU(3) group. The idea was successfully realized in phenomenological interacting boson approximation with SU(6) symmetry.[23,24] And, finally, one notes that the Sp(6,R) configurations are inevitably mixed by the spin-orbit and tensor interaction.

In the version of the symplectic model developed below, the transfer to the mixed irreps results from the dynamical symmetry extension. The latter is achieved by dividing the nucleon system into four interacting subsystems according to the spin and isospin projection of nucleons. The basis states of every subsystem are transformed by the Sp(6,R) irrep, and, consequently, the dynamical symmetry of the whole system is determined by the direct product. The expansion of the latter, besides the dominant Sp(6,R) irrep, incorporates representations with the same number of quanta in the lowest shell state. Besides the vertical mixing of the SU(3) irreps within the Sp(6,R) one, the discussed extension of the Sp(6,R) model space fixes the horizontal mixing. Such mixing is necessary to set a correct structure on the low-energy spectrum region.[21] So one may expect the collective model with horizontal mixture to improve the description of the real nuclei by revising the conception of the intrinsic motion.

As a chief tool to obtain the explicit analytical formulas for the matrix elements of the physical operators, we utilize the generating kernel technique.[5,7,14] Section II begins with a review of some results from Refs. 5 and 7 on making use of shortened coherent states as the generator functions in the framework of the Sp(6,R) model with the regular occupation of the valence shell. Thereupon the extension scheme is considered in its application to the generating kernels for the basis state overlaps and kinetic energy matrix elements.

In Secs. III and IV an algorithm to calculate the generating kernel for the two-body central interaction operator matrix elements is developed. In Sec. V the previous results are modified for the case of tensor nuclear forces. In Sec. VI the formulas derived are detailed for several cases of physical interest. Section VII contains the conclusion.

II. GENERATING INVARIANTS OF EXTENDED SYMPLECTIC MODEL

The version of the generator function method used in the present paper was developed by Vasilevsky *et al.*[5] in application to the conventional Sp(6,R) model. They proposed to make use of a shortened coherent state as a generating kernel for the basis states of the symplectic group irrep. The shortened coherent state, in contrast to the general case, is constructed over the lowest SU(3) irrep only by means of the raising generators of the Sp(6,R) algebra.[25] Hence, it contains nine generating coordinates versus twenty one for the "true" coherent state. The basis states may be selected out of the coherent state (henceforth we omit the word "shortened"), generally speaking, by performing reiterated differentiation with respect to the generating coordinates. By taking the matrix elements of the physical operators between the coherent states, one obtains the corresponding generating kernels which depend only on the set of the generating coordinates. Then, either the standard method of solving the Hill–Wheeler equation[10] or the above pointed differentiation routine reduces the problem to the usual diagonalization of matrices.

The generator function technique is quite practicable to obtain analytical expressions for the matrix elements of various operators between the many-particle oscillator functions with arbitrary values of quantum numbers. Seemingly it is simpler, at least for highly excited states, than the general approach based on the fractional parentage decomposition,[26,27] on account of the latter needs recursive calculations with a large set of Wigner coefficients of the symmetric and unitary groups. Besides, as it was noted in Ref. 7, there exists a way by using the generating kernels to estimate an asymptotic behavior of the corresponding matrix elements at the number of oscillator quanta increasing to infinity, and consequently, to eliminate the convergence problem for the oscillator expansion. However, in the case of essentially antisymmetrized spatial Young pattern, or, which is the same, an extremely irregular shell occupation, making use of the parentage coefficients is vital to construct the coherent state. So we leave the latter case beyond our discussion and appreciate the necessity of a generalized algorithm reasonably compound of both the approaches.

Following Refs. 5 and 7, Sp(6,R) irrep coherent states that generate the oscillator basis can be constructed over the oscillator SU(3) multiplets:

$$|\mathbf{b};\mathbf{u}\rangle = \exp\{\mathrm{Trace}(\mathbf{b}\hat{\mathbf{A}}^{+})\}|\mathbf{u}\rangle , \qquad (2.1)$$

where $\hat{\mathbf{A}}^{+} = \|\hat{\mathbf{A}}^{+}_{rs}\|$, $r, s = \overline{1,3}$ is the symmetric matrix of collective quantum creation operators (here

$$\hat{\mathbf{A}}^{+}_{rs} = \sum_{i=1}^{A-1} \hat{a}^{+}_{ir}\hat{a}^{+}_{is}$$

for the system of A nucleons, $\hat{a}^{+}_{ir}$ are oscillator quantum creation operators), $\mathbf{b} = \|b_{rs}\|$ is the symmetric matrix of collective excitation generator coordinates. The latter are, generally speaking, complex variables. Nevertheless, to fit the generator function one always may assign real values to them. Let the ket (2.1) belong to the $[\sigma_1\sigma_2\sigma_3]$ irrep of Sp(6,R) group. If this state is treated as a generating invariant for the oscillator basis of the minimal approximation of generalized hyperspherical functions method[6] (i.e., for collective wave functions of fixed O($A-1$) symmetry $[f_1f_2f_3]$), the symplectic and orthogonal group irrep indices relation[3] is written as

$$\sigma_i = \tfrac{1}{2}f_i + \tfrac{1}{4}(A-1), \quad i = \overline{1,3}.$$

The SU(3) multiplet $|\mathbf{u}\rangle$ quantum numbers are defined, besides the f_i indices, also by means of SU(3)$\supset$SO(3) reduction ones:

$$N_{\min} = f_1 + f_2 + f_3, \quad \lambda = f_1 - f_2, \quad \mu = f_2 - f_3, \quad LM\alpha ,$$

where $N_{\min}$ is the total number of quanta, $(\lambda\mu)$ is Elliott's notation for the SU(3) numbers,[1] LM and α are the values of orbital angular momentum, its projection and multiplicity index, respectively. The $\mathbf{u} = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$ symbol denotes an orthogonal matrix composed of three vectors of the Cartesian axes orientation. The initial orientation $\mathbf{u} = \mathbf{E}$ (where $\mathbf{E}$ is the unit matrix of the third order) determines the lowest-weight state of the $(\lambda\mu)$ irrep. Three independent generator coordinates, parametrizing the matrix, provide the reduction onto the SO(3) group, e.g., by means of the Peierls–Yoccoz projection technique.[28]

We introduce, following Filippov *et al.*,[7] the single-particle states

$$\langle\mathbf{r}|\mathbf{n},\nu,\mathbf{u}\rangle = (\pi^{3/2}2^{n}\mathbf{n}!)^{-1/2}\zeta_{\nu} \times \exp\left(-\frac{1}{2}\mathbf{r}^{2}\right)\prod_{k=1}^{3} H_{n_k}(\mathbf{u}_k\mathbf{r}) , \qquad (2.2)$$

where $\mathbf{n} = \{n_1,n_2,n_3\}$ are the occupation numbers, $n \equiv n_1 + n_2 + n_3$, $\mathbf{n}! \equiv n_1!n_2!n_3!$, ζ_ν is the spin–isopin function with projection values $\nu = \{\sigma\tau\}$. The distance scale is chosen normed to the oscillator radius $r_0 = \sqrt{\hbar/m\omega}$. If the nucleon shells are filled in a regular manner, i.e., the valence nucleon state is expressed through the Young pattern, fully symmetrized over the antisymmetrization due to a fixed spin and isospin projection, then the single-particle kets (2.2) form a Slater determinant of the whole system state. So the partition numbers f_i are identified with the total numbers of quanta along the Cartesian axes. By passing to Jacobi coordinates in Eq. (2.2) one isolates the center-of-mass factor. The translationally invariant part of the Slater determinant coincides with the SU(3) irrep generator function $|\mathbf{u}\rangle$ of Eq. (2.1).

Now we proceed to abandon the operator form of the Sp(6,R) irrep coherent state (2.1) in favor of the coordinate form. Let us denote the eigenvalues of matrix $\mathbf{b}$ of the collective generator coordinates by β_1, β_2, and β_3. Then the coherent state of the Sp(6,R) irrep with quantum numbers

$$[\sigma_1\sigma_2\sigma_3] = [\tfrac{1}{2}f_1 + \tfrac{1}{4}(A-1), \tfrac{1}{2}f_2 + \tfrac{1}{4}(A-1), \tfrac{1}{2}f_3 + \tfrac{1}{4}(A-1)]$$

is expressed through the translationally invariant part of the Slater determinant composed of the single-particle orbitals

$$\langle\mathbf{r}|\mathbf{b};\mathbf{n},\nu,\mathbf{u}\rangle = \exp\{-\mathbf{r}^{T}\mathbf{b}(\mathbf{E}-\mathbf{b})^{-1}\mathbf{r}\}\langle\mathbf{r}|\mathbf{n},\nu,\mathbf{u}\rangle \qquad (2.3)$$

and supplied with factor $\Pi_k(1-\beta_k)^{-2\sigma_k}$. (The superscript T with vectors and matrices means their transposition). Such a state is defined by three Sp(6,R) quantum numbers and nine generator coordinates. Three of the latter are responsible for the intrinsic state description, and the other for the monopole and quadrupole collective excitations.

As far as the generator function for the basis state is identified with the coherent state, the matrix elements of an arbitrary operator $\hat{F}$ between the coherent states, i.e.,

$$\langle\tilde{\mathbf{b}};i\mathbf{v}|\hat{F}|\mathbf{b};\mathbf{u}\rangle,$$

are treated as the generating kernels. We denote the corresponding generating coordinates of the coherent bra and ket in a different manner to emphasize that conjugated bases are yielded by the independent generator functions.

For a single or two-particle operator $\hat{F}$ the calculation of the matrix elements between the determinant functions $\langle\tilde{\mathbf{b}};\mathbf{v}|$ and $|\mathbf{b};\mathbf{u}\rangle$ can be performed according to the Löwdin algorithm.[29] Describing nucleon interactions through effective two-body potentials we assume that to calculate such matrix elements is quite sufficient for solving the problems of practical significance. We put off the construction of the potential energy operator generating matrix elements until the next section and now turn to the case of one-body operators.

The coherent state overlaps

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle$$

are the simplest generating matrix elements. They contain information on the structure of the basis functions and normalization coefficients. Analytic expressions for the overlaps are to be found[7] by generalizing the Elliott formula[1] for the overlap integrals of definite SU(3)-symmetry oscillator functions:

$$(\mathbf{v}[f_1f_2f_3]|\mathbf{u}[f_1f_2f_3]) = (\mathbf{u}_1\mathbf{v}_1)^{f_1-f_2}([\mathbf{u}_1\mathbf{u}_2][\mathbf{v}_1\mathbf{v}_2])^{f_2-f_3}, \quad (2.4)$$

where $|\mathbf{u}[f_1f_2f_3])$ and $(\mathbf{v}[f_1f_2f_3]|$ are the Slater determinants composed of the corresponding one-particle kets (2.2) and bras. (Square brackets henceforth symbolize the vector product.) Note that a linear transformation of the coordinate system extends the lowest SU(3) irrep coherent state to the Sp(6,R) irrep one.

Every overlap of oscillator orbitals (2.3)

$$\langle\tilde{\mathbf{b}};\tilde{\mathbf{n}},\nu,\mathbf{v}|\mathbf{b};\mathbf{n},\nu,\mathbf{u}\rangle = \langle\tilde{\mathbf{b}};\tilde{\mathbf{n}},\mathbf{v}|\mathbf{b};\mathbf{n},\mathbf{u}\rangle\delta_{\nu\tilde{\nu}}$$

implies the integration of an exponential factor

$$\exp(-\mathbf{r}^T\mathbf{B}\mathbf{r}), \quad (2.5a)$$

$$\mathbf{B} = \mathbf{E} + \mathbf{b}(\mathbf{E}-\mathbf{b})^{-1} + \tilde{\mathbf{b}}(\mathbf{E}-\tilde{\mathbf{b}})^{-1}. \quad (2.5b)$$

One can perform a linear coordinate transformation

$$\mathbf{x} = \mathbf{B}^{1/2}\mathbf{r}, \quad \mathbf{u}' = \mathbf{B}^{-1/2}\mathbf{u}, \quad \mathbf{v}' = \mathbf{B}^{-1/2}\mathbf{v} \quad (2.6)$$

conserving bilinear forms

$$\mathbf{r}^T\mathbf{u} = \mathbf{x}^T\mathbf{u}', \quad \mathbf{r}^T\mathbf{v} = \mathbf{x}^T\mathbf{v}'.$$

There oscillator reper matrices $\mathbf{u}'$ and $\mathbf{v}'$ are no longer orthogonal. To make the Elliott formula (2.4) applicable one employs the Gram–Schmidt orthonormalization technique presented as a triangular transformation

$$\mathbf{u}' = \mathbf{u}''\boldsymbol{\epsilon}, \quad \mathbf{v}' = \mathbf{v}''\tilde{\boldsymbol{\epsilon}}, \quad (2.7)$$

where $\boldsymbol{\epsilon}$ and $\tilde{\boldsymbol{\epsilon}}$ are the upper triangular matrices, $\mathbf{u}''$ and $\mathbf{v}''$ are the new orthogonal ones. The next calculations will need only diagonal elements of the transformation matrices:

$$\epsilon_{11} = |\mathbf{u}_1'|, \quad \epsilon_{22} = \frac{|[\mathbf{u}_1'\mathbf{u}_2']|}{|\mathbf{u}_1'|}, \quad \epsilon_{33} = \frac{|(\mathbf{u}_1'\mathbf{u}_2'\mathbf{u}_3')|}{|[\mathbf{u}_1'\mathbf{u}_2']|},$$
$$\tilde{\epsilon}_{11} = |\mathbf{v}_1'|, \quad \tilde{\epsilon}_{22} = \frac{|[\mathbf{v}_1'\mathbf{v}_2']|}{|\mathbf{v}_1'|}, \quad \tilde{\epsilon}_{33} = \frac{|(\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3')|}{|[\mathbf{v}_1'\mathbf{v}_2']|}. \quad (2.8)$$

The determinant form of the Sp(6,R) irrep coherent states employed and the condition that nucleon shells are regularly filled allow us to obtain the $|\mathbf{b};\mathbf{u}[f_1f_2f_3])$ determinant composed of orbitals (2.3) from the $|\mathbf{u}[f_1f_2f_3])$ determinant using a simple substitution for the parameters of one-particle states (2.2):

$$\langle\mathbf{r}|\mathbf{n},\nu,\mathbf{u}\rangle \rightarrow \epsilon_{11}^{n_1}\epsilon_{22}^{n_2}\epsilon_{33}^{n_3}\langle\mathbf{x}|\mathbf{n},\nu,\mathbf{u}''\rangle. \quad (2.9)$$

After performing an analogous substitution for the bras, one reaches the expression sought for

$$(\tilde{\mathbf{b}};\mathbf{v}[f_1f_2f_3]|\mathbf{b};\mathbf{u}[f_1f_2f_3]) = (\mathbf{v}''[f_1f_2f_3]|\mathbf{u}''[f_1f_2f_3]) \times|\mathbf{B}|^{-A/2}(\epsilon_{11}\tilde{\epsilon}_{11})^{f_1}(\epsilon_{22}\tilde{\epsilon}_{22})^{f_2}(\epsilon_{33}\tilde{\epsilon}_{33})^{f_3}, \quad (2.10)$$

where the multiplier depending on the matrix B determinant represents the Jacobian of transition (2.6) to spatial coordinates $\mathbf{x}$. Taking the relations (2.4), (2.7), (2.8), and (2.10) into account, we obtain

$$(\tilde{\mathbf{b}};\mathbf{v}[f_1f_2f_3]|\mathbf{b};\mathbf{u}[f_1f_2f_3]) = |\mathbf{B}|^{-A/2}(\mathbf{u}_1'\mathbf{v}_1')^{f_1-f_2} \times([\mathbf{u}_1'\mathbf{u}_2'][\mathbf{v}_1'\mathbf{v}_2'])^{f_2-f_3}\{(\mathbf{u}_1'\mathbf{u}_2'\mathbf{u}_3')(\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3')\}^{f_3}, \quad (2.11)$$

where the parentheses with two vector arguments indicate a scalar product, and those with three vector arguments indicate a mixed one. Choosing the new notation

$$\Delta = |\mathbf{B}||\mathbf{E}-\mathbf{b}||\mathbf{E}-\tilde{\mathbf{b}}| = |\mathbf{E}-\mathbf{b}\tilde{\mathbf{b}}| \equiv D(\tilde{\mathbf{b}},\mathbf{b}), \quad (2.12a)$$

$$\mathscr{M} = \Delta\cdot(\mathbf{u}_1'\mathbf{v}_1') = \Delta\cdot\mathbf{v}_1^T\mathbf{B}^{-1}\mathbf{u}_1, \quad (2.12b)$$

$$\mathscr{K} = \Delta\cdot([\mathbf{u}_1'\mathbf{u}_2'][\mathbf{v}_1'\mathbf{v}_2']) = |\mathbf{E}-\mathbf{b}||E-\tilde{\mathbf{b}}|\cdot\mathbf{v}_3^T\mathbf{B}\mathbf{u}_3 \quad (2.12c)$$

(it is easily seen that Δ, $\mathscr{M}$, and $\mathscr{K}$ are polynomials on the matrix elements of $\mathbf{b}$ and $\tilde{\mathbf{b}}$), using

$$(\mathbf{u}_1'\mathbf{u}_2'\mathbf{u}_3')(\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3') = |\mathbf{B}|^{-1}$$

and excluding the center-of-mass motion by normalizing factor $\Delta^{1/2}$, we come from (2.11) to the Sp(6,R) $[\sigma_1\sigma_2\sigma_3]$ irrep coherent state overlap

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle = \frac{\mathscr{M}^{f_1-f_2}\mathscr{K}^{f_2-f_3}}{\Delta^{f_1+(1/2)(A-1)}}. \quad (2.13)$$

Formula (2.13) derived in Ref. 7 is applicable if nucleon shells are occupied in the regular manner. But the regular occupation is preferred only for the lightest nuclei; an increase in the nucleon number results in holes in the shell configurations being energetically preferred.[30,31] To incorporate the irregular shell occupation, as it was discussed in the Introduction, one has to account for the horizontal mixing.

The extended symplectic model proposed in the present paper describes nucleon systems with horizontal mixture in an open shell. Realizing it, one preserves some advantages of the above considered variant of the Sp(6,R) model with regular shell occupation: the usage of a generating invariant in Slater determinant form and, consequently, the existence of a simple transformation connecting the Sp(6,R) and SU(3) coherent states. In the extended model the system of A nucleons is treated as divided into four interacting subsystems, each of them containing nucleons in a fixed spin–isospin state ν. The subsystems are built up in a regular way; the corresponding intrinsic configuration is characterized by the total numbers of quanta $f_{\nu1}, f_{\nu2}, f_{\nu3}$ along the oscillator axes [in the extended SU(3) model Elliott indices $(\lambda_\nu \mu_\nu) = (f_{\nu1} - f_{\nu2}, f_{\nu2} - f_{\nu3})$ are sufficient].

Some comment is required for the previous paragraph. Note that every subsystem is unambiguously characterized by the fully antisymmetric spatial Young pattern. For the p-shell the space symmetry coincides with the intrinsic state U(3) partition $[f_{\nu1}\, f_{\nu2}\, f_{\nu3}]$. For the higher shell one classifies the oscillator quanta, generally speaking, by several partitions. Henceforth we consider the subsystem ν imparted by a fixed $[f_{\nu1}\, f_{\nu2}\, f_{\nu3}]$ partition and, consequently, the whole system labeled by a set of four partitions. The latter condition is based on a physical assumption of the relatively weak coupling of subspaces with the fixed partition set. Really, one can divide the full Hamiltonian as follows:

$$H = H_0 + \sum_{\nu} H_\nu + H_{\rm res} ,$$

where H_0 is the Sp(6,R) invariant contribution of the entire system, H_ν are the corrections due to the independent subsystems, and $H_{\rm res}$ is the residual term corresponding to the interaction of the subsystems. Contrary to the other terms, $H_{\rm res}$ is originated only by the valence nucleon interaction and, for that reason, is expected to be rather small. Hence the cross matrix elements between the states with different partition sets take a nonzero value only on the smallest term of the Hamiltonian.

The formulas for coherent states and their overlaps in the extended model generalize formulas (2.2)–(2.13) quite easily. The single-particle kets of the intrinsic state of the subsystem ν are obtained from states (2.4) by replacing the oscillator reper matrices

$$\mathbf{u} \to \mathbf{u}_\nu ; \quad \langle \mathbf{r}|\mathbf{n},\nu,\mathbf{u}\rangle \to \langle \mathbf{r}|\mathbf{n},\nu,\mathbf{u}_\nu\rangle . \tag{2.14}$$

The matrix $\mathbf{u}_\nu$ is supposed to be constructed from the vector columns $\mathbf{u}_{\nu1}, \mathbf{u}_{\nu2}, \mathbf{u}_{\nu3}$. The bras undergo an analogous modification. The collective matrices $\mathbf{B}$ (2.5b) remain unchanged, reflecting the physical assumption about the defining role of the intrinsic configurations in the structure of a nuclear system. If necessary, an additional mixture in collective occupation may be injected into the model without special difficulties (but this will cause an extraordinary calculation inconvenience in applications to the concrete nuclei).

Returning to the extended Sp(6,R) model generating kernels we generalize formula (2.13) according to the scheme (2.14):

$$\langle \tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle = \frac{\Pi_\nu \mathscr{M}_\nu^{f_{\nu1}-f_{\nu2}} \mathscr{K}_\nu^{f_{\nu2}-f_{\nu3}}}{\Delta^{f_1 + (1/2)(A-1)}} , \tag{2.15a}$$

$$\mathscr{M}_\nu = \Delta \cdot \mathbf{v}_{\nu1}{}^T \mathbf{B}^{-1} u_{\nu1} \equiv M_\nu(\tilde{\mathbf{b}},\mathbf{v};\mathbf{b},\mathbf{u}) , \tag{2.15b}$$

$$\mathscr{K}_\nu = |\mathbf{E}-\mathbf{b}||\mathbf{E}-\tilde{\mathbf{b}}|\cdot \mathbf{v}_{\nu3}^T \mathbf{B}\mathbf{u}_{\nu3} \equiv K_\nu(\tilde{\mathbf{b}},\mathbf{v};\mathbf{b},\mathbf{u}) , \tag{2.15c}$$

where

$$f_i = \sum_\nu f_{\nu i}, \quad i = 1,3 .$$

Now we deal with the kinetic energy operator $\hat{T}$ matrix elements. Let us denote by α the set of the quantum numbers of a normalized basis function. Using the virial theorem for the harmonic oscillator, one can deduce

$$\langle \tilde{\alpha}|\hat{T}|\alpha\rangle = -(\hbar^2/2m)\langle\tilde{\alpha}|\rho^2|\alpha\rangle(1 - 2\delta_{\alpha\tilde{\alpha}}) \tag{2.16}$$

to connect the desired matrix elements with the elements of squared hyperradius operator

$$\rho^2 = \sum_{i=1}^{A} \mathbf{r}_i^2 - \frac{1}{A}\left(\sum_{i=1}^{A} \mathbf{r}_i\right)^2 .$$

A formula for the generating kernels of the ρ^2 matrix in the Sp(6,R) model with a regularly occupied open shell was also derived in Ref. 7. Generalized to the broadened model, it looks like

$$\begin{aligned}\langle \tilde{\mathbf{b}};\mathbf{v}|\rho^2|\mathbf{b};\mathbf{u}\rangle = \Big\{ & f_1 + f_2 + f_3 + \tfrac{3}{2}(A-1) \\ & + \sum_\nu (f_{\nu1} - f_{\nu2}) \frac{\partial}{\partial\gamma} \ln \mathscr{M}_\nu(\gamma) \\ & + \sum_\nu (f_{\nu2} - f_{\nu3}) \frac{\partial}{\partial\gamma} \ln \mathscr{K}_\nu(\gamma) \\ & + \left(f_{\nu1} + \frac{A-1}{2}\right) \frac{\partial}{\partial\gamma} \ln \Delta(\gamma)\Big\}\Big|_{\gamma=0} ,\end{aligned} \tag{2.17a}$$

$$\mathscr{M}_\nu(\gamma) = \frac{M_\nu((1-\gamma)\tilde{\mathbf{b}} + \gamma\mathbf{E},\mathbf{v};(1-\gamma)\mathbf{b} + \gamma\mathbf{E},\mathbf{u})}{M_\nu(\gamma\mathbf{E},\mathbf{E};\gamma\mathbf{E},\mathbf{E})} , \tag{2.17b}$$

$$\mathscr{K}_\nu(\gamma) = \frac{K_\nu((1-\gamma)\tilde{\mathbf{b}} + \gamma\mathbf{E},\mathbf{v};(1-\gamma)\mathbf{b} + \gamma\mathbf{E},\mathbf{u})}{K_\nu(\gamma\mathbf{E},\mathbf{E};\gamma\mathbf{E},\mathbf{E})} , \tag{2.17c}$$

$$\Delta(\gamma) = \frac{D((1-\gamma)\tilde{\mathbf{b}} + \gamma\mathbf{E};(1-\gamma)\mathbf{b} + \gamma\mathbf{E})}{D(\gamma\mathbf{E};\gamma\mathbf{E})} , \tag{2.17d}$$

where the functional dependence of M_ν, K_ν, and D was defined by (2.15b), (2.15c), and (2.12a).

III. POTENTIAL ENERGY MATRIX GENERATING KERNELS. EXPONENTIAL GENERATING FUNCTION FOR PARTIAL EXPANSION

In the present and subsequent sections we derive analytic expressions for the generating kernels of the matrix of Wigner nucleon–nucleon interaction operator

$$\hat{U}_c = \sum_{i<j\leqslant A} w_0 \exp\left[-\frac{\gamma(\mathbf{r}_i - \mathbf{r}_j)^2}{2}\right] \equiv \sum_{i<j\leqslant A} \hat{V}(r_i - r_j) , \tag{3.1}$$

where γ is a parameter settling the interaction range. One can generalize the following results to the case of potentials summing several Gaussian functions in a trivial manner. To work with the potentials including other radial dependence, one has to integrate over the parameter γ with a definite weight function.

In accordance with Löwdin's algorithm,[29] we write the two-body operator $\hat{U}_c$ matrix elements between the determinant functions $\langle\tilde{\mathbf{b}};\mathbf{v}|$ and $|\mathbf{b};\mathbf{u}\rangle$ in the form of

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\hat{U}_c|\mathbf{b};\mathbf{u}\rangle = \langle\tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle \times\sum_{\nu\tilde{\nu}} w_0[W^+_{\nu\tilde{\nu}}(\gamma) - W^-_{\nu\tilde{\nu}}(\gamma)], \quad (3.2a)$$

$$W^+_{\nu\tilde{\nu}}(\gamma) = \iint d\mathbf{r}_1\, d\mathbf{r}_2\, \rho_\nu(r_1,r_1)\, \rho_\nu(r_2,r_2) \times\exp\left[-\frac{\gamma}{2}(\mathbf{r}_1-\mathbf{r}_2)^2\right], \quad (3.2b)$$

$$W^-_{\nu\tilde{\nu}}(\gamma) = \iint d\mathbf{r}_1\, d\mathbf{r}_2\, \rho_\nu(\mathbf{r}_1,\mathbf{r}_2)\, \rho_\nu(\mathbf{r}_2,\mathbf{r}_1) \times\exp\left[-\frac{\gamma}{2}(\mathbf{r}_1-\mathbf{r}_2)^2\right], \quad (3.2c)$$

where $W^+_{\nu\tilde{\nu}}(\gamma)$, $W^-_{\nu\tilde{\nu}}(\gamma)$ are the integrals of direct and exchange nucleon interaction between the subsystems ν and $\tilde{\nu}$, $\rho_\nu(\mathbf{r}_1,\mathbf{r}_2)$ is the single-particle spatial density matrix in the subsystem ν.

We guess the construction of density matrices of single-particle orbitals $\langle\mathbf{r}|\mathbf{b};\mathbf{n},\nu,\mathbf{u}_\nu\rangle$ and $\langle\tilde{\mathbf{b}};\mathbf{n},\nu,\mathbf{u}_\nu|\mathbf{r}\rangle$ to be inexpedient, because the latter are nonorthogonal, if the quantum numbers $\mathbf{n}$ and $\tilde{\mathbf{n}}$ belong to the same shell (in the case $\nu\neq\tilde{\nu}$ the orbitals are orthogonal like the corresponding spin–isospin functions). So we prefer to orthogonalize the employed orbitals beforehand. Then their overlap matrix becomes diagonal, providing undoubted advantages for further transformation.

To solve the stated problem, we use the technique of partial expansion generating functions applied in Ref. 7 to the SU(3) model. One can present the single-particle states (2.2) as a derivative of an exponent with respect to the components of the vector parameter $\mathbf{t}_\nu = \{t_{\nu 1},t_{\nu 2},t_{\nu 3}\}$:

$$\langle\mathbf{r}|\mathbf{n},\nu,\mathbf{u}_\nu\rangle = \pi^{-3/4}\zeta_\nu D(\mathbf{n},\mathbf{t}_\nu) \times\exp\{-\mathbf{t}_\nu^2 + 2\mathbf{r}^T\mathbf{u}_\nu\mathbf{t}_\nu - \tfrac{1}{2}\mathbf{r}^2\}|_{\mathbf{t}_\nu=0}, \quad (3.3a)$$

$$D(\mathbf{n},\mathbf{t}_\nu) = (2^n\mathbf{n}!)^{-1/2}\frac{\partial^{n_1}\partial^{n_2}\partial^{n_3}}{\partial t_{\nu 1}^{n_1}\partial t_{\nu 2}^{n_2}\partial t_{\nu 3}^{n_3}}. \quad (3.3b)$$

The expression (3.3) makes use of the generating function for Hermite polynomials. A passage to an analogous representation for the single-particle orbitals of the symplectic model takes place as a result of the spatial transformation (2.5)–(2.9)

$$\langle\mathbf{r}|\mathbf{b};\mathbf{n},\nu,\mathbf{u}_\nu\rangle \sim \zeta_\nu\pi^{-3/4}D(\mathbf{n},t_\nu) \times\exp\{-\mathbf{t}_\nu^2 + 2\mathbf{x}^T\mathbf{u}''_\nu\mathbf{t}_\nu - \tfrac{1}{2}\mathbf{x}^2\}|_{\mathbf{t}_\nu=0}. \quad (3.4)$$

The factor depending on the elements of the triangular matrix $\boldsymbol{\epsilon}_\nu$ is omitted in formula (3.4), because it contributes only to the coherent state overlap [see (3.2)], but the subsequent calculations involve the direct $W^+_{\nu\tilde{\nu}}(\gamma)$ and exchange $W^-_{\nu\tilde{\nu}}(\gamma)$ integrals.

We define the new generating parameters $\{\tau_{\nu 1},\tau_{\nu 2},\tau_{\nu 3}\} = \boldsymbol{\tau}_\nu$ by the following triangular transformation:

$$\mathbf{t}_\nu = \boldsymbol{\omega}_\nu\boldsymbol{\tau}_\nu, \quad (3.5a)$$

$$\boldsymbol{\omega}_\nu = \begin{pmatrix} 1/(\mathbf{u}''_{\nu 1}\mathbf{u}''_{\nu 1})^{1/2} & -(\mathbf{u}''_{\nu 2}\mathbf{v}''_{\nu 1})/(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} & (\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 3})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \\ 0 & (\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})/(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} & (\mathbf{u}''_{\nu 2}\mathbf{v}''_{\nu 3})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \\ 0 & 0 & (\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \end{pmatrix}. \quad (3.5b)$$

Replacing in the formula (3.4) the differential operator $D(\mathbf{n},\mathbf{t}_\nu)$ (3.3b) by the operator $D(\mathbf{n},\boldsymbol{\tau}_\nu)$, we create new single-particle orbitals

$$\langle\,\overline{\mathbf{r}|\mathbf{b},\mathbf{n},\nu,\mathbf{u}_\nu}\,\rangle \sim \zeta_\nu\pi^{-3/4}D(\mathbf{n},\boldsymbol{\tau}_\nu)\exp\{-\mathbf{t}_\nu^2 + 2\mathbf{x}^T\mathbf{u}''_\nu\mathbf{t}_\nu - \tfrac{1}{2}\mathbf{x}^2\}|_{\mathbf{t}_\nu=0}.$$

It seems instrumental to pass within the exponent to parameters $\boldsymbol{\tau}_\nu$ simultaneously with the oscillator reper transformation which sets an invariant image of the vector $\mathbf{u}''_\nu\mathbf{t}_\nu$:

$$\mathbf{U}_\nu = \mathbf{u}''_\nu\boldsymbol{\omega}_\nu, \quad \mathbf{U}_\nu\boldsymbol{\tau}_\nu = \mathbf{u}''_\nu\mathbf{t}_\nu, \quad (3.6a)$$

$$\langle\,\overline{\mathbf{r}|\mathbf{b},\mathbf{n},\nu,\mathbf{u}_\nu}\,\rangle \sim \pi^{-3/4}\zeta_\nu D(\mathbf{n},\boldsymbol{\tau}_\nu)\exp\{-(\mathbf{U}_\nu\boldsymbol{\tau}_\nu)^2 + 2\mathbf{x}^T\mathbf{U}_\nu\boldsymbol{\tau}_\nu - \tfrac{1}{2}\mathbf{x}^2\}|_{\boldsymbol{\tau}_\nu=0}. \quad (3.6b)$$

In the last relation the orthogonality of the matrix $\mathbf{u}''$ was used. In the same way we define the new bras

$$\langle\,\overline{\tilde{\mathbf{b}};\tilde{\mathbf{n}},\nu,\mathbf{u}_\nu|\mathbf{r}}\,\rangle \sim \pi^{-3/4}\zeta_\nu D(\tilde{\mathbf{n}},\boldsymbol{\sigma}_\nu)\exp\{-(\mathbf{V}_\nu\boldsymbol{\sigma}_\nu)^2 + 2\mathbf{x}^T\mathbf{V}_\nu\boldsymbol{\sigma}_\nu - \tfrac{1}{2}\mathbf{x}^2\}|_{\boldsymbol{\sigma}_\nu=0}, \quad (3.7a)$$

$$\mathbf{V}_\nu = \mathbf{v}''_\nu\tilde{\boldsymbol{\omega}}_\nu, \quad (3.7b)$$

$$\tilde{\boldsymbol{\omega}}_\nu = \begin{pmatrix} 1/(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2} & -(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 2})/(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}(\mathbf{u}_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} & (\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 1})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \\ 0 & (\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})/(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} & (\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 2})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \\ 0 & 0 & (\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})/(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2} \end{pmatrix}. \quad (3.7c)$$

The matrices $\mathbf{U}_\nu$ and $\mathbf{V}_\nu$ may be treated as consisting of vector columns

$$\mathbf{U}_{\nu 1} = \frac{\mathbf{u}''_{\nu 1}}{(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}},$$
$$\mathbf{U}_{\nu 3} = \frac{\mathbf{v}''_{\nu 3}}{(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2}}, \quad \mathbf{U}_{\nu 2} = [\mathbf{V}_{\nu 3}\mathbf{V}_{\nu 1}],$$
$$\mathbf{V}_{\nu 1} = \frac{\mathbf{v}''_{\nu 1}}{(\mathbf{u}''_{\nu 1}\mathbf{v}''_{\nu 1})^{1/2}}, \tag{3.8a}$$
$$\mathbf{V}_{\nu 3} = \frac{\mathbf{u}''_{\nu 3}}{(\mathbf{u}''_{\nu 3}\mathbf{v}''_{\nu 3})^{1/2}}, \quad \mathbf{V}_{\nu 2} = [\mathbf{U}_{\nu 3}\mathbf{U}_{\nu 1}],$$

which, as is easily seen, are connected by the reciprocity relations

$$(\mathbf{U}_{\nu i}\mathbf{V}_{\nu k}) = \delta_{ik}, \quad i,k = \overline{1,3}. \tag{3.8b}$$

The matrices

$$\mathbf{P}_{\nu k} = \mathbf{U}_{\nu k}\mathbf{V}^T_{\nu k}, \quad \mathbf{P}^T_{\nu k} = \mathbf{V}_{\nu k}\mathbf{U}^T_{\nu k} \tag{3.9a}$$

of the outer products of vectors $\mathbf{U}_{\nu k}$ and $\mathbf{V}_{\nu k}$ are projection ones:

$$\mathbf{P}_{\nu i}\mathbf{P}_{\nu k} = \mathbf{P}_{\nu i}\delta_{ik}, \quad \mathbf{P}^T_{\nu i}\mathbf{P}^T_{\nu k} = \mathbf{P}^T_{\nu i}\delta_{ik}, \tag{3.9b}$$

$$\sum_k \mathbf{P}_{\nu k} = \sum_k \mathbf{P}^T_{\nu k} = \mathbf{E}. \tag{3.9c}$$

Equation (3.9c) may also be written with the help of reper matrices

$$\mathbf{U}_\nu \mathbf{V}_\nu{}^T = \mathbf{E}. \tag{3.10}$$

One can show the orthogonality of orbitals (3.6b) and (3.7a). Really,

$$\langle \overline{\mathbf{b};\tilde{\mathbf{n}},\nu,\mathbf{v}_\nu} | \mathbf{b};\mathbf{n},\nu,\mathbf{u}_\nu \rangle \sim \pi^{-3/2} D(\mathbf{n},\boldsymbol{\tau}_\nu) D(\mathbf{n},\boldsymbol{\sigma}_\nu) \times \int d\mathbf{x} \exp\{ -(\mathbf{U}_\nu\boldsymbol{\tau}_\nu)^2 - (\mathbf{V}_\nu\boldsymbol{\sigma}_\nu)^2 + 2\mathbf{x}^T(\mathbf{U}_\nu\boldsymbol{\tau}_\nu + \mathbf{V}_\nu\boldsymbol{\sigma}_\nu) - \mathbf{x}^2\}|_{\boldsymbol{\tau}_\nu = \boldsymbol{\sigma}_\nu = 0}.$$

Integrating and utilizing Eqs. (3.8), one gets

$$\langle \overline{\mathbf{b};\tilde{\mathbf{n}},\nu,\mathbf{u}_\nu} | \mathbf{b};\mathbf{n},\nu,\mathbf{u}_\nu \rangle \sim \delta_{n_1\tilde{n}_2}\delta_{n_2\tilde{n}_2}\delta_{n_3\tilde{n}_3}.$$

The single-particle spatial density matrix built on Slater determinants of the orthogonal orbitals (3.6b) and (3.7a) becomes additive with respect to single-particle states:

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2) = \sum_{\mathbf{n}_\nu} D(\mathbf{n}_\nu,\boldsymbol{\tau}_\nu) D(\mathbf{n}_\nu,\boldsymbol{\sigma}_\nu) \times \rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\tau}_\nu,\boldsymbol{\sigma}_\nu)|_{\boldsymbol{\tau}_\nu=\boldsymbol{\sigma}_\nu=0}, \tag{3.11a}$$

where $\mathbf{n}_\nu$ are the quantum numbers of the occupied single-particle orbitals of the subsystem ν,

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\tau}_\nu,\boldsymbol{\sigma}_\nu) = \pi^{-3/2}\exp\{ -(\mathbf{U}_\nu\boldsymbol{\tau}_\nu)^2 - (\mathbf{V}_\nu\boldsymbol{\sigma}_\nu)^2 + 2\mathbf{x}_1^T\mathbf{U}_\nu\boldsymbol{\tau}_\nu + 2\mathbf{x}_2^T\mathbf{V}_\nu\boldsymbol{\sigma}_\nu - \tfrac{1}{2}(\mathbf{x}_1^2 + \mathbf{x}_2^2)\}. \tag{3.11b}$$

The partial expansion generating function (3.11b) of the density matrix of the subsystem ν contains six generating parameters. Obviously, the generating functions of direct and exchange integrals obtained by substituting (3.11b) into (3.2b) and (3.2c) transform into 12 parametrical ones. Explicit integration taking (3.6) into account leads to

$$W^{\pm}_{\nu\tilde{\nu}}(\gamma) = \frac{|\mathbf{B}|^{1/2}}{|\mathbf{B}+\gamma\mathbf{E}|^{1/2}} \times \sum_{\mathbf{n}_\nu}\sum_{\mathbf{n}_{\tilde{\nu}}} D(\mathbf{n}_\nu,\boldsymbol{\tau}_\nu)D(\mathbf{n}_\nu,\boldsymbol{\sigma}_\nu)D(\mathbf{n}_{\tilde{\nu}}\boldsymbol{\sigma}_{\tilde{\nu}})D(\mathbf{n}_{\tilde{\nu}},\boldsymbol{\tau}_{\tilde{\nu}}) \times \exp(-\mathbf{Y}^T_{\nu\tilde{\nu}}\mathbf{Q}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}}), \tag{3.12a}$$

where $\mathbf{Y}_{\nu\tilde{\nu}}$ is a block vector

$$\mathbf{Y}_{\nu\tilde{\nu}} = \begin{pmatrix} \mathbf{U}_\nu\boldsymbol{\tau}_\nu \\ \mathbf{V}_\nu\boldsymbol{\sigma}_\nu \\ \mathbf{U}_{\tilde{\nu}}\boldsymbol{\tau}_{\tilde{\nu}} \\ \mathbf{V}_{\tilde{\nu}}\boldsymbol{\sigma}_{\tilde{\nu}} \end{pmatrix}, \tag{3.12b}$$

$\mathbf{Q}^{\pm}$ are the block matrices

$$\mathbf{Q}^+ = \begin{pmatrix} \mathbf{q} & \mathbf{q}-\mathbf{E} & -\mathbf{q} & -\mathbf{q} \\ \mathbf{q}-\mathbf{E} & \mathbf{q} & -\mathbf{q} & -\mathbf{q} \\ -\mathbf{q} & -\mathbf{q} & \mathbf{q} & \mathbf{q}-\mathbf{E} \\ -\mathbf{q} & -\mathbf{q} & \mathbf{q}-\mathbf{E} & \mathbf{q} \end{pmatrix},$$
$$Q^- = \begin{pmatrix} \mathbf{q} & -\mathbf{q} & -\mathbf{q} & \mathbf{q}-\mathbf{E} \\ -\mathbf{q} & \mathbf{q} & \mathbf{q}-\mathbf{E} & -\mathbf{q} \\ -\mathbf{q} & \mathbf{q}-\mathbf{E} & \mathbf{q} & -\mathbf{q} \\ \mathbf{q}-\mathbf{E} & -\mathbf{q} & -\mathbf{q} & \mathbf{q} \end{pmatrix}, \tag{3.12c}$$

$$\mathbf{q} = \tfrac{1}{2}\gamma(\mathbf{B}+\gamma\mathbf{E})^{-1}. \tag{3.12d}$$

To derive (3.12), the well-known integral[32] was applied:

$$\int d\mathbf{y}\, e^{-\mathbf{y}^T\mathbf{C}\mathbf{y} + 2\mathbf{z}^T\mathbf{y}} = \left(\frac{\pi^n}{|\mathbf{C}|}\right)^{1/2} \exp(\mathbf{z}^T\mathbf{C}^{-1}\mathbf{z}), \tag{3.13}$$

where $\mathbf{C}$ is a positively defined symmetrical matrix of nth order, and $|\cdots|$ are the determinant bars. Taking (2.5b), (2.12a), and (2.17b) into account, one notes that

$$\frac{|\mathbf{B}|}{|\mathbf{B}+\gamma\mathbf{E}|} = \frac{\Delta}{(1+\gamma)^3\Delta(\gamma)}. \tag{3.14}$$

Equations (3.12) and (3.14) generalize the result of Filippov *et al.*[7] for the SU(3) model with regular shell occupation to the extended Sp(6,R) model. The presence of 12 generating parameters makes the differentiation of the exponential generating function on the right-hand side of (3.12a) sufficiently complicate. It seems expedient to transform the expression obtained to reduce the number of generating parameters. The next section deals with such a program.

IV. POTENTIAL ENERGY MATRIX GENERATING KERNELS. DETERMINANT GENERATING FUNCTION FOR PARTIAL EXPANSION

As one sees by analyzing formula (3.11b) for the partial expansion generating function, the possibility to occupy each of three Cartesian axes with oscillator quanta is provided by two generating parameters. We define a new generating function so that the kth axis occupation numbers $n_{\nu k}$ are generated only with the help of one parameter $g_{\nu k}$, $k = \overline{1,3}$,

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\mathbf{G}_\nu)=\sum_n g_{\nu 1}{}^{n_1} g_{\nu 2}{}^{n_2} g_{\nu_3}{}^{n_3}$$
$$\times D(\mathbf{n},\boldsymbol{\tau}_\nu)D(\mathbf{n},\boldsymbol{\sigma}_\nu)$$
$$\times\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\tau}_\nu,\boldsymbol{\sigma}_\nu)\big|_{\boldsymbol{\tau}_\nu=\boldsymbol{\sigma}_\nu=0}, \quad (4.1)$$

with arbitrary non-negative integer numbers n_1,n_2,n_3. Here the generating parameters are collected not in vectors, as in the previous discussion, but in diagonal matrices

$$\mathbf{G}_\nu=\mathrm{diag}(g_{\nu 1},g_{\nu 2},g_{\nu 3}).$$

Using formula (3.3b), we rewrite (4.1) as follows:

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\mathbf{G}_\nu)=\exp\left\{\frac{1}{2}\left(\frac{\partial}{\partial\boldsymbol{\tau}_\nu}\right)^T\mathbf{G}_\nu\frac{\partial}{\partial\boldsymbol{\sigma}_\nu}\right\}$$
$$\times\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\tau}_\nu,\boldsymbol{\sigma}_\nu)\big|_{\boldsymbol{\tau}_\nu=\boldsymbol{\sigma}_\nu=0}. \quad (4.2)$$

On the right-hand side of (4.2) we denote such variables as $\boldsymbol{\xi},\boldsymbol{\eta}$:

$$\boldsymbol{\tau}_\nu=\tfrac{1}{3}(\boldsymbol{\xi}+\boldsymbol{\eta}),\quad \boldsymbol{\sigma}_\nu=\tfrac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\eta}),$$

and define the vector differential operator

$$\hat{\mathbf{z}}=\frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{G}_\nu^{1/2}\dfrac{\partial}{\partial\boldsymbol{\xi}}\\ i\mathbf{G}_\nu^{1/2}\dfrac{\partial}{\partial\boldsymbol{\eta}}\end{pmatrix}.$$

Then (4.2) passes into

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\mathbf{G}_\nu)$$
$$=e^{\hat{\mathbf{z}}^2}\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\xi}+\boldsymbol{\eta}/2,\ \boldsymbol{\xi}-\boldsymbol{\eta}/2)\big|_{\boldsymbol{\xi}=\boldsymbol{\eta}=0}. \quad (4.3)$$

One can deduce from (3.13) if $\mathbf{C}=\mathbf{E}$

$$e^{\hat{\mathbf{z}}^2}=\pi^{-n/2}\int d\mathbf{y}\exp(-\mathbf{y}^2+2\mathbf{y}^T\mathbf{z}). \quad (4.4)$$

The operator transformation (4.4) is justified when the operator $\hat{\mathbf{z}}$ is restricted with respect to the norm in the investigated function space. In our case the functions (3.11b) are analytic, and the transformation (4.4) really takes place. Performing it on the right side of (4.3) and substituting the auxiliary vector $\mathbf{y}$ as a block

$$\mathbf{y}=\begin{pmatrix}\boldsymbol{\xi}'\\ \boldsymbol{\eta}'\end{pmatrix},$$

one obtains

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\mathbf{G}_\nu)=\pi^{-3}\int\int d\boldsymbol{\xi}'\,d\boldsymbol{\eta}'\exp\left\{-\boldsymbol{\xi}'^2-\boldsymbol{\eta}'^2-\left[\frac{1}{\sqrt{2}}\mathbf{U}_\nu\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'+i\boldsymbol{\eta}')\right]^2-\left[\frac{1}{\sqrt{2}}\mathbf{V}_\nu\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'-i\boldsymbol{\eta}')\right]^2\right.$$
$$\left.+2\mathbf{x}_1^T\left[\frac{1}{\sqrt{2}}\mathbf{U}_\nu\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'+i\boldsymbol{\eta}')\right]+2\mathbf{x}_2^T\left[\frac{1}{\sqrt{2}}\mathbf{V}_\nu\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'-i\boldsymbol{\eta}')\right]-\frac{1}{2}(\mathbf{x}_1^2+\mathbf{x}_2^2)\right\}.$$

Returning to the variables

$$\boldsymbol{\tau}_\nu=\frac{1}{\sqrt{2}}\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'+i\boldsymbol{\eta}'),\quad \boldsymbol{\sigma}_\nu=\frac{1}{\sqrt{2}}\mathbf{G}_\nu^{1/2}(\boldsymbol{\xi}'-i\boldsymbol{\eta}'),$$

we write down the result for the generating function

$$\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\mathbf{G}_\nu)=\left(\frac{i}{\pi}\right)^3(g_{\nu 1}g_{\nu 2}g_{\nu 3})^{-1}$$
$$\times\int\int d\boldsymbol{\tau}_\nu\,d\boldsymbol{\sigma}_\nu\exp(-2\boldsymbol{\sigma}_\nu^T\mathbf{G}_\nu^{-1}\boldsymbol{\tau}_\nu)$$
$$\times\rho_\nu(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\tau}_\nu,\boldsymbol{\sigma}_\nu). \quad (4.5)$$

The generating functions (4.1) and (3.11b) happened to be connected by the integral transformation (4.5), excluding the superfluous generating parameters. By analogy to (4.1), we define, on the basis of (3.12a), the generating function for direct and exchange integrals

$$W^\pm(\gamma;\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})=\frac{|\mathbf{B}|^{1/2}}{|\mathbf{B}+\gamma\mathbf{E}|^{1/2}}\sum_{\mathbf{n}}\sum_{\tilde{\mathbf{n}}}\prod_k g_{\nu k}{}^{n_k}g_{\tilde\nu k}{}^{\tilde n_k}$$
$$\times D(\mathbf{n},\boldsymbol{\tau}_\nu)D(\mathbf{n},\boldsymbol{\sigma}_\nu)D(\tilde{\mathbf{n}},\boldsymbol{\tau}_{\tilde\nu})D(\tilde{\mathbf{n}},\boldsymbol{\sigma}_{\tilde\nu})$$
$$\times\exp(-\mathbf{Y}_{\nu\tilde\nu}^T\mathbf{Q}^\pm\mathbf{Y}_{\nu\tilde\nu}).$$

Then, using (4.5), (3.10), (3.12), and (3.13), we can write

$$W^\pm(\gamma;\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})$$
$$=-\frac{|\mathbf{B}|^{1/2}}{|\mathbf{B}+\gamma\mathbf{E}|^{1/2}|\mathbf{S}_{\nu\tilde\nu}^T\mathbf{Q}^\pm(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})\mathbf{S}_{\nu\tilde\nu}|^{1/2}|\mathbf{G}_\nu\mathbf{G}_{\tilde\nu}|}, \quad (4.6a)$$

where $\mathbf{S}_{\nu\tilde\nu}$ is a quasidiagonal matrix

$$\mathbf{S}_{\nu\tilde\nu}=\begin{pmatrix}\mathbf{U}_\nu&&&\\&\mathbf{V}_\nu&&\\&&\mathbf{U}_{\tilde\nu}&\\&&&\mathbf{V}_{\tilde\nu}\end{pmatrix}, \quad (4.6b)$$

$\mathbf{Q}^\pm(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})$ is a block matrix [see (3.12c)],

$$\mathbf{Q}^\pm(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})$$
$$=\mathbf{Q}^\pm+\begin{pmatrix}\mathbf{O}&\boldsymbol{\Gamma}_\nu{}^{-T}&\mathbf{O}&\mathbf{O}\\ \boldsymbol{\Gamma}_\nu{}^{-1}&\mathbf{O}&\mathbf{O}&\mathbf{O}\\ \mathbf{O}&\mathbf{O}&\mathbf{O}&\boldsymbol{\Gamma}_{\tilde\nu}{}^{-T}\\ \mathbf{O}&\mathbf{O}&\boldsymbol{\Gamma}_{\tilde\nu}{}^{-1}&\mathbf{O}\end{pmatrix}, \quad (4.6c)$$

$$\boldsymbol{\Gamma}_\nu=\mathbf{U}_\nu\mathbf{G}_\nu\mathbf{V}_\nu{}^T,\quad \boldsymbol{\Gamma}_{\tilde\nu}=\mathbf{U}_{\tilde\nu}\mathbf{G}_{\tilde\nu}\mathbf{V}_{\tilde\nu}{}^T. \quad (4.6d)$$

Taking into account [see (3.10)] that

$$|\mathbf{S}_{\nu\tilde\nu}|=1,$$

we focus on the transformation of the determinants $|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})|$ and $|\mathbf{Q}^-(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})|$. Really, these determinants have the 12th order, while the other ones in Eq. (4.6a) have the third order, in accordance with the physical space dimension. We use the $|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})|$ example to consider a possible way to decrease the order of the determinants of interest. The initial form, as one can see from (3.12) and (4.6c), is

$$|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})| = \begin{vmatrix} \mathbf{q} & \mathbf{q}-\mathbf{E}+\mathbf{\Gamma}_\nu^{-T} & -\mathbf{q} & -\mathbf{q} \\ \mathbf{q}-\mathbf{E}+\mathbf{\Gamma}_\nu^{-1} & \mathbf{q} & -\mathbf{q} & -\mathbf{q} \\ -\mathbf{q} & -\mathbf{q} & -\mathbf{q} & \mathbf{q}-\mathbf{E}+\mathbf{\Gamma}_{\tilde\nu}^{-T} \\ -\mathbf{q} & -\mathbf{q} & \mathbf{q}-\mathbf{E}+\mathbf{\Gamma}_{\tilde\nu}^{-1} & \mathbf{q} \end{vmatrix}. \tag{4.7a}$$

With the help of linear operations on block lines and columns that leave the determinant value unchanged, one transforms (4.7a), for example, to

$$|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})| = \begin{vmatrix} 2\mathbf{E}-\mathbf{\Gamma}_\nu^{-T}-\mathbf{\Gamma}_{\tilde\nu}^{-T} & \mathbf{O} & \mathbf{E}-\mathbf{\Gamma}_{\tilde\nu}^{-T} & \mathbf{O} \\ \mathbf{O} & 2\mathbf{E}-\mathbf{\Gamma}_\nu^{-1}-\mathbf{\Gamma}_{\tilde\nu}^{-1} & \mathbf{O} & \mathbf{E}-\mathbf{\Gamma}_{\tilde\nu}^{-1} \\ \mathbf{E}-\mathbf{\Gamma}_\nu^{-T} & \mathbf{O} & \mathbf{q} & \mathbf{q} \\ \mathbf{O} & \mathbf{E}-\mathbf{\Gamma}_\nu^{-1} & \mathbf{q} & \mathbf{q} \end{vmatrix}. \tag{4.7b}$$

The form (4.7b) is fitted better than (4.7) to the Schur's algorithm reducing the calculation of the $2n$th-order determinant to nth-order ones[33]

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|. \tag{4.8}$$

Using formula (4.8) and taking identity

$$(\mathbf{E}-\mathbf{\Gamma}_\nu^{-1})(2\mathbf{E}-\mathbf{\Gamma}_\nu^{-1}-\mathbf{\Gamma}_{\tilde\nu}^{-1})(\mathbf{E}-\mathbf{\Gamma}_{\tilde\nu}^{-1}) = -(\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T)^{-1}, \tag{4.9a}$$

$$\mathbf{H}_\nu = \mathbf{G}_\nu(\mathbf{E}-\mathbf{G}_\nu)^{-1} \tag{4.9b}$$

into account, we obtain a sixth-order determinant

$$\begin{aligned} &|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})| \\ &= |2\mathbf{E}-\mathbf{\Gamma}_\nu^{-1}-\mathbf{\Gamma}_{\tilde\nu}^{-1}||2\mathbf{E}-\mathbf{\Gamma}_\nu^{-1}-\mathbf{\Gamma}_{\tilde\nu}^{-1}| \\ &\times\begin{vmatrix} (\mathbf{V}_\nu\mathbf{H}_\nu\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{U}_{\tilde\nu}^T)^{-1} & \mathbf{q} \\ -(\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T)^{-1}-(\mathbf{V}_\nu\mathbf{H}_\nu\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{U}_{\tilde\nu}^T)^{-1} & (\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T)^{-1} \end{vmatrix}. \end{aligned} \tag{4.10}$$

Application of formula (4.8) together with the identity

$$\begin{aligned} &|2\mathbf{E}-\mathbf{\Gamma}_\nu^{-1}-\mathbf{\Gamma}_{\tilde\nu}^{-1}| \\ &\quad = -|\mathbf{G}_\nu|^{-1}|\mathbf{G}_{\tilde\nu}|^{-1}|\mathbf{E}-\mathbf{G}_\nu||\mathbf{E}-\mathbf{G}_{\tilde\nu}| \\ &\qquad\times|\mathbf{V}_{\tilde\nu}{}^T\mathbf{U}_\nu\mathbf{H}_\nu+\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}{}^T\mathbf{U}_\nu| \end{aligned}$$

transfers (4.10) into

$$\begin{aligned} &|\mathbf{Q}^+(\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu})| \\ &\quad = (-1)^2|\mathbf{G}_\nu|^2|\mathbf{G}_{\tilde\nu}|^2|\mathbf{E}-\mathbf{G}_\nu|^2|\mathbf{E}-\mathbf{G}_{\tilde\nu}|^2 \\ &\qquad\times|\mathbf{E}+(\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T)\mathbf{q} \\ &\qquad+(\mathbf{V}_\nu\mathbf{H}_\nu\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{U}_{\tilde\nu}^T)\mathbf{q}|. \end{aligned} \tag{4.11}$$

Substituting (4.11) into (4.6a) and remembering a definition (3.12d), we reach the following result:

$$\begin{aligned} &W^+(\gamma;\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu}) \\ &= \frac{|\mathbf{B}|^{1/2}}{|\mathbf{E}-\mathbf{G}_\nu||\mathbf{E}-\mathbf{G}_{\tilde\nu}|} \\ &\quad\times|\mathbf{B}+\gamma\mathbf{E}+\frac{\gamma}{2}(\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T) \\ &\quad+\frac{\gamma}{2}(\mathbf{V}_\nu\mathbf{H}_\nu\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{U}_{\tilde\nu}^T)|^{-1/2}. \end{aligned} \tag{4.12}$$

Using the generating function for Laguerre polynomials

$$\frac{1}{1-g}\exp\left(-\frac{gz}{1-g}\right) = \sum_{n=0}^{\infty} g^n L_n(z),$$

we rewrite (4.12) in the form of

$$\begin{aligned} &W^+(\gamma;\mathbf{G}_\nu,\mathbf{G}_{\tilde\nu}) \\ &\quad = |\mathbf{B}|^{1/2}\sum_{\mathbf{n}}\sum_{\tilde{\mathbf{n}}}\prod_k g_{\nu k}{}^{n_k}g_{\tilde\nu k}{}^{\tilde n_k} \\ &\qquad\times L_{n_k}\left(-\frac{\partial}{\partial H_{\nu k}}\right)L_{\tilde n_k}\left(-\frac{\partial}{\partial H_{\tilde\nu k}}\right) \\ &\qquad\times|\mathbf{B}+\gamma\mathbf{E}+\frac{\gamma}{2}(\mathbf{U}_\nu\mathbf{H}_\nu\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{V}_{\tilde\nu}^T) \\ &\qquad+\frac{\gamma}{2}(\mathbf{V}_\nu\mathbf{H}_\nu\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\mathbf{H}_{\tilde\nu}\mathbf{U}_{\tilde\nu}^T)|^{-1/2}|_{\mathbf{H}_\nu=\mathbf{H}_{\tilde\nu}=\mathbf{O}}. \end{aligned} \tag{4.13}$$

In Eq. (4.13) $H_{\nu k}$ and $H_{\tilde\nu k}$ (the elements of diagonal matrices $\mathbf{H}_\nu$ and $\mathbf{H}_{\tilde\nu}$) are treated as independent variables, in contrast to Eq. (4.12) where these symbols denoted the definite functions (4.9b) of the parameters $g_{\nu k}$ and $g_{\tilde\nu k}$.

Comparing (4.13) with (3.12a), one gets the desired formula for direct integrals

$$\begin{aligned} &W^+_{\nu\tilde\nu}(\gamma) \\ &\quad = |\mathbf{B}|^{1/2}\sum_{\mathbf{n}_\nu}\sum_{\mathbf{n}_{\tilde\nu}}\prod_k L_{n_{\nu k}}\left(-\frac{\partial}{\partial H_k}\right)L_{n_{\tilde\nu k}}\left(-\frac{\partial}{\partial \tilde H_k}\right) \\ &\qquad\times|\mathbf{B}+\frac{\gamma}{2}(\mathbf{E}+\mathbf{U}_\nu\mathbf{H}\mathbf{V}_\nu^T+\mathbf{U}_{\tilde\nu}\tilde{\mathbf{H}}\mathbf{V}_{\tilde\nu}^T) \\ &\qquad+\frac{\gamma}{2}(\mathbf{E}+\mathbf{V}_\nu\mathbf{H}\mathbf{U}_\nu^T+\mathbf{V}_{\tilde\nu}\tilde{\mathbf{H}}\mathbf{U}_{\tilde\nu}^T)|^{-1/2}|_{\mathbf{H}=\tilde{\mathbf{H}}=\mathbf{O}}. \end{aligned} \tag{4.14a}$$

The exchange integrals need a more complicated expression:

$$W_{\nu\tilde{\nu}}^{-}(\gamma) = |B|^{1/2} \sum_{\mathbf{n}_\nu} \sum_{\mathbf{n}_{\tilde{\nu}}} \prod_k L_{n_{\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big) L_{n_{\tilde{\nu}k}}\Big(-\frac{\partial}{\partial \tilde{H}_k}\Big)$$
$$\times |(\mathbf{E} + \mathbf{U}_\nu \mathbf{H} \mathbf{V}_\nu{}^T + \mathbf{U}_{\tilde{\nu}} \tilde{\mathbf{H}} \mathbf{V}_{\tilde{\nu}}{}^T)$$
$$\times (\mathbf{E} + \mathbf{V}_\nu \mathbf{H} \mathbf{U}_\nu{}^T + \mathbf{V}_{\tilde{\nu}} \tilde{\mathbf{H}} \mathbf{U}_{\tilde{\nu}}{}^T)|^{-1/2}$$
$$\times |\mathbf{B} + \frac{\gamma}{2}(\mathbf{E} + \mathbf{U}_\nu \mathbf{H} \mathbf{V}_\nu{}^T + \mathbf{U}_{\tilde{\nu}} \tilde{\mathbf{H}} \mathbf{V}_{\tilde{\nu}}{}^T)^{-1}$$
$$+ \frac{\gamma}{2}(\mathbf{E} + \mathbf{V}_{\tilde{\nu}} \tilde{\mathbf{H}} \mathbf{U}_{\tilde{\nu}}{}^T + \mathbf{V}_\nu \mathbf{H} \mathbf{U}_\nu{}^T)^{-1}|^{-1/2}|_{\mathbf{H} = \tilde{\mathbf{H}} = \mathbf{O}}. \tag{4.14b}$$

In Eqs. (4.14), we introduce the symbols for generating parameters simplified as compared to (4.13). One can write these equations in another manner using the projection matrices (3.9)

$$W_{\nu\tilde{\nu}}^{+}(\gamma) = |\mathbf{B}|^{1/2} \sum_{\mathbf{n}_\nu} \sum_{\mathbf{n}_{\tilde{\nu}}} \prod_k L_{n_{\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big) L_{n_{\tilde{\nu}k}}\Big(-\frac{\partial}{\partial \tilde{H}_k}\Big)$$
$$\times \left|\mathbf{B} + \frac{\gamma}{2} \sum_l \Big[\Big(\frac{1}{2} + H_l\Big)(\mathbf{P}_{\nu l} + \mathbf{P}_{\nu l}{}^T) + \Big(\frac{1}{2} + \tilde{H}_l\Big)(\mathbf{P}_{\tilde{\nu} l} + \mathbf{P}_{\tilde{\nu} l}{}^T)\Big]\right|^{-1/2} \Bigg|_{H_k = \tilde{H}_k = 0}, \tag{4.15a}$$

$$W_{\nu\tilde{\nu}}^{-}(\gamma) = |\mathbf{B}|^{1/2} \sum_{\mathbf{n}_\nu} \sum_{\mathbf{n}_{\tilde{\nu}}} \prod_k L_{n_{\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big) L_{n_{\tilde{\nu}k}}\Big(-\frac{\partial}{\partial \tilde{H}_k}\Big)$$
$$\times \left|\sum_l \Big[\Big(\frac{1}{2} + H_l\Big)\mathbf{P}_{\nu l} + \Big(\frac{1}{2} + \tilde{H}_l\Big)\mathbf{P}_{\tilde{\nu} l}\Big]\right|^{-1/2} \left|\sum_l \Big[\Big(\frac{1}{2} + H_l\Big)\mathbf{P}_{\nu l}{}^T + \Big(\frac{1}{2} + \tilde{H}_l\Big)\mathbf{P}_{\tilde{\nu} l}{}^T\Big]\right|^{-1/2}$$
$$\times \left|\mathbf{B} + \frac{\gamma}{2}\Big\{\sum_l \Big[\Big(\frac{1}{2} + H_l\Big)\mathbf{P}_{\nu l} + \Big(\frac{1}{2} + \tilde{H}_l\Big)\mathbf{P}_{\tilde{\nu} l}\Big]\Big\}^{-1}\right.$$
$$\left. + \frac{\gamma}{2}\Big\{\sum_l \Big[\Big(\frac{1}{2} + H_l\Big)\mathbf{P}_{\nu l}{}^T + \Big(\frac{1}{2} + \tilde{H}_l\Big)\mathbf{P}_{\tilde{\nu} l}{}^T\Big]\Big\}^{-1}\right|^{-1/2} \Bigg|_{H_k = \tilde{H}_k = 0}. \tag{4.15b}$$

Equations (2.15a), (3.2a), and (4.15) solve the problem of constructing the generating kernels of the Wigner interaction matrix in the extended symplectic model. As one can see from (4.15), the sought for generating kernels are obtained by the differentiation of the corresponding determinant generating function of partial expansion with respect to, generally speaking, six parameters. (In the case of identical subsystems ν and $\tilde{\nu}$, three independent generating parameters are sufficient, as will be shown in Sec. VI.) Equations (4.15) are much simpler than (3.12) with the exponential generating function: moreover, their determinant form makes them similar to (2.22) and (2.24). The similarity becomes more manifest if one notes that "extra" vectors $\mathbf{U}_{\nu 2}$ and $\mathbf{V}_{\nu 2}$ [see (3.8a)] always may be excluded from (4.15) by means of unit expansion (3.9c), and the matrices $\mathbf{P}_{\nu 1}$ and $\mathbf{P}_{\nu 3}$ may be expressed using the initial generator coordinates and the blocks (2.15):

$$\mathbf{P}_{\nu 1} = \frac{\Delta}{\mathscr{M}_\nu} \mathbf{B}^{-1/2} \mathbf{u}_{\nu 1} \mathbf{v}_{\nu 1}^T \mathbf{B}^{-1/2},$$

$$\mathbf{P}_{\nu 3} = \frac{|E - b||E - \tilde{b}|}{\mathscr{K}_\nu} \mathbf{B}^{1/2} \mathbf{v}_{\nu 3} \mathbf{u}_{\nu 3}^T \mathbf{B}^{1/2}.$$

The correction of Eqs. (4.15) for potentials with spin–isospin dependence and their subsequent analysis in the limiting cases of physical interest will be presented in Sec. VI. And now we proceed to generalize the result of the present section on static nucleon interactions possessing no spherical symmetry.

V. POTENTIAL ENERGY MATRIX GENERATING KERNELS. TENSOR INTERACTION

Let us consider a two-body tensor interaction operator in the A nucleon system

$$\hat{U}_t = \sum_{i<j\leqslant A} \hat{V}(\mathbf{r}_i - \mathbf{r}_j)\hat{S}_{ij}, \tag{5.1a}$$

$$\hat{S}_{ij} = 3\frac{(\mathbf{S}_i, \mathbf{r}_i - \mathbf{r}_j)(\mathbf{S}_j, \mathbf{r}_i - \mathbf{r}_j)}{(\mathbf{r}_i - \mathbf{r}_j)^2} - (\mathbf{S}_i, \mathbf{S}_j)$$
$$= \frac{1}{2}\left[3\frac{(\mathbf{S}_i + \mathbf{S}_j, \mathbf{r}_i - \mathbf{r}_j)^2}{(\mathbf{r}_i - \mathbf{r}_j)^2} - (\mathbf{S}_i + \mathbf{S}_j)^2\right], \tag{5.1b}$$

where $V(\mathbf{r}_i - \mathbf{r}_j)$ are the two-body potentials [see (3.1)], $\mathbf{S}_i$ and $\mathbf{S}_j$ are the spin operators of ith and jth nucleons. Transfering the correction on isospin dependence to Sec. VI, here we deal only with the contribution of unit spin nucleon pairs. In accordance with the radial dependence of the potential (3.1), we define direct and exchange integrals as follows:

$$W_{t,\nu\tilde{\nu}}^{+}(\gamma) = -\frac{3}{4}\Big(\mathbf{S}(1)\frac{\partial}{\partial \mathbf{k}}\Big)^2 \int_\gamma^\infty d\gamma' \int\int d\mathbf{r}_1\, d\mathbf{r}_2$$
$$\times \rho_\nu(\mathbf{r}_1, \mathbf{r}_1)\rho_{\tilde{\nu}}(\mathbf{r}_2, \mathbf{r}_2)$$
$$\times \exp\Big[-\frac{\gamma'}{2}(\mathbf{r}_1 - \mathbf{r}_2)^2 + i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2)\Big]\Big|_{\mathbf{k}=0}$$
$$-\frac{1}{2}(\mathbf{S}(1))^2 W_{\nu\tilde{\nu}}^{+}(\gamma), \tag{5.2a}$$

$$W^{-}_{t,\nu\tilde{\nu}}(\gamma) = -\frac{3}{4}\left(\mathbf{S}(1)\frac{\partial}{\partial \mathbf{k}}\right)^2 \int_\gamma^\infty d\gamma' \int\int d\mathbf{r}_1\, d\mathbf{r}_2 \times \rho_\nu(\mathbf{r}_1,\mathbf{r}_2)\rho_{\tilde{\nu}}(\mathbf{r}_2,\mathbf{r}_1) \times \exp\left[-\frac{\gamma'}{2}(\mathbf{r}_1-\mathbf{r}_2)^2 + i\mathbf{k}(\mathbf{r}_1-\mathbf{r}_2)\right]\Bigg|_{\mathbf{k}=0} - \tfrac{1}{4}(\mathbf{S}(1))^2 W^{+}_{\nu\tilde{\nu}}(\gamma), \quad (5.2\text{b})$$

where $\mathbf{S}(1)$ is a unit spin operator, and $\mathbf{k}$ is the independent vector generating parameter. The introduction of parameter $\mathbf{k}$ (Ref. 34) allows us to generalize the results of Secs. III and IV to spherically nonsymmetric potentials.

The problem (5.2) will be solved if one finds analytic expressions for the integrals

$$W^{+}_{\nu\tilde{\nu}}(\gamma,\mathbf{k}) = \int\int d\mathbf{r}_1\, d\mathbf{r}_2\, \rho_\nu(\mathbf{r}_1,\mathbf{r}_1)\rho_\nu(\mathbf{r}_2,\mathbf{r}_2) \times \exp\left[-\frac{\gamma}{2}(\mathbf{r}_1-\mathbf{r}_2)^2 + i\mathbf{k}(\mathbf{r}_1-\mathbf{r}_2)\right], \quad (5.3\text{a})$$

$$W^{-}_{\nu\tilde{\nu}}(\gamma,\mathbf{k}) = \int\int d\mathbf{r}_1\, d\mathbf{r}_2\, \rho_\nu(\mathbf{r}_1,\mathbf{r}_2)\,\rho_\nu(\mathbf{r}_2,\mathbf{r}_1) \times \exp\left[-\frac{\gamma}{2}(\mathbf{r}_1-\mathbf{r}_2)^2 + i\mathbf{k}(\mathbf{r}_1-\mathbf{r}_2)\right], \quad (5.3\text{b})$$

precisely up to the second order of infinitesimal $\mathbf{k}$. Substituting in (5.3) the spatial density matrices represented as (3.11), one obtains

$$W^{\pm}_{\nu\tilde{\nu}}(\gamma,\mathbf{k}) = \frac{|\mathbf{B}|^{1/2}}{|\mathbf{B}+\gamma\mathbf{E}|^{1/2}} \sum_{n_\nu}\sum_{n_{\tilde{\nu}}} D(\mathbf{n}_\nu,\boldsymbol{\tau}_\nu)D(\mathbf{n}_\nu,\boldsymbol{\sigma}_\nu) \times D(\mathbf{n}_{\tilde{\nu}},\boldsymbol{\tau}_{\tilde{\nu}})D(\mathbf{n}_{\tilde{\nu}},\boldsymbol{\sigma}_{\tilde{\nu}}) \times \exp\left[-\mathbf{Y}^T_{\nu\tilde{\nu}}\mathbf{Q}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}} + i\mathbf{k}^T\mathbf{Z}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}} - \frac{1}{2}\mathbf{k}^T\mathbf{z}\mathbf{k}\right], \quad (5.4\text{a})$$

where $\mathbf{Z}^{\pm}$ are block row matrices

$$\mathbf{Z}^{\pm} = (\mathbf{z}, \pm\mathbf{z}, -\mathbf{z}, \mp\mathbf{z}), \quad (5.4\text{b})$$

$$\mathbf{z} = \mathbf{E} - 2\mathbf{q} = \mathbf{B}(\mathbf{B}+\gamma\mathbf{E})^{-1}. \quad (5.4\text{c})$$

Passing from the representation (5.4) which is based on the exponential generating function to the representation with the determinant generating function, one has to integrate over $\boldsymbol{\tau}_\nu, \boldsymbol{\sigma}_\nu, \boldsymbol{\tau}_{\tilde{\nu}}, \boldsymbol{\sigma}_{\tilde{\nu}}$ [see (4.5)]. The presence of a term linear in $\mathbf{Y}_{\nu\tilde{\nu}}$ within the subintegral exponent causes, according to (3.13), enormously complicated calculations of adjoint matrix to $\mathbf{Q}^{\pm}$.

However, it can be noted that to solve the stated problem, one does not need exact analytic expressions for the integrals $W^{\pm}_{\nu\tilde{\nu}}(\gamma,\mathbf{k})$. The necessary precision $O(|\mathbf{k}|^3)$ would be maintained, if within the exponent one replaces

$$i\mathbf{k}^T\mathbf{Z}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}} \to -\tfrac{1}{2}\mathbf{Y}^T_{\nu\tilde{\nu}}(\mathbf{Z}^{\pm})^T\mathbf{k}\mathbf{k}^T\mathbf{Z}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}}. \quad (5.5)$$

Really, the terms of odd order with respect to $\mathbf{k}$ disappear when integrated with an even weight function $\exp(-\mathbf{Y}^T_{\nu\tilde{\nu}}\mathbf{Q}^{\pm}\mathbf{Y}_{\nu\tilde{\nu}})$; and in the second order in $\mathbf{k}$ the Taylor series of the left- and right-hand side exponents of (5.5) coincide. By taking the structure of the matrices (3.12c) and (5.4b) into account one notices that an admissible expression for $\exp(\frac{1}{2}\mathbf{k}^T\mathbf{z}\mathbf{k})\, W^{\pm}_{\nu\tilde{\nu}}(\gamma,\mathbf{k})$ can be derived by replacing in (3.12a)

$$\mathbf{Q}^{\pm} \to \mathbf{Q}^{\pm}(\mathbf{k}),$$

where the matrices $\mathbf{Q}^{\pm}(\mathbf{k})$ are to be obtained from $\mathbf{Q}^{\pm}$ replacing blocks

$$\mathbf{q} \to \mathbf{q}(\mathbf{k}) = \mathbf{q} + \mathbf{z}\mathbf{k}\mathbf{k}^T\mathbf{z}.$$

Making an analogous replacement in formula (4.11), we generalize (4.15) to

$$W^{+}_{\nu\tilde{\nu}}(\gamma,\mathbf{k}) = \exp\left[-\frac{1}{2}\mathbf{k}^T\mathbf{B}(\mathbf{B}+\gamma\mathbf{E})^{-1}\mathbf{k}\right]|\mathbf{B}|^{1/2} \times \sum_{\mathbf{n}_\nu}\sum_{\mathbf{n}_{\tilde{\nu}}}\prod_k L_{n_{\nu k}}\left(-\frac{\partial}{\partial H_k}\right)L_{n_{\tilde{\nu}k}}\left(-\frac{\partial}{\partial \tilde{H}_k}\right)\Big|\mathbf{B}+\gamma\mathbf{E} + \frac{1}{2}\sum_l\left[H_l(\mathbf{P}_{\nu l}+\mathbf{P}^T_{\nu l}) + \tilde{H}_l(\mathbf{P}_{\tilde{\nu} l}+\mathbf{P}^T_{\tilde{\nu} l})\right]\left[\gamma\mathbf{E} + (\mathbf{B}+\gamma\mathbf{E})^{-1}\mathbf{B}\mathbf{k}\mathbf{k}^T\mathbf{B}\right]\Big|^{-1/2}\Big|_{H_k=\tilde{H}_k=0}, \quad (5.6\text{a})$$

$$W^{-}_{\nu\tilde{\nu}}(\gamma,\mathbf{k}) = \exp\left[-\frac{1}{2}\mathbf{k}^T\mathbf{B}(B+\gamma\mathbf{E})^{-1}\mathbf{k}\right]|\mathbf{B}|^{1/2} \times \sum_{n_\nu}\sum_{n_{\tilde{\nu}}}\prod_k \mathrm{L}_{n_{\nu k}}\left(-\frac{\partial}{\partial H_k}\right)L_{n_{\tilde{\nu}k}}\left(-\frac{\partial}{\partial \tilde{H}_k}\right)\left|\sum_l\left[\left(\frac{1}{2}+H_l\right)\mathbf{P}_{\nu l} + \left(\frac{1}{2}+\tilde{H}_l\right)\mathbf{P}_{\tilde{\nu} l}\right]\right|^{-1} \times \Bigg|\mathbf{B}+\gamma\mathbf{E} + \frac{1}{2}\left[\gamma\mathbf{E}+(\mathbf{B}+\gamma\mathbf{E})^{-1}\mathbf{B}\mathbf{k}\mathbf{k}^T\mathbf{B}\right]\Bigg\{\left[\sum_{l'}\left(\frac{1}{2}+H_{l'}\right)\mathbf{P}_{\nu l'} + \left(\frac{1}{2}+\tilde{H}_{l'}\right)\mathbf{P}_{\tilde{\nu} l'}\right]^{-1} + \left[\sum_{l'}\left(\frac{1}{2}+H_{l'}\right)\mathbf{P}^T_{\nu l'} + \left(\frac{1}{2}+\tilde{H}_{l'}\right)\mathbf{P}^T_{\tilde{\nu} l'}\right]^{-1} - 2\mathbf{E}\Bigg\}\Bigg|^{-1/2}\Bigg|_{H_k=\tilde{H}_k=0}. \quad (5.6\text{b})$$

As a matter of fact, we note that Eqs. (5.6) can be obtained from (4.15) with the same precision of $O(|\mathbf{k}|^3)$ by a simple substitution

$$\mathbf{B}\to\mathbf{B}-\frac{1}{\gamma}\mathbf{B}\mathbf{k}\mathbf{k}^T\mathbf{B}.$$

So the direct and exchange tensor interaction integrals are written as

$$W^{\pm}_{t,\nu\tilde\nu}(\gamma)=-\frac{3}{4}\Big(\mathbf{S}(1)\frac{\partial}{\partial\mathbf{k}}\Big)^2\int_\gamma^\infty d\gamma'\,W^{\pm}_{\nu\tilde\nu}(\gamma',\mathbf{k})|_{\mathbf{k}=0}$$
$$-\tfrac{1}{2}(\mathbf{S}(1))^2W^{\pm}_{\nu\tilde\nu}(\gamma),\qquad(5.7)$$

where the functions $W^{\pm}_{\nu\tilde\nu}(\gamma,\mathbf{k})$ have been defined in (5.6).

In the last section of the present paper the results of Secs. II, IV, and V are generalized to a more realistic situation when the potential magnitudes depend on the nucleon pair spin–isospin numbers; the cases of closed shells and regularly filled open shells are considered in the symplectic and unitary nuclear models.

VI. GENERATING KERNELS OF REALISTIC NUCLEAR HAMILTONIAN MATRIX

Let the nucleon potential of a realistic Hamiltonian take a form of

$$\hat V_c(r)=\sum_{S,T=0}^{1}w_c^{2S+1,2T+1}\exp\Big(-\frac{1}{2}\gamma_c\mathbf{r}^2\Big)\mathscr{P}^S\mathscr{P}^T,\qquad(6.1a)$$

$$\hat V_t(r)=\sum_{T=0}^{1}w_t^{2T+1}\exp\Big(-\frac{1}{2}\gamma_t\mathbf{r}^2\Big)\mathscr{P}^{T}\hat{S},\qquad(6.1b)$$

where S and T are the nucleon pair spin and isospin,

$$\hat S=\frac{3}{2}\frac{(\mathbf{S}\mathbf{r})^2}{\mathbf{r}^2}-\frac{1}{2}\mathbf{S}^2$$

is the tensor operator, $\mathscr{P}^S$ and $\mathscr{P}^T$ are the projection operators. The results of Secs. II, IV, and V allow us to infer that the investigated Hamiltonian (without Coulomb interaction) matrix elements between the generating invariants of the extended $Sp(6,R)$ model take the form of

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\hat H|\mathbf{b};\mathbf{u}\rangle=\langle\tilde{\mathbf{b}};\mathbf{v}|\hat T+\hat U_c+\hat U_t|\mathbf{b};\mathbf{u}\rangle,\qquad(6.2a)$$

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\hat U_c|\mathbf{b};\mathbf{u}\rangle=\langle\tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle\sum_g\sum_{\nu\tilde\nu}W^g_{\nu\tilde\nu}(\gamma_c)\times\sum_{S,T=0}^{1}[C^{S\sigma+\tilde\sigma}_{1/2\,\sigma\,1/2\,\tilde\sigma}C^{T\tau+\tilde\tau}_{1/2\,\tau\,1/2\,\tilde\tau}]^2\times g^{S+T+1}w_c^{2S+1,2T+1},\qquad(6.2b)$$

$$\langle\tilde{\mathbf{b}};\mathbf{v}|\hat U_t|\mathbf{b};\mathbf{u}\rangle=\langle\tilde{\mathbf{b}};\mathbf{v}|\mathbf{b};\mathbf{u}\rangle=\sum_g\sum_{\nu\tilde\nu}\langle S\sigma+\tilde\sigma|W^g_{t,\nu\tilde\nu}(\gamma_t)|S\sigma+\tilde\sigma\rangle|_{S=1}\times\Big(\frac{3}{4}+\sigma\tilde\sigma\Big)\sum_{T=0}^{1}[C^{T\tau+\tilde\tau}_{1/2\,\tau\,1/2\,\tilde\tau}]^2w_t^{2T+1},\qquad(6.2c)$$

where g is the interaction type symbol ($+1$ for the direct interaction, -1 for the exchange one), $\{\sigma\tau\}$ and $\{\tilde\sigma\tilde\tau\}$ are the spin and isospin projections of an individual nucleon for the subsystems ν and $\tilde\nu$, respectively, $C^{jm}_{j_1m_1j_2m_2}$ is the Wigner coefficient of the SU(2) group. The integrals $W^g_{\nu\tilde\nu}(\gamma)$ and $W^g_{t,\nu\tilde\nu}(\gamma)$ are defined by Eqs. (4.15) and (5.7). The matrix elements of the tensor interaction integrals between the unit spin basis states [see (6.2c)] may be calculated by means of the Wigner–Eckart theorem.[35] Equations (2.16) and (2.17) indicate the way to obtain the matrix elements of the kinetic energy operator $\hat T$.

The calculations of the characteristics of the concrete nucleon system on the full basis of the extended symplectic model are expected to be an extraordinarily difficult problem. Nevertheless, one can often meet some applications of physical interest (for example, studies on magical nuclei, "breathing," quadrupole, and "presession" modes in deformed nuclei, nuclei with mixed shell configurations in the unitary model, and so on) where only separate details of the immense world of nuclear motion are observed in their brightest, exposed form. In such cases it is sufficient to employ only a part of the available basis, then Eqs. (6.2) become simplified and make practical utilization easier.

(A) The simplest problem in the framework of shell treatment seems to be the description of nuclei with closed nucleon shells (light magic, etc.). The potential energy operator includes only the central interaction.

We make use of Eqs. (4.15). Then, using the summation formulas for generalized Laguerre polynomials,[36]

$$\sum_{\substack{k_1,k_2,k_3\geq 0\\k_1+k_2+k_3=n}}L_{k_1}(x_1)L_{k_2}(x_2)L_{k_3}(x_3)=L^2_n(x_1+x_2+x_3),\qquad(6.3a)$$

$$\sum_{k=0}^{n}L^\alpha_k(x)=L^{\alpha+1}_n(x),\qquad(6.3b)$$

we note that

$$\sum_{n_\nu}\sum_{n_{\tilde\nu}}\prod_k L_{n_{\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big)L_{n_{\tilde\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big)=L^3_{N_\nu}\Big(-\frac{\partial}{\partial H_1}-\frac{\partial}{\partial H_2}-\frac{\partial}{\partial H_3}\Big)\times L^3_{N_{\tilde\nu}}\Big(-\frac{\partial}{\partial H_1}-\frac{\partial}{\partial H_2}-\frac{\partial}{\partial H_3}\Big),\qquad(6.4)$$

where N_ν is the subsystem ν upper shell number connected with the total numbers of quanta by the relation

$$f_{\nu k}=\frac{(N_\nu+3)!}{(N_\nu-1)!3!},\quad k=\overline{1,3}.$$

Redefining the generating parameters

$$H=\tfrac{1}{2}+\tfrac{1}{3}(H_1+H_2+H_3),\qquad h_1=H_1-H_2,\ h_3=H_2-H_3,\qquad(6.5a)$$

one notes that

$$\frac{1}{2}\sum_l\Big(\frac{1}{2}+H_l\Big)(\mathbf{P}_{\nu l}+\mathbf{P}^T_{\nu l})=H\mathbf{E}+\tfrac{1}{2}h_1(\mathbf{P}_{\nu1}+\mathbf{P}^T_{\nu1})-\tfrac{1}{2}h_3(\mathbf{P}_{\nu3}+\mathbf{P}^T_{\nu3})-\tfrac{1}{3}(h_1-h_3)\mathbf{E},\qquad(6.5b)$$

$$\frac{\partial}{\partial H_1}+\frac{\partial}{\partial H_2}+\frac{\partial}{\partial H_3}=\frac{\partial}{\partial H}.\qquad(6.5c)$$

Substituting (6.4), (6.5b), (6.5c) into (4.15a), one obtains

$$W_{\nu\tilde{\nu}}^{+}(\gamma) = |\mathbf{B}|^{1/2} L_{N_\nu}^{3}\Big(-\frac{\partial}{\partial H}\Big) L_{N_{\tilde{\nu}}}^{3}\Big(-\frac{\partial}{\partial \tilde{H}}\Big) \times |\mathbf{B} + \gamma(H+\tilde{H})\mathbf{E}|^{-1/2}|_{H=\tilde{H}=1/2}\,.$$

Finally we pass to one generating parameter $H+\tilde{H}\to H$ and derive using (3.14):

$$W_{\nu\tilde{\nu}}^{+}(\gamma) = L_{N_\nu}^{3}\Big(-\frac{\partial}{\partial H}\Big) L_{N_{\tilde{\nu}}}^{3}\Big(-\frac{\partial}{\partial H}\Big) \times \Big(\frac{\Delta}{(1+\gamma H)^3\Delta(\gamma H)}\Big)^{1/2}\Big|_{H=1}\,. \tag{6.6a}$$

Similarly, an answer for the exchange integrals follows from (4.15b):

$$W_{\nu\tilde{\nu}}^{-}(\gamma) = L_{N_\nu}^{3}\Big(-\frac{\partial}{\partial H}\Big) L_{N_{\tilde{\nu}}}^{3}\Big(-\frac{\partial}{\partial H}\Big) \times \frac{1}{H^3}\Big(\frac{\Delta}{(1+\gamma/H)^3\Delta(\gamma/H)}\Big)^{1/2}\Big|_{H=1}\,. \tag{6.6b}$$

Equations (6.6) reproduce the result for closed nuclear shells in the Sp(6,R) model obtained by Vasilevsky *et al.*[5] It is seen that the Hamiltonian matrix elements in this case do not depend on the oscillator repers $\mathbf{u}_\nu$, $\mathbf{v}_\nu$ orientation, reflecting the symmetry inherent to the situation.

(B) The structure of the light and almost magic medium nuclei is often defined by the regularly occupied nucleon shells. An admissible quantitative description of such nuclei is achieved in the standard Sp(6,R) model. The Hamiltonian matrix elements on Sp(6,R) irrep coherent states can be obtained from corresponding formulas for the extended model by passing to reper matrices $\mathbf{U}$ and $\mathbf{V}$ universal for all subsystems. Then instead of six generating parameters only three are necessary:

$$1 + H_k + \tilde{H}_k \to H_k, \quad k = \overline{1,3}\,.$$

As a result of the condition (3.10), we can justify the matrix identity

$$\Big(\sum_l H_l \mathbf{P}_l\Big)^{-1} = \sum_l \frac{1}{H_l}\mathbf{P}_l\,.$$

Utilizing the identity together with summation formulas (6.3), we pass from (4.15) to expressions for the direct and exchange integrals in the Sp(6,R) model:

$$W_{\nu\tilde{\nu}}^{g}(\gamma) = \Big\{L_{N_\nu}^{3}\Big(-\frac{\partial}{\partial H_1}-\frac{\partial}{\partial H_2}-\frac{\partial}{\partial H_3}\Big) + \sum_{n_\nu}{}' \prod_k L_{n_{\nu k}}\Big(-\frac{\partial}{\partial H_k}\Big)\Big\} \times \Big\{L_{N_{\tilde{\nu}}}^{3}\Big(-\frac{\partial}{\partial H_1}-\frac{\partial}{\partial H_2}-\frac{\partial}{\partial H_3}\Big) + \sum_{n_{\tilde{\nu}}}{}' \prod_k L_{n_{\tilde{\nu} k}}\Big(-\frac{\partial}{\partial \tilde{H}_k}\Big)\Big\} \times (H_1H_2H_3)^{\frac{1}{2}(g-1)}|\mathbf{B}|^{1/2}\Big|\mathbf{B} + \frac{\gamma}{2}\sum_l H_l^g(\mathbf{P}_l + \mathbf{P}_l^T)\Big|^{-1/2}\Big|_{H_k=0}\,, \tag{6.7}$$

where the symbol Σ'_{n_ν} means a summation only with respect to occupied states of the open shell of the subsystem ν. Note that in the scope of the conventional Sp(6,R) model the integrals of direct and exchange interaction are defined by the similar expressions, contrary to the extended model where the calculation of the exchange integral is much more difficult [cf. (4.14)].

By annihilating all elements of the collective generator coordinate matrices $\mathbf{b}$ and $\tilde{\mathbf{b}}$, except b_{11} and $\tilde{b}_{11}$, the Sp(6,R) generating kernels are reduced to the Sp(2,R) limit of Filippov *et al.*[37]

The conventional Sp(2,R) model, or, that is the same, "the stretched Sp(6,R) approximation,"[10,12] is used to describe the longitudinal quadrupole vibrations.

(C) The light and medium weakly deformed nuclei admit the SU(3) model description[1] of the low-energy spectrum region. The Hamiltonian matrix elements in the extended SU(3) model are obtained by annihilating the matrices $\tilde{\mathbf{b}}$ and $\mathbf{b}$ in Eqs. (6.2) and subsequent projecting onto the states with fixed angular momentum value.

The expressions for the direct and exchange integrals in the unitary model are much simpler than in the symplectic model. This circumstance makes it possible to apply the extended SU(3) model as a first approximation in studying nuclei with the horizontal mixture of the open nucleon shell configurations.

VII. CONCLUSION

In the present paper the extended version of the symplectic and unitary nuclear models is proposed, in the framework of which one can develop consistent microscopic research of the nucleon systems with an arbitrary occupation of the valence shell. We have considered the structure of the extended Sp(6,R) model coherent states generating the basis of many-particle oscillator functions and constructed the microscopic Hamiltonian (including central and tensor nucleon interaction) matrix elements between these states.

The application of the symplectic model with horizontal mixture appreciably broadens the group of questions in the microscopic theory of collective motion of nucleon systems to be investigated (note that the results of the present paper may also be generalized to the other many-body quantum systems with characteristic shell structure, for example, atomic ones).

The possibility is provided to study the dynamics of passages between the different shell configurations of valence nucleons, the influence of these passages on atomic nuclei shape and spectra.

The extended unitary model is accomodated to describe the low-energy region of the spectrum. The comparative simplicity of the model enables us to expect a greater number of analytic results of its utilization, for example, the construction of effective nuclear Hamiltonians. It seems possible to apply the unitary model to microscopic substantiation and correction of the phenomenological nuclear shell theory predictions, for example, the shell state occupation scheme.[30,31] We note that the model indicated realizes a microscopic approach to describe the same processes as the interacting proton and neutron boson model with a dynamical symmetry group $SU(6) \otimes SU(6)$ proposed by Otsuka, Arima, and Iachello[38] for even-even nuclei. Therefore the extended unitary model may be transformed in order to predict microscopic values of the interacting boson model parameters and to estimate the calculational precision of the model.

The determinant form of the Hamiltonian matrix elements obtained in the paper allows a generalization to higher dimension models, $Sp(2d,R)$ and $SU(d)$, $d>3$. These models may be useful in studies of unified, both collective and intrinsic, nuclear dynamics.

ACKNOWLEDGMENTS

The authors are grateful to Yu. F. Smirnov, V. S. Vasilevsky, and Yu. V. Teryoshin for the fruitful discussion of the present work at its different stages.

[1]J. P. Elliott, Proc. R. Soc. A London **245**, 128 (1958); **245**, 562 (1958).
[2]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[3]R. M. Asherova, V. A. Knyr, Yu. F. Smirnov, and V. N. Tolstoy, Yad. Fiz. **21**, 1126 (1975).
[4]G. Rosensteel and D. J. Rowe, Ann. Phys. **96**, 1 (1976); **126**, 198 (1981).
[5]V. S. Vasilevsky, Yu. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980).
[6]G. F. Filippov, V. I. Ovcharenko, and Yu. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (Naukova Dumka, Kiev, 1981) (in Russian).
[7]G. F. Filippov, V. S. Vasilevsky, and L. L. Chopovsky, Fiz. Elem. Chastits At. Yadra **15**, 1338 (1984).
[8]D. R. Peterson and K. T. Hecht, Nucl. Phys. A **344**, 361 (1980).
[9]V. Vanagas, "The Microscopic Theory of Collective Motions in Nuclei," in *Group Theory and its Applications in Physics–1980*, edited by T. H. Seligman (AIP, New York, 1980).
[10]F. Arickx, J. Broeckhove, and E. Deumens, Nucl. Phys. A **377**, 121 (1982).
[11]R. M. Asherova, Yu. F. Smirnov, V. N. Tolstoy, and A. P. Shustov, Nucl. Phys. A **355**, 25 (1981).
[12]P. Park, J. Carvalho, M. Vassanji, D. J. Rowe, and G. Rosensteel, Nucl. Phys. A **414**, 93 (1984).
[13]Y. Suzuki and K. T. Hecht, Nucl. Phys. A **455**, 315 (1986).
[14]O. Castaños, P. Kramer, and M. Moshinsky, J. Math. Phys. **27**, 924 (1986).
[15]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. **128**, 314 (1980); **131**, 1 (1981).
[16]J. P. Draayer and K. J. Weeks, Phys. Rev. Lett. **51**, 1422 (1983).
[17]G. F. Filippov, V. I. Avramenko, and A. M. Sokolov, "Identity of the SU(3) model phenomenological Hamiltonian and the Hamiltonian of nonaxial rotator," preprint ITP-84-174E, Institute for Theoretical Physics, Kiev.
[18]V. S. Vasilevsky and G. F. Filippov, "Sp(6,*R*) model effective Hamiltonian for magic nuclei," in *Group Theoretical Methods in Physics. Proceedings of the third seminar* (Nauka, Moscow, 1986).
[19]K. T. Hecht and J. P. Elliott, Nucl. Phys. A **438**, 29 (1985).
[20]Y. Akiyama, A. Arima, and T. Sebe, Nucl. Phys. A **138**, 273 (1969).
[21]J. P. Draayer, K. J. Weeks, and G. Rosensteel, Nucl. Phys. A **413**, 215 (1984).
[22]F. Ajzenberg-Selove, Nucl. Phys. A **475**, 1 (1987).
[23]A. Arima and F. Iachello, Ann. Phys. **99**, 253 (1976).
[24]A. Leviatan, Ann. Phys. **179**, 201 (1987).
[25]K. T. Hecht, *The Vector Coherent State Method and Its Application to Problems of Higher Symmetries* (Springer, Berlin, 1987).
[26]B. F. Bayman, *Groups and Their Application to Spectroscopy* (NORDITA, Copenhagen, 1960).
[27]V. V. Vanagas, *Algebraic Foundations of Microscopic Nuclear Theory* (Nauka, Moscow, 1988) (in Russian).
[28]R. E. Peierls and J. Yoccoz, Proc. Phys. Soc. A London **70**, 381 (1957).
[29]P. O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[30]M. G. Mayer and J. H. D. Jensen, *Elementary Theory of Nuclear Shell Structure* (Wiley, New York, 1955).
[31]S. G. Nilsson, Mat. Fyz. Medd. Dan. Vid. Selsk. **29**, 1 (1955).
[32]A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, Vol. 1, Elementary Functions (Nauka, Moscow, 1981), p. 594.
[33]F. R. Gantmacher, *Matrix Theory* (Nauka, Moscow, 1967), p. 59.
[34]V. S. Vasilevsky, Yu. V. Teryoshin, and G. F. Filippov, "On realistic Hamiltonian with tensor and spin-orbit interaction for SU(3) model," preprint ITP-86-84E, Institute for Theoretical Physics, Kiev.
[35]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics. Theory and Application* (Addison–Wesley, Reading, MA, 1981).
[36]A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, Vol. 2, Special Functions (Nauka, Moscow, 1983), pp. 638, 640.
[37]G. F. Filippov, V. I. Ovcharenko, and Yu. V. Teryoshin, Yad. Fiz. **33**, 932 (1981).
[38]T. Otsuka, A. Arima, and F. Iachello, Nucl. Phys. A **309**, 1 (1978).

Radiative transfer theory for inhomogeneous media with random extinction and scattering coefficients

Robert M. Manning
NASA Lewis Research Center, 21000 Brookpark Rd., Cleveland, Ohio 44135

(Received 5 July 1988; accepted for publication 31 May 1989)

The small-angle scattering approximation of the scalar radiative transfer equation is examined in the case where the extinction and scattering coefficients have a component that is a deterministic function of position along the propagation path and a component that is a random function of position transverse to the propagation direction. It is found that the resulting stochastic radiative transfer equation can be reduced to a system of two stochastic integrodifferential equations for the average and fluctuating components of the radiant intensity. The system is solved to yield two transfer equations: one that describes the average radiant intensity and one that describes the spatial correlation function of the intensity fluctuations. The integrodifferential equation for the average intensity is then solved and applied to a simple propagation scenario; it is found that the fluctuations in the extinction and scattering coefficients reduce the effects due to the average values of these parameters, and also that the effect of these is greater near the point of observation than near the point of transmission of the radiation. An approximate solution is also derived for the equation giving the correlation function. The equations developed here should find application in problems involving short wavelength electromagnetic wave propagation through media possessing variable characteristics of turbulence and turbidity, such as in plasmas, the atmosphere, and the ocean.

I. INTRODUCTION

Radiative transfer theory, which deals with the propagation of "intensities" (i.e., photometric intensity, neutron flux intensity, etc.) began as a phenomenological theory based on observations of light propagation in foggy atmospheres published by Schuster in 1905.[1] Since that time, radiative transfer theory and its attendant equation, the radiative transfer equation (RTE), have gradually been put on a more rigorous basis[2] and have found application in several diverse areas such as atmospheric and underwater visibility,[3] optics of papers and photographic emulsions,[4] and the propagation of radiant energy in turbulent plasmas,[5] planetary atmospheres, stars, and galaxies.[6] Also, because the RTE is equivalent to the Maxwell–Boltzmann collision equation used in the kinetic theory of gases, it has also been applied to many problems in kinetic theory[7] and in neutron transport theory.[8]

Until very recently, in all of the applications of the various forms of the RTE, an important aspect seems to have been overlooked that can be characteristic of many propagation problems: in particular, many random propagation media that are characterized by quantities (in general, functions of some random process, e.g., scattering and absorption) that can themselves be random quantities, the statistics of which are spatially inhomogeneous throughout the medium. The significance of this circumstance seems to have first been noted by Levermore and co-workers[9,10] and Vanderhaegen[11,12] in the analysis of transport processes in random binary mixtures.

The random nature of the scattering media, usually characterized in transfer theory by spatial averaged extinction and scattering coefficients, may be such that these coefficients themselves are random functions of position and thus also have spatially fluctuating components that can be characterized by prevailing statistical parameters. For example, in the case of imaging through atmospheric aerosols, the concentration of the aerosols can have significant variation across the propagation path, thus adding a fluctuating component to the average extinction and scattering coefficients that would normally be used in the RTE. Or, in the case of neutron propagation through bulk media, the media may have some random distribution of scattering characteristics that can give rise to a scattering coefficient characterized by a spatial average over the media with a spatially varying component. The same scenario can be envisioned in the case of turbulent plasmas. Such circumstances suggest that one should consider the various forms of the RTE with extinction and scattering coefficients that are random functions of position in the propagating medium and modify these forms accordingly.

It is the purpose of this paper to introduce stochastic extinction and scattering coefficients into one form of the RTE, viz., the well known small scattering angle approximation, which finds use in atmospheric and underwater image propagation as well as propagation in turbulent plasmas. The treatment given here is a more general analysis of a more restricted form of the RTE equation than given in the work cited above.[9–12] In Sec. II, Gaussian random functions are introduced into the RTE for the extinction and scattering coefficients and the resulting stochastic RTE is reduced to a system of two stochastic integrodifferential equations that describe the average and fluctuating parts of the radiant intensity. Transfer equations are then obtained in Sec. III that describe the propagation of the average intensity and a statistic of the random intensity, i.e., the spatial correlation func-

 0022-2488/89/102432-09$02.50

tion of the intensity, where the statistics of the propagation parameters are given by spatial correlation functions of extinction and scattering that are taken to have a δ-function component in the direction of propagation. These transfer equations are then solved in Sec. IV; an exact solution is obtained for the average intensity, the properties of which are expounded upon, and an approximate solution is derived for the intensity correlation function. Finally in Sec. V, it is discussed how this stochastic approach to radiative transfer theory can be applied to situations more general than those described by the assumptions made here (i.e., Gaussian statistics of the random extinction field and δ correlation in the propagation direction) and to less restricted forms of the RTE.

II. DEVELOPMENT OF A STOCHASTIC RADIATIVE TRANSFER EQUATION FOR SMALL SCATTERING ANGLES

In an inhomogeneous medium void of volumetric sources, a radiance distribution function (specific intensity) $I(\mathbf{R},\mathbf{n})$ at a point $\mathbf{R}$ in the medium describing transfer in a direction specified by the unit vector $\mathbf{n}$ is described by the general form of the RTE, viz.,

$$\mathbf{n}\cdot\nabla I(\mathbf{R};\mathbf{n}) + \varepsilon(\mathbf{R})I(\mathbf{R};\mathbf{n}) = \int_{4\pi} \sigma_S(\mathbf{R};\mathbf{n};\mathbf{n}')I(\mathbf{R};\mathbf{n}')d\omega, \tag{2.1}$$

where $\varepsilon(\mathbf{R})$ is the position dependent extinction coefficient and $\sigma_S(\mathbf{R};\mathbf{n},\mathbf{n}')$ is the position-dependent generalized volume scattering cross section normalized such that

$$\int_{4\pi} \sigma_S(\mathbf{R};\mathbf{n},\mathbf{n}')d\omega = 1, \tag{2.2}$$

where $d\omega$ is an element of solid angle subtending the scattering angle defined by $\mathbf{n}$ and $\mathbf{n}'$. The scattering cross section is related to the more fundamental scattering coefficient $\sigma(\mathbf{R})$ and the scattering phase function $f(\mathbf{n},\mathbf{n}';\mathbf{R})$, also taken to be position dependent, through the relationship

$$\sigma_S(\mathbf{R};\mathbf{n},\mathbf{n}') = \sigma(\mathbf{R})f(\mathbf{n},\mathbf{n}';\mathbf{R})/4\pi. \tag{2.3}$$

Considering highly anisotropic scattering cases where the scattering takes place predominantly in the small solid angle about the direction $\mathbf{n}$, in particular, when the condition

$$\gamma_0^2 = \int_{4\pi} \gamma^2\sigma_S(\mathbf{R};\mathbf{n},\mathbf{n}')d\omega \ll 1, \quad \gamma = \cos^{-1}(\mathbf{n}\cdot\mathbf{n}'), \tag{2.4}$$

is satisfied, one can expand the unit vector $\mathbf{n}$ into its perpendicular component $\mathbf{n}_\perp$ and its longitudinal component n_z, where $n_z \approx 1$ since

$$n_z = \surd(1 - |\mathbf{n}_\perp|^2) \approx 1 - |\mathbf{n}_\perp|^2/2 \approx 1$$

because $|\mathbf{n}_\perp|$ is small. Upon further assuming that one can write

$$f(\mathbf{n},\mathbf{n}';\mathbf{R}) = f(\mathbf{n}_\perp - \mathbf{n}_\perp';\mathbf{R}),$$

and noting that this function gives an appreciable contribution to the integrand on the right side of Eq. (2.1) only for $|\mathbf{n}_\perp - \mathbf{n}_\perp'| \ll 1$, Eq. (2.1) can be transformed to its small-angle scattering form, i.e.,

$$\left[\frac{\partial}{\partial z} + \mathbf{n}_\perp\cdot\nabla_\mathbf{r} + \varepsilon(\mathbf{r},z)\right]I(\mathbf{r},z;\mathbf{n}_\perp) = \iint_{-\infty}^{\infty} \sigma_S(\mathbf{r},z;|\mathbf{n}_\perp - \mathbf{n}_\perp'|)I(\mathbf{r},z;\mathbf{n}_\perp')d\mathbf{n}_\perp', \tag{2.5}$$

where $\mathbf{R}$ is decomposed into its transverse and longitudinal components, $\mathbf{r}$ and z, respectively. This form of the RTE, along with its prevailing assumptions, finds considerable applications in problems involving electromagnetic wave propagation in random media.

Let $\varepsilon(\mathbf{r},z)$ and $\sigma(\mathbf{r},z)$ be random functions that are taken to be written in the form

$$\begin{aligned}
&\varepsilon(\mathbf{r},z) = \langle\varepsilon(z)\rangle + \tilde{\varepsilon}(\mathbf{r},z), \quad \langle\varepsilon(\mathbf{r},z)\rangle = 0,\\
&\sigma(\mathbf{r},z) = \langle\sigma(z)\rangle + \tilde{\sigma}(\mathbf{r},z), \quad \langle\sigma(\mathbf{r},z)\rangle = 0,\\
&\langle\varepsilon(z)\rangle,\langle\sigma(z)\rangle \geqslant |\tilde{\varepsilon}(\mathbf{r},z)|,|\tilde{\sigma}(\mathbf{r},z)|,
\end{aligned} \tag{2.6}$$

where the ensemble averages $\langle\cdots\rangle$ are deterministic functions only of the longitudinal coordinate z, and the fluctuating parts, $\tilde{\varepsilon}(\mathbf{r},z)$ and $\tilde{\sigma}(\mathbf{r},z)$, are zero-mean Gaussian random functions only in the variable $\mathbf{r}$ but can also have deterministic factors in the coordinate z. Substituting Eq. (2.6) into Eqs. (2.5) and (2.3) and rearranging terms yields the relationship

$$\Re I = \Im I, \tag{2.7}$$

where $\Re$ and $\Im$ are, respectively, the deterministic and stochastic radiative transfer operators defined by

$$\Re I \equiv \left[\frac{\partial}{\partial z} + \mathbf{n}_\perp\cdot\nabla_\mathbf{r} + \langle\varepsilon(z)\rangle\right]I - \iint_{-\infty}^{\infty} \langle\sigma_S(z;|\mathbf{n}_\perp - \mathbf{n}_\perp'|)\rangle I'\,d\mathbf{n}_\perp' \tag{2.8}$$

and

$$\Im I = -\tilde{\varepsilon}(\mathbf{r},z)I + \iint_{-\infty}^{\infty} \tilde{\sigma}_S(\mathbf{r},z;|\mathbf{n}_\perp - \mathbf{n}_\perp'|)I'\,d\mathbf{n}_\perp', \tag{2.9}$$

where $I \equiv I(\mathbf{r},z;\mathbf{n}_\perp)$ and $I' \equiv I(\mathbf{r},z;\mathbf{n}_\perp')$. Averaging Eq. (2.7) and noting the deterministic nature of the operator $\Re$ simply gives

$$\Re\langle I\rangle = \langle\Im I\rangle, \tag{2.10}$$

the formal solution of which yields the average intensity $\langle I\rangle$ and can be taken to define that quantity. Decomposing the total intensity into its average and fluctuating components, viz.,

$$I(\mathbf{r},z;\mathbf{n}_\perp) = \langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle + \tilde{I}(\mathbf{r},z;\mathbf{n}_\perp), \tag{2.11}$$

and substituting this expression into Eq. (2.7), averaging, and using Eq. (2.10) yield the fact that $\Re\langle\tilde{I}\rangle = 0$ and thus $\langle\tilde{I}\rangle = 0$.

It now remains to develop an expression for the random part of the total intensity, viz., $\tilde{I}(\mathbf{r},z;\mathbf{n}_\perp)$, that together with Eqs. (2.10) and (2.11) will give a closed system of equations for the problem. Solving Eq. (2.11) for $\tilde{I}$, applying the operator $\Re$ to the result, averaging, and employing Eqs. (2.7) and (2.10) yield

$$\Re\tilde{I} = \Im I - \langle\Im I\rangle. \tag{2.12}$$

Consider now the term $\langle \Im I \rangle$. Substituting Eq. (2.11) into Eq. (2.9), averaging, and using the fact that $\langle \tilde{\varepsilon}(\mathbf{r},z)\rangle = \langle \tilde{\sigma}(\mathbf{r},z)\rangle = 0$, one finds that $\langle \Im I \rangle = \langle \Im \tilde{I} \rangle$. Thus

$$\Re \tilde{I} = \Im I - \langle \Im \tilde{I} \rangle. \tag{2.13}$$

Equations (2.10) and (2.13) form a system of stochastic integrodifferential equations that collectively form the stochastic radiative theory to describe propagation problems in situations where the extinction and scattering coefficients are Gaussian random functions. The system of equations (2.10) and (2.13) can be solved for $\langle I \rangle$ and $\tilde{I}$; in the latter case, since $\tilde{I}$ is a random function, one can only obtain expressions of the various statistical quantities that govern the intensity fluctuations. This forms the subject of the following section.

III. SOLUTION OF THE SYSTEM OF STOCHASTIC INTEGRODIFFERENTIAL TRANSFER EQUATIONS

A. Transfer equation for the average intensity

The solution of the system of equations commences with the first equation, i.e., Eq. (2.10), which, written out in full, is

$$\left[\frac{\partial}{\partial z} + \mathbf{n}_\perp \cdot \nabla_r + \langle \varepsilon(z)\rangle\right]\langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle - \iint_{-\infty}^{\infty} \langle \sigma_S(z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\rangle \langle I(\mathbf{r},z;\mathbf{n}'_\perp)\rangle d\mathbf{n}'_\perp$$
$$= -\left\langle \tilde{\varepsilon}(\mathbf{r},z)\tilde{I}(\mathbf{r},z;\mathbf{n}_\perp) + \iint_{-\infty}^{\infty} \tilde{\sigma}_S(\mathbf{r},z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\tilde{I}(\mathbf{r},z;\mathbf{n}'_\perp) d\mathbf{n}'_\perp \right\rangle. \tag{3.1}$$

Since no random function in the variable $\mathbf{n}_\perp$ has been admitted into the problem, it is permissible to apply the Fourier transform in this variable to the equation. Thus, defining

$$\langle J(\mathbf{r},z;\mathbf{q})\rangle = \iint_{-\infty}^{\infty} \langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle \exp(-i\mathbf{q}\cdot\mathbf{n}_\perp) d\mathbf{n}_\perp,$$
$$\tilde{J}(\mathbf{r},z;\mathbf{q}) = \iint_{-\infty}^{\infty} \tilde{I}(\mathbf{r},z;\mathbf{n}_\perp)\exp(-i\mathbf{q}\cdot\mathbf{n}_\perp) d\mathbf{n}_\perp, \tag{3.2}$$

Eq. (3.1) becomes

$$\left[\frac{\partial}{\partial z} + i\nabla_r \cdot \frac{\partial}{\partial \mathbf{q}} + \langle \varepsilon(z)\rangle\right]\langle J(\mathbf{r},z;\mathbf{q})\rangle - \langle \Sigma_S(z,\mathbf{q})\rangle \langle J(\mathbf{r},z;\mathbf{q})\rangle = \langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle, \tag{3.3}$$

where

$$\tilde{\phi}(\mathbf{r},z;\mathbf{q}) = -\tilde{\varepsilon}(\mathbf{r},z) + \tilde{\Sigma}_S(\mathbf{r},z;\mathbf{q}),$$
$$\langle \Sigma_S(z;\mathbf{q})\rangle = \langle \sigma(z)\rangle P(\mathbf{q},z),$$
$$\tilde{\Sigma}_S(\mathbf{r},z;\mathbf{q}) = \tilde{\sigma}(\mathbf{r},z)P(\mathbf{q},z), \tag{3.4}$$
$$P(\mathbf{q},z) = \frac{1}{4\pi}\iint_{-\infty}^{\infty} f(\mathbf{n}_\perp,z)\exp(-i\mathbf{q}\cdot\mathbf{n}_\perp) d\mathbf{n}_\perp.$$

Since $\tilde{\varepsilon}$ and $\tilde{\Sigma}_S$ are Gaussian random functions, so, too, is $\tilde{\phi}$. One can now employ the Novikov theorem[13] to evaluate the product $\langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle$. This theorem states that for a zero-mean Gaussian random function $f(\mathbf{R})$ and a corresponding functional $G[f]$,

$$\langle f(\mathbf{R})G[f]\rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \langle f(\mathbf{R})f(\mathbf{R}')\rangle \left\langle \frac{\delta G}{\delta f(\mathbf{R}')}\right\rangle d\mathbf{R}',$$

where $\delta/\delta f(\mathbf{R}')$ is a variational derivative and the integration is taken over the entire space that defines $\mathbf{R}$. It is at this point where the statistics that govern $\tilde{\phi}$ enter into the problem. In the case of the right-hand member of Eq. (3.3), application of the theorem gives

$$\langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle = \int\int\int B_\phi(\mathbf{r},\mathbf{r}';z,z';\mathbf{q},\mathbf{q}') \times \left\langle \frac{\delta \tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta \tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}\right\rangle d\mathbf{r}'\, dz'\, d\mathbf{q}', \tag{3.5}$$

where

$$B_\phi(\mathbf{r},\mathbf{r}';z,z';\mathbf{q},\mathbf{q}') = \langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')\rangle$$

is the spatial correlation function of the linear combination of the random propagation parameters. As shown in Appendix A, for a statistically homogeneous (in the coordinate $\mathbf{r}$) correlation function of $\tilde{\phi}$ that has a δ-function component in the longitudinal z direction, and with the help of Eq. (2.13), one finds that

$$\langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle = A_\phi(0,z;\mathbf{q})\langle J(\mathbf{r},z;\mathbf{q})\rangle, \tag{3.6}$$

where $A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q})$ is the two-dimensional (transverse) correlation function related to the full three-dimensional one, i.e., $B_\phi(\mathbf{r}-\mathbf{r}',z-z';\mathbf{q},\mathbf{q}')$, via

$$A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q},\mathbf{q}') = \int_0^L B_\phi(\mathbf{r}-\mathbf{r}',z-z';\mathbf{q},\mathbf{q}')dz', \tag{3.7a}$$

where

$$A_\phi(\mathbf{r}-\mathbf{r}',z,z';\mathbf{q},\mathbf{q}') = A_{\varepsilon\varepsilon}(\mathbf{r}-\mathbf{r}') - A_{\varepsilon\sigma}(\mathbf{r}-\mathbf{r}')P(\mathbf{q}',z') - A_{\sigma\varepsilon}(\mathbf{r}-\mathbf{r}')P(\mathbf{q},z) + A_{\sigma\sigma}(\mathbf{r}-\mathbf{r}')P(\mathbf{q},z)P(\mathbf{q}',z'), \tag{3.7b}$$

with

$$A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q},\mathbf{q}') \equiv A_\phi(\mathbf{r}-\mathbf{r}',z,z;\mathbf{q},\mathbf{q}')$$

and

$$A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q}) \equiv A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q},\mathbf{q})$$

which relates the two-dimensional correlation of the composite random function $\tilde{\phi}$ to those of the fundamental extinction and scattering parameters.

It is very important to note that use of the δ-correlation assumption, as pointed out in Appendix A, places specific requirements on the spatial extent and level of the fluctuations $\tilde{\varepsilon}(\mathbf{r},z)$ and $\tilde{\sigma}(\mathbf{r},z)$; in particular, letting l_0 and L_0 denote, respectively, the smallest and largest spatial extent of the fluctuations, and letting $k = 2\pi/\lambda$ be the wave number of the wave field of wavelength λ, sufficient conditions to be satisfied are $\lambda \ll l_0$, $L \gg L_0$, $\lambda\alpha \ll 1$, and

$$\langle(\tilde{\varepsilon}(\mathbf{r},z))^2\rangle, \langle(\tilde{\sigma}(\mathbf{r},z))^2\rangle \ll 1/(kL_0),$$

where L is the total length of propagation and α is the coefficient of absorption, $\alpha = \langle \varepsilon \rangle - \langle \sigma \rangle$.

Substituting Eq. (3.6) into Eq. (3.3) gives an equation closed in the quantity $\langle J(\mathbf{r},z;\mathbf{q})\rangle$. Taking the inverse Fourier transform of this equation yields a radiative transfer equation for the average radiant intensity, viz.,

$$\left[\frac{\partial}{\partial z} + \mathbf{n}_\perp\cdot\nabla_\mathbf{r} + \langle \varepsilon(z)\rangle\right]\langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle - \iint_{-\infty}^{\infty} \langle \sigma_S(z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\rangle \langle I(\mathbf{r},z;\mathbf{n}'_\perp)\rangle d\mathbf{n}'_\perp = \iint_{-\infty}^{\infty} H_\phi(0,z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\langle I(\mathbf{r},z;\mathbf{n}'_\perp)\rangle d\mathbf{n}'_\perp, \tag{3.8}$$

where the source term is the scattering of the average intensity field that results from the extinction and scattering parameter fluctuations incorporated in the volume scattering factor $H_\phi(\mathbf{r}-\mathbf{r}',z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)$, which is given by

$$H_\phi(\mathbf{r},z;\mathbf{n}_\perp) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} A_\phi(\mathbf{r},z;\mathbf{q})\exp(i\mathbf{n}_\perp\cdot\mathbf{q})d\mathbf{q}. \tag{3.9}$$

Before discussing the solution of Eq. (3.8), Eqs. (2.10) and (2.13) will now be solved to give a relation, companion to that of Eq. (3.8), governing the correlation of the intensity fluctuations $\tilde{I}$.

B. Transfer equation for the correlation of the intensity fluctuations

The second equation of the system obtained in Sec. II is, upon using the definitions of Eqs. (2.8) and (2.9) in Eq. (2.13),

$$\left[\frac{\partial}{\partial z} + \mathbf{n}_\perp\cdot\nabla_\mathbf{r} + \langle \varepsilon(z)\rangle\right]\tilde{I}(\mathbf{r},z;\mathbf{n}_\perp) - \iint_{-\infty}^{\infty} \langle \sigma_S(z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\rangle \tilde{I}(\mathbf{r},z;\mathbf{n}'_\perp)d\mathbf{n}'_\perp$$
$$= -\tilde{\varepsilon}(\mathbf{r},z)I(\mathbf{r},z;\mathbf{n}_\perp) + \iint_{-\infty}^{\infty} \tilde{\sigma}_S(\mathbf{r},z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)I(\mathbf{r},z;\mathbf{n}'_\perp)d\mathbf{n}'_\perp + \langle \tilde{\varepsilon}(\mathbf{r},z)\tilde{I}(\mathbf{r},z;\mathbf{n}_\perp) - \iint_{-\infty}^{\infty} \tilde{\sigma}_S(\mathbf{r},z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)\tilde{I}(\mathbf{r},z;\mathbf{n}'_\perp)d\mathbf{n}'_\perp\rangle. \tag{3.10}$$

As was done with Eq. (3.1), this equation can be Fourier transformed with respect to the variable $\mathbf{n}_\perp$. Thus, remembering the definitions of Eq. (3.2), one can transform Eq. (3.10) and rearrange terms to obtain

$$\frac{\partial \tilde{J}(\mathbf{r},z;\mathbf{q})}{\partial z} = -i\nabla_\mathbf{r}\cdot\frac{\partial \tilde{J}(\mathbf{r},z;\mathbf{q})}{\partial \mathbf{q}} + \langle \phi(z;\mathbf{q})\rangle \tilde{J}(\mathbf{r},z;\mathbf{q}) + \tilde{\phi}(\mathbf{r},z;\mathbf{q})J(\mathbf{r},z;\mathbf{q}) - \langle \tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle, \tag{3.11}$$

where, in addition to the quantities already defined,

$$\langle \phi(z;\mathbf{q})\rangle \equiv -\langle \varepsilon(z)\rangle + \langle \Sigma_S(z;\mathbf{q})\rangle.$$

The formal solution of Eq. (3.11) is a random function and it is therefore desired to obtain some statistical characterization of it, e.g., the correlation function

$$B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2) = \langle \tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle.$$

To get an expression for this function from Eq. (3.11), one writes Eq. (3.11) in the variables $\mathbf{r}_1$ and $\mathbf{q}_1$, multiplies the resulting equation by $\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)$, and adds to this result an identical equation with the subscripts of $\mathbf{r}_1$ and $\mathbf{r}_2$ and those of $\mathbf{q}_1$ and $\mathbf{q}_2$ interchanged. Writing the total Fourier transformed intensity in terms of its average and fluctuating parts, i.e., $J(\mathbf{r},z;\mathbf{q}) = \langle J(\mathbf{r},z;\mathbf{q})\rangle + \tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)$, this entire result can be simplified to give

$$\frac{\partial B(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2)}{\partial z} = -i\left[\nabla_{\mathbf{r}_1}\cdot\frac{\partial}{\partial \mathbf{q}_1} + \nabla_{\mathbf{r}_2}\cdot\frac{\partial}{\partial \mathbf{q}_2}\right]B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2) + (\langle \phi(z;\mathbf{q}_1)\rangle + \langle \phi(z;\mathbf{q}_2)\rangle)B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2)$$
$$+ \langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle\langle J(\mathbf{r}_1,z;\mathbf{q}_1)\rangle + \langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\rangle$$
$$+ \langle \tilde{\phi}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\rangle\langle J(\mathbf{r}_2,z;\mathbf{q}_2)\rangle + \langle \tilde{\phi}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle. \tag{3.12}$$

One must now find expressions for the ensemble averages of the products that appear in the last four terms of Eq. (3.12). As shown in Appendix B, one can derive the following relationships making use of the Novikov theorem and the results of Appendix A:

$$\langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle = A_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{q}_1,\mathbf{q}_2)\langle J(\mathbf{r}_2,z;\mathbf{q}_2)\rangle, \tag{3.13}$$

$$\langle \tilde{\phi}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_2)\rangle = A_\phi(\mathbf{r}_2-\mathbf{r}_1,z;\mathbf{q}_2,\mathbf{q}_1)\langle J(\mathbf{r}_1,z;\mathbf{q}_1)\rangle, \tag{3.14}$$

$$\langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\rangle = [A_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{q}_1,\mathbf{q}_2) + A_\phi(0,z;\mathbf{q}_1)]B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2), \tag{3.15}$$

$$\langle \tilde{\phi}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle = [A_\phi(\mathbf{r}_2-\mathbf{r}_1,z;\mathbf{q}_2,\mathbf{q}_1) + A_\phi(0,z;\mathbf{q}_2)]B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2). \tag{3.16}$$

If, in addition to the assumption introduced earlier of homogeneity of the statistics of $\tilde{\phi}$, one admits the additional assumption that isotropy (in the coordinate $\mathbf{r}$) prevails, i.e., $A_\phi(\mathbf{r}_2-\mathbf{r}_1;z;\mathbf{q}_2,\mathbf{q}_1) = A_\phi(\mathbf{r}_1-\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2)$, one obtains, after substitution of Eqs. (3.13)–(3.16) into Eq. (3.12) and a rearrangement of terms,

$$\left[\frac{\partial}{\partial z} + i\left(\nabla_{r_1}\cdot\frac{\partial}{\partial \mathbf{q}_1} + \nabla_{r_2}\cdot\frac{\partial}{\partial \mathbf{q}_2}\right) + 2\langle\varepsilon(z)\rangle\right]B_J$$
$$= (\langle\Sigma_S(z;\mathbf{q})\rangle + \langle\Sigma_S(z;\mathbf{q}_2)\rangle)B_J + [A_\phi(0,z;\mathbf{q}_1) + 2A_\phi(\mathbf{r}_1-\mathbf{r}_2;\mathbf{q}_1,\mathbf{q}_2) + A_\phi(0,z;\mathbf{q}_2)]B_J$$
$$+ 2A_\phi(\mathbf{r}_1-\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2)\langle J(\mathbf{r}_1,z;\mathbf{q}_1)\rangle\langle J(\mathbf{r}_2,z;\mathbf{q}_2)\rangle. \quad (3.17)$$

It is now desired to relate the correlation function B_J to that of the intensity fluctuations $B_I = B_I(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{n}_{\perp 1},\mathbf{n}_{\perp 2}) \equiv \langle\tilde{I}(\mathbf{r}_1,z;\mathbf{n}_{\perp 1})\tilde{I}(\mathbf{r}_2,z;\mathbf{n}_{\perp 2})\rangle$.

From the second relationship of Eq. (3.2) and the definition of the correlation function B_J, one has the Fourier transform pair

$$B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2) = \int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} B_I(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{n}_{\perp 1},\mathbf{n}_{\perp 2})\exp(-i\mathbf{q}_1\cdot\mathbf{n}_{\perp 1} - i\mathbf{q}_2\cdot\mathbf{n}_{\perp 2})d\mathbf{n}_{\perp 1}\, d\mathbf{n}_{\perp 2},$$
$$B_I(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{n}_{\perp 1},\mathbf{n}_{\perp 2}) = \left(\frac{1}{2\pi}\right)^4\int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} B_J(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2)\exp(i\mathbf{q}_1\cdot\mathbf{n}_{\perp 1} + i\mathbf{q}_2\cdot\mathbf{n}_{\perp 2})d\mathbf{q}_1\, d\mathbf{q}_2 \quad (3.18)$$

that exists between the correlations of $\tilde{J}(\mathbf{r},z;\mathbf{q})$ and the random intensity component $\tilde{I}(\mathbf{r},z;\mathbf{q})$. Thus applying the inverse transform of Eq. (3.18) to Eq. (3.17) yields a transfer equation that governs the correlation of intensity fluctuations, viz.,

$$\left[\frac{\partial}{\partial z} + \mathbf{n}_{\perp 1}\cdot\nabla_{r_1} + \mathbf{n}_{\perp 2}\cdot\nabla_{r_2} + 2\langle\varepsilon(z)\rangle\right]B_I - \int\int_{-\infty}^{\infty}\langle\sigma_S(z;|\mathbf{n}_{\perp 1}-\mathbf{n}'_{\perp 1}|)\rangle B_I(\mathbf{n}'_{\perp 1})d\mathbf{n}'_{\perp 1}$$
$$-\int\int_{-\infty}^{\infty}\langle\sigma_S(z;|\mathbf{n}_{\perp 2}-\mathbf{n}'_{\perp 2}|)\rangle B(\mathbf{n}'_{\perp 2})d\mathbf{n}'_{\perp 2} - \int\int_{-\infty}^{\infty} H_\phi(0,z;|\mathbf{n}_{\perp 1}-\mathbf{n}'_{\perp 1}|)B_I(\mathbf{n}'_{\perp 1})d\mathbf{n}'_{\perp}$$
$$-2\int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} H_\phi(\mathbf{r}_1-\mathbf{r}_2,z;|\mathbf{n}_{\perp 1}-\mathbf{n}'_{\perp 1}|,\ |\mathbf{n}_{\perp 2}-\mathbf{n}'_{\perp 2}|)B_I(\mathbf{n}'_{\perp 1},\mathbf{n}'_{\perp 2})d\mathbf{n}'_{\perp 1}\, d\mathbf{n}'_{\perp 2} - \int\int_{-\infty}^{\infty} H_\phi(0,z;|\mathbf{n}_{\perp 2}-\mathbf{n}'_{\perp 2}|)B_I(\mathbf{n}'_{\perp 2})d\mathbf{n}'_{\perp 2}$$
$$= 2\int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} H_\phi(\mathbf{r}_1-\mathbf{r}_2,z;|\mathbf{n}_{\perp 1}-\mathbf{n}'_{\perp 1}|,\ |\mathbf{n}_{\perp 2}-\mathbf{n}'_{\perp 2}|)\langle I(\mathbf{n}'_{\perp 1})\rangle\langle I(\mathbf{n}'_{\perp 2})\rangle d\mathbf{n}'_{\perp 1}\, \mathbf{n}'_{\perp 2}, \quad (3.19)$$

where, for the convenience of notation, $B_I(\mathbf{n}'_{\perp 1}) \equiv B_I(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{n}'_{\perp 1},\mathbf{n}_{\perp 2})$, etc., and $\langle I(\mathbf{n}'_{\perp 1})\rangle \equiv \langle I(\mathbf{r}_1,z;\mathbf{n}'_{\perp 1})\rangle$, etc., $H_\phi(\mathbf{r}-\mathbf{r}',z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)$ is as given in Eq. (3.9) and

$$H_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{n}_{\perp 1},\mathbf{n}_{\perp 2}) = \left(\frac{1}{2\pi}\right)^4\int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} A_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{q}_1,\mathbf{q}_2) \times\exp(i\mathbf{q}_1\cdot\mathbf{n}_{\perp 1} + i\mathbf{q}_2\cdot\mathbf{n}_{\perp 2})d\mathbf{q}_1\, d\mathbf{q}_2. \quad (3.20)$$

The interpretation of Eq. (3.19) is straightforward; the second and third members on the left side of the equation describe the average contributions of the scattering into the direction $\mathbf{n}_{\perp 1}$ and $\mathbf{n}_{\perp 2}$, respectively; the fourth and sixth members describe the contribution that the fluctuations of extinction and scattering have on radiation propagating into the directions $\mathbf{n}_{\perp 1}$ and $\mathbf{n}_{\perp 2}$, respectively; and the fifth member gives the effect of the cross correlated fluctuations at two points $\mathbf{r}_1$ and $\mathbf{r}_2$ of the extinction and scattering on propagation into the directions $\mathbf{n}_{\perp 1}$ and $\mathbf{n}_{\perp 2}$. The source term of Eq. (3.19) is the cross-correlated fluctuations of scattering and extinction at $\mathbf{r}_1$ and $\mathbf{r}_2$, due to the average fields $\langle I(\mathbf{r}_1,z;\mathbf{n}'_\perp)\rangle$ and $\langle I(\mathbf{r}_2,z;\mathbf{n}'_{\perp 2})\rangle$, into the directions $\mathbf{n}_{\perp 1}$ and $\mathbf{n}_{\perp 2}$.

IV. SOLUTION OF THE TRANSFER EQUATIONS FOR THE AVERAGE INTENSITY AND THE CORRELATION FUNCTION OF THE ASSOCIATED FLUCTUATIONS

The solution to Eq. (3.8), the stochastic transfer equation for the average radiant intensity, is straightforward.[14] Fourier transforming the equation with respect to the variables $\mathbf{r}$ and $\mathbf{n}_\perp$, solving the resulting first-order differential equation via the method of characteristics,[15] employing Eq. (3.7b), and rearranging terms yield

$$F(\boldsymbol{\kappa},L;\mathbf{q}) = F(\boldsymbol{\kappa},0;\mathbf{q}+\boldsymbol{\kappa}L)\exp\left[-\int_0^L\{\langle\varepsilon(z)\rangle - A_{\varepsilon\varepsilon}(0,z)\}dz + \int_0^L\{\langle\sigma(z)\rangle - 2A_{\varepsilon\sigma}(0,z)\}P(\mathbf{q}+\boldsymbol{\kappa}(L-z),z)dz + \int_0^L A_{\sigma\sigma}(0,z)P^2(\mathbf{q}+\boldsymbol{\kappa}(L-z),z)dz\right], \quad (4.1)$$

where

$$\langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle = \left(\frac{1}{2\pi}\right)^4\int\int_{-\infty}^{\infty}\int\int_{-\infty}^{\infty} F(\boldsymbol{\kappa},z;\mathbf{q}) \times\exp(i\boldsymbol{\kappa}\cdot\mathbf{r} + i\mathbf{q}\cdot\mathbf{n}_\perp)d\boldsymbol{\kappa}\, d\mathbf{q} \quad (4.2)$$

and L is the total propagation distance in the turbulent medium. As can be easily seen from this solution, the form is analogous to that of the well nonstochastic small scattering angle case, and thus possesses the same overall properties, but with some obvious differences in its composition. The first two terms in the exponential of Eq. (4.1) indicate that the random fluctuations of the extinction and scattering within the medium, characterized by the two-dimensional correlation functions $A_{\varepsilon\varepsilon}$ and $A_{\varepsilon\sigma}$, tend to negate the effects presented by the attendant average quantities, i.e., $\langle\varepsilon(z)\rangle$ and $\langle\sigma(z)\rangle$. The implications that this has on particular propagation problems are specific to those problems and will be left to future investigations dealing with these specific problems. However, a simple generic example will be given here that demonstrates the implications that result due to the two major aspects that have been incorporated into the present theory, i.e., the random fluctuations of the extinction and scattering parameters and the variations of these parameters along the longitudinal propagation path.

At the outset, consider, instead of the radiant intensity, the illumination distribution $S(\mathbf{r},z)$ at a point $\mathbf{r}$ a distance z from the source. In the case, for example, of a narrow directional beam of electromagnetic radiation undergoing strong anisotropic scattering, which is that assumed to prevail in the small scattering angle case here, one has

$$S(\mathbf{r},z) = \int_{4\pi} \langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle \cos(\mathbf{n}_\perp\cdot\Omega)\,d\Omega \approx \iint_{-\infty}^{\infty} \langle I(\mathbf{r},z;\mathbf{n}_\perp)\rangle\,d\mathbf{n}_\perp, \tag{4.3}$$

where Ω is an element of solid angle centered at $\mathbf{r}$ extending toward the source. Applying this to Eq. (4.2) gives

$$S(\mathbf{r},z) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} F(\boldsymbol{\kappa},z;0)\exp(i\boldsymbol{\kappa}\cdot\mathbf{r})\,d\boldsymbol{\kappa}. \tag{4.4}$$

Thus one now need only consider the simpler case of Eq. (4.1) evaluated at $\mathbf{q}=0$ and still have a physically meaningful quantity.

Let the scattering and extinction take place along a fractional length Δ of the propagation path that begins at some intermediate point z_0 and let $\langle\varepsilon(z)\rangle$ and $\langle\sigma(z)\rangle$ as well as the fluctuations in these quantities, represented by the two-dimensional correlation functions $A_{ij}(0,z)$, $i,j=\varepsilon,\sigma$, be constant over the length Δ. Such a distance can be modeled by

$$\begin{aligned}&\langle\varepsilon(z)\rangle = \langle\varepsilon\rangle\zeta(z), \quad \langle\sigma(z)\rangle = \langle\sigma\rangle\zeta(z),\\ &A_{ij}(0,z) = A_{ij}(0)\zeta(z),\\ &\zeta(z) = \Theta(z-z_0) - \Theta(z-(z_0-\Delta)),\end{aligned} \tag{4.5}$$

where $\Theta(z)$ is the Heaviside function and, by construction of the problem, $0<z_0<L-\Delta$, where the origin is taken to be placed at the source of radiation. Applying these restrictions and Eq. (4.5) to Eq. (4.1) yields

$$\begin{aligned}F(\boldsymbol{\kappa},L;0) = F(\boldsymbol{\kappa},0;\boldsymbol{\kappa}L)\exp\Big[&-\langle\varepsilon\rangle\Delta + A_{\varepsilon\varepsilon}(0)\Delta\\ &+\{\langle\sigma\rangle - 2A_{\varepsilon\sigma}(0)\}\int_{z_0}^{z_0+\Delta} P(\boldsymbol{\kappa}(L-z))dz\\ &+A_{\sigma\sigma}(0)\int_{z_0}^{z_0+\Delta} P^2(\boldsymbol{\kappa}(L-z))dz\Big].\end{aligned} \tag{4.6}$$

Let the scattering phase function $f(\mathbf{n}_\perp,z)$ also be constant in the region Δ and, to keep the problem simple, let it be given by[16]

$$f(\mathbf{n}_\perp) = 4\alpha_p W_0 \exp(-\alpha_p n_\perp^2), \tag{4.7}$$

where W_0 is the single particle albedo and $\alpha_p \approx D^2/\lambda^2$ for a particle of diameter D scattering radiation of wavelength λ. Substituting Eq. (4.7) into the last relation of Eq. (3.4) and evaluating the integral in the polar plane of $\mathbf{n}_\perp$ gives

$$P(\mathbf{q}) = W_0\exp(-q^2/4\alpha_p). \tag{4.8}$$

Substituting Eq. (4.8) into Eq. (4.6) and performing the integrals yield the result

$$\begin{aligned}F(\boldsymbol{\kappa},L;0) = F(\boldsymbol{\kappa},0;\boldsymbol{\kappa}L)\exp\big[&(-\langle\varepsilon\rangle + A_{\varepsilon\varepsilon}(0))\Delta + (\sqrt{\pi}W_0/2C_1(\boldsymbol{\kappa}))\{(\langle\sigma\rangle - 2A_{\varepsilon\sigma}(0))(\Phi(C_1(\boldsymbol{\kappa})(L-z_0))\\ &-\Phi(C_1(\boldsymbol{\kappa})(L-z_0-\Delta))) + (W_0C_1(\boldsymbol{\kappa})/C_2(\boldsymbol{\kappa}))A_{\sigma\sigma}(0)(\Phi(C_2(\boldsymbol{\kappa})(L-z_0)) - \Phi(C_2(\boldsymbol{\kappa})(L-z_0-\Delta)))\}\big],\end{aligned} \tag{4.9}$$

where $\Phi(\cdots)$ is the error integral, $C_1(\boldsymbol{\kappa})\equiv\boldsymbol{\kappa}/(2\sqrt{\alpha_p})$, and $C_2(\boldsymbol{\kappa})\equiv\boldsymbol{\kappa}/\sqrt{(2\alpha_p)}$. Two extreme cases of this situation will now be considered where $\boldsymbol{\kappa}\gg 0$ as well as C_1, $C_2\gg 0$ is assumed.

Taking $z_0=0$ places the scattering layer at the source of the radiation and, when the condition $L\gg\Delta$ prevails, one finds, by the fact that $\Phi(\cdots)\sim 1$ for large values of its argument, that Eq. (4.9) becomes

$$F(\boldsymbol{\kappa},L;0) = F(\boldsymbol{\kappa},0;\boldsymbol{\kappa}L)\exp[(-\langle\varepsilon\rangle + A_{\varepsilon\varepsilon}(0))\Delta], \tag{4.10}$$

showing that only the $\boldsymbol{\kappa}$-independent extinction modifies the propagating spatial spectrum $F(\boldsymbol{\kappa},L;0)$. However, at the other extreme where $\Delta = L - z_0$, i.e., where the scattering layer is at the point of observation, Eq. (4.9) becomes

$$\begin{aligned}F(\boldsymbol{\kappa},L;0) = F(\boldsymbol{\kappa},0;\boldsymbol{\kappa}L)\exp\big[&(-\langle\varepsilon\rangle + A_{\varepsilon\varepsilon}(0))\Delta + (\sqrt{\pi}W_0/2C_1(\boldsymbol{\kappa}))\\ &\times\{(\langle\sigma\rangle - 2A_{\varepsilon\sigma}(0))(\Phi(C_1(\boldsymbol{\kappa})L) + (W_0C_1(\boldsymbol{\kappa})/C_2(\boldsymbol{\kappa}))A_{\sigma\sigma}(0)\Phi(C_2(\boldsymbol{\kappa})L)\}\big],\end{aligned} \tag{4.11}$$

which shows the large effect on the modification of the spectrum $F(\boldsymbol{\kappa},L;0)$ which, via the relation of Eq. (4.4), indicates that the illumination distribution is broadened. Thus, as is well known, the scattering layer has its strongest effect when the layer, in particular, the fluctuations A_{ij} of the scattering parameters, is close to the point of observation.

The solution of Eq. (3.19) is not so straightforward and an approximate expression will be derived. The form of Eq. (3.19) is not amenable to the Fourier convolution solution that was applied to Eq. (3.8) due to the presence of the term $\mathbf{r}_1 - \mathbf{r}_2$

in the arguments of two of the H_ϕ factors. However, if one limits the desired solutions to those values of $\mathbf{r}_1$ and $\mathbf{r}_2$, where $\mathbf{r}_1$, $\mathbf{r}_2 < \rho_c$, where ρ_c is the characteristic length scale of the fluctuations A_{ij}, one can assume that $H_\phi(\mathbf{r}-\mathbf{r}',z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|) \approx H_\phi(0,z;|\mathbf{n}_\perp - \mathbf{n}'_\perp|)$ thus lending to Eq. (3.19) a solution (albeit, an approximate one) via Fourier convolution and the method of characterisics as used earlier. The result is

$$G(\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2;L;\mathbf{q}_1,\mathbf{q}_2) = 2\int_0^L A_\phi(0,z;\mathbf{q}_1(z),\mathbf{q}_2(z))\langle F(\boldsymbol{\kappa}_1,z;\mathbf{q}_1(z))\rangle\langle F(\boldsymbol{\kappa}_2,z;\mathbf{q}_2(z))\rangle$$
$$\times\exp\Big[-\int_z^L\{2\langle\epsilon(z')\rangle - \langle\sigma(z')\rangle(P(\mathbf{q}_1(z'),z') + P(\mathbf{q}_2(z'),z'))$$
$$-(A_\phi(0,z';\mathbf{q}_1(z')) + 2A_\phi(0,z';\mathbf{q}_1(z'),\mathbf{q}_2(z')) + A_\phi(0,z';\mathbf{q}_2(z')))\}dz'\Big]dz, \tag{4.12}$$

where

$$B_I(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{n}_{\perp 1},\mathbf{n}_{\perp 2}) = \left(\frac{1}{2\pi}\right)^8 \int\!\!\int_{-\infty}^{\infty}\!\!\int\!\!\int_{-\infty}^{\infty}\!\!\int\!\!\int_{-\infty}^{\infty}\!\!\int\!\!\int_{-\infty}^{\infty} G(\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2;z;\mathbf{q}_1,\mathbf{q}_2)\exp(i\mathbf{r}_1\cdot\boldsymbol{\kappa}_1 + i\mathbf{q}_1\cdot\mathbf{n}_{\perp 1} + i\mathbf{r}_2\cdot\boldsymbol{\kappa}_2 + i\mathbf{q}_2\cdot\mathbf{n}_{\perp 2})d\mathbf{r}_1\, d\mathbf{n}_{\perp 1}\, d\mathbf{r}_2\, d\mathbf{n}_{\perp 2}, \tag{4.13}$$

with $\mathbf{q}_1(z) = \mathbf{q}_1 + \boldsymbol{\kappa}_1(L-z)$ and similarly for $\mathbf{q}_2$. The properties of this solution will not be dealt with here; its analysis would be greatly facilitated if the form of the correlation functions A_{ij} is known, which, of course, is specific to the particular propagation problem.

V. LIMITS OF APPLICABILITY OF THE RESULTS AND THEIR POSSIBLE EXTENSION

Summarizing the assumptions made throughout the foregoing, the results obtained here hold for scattering media characterized by overall propagation parameters that are taken to be zero-mean Gaussian random variables that are δ-correlated in the direction of propagation and propagation situations that are described by the small-angle scattering approximation to the radiative transfer equation and, of course, satisfy the restrictions attendant in the use of δ correlation. Such scenarios are applicable to many propagation problems. There are, however, two major extensions that can be made to the theory. The first logical step would be to apply the small-angle scattering formulation, i.e., Eq. (2.5), to situations where the medium fluctuations are governed by arbitrary statistics rather than the special Gaussian case considered here. This, of course, prohibits the use of the Novikov theorem and requires the use of an approach based on characteristic functionals of the propagation statistics. Retention of the assumption of the δ-correlatedness of the medium fluctuations could be justified since the requirements of the small-angle formulation are a subset of those of δ correlation. This will form the subject of a future publication.

The second extension would be to employ the full form of the radiative transfer equation, i.e., Eq. (2.1), in a similar analysis. One would necessarily need to use an approach different to the one given here since the use of the causality condition, employed in Appendix A, Eq. (A4), which greatly facilitated the analysis there, is no longer valid. In fact, questions of causality in general could be expected to hamper such a development. It is possible to simply introduce stochastic descriptions of extinction and scattering coefficients into the well known treatments of Eq. (2.1) in an analysis involving characteristic functionals. This, however, will require further investigation.

APPENDIX A: DERIVATION OF EQ. (3.6)

The evaluation of the Novikov theorem in this case is facilitated by first applying the realistic assumptions that (1) statistical homogeneity in the spatial coordinate $\mathbf{r}$ of $\tilde{\phi}$ prevails and (2) that the resulting correlation function is δ-correlated in the longitudinal z direction. The latter assumption places restrictions on the fluctuations of the medium.[17] These conditions, which are sufficient ones, are $\lambda \ll l_0$, $L \gg L_0$, $\lambda\alpha \ll 1$, and $\langle(\tilde{\epsilon}(\mathbf{r},z))^2\rangle, \langle(\tilde{\sigma}(\mathbf{r},z))^2\rangle \ll 1/(kL_0)$, where $\alpha = \langle\epsilon\rangle - \langle\sigma\rangle$ is the coefficient of absorption, L is the total distance of propagation, L_0 is the largest spatial extent of the fluctuations (sometimes called the outer scale), l_0 is that of the smallest spatial extent (sometimes called the inner scale), and k is the wave number, $k = 2\pi/\lambda$, where λ is the wavelength. Thus one has the following development:

$$B_\phi(\mathbf{r},\mathbf{r}';z,z';\mathbf{q},\mathbf{q}') = B_\phi(\mathbf{r}-\mathbf{r}',z,z';\mathbf{q},\mathbf{q}')$$
$$= B_{\epsilon\epsilon}(\mathbf{r}-\mathbf{r}',z,z') - B_{\epsilon\sigma}(\mathbf{r}-\mathbf{r}',z,z')P(\mathbf{q}',z') - B_{\sigma\epsilon}(\mathbf{r}-\mathbf{r}',z,z')P(\mathbf{q},z) + B_{\sigma\sigma}(\mathbf{r}-\mathbf{r}',z,z')P(\mathbf{q},z)P(\mathbf{q}',z')$$
$$= \{A_{\epsilon\epsilon}(\mathbf{r}-\mathbf{r}',z,z') - A_{\epsilon\sigma}(\mathbf{r}-\mathbf{r}',z,z')P(\mathbf{q}',z') - A_{\sigma\epsilon}(\mathbf{r}-\mathbf{r}',z,z')P(\mathbf{q},z) + A_{\sigma\sigma}(\mathbf{r}-\mathbf{r}',z,z')$$
$$\times P(\mathbf{q},z)P(\mathbf{q}',z')\}\delta(z-z'), \tag{A1}$$

where the definition of $\tilde{\phi}(\mathbf{r},z;\mathbf{q})$ from Eq. (3.4) was used in the expansion of the correlation function B_ϕ, thus giving rise to the corresponding correlation and cross-correlation functions B_{ij}, for $i,j = \varepsilon,\sigma$. The δ-correlation component is then factored out of the B_{ij} functions giving rise to associated two-dimensional correlation functions A_{ij}. Noting that the absolute z dependence exists in the factors $P(\mathbf{q},z)$, one can define a composite two-dimensional correlation $A_\phi(\mathbf{r}-\mathbf{r}',z,z';\mathbf{q},\mathbf{q}')$ to represent the quantity within the braces of Eq. (A1). Integration of Eq. (A1) over the coordinate z' gives Eq. (3.7a).

Substituting Eq. (A1) into Eq. (3.5) and performing the z' integration gives

$$\langle\tilde{\phi}(\mathbf{r},z;\mathbf{q})\tilde{J}(\mathbf{r},z;\mathbf{q})\rangle = \iint A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q},\mathbf{q}')\left\langle\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')}\right\rangle d\mathbf{r}'\,d\mathbf{q}'. \quad \text{(A2)}$$

It now remains to determine the average of the variational derivative as indicated in Eq. (A2). To this end, one must now consider the second equation of the system of equations obtained in Sec. II, viz., Eq. (2.13). Using the definitions of Eqs. (2.8) and (2.9) in Eq. (2.13) and applying the Fourier transform defined by Eq. (3.2) [for the same reasons it was allowed to be applied to Eq. (3.1)], solving for the derivative $\partial\tilde{J}(\mathbf{r},z;\mathbf{q})/\partial z$, and integrating along the z coordinate from 0 to z, one obtains

$$\begin{aligned}\tilde{J}(\mathbf{r},z;\mathbf{q}) - \tilde{J}(\mathbf{r},0;\mathbf{q}) &= -i\int_0^z \nabla_{\mathbf{r}}\cdot\frac{\partial\tilde{J}(\mathbf{r},z';\mathbf{q})}{\partial\mathbf{q}}\,dz' + \int_0^z\langle\phi(z';\mathbf{q})\rangle\tilde{J}(\mathbf{r},z';\mathbf{q})dz' \\ &\quad + \int_0^z\tilde{\phi}(\mathbf{r},z';\mathbf{q})J(\mathbf{r},z';\mathbf{q})dz' \\ &\quad - \int_0^z\langle\tilde{\phi}(\mathbf{r},z';\mathbf{q})\tilde{J}(\mathbf{r},z';\mathbf{q})\rangle dz'. \end{aligned} \quad \text{(A3)}$$

Formally, one can solve this equation iteratively and find that the quantity $\tilde{J}(\mathbf{r},z;\mathbf{q})$ is a function of the coordinate z for only those values z' such that $z'<z$. The values z' of z, where $z'>z$, do not enter into the solution. Thus fluctuations of the composite propagation parameter $\tilde{\phi}(\mathbf{r},z;\mathbf{q})$ at positions $z'<z$ can only influence the quantity $\tilde{J}(\mathbf{r},z;\mathbf{q})$; for the variational derivatives that must be considered in the sequel, this "causality condition" implies

$$\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')} = 0, \quad z'>z. \quad \text{(A4)}$$

Taking the variational derivative of Eq. (A3) and employing the fact that

$$\frac{\delta\tilde{\phi}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')} = \delta(\mathbf{r}-\mathbf{r}')\delta(z-z')\delta(\mathbf{q}-\mathbf{q}'), \quad \text{(A5)}$$

one obtains

$$\begin{aligned}\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')} &= -i\int_{z'}^z\nabla_{\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{q}}\left(\frac{\delta\tilde{J}(\mathbf{r},z'';\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}\right)dz'' + \int_{z'}^z\langle\phi(z'';\mathbf{q})\rangle\frac{\delta\tilde{J}(\mathbf{r},z'';\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}dz'' \\ &\quad + \int_{z'}^z\left[\delta(\mathbf{r}-\mathbf{r}')\delta(z''-z')\delta(\mathbf{q}-\mathbf{q}')J(\mathbf{r},z'';\mathbf{q}) + \tilde{\phi}(\mathbf{r},z'';\mathbf{q})\frac{\delta\tilde{J}(\mathbf{r},z'';\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}\right]dz'' \\ &\quad - \int_{z'}^z\left[\left\langle\delta(\mathbf{r}-\mathbf{r}')\delta(z''-z')\delta(\mathbf{q}-\mathbf{q}')\tilde{J}(\mathbf{r},z'';\mathbf{q}) + \tilde{\phi}(\mathbf{r},z'';\mathbf{q})\frac{\delta\tilde{J}(\mathbf{r},z'';\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}\right\rangle\right]dz''. \end{aligned} \quad \text{(A6)}$$

Finally, noting that the variational derivative desired in Eq. (A2) is related to the one given in Eq. (A6) via the relation

$$\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')} = \lim_{z'\to z}\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z';\mathbf{q}')}, \quad \text{(A7)}$$

and using the fact that the first, second, fourth, and sixth terms of Eq. (A6) converge to zero upon evoking the limit indicated in Eq. (A7), and that the fifth term gives zero since $\langle\tilde{I}\rangle = 0$, thus making $\langle\tilde{J}\rangle = 0$, one obtains

$$\begin{aligned}\frac{\delta\tilde{J}(\mathbf{r},z;\mathbf{q})}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')} &= \lim_{z'\to z}\int_{z'}^z\delta(\mathbf{r},\mathbf{r}')\delta(z''-z')\delta(\mathbf{q}-\mathbf{q}')J(\mathbf{r},z'';\mathbf{q})dz'' \\ &= \delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{q}-\mathbf{q}')J(\mathbf{r},z;\mathbf{q}). \end{aligned} \quad \text{(A8)}$$

Taking the ensemble average of this expression, substituting the result into Eq. (A2), and performing the required integrals, one obtains the result of Eq. (3.6) noting that the quantity $A_\phi(\mathbf{r}-\mathbf{r}',z;\mathbf{q},\mathbf{q}')$ becomes independent of the $\mathbf{q}'$ variable.

APPENDIX B: DERIVATION OF EQS. (3.13)–(3.16)

The derivation of Eq. (3.13) commences with the Novikov theorem, which, in the case of the δ-correlated random functions $\tilde{\phi}$, gives

$$\langle\tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle = \iint A_\phi(\mathbf{r}_1\mathbf{r}',z;\mathbf{q}_1\mathbf{q}') \times\left\langle\frac{\delta\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')}\right\rangle d\mathbf{r}'\,d\mathbf{q}'. \quad \text{(B1)}$$

This expression can also be obtained directly from Eq. (A2) by evaluating $\tilde{J}$ at the points $\mathbf{r}_2$ and $\mathbf{q}_2$. One must now evaluate the indicated variational derivative. This, too, can be obtained directly from a previous result in Appendix A, in particular, Eq. (A8), simply by letting $\mathbf{r}=\mathbf{r}_2$ and $\mathbf{q}=\mathbf{q}_2$, viz.,

$$\frac{\delta\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')} = \delta(\mathbf{r}_2-\mathbf{r}')\delta(\mathbf{q}_2-\mathbf{q}')J(\mathbf{r}_2,z;\mathbf{q}_2). \quad \text{(B2)}$$

Taking the ensemble average of this result, substituting it into Eq. (B1), and performing the required integrations yield

$$\langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\rangle = A_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{q}_1,\mathbf{q}_2)\langle J(\mathbf{r}_2,z;\mathbf{q}_2)\rangle,$$

which is the result of Eq. (3.13).

Equation (3.14) follows similarly and can be obtained from Eq. (3.13) by a simple transcription of $\mathbf{r}_1$ and $\mathbf{r}_2$ and $\mathbf{q}_1$ and $\mathbf{q}_2$.

The derivation of Eq. (3.15) requires one to employ the Novikov theorem once again, which gives, for the δ-correlated assumption made in the foregoing,

$$\langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\rangle = \int\int A_\phi(\mathbf{r}_1-\mathbf{r}',z;\mathbf{q}_1,\mathbf{q}_2) \times \left\langle \frac{\delta\{\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\}}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')}\right\rangle d\mathbf{r}'\,d\mathbf{q}'. \tag{B3}$$

Expanding the variational derivative of the product and using the result of Eq. (B2) yield

$$\frac{\delta\{\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\}}{\delta\tilde{\phi}(\mathbf{r}',z;\mathbf{q}')} = \delta(\mathbf{r}_2-\mathbf{r}')\delta(\mathbf{q}_2-\mathbf{q}')\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1) + \delta(\mathbf{r}_1-\mathbf{r}')\delta(\mathbf{q}_1-\mathbf{q}')\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2). \tag{B4}$$

Substituting this into Eq. (B3) and performing the integrations and using the various definitions given earlier give

$$\langle \tilde{\phi}(\mathbf{r}_1,z;\mathbf{q}_1)\tilde{J}(\mathbf{r}_2,z;\mathbf{q}_2)\tilde{J}(\mathbf{r}_1,z;\mathbf{q}_1)\rangle = [A_\phi(\mathbf{r}_1-\mathbf{r}_2,z;\mathbf{q}_1,\mathbf{q}_2)+A_\phi(0,z;\mathbf{q}_1)] \times B_j(\mathbf{r}_1,\mathbf{r}_2;z;\mathbf{q}_1,\mathbf{q}_2),$$

which is Eq. (3.15).

Equation (3.16) follows a similar way and can be obtained directly by transcribing $\mathbf{r}_1$ and $\mathbf{r}_2$ and $\mathbf{q}_1$ and $\mathbf{q}_2$ along with the application of the assumption of isotropy of the $\tilde{\phi}$ statistics.

[1]A. Schuster, Astrophys. J. **21**, 1 (1905).
[2]Yu. N. Barabanenkov, Sov. Phys. Usp. **18** (9), 673 (1975).
[3]For a good review, see R. L. Fante, Proc. IEEE **68**, 1424 (1980).
[4]G. V. Rozenberg, Sov. Phys. Usp. **69** (5), 666 (1959).
[5]K. Watson, J. Math. Phys. **10**, 688 (1969).
[6]S. Chandrasekhar, *Radiative Transfer* (Oxford, U. P., London, 1950).
[7]M. M. R. Williams, *Mathematical Methods in Particle Transport Theory* (Wiley, New York, 1971).
[8]G. I. Bell and S. Glasstone, *Nuclear Reactor Theory* (Van Nostrand Reinhold, New York, 1970).
[9]C. D. Levermore, G. C. Pomraning, D. L. Sanzo, and J. Wong, J. Math. Phys. **27**, 2526 (1986).
[10]C. D. Levermore, J. Wong, and G. C. Pomraning, J. Math. Phys. **29**, 995 (1988).
[11]D. Vanderhaegen, J. Quant. Spectros. Radiat. Transfer **36**, 557 (1986).
[12]D. Vanderhaegen, J. Quant. Spectros. Radiat. Transfer **39**, 333 (1988).
[13]E. A. Novikov, Sov. Phys. JETP **20** (5), 1290 (1965).
[14]See, for example, D. Arnush, J. Opt. Soc. Am. **62** (9), 1109 (1972).
[15]A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[16]A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Vol. 1, *Single Scattering and Transport Theory* (Academic, New York, 1978), p. 238.
[17]V. I. Klyatskin, Sov. Phys. JETP **30** (3), 520 (1970).

Intervals of electrohydrodynamic Rayleigh–Taylor instability. I. Effect of a tangential periodic field

Elsayed F. Elshehawey[a)]
Department of Mathematics and Computer Sciences, Faculty of Science, U.A.E. University, Al-Ain P.O. Box 15551, United Arab Emirates

(Received 25 January 1989; accepted for publication 17 May 1989)

The intervals of electrohydrodynamic Rayleigh–Taylor instability influenced by a periodic tangential field are considered. It is shown that a linear model of the interface is governed by Hill's differential equation. Characteristic values and intervals of stability are discussed. The special case of the Mathieu differential equation type is obtained.

I. INTRODUCTION

Problems of electrohydrodynamic stability have been considered by many authors. (See Melcher,[1,2] Woodson and Melcher,[3] Mohamed and Elshehawey,[4–6] Elshehawey,[7] Elshehawey *et al.*[8] El Dabe *et al.*,[9] and Mohamed *et al.*[10,11] and the references therein.) There are some physical situations when one needs a limited band of wavenumbers to achieve instability; at the same time, for values of wavenumbers less or greater than this band stability is required. For example, in biophysics[12] the cell membrane is formed by a number of adjacent cells if they are subjected to a periodic field. Also, the membrane breaks down if a field at a given strength is applied to it. Also, if a force varying periodically with time acts on a mass in such a manner that the force tends to move the mass back into a position of equilibrium in proportion to the dislocation of the mass, one might expect the mass to be confined to a neighborhood of the position of equilibrium. In particular, once the force is strong enough to achieve this effect, one would expect a stronger force to be even more efficient for this purpose. An increase of the restraining force may cause the mass to oscillate with wider and wider amplitudes. The theory of the intervals of instability provides the precise description of this phenomenon.

In the present paper, we shall confine ourselves to giving a general description of the so-called regions of absolute stability since in most previous cases, the results are based on numerical computations.

II. FORMULATION OF THE PROBLEM

Consider two semi-infinite dielectric inviscid fluids separated by the plane $y=0$. The upper and lower densities of the fluids are $\rho^{(2)}$ and $\rho^{(1)}$, respectively. The fluids are influenced by a periodic electric field

$$\mathbf{E}_0 = E^*E_0(t)\mathbf{e}_x, \quad \int_0^\pi E_0^2(t)\,dt = 0, \tag{2.1}$$

where $\mathbf{e}_x$ is the unit vector in the x direction. We shall consider all "functions $E_0^2(t)$ of class P" which are defined by

$$\left[\pi\int_0^\pi |E_0^2(t)|^p\,dt\right]^{1/p} = 1, \tag{2.2}$$

where $p = 1,2,3,\ldots$, or $p=\infty$. If $p=\infty$, (2.2) means that

$$\max|E_0^2(t)| = 1. \tag{2.3}$$

We assume that $E_0^2(t)$ is continuous except for a finite number of points, where $E_0^2(t)$ may have a jump. We dimensionalize the various quantities using the characteristic length $L=(t/\rho^{(1)}g)^{1/2}$ and the characteristic time $(L/g)^{1/2}$. The motion considered here is irrotational and there exists a velocity potential ϕ such that $\mathbf{v}=\nabla\phi$.

The velocity potential ϕ satisifies

$$\frac{\partial^2\phi^{(2),(1)}}{\partial x^2} + \frac{\partial^2\phi^{(2),(1)}}{\partial y^2} = 0 \tag{2.4}$$

such that

$$|\nabla\phi^{(2)}| \to 0, \quad \text{as } y\to\infty, \tag{2.5}$$

$$|\nabla\phi^{(1)}| \to 0, \quad \text{as } y\to-\infty. \tag{2.6}$$

The surface deflection is expressed as

$$y = \xi(x,t). \tag{2.7}$$

Then,

$$\mathbf{n} = \nabla F/[\nabla F\cdot\nabla F]^{1/2} = -\xi_x(\xi_x^2+1)^{-1/2}\mathbf{e}_x + (\xi_x^2+1)^{-1/2}\mathbf{e}_y \tag{2.8}$$

is the unit normal $\mathbf{n}$ to the surface, and $F=0$ is the equation to the surface of separation.

The condition that the interface is moving with the fluid leads to

$$\xi_t - \phi_y^{(2),(1)} + \phi_x^{(2),(1)}\xi_x = 0, \quad \text{at } y=\xi. \tag{2.9}$$

We assume that the quasistatic approximation is valid and we introduce the electrostatic potential $\Psi^{(2),(1)}$ such that

$$\mathbf{E}^{(2),(1)} = E^*E_0(t)\mathbf{e}_x - \nabla\Psi^{(2),(1)}. \tag{2.10}$$

Therefore, the differential equation satisfied by $\Psi^{(2),(1)}$ is the Laplace equation

$$\frac{\partial^2\Psi^{(2),(1)}}{\partial x^2} + \frac{\partial^2\Psi^{(2),(1)}}{\partial y^2} = 0 \tag{2.11}$$

along with the following boundary conditions.

(i) The tangential component of the electric field is continuous at the interface

$$\mathbf{n}\wedge(\mathbf{E}^{(2)} - \mathbf{E}^{(1)}) = 0, \quad \text{at } y=\xi, \tag{2.12}$$

which leads to

$$\xi_x(\Psi_y^{(2)} - \Psi_y^{(1)}) + (\Psi_x^{(2)} - \Psi_x^{(1)}) = 0, \quad \text{at } y=\xi. \tag{2.13}$$

[a)] On leave from the Department of Mathematics, Faculty of Education, Ain Shams University, Heliopolis, Cairo, Egypt.

 0022-2488/89/102441-04$02.50

(ii) The normal electric displacement is continuous at the interface $y=\xi(x,t)$:

$$\mathbf{n}\cdot(\bar{\epsilon}^{(2)}\mathbf{E}^{(2)}-\bar{\epsilon}^{(1)}\mathbf{E}^{(1)})=0, \quad \text{at } y=\xi \tag{2.14}$$

and hence,

$$\xi_x(\bar{\epsilon}^{(2)}\Psi_x^{(2)}-\bar{\epsilon}^{(1)}\Psi_x^{(1)})-(\bar{\epsilon}^{(2)}\Psi_y^{(2)}-\bar{\epsilon}^{(1)}\Psi_y^{(1)}) =\xi_x E^*E_0(t)(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)}), \quad \text{at } y=\xi, \tag{2.15}$$

where $\bar{\epsilon}^{(2),(1)}$ is the dielectric constant.

(iii) The normal hydrodynamic stress is balanced by the normal electric stress. The balance condition is then[2]

$$\begin{aligned}
&\phi_t^{(1)}-\rho\phi_t^{(2)}+(1-\rho)\xi+\tfrac{1}{2}(\phi_x^{(1)^2}-\rho\phi_x^{(2)^2})+\tfrac{1}{2}(\phi_y^{(1)^2}-\rho\phi_y^{(2)^2})\\
&\quad=\xi_{xx}(1+\xi_x^2)^{-3/2}+E^*E_0(t)(\bar{\epsilon}^{(2)}\Psi_x^{(2)}-\bar{\epsilon}^{(1)}\Psi_x^{(1)})-\tfrac{1}{2}(\bar{\epsilon}^{(2)}\Psi_x^{(2)^2}-\bar{\epsilon}^{(1)}\Psi_x^{(1)^2})+\tfrac{1}{2}(\bar{\epsilon}^{(2)}\Psi_y^{(2)^2}-\bar{\epsilon}^{(1)}\Psi_y^{(1)^2})\\
&\qquad+2\xi_x E^*E_0(t)(\bar{\epsilon}^{(2)}\Psi_y^{(2)}-\bar{\epsilon}^{(1)}\Psi_y^{(1)})-2\xi_x(\bar{\epsilon}^{(2)}\Psi_x^{(2)}\Psi_y^{(2)}-\bar{\epsilon}^{(1)}\Psi_x^{(1)}\Psi_y^{(1)})+\xi_x^2E^{*2}E_0^2(t)(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})\\
&\qquad-2E^*E_0(t)\,\xi_x^2(\bar{\epsilon}^{(2)}\Psi_x^{(2)}-\bar{\epsilon}^{(1)}\Psi_x^{(1)})+\xi_x^2(\bar{\epsilon}^{(2)}\Psi_x^{(2)^2}-\bar{\epsilon}^{(1)}\Psi_x^{(1)^2})-\xi_x^2(\bar{\epsilon}^{(2)}\Psi_y^{(2)^2}-\bar{\epsilon}^{(1)}\Psi_y^{(1)^2})\\
&\qquad-2\xi_x^3E^*E_0(t)(\bar{\epsilon}^{(2)}\Psi_y^{(2)}-\bar{\epsilon}^{(1)}\Psi_y^{(1)})+2\xi_x^3(\bar{\epsilon}^{(2)}\Psi_x^{(2)}\Psi_y^{(2)}-\bar{\epsilon}^{(1)}\Psi_x^{(1)}\Psi_y^{(1)})+O(\xi_x^4), \quad \text{at } y=\xi(x,t).
\end{aligned} \tag{2.16}$$

Equations (2.4)–(2.16) will be solved using the method of multiple scale.[13]

We introduce the scales X_n and T_n defined by

$$X_n=\epsilon^n x, \quad T_n=\epsilon^n t. \tag{2.17}$$

We may also expand ξ, $\Psi^{(2),(1)}$, and $\phi^{(2),(1)}$ in the form

$$\xi(x,t)=\sum_{n=1}^{3}\epsilon^n\xi_n(X_0,X_1,X_2;T_0,T_1,T_2)+O(\epsilon^4), \tag{2.18}$$

$$\Psi_n^{(2),(1)}(x,y,t)=\sum_{n=1}^{3}\epsilon^n\Psi_n^{(2),(1)}(X_0,X_1,X_2;y;T_0,T_1,T_2)+O(\epsilon^4), \tag{2.19}$$

$$\phi_n^{(2),(1)}(x,y,t)=\sum_{n=1}^{3}\epsilon^n\phi_n^{(2),(1)}(X_0,X_1,X_2;y;T_0,T_1,T_2)+O(\epsilon^4). \tag{2.20}$$

We now substitute from Eqs. (2.18)–(2.20) into (2.4)–(2.16) and equate coefficients of like powers of ϵ.

The problem considered here is the intervals of the linear electrohydrodynamic stability of a single interface stressed by a tangential periodic electric field. The effect of nonlinearity on the problem at hand will not be discussed here and will be the subject of a subsequent paper.

The solution of the first-order problem for traveling waves with respect to the variable X_0 that decays far from the interface is

$$\xi_1=D(X_1,X_2;t)e^{ikX_0}+\bar{D}(X_1,X_2;t)e^{-ikX_0}, \tag{2.21}$$

$$\Psi_1^{(2)}=[iE^*E_0(t)(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})/(\bar{\epsilon}^{(2)}+\bar{\epsilon}^{(1)})]\times[D(X_1,X_2;t)e^{ikX_0-ky}-\bar{D}e^{-ikX_0-ky}], \tag{2.22}$$

$$\Psi_1^{(1)}=[iE^*E_0(t)(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})/(\bar{\epsilon}^{(2)}+\bar{\epsilon}^{(1)})]\times[D(X_1,X_2;t)e^{ikX_0+ky}-\bar{D}e^{-ikX_0+ky}], \tag{2.23}$$

$$\phi_1^{(2)}=-\frac{1}{k}\frac{\partial D}{\partial t}e^{ikX_0-ky}-\frac{1}{k}\frac{\partial\bar{D}}{\partial t}e^{-ikX_0-ky}, \tag{2.24}$$

$$\phi_1^{(1)}=\frac{1}{k}\frac{\partial D}{\partial t}e^{ikX_0+ky}+\frac{1}{k}\frac{\partial\bar{D}}{\partial t}e^{-ikX_0+ky}, \tag{2.25}$$

$$\frac{\partial^2D}{\partial t^2}+\frac{k}{(1+\rho)}\times\left[1-\rho+k^2+\frac{kE^{*2}E_0^2(t)(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})^2}{(\bar{\epsilon}^{(2)}+\bar{\epsilon}^{(1)})}\right]D=0. \tag{2.26}$$

As a special case of Eq. (2.26), which is the well-known dispersion relation, if we replace the periodic electric field by a constant field, we obtain the same result given in Ref. 4 for the linear system, where

$$K_c=\sqrt{\rho-1}(\cosh\theta_E-\sinh\theta_E),$$
$$\sinh\theta_E=\alpha_E/2\sqrt{\rho-1},$$
$$\alpha_E=E^{*2}(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})^2/(\bar{\epsilon}^{(2)}+\bar{\epsilon}^{(1)}).$$

III. THE HILL EQUATION

We put Eq. (2.26) in the standard form

$$\frac{\partial^2D}{\partial t^2}+[\lambda+Q(t)]D=0, \tag{3.1}$$

where

$$\lambda=[k/(1+\rho)](1-\rho+k^2), \tag{3.2}$$

$$Q(t)=\beta E_0^2(t), \tag{3.3}$$

$$\beta=\frac{k^2E^{*2}(\bar{\epsilon}^{(2)}-\bar{\epsilon}^{(1)})^2}{(1+\rho)(\bar{\epsilon}^{(2)}+\bar{\epsilon}^{(1)})}, \quad Q(t+\pi)=Q(t), \tag{3.4}$$

where λ is a parameter depending on the ratio density ρ and wavenumber k. Here, $\beta>0$ and $Q(t)$ is a real periodic function of t with period π.

Here, we determine the values of λ for which the solutions of the Hill equation (3.1) are stable. (See Refs. 14–17 and the references therein.) Following the methods of Magnus and Winkler,[16] one can show that for the given Hill equation (3.1) there belong two monotonically increasing infinite sequences of real numbers

$$\lambda_0, \lambda_1, \lambda_2, \ldots, \tag{3.5}$$

$$\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4, \ldots \tag{3.6}$$

such that Eq. (3.1) has a solution of period π iff $\lambda = \lambda_n$, $n = 0, 1, 2, \ldots$ and a solution of period 2π iff $\lambda = \lambda'_n$, $n = 1, 2, 3, \ldots$.

The λ_n and λ'_n satisfy the inequalities

$$\lambda_0 < \lambda'_1 \leqslant \lambda'_2 < \lambda_1 \leqslant \lambda_2 < \lambda'_3 \leqslant \lambda'_4 < \lambda_3 \leqslant \lambda_4 < \cdots \tag{3.7}$$

and the relations

$$\lim_{n\to\infty} \lambda_n^{(-1)} = 0, \qquad \lim_{n\to\infty} (\lambda'_n)^{-1} = 0. \tag{3.8}$$

The solutions of (3.1) are stable in the intervals

$$(\lambda_0, \lambda'_1), \quad (\lambda'_2, \lambda_1), \quad (\lambda_2, \lambda'_3), \quad (\lambda'_4, \lambda_3), \ldots . \tag{3.9}$$

At the endpoints of the intervals (3.9) the solutions of (3.1) are in general unstable: This is always true for $\lambda = \lambda_0$. The solutions of (3.1) are stable for $\lambda = \lambda_{2n+1}$ or $\lambda = \lambda_{2n+2}$ iff $\lambda_{2n+1} = \lambda_{2n+2}$ and they are stable for $\lambda = \lambda'_{2n+1}$ or $\lambda = \lambda'_{2n+2}$ iff $\lambda'_{2n+1} = \lambda'_{2n+2}$. For complex values of λ, Eq. (3.1) always has unstable solutions and it cannot happen since ρ and k are real here [see Eq. (3.2)].

The λ_n are the roots of $\Delta(\lambda) = 2$ and the λ'_n are those of $\Delta(\lambda) = -2$, where

$$\Delta(\lambda) = D_1(\pi, \lambda) + D'_2(\pi, \lambda). \tag{3.10}$$

The intervals of instability $(-\infty, \lambda_0)$ will always be present (the zeroth interval of instability) and we define (λ'_1, λ'_2) as the first interval of instability.

We observe that neither an interval of stability nor an interval of instability can ever shrink to a point. The intervals of stability can never disappear, but two of them can combine to a single one if $\lambda_{2n+1} = \lambda_{2n+2}$ or $\lambda'_{2n+1} = \lambda'_{2n+2}$. However, the interval of instability (with the exception of the zeroth intervals) may disappear altogether. This takes place if $Q(t) = \beta E_0^2(t)$ is a constant (i.e., for the case of the constant tangential electric field given in Ref. 6).

A region in the real λ, β plane will be called a region of absolute stability for functions of class p if for any point in this region (3.1) will have stable solutions for all functions $Q(t) = \beta E_0^2(t)$, where $E_0^2(t)$ belongs to the class p.

Let $n = 0, 1, 2, \ldots$. The region of absolute stability for the functions $E_0^2(t)$ of class 1 is bounded by the curves

$$\begin{aligned} \beta_{n+1} &= \pm [4(n+1)\sqrt{\lambda}/\pi] \cot[\pi\sqrt{\lambda}/2(n+1)], \\ & \quad n^2 < \lambda < (n+1)^2, \\ \beta_n &= \pm 2\lambda(1 - n/\sqrt{\lambda}), \quad \lambda > 1, \quad n \geqslant 1, \\ \lambda &= 0, \quad \text{for } n = 0 \text{ (i.e., } \rho = 1 + k^2), \end{aligned} \tag{3.11}$$

and is such that none of these curves is contained in its interior.

The open region bounded by the curves (3.11) is maximal; for any point outside or on the boundary of this region, there exists a function $E_0^2(t)$ of class 1 such that not all solutions of (3.1) are bounded. Also, let m be a real variable, $0 < m^2 < 1$, and let

$$M = \int_0^{\pi/2} \frac{ds}{\sqrt{1 - m^2 \sin^2 s}}, \quad E = \int_0^{\pi/2} \sqrt{1 - m^2 \sin^2 s}\, ds. \tag{3.12}$$

Then the curves defined for $n = 0, 1, 2, \ldots$ by

$$\begin{aligned} \beta_{n+1} &= \pm 8.3^{-1/2}\pi^{-2}(n+1)^2 M [M^2(m^2 - 1) \\ &\quad + 2ME(2 - m^2) - 3E^2]^{1/2}, \end{aligned} \tag{3.13}$$

$$\begin{aligned} \lambda_{n+1} &= 4\pi^{-2}(n+1)^2[M^2(m^2-1) \\ &\quad + 2ME], \quad \lambda > 0, \end{aligned} \tag{3.14}$$

bound the region of absolute stability of the functions of class 2. The boundary points do not belong to the region since for

$$\begin{aligned} \lambda + \beta E_0^2(t) &= 4\pi^{-2}(n+1)^2 M^2(1 + m^2) \\ &\quad - 8\pi^{-2}(n+1)^2 m^2 M^2 \\ &\quad \times \mathrm{sn}^2(2(n+1)Mt/\pi), \end{aligned} \tag{3.15}$$

the differential equation (3.1) has only one periodic solution (and therefore, at least one unbounded solution).

The periodic solution (with period π or 2π) is

$$D_p = \mathrm{sn}\, \tau, \quad \tau = 2(n+1)Mt/\pi, \tag{3.16}$$

where sn τ is the Jacobian elliptic function with module m and period $4M$.

Also, for the functions of class ∞, the region of absolute stability is bounded by the curves

$$\begin{aligned} &(\lambda_{n+1} + \beta_{n+1})^{1/2} \tan\left[\pi\sqrt{\lambda_{n+1} + \beta_{n+1}}/4(n+1)\right] \\ &\quad = (\lambda_{n+1} - \beta_{n+1})^{1/2} \\ &\quad \times \cot\left[\pi\sqrt{\lambda_{n+1} - \beta_{n+1}}/4(n+1)\right], \end{aligned} \tag{3.17}$$

where $n = 0, 1, 2, \ldots$ and where the region does not contain any of these curves in its interior. If one of the square roots should be imaginary, the functions tan and cot have to be replaced by the corresponding hyperbolic functions, i.e.,

$$E^{*4} > (1 - \rho + k^2)^2(\tilde{\epsilon}^{(2)} + \tilde{\epsilon}^{(1)})^2/k^2(\tilde{\epsilon}^{(2)} - \tilde{\epsilon}^{(1)})^4. \tag{3.18}$$

Also, if a and b are real numbers and

$$a^2 \leqslant \lambda + \beta E_0^2(t) \leqslant b^2, \tag{3.19}$$

then the solutions of (3.1) will be stable for all possible $\lambda + \beta E_0^2(t)$ satisfying this condition iff the interval (a^2, b^2) does not contain the square of an integer.

IV. THE MATHIEU EQUATION

If we take $\lambda = [k/(1+\rho)](1 - \rho + k^2)$ as (3.2) and

$$E_0^2(t) = \cos 2t, \tag{4.1}$$

$$q = \frac{-k^2 E^{*2}(\tilde{\epsilon}^{(2)} - \tilde{\epsilon}^{(1)})^2}{2(1+\rho)(\tilde{\epsilon}^{(2)} + \tilde{\epsilon}^{(1)})} = -\frac{\beta}{2}, \tag{4.2}$$

Eq. (3.1) becomes

$$\frac{\partial^2 D}{\partial t^2} + [\lambda - 2q\cos 2t]D = 0, \tag{4.3}$$

which is the Mathieu differential equation.[18]

According to the Floquet theorem,[17] the general periodic solution of the Mathieu differential equation given by (4.3) can be written as

$$D_p(X_1, X_2, t) = F_1(X_1, X_2)e^{\mu t}H(t) + F_2(X_1, X_2)e^{-\mu t}H(-t), \tag{4.4}$$

where $H(t)$ is a periodic function in t of period 2π or π; $F_1(X_1, X_2)$, $F_2(X_1, X_2)$ are arbitrary constants; and μ is a parameter given by the relation

$$\sin^2 i\mu\pi = \Delta(0)\sin^2 \tfrac{1}{2}\pi\sqrt{\lambda}. \tag{4.5}$$

Here, $\Delta(0)$ is an infinite Hill determinant depending on λ and q (see Ref. 18) and takes the form

$$\Delta(0) \simeq 1 - \pi\lambda^2 \cot \tfrac{1}{2}\pi\lambda^2/4\sqrt{\lambda}\,(\lambda - 1). \tag{4.6}$$

It is seen from Eq. (4.4) that if μ is pure imaginary, the solution for D_p will be bounded as $t \to \infty$ and the system is stable. The characteristic curves of the Mathieu functions and the regions of stability and instability are discussed in Ref. 18. In the (λ,q) plane, the regions in which the values of λ and q yield imaginary values of μ are the stable regions.[18] On the other hand, if μ is real, the solution for D_p will tend to ∞ as $t \to \infty$.

The unstable regions [in the (λ,q) plane] are the regions in which the values of λ and q correspond to real values of μ; the boundary curves of these regions are symmetric about the λ axis. On the other hand, we assume that

$$q = -k^2E^{*2}(\tilde{\epsilon}^{(2)} - \tilde{\epsilon}^{(1)})^2/2(1+\rho)(\tilde{\epsilon}^{(2)} + \tilde{\epsilon}^{(1)})$$

are small (which is a good approximation to high-frequency fields or large wavenumbers). Then following Morse and Feshbach[14] one can show that the solution of Eq. (4.3) will be bounded at $t \to \infty$ provided that q and λ satisfy the inequality

$$4q^2 - 32(1-\lambda)q + 32\lambda(1-\lambda) > 0 \tag{4.7}$$

or

$$\frac{k^4E^{*4}(\tilde{\epsilon}^{(2)} - \tilde{\epsilon}^{(1)})^4}{(1+\rho)^2(\tilde{\epsilon}^{(2)} + \tilde{\epsilon}^{(1)})^2} + \frac{16k^2E^{*2}(\tilde{\epsilon}^{(2)} - \tilde{\epsilon}^{(1)})^2(1-\lambda)}{(1+\rho)(\tilde{\epsilon}^{(2)} + \tilde{\epsilon}^{(1)})} + 32\lambda(1-\lambda) > 0. \tag{4.8}$$

Also, if $\beta E_0^2(t) + \lambda > 0$ and

$$\int_0^\pi [\beta E_0^2(t) + \lambda]^2\, dt < \left(\frac{64}{(3\pi^2)}\right)\left\{\int_0^{\pi/2} \frac{ds}{\sqrt{1+\sin^2 s}}\right\}^4 = \frac{1}{12}\left[\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{3}{4})}\right]^4, \tag{4.9}$$

then the solutions of the Mathieu equation (4.3) are stable.

ACKNOWLEDGMENTS

The author wishes to express his cordial thanks to Professor A. A. Mohamed for helpful discussions. He is also grateful to the referees and the editor for their interesting suggestions and comments.

[1]J. R. Melcher, *Field Coupled Surface Waves* (MIT, Cambridge, 1963).
[2]J. R. Melcher, *Continuum Electromechanics* (MIT, Cambridge, 1981).
[3]H. H. Woodson and J. R. Melcher, *Electromechanical Dynamics* (Wiley, New York, 1968), Part II.
[4]A. A. Mohamed and E. F. Elshehawey, "Nonlinear electrohydrodynamic Rayleigh–Taylor instability. I. A perpendicular field in the absence of surface charges," J. Fluid Mech. **129**, 473 (1983).
[5]A. A. Mohamed and E. F. Elshehawey, "Nonlinear electrohydrodynamic Rayleigh–Taylor instability. II. A perpendicular field producing surface charge, Phys. Fluids **26**, 1724 (1983).
[6]A. A. Mohamed and E. F. Elshehawey," Nonlinear electrohydrodynamic Rayleigh–Taylor instability. III. Effect of a tangential field," Arabian J. Sci. Engrg. **9**, 345 (1984).
[7]E. F. Elshehawey, "Electrohydrodynamic solitons in Kelvin–Helmholtz flow. The case of a normal field in the absence of surface charges," Q. Appl. Math. **XLIII**, 483 (1986).
[8]E. F. Elshehawey, Y. O. ElDib, and A. A. Mohamed, "Electrohydrodynamic stability of a fluid layer. I. Effect of a tangential field," Nuovo Cimento D **6**, 291 (1985).
[9]N. T. ElDabe, E. F. Elshehawey, G. M. Moatimid, and A. A. Mohamed, "Electrohydrodynamic stability of two cylindrical interfaces under the influence of a tangential periodic electric field," J. Math. Phys. **26**, 2072 (1985).
[10]A. A. Mohamed, E. F. Elshehawey, and Y. O. ElDib, "Electrohydrodynamic stability of a fluid layer. II. Effect of a normal electric field," J. Chem. Phys. **85**, 445 (1986).
[11]A. A. Mohamed, E. F. Elshehawey, and Y. O. ElDib, Electrohydrodynamic stability of a fluid layer. Effect of a tangential periodic field," Nuovo Cimento D **8**, 177 (1986).
[12]U. Zimmerman, Biochim. Biophys. Acta **694**, 227 (1982).
[13]A. H. Nayfeh, *Perturbation Methods* (Wiley Interscience, New York, 1973).
[14]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I.
[15]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[16]W. Magnus and S. Winkler, *Hill's Equation* (Dover, New York, 1979).
[17]E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1927).
[18]N. W. McLanchlan, *Theory and Applications of the Mathieu Functions* (Clarendon, Oxford, 1947).

Global solution of the Boltzmann equation for rigid spheres and initial data close to a local Maxwellian

A. Palczewski and G. Toscani
Dipartimento di Matematica, Università di Ferrara, 44100 Ferrara, Italy

(Received 18 November 1988; accepted for publication 10 May 1989)

In this paper, which is an extension of the previous paper of one of the authors [Arch. Rat. Mech. Anal. **102**, 231 (1988)], a global existence theorem is presented for the Boltzmann equation for initial data close to a local Maxwellian of a special type. Contrary to the previous result, the global existence is proved for rigid spheres and "hard" potentials with angular cutoff.

I. INTRODUCTION

After recent results of DiPerna and Lions,[1,2] the main problem of the Boltzmann equation, i.e., the existence of global solutions, seems to be solved. On the other hand, the problem of the uniqueness of the solution is open, and the asymptotic behavior of the gas density has been only partially investigated in a recent paper by Arkeryd.[3] These facts still make different approaches justified.

In the present paper we consider a particular case when initial data are not small but close to a given local Maxwellian. We prove global existence and uniqueness for a corresponding initial value problem. The paper is a continuation of the previous paper of one of the authors[4] when similar results were obtained for soft interactions and Maxwellian potentials with angular cutoff.

In the present paper we consider a gas of rigid spheres. Since, as it was observed earlier (cf. Illner and Shinbrot[5]), rigid spheres create some peculiar technical difficulties, we begin with the analysis of the Boltzmann equation for "hard" potentials with angular cutoff. Then, the rigid spheres result is obtained as a limit of solutions for hard potentials. The proof of existence is similar to that used in Ref. 4 and is based on the Kaniel and Shinbrot iteration scheme.[5] What is worth noting is the fact that although we start with initial data close to a Maxwellian there is no trend to equilibrium. This shows a substantial difference between local and global Maxwellians for which the trend to equilibrium always holds.

II. THE BOLTZMANN EQUATION

Our aim is to solve an initial value problem of the nonlinear Boltzmann equation. In the absence of an external force field and with initial data defined in the whole space this problem can be written as follows:

$$\frac{\partial f}{\partial t} + v\cdot \text{grad}_x f = J(f,f), \quad x\in R^3, \quad v\in R^3, \quad t\in R^+, \tag{1a}$$

$$f(x,v,0) = \varphi(x,v). \tag{1b}$$

For gas particles interacting by cutoff potentials or rigid sphere interactions the collision operator J can be split into gain and loss terms:

$$\frac{\partial f}{\partial t} + v\cdot \text{grad}_x f = Q(f,f) - fR(f), \tag{2}$$

where

$$Q(f,g)(x,v,t) = \frac{1}{2}\int_D B(q,\theta)[f(v_1')g(v') + f(v')g(v_1')]dv_1\, d\epsilon\, d\theta, \tag{3}$$

$$R(f) = \int_D B(q,\theta) f(v_1) dv_1\, d\epsilon\, d\theta. \tag{4}$$

Here (v,v_1) and (v',v_1') denote the precollisional and postcollisional velocities of the two colliding particles and $q = v_1 - v$ is the relative velocity. Here $\psi = \pi - 2\theta$ is the scattering angle of the binary collision and ϵ is the azimuthal angle of the plane in which the collision takes place. Consequently $D = R^3 \times [0,\pi/2] \times [0,2\pi]$.

To begin with, let us give essential notation and preliminary results. First, let us denote

$$f^{\#}(x,v,t) = f(x+vt,v,t).$$

Then Eq. (1) can be written as follows:

$$\frac{\partial f^{\#}}{\partial t} + f^{\#}R^{\#}(f) = Q^{\#}(f,f), \tag{5a}$$

$$f^{\#}(0) = \varphi(x,v). \tag{5b}$$

Following the approach of Kaniel and Shinbrot we will solve Eq. (5) by the iterative scheme,

$$\frac{\partial l_n^{\#}}{\partial t} + l_n^{\#} R^{\#}(u_{n-1}) = Q^{\#}(l_{n-1},l_{n-1}), \tag{6a}$$

$$\frac{\partial u_n^{\#}}{\partial t} + u_n^{\#} R^{\#}(l_{n-1}) = Q^{\#}(u_{n-1},u_{n-1}), \tag{6b}$$

$$l_n(0) = u_n(0) = \varphi(x,v). \tag{6c}$$

To begin iterations we need a pair of functions (l_0,u_0). Following Kaniel and Shinbrot we say that such a pair satisfies the beginning condition if

$$0 \leqslant l_0(t) \leqslant l_1(t) \leqslant u_1(t) \leqslant u_0(t).$$

It is well known from the results of Kaniel and Shinbrot[6] that if $\varphi \geqslant 0$ and (l_0,u_0) satisfy the beginning condition then the sequence of iterative solutions l_n is increasing and u_n decreasing and both converge (in the L^1 sense) to a mild solution of (5).

 0022-2488/89/102445-06$02.50

III. GLOBAL EXISTENCE FOR HARD POTENTIALS

In this section we shall prove that for initial data close to a special local Maxwellian it is possible to construct two functions u_0 and l_0 which satisfy the beginning condition. As we know, local Maxwellians are given in the form

$$\omega(x,v,t) = (2\pi T)^{-3/2} n \exp(-|v-u|^2/2T), \quad (7)$$

where $n = n(x,t)$, $u = u(x,t)$, and $T = T(x,t)$ are the fluid-dynamical parameters denoting the mass density, mean velocity, and temperature of a gas (for a general discussion of local Maxwellians see Truesdell and Muncaster[7]).

For the purpose of this paper we shall assume u and T to be constant and $n = n(x) = \exp(-\alpha x^2)$. Hence our local Maxwellian has the form

$$\omega(x,v) = c \exp(-\alpha x^2)\exp(-\beta v^2). \quad (8)$$

Given a steady Maxwellian ω in the form (8) we want to solve the Cauchy problem (5) in space $L^1(R^6)$, or strictly speaking, in the subspace $\{f(x,v)\epsilon L^1(R^6)\colon 0 \leqslant f(x,v) \leqslant c\omega(x,v)\}$ with an initial value $\varphi(x,v)$, which is close to ω in the sense that

$$(1-\epsilon)\omega(x,v) \leqslant \varphi(x,v) \leqslant (1+\epsilon)\omega(x,v), \quad (9)$$

with ϵ sufficiently small.

As we know from the previous section, the essential step in solving the Cauchy problem (5) is the construction of the beginning condition. To this purpose, let us set

$$\begin{aligned} u_0^{\#}(x,v,t) &= (1+\epsilon(t))\omega(x,v), \\ l_0^{\#}(x,v,t) &= (1-\epsilon(t))\omega(x,v), \end{aligned} \quad (10)$$

where the function $\epsilon(t)$ will be specified later.

Due to (8), we obtain for any constants c_1, c_2, c_3, and c_4,

$$\begin{aligned} &Q^{\#}(c_1 l_0 + c_2 u_0, c_3 l_0 + c_4 u_0) \\ &\quad = (c_1 l_0 + c_2 u_0)^{\#} R^{\#}(c_3 l_0 + c_4 u_0). \end{aligned} \quad (11)$$

Applying (11) to the iteration scheme we obtain

$$\begin{aligned} l_1^{\#}(t) &= \varphi + \int_0^t [l_0^{\#}(s) R^{\#}(l_0)(s) \\ &\qquad - l_1^{\#}(s) R^{\#}(u_0)(s)] ds, \\ u_1^{\#}(t) &= \varphi + \int_0^t [u_0^{\#}(s) R^{\#}(u_0)(s) \\ &\qquad - u_1^{\#}(s) R^{\#}(l_0)(s)] ds, \end{aligned} \quad (12)$$

so that $l_1 - l_0$ and $u_0 - u_1$ satisfy the following equations:

$$\begin{aligned} &(l_1 - l_0)^{\#}(t) \\ &\quad = \varphi - l_0^{\#}(t) - \int_0^t l_0^{\#}(s) R^{\#}(u_0 - l_0)(s) ds \\ &\qquad - \int_0^t (l_1 - l_0)^{\#}(s) R^{\#}(u_0)(s) ds, \\ &(u_0 - u_1)^{\#}(t) \\ &\quad = u_0^{\#}(t) - \varphi - \int_0^t u_0^{\#}(s) R^{\#}(u_0 - l_0)(s) ds \\ &\qquad - \int_0^t (u_0 - u_1)^{\#}(s) R^{\#}(l_0)(s) ds. \end{aligned} \quad (13)$$

In order to prove that Eqs. (13) possess positive solutions it is sufficient to show that the following inequalities hold:

$$u_0^{\#}(t) - \varphi - \int_0^t u_0^{\#}(s) R^{\#}(u_0 - l_0)(s) ds \geqslant 0 \quad (14)$$

and

$$\varphi - l_0^{\#}(t) - \int_0^t l_0^{\#}(s) R^{\#}(u_0 - l_0)(s) ds \geqslant 0. \quad (15)$$

Recalling the definitions of $u_0^{\#}$ and $l_0^{\#}$ and (9) we shall verify (14) and (15) if $\epsilon(t)$ satisfies

$$\epsilon + d\int_0^t \epsilon(s)(1+\epsilon(s))F(x,v,s) ds \leqslant \epsilon(t), \quad (16)$$

where d is a given constant, $\epsilon(t) \leqslant 1$ and

$$\begin{aligned} F(x,v,t) = &\int_D \beta_s(\theta) q^{(s-4)/s} \exp(-\alpha|x - qs|^2) \\ &\times \exp(-\beta v_1^2) dv_1\, d\theta\, d\epsilon. \end{aligned} \quad (17)$$

To prove that inequality (16) has solutions we first have to analyze the function $F(x,v,t)$.

Lemma 1: Let $\beta_s(\theta)$ satisfy the "cutoff" hypothesis

$$\int_0^{\pi/2} \beta_s(\theta) d\theta \leqslant c. \quad (18)$$

Then for hard interactions $4 < s < \infty$ the following estimate holds:

$$\sup_{x,v \in R^n} \int_{t_0}^{t_1} F(x,v,r) dr \leqslant c_P \left(\int_{t_0}^{t_1} (\alpha r^2 + \beta)^{(2-s)/4} dr \right)^{4/s} \quad (19)$$

for every $t_0, t_1 \geqslant 0$ and c_P independent of s.

Proof: First, let us observe that

$$\begin{aligned} &-\alpha|x - qt|^2 - \beta v_1^2 \\ &\quad = -[(\alpha t^2 + \beta) q^2 - 2(\alpha x t - \beta v)\cdot q + \alpha x^2 + \beta v^2] \\ &\quad = -[(\alpha t^2 + \beta)^{1/2} q - (\alpha x t - \beta v)/(\alpha t^2 + \beta)^{1/2}]^2 \\ &\qquad - \alpha\beta(x + vt)^2/(\alpha t^2 + \beta). \end{aligned}$$

Using this equality and estimate (18) we get

$$\begin{aligned} &F(x,v,t) \\ &\quad = \int_D \beta_s(\theta) q^{(s-4)/s} \\ &\qquad \times \exp(-\alpha|x - qs|^2)\exp(-\beta v_1^2) dv_1\, d\theta\, d\epsilon \\ &\quad \leqslant c(\alpha t^2 + \beta)^{(2-2s)/s} \frac{\exp - \alpha\beta(x+vt)^2}{(\alpha t^2 + \beta)} \\ &\qquad \times \int_{R^3} q^{(s-4)/s} \exp - \left[\frac{q - (\alpha x t - \beta v)}{(\alpha t^2 + \beta)^{1/2}} \right]^2 dq. \end{aligned}$$

Because $(s-4)/s < 1$, we can use the triangle inequality:

$$(|a| + |b|)^{(s-4)/s} \leqslant |a|^{(s-4)/s} + |b|^{(s-4)/s}$$

to estimate the last integral, obtaining

$$\int_{R^3} q^{(s-4)/s} \exp - \left[\frac{q-(\alpha xt - \beta v)^2}{(\alpha t^2+\beta)^{1/2}} \right] dq$$

$$\leqslant \int_{R^3} \left\{ \left| \frac{q-(\alpha xt-\beta v)}{(\alpha t^2+\beta)^{1/2}} \right|^{(s-4)/s} + \left| \frac{(\alpha xt-\beta v)}{(\alpha t^2+\beta)^{1/2}} \right|^{(s-4)/s} \right\} \times \exp - \left[\frac{q-(\alpha xt-\beta v)}{(\alpha t^2+\beta)^{1/2}} \right]^2 dq$$

$$= \int_{R^3} \left| \frac{(\alpha xt-\beta v)}{(\alpha t^2+\beta)^{1/2}} \right|^{(s-4)/s} \exp(-q^2) dq + \int_{R^3} q^{(s-4)/s} \exp(-q^2) dq$$

$$\leqslant c \left(1 + \left| \frac{(\alpha xt-\beta v)}{(\alpha t^2+\beta)^{1/2}} \right|^{(s-4)/s} \right).$$

Hence we obtain

$$F(x,v,t) \leqslant c(1+|(\alpha xt-\beta v)/(\alpha t^2+\beta)^{1/2}|^{(s-4)/s}) \times (\alpha t^2+\beta)^{(2-2s)/s} \times \exp\{-\alpha\beta(x+vt)^2/(\alpha t^2+\beta)\} = P(x,v,t)(\alpha t^2+\beta)^{-(s+2)/s}, \tag{20}$$

where

$$P(x,v,t) = c\{(\alpha t^2+\beta)^{(4-s)/s} + |(\alpha xt-\beta v)/(\alpha t^2+\beta)^{3/2}|^{(s-4)/s})\} \times \exp\{-\alpha\beta(x+vt)^2/(\alpha t^2+\beta)\}.$$

Using inequality (20) we get

$$\int_{t_0}^{t_1} F(x,v,r)dr \leqslant \left\{ \int_{t_0}^{t_1} P(x,v,r)^{s/(s-4)} dr \right\}^{(s-4)/s} \times \left\{ \int_{t_0}^{t_1} (\alpha r^2+\beta)^{-(s+2)/4} dr \right\}^{4/s}.$$

To end the proof we have to show that the integral of $P(x,v,t)$ is bounded independently of s. Indeed

$$\left\{ \int_{t_0}^{t_1} P(x,v,r)^{s/(s-4)} dr \right\}^{(s-4)/s} \leqslant c \left\{ \int_{t_0}^{t_1} (\alpha r^2+\beta)^{-1} dr \right\}^{(s-4)/s} + c \left\{ \int_{t_0}^{t_1} \left| \frac{(\alpha xr-\beta v)}{(\alpha r^2+\beta)^{3/2}} \right| \times \exp\left[\frac{-\alpha\beta s(x+vr)^2}{(s-4)(\alpha r^2+\beta)} \right] dr \right\}^{(s-4)/s}.$$

The first term in the right-hand side is bounded. To prove that also the second is bounded we make the following estimate:

$$|(\alpha xr-\beta v)/(\alpha r^2+\beta)^{3/2}| \times \exp[-\alpha\beta s(x+vr)^2/(s-4)(\alpha r^2+\beta)]$$
$$\leqslant (\alpha|x|r+\beta|v|)/(\alpha r^2+\beta)^{3/2} \times \exp[-\alpha\beta s(|x|-|v|r)^2/(s-4)(\alpha r^2+\beta)]$$
$$= \frac{d}{dr}\{(|x|-|v|r)/(\alpha r^2+\beta)\} \times \exp[-\alpha\beta s(|x|-|v|r)^2/(s-4)(\alpha r^2+\beta)].$$

With this inequality we obtain:

$$\int_{t_0}^{t_1} \left| \frac{(\alpha xr-\beta v)}{(\alpha r^2+\beta)^{3/2}} \right| \exp\left[\frac{-\alpha\beta s(x+vr)^2}{(s-4)(\alpha r^2+\beta)} \right] dr \leqslant \int_{-\infty}^{\infty} \exp\left[\frac{-\alpha\beta s r^2}{(s-4)} \right] dr \leqslant c,$$

which shows that also the second term is bounded.

Using the above lemma we can prove the following theorem.

Theorem 1: Let $4 < s < \infty$. There exists $\epsilon(t)$ which satisfies inequality (16). Then l_0 and u_0 defined by (10) are the beginning condition for the system of equations (6) and the limits of sequences l_n and u_n coincide, i.e.,

$$\|u-l\|(t) = 0 \tag{21}$$

for every $t \geqslant 0$.

Here, as in the rest of the paper $\|\cdot\|$ denotes the L^1 norm.

Proof: Let us consider the following auxiliary problem:

$$w_n(t) = 2\epsilon + M\int_0^t K_n(t,u) w_n(u) du, \tag{22}$$

where

$$K_n(t,u) = -\frac{\partial}{\partial u}\left[\int_u^t (\alpha r^2+\beta)^{-(s+2)/4} dr + n^{-s/4} \right]^{4/s}.$$

The integral kernel K_n is bounded and continuous. Thus Eq. (22) has a bounded, differentiable solution. Differentiating this equation we obtain

$$w_n'(t) = M\int_0^t -\frac{\partial^2}{\partial u\,\partial t}\left[\int_u^t (\alpha r^2+\beta)^{-(s+2)/4} dr + n^{-s/4} \right]^{4/s} w_n(u) du + \frac{(\alpha t^2+\beta)^{-(s+2)/4}}{n^{(4-s)/4} s/4} M w_n(t).$$

Then integrating by parts we obtain

$$w_n'(t) = 2\epsilon M \frac{d}{dt}\left[\int_0^t (\alpha r^2+\beta)^{-(s+2)/4} dr + n^{-s/4} \right]^{4/s} + M\int_0^t \frac{\partial}{\partial t}\left[\int_u^t (\alpha r^2+\beta)^{-(s+2)/4} dr + n^{-s/4} \right]^{4/s} w_n'(u) du. \tag{23}$$

Equation (23) implies that $w_n'(t) \geqslant 0$, thus $w_n(t)$ is an increasing function of t. Hence tending with t to infinity we obtain from (22),

$$w_n(t) \leqslant 2\epsilon M \int_0^\infty K_n(u) w_n(u) du. \tag{24}$$

Now, if $\gamma_n(t,s)$ is the resolvent kernel associated to $K_n(t,s)$, $[\gamma(t,s)$ is the resolvent kernel associated to $K(t,s)]$, the method of successive approximations leads to the following bound for the solution:

$$w_n(t) \leqslant w_n(\infty) = 2\epsilon\left(1 + \int_0^\infty \gamma_n(\infty,s)ds\right)$$
$$\leqslant 2\epsilon\left(1 + \int_0^\infty \gamma(\infty,s)ds\right). \quad (25)$$

On the other hand, Corollary 1 of Ref. 8, p. 53, ensures that $\int_0^\infty \gamma(\infty,s)ds$ is bounded. Hence, choosing ϵ sufficiently small we can make $w_n(t) < 1$.

Integrating (22) by parts we obtain

$$w_n(t) = 2\epsilon + 2\epsilon M\left[\int_0^t (\alpha r^2 + \beta)^{-(s+2)/4} dr + n^{-s/4}\right]^{4/s} - \frac{M}{nw_n} + M\int_0^t \left[\int_u^t (\alpha r^2 + \beta)^{-(s+2)/4} dr + n^{-s/4}\right]^{4/2} w_n'(u)du.$$

Since w_n' is positive we can drop the term $n^{-s/4}$ to obtain

$$w_n(1 + M/n) \geqslant 2\epsilon + 2\epsilon M\left[\int_0^t (\alpha r^2 + \beta)^{-(s+2)/4} dr\right]^{4/s} + M\int_0^t \left[\int_u^t (\alpha r^2 + \beta)^{-(s+2)/4} dr\right]^{4/s} w_n'(u)du.$$

Then, due to Lemma 1, we have

$$w_n(1 + M/n) \geqslant 2\epsilon + \frac{2\epsilon M}{c_p}\int_0^t F(x,v,r)dr + \frac{M}{c_p}\int_0^t \left[\int_u^t F(x,v,r)dr\right] w_n'(u)du$$
$$\geqslant 2\epsilon + \frac{M}{c_p}\int_0^t F(x,v,u)w_n(u)du$$
$$\geqslant 2\epsilon + \frac{M}{2c_p}\int_0^t F(x,v,u) \times w_n(u)(1 + w_n(u))du.$$

Choosing M such that $M/4c_p > d$ [cf. (16)] and $n \geqslant M$ we obtain:

$$w_n(t) \geqslant \epsilon + d\int_0^t F(x,v,u)w_n(u) \times (1 + w_n(u))du.$$

Therefore, if $n > M = 4c_p d$, $w_n(t)$ satisfies inequality (16).

From Eqs. (6) it follows that $\|l\|(t)$ and $\|u\|(t)$ are bounded. Then subtracting (6b) from (6a) and tending with n to infinity we get the inequality

$$(u - l)^\#(t) \leqslant c\omega(x,v)\int_0^t \|u - l\|(r) \times \int_D \beta_S(\theta)\omega(x - qr,v_1)dv_1\, d\theta\, d\epsilon\, dr. \quad (26)$$

Applying to (26) the Holder inequality and using (18) we obtain

$$\|u - l\|(t) \leqslant c\int_0^t \|u - l\|(r)F(x,v,r)dr$$
$$\leqslant c\left[\int_0^t (\|u - l\|(r))^{s/4} dr\right]^{4/s} \times \left[\int_0^t F(x,v,r)^{s/(s-4)} dr\right]^{(s-4)/s}$$
$$\leqslant c\left[\int_0^t (\|u - l\|(r))^{s/4} dr\right]^{4/s}.$$

Hence by the Gronwall lemma we get (21). ∎

IV. EXISTENCE OF SOLUTIONS FOR RIGID SPHERES

The analysis made in the preceding section does not take into account the rigid spheres model. In fact, for $s = \infty$ inequality (16) does not hold for any $\epsilon(t)$. On the other hand, the smallness condition (25) holds also for $s = \infty$. This enables us to hope that the result which is true for $s < \infty$ can be extended to rigid spheres.

To solve the initial value problem (1) for rigid spheres let us take initial data in the form

$$(l - \epsilon_0)\omega(x,v) \leqslant \varphi(x,v) \leqslant (1 + \epsilon_0)\omega(x,v), \quad (27)$$

where ϵ_0 is given by the following expression:

$$\epsilon_0^{-1} = 2\sup_s \exp\left\{M\left[\int_0^\infty (\alpha r^2 + \beta)^{-(s+2)/4} dr + n^{-s/4}\right]^{4/s}\right\}.$$

Let us now introduce a sequence of collision kernels which approximate the rigid sphere kernel:

$$B_n(q,\theta) = A\cos\theta\sin\theta q^{1-\exp(-n^2)} \quad (28)$$

and let $f_n(x,v,t)$ be a solution of the Boltzmann equation (1) with collision kernel (28) and initial data (27). Then the following lemma holds.

Lemma 2: The sequence $f_n(x,v,t)$ is a Cauchy sequence in $L^1(R^6)$ for every $t \geqslant 0$.

Proof: Let us denote by Q_n and R_n the collision operators whose kernels are given by (28). Then for given $n,m \geqslant n_0$ with $m > n$, we have

$$f_n^\#(t) = \varphi + \int_0^t Q_n^\#(f_n,f_n)(s)ds - \int_0^t f_n^\#(s)\, R_n^\#(f_n)(s)ds,$$

$$f_m^\#(t) = \varphi + \int_0^t Q_m^\#(f_m,f_m)(s)ds - \int_0^t f_m^\#(s)\, R_m^\#(f_m)(s)ds.$$

Subtracting these two equations we obtain

$$\int_{R^6} |f_m - f_n|^{\#}(t) dx\, dv \leqslant 2\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_m,f_m) - Q_n^{\#}(f_n,f_n)](s) ds \right| dx\, dv$$

$$\leqslant 2\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_m,f_m) - Q_m^{\#}(f_n,f_n)](s) ds \right| dx\, dv + 2\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_n,f_n) - Q_n^{\#}(f_n,f_n)](s) ds \right| dx\, dv. \quad (29)$$

Consider the first of the above two integrals:

$$\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_m,f_m) - Q_m^{\#}(f_n,f_n)](s) ds \right| dx\, dv \leqslant c \int_{R^6} dx\, dv \int_0^t ds \left| f_m^{\#} - f_n^{\#} \right|(s) \int_{R^3} dv_1 q^{1-\exp(-m^2)} \omega(x - qs, v_1).$$

We split the last integral into two parts corresponding to $|v| > n$ and $|v| \leqslant n$. Then we obtain

$$\int_{R^3} dx \int_{|v|>n} dv \int_0^t ds |f_m^{\#} - f_n^{\#}|(s) \int_{R^3} dv_1 q^{1-\exp(-m^2)} \omega(x - qs, v_1)$$

$$\leqslant c \int_{R^3} dx \int_{|v|>n} dv\, \omega(x,v) \int_{R^3} dv_1\, q^{-\exp(-m^2)} \exp(-\beta v_1^2) \leqslant c \int_{|v|>n} dv \exp(-\beta v^2) \leqslant c \exp(-\beta n^2/2).$$

For the part with $|v| \leqslant n$ we have

$$\int_{R^3} dx \int_{|v|\leqslant n} dv \int_0^t ds |f_m^{\#} - f_n^{\#}|(s) \int_{R^3} dv_1 q^{1-\exp(-m^2)} \omega(x - qs, v_1) \leqslant c(1+n) \int_0^t ds \int_{R^6} dx\, dv |f_m^{\#} - f_n^{\#}|(s).$$

To estimate the second integral in (29) we utilize the fact that $f_n^{\#}(x,v,t) \leqslant 2\omega(x,v)$. Then we proceed as follows:

$$\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_n,f_n) - Q_n^{\#}(f_n,f_n)](s) ds \right| dx\, dv$$

$$\leqslant c \int_{R^6} dx\, dv \int_0^t ds \int_{R^3} dv_1 |q^{1-\exp(-m^2)} - q^{1-\exp(-n^2)}| \omega(x - qs, v_1) \omega(x,v)$$

$$\leqslant c \int_{R^6} dx\, dv \exp\left(\frac{-\alpha x^2 - \beta v^2}{2}\right) \int_0^t ds \exp(-\alpha |x - qs|^2) q \int_{R^3} dq |q^{-\exp(-m^2)} - q^{-\exp(-n^2)}| \exp\left(\frac{-\beta q^2}{4}\right)$$

$$\leqslant c \int_{R^3} dq |q^{-\exp(-m^2)} - q^{-\exp(-n^2)}| \exp\left(\frac{-\beta q^2}{4}\right).$$

In the above inequality we used the fact that $\beta v_1^2 + \beta v^2/2 \geqslant \beta q^2/4$. For the last integral we have

$$\int_{R^3} dq |q^{-\exp(-m^2)} - q^{-\exp(-n^2)}| \exp\left(\frac{-\beta q^2}{4}\right) = 4\pi \int_0^\infty dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \exp\left(\frac{-\beta q^2}{4}\right)$$

$$= 4\pi \int_0^1 dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \times \exp\left(\frac{-\beta q^2}{4}\right) + 4\pi \int_1^\infty dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \exp\left(\frac{-\beta q^2}{4}\right).$$

The first integral can be evaluated straightforwardly,

$$\int_0^1 dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \exp\left(\frac{-\beta q^2}{4}\right) \leqslant \int_0^1 dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \leqslant c\{\exp(-n^2) - \exp(-m^2)\}.$$

To estimate the second integral let us observe that for $q \geqslant 1$ we have $q^{p-1} \leqslant pq^{p+1}, p > 0$. Then

$$\int_1^\infty dq (q^{2-\exp(-m^2)} - q^{2-\exp(-n^2)}) \exp\left(\frac{-\beta q^2}{4}\right) = \int_1^\infty dq\, q^{2-\exp(-n^2)} \times (q^{\exp(-n^2)-\exp(-m^2)} - 1) \exp\left(\frac{-\beta q^2}{4}\right)$$

$$\leqslant [\exp(-n^2) - \exp(-m^2)] \int_1^\infty dq\, q^3 \exp\left(\frac{-\beta q^2}{4}\right).$$

Summarizing all the above estimates we obtain

$$\int_{R^6} \left| \int_0^t [Q_m^{\#}(f_n,f_n) - Q_n^{\#}(f_n,f_n)](s) ds \right| dx\, dv \leqslant c \times [\exp(-n^2) - \exp(-m^2)].$$

Hence from (29) the following estimate can be derived:

$$\|f_m - f_n\|(t) \leqslant c \exp(-an^2) + c(1+n) \int_0^t \|f_m - f_n\|(s) ds.$$

Applying the Gronwall lemma we have

$$\|f_m - f_n\|(t) \leqslant c \exp(-an^2)\exp[c(1+n)t].$$

Hence tending with n_0 to infinity we obtain the assertion of the lemma.

Now we can prove our main result.

Theorem 2: Let $f(x,v,t)$ be the limit of the sequence $f_n(x,v,t)$ which exists due to Lemma 2. Then $f(x,v,t)$ is a unique solution of the Boltzmann equation for rigid spheres.

Proof: First, we will show that $f(x,v,t)$ solves the Boltzmann equation in the sense of L^1. To this end let us observe that since $f_n(x,v,t) \leqslant 2\omega(x,v)$, then also $f(x,v,t) \leqslant 2\omega(x,v)$. Hence we have

$$\begin{aligned}
&\left\| f^{\#}(t) - \varphi - \int_0^t J^{\#}(f,f)(s)ds \right\| \\
&\leqslant \left\| f^{\#}(t) - \varphi - \int_0^t J^{\#}(f,f)(s)ds - \left(f_n^{\#}(t) - \varphi - \int_0^t J_n^{\#}(f_n,f_n)(s)ds \right) \right\| \\
&\leqslant \| f^{\#}(t) - f_n^{\#}(t) \| + \left\| \int_0^t J^{\#}(f,f)(s)ds - \int_0^t J_n^{\#}(f_n,f_n)(s)ds \right\|.
\end{aligned}$$

The first term tends to zero by definition. To prove that the second term converges to zero we have to apply similar estimates as in Lemma 2, but with $m = \infty$.

To show that $f(x,v,t)$ is a unique solution of the Boltzmann equation for rigid spheres let us assume that there exists another solution $g(x,v,t)$ for which the estimate $g(x,v,t) \leqslant c\omega(x,v)$ holds. Then subtracting $g(x,v,t)$ from $f(x,v,t)$ we obtain

$$(f-g)^{\#}(t) \leqslant c\omega(x,t) \int_0^t \|f-g\|(r) \times \int_D \beta_s(\theta)\omega(x-qr,v_1) \times dv_1\, d\theta\, d\epsilon\, dr.$$

Applying the Holder inequality we obtain:

$$\begin{aligned}
\|f-g\|(t) &\leqslant c\int_0^t \|f-g\|(r) F(x,v,r)dr \\
&\leqslant c\left[\int_0^t (\|f-g\|(r))^{s/4}\, dr\right]^{4/s} \times \left[\int_0^t F(x,v,r)^{s/(s-4)}\, dr\right]^{(s-4)/s} \\
&\leqslant c\left[\int_0^t (\|f-g\|(r))^{s/4}\, dr\right]^{4/s}.
\end{aligned}$$

Hence by the Gronwall lemma we obtain

$$\|f-g\|(t) = 0.$$

ACKNOWLEDGMENTS

One of the authors (AP) would like to acknowledge CNR for sponsoring his stay in Italy and the hospitality of the Universities of Ferrara and Pavia where this work has been completed.

This research has been partially supported by CNR—Gruppo Nazionale per la Fisica Matematica. AP has also been partially supported by the Ministry of National Education of Poland under Grant No. C.P.B.P. 01.02 and GT has been partially supported by the Italian Ministry of Education (M.P.I.).

[1]R. DiPerna and P. L. Lions, "Solutions globales de l'equation de Boltzmann," C. R. Acad. Sci. Paris, Ser. I **306**, 343 (1988).

[2]R. J. DiPerna and P. L. Lions, "On the Cauchy problem for the Boltzmann equation: Global existence and weak stability," preprint, 1988.

[3]L. Arkeryd, "On the long time behavior of the Boltzmann equation in a periodic box," preprint No. 23, Chalmers University of Technology, Göteborg, 1988.

[4]G. Toscani, "Global solution of the initial value problem for the Boltzmann equation near a local Maxwellian," Arch. Rat. Mech. Anal. **102**, 231 (1988).

[5]R. Illner and M. Shinbrot, "Global existence for a rare gas in an infinite vacuum," Commun. Math. Phys. **95**, 217 (1984).

[6]S. Kaniel and M. Shinbrot, "The Boltzmann equation 1: Uniqueness and local existence," Commun. Math. Phys. **58**, 65 (1978).

[7]C. Truesdell and R. G. Muncaster, *Fundamentals of Maxwell's Kinetic Theory of a Simple Monoatomic Gas* (Academic, New York, 1980).

[8]C. Cordoneanu, *Integral Equations and Stability of Feedback Systems* (Academic, New York, 1973).

ERRATUM

Erratum: The hyperspin structure of unitary groups [J. Math. Phys. 29, 978 (1988)]

Christian Holm
Institut für Theoretische Physik A, TU Clausthal, 3392 Clausthal, Federal Republic of Germany

(Received 4 April 1989; accepted for publication 19 April 1989)

Since the publication of the above article subsequent research,[1] motivated by a comment of Borowiec,[2] has shown that the hyper-Christoffel connection [p. 979, (2.9)] does not transform like a connection, and even worse, does not exist for most geometries. Subsequently formula (2.11) should be ignored.

From the metricity condition (1) [p. 984, (4.2)] only follows that $w^a{}_a = 0$, it is not equivalent to (1). This in turn invalidates the lemma of Eq. (4.5), as well as Eqs. (4.8) and (4.11). The existence question of a torsion-free metric connection is dealt with in Ref. 1. There it is shown that a general Bergmann manifold does not possess such a connection.

Nevertheless such a connection does exist for the geometry of the unitary groups, so that the main results of the paper remain valid.

[1]C. Holm, "On the connection in Bergmann manifolds," to be published in Int. J. Theor. Phys.

[2]A. Borowiec, "Some comment on geometry of hyperspin manifold," Wroclav preprint, 1988.

