Meixner–Pollaczek polynomials and the Heisenberg algebra

Tom H. Koornwinder
Centre for Mathematics and Computer Science, P. O. Box 4079, 1009 AB Amsterdam, The Netherlands

(Received 12 July 1988; accepted for publication 9 November 1988)

An alternative proof is given for the connection between a system of continuous Hahn polynomials and identities for symmetric elements in the Heisenberg algebra, which was first observed by Bender, Mead, and Pinsky [Phys. Rev. Lett. **56**, 2445 (1986); J. Math. Phys. **28**, 509 (1987)]. The continuous Hahn polynomials turn out to be Meixner–Pollaczek polynomials. Use is made of the connection between Laguerre polynomials and Meixner–Pollaczek polynomials, the Rodrigues formula for Laguerre polynomials, an operational formula involving Meixner–Pollaczek polynomials, and the Schrödinger model for the irreducible unitary representations of the three-dimensional Heisenberg group.

I. INTRODUCTION

In two recent papers[1,2] Bender, Mead, and Pinsky discussed the connection between certain continuous Hahn polynomials and symmetrizations of elements in the Heisenberg algebra. They showed that, if

$$[q,p]=i$$

and $T_{m,n}$ is the sum of all possible terms containing m factors of p and n factors of q, then

$$T_{n,n}=\mathrm{const}\, S_n(T_{1,1}), \tag{1.1}$$

for some polynomial S_n of degree n, which turns out to be the orthogonal polynomial of degree n on $\mathbb{R}$ with respect to the weight function $x \mapsto 1/\mathrm{ch}(\pi x/2)$. However, the actual proof of this result is not very clear from these two papers.

In the present paper we give an alternative proof of (1.1). First, in Sec. II, we observe a transformation connecting certain continuous Hahn polynomials, in particular, the above polynomials S_n to certain Meixner–Pollaczek polynomials. Next, in Sec. III we use a Mellin transform relating Laguerre polynomials and Meixner–Pollaczek polynomials and the Rodrigues formula for Laguerre polynomials in order to derive an operational formula involving Meixner–Pollaczek polynomials. Finally, in Sec. IV we use this operational formula in order to derive formula (1.1). Here we make use of the Schrödinger model for the irreducible unitary representations of the Heisenberg group.

II. ON CONTINUOUS HAHN POLYNOMIALS EXPRESSIBLE AS MEIXNER–POLLACZEK POLYNOMIALS

Continuous Hahn polynomials are defined by

$$p_n(x;a,b,c,d):=i^n\frac{(a+c)_n(a+d)_n}{n}\times {}_3F_2\left(\begin{matrix}-n,n+a+b+c+d-1,a+ix\\ a+c,a+d\end{matrix};1\right). \tag{2.1}$$

If $c=\bar a$, $d=\bar b$ and $\mathrm{Re}\, a$, $\mathrm{Re}\, b>0$, then they are orthogonal on $(-\infty,\infty)$ with respect to the weight function

$$w(x):=\Gamma(a+ix)\Gamma(b+ix)\Gamma(c-ix)\Gamma(d-ix). \tag{2.2}$$

See Refs. 3 and 4, but read $a+ix$ instead of $a-ix$ in formula (3) of Ref. 4.

Meixner–Pollaczek polynomials are defined by

$$P_n^{(a)}(x;\phi):=e^{in\phi}\,{}_2F_1(-n,a+ix;2a;1-e^{-2i\phi}). \tag{2.3}$$

If $a>0$ and $0<\phi<\pi$, they are orthogonal on $(-\infty,\infty)$ with respect to the weight function

$$w(x)=e^{(2\phi-\pi)x}|\Gamma(a+ix)|^2. \tag{2.4}$$

See Refs. 5 and 6 and, for standardized notation, the Appendix of Ref. 7.

For $a=c=b-\frac12=d-\frac12>0$ the weight function (2.2) becomes

$$w(x)=2^{-4a+2}\pi|\Gamma(2a+2ix)|^2. \tag{2.5}$$

On comparing with (2.4) we conclude that

$$p_n(x;a,a+\tfrac12,a,a+\tfrac12)=\mathrm{const}\,P_n^{(2a)}(2x;\tfrac12\pi).$$

The constant can be computed by comparing coefficients of x^n. We obtain

$$p_n(x;a,a+\tfrac12,a,a+\tfrac12)=[(2a)_n(2a+\tfrac12)_n/n!]\times P_n^{(2a)}(2x;\tfrac12\pi). \tag{2.6}$$

In terms of hypergeometric functions this formula reads

$${}_3F_2\left(\begin{matrix}-n,n+4a,a+ix\\ 2a,2a+\frac12\end{matrix}\middle|\,1\right)={}_2F_1(-n,2a+2ix;4a;2). \tag{2.7}$$

This identity can also be obtained from Ref. 8,

$${}_4F_3\left(\begin{matrix}a,b,n+2c,-n\\ a+b+\frac12,c,c+\frac12\end{matrix};1\right)={}_3F_2\left(\begin{matrix}2a,2b,-n\\ a+b+\frac12,2c\end{matrix};1\right), \tag{2.8}$$

by letting $b\to\infty$.

For $a:=\frac14$ the weight function (2.5) becomes

$$w(x)=2\pi^2/\mathrm{ch}(2\pi x).$$

In particular, we find for the polynomials S_n introduced in Sec. I, which were identified with special continuous Hahn polynomials in Ref. 2, that they can be written as Meixner–Pollaczek polynomials:

$$S_n(x)=\mathrm{const}\,P_n^{(1/2)}(\tfrac12 x,\tfrac12\pi). \tag{2.9}$$

 0022-2488/89/040767-03$02.50

III. AN OPERATIONAL FORMULA INVOLVING MEIXNER–POLLACZEK POLYNOMIALS

Recall that we can obtain the *Mellin transform pair*,

$$
\begin{aligned}
G(\lambda) &= \int_0^\infty F(\tau)\tau^{-1-i\lambda}\,d\tau,\\
F(\tau) &= (2\pi)^{-1}\int_{-\infty}^\infty G(\lambda)\tau^{i\lambda}\,d\lambda,
\end{aligned}
\tag{3.1}
$$

from the *Fourier transform pair*,

$$
\begin{aligned}
g(\lambda) &= \int_{-\infty}^\infty f(t)e^{-2\pi i\lambda t}\,dt,\\
f(t) &= \int_{-\infty}^\infty g(\lambda)e^{2\pi\lambda t}\,d\lambda,
\end{aligned}
\tag{3.2}
$$

by making the substitutions

$$\tau = e^{2\pi t},\quad F(\tau)=f(t),\quad G(\lambda)=2\pi g(\lambda)$$

in (3.2). In particular, Mellin inversion in (3.1) is valid if the function $t\mapsto F(e^{2\pi t})$ belongs to the class $\mathscr{S}$ of rapidly decreasing C^∞ function on $\mathbb{R}$. If F_1, F_2 are two such functions and G_1, G_2 their Mellin transforms then we have the *Parseval formula*

$$\int_0^\infty F_1(\tau)\,\overline{F_2(\tau)}\,\frac{d\tau}{\tau} = \int_{-\infty}^\infty G_1(\lambda)\,\overline{G_2(\lambda)}\,\frac{d\lambda}{2\pi}. \tag{3.3}$$

Proposition 3.1: For $a>0$ and $0<\phi<\pi$ Laguerre polynomials $x\mapsto L_n^{2a-1}(x)$ and Meixner–Pollaczek polynomials $\lambda\mapsto P_n^{(a)}(\lambda;\phi)$ are mapped onto each other by the Mellin transform in the following way:

$$
\begin{aligned}
&\int_0^\infty \frac{n!e^{-in\phi}}{(2a)_n}e^{-(1/2)x(1+i\cot\phi)}x^a L_n^{2a-1}(x)x^{-1-i\lambda}\,dx\\
&\quad = e^{(ia-\lambda)[\phi-(1/2)\pi]}(2\sin\phi)^{a-i\lambda}\Gamma(a-i\lambda)P_n^{(a)}(\lambda;\phi).
\end{aligned}
\tag{3.4}
$$

Proof: The left-hand side can be rewritten as

$$
\begin{aligned}
&e^{-in\phi}\sum_{k=0}^n \frac{(-n)_k}{(2a)_k k!}\int_0^\infty e^{-(1/2)x(1+i\cot\phi)}x^{k+a-i\lambda-1}\,dx\\
&\quad = e^{-in\phi}\sum_{k=0}^n \frac{(-n)_k}{(2a)_k k!}\,\frac{\Gamma(a-i\lambda+k)}{(\frac12+\frac12 i\cot\phi)^{a-i\lambda+k}}\\
&\quad = e^{-in\phi}\Gamma(a-i\lambda)(1-e^{2i\phi})^{a-i\lambda}\\
&\qquad\times {}_2F_1(-n,a-i\lambda;2a;1-e^{2i\phi})\\
&\quad = e^{in\phi}\Gamma(a-i\lambda)(1-e^{2i\phi})^{a-i\lambda}\\
&\qquad\times {}_2F_1(-n,a+i\lambda;2a;1-e^{-2i\phi}),
\end{aligned}
$$

which can be rewritten as the right-hand side of (3.4). □

It is possible[9,10] to give an interpretation of the above proposition in the context of matrix elements of discrete series representations of SL(2,$\mathbb{R}$).

Corollary 3.2: For $a>0$ and $0<\phi<\pi$ Laguerre polynomials can be expressed by the differentiation formula

$$
\begin{aligned}
&\frac{n!e^{-in\phi}}{(2a)_n}e^{(1/2)x(1+i\cot\phi)}x^a L_n^{2a-1}(x)\\
&\quad = P_n^{(a)}\Big(-ix\frac{d}{dx};\phi\Big)(e^{-(1/2)x(1+i\cot\phi)}x^a).
\end{aligned}
\tag{3.5}
$$

Proof: In the left-hand side of (3.4) Mellin transform is taken of a function that belongs to the class $\mathscr{S}$ as a function of t, where $x=e^t$. Hence we can apply Mellin inversion [cf. (3.1)] and we can write the left-hand side of (3.5) as

$$
\begin{aligned}
&(2\pi)^{-1}\int_{-\infty}^\infty e^{(ia-\lambda)[\phi-(1/2)\pi]}(2\sin\phi)^{a-i\lambda}\\
&\qquad\times\Gamma(a-i\lambda)P_n^{(a)}(\lambda;\phi)x^{i\lambda}\,d\lambda\\
&\quad = P_n^{(a)}\Big(-ix\frac{d}{dx};\phi\Big)\\
&\qquad\times[e^{(ia-\lambda)[\phi-(1/2)\pi]}(2\sin\phi)^{a-i\lambda}\\
&\qquad\times\Gamma(a-i\lambda)x^{i\lambda}],
\end{aligned}
$$

which equals the right-hand side of (3.5). □

By substitution of the Rodrigues formula

$$n!e^{-x}x^aL_n^a(x) = \Big(\frac{d}{dx}\Big)^n(e^{-x}x^{n+a})$$

into (3.5) we obtain

$$
\begin{aligned}
&\Big(\frac{d}{dx}\Big)^n(e^{-x}x^{n+2a-1})\\
&\quad = (2a)_n e^{in\phi}e^{-(1/2)x(1-i\cot g\phi)}x^{a-1}\\
&\qquad\times P_n^{(a)}\Big(-ix\frac{d}{dx},\phi\Big)[e^{-(1/2)x(1+i\cot\phi)}x^a].
\end{aligned}
\tag{3.6}
$$

In particular, for $\phi=\frac12\pi$ and $a=\frac12$ we obtain

$$
\begin{aligned}
&\Big(i\frac{d}{dx}\Big)^n(e^{-x}x^n)\\
&\quad = n!e^{-(1/2)x}P_n^{(1/2)}\Big(ix\frac{d}{dx}+\frac12 i,\frac12\pi\Big)[e^{-(1/2)x}].
\end{aligned}
$$

Hence for arbitrary $\nu\in\mathbb{C}$,

$$
\begin{aligned}
&e^{i\nu x}\Big(i\frac{d}{dx}\Big)^n(x^ne^{-2i\nu x})\\
&\quad = n!P_n^{(1/2)}\Big(ix\frac{d}{dx}+\frac12 i,\frac12\pi\Big)[e^{-i\nu x}].
\end{aligned}
\tag{3.7}
$$

IV. PROOF OF THE BENDER–MEAD–PINSKY RESULT

Consider the *Heisenberg group* H_1, which is $\mathbb{R}^3$ equipped with the multiplication rule

$$
\begin{aligned}
&(\xi,\eta,\tau)(\xi',\eta',\tau')\\
&\quad = (\xi+\xi',\eta+\eta',\tau+\tau'+\tfrac12(\xi'\eta-\xi\eta')).
\end{aligned}
\tag{4.1}
$$

Let $\lambda\in\mathbb{R}\setminus\{0\}$ and let π_λ denote the unique (up to equivalence) irreducible unitary representation of H_1 such that

$$\pi_\lambda(0,0,\tau) = e^{i\lambda\tau}I,\quad \tau\in\mathbb{R}.$$

Then, with $\mu:=|\lambda|^{1/2}$ and $\epsilon:=\operatorname{sgn}(\lambda)$, π_λ can be realized on $L^2(\mathbb{R})$ by

$$
\begin{aligned}
&(\pi_\lambda(\xi,\eta,\tau)f)(x)\\
&\quad = e^{i\mu\xi x}e^{i\mu^2[\epsilon\tau+(\frac12)\xi\eta]}f(x+\mu\eta),\quad f\in L^2(\mathbb{R}).
\end{aligned}
\tag{4.2}
$$

Let X and Y be the infinitesimal generators of the one-parameter subgroups of elements $(\xi,0,0)$ and $(0,\eta,0)$, respectively. Let σ denote the symmetrization mapping[11] from the symmetric algebra to the universal enveloping algebra of the Lie algebra of H_1, i.e.,

$$\sigma(X_1\cdots X_k) := \frac{1}{k!}\sum_s X_{s(1)}\cdots X_{s(k)}, \tag{4.3}$$

where s runs over all permutations of $\{1,...,k\}$. Let f be a C^∞ function locally defined on $\mathbb{R}$. Then

$$(\pi_\lambda(X)f)(x) = i\mu x f(x),$$
$$(\pi_\lambda(Y)f)(x) = \mu f'(x),$$

and

$$(\pi_\lambda(\sigma(X^nY^n))f)(x)$$
$$= \left(\frac{\partial}{\partial\xi}\right)^n\left(\frac{\partial}{\partial\eta}\right)^n (e^{i\mu\xi x}e^{i\mu^2[\epsilon\tau + (1/2)\xi\eta]}f(x+\mu\eta))\Big|_{\xi,\eta,\tau=0}$$
$$= \left(i\mu\frac{\partial}{\partial\eta}\right)^n\left(\left(x+\frac{1}{2}\mu\eta\right)^n f(x+\mu\eta)\right)\Big|_{\eta=0}.$$

Hence

$$(\pi_\lambda(\sigma(X^nY^n))f)(x)$$
$$= |\lambda|^n\left[i\frac{\partial}{\partial y}\right]^n$$
$$\times\left(\left(x+\frac{1}{2}y\right)^n f(x+y)\right)\Big|_{y=0}. \tag{4.4}$$

For $n = 1$ this simplifies to

$$(\pi_\lambda(\sigma(XY))f)(x) = |\lambda|\left(ix\frac{\partial}{\partial x}+\frac{1}{2}i\right)f(x). \tag{4.5}$$

Let

$$f_v(x) := e^{-ivx}.$$

Then we obtain from (4.4), (3.7), and (4.5) that

$$(\pi_\lambda(\sigma(X^nY^n))f_v)(x)$$
$$= |\lambda|^n\left(i\frac{\partial}{\partial y}\right)^n\left(\left(x+\frac{1}{2}y\right)^n e^{-iv(x+y)}\right)\Big|_{y=0}$$
$$= 2^{-n}|\lambda|^n e^{ivx}\left(i\frac{\partial}{\partial x}\right)^n(x^n e^{-2ivx})$$
$$= 2^{-n}n!|\lambda|^n P_n^{(1/2)}\left(ix\frac{d}{dx}+\frac{1}{2}i,\frac{1}{2}\pi\right)[e^{-ivx}]$$
$$= 2^{-n}n!|\lambda|^n P_n^{(1/2)}(|\lambda|^{-1}\pi_\lambda(\sigma(XY)),\tfrac{1}{2}\pi)[f_v(x)].$$

Hence by integrating both sides against suitable functions of v, we obtain

$$\pi_\lambda(\sigma(X^nY^n)) = 2^{-n}n!|\lambda|^n P_n^{(1/2)}(|\lambda|^{-1}\pi_\lambda(\sigma(XY)),\tfrac{1}{2}\pi). \tag{4.6}$$

In view of (2.9) and (4.3) this becomes for $\lambda = 1$ the result (1.1) of Ref. 1.

ACKNOWLEDGMENT

The author thanks G. Gasper for a reference to (2.8).

[1] C. M. Bender, L. R. Mead, and S. S. Pinsky, Phys. Rev. Lett. **56**, 2445 (1986).
[2] C. M. Bender, L. R. Mead, and S. S. Pinsky, J. Math. Phys. **28**, 509 (1987).
[3] N. M. Atakishiyev and S. K. Suslov, J. Phys. A **18**, 1583 (1985).
[4] R. Askey, J. Phys. A **18**, L1017 (1985).
[5] J. Meixner, J. London Math. Soc. **9**, 6 (1934).
[6] F. Pollaczek, C. R. Acad. Sci. Paris **230**, 1563 (1950).
[7] R. Askey and J. Wilson, Mem. Am. Math. Soc. **54**, 319 (1985).
[8] W. N. Bailey, Proc. London Math. Soc. **29**, 495 (1929).
[9] T. H. Koornwinder, "Group theoretic interpretations of Askey's scheme of hypergeometric orthogonal polynomials," "Orthogonal polynomials and their applications," *Lecture Notes in Mathematics*, Vol. 1329, edited by M. Alfaro *et al.* (Springer, Berlin, 1988), pp. 46–72.
[10] D. Basu and K. B. Wolf, J. Math. Phys. **23**, 189 (1982).
[11] S. Helgason, *Groups and Geometric Analysis* (Academic, New York, 1984).

Heat kernel expansion on a generalized cone

J. S. Dowker
Department of Theoretical Physics, The University of Manchester, Manchester, M139PL England

(Received 4 October 1988; accepted for publication 7 December 1988)

It is shown that the short-time expansion of the integrated heat kernel on a locally flat generalized cone $C(N)$, as defined by Cheeger, consists of just the Weyl volume term and the constant term. This latter is explicitly evaluated when N is a lens space, S^d/Z_m (for odd d), elliptic space S^d/Z_2 (for all d), and any of the three-dimensional, homogeneous space forms S^3/Γ. Agreement is found with the corresponding expansion on the orbifold version, T^4/Z_2, of the $K3$ surface, and, in fact, with all T^D/Z_2.

I. INTRODUCTION

In Ref. 1, we discussed the heat kernel expansion on a polyhedron, considered as a collection of ordinary, two-dimensional, conical singularities. In this article, we wish to analyze a similar situation but with the subsitution of a "generalized cone," $C(N)$. This is defined by Cheeger[2] as the space $R^+\times N$ with the "hyperspherical polar" metric

$$ds^2 = dr^2 + r^2\,d\Sigma^2, \tag{1}$$

where $d\Sigma^2$ is the metric on the manifold N, and r runs from 0 to infinity, the point $r=0$ being generally a singular point—the apex of the cone. If N is the sphere of unit radius S^d, $C(N)$ is just R^{d+1}. For $N=S^1$, of any radius, $C(N)$ is the ordinary cone.

Cheeger[2] discusses precisely the heat kernel question but his analysis is very complete and it was thought that a simpler treatment of some special cases might be useful.

Some "physical" motivation will be found in Sec. VI. As a simple extension, which often proves useful, we could add d' extra Euclidean dimensions $\mathbf{z}$, and write the metric as

$$ds^2 = dr^2 + r^2\,d\Sigma^2 + d\mathbf{z}^2. \tag{2}$$

In this paper we set d' equal to zero.

II. THE HEAT KERNEL

In this work we consider only scalar fields (functions) and write the heat equation as

$$\left[\frac{\partial}{\partial t} - \Delta_C\right]K(\mathbf{x},\mathbf{x}';t) = \delta(t)\delta(\mathbf{x},\mathbf{x}'),$$

where $\mathbf{x},\mathbf{x}'$ are points in $C(N)$. Formally, $K=\exp(t\Delta_C)$, $t>0$.

The metric (1) allows a separation of variables since the Laplacian can be written

$$\Delta_c = \frac{\partial^2}{\partial r^2} + \frac{d}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\Delta_N,$$

where Δ_N is the Laplacian on N.

The eigenfunction form of the heat kernel is then easily manipulated into

$$K(r,q,r',q';t) = (rr')^{(1-d)/2}\frac{1}{2t}e^{-(r^2+r'^2)/4t} \times\sum_n I_{\nu_n}\left(\frac{rr'}{2t}\right)\phi_n^*(q)\phi_n(q'), \tag{3}$$

where the ϕ_n are eigenfunctions of Δ_N,

$$\Delta_N\phi_N = -\lambda_n^2\phi_n.$$

The ν_n and λ_n are related by

$$\nu_n = (\lambda_n^2 + (d-1)^2/4)^{1/2}. \tag{4}$$

We remark that (3) is partly in classical path and partly in eigenfunction form.

III. THE SPHERICAL CASE

In order to take the analysis further, in explicit form, we need a special choice for N. Clearly, one important class of manifolds is that of homogeneous spaces, G/H. In particular, we might consider those of rank 1. However, we leave these aside, in their generality, and simply look at the special case of spheres or rather those manifolds locally isometric to spheres. Thus we choose $N=S^d/\Gamma$, where Γ is a discrete group of isometries of S^d.

Moreover, if the spheres are of unit radius, the cones $C(N)$ will be flat, except at the apex, since they will be locally isometric to R^{d+1}. This is the case we exclusively discuss in this paper.

Conical spaces such as these occur in general relativity.[3] Because of the local flatness, we would expect certain simplifications. Indeed, the eigenvalues on the sphere S^d are well known to be

$$\begin{aligned}\lambda_n &= (n-1)(n+d-2)\\ &= (n+(d-3)/2)^2 - (d-1)^2/4 \quad (n=1,2,\ldots).\end{aligned}$$

Hence, from (4),

$$\nu_n = n + (d-3)/2. \tag{5}$$

It is no surprise that the term $(d-1)^2/4$ is just the quantity ξR [R is the scalar curvature and $\xi=(d-1)/4d$] needed to make the operator $-\Delta_N+\xi R$ conformally covariant. (If d is odd, this leads to a Huygens principle for the wave operator $\partial^2/\partial t^2 - \Delta_N + \xi R$, a fact well known to physicists for a long time and, more recently, to mathematicians.[4])

In order to illustrate the general method, it is sufficient to consider the case when $N=S^3/\Gamma$ so that, technically, we can use the isometry $S^3\sim SU(2)$.

The structure of Γ is well known[5] to be the product $\Gamma_L\times\Gamma_R$ acting on the coordinate q, thought of as an element of SU(2). For homogeneity, one of these components must be the trivial identity group. We choose $\Gamma=\Gamma_L$ and, furthermore, we shall look closely at lens spaces, $\Gamma=Z_m$. This is a case mentioned by Cheeger,[2] but he does not give the explicit formulas.

 0022-2488/89/040770-04$02.50

IV. S^3/Γ CONES

Just as in the theory[6] of the heat kernel (or Schrödinger propagator) on S^3/Γ, the K of (3) can be found from that on $N = S^3$, which is the universal covering space of S^3/Γ, by adding up the contributions from the preimages of the final point q', say. Equivalently, one can use the eigenfunctions on S^3/Γ expressed as the periodized sums,

$$\phi^{(j)}_{rs}(q) = \left[\frac{2j+1}{|\Gamma|}\right]^{1/2} \frac{1}{2\pi^2}\sum_{\gamma} D^{(j)}_{rs}(\gamma q),$$

where the $D^{(j)}$ are the standard SU(2) representation matrices. The relation between the angular momentum number and the label n of (5) is $n = 2j+1$; $|\Gamma|$ is the order of Γ.

The answer is

$$K(r,q,r',q';t) = \frac{1}{4\pi^2 trr'} e^{-(r^2+r'^2)/4t} \sum_{n,\gamma} I_n\left(\frac{rr'}{2t}\right) n\chi^{(j)}(\gamma q' q^{-1}), \quad (6)$$

$\chi^{(j)}(g)$ is the character of the group element g in the representation labeled by j. Homogeneity says that we can set $q = e$, the unit element,[7] and we see this in (6).

Explicitly, we have

$$\chi^{(j)}(g) = \sin(n\theta_g)/\sin\theta_g,$$

where θ_g is the geodesic distance between the origin (unit element) and the point g. If we use Euler angles ϑ, φ, and ψ as coordinates for g, then we can always rotate the S^3 so that ϑ and φ are constant on the geodesic connecting e and g and we then see that $\theta_g = \psi/2$; θ_g ranges from 0 to 2π.

For us, interest centers on the coincidence limit $r' = r$, $q' = q$. Denote this by $K(r,q;t)$ and from (6) find

$$K(r,q;t) = \frac{1}{4\pi^2 tr^2} e^{-r^2/2t} \sum_{n,\gamma} I_n \frac{r^2}{2t} n\chi^{(j)}(\gamma). \quad (7)$$

Although this is a useful formula, it is not convenient if we wish to integrate it over the cone $C(N)$. We rewrite it, initially, as

$$K(r,q;t) = \frac{1}{4\pi^2 tr^2} e^{-r^2/2t} \sum_{\gamma} \frac{\partial}{\partial\cos\theta_\gamma} \sum_{n=0}^{\infty} I_n \frac{r^2}{2t}\cos(n\theta_\gamma), \quad (8)$$

where we have extended the sum down to $n = 0$, as we may. The second summation is recognized as the corresponding quantity in the S^1 case, treated earlier.[1]

We have here an example of the method whereby application of the operator $\partial/\partial s^2$ (where s is the geodesic distance) produces quantities on a space of two higher dimensions. In flat space, this was used by Hadamard[8] and on spheres by us[9] and others more recently.[10] As we have done it here, the formula for the sphere follows from that on flat space. In fact, since the metric (1) is locally flat, we could have begun with the standard Gaussian heat kernel on R^4 written in (r,q) coordinates and then applied a preimage sum to this. In this way, the Bessel functions could have been avoided. Section V contains the details of this method.

All these statements extend to the full propagator $K(r,q,r',q';t)$. We do not write out the expressions but concentrate on the coincidence limit (8). Thus, either directly from the Gaussian "classical paths" form or by applying a transformation to (8), we find

$$k(r,q;t) = \frac{1}{16\pi^2 t^2} + \frac{1}{8\pi^2 tr^2} \sum_{\gamma\neq e} \frac{\partial}{\partial\cos\theta_\gamma} \times\exp\left(-(1-\cos\theta_\gamma)\frac{r}{2t}\right). \quad (9)$$

The first term comes from the Gaussian kernel on R^4 and, when integrated over $C(N)$, will produce the usual Weyl volume divergence. The remainder gives, on integration, the time-independent expression,

$$K'(t) = \frac{1}{8|\Gamma|}\sum_{\gamma\neq e}\csc^2(\theta_\gamma). \quad (10)$$

This, then, is the required constant term, $16\pi^2 c_2$, in the expansion of the integrated heat kernel. It is easily evaluated for the different Γ. Thus, for lens spaces S^3/Z_m, where $\theta_\gamma = 2\pi k/m$ $(k = 0,...,m-1)$, we find

$$16\pi^2 c_2 = \frac{(m^2+11)(m^2-1)}{360m}, \quad (11)$$

while, for the double dihedral group, D'_m, (S^3/D'_m = "prism space"), calculation yields[11]

$$16\pi^2 c_2 = \frac{(16m^4 + 40m^2 + 360m - 11)}{1440m}.$$

For the remaining groups, T', O', and Y', the values of $16\pi^2 c_2$ are 1505/1728, 4529/3456, and 87109/43200, respectively.

As a general conclusion, we see that the (asymptotic) expansion of the integrated heat kernel on the generalized cone $C(S^3/\Gamma)$ consists of just two terms, the Weyl volume term and the constant term. Curiously, there is no term proportional to $1/t$, as might have been expected.

V. HIGHER SPHERES

The method we employ here is the alternative one mentioned in Sec. IV, which starts from the standard Gaussian heat kernel on R^{d+1},

$$K_0(\mathbf{x},\mathbf{x}';t) = [1/(4\pi t)^{(d+1)/2}]\exp(-(\mathbf{x}-\mathbf{x}')^2/4t).$$

The coordinates $\mathbf{x}$ can be replaced by r and q, where q is the coordinate on S^d.

We are interested in the coincidence limit of the kernel when S^d is replaced by S^d/Γ. By homogeneity this will be independent of the position on the sphere and will be simply a function of r and θ_γ (as well as of t), where θ_γ is the geodesic distance (on the sphere) between the north pole, say, and its image under $\gamma\in\Gamma$.

Specifically we are saying that the kernel we require is given by the preimage sum

$$K(\mathbf{x},\mathbf{x}';t) = \sum_{\gamma} K_0(\mathbf{x},\gamma\mathbf{x}';t), \quad (12)$$

where the action of Γ on $\mathbf{x}$ is defined by $\gamma\mathbf{x} = \gamma(r,q) = (r,\gamma q)$ and q and q' are restricted to a fundamental domain of Γ on S^d.

Setting $\mathbf{x}'$ equal to $\mathbf{x}$, the geometry enables us to write

$$K_0(\mathbf{x},\gamma\mathbf{x};t) = [1/(4\pi t)^{(d+1)/2}]\exp(-(1-\cos\theta_\gamma)r^2/2t).$$

When summed over γ, this is in agreement with (9) and is an easier derivation. As before, $\gamma = e$ gives the Weyl volume term.

It is straightforward to integrate the $\gamma\neq e$ terms over the cone to obtain the generalization of (10),

$$K'(t)=\frac{1}{2^{d+1}|\Gamma|}\sum_{\gamma\neq e}\csc^{d+1}\left(\frac{\theta_\gamma}{2}\right). \tag{13}$$

If d is even, the only possibility for Γ is Z_2 giving elliptic space. Then substituting $\theta_\gamma=\pi$ gives $K'(t)=1/2^{d+2}$, for all d.

For odd d, the classification of *homogeneous* S^d/Γ is given by Wolf (Ref. 12, Corollary 2.7.2). For all odd d there are lens spaces, with $\Gamma=Z_m$, and for $d=3,7,11,\ldots$, additionally, $\Gamma=D'_m$, T', O', and Y'. The case of $d=3$, discussed in Sec. IV, is thus adequate. Nevertheless, we shall evaluate $K'(t)$ for the general dimension in the case of lens spaces. The other values can be easily calculated, if desired.

For $\Gamma=Z_m$, the θ_γ are as before, i.e., $\theta_\gamma=2\pi k/m$ ($k=0,1,\ldots,m-1$). The calculation of the sums of powers of cosecants goes back to Euler.[13] He used purely trigonometric methods but there are, of course, many other ways of proceeding. Information can be found in Refs. 14. A particularly easy way is to use contour integration, although we have not seen this method in print.

The calculations in Ref. 15 produce the identity

$$\sum_{k=1}^{m-1}\cos\left(\frac{2\pi rk}{m}\right)\csc^{2N}\left(\frac{\pi k}{m}\right)$$

$$=\frac{1}{4\pi}(-2)^N m W_{2N}\left(\frac{r}{m}-\frac{1}{2}\right) \tag{14}$$

in terms of the polynomials $W_N(\delta)$ defined in Ref. 15 using Bernoulli polynomials. We do not give their general form here but note that (14) includes all the known results. The cosine coefficient is a "twisting," with r taking integer values from 0 to $m-1$.

The sums that we need are those for $r=0$. The result for $N=2$ has been given in Sec. IV and we simply list now, for $d=5$, 7, and 9, the expressions for $K'(t)$ obtained by combining (13) and (14).

$$K'(t)=\frac{1}{60\,480m}(m^2-1)(2m^4+23m^2+191),\quad d=5$$

$$=\frac{1}{3\,628\,800m}(m^2-1)(m^2+11)(3m^4+10m^2+227),\quad d=7$$

$$=\frac{1}{95\,800\,320m}(m^2-1)(2m^8++35m^6+321m^4+2125m^2+14797),\quad d=9.$$

All these results can be obtained using the expressions in Cheeger's paper[2] but we believe our methods to be more direct, if less rigorous.

VI. DISCUSSION AND CONCLUSION

We have evaluated the short-time expansion of the integrated heat kernel on the generalized cone $C(N)$, where N is a homogeneous spherical space form of unit radius.

The expansion consists of only two terms—the Weyl volume term and the term independent of the time.

Cheeger's general result (Ref. 2, Theorem 7.2) is that, on a piecewise flat pseudomanifold, the expansion ceases after the constant term. Our conclusion agrees with this because such a pseudomanifold can be thought of, roughly, as a complex of flat generalized cones of dimensions up to the dimension n of the pseudomanifold. The heat kernel takes contributions from all these cones and so contains only negative powers of t from $t^{-n/2}$ up to t^0, in steps in $t^{1/2}$.

Explicit expressions have been given for the lens spaces, S^d/Z_m (d odd), S_d/Z_2 (d even) and the three-dimensional space forms.

In particular, for the elliptic case S^3/Z_2, we obtain a value of $\frac{1}{32}$ for the constant term. It is possible to join 16 of these generalized cones to make a closed surface on which the constant term would thus be $\frac{1}{2}$. This surface is the "orbifold" T^4/Z_2 and it is an easy exercise to check that the heat kernel on this surface, constructed as a theta function, has the required expansion. (See the Appendix.) The famous $K3$ surface is obtained if the 16 singular points are blown up by smoothly patching in metrics of the Eguchi–Hanson type.[3]

The interest in such constructions stems, partly, from string theory and the attempt to compactify into orbifolds, for example, into the Z orbifold.[16]

It would also seem possible to construct locally flat, compact combinations of generalized cones, only some of which would be orbifolds since they would not be factored flat spaces, in general. Nevertheless, they could possess symmetry. The cube is not an orbifold. Should one consider compactification onto regular, nonorbifold, generalized polyhedra?

Another physical possibility is to extend the metric as in (2) and consider these as describing some sort of "linear" defect in higher dimensional space-time, similar to the idealized cosmic string (which corresponds to the ordinary cone) or to solutions in $1+2$-dimensional gravity. It would then be possible to calculate the ensuing vacuum polarization, for example.

From the formal point of view, there are a number of extensions that need to be carried out. For example, the class of functions could be generalized to include forms[2] or spinors. One could also include "twisted" fields, which are allowed because of the nontriviality of the first fundamental group, $\pi_1(C)$. Characters belonging to $\mathrm{Hom}(\pi_1,U(1))$ then occur[6] in Eq. (12) and the sums (14) for nonzero r would be needed. Some relevant calculations have been performed in Ref. 17. In addition, a general gauge group G might be considered, requiring $\mathrm{Hom}(\pi_1,G)$.[18]

In a different vein, it would be interesting to look at classical motion and quantum scattering on generalized cones.

ACKNOWLEDGMENTS

I would like to thank Gary Gibbons for reawakening my interest in this topic and for information. Thanks also go to

Simon Jadhav for providing the double dihedral group resuls. I wish also to thank the organizers of the 9th U. K. Institute for Theoretical High-Energy Physicists held at the University of St. Andrews, 1988, where part of this work was performed.

APPENDIX

We require the constant term in the heat kernel expansion on the T^D/Z_2 orbifold. Although it is straightforward to use the theta function form of the heat kernel itself, we shall be a little roundabout and approach the topic via zeta functions, the constant term being related to the value of the zeta function at zero argument by

$$\zeta(0) = \text{const term} - \text{no. of zero modes}. \qquad \text{(A1)}$$

We take the torus T^D to be the unit one so that the eigenfunctions of the Laplacian are porportional to $\exp(i\mathbf{n}\cdot\boldsymbol{\vartheta})$, where the integers n_i range from $-\infty$ to $+\infty$. The corresponding zeta function $\zeta_{T^D}(s)$ is then of Epstein form,

$$\zeta_{T^D}(s) = \sum_{-\infty}^{+\infty}{}'(\mathbf{n}^2)^{-s},$$

where the sum is over all $\mathbf{n}$, except $\mathbf{0}$ (this restriction is denoted by the prime).

The torus is a smooth, flat, compact manifold so that from (A1) we have

$$\zeta_{T^D}(0) = -1 \qquad \text{(A2)}$$

because the constant term vanishes and the only zero mode is $\mathbf{n} = \mathbf{0}$.

On T^D/Z_2 the modes are required, by definition, to be symmetric under the reflection $\boldsymbol{\vartheta} \to -\boldsymbol{\vartheta}$ and so are proportional to $\cos(\mathbf{n}\cdot\boldsymbol{\vartheta})$. The degeneracies are slightly altered. All possible modes are covered by the label values $n_1 = 0,1,2,\ldots$ and $n_i = 0, \mp 1, \mp 2, \ldots$ for $i \neq 1$. The zeta function easily follows,

$$\zeta_{T^D/Z_2}(s) = \tfrac{1}{2}\zeta_{T^D}(s). \qquad \text{(A3)}$$

The $\frac{1}{2}$ is not quite just a volume factor.

Again, there is only one zero mode on T^D/Z_2, so that, from (A1)–(A3)

$$\text{const term on } T^D/Z_2 = \zeta_{T^D/Z_2}(0) + 1 = \tfrac{1}{2}\zeta_{T^D}(0) + 1 = \tfrac{1}{2}.$$

Antisymmetric modes give $-\frac{1}{2}$ since now there is no zero mode.

Because of (A3), all the other heat kernel coefficients vanish, apart from the first one (for which the $\frac{1}{2}$ *is* a volume effect).

There are 2^D fixed points on T^D/Z_2, so that the contribution of each to the constant term in the heat kernel expansion is $1/2^{D+1}$. This agrees with that found for the generalized cone $C(S^{D-1}/Z_2)$ in Sec. V.

[1]J. S. Dowker, Phys. Rev. D **36**, 620 (1987); See also J. S. Dowker J. Math. Phys. **28**, 33 (1987).
[2]J. Cheeger, J. Diff. Geometry **18**, 575 (1983); See also J. Cheeger, Proc. Natl. Acad. Sci. USA **76**, 2103 (1979).
[3]G. W. Gibbons and C. N. Pope, Comm. Math. Phys. **66**, 267 (1979).
[4]P. Lax and R. Phillips, Comm. Pure Appl. Math. **31**, 415 (1978).
[5]W. Seifert and H. Threlfall, Math. Ann. **104**, 1 (1930).
[6]J. S. Dowker, J. Phys. A **5**, 936 (1972).
[7]J. S. Dowker, Ann. Phys. **62**, 361 (1971).
[8]J. Hadamard, *Lectures on Cauchy's Problem* (Yale U. P., New Haven, CT, 1923).
[9]J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 224, 3224 (1976).
[10]J. Cheeger and M. Taylor, Comm. Pure Appl. Math. **35**, 275, 487 (1982); A. Anderson, Phys. Rev. D **37**, 536 (1988).
[11]S. Jadhav (unpublished).
[12]J. Wolf, *Spaces of Constant Curvature* (McGraw-Hill, New York, 1967).
[13]L. Euler, *Introduction in analysin infinitorum* (Lausanne, 1748), Vol. I.
[14]T. J. I'A Bromwich, *Theory of Infinite Series* (MacMillan, London, 1908); E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975); Problem 69-14, SIAM Rev. **13**, 116 (1969).
[15]J. S. Dowker, Phys. Rev. D **36**, 3095, 3742 (1987).
[16]L. Dixon, J. A. Harvey, C. Vafa, and E. Witten, Nucl. Phys. B **261**, 678 (1985); P. Candelas, G. Horowitz, A. Strominger, and E. Witten, *ibid.* **258**, 46 (1985).
[17]J. S. Dowker and R. Banach, J. Phys. A **11**, 2255 (1978); J. S. Dowker and S. Jadhav (in preparation).
[18]R. Banach and J. S. Dowker, J. Phys. A **12**, 2527, 2545 (1979); R. Banach, J. Phys. A **13**, 2179 (1980).

Addition theorems for B functions and other exponentially declining functions

E. Joachim Weniger and E. Otto Steinborn
Institut für Physikalische und Theoretische Chemie, Universität Regensburg, D-8400 Regensburg, West Germany

(Received 27 January 1988; accepted for publication 26 October 1988)

In this paper addition theorems are derived for a special class of exponentially declining functions, the so-called B functions $B_{n,l}^{m}(\alpha,\mathbf{r})$ [E. Filter and E. O. Steinborn, Phys. Rev. A **18**, 1 (1978)]. Although these B functions have a relatively complicated analytical structure they nevertheless have some mathematical properties that are particularly advantageous in connection with multicenter problems. Also, all the commonly occurring exponentially declining functions like, for instance, bound-state hydrogen eigenfunctions and Slater-type functions, can be expressed by simple finite sums of B functions. Consequently, addition theorems for these functions can also be written down immediately. The various addition theorems for B functions are derived by applying suitable generating differential operators to the well-known addition theorem of the special B function $B_{-l,l}^{m}$, which is that solution of the modified Helmholtz equation that is irregular at the origin and regular at infinity [E. J. Weniger and E. O. Steinborn, J. Math. Phys. **26**, 664 (1985)]. All differentiations, which have to be done in this approach, can be expressed in closed form leading to comparatively compact addition theorems.

I. INTRODUCTION

In the mathematical treatment of physical problems it is often necessary to transform wave functions and operators, which may depend upon the coordinates of more than one particle, in such a way that the coordinates of the pertaining particles appear in a computationally more convenient and accessible form. In most cases this requires a separation of variables. Unfortunately, there is no straightforward way to accomplish this task for all functions and operators of physical interest. Problems of that kind, which are quite common in systems with a Coulombic interaction, are particularly troublesome in electronic structure calculations of molecules and solids. If in an orbital approximation the LCAO–MO ansatz is used, it is not only necessary to deal with the coordinates of many electrons but the wave functions are also defined with respect to the coordinates of different nuclei. Similar multicenter problems also occur in the theory of intermolecular forces or in scattering theory.

The aforementioned separation of variables can be accomplished with the help of so-called addition theorems. An addition theorem is a series expansion of a given function $f(\mathbf{x}+\mathbf{y})$ with $\mathbf{x},\mathbf{y}\in\mathbb{R}^3$ in terms of other functions that only depend upon either $\mathbf{x}$ or $\mathbf{y}$. The series expansion may either converge pointwise or with respect to the norm of a suitable Hilbert space. However, in this paper, only addition theorems that converge pointwise will be treated. The probably best known example of such an addition theorem is the Laplace expansion of the Coulomb potential in terms of spherical harmonics Y_l^m,

$$\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\frac{r_<^l}{r_>^{l+1}}Y_l^{m*}\left(\frac{\mathbf{r}_1}{r_1}\right)Y_l^m\left(\frac{\mathbf{r}_2}{r_2}\right),$$

$$r_<=\min(r_1,r_2),\quad r_>=\max(r_1,r_2)\,. \tag{1.1}$$

There is a vast literature on addition theorems. Particularly well studied are the addition theorems of the solutions of the homogeneous Laplace equation,[1–12] of the homogeneous Helmholtz equation,[10,12–19] and of the homogeneous modified Helmholtz equation.[12,17–20] This large number of references not only reflects the physical importance of the solutions of these equations but also the relative ease with which the addition theorems of these functions can be derived. The derivation of addition theorems for other functions, which are not solutions of the equations listed above, turns out to be, in most cases, significantly harder. Ruedenberg[21] and Silverstone[22] showed how addition theorems of relatively general functions may be derived with the help of Fourier transformation. Another general method for the derivation of addition theorems is Löwdin's α function technique,[23] which was used and also extended by numerous other authors.[24–34] Since exponentially declining functions are well suited to serve as basis functions in electronic structure calculations of atoms, molecules, and solids, and since in this context the notorious molecular multicenter integrals cannot be avoided, there is also a large number of articles dealing with the relatively complicated addition theorems of Slater-type functions.[22,23,26,28,29,35–39]

In this paper, we shall derive addition theorems for another class of exponentially declining functions, the so-called B functions.[40] This choice may appear to be somewhat surprising since B functions have a relatively complicated mathematical structure, and at first sight it is by no means obvious that anything can be gained by considering B functions instead of the apparently much simpler Slater-type functions. However, as will be discussed later, B functions have some remarkable mathematical properties that give them a unique position among exponentially declining functions and make them especially useful for molecular calculations. As a result of these advantageous properties it is not

 0022-2488/89/040774-11$02.50

only relatively easy to derive addition theorems for B functins, but these addition theorems also have comparatively simple structures.

Our derivation of addition theorems of B functions closely resembles the zeta function method of Barnett and Coulson.[35] The starting point of this approach is the well-known addition theorem of the Yukawa potential $e^{-\alpha r}/r$. If a given exponentially declining function can be generated by applying a suitable differential operator to the Yukawa potential, then it is at least in principle possible to derive an addition theorem for this function by applying this generating differential operator to the addition theorem of the Yukawa potential. Unfortunately, this idea did not lead to complete success in the case of Slater-type functions. Because of technical problems it was simply not possible to perform the differentiations in closed form. Consequently, the coefficients of the zeta function expansion could only be computed recursively.[35,36] In the case of B functions, however, these problems can be overcome. We shall show in this paper that all differentiations can easily be expressed in closed form if the generating differential operator of a B function is applied to the addition theorem of the Yukawa potential. In our opinion, the relative ease with which this can be accomplished again indicates that, because of their advantageous mathematical properties, B functions indeed assume an exceptional position among exponentially declining functions.

In this context, it should be mentioned that all the commonly occurring exponentially declining functions, like, for instance, Slater-type functions or bound-state hydrogen eigenfunctions, may be expressed by simple finite sums of B functions. Consequently, with the help of the results of this paper addition theorems for all these functions can be written down immediately.

II. DEFINITIONS

For the commonly occurring special functions of mathematical physics we shall use the notations and conventions of Magnus, Oberhettinger, and Soni[41] unless explicitly stated. Hereafter, this reference will be denoted as MOS in the text.

For the spherical harmonics $Y_l^m(\vartheta,\varphi)$ we use the phase convention of Condon and Shortley,[42] i.e., they are defined by the expression

$$Y_l^m(\vartheta,\varphi) = i^{m+|m|}\left[\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}\right]^{1/2} P_l^{|m|}(\cos\vartheta)e^{im\varphi}. \tag{2.1}$$

Here, $P_l^{|m|}(\cos\vartheta)$ is an associated Legendre polynomial,

$$\begin{aligned} P_l^m(x) &= (1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}\frac{(x^2-1)^l}{2^l l!} \\ &= (1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x) . \end{aligned} \tag{2.2}$$

For the regular and irregular solid harmonics we write

$$\mathcal{Y}_l^m(\mathbf{r}) = r^l Y_l^m(\vartheta,\varphi) , \tag{2.3}$$

$$\mathcal{Z}_l^m(\mathbf{r}) = r^{-l-1} Y_l^m(\vartheta,\varphi) . \tag{2.4}$$

It is important to note that the regular solid harmonic $\mathcal{Y}_l^m$ is a homogeneous polynomial of degree l in the Cartesian components x, y, and z of $\mathbf{r}$,[43]

$$\mathcal{Y}_l^m(\mathbf{r}) = \left[\frac{2l+1}{4\pi}(l+m)!(l-m)!\right]^{1/2} \times \sum_{k\geq 0} \frac{(-x-iy)^{m+k}(x-iy)^k z^{l-m-2k}}{2^{m+2k}(m+k)!k!(l-m-2k)!} . \tag{2.5}$$

In Eq. (2.5) the Cartesian components of $\mathbf{r} = (x,y,z)$ can be replaced by the Cartesian components of $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. This yields the differential operator $\mathcal{Y}_l^m(\nabla)$ that is also a spherical tensor of rank l, which we shall call the spherical tensor gradient. A discussion of the properties of this differential operator $\mathcal{Y}_l^m(\nabla)$ and a survey of the relevant literature can be found in articles by Niukkanen,[44] Rashid,[45] and ourselves.[12,46]

For the integral of the product of three spherical harmonics over the surface of the unit sphere in $\mathbb{R}^3$, the so-called Gaunt coefficient, we write

$$\langle l_3 m_3 | l_2 m_2 | l_1 m_1 \rangle = \int Y_{l_3}^{m_3 *}(\Omega) Y_{l_2}^{m_2}(\Omega) Y_{l_1}^{m_1}(\Omega) d\Omega . \tag{2.6}$$

The Gaunt coefficients linearize the product of two spherical harmonics,

$$Y_{l_1}^{m_1}(\Omega) Y_{l_2}^{m_2}(\Omega) = \sum_{l=l_{\min}}^{l_{\max}}{}^{(2)} \langle l m_1+m_2 | l_1 m_1 | l_2 m_2 \rangle Y_l^{m_1+m_2}(\Omega) . \tag{2.7}$$

The symbol $\Sigma^{(2)}$ indicates that the summation proceeds in steps of 2. The summation limits in Eq. (2.7), which follow from the selection rules satisfied by the Gaunt coefficient, are given by[47]

$$l_{\max} = l_1 + l_2 ,$$

$$l_{\min} = \begin{cases} \max(|l_1-l_2|,|m_1+m_2|), \\ \quad \text{if } l_{\max} + \max(|l_1-l_2|,|m_1+m_2|) \text{ is even,} \\ \max(|l_1-l_2|,|m_1+m_2|)+1, \\ \quad \text{if } l_{\max} + \max(|l_1-l_2|,|m_1+m_2|) \text{ is odd.} \end{cases} \tag{2.8}$$

III. SOME PROPERTIES OF *B* FUNCTIONS

In this section we shall review those mathematical properties of B functions that are relevant for the derivation of addition theorems. More complete treatments of the properties of B functions were given elsewhere.[46,48–52]

If $K_\nu(z)$ is a modified Bessel function of the second kind (MOS, p. 66), the reduced Bessel function $\hat{k}_\nu(z)$ is defined by

$$\hat{k}_\nu(z) = (2/\pi)^{1/2} z^\nu K_\nu(z) . \tag{3.1}$$

If the order ν of the reduced Bessel function is negative, we may use

$$\hat{k}_{-\nu}(z) = z^{-2\nu}\hat{k}_\nu(z) . \tag{3.2}$$

This relationship follows from a symmetry property of the modified Bessel function $K_\nu(z)$ (MOS, p. 67).

If the order ν is half integral and positive, $\nu = n + 1/2$ and $n \in \mathbb{N}_0$, a reduced Bessel function may be written as an exponential multiplied by a terminating hypergeometric series ${}_1F_1$,[50]

$$\hat{k}_{n+1/2}(z) = 2^n (1/2)_n e^{-z} \, {}_1F_1(-n; -2n; 2z), \quad n \geqslant 0, \tag{3.3}$$

with $(1/2)_n$ being a Pochhammer symbol (MOS, p. 3). We found out some time ago that the hypergeometric polynomial in Eq. (3.3) has also been investigated independently in the mathematical literature.[53] There, the notation

$$\Theta_n(z) = e^z \hat{k}_{n+1/2}(z) \tag{3.4}$$

is used. Together with some other, closely related polynomials the $\Theta_n(z)$ are called Bessel polynomials. They were applied in such diverse fields as number theory, statistics, and the analysis of complex electrical networks.[53]

As an anisotropic generalization of the reduced Bessel function with half integral order, the so-called B function was introduced,

$$B^m_{n,l}(\alpha,\mathbf{r}) = [2^{n+l}(n+l)!]^{-1} \hat{k}_{n-1/2}(\alpha r) \mathscr{Y}^m_l(\alpha \mathbf{r}), \quad \alpha \in \mathbb{R}_+ . \tag{3.5}$$

Because of the factorial $(n+l)!$, which occurs in the denominator of Eq. (3.5), B functions are defined in the sense of classical analysis only if $n + l \geqslant 0$ holds. However, it could be shown that the definition of a B function according to Eq. (3.5) remains meaningful even if n is a negative integer such that $n + l < 0$ holds. In this case, B "functions" can be identified with derivatives of the three-dimensional delta function,[46]

$$B^m_{-n-l,l}(\alpha,\mathbf{r}) = (4\pi/\alpha^{l+3})(2l-1)!![1 - \alpha^{-2}\nabla^2]^{n-1} \delta^m_l(\mathbf{r}), \quad n \geqslant 1 . \tag{3.6}$$

The spherical delta function δ^m_l is defined by

$$\delta^m_l(\mathbf{r}) = (-1)^l [(2l-1)!!]^{-1} \mathscr{Y}^m_l(\nabla) \delta(\mathbf{r}) . \tag{3.7}$$

The B functions have a remarkable property that seems to be unique among exponentially declining functions: It is extremely easy to generate anisotropic B functions by differentiating scalar B functions. One only has to apply the spherical tensor gradient $\mathscr{Y}^m_l(\nabla)$ to a scalar B function in order to obtain a nonscalar B function,[46,49] i.e., a spherical tensor of rank l,

$$B^m_{n,l}(\alpha,\mathbf{r}) = (-\alpha)^{-l} (4\pi)^{1/2} \mathscr{Y}^m_l(\nabla) B^0_{n+l,0}(\alpha,\mathbf{r}) . \tag{3.8}$$

In the case of other exponentially declining atomic orbitals the corresponding generating differential operators are significantly more complicated.[54]

Of particular importance is the special case $n = -l$ in Eq. (3.8) because it connects the functions $B^m_{-l,l}$ and $B^0_{0,0}$. The function

$$B^m_{-l,l}(\alpha,\mathbf{r}) = \hat{k}_{l+1/2}(\alpha r) \mathscr{Z}^m_l(\alpha \mathbf{r}) \tag{3.9}$$

is the solution of the homogeneous modified Helmholtz equation that is irregular at the origin and regular at infinity,[12] and $B^0_{0,0}$ is essentially the Yukawa potential,

$$e^{-\alpha r}/r = (4\pi)^{1/2} \alpha B^0_{0,0}(\alpha,\mathbf{r}) . \tag{3.10}$$

This fact enabled us to simplify the derivation of the addition theorem of the modified Helmholtz harmonic $B^m_{-l,l}$ considerably.[12] We only had to apply the spherical tensor gradient $\mathscr{Y}^m_l(\nabla)$ to the well-known (MOS, p. 107) addition theorem of the Yukawa potential. The differentiations posed no problems and could all be done in closed form.

Starting from the modified Helmholtz harmonic $B^m_{-l,l}$ all higher B functions $B^m_{n,l}$ with $n > -l$ can be generated. One only has to differentiate $B^m_{-l,l}$ with respect to the scaling parameter α,[49]

$$B^m_{n,l}(\alpha,\mathbf{r}) = \frac{\alpha^{2n+l-1}}{(-2)^{n+l}(n+l)!} \left(\frac{1}{\alpha} \frac{\partial}{\partial \alpha} \right)^{n+l} \alpha^{l+1} B^m_{-l,l}(\alpha,\mathbf{r}) . \tag{3.11}$$

Because of the differential operator $\alpha^{-1}\partial/\partial\alpha$ this generating differential operator for B functions appears to be much more complicated than analogous generating differential operators[49] for Slater-type functions which only contain the differential operator $\partial/\partial\alpha$. However, this is not the case if Bessel functions are to be differentiated because then $\alpha^{-1}\partial/\partial\alpha$ can be applied quite easily whereas $\partial/\partial\alpha$ is very inconvenient. In fact, Eq. (3.11) will be one of the central relationships of this paper because we shall derive addition theorems for B functions by applying the generating differential operator in Eq. (3.11) to the addition theorem of the modified Helmholtz harmonics.

It follows from Eqs. (3.3) and (3.5) that a B function has a relatively complicated structure. However, it could be shown by ourselves[48,49] and shortly afterwards also by Niukkanen[55] that the Fourier transform of a B functions is of exceptional simplicity:

$$\bar{B}^m_{n,l}(\alpha,\mathbf{p}) = (2\pi)^{-3/2} \int e^{-i\mathbf{p}\cdot\mathbf{r}} B^m_{n,l}(\alpha,\mathbf{r}) d^3\mathbf{r} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\alpha^{2n+l-1}}{[\alpha^2 + p^2]^{n+l+1}} \mathscr{Y}^m_l(-i\mathbf{p}) . \tag{3.12}$$

The Fourier transforms of other exponentially declining functions such as Slater-type functions or bound-state hydrogen eigenfunctions are significantly more complicated. In articles by Niukkanen[55] and ourselves[49,56] it was shown that the Fourier transforms of all commonly occurring exponentially declining functions can be expressed as simple finite sums of Fourier transforms of B functions. Since Fourier transformation is a linear operation, this implies that the commonly occurring exponentially declining functions can be written as simple finite sums of B functions. If $\chi^m_{n,l}$ is an unnormalized Slater-type function,

$$\chi^m_{n,l}(\alpha,\mathbf{r}) = (\alpha r)^{n-1} e^{-\alpha r} Y^m_l(\vartheta,\varphi), \quad n - 1 \geqslant l, \tag{3.13}$$

then we can write[40]

$$\chi^m_{n,l}(\alpha,\mathbf{r}) = \sum_{p=p_{\min}}^{n-l} \frac{(-1)^{n-l-p}(n-l)!\, 2^{l+p}(l+p)!}{(2p-n+l)!(2n-2l-2p)!!} B^m_{p,l}(\alpha,\mathbf{r}),$$

$$p_{\min} \begin{cases} (n-l)/2, & \text{if } n-l \text{ is even}, \\ (n-l-1)/2, & \text{if } n-l \text{ is odd}. \end{cases} \tag{3.14}$$

Let $W^m_{n,l}(Z,\mathbf{r})$ be a bound-state eigenfunction of a hydrogenlike ion with nuclear charge Z,

$$W^m_{n,l}(Z,\mathbf{r}) = \{2Z/n\}^{3/2}\{(n-l-1)!/[2n(n+l)!]\}^{1/2} \times e^{-Zr/n} L^{(2l+1)}_{n-l-1}(2Zr/n)\mathscr{Y}^m_l(2Z\mathbf{r}/n). \quad (3.15)$$

Here, $L^{(2l+1)}_{n-l-1}$ is a generalized Laguerre polynomial (MOS, pp. 239–249). Then we can write[56]

$$W^m_{n,l}(Z,\mathbf{r}) = \left(\frac{2Z}{n}\right)^{3/2} \frac{2^{l+1}}{(2l+1)!!}\left\{\frac{n(n+l)!}{2(n-l-1)!}\right\}^{1/2} \times \sum_{t=0}^{n-l-1} \frac{(-n+l+1)_t(n+l+1)_t}{t!(l+3/2)_t} \times B^m_{t+1,l}\left(\frac{Z}{n},\mathbf{r}\right). \quad (3.16)$$

The following set of functions, which was introduced in atomic and molecular calculations by Hylleraas[57] and by Shull and Löwdin,[58] is complete and orthonormal in the Hilbert space $L^2(\mathbb{R}^3)$,

$$\Lambda^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2}\{(n-l-1)!/(n+l+1)!\}^{1/2} \times e^{-\alpha r} L^{(2l+2)}_{n-l-1}(2\alpha r)\mathscr{Y}^m_l(2\alpha\mathbf{r}). \quad (3.17)$$

These functions can be written as[56,59]

$$\Lambda^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2} 2^l \frac{(2n+1)}{(2l+3)!!}\left\{\frac{(n+l+1)!}{(n-l-1)!}\right\}^{1/2} \times \sum_{t=0}^{n-l-1} \frac{(-n+l+1)_t(n+l+2)_t}{t!(l+5/2)_t} \times B^m_{t+1,l}(\alpha,\mathbf{r}). \quad (3.18)$$

Closely related to the hydrogen eigenfunctions are the following functions that were already used in 1928 by Hylleraas[60] and which are commonly called Coulomb Sturmians or simply Sturmians,[61]

$$\Psi^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2}\{(n-l-1)!/[2n(n+l)!]\}^{1/2} \times e^{-\alpha r} L^{(2l+1)}_{n-l-1}(2\alpha r)\mathscr{Y}^m_l(2\alpha\mathbf{r}). \quad (3.19)$$

Comparison of Eqs. (3.15) and (3.19) yields[56]

$$\Psi^m_{n,l}(Z/n,\mathbf{r}) = W^m_{n,l}(Z,\mathbf{r}). \quad (3.20)$$

Hence we only have to replace Z/n in Eq. (3.16) by α in order to obtain a representation of Sturmians in terms of B functions,[56]

$$\Psi^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2} \frac{2^{l+1}}{(2l+1)!!}\left\{\frac{n(n+l)!}{2(n-l-1)!}\right\}^{1/2} \times \sum_{t=0}^{n-l-1} \frac{(-n+l+1)_t(n+l+1)_t}{t!(l+3/2)_t} \times B^m_{t+1,l}(\alpha,\mathbf{r}). \quad (3.21)$$

It is, in fact, by no means obvious that the normalization constants of bound-state hydrogen eigenfunctions and Sturmians should be identical. Hydrogen eigenfunctions are normalized with respect to the norm of the Hilbert space $L^2(\mathbb{R}^3)$, whereas Sturmians are normalized with respect to the norm of the Sobolev space $W^{(1)}_2(\mathbb{R}^3)$.[56,62]

There is another, very important difference. Bound-state hydrogen eigenfunctions are an orthonormal set in $L^2(\mathbb{R}^3)$ which is incomplete without the inclusion of the continuum eigenfunctions, whereas Sturmians are a complete orthonormal set in $W^{(1)}_2(\mathbb{R}^3)$. The importance of the completeness of a set of functions in the Sobolev space $W^{(1)}_2(\mathbb{R}^3)$ for the convergence of the Ritz variational procedure was emphasized by Klahn and Bingel.[63]

The following set of functions is also complete in the Hilbert space $L^2(\mathbb{R}^3)$ and satisfies a biorthogonality relation with Sturmians,[56]

$$\Phi^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2}\left\{\frac{n(n-l-1)!}{2(n+l)!}\right\}^{1/2} \frac{e^{-\alpha r}}{\alpha r} \times L^{(2l+1)}_{n-l-1}(2\alpha r)\mathscr{Y}^m_l(2\alpha\mathbf{r}), \quad (3.22)$$

$$\int_{\mathbb{R}^3} \Phi^{m*}_{n,l}(\alpha,\mathbf{r})\Psi^{m'}_{n',l'}(\alpha,\mathbf{r})d^3\mathbf{r} = \delta_{nn'}\delta_{ll'}\delta_{mm'}. \quad (3.23)$$

These functions were introduced in connection with some weakly convergent expansions of a plane wave[56] which are closely related to the Shibuya–Wulfman expansion[64] of a plane wave in terms of four-dimensional spherical harmonics. They can also be expressed in terms of B functions,[56]

$$\Phi^m_{n,l}(\alpha,\mathbf{r}) = (2\alpha)^{3/2} \frac{2^l}{(2l+1)!!}\left\{\frac{n(n+l)}{2(n-l-1)!}\right\}^{1/2} \times \sum_{t=0}^{n-l-1} \frac{(-n+l+1)_t(n+l+1)_t}{t!(l+3/2)_t} \times B^m_{t,l}(\alpha,\mathbf{r}). \quad (3.24)$$

IV. ADDITION THEOREMS FOR *B* FUNCTIONS

The fundamental relationship for the derivation of addition theorems of B functions is the well-known addition theorem (MOS, p. 107) of the Yukawa potential,

$$\frac{e^{-\alpha w}}{w} = (r\rho)^{-1/2}\sum_{l=0}^{\infty}(2l+1)P_l(\cos\psi) \times I_{l+1/2}(\alpha\rho)K_{l+1/2}(\alpha r),$$
$$w = [r^2+\rho^2-2r\rho\cos\psi]^{1/2}, \quad 0<\rho<r. \quad (4.1)$$

Here, $I_{l+1/2}$ is a modified Bessel function of the first kind (MOS, p. 66). If we now use the so-called addition theorem of the spherical harmonics,

$$P_l(\cos\vartheta) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y^{m*}_l\left(\frac{\mathbf{x}}{x}\right)Y^m_l\left(\frac{\mathbf{y}}{y}\right),$$
$$\cos\vartheta = \mathbf{x}\cdot\mathbf{y}/xy, \quad (4.2)$$

and introduce B functions, we can rewrite Eq. (4.1) in the following way:

$$B^0_{0,0}(\alpha,\mathbf{r}_< + \mathbf{r}_>) = (2\pi^2)^{1/2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l(\alpha r_<)^{-l-1/2} \times I_{l+1/2}(\alpha r_<)\mathscr{Y}^m_l(\alpha\mathbf{r}_<)B^m_{-l,l}(\alpha,\mathbf{r}_>). \quad (4.3)$$

We assume that $\mathbf{r}_<$ is smaller in magnitude than $\mathbf{r}_>$, i.e., $|\mathbf{r}_<|<|\mathbf{r}_>|$.

Starting from Eq. (4.3), addition theorems for B functions with arbitrary values of n, l, and m can be derived by applying suitable generating differential operators. Here, we essentially have two different alternatives. We can first differentiate with respect to the scaling parameter α according to Eq. (3.11) and then apply the spherical tensor gradient $\mathcal{Y}_l^m(\nabla)$ according to Eq. (3.8), or, we can first apply the spherical tensor gradient and then differentiate with respect to the scaling parameter α. In our opinion, the second alternative is to be preferred. This is partly due to the fact that we do not have to do the first step anymore. We could show previously that the application of the spherical tensor gradient to the addition theorem (4.3) gives the addition theorem of the modified Helmholtz harmonics $B^m_{-l,l}$,[12]

$$
\begin{aligned}
B^m_{-l,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)
&= (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} (\alpha r_<)^{-l_1-1/2} \\
&\quad \times I_{l_1+1/2}(\alpha r_<) \mathcal{Y}_{l_1}^{m_1 *}(\alpha \mathbf{r}_<) \\
&\quad \times \sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)} \langle l_2 m+m_1 | lm | l_1 m_1 \rangle B^{m+m_1}_{-l_2,l_2}(\alpha,\mathbf{r}_>) .
\end{aligned}
\tag{4.4}
$$

The summation limits $l_2^{\min}$ and $l_2^{\max}$ are given in Eq. (2.8). In order to simplify the application of the generating differential operator (3.11) this addition theorem (4.4) is first rewritten in the following way:

$$
\begin{aligned}
B^m_{-l,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)
&= 4\pi \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} (\alpha r_<)^{-1/2} \\
&\quad \times I_{l_1+1/2}(\alpha r_<) Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right) \\
&\quad \times \sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)} \langle l_2 m+m_1 | lm | l_1 m_1 \rangle (\alpha r_>)^{-1/2} \\
&\quad \times K_{l_2+1/2}(\alpha r_>) Y_{l_2}^{m_2}\left(\frac{\mathbf{r}_>}{r_>}\right) .
\end{aligned}
\tag{4.5}
$$

If we now combine Eqs. (3.11) and (4.5) we obtain

$$
\begin{aligned}
B^m_{n,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)
&= 4\pi \frac{\alpha^{2n+l-1}(r_< r_>)^{-1/2}}{(-2)^{n+l}(n+l)!} \\
&\quad \times \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right) \\
&\quad \times \sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)} \langle l_2 m+m_1 | lm | l_1 m_1 \rangle Y_l^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right) \\
&\quad \times \left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\} .
\end{aligned}
\tag{4.6}
$$

In order to be able to do the remaining differentiations we need the following differential properties of modified Bessel functions (MOS, p. 67),

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^\nu I_\nu(z) = z^{\nu-m} I_{\nu-m}(z) , \tag{4.7}$$

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^{-\nu} I_\nu(z) = z^{-\nu-m} I_{\nu+m}(z) , \tag{4.8}$$

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^\nu K_\nu(z) = (-1)^m z^{\nu-m} K_{\nu-m}(z) , \tag{4.9}$$

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^{-\nu} K_\nu(z) = (-1)^m z^{-\nu-m} K_{\nu+m}(z) . \tag{4.10}$$

In addition, we use

$$\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^n = 2^n \left(\frac{\partial}{\partial(\alpha^2)}\right)^n . \tag{4.11}$$

Because of this relationship it is possible to use the Leibniz formula for the differentiation of a product repeatedly. For instance, we can write

$$
\begin{aligned}
&\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\} \\
&= 2^{n+l} \left\{\frac{\partial}{\partial(\alpha^2)}\right\}^{n+l} \{[\alpha^{l_1+1/2} I_{l_1+1/2}(\alpha r_<)] \\
&\quad \times [\alpha^{l_2+1/2} K_{l_2+1/2}(\alpha r_>)] \alpha^{l-l_1-l_2-1}\}
\end{aligned}
\tag{4.12}
$$

$$
\begin{aligned}
&= \sum_{q=0}^{n+l} \binom{n+l}{q} \left\{\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{n+l-q} \alpha^{l_1+1/2} I_{l_1+1/2}(\alpha r_<)\right\} \\
&\quad \times \sum_{s=0}^{q} \binom{q}{s} \left\{\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{q-s} \alpha^{l_2+1/2} K_{l_2+1/2}(\alpha r_>)\right\} \\
&\quad \times 2^s \left\{\frac{\partial}{\partial(\alpha^2)}\right\}^s \alpha^{l-l_1-l_2-1} .
\end{aligned}
\tag{4.13}
$$

Now, all differentiations can be done quite easily. If we use Eqs. (4.7) and (4.9) we obtain

$$
\begin{aligned}
&\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{n+l-q} \alpha^{l_1+1/2} I_{l_1+1/2}(\alpha r_<) \\
&= \alpha^{l_1-n-l+q+1/2} r_<^{n+l-q} I_{l_1-n-l+q+1/2}(\alpha r_<) ,
\end{aligned}
\tag{4.14}
$$

$$
\begin{aligned}
&\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{q-s} \alpha^{l_2+1/2} K_{l_2+1/2}(\alpha r_>) \\
&= (-1)^{q-s} \alpha^{l_2-q+s+1/2} r_>^{q-s} K_{l_2-q+s+1/2}(\alpha r_>) .
\end{aligned}
\tag{4.15}
$$

We also find

$$
\begin{aligned}
&2^s \left\{\frac{\partial}{\partial(\alpha^2)}\right\}^s \alpha^{l-l_1-l_2-1} \\
&= (-2)^s \left(\frac{l_1+l_2-l+1}{2}\right)_s \alpha^{l-l_1-l_2-2s-1} .
\end{aligned}
\tag{4.16}
$$

Combination of Eqs. (4.12)–(4.16) gives

$$
\begin{aligned}
&\left(\frac{1}{\alpha}\frac{\partial}{\partial \alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\} \\
&= \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} r_<^{n+l-q} I_{l_1-n-l+q+1/2}(\alpha r_<) \\
&\quad \times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(\frac{l_1+l_2-l+1}{2}\right)_s \frac{r_>^{q-s}}{\alpha^{n+s}} \\
&\quad \times K_{l_2-q+s+1/2}(\alpha r_>) .
\end{aligned}
\tag{4.17}
$$

Inserting this into Eq. (4.6) gives us an addition theorem for B functions:

$$B^m_{n,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= \frac{4\pi}{(-2)^{n+l}(n+l)!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times \sum_{l_2=l_2^{\min}}^{l_2^{\max}} {}^{(2)} \langle l_2 m + m_1 | l m | l_1 m_1 \rangle Y_{l_2}^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} (\alpha r_<)^{n+l-q-1/2}$$
$$\times I_{l_1-n-l+q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(\frac{l_1+l_2-l+1}{2}\right)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l_2-q+s+1/2}(\alpha r_>) . \quad (4.18)$$

Alternative versions of the B function addition theorem can be derived if the expression, which is to be differentiated with respect to α, is factorized differently. For instance, we can choose

$$\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)$$
$$= [\alpha^{-l_1-1/2} I_{l_1+1/2}(\alpha r_<)]$$
$$\times [\alpha^{-l_2-1/2} K_{l_2+1/2}(\alpha r_>)] \alpha^{l+l_1+l_2+1} . \quad (4.19)$$

Repeated application of the Leibniz formula in connection with Eqs. (4.8) and (4.10) yields

$$\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\}$$
$$= \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} r_<^{n+l-q} I_{l_1+n+l-q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(-\frac{l_1+l_2+l+1}{2}\right)_s \frac{r_>^{q-s}}{\alpha^{n+s}}$$
$$\times K_{l_2+q-s+1/2}(\alpha r_>) . \quad (4.20)$$

Inserting Eq. (4.20) into Eq. (4.6) gives another version of the addition theorem for B functions:

$$B^m_{n,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= \frac{4\pi}{(-2)^{n+l}(n+l)!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times \sum_{l_2=l_2^{\min}}^{l_2^{\max}} {}^{(2)} \langle l_2 m + m_1 | l m | l_1 m_1 \rangle Y_{l_2}^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} (\alpha r_<)^{n+l-q-1/2}$$
$$\times I_{l_1+n+l-q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(-\frac{l_1+l_2+l+1}{2}\right)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l_2+q-s+1/2}(\alpha r_>) . \quad (4.21)$$

Next we choose the following factorization:

$$\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)$$
$$= [\alpha^{l_1+1/2} I_{l_1+1/2}(\alpha r_<)]$$
$$\times [\alpha^{-l_2-1/2} K_{l_2+1/2}(\alpha r_>)] \alpha^{l-l_1+l_2} . \quad (4.22)$$

Repeated application of the Leibniz formula in combination with Eqs. (4.7) and (4.10) yields

$$\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\}$$
$$= \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} r_<^{n+l-q} I_{l_1-n-l+q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{\min(q,\Delta l_1)} 2^s \binom{q}{s} (-\Delta l_1)_s \frac{r_>^{q-s}}{\alpha^{n+s}} \quad (4.23)$$
$$\times K_{l_2+q-s+1/2}(\alpha r_>) ,$$
$$\Delta l_1 = (l - l_1 + l_2)/2 .$$

Because of the selection rules satisifed by the Gaunt coefficients in Eq. (4.6), Δl_1 is always either a positive integer or zero. If we now insert Eq. (4.23) into Eq. (4.6) we obtain another addition theorem for B functions:

$$B^m_{n,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= \frac{4\pi}{(-2)^{n+l}(n+l)!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times \sum_{l_2=l_2^{\min}}^{l_2^{\max}} {}^{(2)} \langle l_2 m + m_1 | l m | l_1 m_1 \rangle Y_{l_2}^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} (\alpha r_<)^{n+l-q-1/2}$$
$$\times I_{l_1-n-l+q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{\min(q,\Delta l_1)} 2^s \binom{q}{s} (-\Delta l_1)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l_2+q-s+1/2}(\alpha r_>) . \quad (4.24)$$

Finally, we choose the following factorization:

$$\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)$$
$$= [\alpha^{-l_1-1/2} I_{l_1+1/2}(\alpha r_<)]$$
$$[\alpha^{l_2+1/2} K_{l_2+1/2}(\alpha r_>)] \alpha^{l+l_1-l_2} . \quad (4.25)$$

Repeated application of the Leibniz formula in connection with Eqs. (4.8) and (4.9) yields:

$$\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^{n+l} \{\alpha^l I_{l_1+1/2}(\alpha r_<) K_{l_2+1/2}(\alpha r_>)\}$$
$$= \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} r_<^{n+l-q} I_{l_1+n+l-q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{\min(q,\Delta l_2)} 2^s \binom{q}{s} (-\Delta l_2)_s \frac{r_>^{q-s}}{\alpha^{n+s}} \quad (4.26)$$
$$\times K_{l_2-q+s+1/2}(\alpha r_>) ,$$
$$\Delta l_2 = (l + l_1 - l_2)/2 .$$

Because of the selection rules satisfied by the Gaunt coefficients in Eq. (4.6), Δl_2 is always either a positive integer or zero. If we now insert Eq. (4.26) into Eq. (4.6) we obtain another addition theorem for B functions:

$$B^{m}_{n,l}(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= \frac{4\pi}{(-2)^{n+l}(n+l)!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} Y_{l_1}^{m_1 *}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times \sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)} \langle l_2 m + m_1 | lm | l_1 m_1 \rangle Y_{l_2}^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n+l} (-1)^q \binom{n+l}{q} (\alpha r_<)^{n+l-q-1/2}$$
$$\times I_{l_1+n+l-q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{\min(q,\Delta l_2)} 2^s \binom{q}{s} (-\Delta l_2)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l_2-q+s+1/2}(\alpha r_>) . \qquad (4.27)$$

Setting $l=0$ in the addition theorems (4.18), (4.21), (4.24), and (4.27), and observing

$$\langle l_2 m_2 | 00 | l_1 m_1 \rangle = (4\pi)^{-1/2} \delta_{l_1 l_2} \delta_{m_1 m_2} \qquad (4.28)$$

gives us the analogous addition theorems for reduced Bessel functions. However, unlike the B function addition theorems which have roughly the same complexity, these addition theorems for reduced Bessel functions differ with respect to the number of inner sums. For instance, from Eqs. (4.18) and (4.21) we obtain

$$\hat{k}_{n-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (-1)^n 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n} (-1)^q \binom{n}{q} (\alpha r_<)^{n-q-1/2}$$
$$\times I_{l-n+q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(l+\frac{1}{2}\right)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l-q+s+1/2}(\alpha r_>) \qquad (4.29)$$
$$= (-1)^n 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n} (-1)^q \binom{n}{q} (\alpha r_<)^{n-q-1/2}$$
$$\times I_{l+n-q+1/2}(\alpha r_<)$$
$$\times \sum_{s=0}^{q} 2^s \binom{q}{s} \left(-l-\frac{1}{2}\right)_s (\alpha r_>)^{q-s-1/2}$$
$$\times K_{l+q-s+1/2}(\alpha r_>) . \qquad (4.30)$$

In the same way we obtain from Eqs. (4.24) and (4.27)

$$\hat{k}_{n-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (-1)^n 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n} (-1)^q \binom{n}{q} (\alpha r_<)^{n-q-1/2}$$
$$\times I_{l-n+q+1/2}(\alpha r_<)(\alpha r_>)^{q-1/2} K_{l+q+1/2}(\alpha r_>) \qquad (4.31)$$
$$= (-1)^n 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{q=0}^{n} (-1)^q \binom{n}{q} (\alpha r_<)^{n-q-1/2}$$
$$\times I_{l+n-q+1/2}(\alpha r_<)(\alpha r_>)^{q-1/2} K_{l-q+1/2}(\alpha r_>) . \qquad (4.32)$$

Obviously, the addition theorems (4.29) and (4.30) have a more complicated structure than the addition theorems (4.31) and (4.32).

V. RELATION TO PREVIOUS WORK

Our starting point is the following expansion of the modified Bessel function of the second kind in terms of Gegenbauer polynomials (MOS, p. 107):

$$(\alpha|\mathbf{r}_< - \mathbf{r}_>|)^\nu K_\nu(\alpha|\mathbf{r}_< - \mathbf{r}_>|)$$
$$= 2^{-\nu}\Gamma(-\nu)(\alpha r_<)^\nu(\alpha r_>)^\nu$$
$$\times \sum_{m=0}^{\infty} (m-\nu) C_m^{-\nu}(\cos\psi)$$
$$\times I_{m-\nu}(\alpha r_<) K_{m-\nu}(\alpha r_>) ,$$
$$\cos\psi = \mathbf{r}_< \cdot \mathbf{r}_> / r_< r_> , \quad \nu \notin \mathbb{N}_0 . \qquad (5.1)$$

This expansion can easily be reformulated as an addition theorem if the Gegenbauer polynomials in Eq. (5.1) can be expressed in terms of Legendre polynomials. For that purpose, we use[65]

$$C_m^{-\nu}(\cos\psi) = \sum_{s=0}^{[m/2]} \frac{(-\nu)_{m-s}(-\nu-1/2)_s}{s!(3/2)_{m-s}}$$
$$\times (2m-4s+1) P_{m-2s}(\cos\psi) . \qquad (5.2)$$

Here, $[m/2]$ is the integral part of $m/2$, i.e., the largest integer n satisfying $n \leqslant m/2$. If we insert Eq. (5.2) into Eq. (5.1) and introduce a new summation variable $l = m - 2s$, we obtain the following expansion of the modified Bessel function of the second kind in terms of Legendre polynomials:

$$(\alpha|\mathbf{r}_< - \mathbf{r}_>|)^\nu K_\nu(\alpha|\mathbf{r}_< - \mathbf{r}_>|)$$
$$= 2^{-\nu}\Gamma(-\nu)(\alpha r_<)^\nu(\alpha r_>)^\nu$$
$$\times \sum_{l=0}^{\infty} (2l+1) P_l(\cos\psi)$$
$$\times \sum_{s=0}^{\infty} \frac{(-\nu)_{l+s}(-\nu-1/2)_s}{s!(3/2)_{l+s}} (l+2s-\nu)$$
$$\times I_{l+2s-\nu}(\alpha r_<) K_{l+2s-\nu}(\alpha r_>) . \qquad (5.3)$$

In the next step we use Eqs. (3.1) and (4.2). This gives us an addition theorem of the reduced Bessel functions with nonintegral but otherwise arbitrary orders ν:

$$\hat{k}_\nu(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (32\pi)^{1/2} 2^{-\nu}\Gamma(-\nu)(\alpha r_<)^\nu(\alpha r_>)^\nu$$
$$\times \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{s=0}^{\infty}\frac{(-\nu)_{l+s}(-\nu-1/2)_s}{s!(3/2)_{l+s}}$$
$$\times (l+2s-\nu) I_{l+2s-\nu}(\alpha r_<) K_{l+2s-\nu}(\alpha r_>),$$
$$\nu \notin \mathbb{N}_0. \tag{5.4}$$

If we now assume that the order ν is half-integral, $\nu = n - \frac{1}{2}$, $n \in \mathbb{N}_0$, the innermost series in Eq. (5.4) terminates after a finite number of terms and we obtain an addition theorem for a reduced Bessel function with half-integral order:

$$\hat{k}_{n-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (-1)^n 8\pi[(2n-1)!!]^{-1}(\alpha r_<)^{n-1/2}(\alpha r_>)^{n-1/2}$$
$$\times \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \sum_{s=0}^{n}\frac{(-n)_s(1/2-n)_{l+s}}{s!(3/2)_{l+s}}$$
$$\times (l+2s-n+1/2) I_{l+2s-n+1/2}(\alpha r_<)$$
$$\times K_{l+2s-n+1/2}(\alpha r_>). \tag{5.5}$$

This relationship is equivalent to an addition theorem derived previously by Steinborn and Filter.[40]

With respect to its structure, the addition theorem (5.5) differs significantly from the alternative versions which were derived in the last section. This is not necessarily surprising since the derivations were completely different. However, we want to show in the sequel that the addition theorem (5.5) can nevertheless be obtained by differentiating the addition theorem of the Yukawa potential, Eq. (4.3), with respect to the scaling parameter α. First, we consider in Eq. (3.11) the case $l = 0$, which yields

$$\hat{k}_{n-1/2}(\alpha r) = (-1)^n \alpha^{2n-1}\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^n [\alpha \hat{k}_{-1/2}(\alpha r)]. \tag{5.6}$$

Then, we rewrite the addition theorem of the Yukawa potential, Eq. (4.3), in the following way:

$$\hat{k}_{-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= 4\pi(\alpha r_<)^{-1/2}(\alpha r_>)^{-1/2}$$
$$\times \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times I_{l+1/2}(\alpha r_<) K_{l+1/2}(\alpha r_>). \tag{5.7}$$

Combination of Eqs. (5.6) and (5.7) yields

$$\hat{k}_{n-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (-1)^n 4\pi\alpha^{2n-1}(r_< r_>)^{-1/2}$$
$$\times \sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right) Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)$$
$$\times \left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^n \{I_{l+1/2}(\alpha r_<) K_{l+1/2}(\alpha r_>)\}. \tag{5.8}$$

Next, we need the following integral representation for the product of two modified Bessel functions (MOS, p. 98):

$$2 I_\nu(2az^{1/2}) K_\nu(2bz^{1/2}) = \int_0^\infty t^{-1} e^{-zt-(a^2+b^2)/t} I_\nu\left(\frac{2ab}{t}\right) dt,$$
$$a < b, \quad \mathrm{Re}(z) > 0. \tag{5.9}$$

With the help of the substitutions $x = 2z^{1/2}$ and $s = 1/t$ we obtain

$$I_\nu(ax) K_\nu(bx)$$
$$= \frac{1}{2}\int_0^\infty s^{-1} e^{-(a^2+b^2)s - x^2/4s} I_\nu(2abs) ds, \quad a < b. \tag{5.10}$$

Combination of Eqs. (4.11) and (5.10) yields

$$\left(\frac{1}{x}\frac{\partial}{\partial x}\right)^n \{I_\nu(ax) K_\nu(bx)\}$$
$$= \frac{(-1)^n}{2^{n+1}}\int_0^\infty s^{-n-1} e^{-(a^2+b^2)s - x^2/4s} I_\nu(2abs) ds. \tag{5.11}$$

If we compare Eqs. (5.1) and (5.11) we observe that the integral representation in Eq. (5.11) can be reduced to a sum of products of the type $I_\mu(ax)K_\mu(bx)$ provided that we are able to express the function $s^{-n} I_\nu(2abs)$ in terms of functions of the type $I_\mu(2abs)$. For that purpose we use[66]

$$\left(\frac{z}{2}\right)^{\mu-\nu} J_\nu(z)$$
$$= \sum_{m=0}^{\infty}\frac{\Gamma(\mu+m)\Gamma(\nu-\mu+1)(\mu+2m)}{m!\Gamma(\nu-\mu-m+1)\Gamma(\nu+m+1)}$$
$$\times J_{\nu+2m}(z). \tag{5.12}$$

If we now use (MOS, p. 66)

$$J_\nu(iz) = e^{i\nu\pi/2} I_\nu(z), \tag{5.13}$$

we find

$$\left(\frac{z}{2}\right)^{\mu-\nu} I_\nu(z)$$
$$= \sum_{m=0}^{\infty}(\mu+2m)\frac{\Gamma(\mu+m)(\mu-\nu)_m}{m!\Gamma(\nu+m+1)} I_{\mu+2m}(z). \tag{5.14}$$

If we have $\mu = \nu - k$, $k \in \mathbb{N}_0$, the infinite series in Eq. (5.14) terminates yielding

$$z^{-k} I_\nu(z)$$
$$= 2^{-k}\sum_{m=0}^{k}(\nu+2m-k)\frac{\Gamma(\nu+m-k)(-k)_m}{m!\Gamma(\nu+m+1)}$$
$$\times I_{\nu+2m-k}(z). \tag{5.15}$$

From this relationship we may immediately deduce

$$s^{-n} I_\nu(2abs)$$
$$= (ab)^n \sum_{t=0}^{n}(\nu+2t-n)\frac{\Gamma(\nu-n+t)(-n)_t}{t!\Gamma(\nu+t+1)}$$
$$\times I_{\nu+2t-n}(2abs). \tag{5.16}$$

If we combine Eqs. (5.10), (5.11), and (5.16), we find a differentiation formula for the product of modified Bessel functions which seems to be new:

$$\left(\frac{1}{x}\frac{\partial}{\partial x}\right)^n\{I_\nu(ax)K_\nu(bx)\}$$
$$=\frac{(-ab)^n}{2^n}\sum_{t=0}^{n}\frac{\Gamma(\nu-n+t)(-n)_t}{\Gamma(\nu+t+1)t!}$$
$$(\nu+2t-n)I_{\nu+2t-n}(ax)K_{\nu+2t-n}(bx), \quad (5.17)$$
$$a<b.$$

With the help of this relationship the remaining differentiations in Eq. (5.8) can be done in closed form and we obtain

$$\hat{k}_{n-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$=(4\pi/2^n)(\alpha r_<)^{n-1/2}(\alpha r_>)^{n-1/2}$$
$$\times\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l Y_l^{m*}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times Y_l^m\left(\frac{\mathbf{r}_>}{r_>}\right)\sum_{t=0}^{n}\left(l+2t-n+\frac{1}{2}\right)$$
$$\times\frac{(-n)_t\Gamma(l-n+t+1/2)}{t!\Gamma(l+t+3/2)}I_{l+2t-n+1/2}(\alpha r_<)$$
$$\times K_{l+2t-n+1/2}(\alpha r_>). \quad (5.18)$$

Now we only need the relationship

$$\frac{\Gamma(l-n+t+\frac{1}{2})}{\Gamma(l+t+\frac{3}{2})}=\frac{(-1)^n 2^{n+1}}{(2n-1)!!}\frac{(\frac{1}{2}-n)_{l+t}}{t!(3/2)_{l+t}} \quad (5.19)$$

to see that Eqs. (5.5) and (5.15) are indeed identical.

It is, of course, possible to derive alternative addition theorems for B functions by applying the spherical tensor gradient $\mathcal{Y}_l^m(\nabla)$ to Eq. (5.18). However, this would lead to expressions less compact than the addition theorems derived in the last section. Consequently, we shall not consider them explicitly here. In the same way, we could apply the spherical tensor gradient to the addition theorems (4.29)–(4.32) for reduced Bessel functions. But, again we do not see that the results would be more compact than the B function addition theorems we already know.

All addition theorems in this paper have the following general structure:

$$f_l^m(\mathbf{r}_< + \mathbf{r}_>)=\sum_{l_1=0}^{\infty}\sum_{m_1=-l_1}^{l_1}(-1)^{l_1}Y_{l_1}^{m_1*}\left(\frac{\mathbf{r}_<}{r_<}\right)$$
$$\times\sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)}\langle l_2 m+m_1|lm|l_1 m_1\rangle$$
$$\times F_{l_1 l_2}^l(r_<,r_>)Y_{l_2}^{m+m_1}\left(\frac{\mathbf{r}_>}{r_>}\right). \quad (5.20)$$

Since the function f_l^m is, by assumption, an irreducible spherical tensor,

$$f_l^m(\mathbf{r})=\varphi_l(r)Y_l^m(\mathbf{r}/r), \quad (5.21)$$

the structure of the angular momentum sums in Eq. (5.20) containing spherical harmonics and Gaunt coefficients is determined by group theory.[21,22,67] Only the "radial functions" $F_{l_1 l_2}^l(r_<,r_>)$ are specific for the function f_l^m that is to be expanded.

In the literature, several general methods for the determination of the $F_{l_1 l_2}^l(r_<,r_>)$ are described. For instance, Ruedenberg[21] and Silverstone[22] showed how an addition theorem for an irreducible spherical tensor f_l^m can be obtained with the help of Fourier transformation. In this approach, $F_{l_1 l_2}^l(r_<,r_>)$ is given as a radial integral involving two Bessel functions of the first kind and the radial part of the Fourier transform of f_l^m. Apart from some more or less trivial cases, the remaining radial integrals are extremely complicated and it is very hard to obtain explicit expressions for them.

Another possibility for the determination of the radial functions $F_{l_1 l_2}^l(r_<,r_>)$ in Eq. (5.20) consists of the rearrangement of Taylor expansions. This was discussed by Santos,[68] Bayman,[69] and Niukkanen.[44] In this approach, $F_{l_1 l_2}^l(r_<,r_>)$ is given as an infinite series involving powers and differential operators which act upon the radial part of f_l^m. Again, only relatively simple addition theorems could be derived with the help of this approach.

The radial functions $F_{l_1 l_2}^l(r_<,r_>)$ in Eq. (5.20) are uniquely determined since they serve as coefficients in orthogonal expansions. However, in view of the complexity of these $F_{l_1 l_2}^l(r_<,r_>)$ and since several different approaches, which all involve complicated mathematical operations, are, at least in principle, available for their determination, it is likely that also several different explicit expressions for a radial function $F_{l_1 l_2}^l(r_<,r_>)$ can be obtained, if such an explicit expression can be derived at all. The numerous different versions of the addition theorems of B functions, which either have been derived already or which probably still can be derived, are a striking example for the nonuniqueness of the functional form of these $F_{l_1 l_2}^l(r_<,r_>)$.

In view of this multitude of different addition theorems for the same function it would, of course, be desirable to evaluate the relative merits and disadvantages of the different addition theorems for the same function. However, we believe that a context-free evaluation of the different versions of an addition theorem will only lead to conclusions that are more or less trivial and that not much insight can be gained that way. For instance, it is immediately obvious that the addition theorems (4.29) and (4.30) for reduced Bessel functions are more complicated than the addition theorems (4.31), (4.32), and (5.5), since they contain two nested inner sums instead of one, but we are definitely not able to differentiate among the remaining three addition theorems (4.31), (4.32), and (5.5) on the basis of general considerations alone, i.e., without explicitly specifying the context in which they are to be applied.

Because of the same reasons it is also relatively hard to compare the different variants of the addition theorem for Slater-type functions not only among themselves but also with the addition theorems for B functions which because of Eq. (3.13) can be used for the construction of new addition theorems for Slater-type functions. A further complication is that according to Silverstone and Moats[27] many addition theorems, which are based upon Löwdin's α function method,[23] are not of the form of Eq. (5.20). This is due to the fact that in the original form of Löwdin's α function method a special orientation of the coordinate system was assumed that leads to some simplifications among the spherical harmonics and the Gaunt coefficients in expansions of the type of Eq. (5.20). Consequently, the addition theorems for

Slater-type functions published by Sharma,[24,26] Duff,[25] Jones and Weatherford,[28] and Rashid[29,30] are not directly comparable with the addition theorems of this paper since they assume Löwdin's[23] simplifying orientation of the coordinate system and are therefore somewhat less general.

It is also a typical feature of the addition theorems mentioned above that they do not contain modified Bessel functions as the addition theorems of this paper do. Instead they contain products of the type $e^{\pm \alpha r_<} e^{-\alpha r_>}$ multiplied by powers of $r_<$ and $r_>$. The equivalence of these two different approaches can be proved quite easily. We only need the following representations for modified Bessel functions with half-integral orders[70]:

$$\begin{aligned} I_{\pm(n+1/2)}(z) &= (2\pi z)^{-1/2}\Big\{ e^{z} \sum_{\nu=0}^{n} \frac{(-1)^{\nu}(n+\nu)!}{\nu!(n-\nu)!(2z)^{\nu}} \\ &\quad \mp (-1)^{n} e^{-z} \sum_{\nu=0}^{n} \frac{(n+\nu)!}{\nu!(n-\nu)!(2z)^{\nu}} \Big\}, \end{aligned} \tag{5.22}$$

$$K_{n+1/2}(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z} \sum_{\nu=0}^{n} \frac{(n+\nu)!}{\nu!(n-\nu)!(2z)^{\nu}}. \tag{5.23}$$

Most closely related with our approach is an addition theorem for Slater-type functions that was derived by Silverstone[22] using Fourier transformation and that also contains products of modified Bessel functions I and K. In fact, it can be shown that Silverstone's addition theorem[22] can also be obtained by differentiating the addition theorem of the modified Helmholtz harmonics $B^{m}_{-l,l'}$, Eq. (4.4). The generating differential operator, which has to be used, can be obtained by combining[49]

$$\chi^{m}_{n,l}(\alpha,\mathbf{r}) = \alpha^{n-1}\left(-\frac{\partial}{\partial\alpha}\right)^{n-l-1} \alpha^{-l}\chi^{m}_{l+1,l}(\alpha,\mathbf{r}) \tag{5.24}$$

and

$$\begin{aligned} B^{m}_{1,l}(\alpha,\mathbf{r}) &= [2^{l+1}(l+1)!]^{-1}\chi^{m}_{l+1,l}(\alpha,\mathbf{r}) \\ &= \frac{\alpha^{l+1}}{(-2)^{l+1}(l+1)!}\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^{l+1} \alpha^{l+1} B^{m}_{-l,l}(\alpha,\mathbf{r}). \end{aligned} \tag{5.25}$$

This yields

$$\begin{aligned} \chi^{m}_{n,l}(\alpha,\mathbf{r}) &= (-\alpha)^{n-1}\left(\frac{\partial}{\partial\alpha}\right)^{n-l-1} \alpha\left(\frac{1}{\alpha}\frac{\partial}{\partial\alpha}\right)^{l+1} \alpha^{l+1} \\ &\quad \times B^{m}_{-l,l}(\alpha,\mathbf{r}). \end{aligned} \tag{5.26}$$

As we could show in this paper, the differential operator $(\alpha^{-1}\partial/\partial\alpha)$ poses no particular problems. However it is by no means a simple task to obtain closed-form expressions if one has to apply higher powers of $\partial/\partial\alpha$ to products of modified Bessel functions as they occur in the addition theorems for $B^{m}_{1,l}$. It is probably much simpler to proceed as we suggested, i.e., to combine Eq. (3.14) with one of the addition theorems for B functions.

Finally, we would like to remark that we are not aware of any reference dealing with pointwise convergent addition theorems of the other sets of exponentially declining functions, which are defined in Eqs. (3.15), (3.17), (3.19), and (3.22). The addition theorems of B functions in combination with Eqs. (3.16), (3.18), (3.21), and (3.24) allow a convenient and systematic construction of addition theorems for these functions. This should simplify the applications of these exponentially declining functions in all situations in which multicenter problems can occur, especially in LCAO–MO calculations of molecules, clusters, and solids.

ACKNOWLEDGMENT

E.O.S. thanks the Fonds der Chemischen Industrie for financial support.

[1] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea, New York, 1965), Chap. IV.
[2] M. E. Rose, J. Math. Phys. (Cambridge, MA) **37**, 215 (1958).
[3] M. I. Seaton, Proc. Phys. Soc. London **77**, 184 (1961).
[4] Y.-N. Chiu, J. Math. Phys. **5**, 283 (1964).
[5] R. A. Sack, J. Math. Phys. **5**, 252 (1964).
[6] J. P. Dahl and M. P. Barnett, Mol. Phys. **9**, 175 (1965).
[7] E. O. Steinborn, Chem. Phys. Lett. **3**, 671 (1969).
[8] E. O. Steinborn and K. Ruedenberg, Adv. Quantum Chem. **7**, 1 (1973).
[9] R. A. Sack, SIAM J. Math. Anal. **5**, 774 (1974).
[10] B. R. Judd, *Angular Momentum Theory for Diatomic Molecules* (Academic, New York, 1975), Chap. 5.
[11] R. J. A. Tough and A. J. Stone, J. Phys. A **10**, 1261 (1977).
[12] E. J. Weniger and E. O. Steinborn, J. Math. Phys. **26**, 664 (1985).
[13] B. Friedman and J. Russek, Q. Appl. Math. **20**, 13 (1954).
[14] S. Stein, Q. Appl. Math. **19**, 15 (1961).
[15] O. R. Cruzan, Q. Appl. Math. **20**, 33 (1962).
[16] M. Danos and L. C. Maximon, J. Math. Phys. **6**, 766 (1965).
[17] R. Nozawa, J. Math. Phys. **7**, 1841 (1966).
[18] A. K. Rafiqullah, J. Math. Phys. **12**, 549 (1971).
[19] E. O. Steinborn and E. Filter, Int. J. Quantum Chem. Symp. **9**, 435 (1975).
[20] P. J. A. Buttle and L. J. B. Goldfarb, Nucl. Phys. **78**, 409 (1966).
[21] K. Ruedenberg, Theor. Chim. Acta **7**, 359 (1967).
[22] H. J. Silverstone, J. Chem. Phys. **47**, 537 (1967).
[23] P. O. Löwdin, Adv. Phys. **5**, 96 (1956).
[24] R. R. Sharma, J. Math. Phys. **9**, 505 (1968).
[25] K. J. Duff, Int. J. Quantum Chem. **5**, 111 (1971).
[26] R. R. Sharma, Phys. Rev. A **13**, 517 (1976).
[27] H. J. Silverstone and R. K. Moats, Phys. Rev. A **16**, 1731 (1977).
[28] H. W. Jones and C. A. Weatherford, Int. J. Quantum Chem. Symp. **12**, 483 (1978).
[29] M. A. Rashid, J. Math. Phys. **22**, 271 (1981).
[30] M. A. Rashid, in *ETO Multicenter Molecular Integrals*, edited by C. A. Weatherford and H. W. Jones (Reidel, Dordrecht, 1982), p. 61.
[31] N. Suzuki, J. Math. Phys. **25**, 1133, 3135 (E) (1984).
[32] A. A. Antone, J. Math. Phys. **26**, 940 (1985).
[33] N. Suzuki, J. Math. Phys. **26**, 3193 (1985).
[34] N. Suzuki, J. Math. Phys. **28**, 769 (1987).
[35] M. P. Barnett and C. A. Coulson, Philos. Trans. R. Soc. London Ser. A **243**, 221 (1951).
[36] M. P. Barnett, in *Methods of Computational Physics*, edited by B. Alder, S. Fernbach and M. Rotenberg (Academic, New York, 1963), Vol. 2, p. 95.
[37] F. E. Harris and H. H. Michels, J. Chem. Phys. **43**, S 165 (1965).
[38] F. E. Harris and H. H. Michels, Adv. Chem. Phys. **13**, 205 (1967).
[39] E. O. Steinborn and E. Filter, Theor. Chim. Acta **38**, 273 (1975).
[40] E. Filter and E. O. Steinborn, Phys. Rev. **18**, 1 (1978).
[41] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966). This reference will be denoted as MOS in the text.
[42] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U. P., Cambridge, England, 1970), p. 48.
[43] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981), p. 71, Eq. (3.153).
[44] A. W. Niukkanen, J. Math. Phys. **24**, 1989 (1983).
[45] M. A. Rashid, J. Math. Phys. **27**, 549 (1986).
[46] E. J. Weniger and E. O. Steinborn, J. Math. Phys. **24**, 2553 (1983).

[47]E. J. Weniger and E. O. Steinborn, Comput. Phys. Commun. **25**, 149 (1982).
[48]E. J. Weniger, Ph.D. thesis, Universität Regensburg, 1982. A short abstract of this thesis was published in Zentralbl. Math. **523**, 444 (1984), abstract No. 523-65015.
[49]E. J. Weniger and E. O. Steinborn, J. Chem. Phys. **78**, 6121 (1983).
[50]E. J. Weniger and E. O. Steinborn, Phys. Rev. A **28**, 2026 (1983).
[51]E. J. Weniger, J. Grotendorst, and E. O. Steinborn, Phys. Rev. A **33**, 3688 (1986).
[52]J. Grotendorst, E. J. Weniger, and E. O. Steinborn, Phys. Rev. A **33**, 3706 (1986).
[53]E. Grosswald, *Bessel Polynomials* (Springer, Berlin, 1978), and references therein.
[54]A. W. Niukkanen, J. Math. Phys. **25**, 698 (1984).
[55]A. W. Niukkanen, Int. J. Quantum Chem. **25**, 941 (1984).
[56]E. J. Weniger, J. Math. Phys. **26**, 276 (1985).
[57]E. A. Hylleraas, Z. Phys. **54**, 347 (1929).
[58]H. Shull and P. O. Löwdin, J. Chem. Phys. **23**, 1392 (1955); P. O. Löwdin and H. Shull, Phys. Rev. **101**, 1730 (1956).
[59]E. Filter and E. O. Steinborn, J. Math. Phys. **21**, 2725 (1980).
[60]E. A. Hylleraas, Z. Phys. **48**, 469 (1928).
[61]M. Rotenberg, Adv. At. Mol. Phys. **6**, 233 (1970).
[62]S. L. Sobolev, *Applications of Functional Analysis in Mathematical Physics* (Am. Math. Soc., Providence, RI, 1963).
[63]B. Klahn and W. Bingel, Theor. Chim. Acta **44**, 9 (1977); **44**, 27 (1977).
[64]T.-I. Shibuya and C. Wulfman, Proc. R. Soc. London Ser. A **286**, 376 (1965).
[65]E. D. Rainville, *Special Functions* (Chelsea, New York, 1960). See p. 284, Exercise 4.
[66]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, England, 1944), p. 139.
[67]E. O. Steinborn and E. Filter, Theor. Chim. Acta **52**, 189 (1979).
[68]F. D. Santos, Nucl. Phys. A **212**, 341 (1973), Appendix 1.
[69]B. F. Bayman, J. Math. Phys. **19**, 2558 (1978), Sec. III.
[70]See Ref. 66, p. 80.

On the dynamics of singular, continuous systems

Y. Güler
Physics Department, Middle East Technical University, Ankara, Turkey

(Received 27 July 1988; accepted for publication 30 November 1988)

The Hamilton–Jacobi theory of a special type of singular continuous systems is investigated by the equivalent Lagrangians method. The Hamiltonian is constructed in such a way that the constraint equations are involved in the canonical equations implicitly. The Hamilton–Jacobi partial differential equation is set up in a similar manner to the regular case.

I. INTRODUCTION

Studies on singular systems started around the 1950's. Dirac[1] first investigated the discrete singular systems setting up the basic structure. Bergmann[2] and his collaborators stressed on the relation between invariance principles and constraints in field theories. In fact, their efforts were to construct a Hamiltonian approach of general relativity to quantize the theory since the Einstein's theory of gravitation is a singular theory due to its general covariance. Singular field theories became the center of interest for physicists especially after the pioneering work of Faddeev[3] who introduced the Feynman path integral quantization of singular systems. Nowadays singular systems find a very wide range of applications in theoretical physics. An invariance under a global gauge transformation implies a singular theory. Hence, starting from the electromagnetic theory, all gauge theories have singular nature. String theories, which are hot research subjects of our time, are other typical examples of singular field theories.

The aim of this work is to obtain a valid and consistent Hamilton–Jacobi theory of a special type of singular continuous systems. In fact, this work is a continuation of previous papers[4,5] in which we studied the regular fields and singular discrete systems. The main trend in the treatment of singular systems is, following Dirac, to start with the Lagrangian and then pass to phase-space-defining canonical momenta. Equations of motion are given by Poisson brackets defined in the full phase space. Unfortunately this treatment does not lead to a valid Hamilton–Jacobi theory that is essential for the dynamics of any system. Thus we need the necessity of elaborating singular continuous systems by the equivalent Lagrangians method.

This paper is arranged as follows: In Sec. II the Hamiltonian of a singular continuous system is determined by the method of equivalent Lagrangians, and the first set of the canonical equation is obtained. In Sec. III an alternative approach for singular systems is displayed. In Sec. IV it is proved that the Hesse determinant of the Hamiltonian has the rank $4n-p$. A general discussion is given in Sec. V.

II. EQUIVALENT LAGRANGIANS METHOD

Singular systems are defined as those systems for which the Hesse determinant

$$\left|\frac{\partial^2 L}{\partial \dot q_1\, \partial \dot q_j}\right|, \quad i,j=1,\ldots,n, \tag{2.1}$$

has the rank $n-p$, where $p<n$. The first step in the treatment of singular systems is to demonstrate that any singular system may be considered as a system of p constraints. In other words, the dynamics of singular systems is the dynamics of constrained systems. Let

$$f_i(q_j,\dot q_j,t)=\frac{\partial L}{\partial \dot q_i}. \tag{2.2}$$

Since the rank of the Hesse determinant is $n-p$, p of the functions f_i may be expressed as

$$f_\alpha = K_\alpha(f_a), \quad \alpha=1,\ldots,p, \quad a=p+1,\ldots,n. \tag{2.3}$$

Thus, one may treat p equations (2.3) as the constraint equations, i.e.,

$$G_\alpha(t,q_i,\dot q_i)=\frac{\partial L}{\partial \dot q_\alpha}-K_\alpha\left(\frac{\partial L}{\partial \dot q_a}\right)=0. \tag{2.4}$$

The Hesse determinant of a continuous system[4] is a $4n\times 4n$ determinant formed by the partial derivatives of $L(\Phi_i,\partial_\mu \Phi_i x_\upsilon)$ with respect to

$$\dot\Phi_k\equiv\frac{\partial\Phi_i}{\partial x_0},\quad \dot\Phi_l\equiv\frac{\partial\Phi_i}{\partial x_1},\quad \dot\Phi_m\equiv\frac{\partial\Phi_i}{\partial x_2},$$
$$\dot\Phi_r\equiv\frac{\partial\Phi_i}{\partial x_3},\quad i,k,l,m,r=1,\ldots,n. \tag{2.5}$$

A continuous, singular system has a Lagrangian L such that the rank of determinant

$$\left|\frac{\partial^2 L}{\partial\dot\Phi_\rho\,\partial\dot\Phi_\sigma}\right|,\quad \rho,\sigma=1,\ldots,4n, \tag{2.6}$$

is $4n-p$, where $p<4n$.

The investigation of systems with constraints had been done in a previous work.[5] Another version of the same method will be employed in this work. The starting point of this method is the function

$$M(x_\upsilon,\Phi_i,\dot\Phi_\rho,\lambda_\alpha)=L(x_\upsilon,\Phi_i,\dot\Phi_\rho,)+\lambda_\alpha G_\alpha, \tag{2.7}$$

where i runs from 1 to n and υ from 0 to 3.

In this work we will be interested in cases for which

$$\left|\frac{\partial G_\alpha}{\partial\dot\Phi_\beta}\right|\neq 0,\quad \alpha,\beta=1,\ldots,p. \tag{2.8}$$

This definition reduces the variational problem to an ordinary calculus problem. Equivalence of Lagrangians L' and L where

$$L'=L-\frac{dS_\mu}{dx_\mu} \tag{2.9}$$

implies that the necessary condition to have a local mini-

 0022-2488/89/040785-04$02.50

mum of L' in a certain neighborhood of $\dot{\Phi}_\rho = \eta_\rho(\Phi_j, x_\nu)$ with constraint equations (2.4) leads to the fundamental equations

$$\frac{\partial S_\mu}{\partial \Phi_i} = \frac{\partial M(x_\nu, \Phi_i, \lambda_\alpha, \dot{\Phi}_\sigma = \eta_\sigma)}{\partial \dot{\Phi}_\rho}, \tag{2.10}$$

$$\frac{\partial S_\mu}{\partial x_\mu} = M(x_\nu, \Phi_i, \lambda_\alpha \dot{\Phi}_\sigma = \eta_\sigma) - \eta_\rho \frac{\partial M(x_\nu, \Phi_i, \lambda_\alpha \dot{\Phi}_\sigma = \eta_\sigma)}{\partial \dot{\Phi}_\rho}. \tag{2.11}$$

One should notice that the theory depends on multipliers λ_α. Hence we will treat the problem as if there are $5n+4+p$ variables Φ_i, $\dot{\Phi}_\rho$, x_ν, and λ_α. The existence of constraints actually reduces the number of independent variables to $5n+4$.

To pass to phase space one introduces the canonical momenta as

$$p_\rho \equiv p_{\mu i} = \frac{\partial M}{\partial(\partial \Phi_i / \partial x_\mu)} \equiv \frac{\partial M}{\partial(\dot{\Phi}_\rho)}, \quad \rho = 1,\ldots,4n, \tag{2.12}$$

$$p_\alpha = \frac{\partial M}{\partial \lambda_\alpha} = G_\alpha = 0. \tag{2.13}$$

Since the determinant of the matrix

$$\begin{pmatrix} \dfrac{\partial^2 M}{\partial \dot{\Phi}_\rho \, \partial \dot{\Phi}_\sigma} & \dfrac{\partial G_\alpha}{\partial \dot{\Phi}_\beta} \\ \dfrac{\partial G_\gamma}{\partial \dot{\Phi}_\delta} & 0 \end{pmatrix}, \quad \partial, \delta = 1,\ldots,p \tag{2.14}$$

is not zero one may solve Eqs. (2.12) and (2.13) for $\dot{\Phi}_\rho$ and λ_α as

$$\dot{\Phi}_\rho = \varphi_\rho(\Phi_i, p_\sigma, p_\alpha, x_\nu), \quad \lambda_\alpha = \chi_\alpha(\Phi_i, p_\sigma, p_\alpha, x_\nu). \tag{2.15}$$

In this notation the fundamental equations read as

$$P_\rho = \frac{\partial S_\mu}{\partial \Phi_i}, \tag{2.16}$$

$$\frac{\partial S_\mu}{\partial x_\mu} = M - p_\rho \psi_\rho. \tag{2.17}$$

Defining the Hamiltonian as

$$H_1(x_\nu, \Phi_i, p_\rho, p_\alpha) = -M(x_\nu, \dot{\Phi}_\rho = \varphi_\rho, \lambda_\alpha = \chi_\alpha) + p_\rho \varphi_\rho + p_\alpha \chi_\alpha, \tag{2.18}$$

some partial derivatives of H_1 may be evaluated as

$$\frac{\partial H_1}{\partial x_\nu} = -\frac{\partial M}{\partial x_\nu} - \frac{\partial M}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\rho}{\partial x_\nu} - \frac{\partial M}{\partial \lambda_\alpha} \frac{\partial \chi_\alpha}{\partial x_\nu} + p_\rho \frac{\partial \varphi_\rho}{\partial x_\nu} + p_\alpha \frac{\partial \chi_\alpha}{\partial x_\nu} = -\frac{\partial M}{\partial x_\nu}, \tag{2.19}$$

$$\frac{\partial H_1}{\partial \Phi_i} = -\frac{\partial M}{\partial \Phi_i} - \frac{\partial M}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\rho}{\partial \Phi_i} - \frac{\partial M}{\Phi \lambda_\alpha} \frac{\partial \chi_\alpha}{\partial \Phi_i} + p_\rho \frac{\partial \varphi_\rho}{\partial \Phi_i} + p_\alpha \frac{\partial \chi_\alpha}{\partial \Phi_i} = -\frac{\partial M}{\partial \Phi_i}, \tag{2.20}$$

$$\frac{\partial H_1}{\partial p_\rho} = -\frac{\partial M}{\partial \dot{\Phi}_\sigma} \frac{\partial \varphi_\sigma}{p_\rho} - \frac{\partial M}{\partial \lambda_\alpha} \frac{\partial \chi_\alpha}{\partial p_\rho} + \Phi_\rho p_\sigma \frac{\partial \varphi_\sigma}{\partial p_\rho} + p_\alpha \frac{\partial \chi_\alpha}{\partial p_\rho} = \dot{\Phi}_\rho. \tag{2.21}$$

Actually (p) equations (1.13) reduce the Hamiltonian H_1 to

$$H(x_\nu, \Phi_i, p_\rho) \equiv H_1(x_\nu, \Phi_i, p_\rho, p_\alpha = 0) = -M + p_\rho \varphi_\rho \tag{2.22}$$

A similar calculation shows that Eqs. (2.19)–(2.21) are valid for H also, i.e.,

$$\frac{\partial H}{\partial x_\nu} = -\frac{\partial M}{\partial x_\nu}, \quad \frac{\partial H}{\partial \Phi_i} = -\frac{\partial M}{\partial \Phi_i}, \quad \frac{\partial H}{\partial p_\rho} = \dot{\Phi}_\rho. \tag{2.23}$$

There is still a third way to define a Hamiltonian H that is completely equivalent to (2.18). It is obtained by replacing the expression (1.7) in (2.18). The explicit form is

$$H(x_\nu, \Phi_i, p_\rho) = -L(x_\nu, \Phi_i, \dot{\Phi}_\rho = \varphi_\rho) + p_\rho \Phi_\rho. \tag{2.24}$$

Although H is similar to the Hamiltonian of a regular system one should keep in mind that the generalized momenta p_ρ are defined by the function M not by L.

The definitions (2.18) and (2.24) are completely equivalent dynamically. In other words both definitions yield to the same partial derivatives (2.19)–(2.21). In fact,

$$\frac{\partial H_1}{\partial x_\nu} = \frac{\partial L}{\partial x_\nu} - \frac{\partial L}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\rho}{\partial x_\nu} - \lambda_\alpha \times \left(\frac{\partial G_\alpha}{\partial x_\nu} + \frac{\partial G_\alpha}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\rho}{\partial x_\nu} \right) + p_\rho \frac{\partial \varphi_\rho}{\partial x_\nu}, \tag{2.25}$$

$$\frac{\partial H_1}{\partial \Phi_i} = \frac{\partial L}{\partial \Phi_i} - \frac{\partial L}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\rho}{\partial \Phi_i} - \lambda_\alpha \times \left(\frac{\partial G_\alpha}{\partial \Phi_i} + \frac{\partial G_\alpha}{\partial \dot{\Phi}_\rho} \frac{\partial \varphi_\alpha}{\partial \Phi_i} \right) + p_\rho \frac{\partial \varphi_\rho}{\partial \Phi_i}, \tag{2.26}$$

$$\frac{\partial H}{\partial p_\rho} = -\frac{\partial L}{\partial \dot{\Phi}_\sigma} \frac{\partial \varphi_\sigma}{\partial p_\rho} + \Phi_\rho + p_\sigma \frac{\partial \varphi_\sigma}{\partial p_\rho} - \lambda_\alpha \times \left(\frac{\partial G_\alpha}{\partial \dot{\Phi}_\sigma} \frac{\partial \varphi_\sigma}{\partial p_\rho} \right). \tag{2.27}$$

Due to the constraint equations

$$G_\alpha(x_e, \Phi_i, \dot{\Phi}_\rho = \varphi_\rho) = 0, \tag{2.28}$$

terms in the parentheses are zero, thus

$$\frac{\partial H_1}{\partial x_\nu} = \frac{\partial H}{\partial x_\nu} = -\frac{\partial M}{\partial x_\nu}, \tag{2.29}$$

$$\frac{\partial H_1}{\partial \Phi_i} = \frac{\partial H}{\partial \Phi_i} = -\frac{\partial M}{\partial \Phi_i}, \tag{2.30}$$

$$\frac{\partial H_1}{\partial p_\rho} = \varphi_\rho = \frac{\partial H}{\partial p_\rho}. \tag{2.31}$$

III. AN ALTERNATIVE APPROACH FOR SINGULAR SYSTEMS

The condition (2.8) implies that the constraint equations

$$G_\alpha(x_\nu, \Phi_i, \dot{\Phi}_\rho) = 0 \tag{3.1}$$

may be solved for $\dot{\Phi}_\alpha$ in terms of x_υ,Φ_i, and $\dot{\Phi}_a$ as

$$\dot{\Phi}_\alpha = \Psi_\alpha(x_\upsilon,\Phi_i,\dot{\Phi}_a), \quad a = p+1,\ldots,4n. \tag{3.2}$$

Thus instead of the Lagrangian $L(x_\upsilon,\Phi_i,\dot{\Phi}_\rho)$ one may as well start with the Lagrangian

$$\overline{L}(x_\upsilon,\Phi_i,\dot{\Phi}_a) = L(x_\upsilon,\Phi_i,\dot{\Phi}_a,\dot{\Phi}_\alpha = \Psi_\alpha). \tag{3.3}$$

It is obvious that the constraint equations are introduced in $\overline{L}$ implicity. In this approach the system defined by the Lagrangian $L(x_\upsilon,\Phi_i,\dot{\Phi}_\rho)$ and (p) constraint equations (3.1) is considered as a system described by the regular Lagrangian $\overline{L}$ that has nonvanishing Hesse determinant.

To pass the phase space one should be able to express $\dot{\Phi}_\rho$ in terms of x_υ,Φ_i, and p_ρ, where p_ρ's are defined as (1.12)

$$p_\alpha = \frac{\partial M}{\partial \dot{\Phi}_\alpha} = \frac{\partial L}{\partial \dot{\Phi}_\alpha} + \lambda_\beta \frac{\partial G_\beta}{\partial \dot{\Phi}_\alpha}, \tag{3.4}$$

$$p_a = \frac{\partial M}{\partial \dot{\Phi}_a} = \frac{\partial L}{\partial \dot{\Phi}_a} + \lambda_\beta \frac{\partial G_\beta}{\partial \dot{\Phi}_a}. \tag{3.5}$$

Contracting (3.4) with $\partial\Psi_\alpha/\partial\dot{\Phi}_a$ and adding it to (3.5) one obtains

$$p_a = \left(\frac{\partial L}{\partial \dot{\Phi}_a} + \frac{\partial L}{\partial \dot{\Phi}_\alpha}\frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}\right) + \lambda_\beta \times\left(\frac{\partial G_\beta}{\partial \dot{\Phi}_a} + \frac{\partial G_\beta}{\partial \dot{\Phi}_\alpha}\frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}\right) - p_\alpha \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}. \tag{3.6}$$

The identities

$$G_\alpha(x_\upsilon,\Phi_i,\dot{\Phi}_\alpha = \Psi_\alpha,\dot{\Phi}_a) = 0 \tag{3.7}$$

make the second parentheses equal to zero. So, one gets $4n-p$ equations

$$p_a = \frac{\partial \overline{L}}{\partial \dot{\Phi}_a} - p_\alpha \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}, \tag{3.8}$$

which are independent from λ_α. Solving them for $\dot{\Phi}_a$'s as

$$\dot{\Phi}_a = \varphi_a(x_\upsilon,\Phi_i,p_\rho), \tag{3.9}$$

we have achieved to express $\dot{\Phi}_a$ in terms of p_ρ. Besides, these expressions make it possible to write $\dot{\Phi}_\alpha$'s in terms of p_ρ as

$$\dot{\Phi}_\alpha = \varphi_\alpha(x_\upsilon,\Phi_i,\dot{\Phi}_a = \varphi_a) = \varphi_\alpha(x_\upsilon,\Phi_i,p_\rho). \tag{3.10}$$

In this approach the definition of the Hamiltonian reads as

$$H(x_\upsilon,\Phi_i,p_\rho) = -\overline{L}(x_\upsilon,\Phi_i,\dot{\Phi}_a = \varphi_a) + p_a\varphi_a + p_\alpha\varphi_\alpha. \tag{3.11}$$

This definition again leads us to the first set of canonical equations. In fact,

$$\frac{\partial H}{\partial p_a} = -\frac{\partial \overline{L}}{\partial \dot{\Phi}_b}\frac{\partial \varphi_b}{\partial p_a} + p_a + p_\alpha \frac{\partial \varphi_\alpha}{\partial p_a} + p_b \frac{\partial \varphi_b}{\partial p_a}. \tag{3.12}$$

Making use of (3.8) one expresses $\partial\overline{L}/\partial\dot{\Phi}_b$ as

$$\frac{\partial \overline{L}}{\partial \dot{\Phi}_b} = p_b + p_\alpha \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_b}. \tag{3.13}$$

Hence (3.12) takes the form

$$\frac{\partial H}{\partial p_a} = \varphi_a = \dot{\Phi}_a. \tag{3.14}$$

In the same manner, one may show that

$$\frac{\partial H}{\partial p_\alpha} = \varphi_\alpha = \dot{\Phi}_\alpha. \tag{3.15}$$

Now to check the equivalence of the Lagrangians L and $\overline{L}$ one should demonstrate that the Hamiltonian obtained from $\overline{L}$ with p constraint equations

$$\overline{\Gamma}_\alpha = \dot{\Phi}_\alpha - \Psi_\alpha = 0 \tag{3.16}$$

coincides with the Hamiltonian (3.11). This coincidence will be exhibited in the following section.

As in the previous case one should start with the function

$$\overline{M}(x_\upsilon,\Phi_i,\dot{\Phi}_\rho,\overline{\lambda}_\alpha) = \overline{L}(x_\upsilon,\Phi_i,\dot{\Phi}_\rho) + \overline{\lambda}_\alpha\Gamma_\alpha. \tag{3.17}$$

By introducing the canonical momenta as

$$\overline{p}_a = \frac{\partial \overline{M}}{\partial \dot{\Phi}_a} = \frac{\partial \overline{L}}{\partial \dot{\Phi}_a} - \overline{\lambda}_\alpha \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}, \tag{3.18}$$

$$\overline{p}_\alpha = \frac{\partial \overline{M}}{\partial \dot{\Phi}_\alpha} = \overline{\lambda}_\alpha, \tag{3.19}$$

one arrives at the relation

$$\overline{p}_a = \frac{\partial \overline{L}}{\partial \dot{\Phi}_a} - \overline{p}_\alpha \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a}. \tag{3.20}$$

By comparing relations (3.20) and (3.8) one deduces that two definitions of generalized momenta p_ρ and $\overline{p}_\rho$ are the same, i.e.,

$$\overline{p}_a = p_a, \quad \overline{p}_\alpha = p_\alpha. \tag{3.21}$$

Hence one may solve Eqs. (3.20) for $\dot{\Phi}_a$ as

$$\begin{aligned} &\dot{\Phi}_a = \varphi_a(x_\upsilon,\Phi_i,p_\rho),\\ &\dot{\Phi}_\alpha = \Psi_\alpha(x_\upsilon,\Phi_i,\dot{\Phi}_a = \Psi_a) = \varphi_\alpha(x_\upsilon,\Phi_i,p_\rho). \end{aligned} \tag{3.22}$$

The Hamiltonian derived from $\overline{L}$ is

$$\begin{aligned} \overline{H}(x_\upsilon,\Phi_i,\overline{p}_\rho) &= -\overline{L}(x_\upsilon,\Phi_i,\dot{\Phi}_a = \varphi_a)\\ &\quad + \overline{p}_\alpha\Psi_\alpha + p_a\varphi_a\\ &= -\overline{L} + p_a\varphi_a + p_\alpha\varphi_\alpha = H. \end{aligned} \tag{3.23}$$

This treatment reveals the fact that the Hamiltonian of a constrained system is independent from the Lagrangian that we start with. In this sense both Lagrangians L and $\overline{L}$ are completely equivalent.

IV. DETERMINATION OF THE HESSE DETERMINANT OF $H(x_\upsilon,\Phi_i,p_\rho)$

Another point to be specified is that the Hamiltonian of a constrained system with p constraint equations has the rank $4n-p$. To demonstrate this let us start with Eqs. (3.14), which are the first set of canonical equations of motion. Due to the condition

$$\left|\frac{\partial \varphi_a}{\partial p_b}\right| \neq 0 \tag{4.1}$$

the identities

$$\varphi_a = \dot{\Phi}_a = \frac{\partial H}{\partial p_a} \tag{4.2}$$

imply that

$$\left|\frac{\partial^2 H}{\partial p_a \partial p_b}\right| \neq 0. \tag{4.3}$$

Besides

$$\dot{\Phi}_\alpha = \varphi_\alpha = \frac{\partial H}{\partial p_\alpha} = \Psi_\alpha\left(x_v, \Phi_i, \dot{\Phi}_a = \frac{\partial H}{\partial p_a}\right). \tag{4.4}$$

By taking the partial derivatives of the both sides of Eq. (4.4) with respect to p_ρ one obtains

$$\frac{\partial^2 H}{\partial p_\alpha \partial p_\rho} = \frac{\partial \Psi_\alpha}{\partial \dot{\Phi}_a} \frac{\partial^2 H}{\partial p_a \partial p_\rho}, \tag{4.5}$$

which implies that the rank of the determinant

$$\left|\frac{\partial^2 H}{\partial p_\rho\, \partial p_\sigma}\right| \tag{4.6}$$

is at most $4n - p$. Due to the condition (4.3) we deduce that the rank of the Hesse determinant is actually $4n - p$.

V. CONCLUSION

As it was stated in the Introduction, our main goal is to construct a valid Hamilton–Jacobi theory of singular, continuous systems. To achieve this goal we start with the equivalent Lagrangian L' and the necessary condition to have a local minimum of L' leads us to the fundamental equations (2.10) and (2.11). The phase-space treatment reduces the variational problem to the solution of the Hamilton–Jacobi partial differential equation (HJPDE),

$$\frac{\partial S_\mu}{\partial x_\mu} = -H\left(\Phi_i, x_v, \frac{\partial S_\mu}{\partial \Phi_i}\right). \tag{5.1}$$

We stress on the fact that the definition of the Hamiltonian is independent of the Lagrangian that we start with. In other words, both Lagrangians L and $\overline{L}$ are dynamically equivalent. The first set of the canonical equations (3.14) and (3.15) directly follow from the definition of the Hamiltonian just like in the regular case. One point that should be clarified is that the constraint equations $G_\alpha = 0$ are implicitly included in (3.14) and (3.15). Thus the rest of the theory is completely the same as in the regular case. If S are any set of solutions of the HJPDE then the solutions of Eqs. (3.14) and (3.15) are the extremals of the action. This formulation leads us to the second set of the canonical equations

$$\frac{\partial p_{\mu i}}{\partial x_\mu} = \frac{\partial H}{\partial \Phi_i} \quad \forall\, \mu, i \tag{5.2}$$

as in the regular case. The only difference between the regular and the singular cases is in the definition of the generalized momenta. In the singular case they are defined as

$$p_\rho = \frac{\partial M}{\partial \dot{\Phi}_{\ddot{\rho}}} \tag{5.3}$$

contrary to the usual definition

$$p_\rho = \frac{\partial L}{\partial \dot{\Phi}_\rho}. \tag{5.4}$$

Of course one should keep in mind that all of these calculations are based on the condition (2.8). The most general case will be studied in a subsequent paper.

[1] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964).

[2] P. G. Bergmann, Phys. Rev. **75**, 680 (1949).

[3] L. D. Faddeev, Theor. Math. Phys. **1**, 1 (1970).

[4] Y. Güler, "Hamilton–Jacobi theory of continuous systems," Nuovo Cimento B **100**, 251 (1987).

[5] Y. Güler, "Hamilton–Jacobi theory of discrete, regular constrained systems," Nuovo Cimento B **100**, 267 (1987).

Positive (2+1)-dimensional exact shock waves solutions to the Broadwell model

Henri Cornille
Service de Physique Théorique de Saclay, Laboratoire de L'Institut de Recherche Fondamentale du Commissariat à l'Energie Atomique, 91191 Gif-sur-Yvette Cedex, France

(Received 12 September 1988; accepted for publication 7 December 1988)

The exact solutions for the discrete Boltzmann models are sums of similarity shock waves: at least three for the (2 + 1)-dimensional solutions (two spatial coordinates). Here, for the three-dimensional, six velocities, Broadwell model shock waves solutions are also constructed. The difficult problem is the determination of solutions with positive density. It is proven that in the arbitrary parameters space, from which the solutions are constructed, there exists a domain leading to positivity. As an illustration, a numerical example is discussed. In the spatial coordinate plane both shock waves along one axis and bumps along the other are observed. The movement of these structures with time is discussed.

I. INTRODUCTION

For the discrete Boltzmann models,[1] the velocity $\mathbf{V}$ can only take discrete values $\mathbf{v}_i$, $|v_i| = 1$, $i = 1,...,2p$ with p couples of opposite velocities $\mathbf{v}_{2i} + \mathbf{v}_{2i} = 0$, $i = 1,...,p$. To each velocity $\mathbf{v}_i$ is associated a density $N_i(\mathbf{x},t)$ ($\mathbf{x}$ spatial coordinates $x_1,x_2,...,x_q$).

The simplest solutions are the similarity shock waves solutions

$$N_i = n_{0i} + n_i/(1 + d\exp(\rho t + \boldsymbol{\gamma}\cdot\mathbf{x})), \quad d>0, \tag{1.1}$$

with $\boldsymbol{\gamma}\cdot\mathbf{x} = \Sigma\gamma_q x_q$, while exact multidimensional solutions have been found simply as sums of such solutions

$$N_i = n_{0i} + \sum_{j=1}^{j_{\max}} \frac{n_{ji}}{D_j}, \quad D_j = 1 + d_j \exp(\rho_j t + \boldsymbol{\gamma}_j\cdot\mathbf{x}), \tag{1.2}$$

with $j_{\max} = 3$ for the (2 + 1)-dimensional solutions (two spatial coordinates).

Three classes of exact (2 + 1)-dimensional solutions are known: (i) a solution relaxing toward nonuniform Maxwellians,[2] (ii) semiperiodic solutions,[2] (iii) shock waves.[3] Three models have been investigated: the $4\mathbf{v}_i$ planar model,[1] the $6\mathbf{v}_i$ three-dimensional Broadwell model,[1] and as a generalization, in a p-dimensional space, an hypercubic model.[2]

Unfortunately, shock wave solutions, which physically are the most interesting ones, have not been obtained for the Broadwell model. This model is the most popular one and the most studied. Multidimensional solutions are more difficult to handle for the Broadwell model than for the $4\mathbf{v}_i$ one (for which shock wave solutions are known,[3]) the main reason being that the Broadwell model has two independent quadratic collisions terms. With two collision terms we double the number of constraints and it is not clear that we still have sufficient freedom for the construction of solutions. Another difficulty, present for all multidimensional solutions, is the necessity to build up positive densities. When the number of densities increases (four for the $4v_i$ model and six for the Broadwell one) then the number of constraints for positivity increases too.

Our goal is not to construct the most general shock wave solutions but only to prove that such solutions exist. Due to the two abovementioned difficulties, we simplify, as much as possible, our class of solutions.

For the Broadwell model, the three velocities $\mathbf{v}_{2i-1}$, $i = 1,2,3$ are along the three $0\mathbf{x}_i$ axis and the six densities satisfy the equations

$$\begin{aligned}
&N_{1t} + N_{1x_1} \\
&\quad = N_{2t} - N_{2x_1} = \mathrm{Col}_1 = N_3N_4 + N_5N_6 - 2N_1N_2, \\
&N_{3t} + N_{3x_2} \\
&\quad = N_{4t} - N_{4x_2} = \mathrm{Col}_3 = N_1N_2 + N_5N_6 - 2N_3N_4, \\
&N_{5t} + N_{5x_3} \\
&\quad = N_{6t} - N_{6x_3} = -N_{1t} - N_{3t} - N_{1x_1} - N_{3x_2}.
\end{aligned} \tag{1.3}$$

Here, we construct a simple class of solutions with four independent densities: $N_6 = N_4$, $N_5 = N_3$, and

$$\begin{aligned}
&\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} n_{01} \\ n_{02} \end{pmatrix} + \begin{pmatrix} n_{11} & n_{12} \\ n_{12} & n_{11} \end{pmatrix}\begin{pmatrix} 1/D_1 \\ 1/D_2 \end{pmatrix} + \frac{1}{D_3}\begin{pmatrix} n_{31} \\ n_{32} \end{pmatrix}, \\
&\begin{pmatrix} N_3 \\ N_4 \end{pmatrix} = \begin{pmatrix} n_{03} \\ n_{04} \end{pmatrix} + \begin{pmatrix} n_{13} & n_{14} \\ n_{14} & n_{13} \end{pmatrix}\begin{pmatrix} 1/D_1 \\ 1/D_2 \end{pmatrix} + \frac{1}{D_3}\begin{pmatrix} n_{33} \\ n_{34} \end{pmatrix}, \\
&\boldsymbol{\gamma}_1\cdot\mathbf{x} = \gamma_1 x_1 + \gamma_2(x_2 + x_3) = y, \\
&\boldsymbol{\gamma}_2\cdot\mathbf{x} = -y, \quad \boldsymbol{\gamma}_3 x = \gamma_{31}x_1 + \gamma_{32}(x_2 + x_3).
\end{aligned} \tag{1.4}$$

We note that in the three spatial coordinates, the solutions depend only on two coordinates x_1 and $x_2 + x_3$ so that we can discuss as well the properties in an (x,y) coordinate plane ($\gamma_1\gamma_{32} \neq \gamma_2\gamma_{31}$).

The positivity study is simple in a one-dimensional coordinate space. If the two asymptotic shock limits along the x (or y) axis are positive, then we can manage the $d_j > 0$ parameters in D_j so that positivity is preserved along the whole x (or y) axis. In a two-dimensional coordinate space we must control the positivity in all the asymptotic x,y directions of the plane. However a generalization of the one-dimensional result remains. The asymptotic shock limits become plateaus in the plane, and it was proved (Appendix A of Ref. 3) that, if these limits are positive, we can still choose d_j such that positivity be preserved in the whole x,y plane.

For each density N_i of the (1.4) type, there exist four asymptotic plateaus that must be positive,

$$\Sigma_i^{2j} = n_{0i} + n_{ji} > 0,$$
$$\Sigma_i^{3j} = \Sigma_i^{2j} + n_{3i} > 0, \quad i = 1,...,4 \quad j = 1,2. \tag{1.5}$$

The principle of the positivity proof is then simple. We must express the 16 Σ_i in terms of the arbitrary parameters that built-up the solutions and prove that there exists a positivity domain in the arbitrary parameters space. For (1.4) solutions we put $n_{01} = 1$, the two arbitrary parameters being

$$(s = n_{12}/n_{11}, n_{02}) \tag{1.6}$$

and we restrict our positivity study to $s \in [0.9,1]$.

In Sec. II we contruct a class of positive solutions of the (1–4) type (all the cumbersome details are written down in the Appendix). The main ingredient that allows an analytical proof (already present for the $4\mathbf{v}_i$ model Ref. 3) is the following: all asymptotic plateaus are factorized out as a product of two first-order n_{03} factors of the type

$$n_{03}\Sigma_i = \Omega_i(s)(n_{03} - B_m(s)n_{01})(n_{03} - n_{02}B_{m'}(s)). \tag{1.7}$$

Thus, to verify the positivity of Σ_i, it is sufficient to check the positivity of each factor. The n_{03}^2 coefficients are s dependent while the roots are n_{0i} $i = 1,2$ multiplied by s-dependent factors. In principle we have 32 such B_m, however, invariance properties allow us to calculate only six of them and to deduce the others. First for each Σ_i^{kj}, i,j,k, fixed we determine the n_{03} interval in which the Σ_i^{kj} are positive and we study the intersections of these 16 n_{03} intervals. Second n_{03} being constructed from (n_{02},s) the final result (Theorem 4) is of the following type: if $s \in (0.9,1)$ and n_{02} has well-defined s-dependent lower and upper bounds then the 16 Σ_i^{kj} are positive [see Fig. 1 for the (s,n_{0i}) positivity domain].

As an illustration, in Sec. II, we study a numerical example, with physical structures different from those of the examples studied in the $4\mathbf{v}_i$ model.[3] Similarity shock waves are, in coordinate space, like kinks. For the solutions (1.4) or those of Ref. 3, the first two components depend on only one spatial coordinate $D_j = 1 + d_j \exp \tau_j y$ (at $t = 0$). Here $\tau_1 = -\tau_2 = 1$ while in Ref. 3 the two τ_j have the same sign. These solutions are sums of two kinks in the y variable and one (third component) in the x variable. If, like here, the two τ_j are opposite, the y dependence is no longer like a shock wave sum of two kinks but like a sum of a kink and an antikink. So for the present solutions (1.4) we observe a shock wave along the x axis and a bump along the y axis. Deformations of these initial time structures, when the time is growing, are observed: the shock part moves in the coordinate space with its shape practically unchanged, while the bump spreads out.

II. CONSTRUCTION OF POSITIVE SOLUTIONS

All details are given in the Appendix and we briefly report the main results. For (1.4) type of solutions, due to $\mathrm{Col}_1 + 2\mathrm{Col}_3 = 0$ and $N_i(x_1, x_2 + x_3, t)$, simplifications occur in the Broadwell model quantities (A1). We construct the (2 + 1)-dimensional solutions in two successive stages: first (1 + 1)-dimensional solutions as sums $n_{0i} + \Sigma n_{ji}/D_j$, $j = 1,2$ of the two first components and second we add the third component n_{3i}/D_3.

A. Algebraic construction of the two first components

There exist eight relations [written down in (A2)] for 11 parameters $n_{0i}, n_{1i}, \rho, \gamma_j$, leaving three arbitrary ones. We choose

$$n_{01} = 1, \quad s = n_{12}/n_{11}, \; n_{02} \tag{2.1}$$

as the arbitrary parameters and construct all others in successive steps.

(i) First we define s-dependent intermediate parameters $\bar{n}_{1i} = n_{1i}/n_{11}$, $\mathscr{S} = \bar{n}_{13} + \bar{n}_{14}$, for $i = 3,4$,

$$2\mathscr{S}(1+s) = -s + \sqrt{\delta}, \quad \delta = s^2 + 4(1+s^2)(1+s)^2,$$
$$2\bar{n}_{13} = \mathscr{S} + \sqrt{1 + s^2 + \mathscr{S}s/(1+s)}, \quad \bar{n}_{14} = \mathscr{S} - \bar{n}_{13}. \tag{2.2}$$

(ii) Second $n_{04} = n_{01}n_{02}/n_{03}$ while n_{03} and n_{11} can be obtained from the arbitrary parameters

$$2n_{03} = -\mu + \sqrt{\mu^2 + 4n_{02}n_{01}},$$
$$\mu(\bar{n}_{14} - \bar{n}_{13}) = (n_{01} - n_{02})(1 - y),$$
$$n_{11}A_1 = n_{01}s + n_{02} - n_{03}\bar{n}_{14} - n_{04}\bar{n}_{13},$$
$$A_1 = \bar{n}_{13}\bar{n}_{14} - s. \tag{2.3}$$

(iii) Third, with the help of (2.2) and (2.3), we construct all n_{1i} parameters from (2.1): $n_{12} = sn_{11}$, $n_{1i} = \bar{n}_{1i}n_{11}$, $i = 3,4$, and the frequency ρ and wave vector γ_1, γ_2 from the n_{1i} [see (A3)].

B. Algebraic construction of the third component

There exist seven parameters $n_{3i}, \rho_3, \gamma_{3j}$ and seven relations (A7). This means that we must build up these parameters from the arbitrary ones s, n_{02}.

(i) First we define s-dependent intermediate parameters $z = n_{32}/n_{31}$, $\bar{n}_{3i} = n_{3i}/n_{31}$, $i = 3,4$, $S = \bar{n}_{33} + \bar{n}_{34}$,

$$2z = -\nu + \sqrt{\nu^2 - 4},$$
$$(\nu - 1/s)(1 + s + \mathscr{S}/2)$$
$$= (1+s)(s + 1/2) + 0.5\mathscr{S}(s + 2 + 3s/(1+s^2)),$$
$$S = (1+y)(1+z)/\mathscr{S},$$
$$2\bar{n}_{33} = S + \sqrt{S^2 + 2Sz/(1+z)}, \quad \bar{n}_{34} = S - \bar{n}_{33}. \tag{2.4}$$

(ii) Second we obtain n_{31} as a function of s, n_{02},

$$n_{31}A_3 = n_{01}z + n_{02} - n_{02}\bar{n}_{34} - n_{04}\bar{n}_{33}, \quad A_3 = \bar{n}_{33}\bar{n}_{34} - z. \tag{2.5}$$

(iii) Third, with the help of (2.4) and (2.5) we construct all n_{3i} parameters from (2.1): $n_{32} = zn_{31}$, $n_{3i} = \bar{n}_{3i}n_{31}$, $i = 3,4$, and the frequency ρ_3 and wave vector γ_{31}, γ_{32} from the n_{3i} [see (A8)].

C. Shock limits $\Sigma_i^{2j} = n_{0i} + n_{ji}, \Sigma_i^{3j} = \Sigma_i^{2j} + n_{3i}$ (see Tables I and II)

1. There exists an interesting property (already present[3] for the $4\mathbf{v}_i$ model): all the Σ_i (we omit the upperscripts) in the present class can be written in a factorized form

$$n_{03}\Sigma_i = \Omega_i(s)(n_{03} - n_{01}B_k(s))(n_{03} - n_{02}B_{k'}(s)). \tag{2.6}$$

For these quadratic n_{03} polynomials, the coefficients of n_{03}^2 are s dependent while the roots are of the type n_{01} (or n_{02}) multiplied by s-dependent terms. Consequently we can easily study the n_{03} intervals in which $n_{03}\Sigma_i$ is positive. We want to understand this factorization (2.6). We remark that from the n_{11},n_{31} expressions, all n_{ji} and consequently all Σ_i are linear combination of the n_{0i} with s-dependent coefficients,

$$\Sigma_i = \sum_{p=1}^{4} \lambda_p(s) n_{0p} . \tag{2.7a}$$

Recalling that $n_{04} = n_{02}n_{01}/n_{03}$, it turns out that the following identity:

$$\lambda_1\lambda_2 = \lambda_3\lambda_4 \tag{2.7b}$$

is also satisfied in our solutions. Then the structure (2.6) holds for the Σ_i. This factorization holds for the 16 Σ_i and we give four examples in subsection 3 of the Appendix [see (A13)]. Here we report two other examples. We begin with $\Sigma_2^{21} = n_{02} + n_{12} = n_{02} + sn_{11}$ and substitute (2.3) for n_{11},

$$A_1\Sigma_2^{21} = n_{01}s^2 + n_{02}\bar{n}_{13}\bar{n}_{14} - n_{03}\bar{n}_{14}s - n_{04}\bar{n}_{13}s , \tag{2.8a}$$

and note that the identity (2.7b) is trivial. We go on with $\Sigma_2^{31} = \Sigma_2^{21} + n_{32}$ and substitute (2.5),

$$\Sigma_2^{31} = \Sigma_2^{21} + (n_{01}z^2 + n_{02}z - n_{03}\bar{n}_{34}z - n_{04}\bar{n}_{33}z)/A_3 , \tag{2.8b}$$

with Σ_2^{21} in (2.8a). The identity (2.7b) becomes

$$(\bar{n}_{34}A_1z + s\bar{n}_{14}A_3)(s\bar{n}_{13}A_3 + \bar{n}_{33}zA_1)$$
$$= (s^2A_3 + z^2A_1)(\bar{n}_{13}\bar{n}_{14}A_3 + zA_1) ,$$

which is easily verified using the relation (A10) for $s + z$.

2. Other useful relations that simplify the calculations of the Σ_i are provided by invariance properties (yet present in the $4v_i$ model). For the exchange $1\leftrightarrow2$ or $\Sigma_1\leftrightarrow\Sigma_2$ we must perform both $n_{01}\leftrightarrow n_{02}$ and $n_{j1}\leftrightarrow n_{j2}$. For the intermediate parameters this gives

$$(1\leftrightarrow2)\rightarrow(s\leftrightarrow1/s,\bar{n}_{1i}\leftrightarrow\bar{n}_{1i}/s,z\leftrightarrow1/z,\bar{n}_{3i}\leftrightarrow\bar{n}_{3i}/s) . \tag{2.9}$$

As an application for Σ_i^{2j}, $i = 1,2$, we see, for the roots, that $B_1 = \bar{n}_{13}\rightarrow\widetilde{B}_1 = \bar{n}_{13}/s$ and $B_2 = 1/\bar{n}_{14}\rightarrow\widetilde{B}_2 = s/\bar{n}_{14}$. With (2.9) we find $A_1\rightarrow A_1/s^2$, $A_3\rightarrow A_3/z^2$, and for instance for the root $n_{02}B_4$ of Σ_1^{31}

$$B_4 = (A_1 + A_3)/(\bar{n}_{34}A_1 + \bar{n}_{14}A_3)$$
$$\rightarrow\widetilde{B}_4 = (s^2A_3 + z^2A_1)/(\bar{n}_{34}zA_1 + \bar{n}_{14}sA_3) .$$

We can verify that the same transformations (2.9) applied to B_m give $\widetilde{B}_m$, $m = 1,\ldots,6$.

3. A third simplification (common to the $4v_i$ model) comes from the fact that Σ_i for $i = 1,2$ and $i = 3,4$ (k,j fixed) have common roots. For instance, $\Sigma_1^{21} = \Sigma_3^{21}$ if $n_{01} + n_{11} - n_{03} - n_{13} = 0$ or $n_{01} - n_{03} + n_{11}(1 - \bar{n}_{13}) = 0$ or with (2.3) for n_{11}:

$$(n_{01}\bar{n}_{13} - n_{03})(n_{03}(\bar{n}_{14} - s) + n_{02}(\bar{n}_{13} - 1)) = 0 .$$

The first factor gives the common root $n_{01}B_1$.

In conclusion, for the 16 Σ_i, instead of the 32 possible roots, only $n_{01}B_k$, $k = 1,3,5$, and $n_{02}B_{k'}$, $k' = 2,4,6$, are really independent.

D. Signs and bounds for the intermediate parameters $\bar{n}_{1i},\bar{n}_{3i},z$, when s belongs to [0.9,1]

In the sequel we limit s to the interval (0.9,1). In subsection 4 of the Appendix we study the square-roots representations (2.2) for $\mathscr{S},\bar{n}_{13}$ and (2.4) for $z,\bar{n}_{34}$. We find that $\mathscr{S}\bar{n}_{13}, -\bar{n}_{14}$ are positive and increase between their values at $s = 0.9$ and $s = 1$. We prove Theorem 1 with the results: $\bar{n}_{13}>0$, $\bar{n}_{14}<0$, $z<0$, $\bar{n}_{33}>0$, $\bar{n}_{34}>0$, $A_1<0$, $A_3>0$, $\bar{n}_{14} + \bar{n}_{34}>0$, $\bar{n}_{13} - \bar{n}_{33}>0$, and the numerical bounds.

$$1.33\leqslant\bar{n}_{13}\leqslant1.4, \quad -0.21\leqslant\bar{n}_{14}\leqslant-0.2 ,$$
$$0.52<\bar{n}_{33}<0.63, \quad 0.43<\bar{n}_{34}\leqslant0.54 .$$

E. n_{03} intervals for positive Σ_i^{kj}, $s\in$[0.9,1] and $n_{0i}>0$

In subsection 5 of the Appendix, for each of the 16 Σ_i written down like (2.6) in Tables I and II, from the signs of $\Omega_i(s),B_k(s),B_{k'}(s)$ we determine the n_{03} interval for which $\Sigma_i>0$. The intersections of these 16 intervals lead in Theorem 3 to the positivity domain:

$$\mathrm{Sup}(n_{02}/\widetilde{B}_1,n_{01}B_5) < n_{03} < \inf(n_{01}B_1,n_{02}B_4) . \tag{2.10}$$

F. ($n_{02},s\in$[0.9,1]) domain for $\Sigma_i^{kj}>0$

Here, n_{03} is constructed from the (n_{02},s) arbitrary parameters with the representation (2.3): $2n_{03} = -\mu + \sqrt{\mu^2 + 4n_{02}}$, μ being s and n_{02} dependent. Inequalities of the type $n_{03}\lesseqgtr Xn_{02}$, $n_{03}\gtreqless n_{01}\overline{X}$, with X and $\overline{X}$ s = dependent positive functions can be expressed in the (s,n_{02}) variables. This is done in Lemma 12 of subsection 6 of the Appendix, leading, respectively, to $n_{02}\lesseqgtr F(X)$, $n_{02}\gtreqless G(\overline{X})$ with F, G given by

TABLE I. $\overline{\Sigma}_i^{2j} = n_{03}\Sigma_i^{2j}/\Omega_i^{2j}$, $s\in(0.9,1)$, $(\bar{n}_{1i}\rightarrow\bar{n}_{1i}/s)\rightarrow(B_k\rightarrow\widetilde{B}_k)$.

$j = 1$ $\quad \overline{\Sigma}_i^{21}\rightarrow\overline{\Sigma}_i$, $\quad \Omega_i^{21}\rightarrow\Omega_i$

$\overline{\Sigma}_1 = (n_{03} - n_{01}B_1)(n_{03} - n_{02}B_2)$ $\quad \overline{\Sigma}_2 = (n_{03} - n_{01}\widetilde{B}_2)(n_{03} - n_{02}\widetilde{B}_1)$

$\overline{\Sigma}_3 = (n_{03} - n_{01}B_1)(n_{03} - n_{02}\widetilde{B}_1)$ $\quad \overline{\Sigma}_4 = (n_{03} - n_{01}\widetilde{B}_2)(n_{03} - n_{01}B_2)$

$B_1 = \bar{n}_{13}>0$, $\widetilde{B}_1 = B_1/s>0$, $B_2 = 1/\bar{n}_{14}<0$, $\widetilde{B}_2 = sB_2<0$, $A_1 = \bar{n}_{14}\bar{n}_{13} - s<0$, $\Omega_1 = -\bar{n}_{14}/A_1<0$

$\Omega_2 = s\Omega_1<0$, $\Omega_3 = -sA_1>0$, $\Omega_4 = \bar{n}_{14}\Omega_1>0$

$j = 2$ $\quad \overline{\Sigma}_i^{22}\rightarrow\overline{\Sigma}_i$, $\quad \Omega_i^{22}\rightarrow\Omega_i$

$\overline{\Sigma}_1 = (n_{03} - n_{01}/\widetilde{B}_2)(n_{03} - n_{02}/\widetilde{B}_1)$ $\quad \overline{\Sigma}_2 = (n_{03} - n_{01}/B_1)(n_{03} - n_{02}/B_2)$

$\overline{\Sigma}_3 = (n_{03} - n_{01}/\widetilde{B}_2)(n_{03} - n_{02}/B_2)$ $\quad \overline{\Sigma}_4 = (n_{03} - n_{01}/B_1)(n_{03} - n_{02}/\widetilde{B}_1)$

$\Omega_1 = -\bar{n}_{13}s/A_1>0$, $\Omega_2 = s\Omega_1>0$, $\Omega_3 = -s/A_1>0$, $\Omega_4 = n_{13}^{-2}/(-A_1)>0$

TABLE II. $\overline{\Sigma}_i^{3j} = n_{03}(-A_1A_3)\Sigma_i^{3j}/\Omega_i^{3j}$, $s\in(0.9,1)$, $(\bar{n}_{1i}\to\bar{n}_{1i}/s,\ \bar{n}_{3i}\to\bar{n}_{3i}/z)\to(B_k\to\widetilde{B}_k)$.

$j=1 \quad \overline{\Sigma}_i^{31}\to\overline{\Sigma}_i, \quad \Omega_i^{31}\to\Omega_i$

$$\overline{\Sigma}_1 = (n_{03}-n_{01}B_3)(n_{03}-n_{02}B_4) \quad \overline{\Sigma}_2 = (n_{03}-n_{01}\widetilde{B}_4)(n_{03}-n_{02}\widetilde{B}_3)$$

$$\overline{\Sigma}_3 = (n_{03}-n_{01}\widetilde{B}_4)(n_{03}-n_{02}B_4) \quad \overline{\Sigma}_4 = (n_{03}-n_{01}B_3)(n_{03}-n_{02}\widetilde{B}_2)$$

$$z<0,\ \bar{n}_{3i}>0\ i=3,4,\ A_3=\bar{n}_{33}\bar{n}_{34}-z>0,\ -A_1A_3>0,\ \Omega_1=\bar{n}_{34}A_1+\bar{n}_{14}A_3<0,\ \Omega_2=\bar{n}_{34}zA_1+\bar{n}_{14}sA_3>0$$

$$\Omega_3=-3sz/4>0,\ \Omega_4=A_1\bar{n}_{34}^2+A_3\bar{n}_{14}^2<0,\ B_3=\Omega_3/\Omega_1=\Omega_2/\Omega_4<0,\ B_4=(A_1+A_3)/\Omega_1=(\bar{n}_{13}A_3+\bar{n}_{33}A_1)/\Omega_3>0$$

$$\widetilde{B}_3=\Omega_3/\Omega_2=\Omega_1/\Omega_4>0,\ \widetilde{B}_4=(s^2A_3+z^2A_1)/\Omega_2=(\bar{n}_{13}sA_3+\bar{n}_{33}zA_1)\Omega_3>0$$

$j=2 \quad \overline{\Sigma}_i^{32}\to\overline{\Sigma}_i, \quad \Omega_i^{32}\to\Omega_i$

$$\overline{\Sigma}_1 = (n_{03}-n_{01}B_5)(n_{03}-n_{02}B_6) \quad \overline{\Sigma}_2 = (n_{03}-n_{01}\widetilde{B}_6)(n_{03}-n_{02}\widetilde{B}_5)$$

$$\overline{\Sigma}_3 = (n_{03}-n_{01}\widetilde{B}_6)(n_{03}-n_{02}B_6) \quad \overline{\Sigma}_4 = (n_{03}-n_{01}B_5)(n_{03}-n_{02}\widetilde{B}_5)$$

$$\Omega_1=A_1\bar{n}_{34}+s\bar{n}_{13}A_3>0, \quad \Omega_2=A_1z\bar{n}_{34}+A_3\bar{n}_{13}>0, \quad \Omega_3=-3sz/4>0, \quad \Omega_4=A_1\bar{n}_{34}^2+A_3\bar{n}_{13}^2>0$$

$$B_5=\Omega_3/\Omega_1=\Omega_2/\Omega_4>0, \quad \widetilde{B}_5=\Omega_3/\Omega_2=\Omega_1/\Omega_4>0, \quad B_6=(s^2A_3+A_1)/\Omega_1=(\bar{n}_{33}A_1+s\bar{n}_{14}A_3)\Omega_3<0$$

$$\widetilde{B}_6=(z^2A_1+A_3)/\Omega_2=(z\bar{n}_{33}A_1+A_3\bar{n}_{14})/\Omega_3>0, \quad \Omega_3=zA_1+\bar{n}_{13}\bar{n}_{14}A_3=A_1\bar{n}_{33}\bar{n}_{34}+sA_3$$

$$F(X)=\frac{\bar{n}_{13}-\bar{n}_{14}+(1-s)X}{X[(\bar{n}_{13}-\bar{n}_{14})+1-s]},\qquad G(\overline{X})=\frac{\overline{X}^2(\bar{n}_{13}-\bar{n}_{14})-\overline{X}(1-s)}{\bar{n}_{13}-\bar{n}_{14}-\bar{n}\sim X(1-s)}. \tag{2.11}$$

Then the positivity domain (2.9) can be written down in the n_{02}, s plane (see Theorem 3′ in the Appendix) as

$$\mathrm{Sup}(F(B_1),G(B_5))<n_{02}<\inf(G(B_1),F(s/B_1)). \tag{2.10'}$$

It remains to determine the sup of the lower bounds and the inf of the upper bounds. This is done in Lemma 14 and we obtain the final positivity domain (see Theorem 4 in the Appendix) $\Sigma_i>0$ if

$$F(B_4)<n_{02}<G(B_1),\quad s\in[0.9,1],\quad n_{01}=1,$$

$$G(B_1)=\frac{\bar{n}_{13}(s-1+\bar{n}_{13}(\bar{n}_{13}-\bar{n}_{14}))}{s\bar{n}_{13}-\bar{n}_{14}},$$

$$F(B_4)=\frac{\bar{n}_{13}-\bar{n}_{14}+B_4(1-s)}{B_4(B_4(\bar{n}_{13}-\bar{n}_{14})+1-s)},$$

and B_4 is written down in Table II. The corresponding (s,n_{02}) domain is presented in Fig. 1.

III. PHYSICAL DISCUSSION AND NUMERICAL RESULTS

In addition to the N_i, we introduce the total mass $M=\Sigma N_i$, and instead of the original x_1,x_2+x plane, we consider the x,y plane defined in (1.4):

$$M(x,y,t)=m_0+\sum_j\frac{m_j}{D_j},$$

$$m_j=n_{j1}+n_{j2}+2(n_{j3}+n_{j4}),\quad j=1,2,3,$$
$$m_0=n_{01}+n_{02}+2(n_{03}+n_{04}),\quad m_1=m_2=m. \tag{3.1}$$

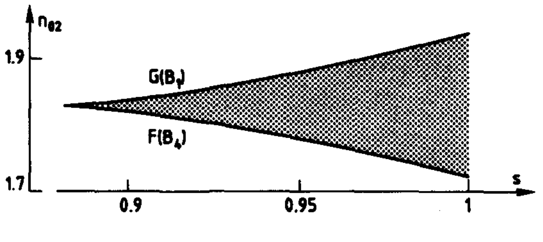


FIG. 1. The (s,n_{02}) domain for positive Σ_i^{kj}.

For simplicity we assume $d_2=d_1=d$. In such a case $D_2=D_1(-y)$, M is even in y while this property does not hold for N_1 and N_2. We introduce $Q=1/D_1+1/D_2$ and rewrite M

$$Q(y,t)=1/(1+\bar{d}e^y)+1/(1+\bar{d}e^{-y}),\quad \bar{d}=d\exp(\rho t),$$
$$M=m_0+mQ(y,t)+m_3/(1+\bar{d}_3e^x),\quad \bar{d}_3=d_3\exp(\rho_3t). \tag{3.2}$$

For t fixed, the equidensity lines $x(y)$, corresponding to $M(x,y,t=\text{const})=\text{const}$ are easily constructed from (3.2),

$$x=-\log\bar{d}_3+\log(m_3/(m_0+mQ-M)-1), \tag{3.3}$$

x being real, the argument of the log must be positive.

We choose an example with the following values for the arbitrary parameters:

$$n_{01}=1,\quad s=0.97,\quad n_{02}=1.916, \tag{3.4a}$$

which satisfy Theorem 4. We deduce for the other parameters

$$n_{1i}=-0.998,-0.968,-1.375,\ 0.2084,$$
$$n_{3i}=0.0818,-0.0288,\ 0.0457,\ 0.0436$$
$$i=1,2,3,4;\ \gamma_1=-0.04,\ \gamma_2=-3,\ \gamma_{31}=0.2,$$
$$\gamma_{32}\equiv0.002,\ m_0=8.454,\ m_1=-4.298,$$
$$m_3=0.2315,\ \rho=2.5,\ \rho_3=-0.09. \tag{3.4b}$$

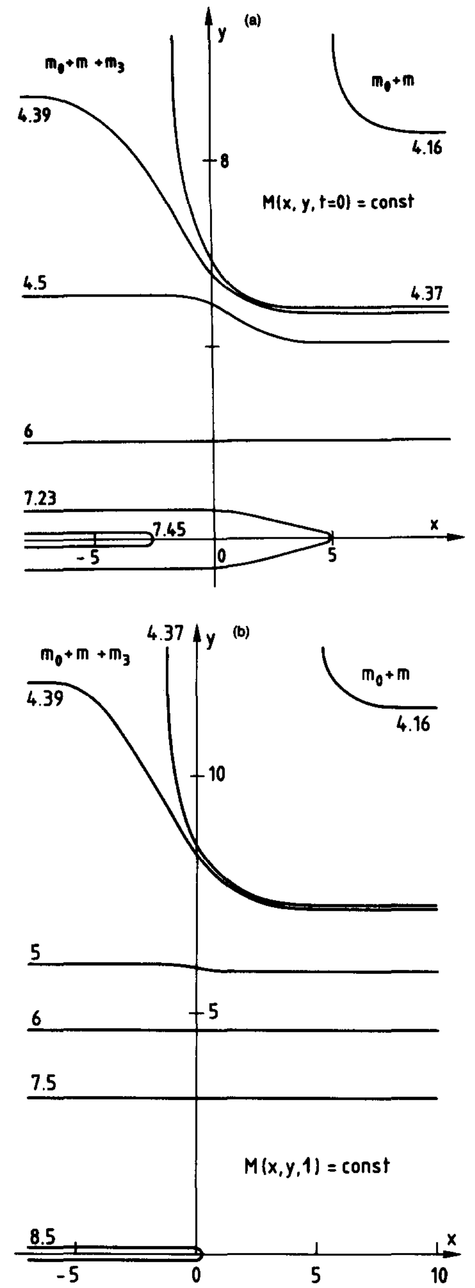
With these values, one can verify that the Σ_i^{kj} are positive. For the d_j we choose

$$d=6,\quad d_3=0.5, \tag{3.4c}$$

M being symmetric with respect to the x axis, we restrict our study to the half-plane $y>0$.

A. Equidensity lines $M(x,y,t=0)=\text{const}$ [Fig. 2(a)]

The asymptotic plateaus (x and y large) are $m_0+m=4.16$ in the right half-plane $x>0$ and



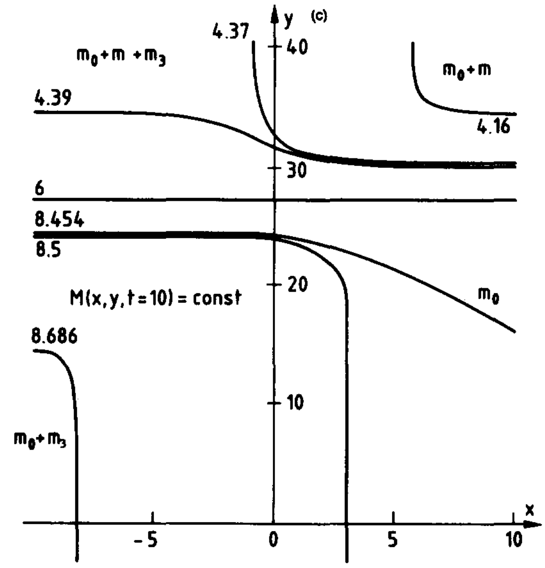


FIG. 2. (a) Equidensity lines at $t = 0$ for the total mass M that is even in the y variable. (b),(c) Equidensity lines at $t = 1$ and 10 for the total mass.

$m_0 + m + m_3 = 4.38$ in the left $x < 0$ one, m_3 being the difference between the two shock limits. Similarly, along any line parallel to the x axis (y fixed), the difference between the two shock limits is $m_3 > 0$. This shock structure is provided by the third similarity component m_3/D_3 with a downstream domain at the right and an upstream one at the left. In this upstream domain the shock limits, when $x \to -\infty$, are $m_0 + mQ(y,0) + m_3(m_0 + mQ(y,0)$ in the downstream domain when $x \to \infty$). If $d > 1$ then $Q < 1$ and, due to $m < 0$, the asymptotic plateaus $m_0 + m + m_3$ provided the highest equidensity lines. If (like in the present numerical example) $d > 1$, then $Q < 1$, the highest values $m_0 + 2m/(1 + d) + m_3$ are obtained along the x axis ($y = 0$). Then $Q(y,0)$ increases with y and the shock limits decrease when $|y|$ is increasing. Consequently, for the profiles parallel to the y axis, we observe a bump in a strip parallel to the x axis. Similarity shock waves behave like kinks. One can say that these two-spatial-dimensional solutions are the superposition of a kink in the x variable (shock structure) and both a kink and an antikink in the y variable (bump) (as we shall see later for the M profiles at fixed y and at fixed x).

B. Movement with t of the equidensity lines $M = \text{const}$ [Figs. 2(b)–(c)]

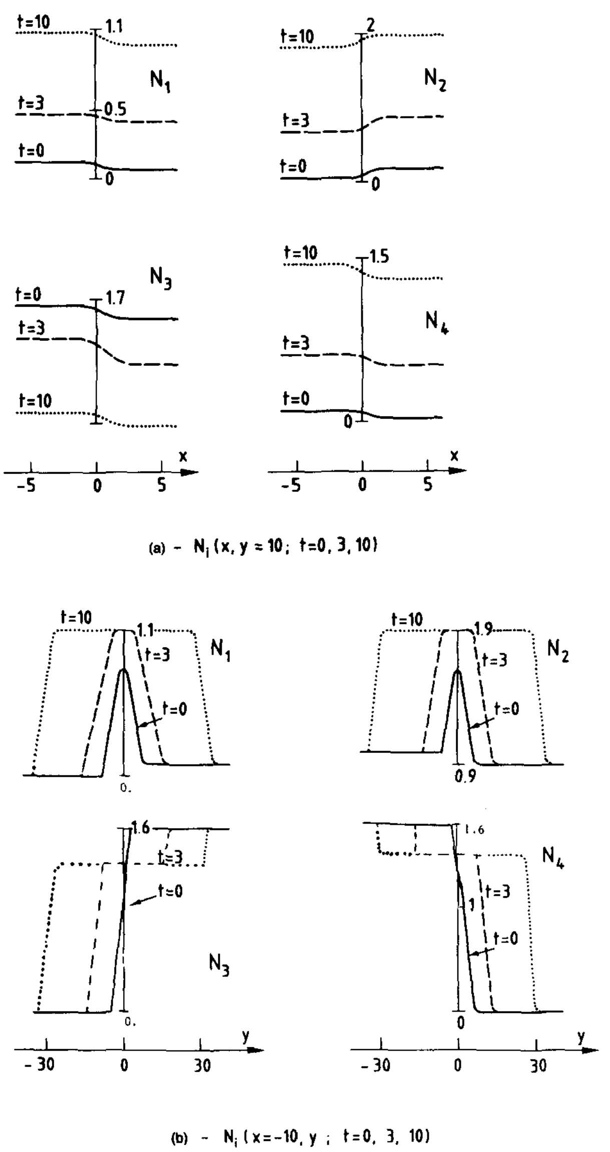
We first discuss the movement of the bump. When t increases, ρ being positive, $\bar{d}$ increases while $|m|/(1 + \bar{d})$ decreases. Due to $m < 0$, the two shock limits ($x \to \pm\infty$), around the x axis, increase. The height of the bump around the x axis increases up to the Maxwellian value $m_0 + m_3$ for $x < 0$ and m_0 for $x > 0$. On the contrary, for profiles parallel to the x axis but with $|y|$ very large, the asymptotic lower plateaus $m_0 + m + m_3$ and $m_0 + m_3$ must reappear. Summarizing, the bump spreads out in a large strip parallel to the

x axis, pushing the asymptotic plateaus to higher and higher $|y|$ values.

Second, we discuss the x-dependent shock due to the third similarity component, ρ_3 being negative, $\bar{d}_3$ varies slowly from its initial 1/2 value up to 1 for infinite time. The shock structure does not change very much, keeping the shift m_3 for the asymptotic values. It is useful to look at a large $t = 10$ fixed value, while (x,y) are increasing. For not too large y values ($|y| < 20$), $M \simeq m_0 + m_3/(1 + e^x)$, the equidensity lines are parallel to the y axis and the two asymptotic limits are either the Maxwellian $m_0 + m_3$ or the other shock limit m_0. On the contrary, for larger y values ($|y| > 20$), the initial time structures reappear, with equidensity lines parallel to the x axis and we observe the two asymptotic plateaus $m_0 + m + m_3, m_0 + m$.

C. Densities N_i and total mass M at fixed y and fixed x

In Fig. 3(a) for N_i and Fig. 3(c) for M we present profiles parallel to the x axis at $y = 10$ fixed. They correspond to the third similarity shock wave structure with a shift of these structures when the time is growing. On the contrary, in Fig. 3(b), for N_i profiles parallel to the y axis with $x = -10$ fixed we observe the two shock limits and in Fig. 3(c) for M the spreading of the bump when the time increases.



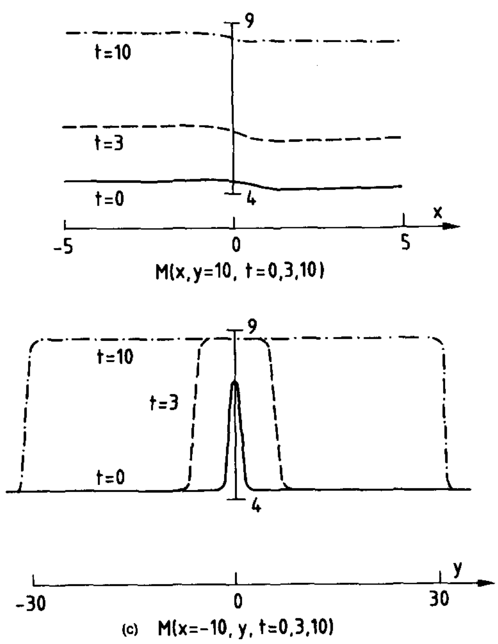


FIG. 3. Densities N_i and total mass M at fixed $y = 10$ and at fixed $x = -10$.

APPENDIX: CONSTRUCTION OF POSITIVE SHOCK WAVES SOLUTIONS FOR THE BROADWELL MODEL

For the $N_i(x_1, x_2 + x_3, t)$ solutions with $N_6 = N_4, N_5 = N_3$, due to $\mathrm{Col}_1 + 2\mathrm{Col}_3 = 0$ and $N_{ix_2} = N_{ix_3}$, simplifications occur for the Broadwell model:

$$\begin{aligned}(N_{1t} + N_{1x_1})/2 &= (N_{2t} - N_{2x_1})/2\\ &= -N_{3t} - N_{3x_2}\\ &= -N_{4t} + N_{4x_2} = N_3N_4 - N_1N_2\,. \end{aligned} \tag{A1}$$

We build up first the sums of the two first components $n_{0i} + \Sigma n_{ji}/D_j, j = 1,2$ and add the third one n_{3i}/D_3. We substitute the ansatz (1.4) into (A1).

1. Relations and construction for the two first components

We find the relations

$$\begin{aligned}n_{11}(\rho + \gamma_1)/2 &= n_{12}(\rho - \gamma_1)/2 = -n_{13}(\rho + \gamma_2)\\ &= n_{14}(\gamma_2 - \rho) = n_{13}n_{14} - n_{11}n_{12}\end{aligned}$$

$$= n_{01}n_{12} + n_{02}n_{11} - n_{03}n_{14} - n_{04}n_{13}, \quad \text{(A2a)}$$

$$n_{03}n_{04} = n_{01}n_{02}, \quad n_{11}^2 + n_{12}^2 = n_{13}^2 + n_{14}^2,$$
$$(n_{12} - n_{11})(n_{01} - n_{02}) = (n_{14} - n_{13})(n_{03} - n_{04}). \quad \text{(A2b)}$$

We have 11 parameters $n_{0i}, n_{1i}, \rho, \gamma_j$, eight relations leaving one scaling fixed parameter and two arbitrary ones chosen to be:

$$(n_{01} = 1, \quad s = n_{12}/n_{11}, n_{02}).$$

We construct in successive stages all parameters from s and n_{02}: (i) First, ρ, γ_j can be deduced from the n_{1i} alone,

$$\begin{aligned} &\rho n_{11}n_{12} = (n_{13}n_{14} - n_{11}n_{12})(n_{11} + n_{12}), \\ &\gamma_1(n_{11} + n_{12}) = \rho(n_{12} - n_{11}), \\ &\gamma_2(n_{13} + n_{14}) = \rho(n_{14} - n_{13}) \end{aligned} \quad \text{(A3)}$$

from which we deduce a second relation between the n_{1i}: $n_{11}n_{12}(n_{13} + n_{14}) + 2n_{13}n_{14}(n_{11} + n_{12}) = 0$. (ii) Second, we define intermediate parameters

$$\bar{n}_{1i} = n_{1i}/n_{11}, \quad i = 3,4, \quad \mathscr{P} = \bar{n}_{13}\bar{n}_{14}, \quad \mathscr{S} = \bar{n}_{13} + \bar{n}_{14},$$

which, due to the two n_{1i} relations, are s-dependent functions:

$$\begin{aligned} &\mathscr{S}s + 2(1+s)\mathscr{P} = 0, \quad \mathscr{S}^2 + \mathscr{S}s/(1+s) = 1 + s^2, \\ &2(1+s)\mathscr{S} = -s + \sqrt{\delta}, \\ &\delta = s^2 + 4(1+s^2)(1+s)^2, \\ &2\bar{n}_{13} = \mathscr{S} + \sqrt{1 + s^2 + \mathscr{S}s/(1+s)}, \\ &\bar{n}_{14} = \mathscr{S} - \bar{n}_{13}. \end{aligned} \quad \text{(A4)}$$

(iii) Third, we see from (A2b) that n_{04} is known from n_{oi} $i \neq 4$ and for n_{03} we get

$$\begin{aligned} &n_{03}^2 + \mu n_{03} - n_{02} = 0, \\ &2n_{03} = -\mu + \sqrt{\mu^2 + 4n_{02}}, \\ &\mu = (1 - n_{02})(1 - s)/(\bar{n}_{14} - \bar{n}_{13}). \end{aligned} \quad \text{(A5)}$$

(iv) Fourth, we have two equivalent relations for n_{11} (we define $A_1 = \mathscr{P} - s$),

$$\begin{aligned} &n_{01}s + n_{02} - n_{03}\bar{n}_{14} - n_{03}\bar{n}_{13} \\ &\quad = A_1 n_{11} = n_{01} + n_{02}s - n_{03}\bar{n}_{13} - n_{04}\bar{n}_{14}, \end{aligned} \quad \text{(A6)}$$

from which we can reconstruct all parameters: $n_{12} = sn_{11}, n_{1i} = \bar{n}_{1i}n_{11}, \rho, \gamma_j$.

2. Relations and constructions with the third component included

We find seven relations and seven parameters $n_{3i}, \rho_3, \gamma_{3i}$ that must be constructed with (s, n_{02}):

$$\begin{aligned} &n_{31}(\rho_3 + \gamma_{31})/2 = n_{32}(\rho_3 - \gamma_{31})/2 = -n_{33}(\rho_3 + \gamma_{32}) \\ &\quad = n_{34}(\gamma_{32} - \rho_3) = n_{33}n_{34} - n_{31}n_{32} \\ &\quad = n_{01}n_{32} + n_{02}n_{31} - n_{03}n_{34} - n_{04}n_{33}, \\ &n_{31}n_{12} + n_{32}n_{11} = n_{34}n_{13} + n_{33}n_{14}, \\ &n_{31}n_{11} + n_{32}n_{12} = n_{34}n_{14} + n_{33}n_{13}. \end{aligned} \quad \text{(A7)}$$

(i) Still ρ_3, γ_{3j} are known from the n_{3i}:

$$\begin{aligned} &\rho_3 n_{31}n_{32} = (n_{31} + n_{32})(n_{33}n_{34} - n_{31}n_{32}), \\ &\gamma_{31}(n_{32} + n_{31}) = \rho_3(n_{32} - n_{31}), \\ &\gamma_{32}(n_{34} + n_{33}) = \rho_3(n_{34} - n_{33}), \end{aligned} \quad \text{(A8)}$$

from which we obtain in (A7) a third n_{3i} relation

$$n_{31}n_{32}(n_{33} + n_{34}) + 2n_{33}n_{34}(n_{31} + n_{32}) = 0.$$

(ii) We define intermediate parameters that are s-dependent functions:

$$\begin{aligned} &z = n_{32}/n_{31}, \quad \bar{n}_{3i} = n_{3i}/n_{31}, \\ &i = 3,4, \quad P = \bar{n}_{33}\bar{n}_{34}, \quad S = \bar{n}_{33} + \bar{n}_{34}, \end{aligned} \quad \text{(A9)}$$

and satisfy the relations:

$$\begin{aligned} &s + z = \bar{n}_{34}\bar{n}_{13} + \bar{n}_{33}\bar{n}_{14}, \quad 1 + zs = \bar{n}_{33}\bar{n}_{13} + \bar{n}_{34}\bar{n}_{14}, \\ &zS + 2(1+z)P = 0, \quad P\mathscr{P} = sz/4, \\ &S\mathscr{S} = (1+z)(1+s), \quad z^2 + 1 + vz = 0, \\ &2z = -v + \sqrt{v^2 - 4}, \\ &v(\bar{n}_{13} - s\bar{n}_{14})(s\bar{n}_{13} - \bar{n}_{14}) \\ &\quad = (1+s)(\bar{n}_{13} - \bar{n}_{14})^2\mathscr{S}/2 + (\bar{n}_{13} - s\bar{n}_{14})^2 \\ &\qquad + (\bar{n}_{14} - s\bar{n}_{13})^2, \\ &2\bar{n}_{33} = S + \sqrt{S^2 - 4P}, \quad \bar{n}_{34} = S - \bar{n}_{33}. \end{aligned} \quad \text{(A10)}$$

(iii) We get n_{31} and define $A_3 = P - z$,

$$n_{31}A_3 = n_{01}z + n_{02} - n_{03}\bar{n}_{34} - n_{04}\bar{n}_{33}. \quad \text{(A11)}$$

Finally all parameters are deduced from $(s, n_{02}): \to n_{32} = zn_{31}, \quad n_{3i} = \bar{n}_{3i}n_{31}, \rho_3, \gamma_{3j}$.

3. Construction of the 16 shock limits $\Sigma_i^{2j} = n_{0i} + n_{ji}, \Sigma_i^{3j} = \Sigma_i^{2j} + n_{3i}$

From (A6)–(A10) we see that the n_{ji} and Σ_i are linear combination of the n_{0i} with s-dependent factors. Eliminating $n_{04} = n_{01}n_{02}/n_{03}$ then $n_{03}\Sigma_i$ become quadratic n_{03} polynomials. It is remarkable that all roots are of the type n_{01} or n_{02} multiplied by s-dependent functions. This comes from an identity satisfied by the coefficients of n_{0i} at the linear Σ_i level:

$$\begin{aligned} &\Sigma_i = \sum_{p=1}^{4} n_{0p}d_p, \\ &\quad \text{if } d_1d_2 = d_3d_4 \to n_{03}\Sigma_i \\ &\quad = d_3(n_{03} + n_{01}d_1/d_3)(n_{03} + n_{02}d_2/d_3). \end{aligned} \quad \text{(A12)}$$

For each (p, j) family of Σ_i we explicit one case with $i = 1$:

$$\begin{aligned} &\Sigma_1^{21} = n_{01} + n_{11} \\ &\quad = (-\bar{n}_{14}n_{03} - n_{04}\bar{n}_{13} + n_{02} + n_{01}\bar{n}_{13}\bar{n}_{14})/A_1, \\ &\Sigma_1^{22} = n_{01} + n_{21} = n_{01} + n_{12} \\ &\quad = (n_{01}\mathscr{P} + n_{02}s^2 - n_{03}s\bar{n}_{13} - n_{04}s\bar{n}_{14})/A_1, \\ &\Sigma_1^{31} = n_{01} + n_{11} + n_{31} \\ &\quad = -n_{03}(\bar{n}_{34}/A_3 + \bar{n}_{14}/A_1) - n_{04}(\bar{n}_{33}/A_3 + \bar{n}_{13}/A_1) \\ &\qquad + n_{01}(\mathscr{P}/A_1 + z/A_3) + n_{02}(1/A_1 + 1/A_3), \\ &\Sigma_1^{32} = \Sigma_1^{22} + n_{31} \\ &\quad = -n_{03}(s\bar{n}_{13}/A_1 + \bar{n}_{34}/A_3) - n_{04}(s\bar{n}_{14}/A_1 + \bar{n}_{33}/A_3) \\ &\qquad + n_{01}(\mathscr{P}/A_1 + z/A_3) + n_{02}(s^2/A_1 + 1/A_3). \end{aligned} \quad \text{(A13)}$$

For Σ_1^{2j} the identity $d_1d_2 = d_3d_4$ is trivial, for Σ_1^{3j} we must use the relations $s+z$ and $1+sz$ written down in (A10). The same tools occur for the other $i=2,3,4$ values of Σ_i^{pj}. The explicit expressions are written down in Tables I and II,

$$n_{03}\Sigma_i = \Omega_i(s)(n_{03} - B_k(s)n_{01})(n_{03} - B_{k'}(s)n_{02}).$$

4. Bounds for the intermediate parameters $\bar{n}_{1i},\bar{n}_{3i},z$ when s belongs to (0.9,1)

From (A4)–(A10) we obtain bounds for the s-dependent functions when $s\in(0.9,1)$.

Lemma 1: $1<\mathscr{S}<1+s$ and s is increasing. From (A4) the inequalities are equivalent to $2+3s<\sqrt{\delta}<2(1+s)^2+s$ or $s^2+s-1>0$, $s>0$. We deduce $\mathscr{S}<1+s<2s(1+s)$.[2] From the quadratic $\mathscr{S}$ equation we find the derivative:

$$\mathscr{S}_s = (2s - \mathscr{S}/(+s)^2)(2\mathscr{S}+s/(1+s))>0,$$
$$\mathscr{S}(0.9) = 1.12\leqslant\mathscr{S}\leqslant S(1) = 1.186. \quad \text{(A14)}$$

Lemma 2: $\bar{n}_{13}>0$ and $-\bar{n}_{14}>0$ are increasing s functions. From (A4)–(A14) we get $\mathscr{P}<0$, $\bar{n}_{13}>0$, $-\bar{n}_{14}>0$ and for the n_{03} derivative

$$2(\bar{n}_{13})_s = \mathscr{S}_s + (2s + s\mathscr{S}_s/(1+s) + \mathscr{S}/(1+s^2))/(4\bar{n}_{13} - 2\mathscr{S})>0.$$

The $-\bar{n}_{14}$ derivative is more complicated. From $-1/\bar{n}_{14} = (1+s)/s + ((1+s)/s)\sqrt{1+2s/\mathscr{S}(1+s)}$ it is sufficient that the second term be decreasing. The derivative of this term has a sign given by

$$s^2(\sqrt{\delta}+1) - 2(1+s^2)(1+s)^2 - (s+1)s^2\delta_s/2\sqrt{\delta}$$

with δ given in (A4). The last term is negative, the sum of the two first terms being also negative it follows that $-\bar{n}_{14}$ is increasing,

$$\bar{n}_{13}(0.9) = 1.3303\leqslant\bar{n}_{13}\leqslant\bar{n}_{13}(1) = 1.398,$$
$$\bar{n}_{14}(1) = -0.2121\leqslant\bar{n}_{14}\leqslant\bar{n}_{14}(0.9) = -0.201. \quad \text{(A15)}$$

Lemma 4: $-z$ is a decreasing v function and if $v_{\inf}<v<v_{\sup}$ then $-z(v_{\sup})<-z(v)<-z(v_{\inf})$. From (A4) we get $v>0, z<0$, and $(-z)_v = z/\sqrt{v^2-4}<0$. For $v_{\sup}$ we write:

$$v = 1/s + ((1+s)(s+1/2) + \mathscr{S}(s+2+3s/(1+s^2))/2)/(1+s+\mathscr{S}/2)$$

deduced from (A10). We get an upper bound equal to 3.4117 by substituting $s=1$, $\mathscr{S}=\mathscr{S}(1)$ in the numerator and $s=0.9$, $\mathscr{S}(0.9)$ in the denominator. We find a lower bound equal to $v(1) = (7+\sqrt{33})/4$ because the sign of $v(s)-v(1)$ is given by

$$(\bar{n}_{13}-\bar{n}_{14})^2(\sqrt{\delta} - y\sqrt{33}) + 2(1-s)\mathscr{S}(\sqrt{33} - s(3+\sqrt{33})/4)>0,$$
$$0.323 = -z(v=3.4117) < -z < -z(s=1) = 0.353. \quad \text{(A16)}$$

Lemma 5: $\bar{\mathscr{S}} = \mathscr{S}/(1+s)$ is a decreasing s function. We find for the derivative:

$$\bar{\mathscr{S}}_s = -(1-s)(2+\bar{\mathscr{S}})/(1+s)^2(2\bar{\mathscr{S}}(1+s)+s)<0.$$

So $\bar{\mathscr{S}}(1)\leqslant\bar{\mathscr{S}}<S(0.9)$, and (A10): $2P = -z/\bar{\mathscr{S}}$ applying lower and upper bounds (A14)–(A16)

$$0.2726<\mathscr{P}\leqslant P(1) = 0.2976. \quad \text{(A17)}$$

Lemma 6: $\bar{n}_{34}<\bar{n}_{33}$ and $\bar{n}_{34}(1) = \bar{n}_{33}(1)$. From the (A9) $s+z$, $1+sz$ relations we find: $\bar{n}_{33}-\bar{n}_{34} = (1-s)(1-z)/(\bar{n}_{13}-\bar{n}_{14})>0$. It follows $\bar{n}_{34}^2\leqslant P(1), \bar{n}_{33}^2>0.2726$,

$$\bar{n}_{34}\leqslant\bar{n}_{34}(s=1) = 0.54545, \quad \bar{n}_{33}>0.52092,$$
$$\bar{n}_{33}-\bar{n}_{34}<0.0884, \quad \bar{n}_{33}<0.6338, \quad \bar{n}_{34}>0.4325. \quad \text{(A18)}$$

For the third inequality we use $(1-s)(1-z)<0.135$ and for two last ones the two first results.

Theorem 1: For $s\in(0.9,1)$ the intermediate parameters satisfy: $\bar{n}_{13}>0, \bar{n}_{14}<0$, $z<0, \bar{n}_{33}>0$, $\bar{n}_{34}>0$, $\mathscr{S}>0$, $\mathscr{P}<0$, $S>0$, $P>0$, $\bar{n}_{14}+\bar{n}_{34}>0$, $\bar{n}_{13}-\bar{n}_{33}>0$, $A_1 = \mathscr{P} - s<0$, $A_3 = P-z>0$ and the numerical bounds (A15)–(A18).

5. Positivity n_{03} intervals for Σ_i^{kj} with $s\in(0.9,1)$ and $n_{0i}>0$ (see Tables I and II)

The s-dependent roots $B_k,\tilde{B}_k$ are built-up with $\bar{n}_{ji},A_1,A_3,s,z$ (see Theorem 1 for the signs). For k,j fixed we first study the n_{03} intervals with $\Sigma_i>0$ and second their intersections for k,j varying.

Lemma 8: If $0<n_{03}<\inf(n_{01}B_1,n_{02}\tilde{B}_1)$ then $\Sigma_i^{21}>0$, if $n_{03}>\sup(n_{02}/\tilde{B}_1,n_{01}/B_1)$ then $\Sigma_i^{22}>0$. We find that $B_1,\tilde{B}_1$ are positive, $B_2,\tilde{B}_2$ negative, while Ω_i^{21} are negative for $i=1,2$ positive for $i=3,4$ and for Ω_i^{22} they are all positive.

Lemma 9: All Σ_i^{31} are positive if $0<n_{03}<\inf(n_{02}B_4,n_{01}\tilde{B}_4)$. From signs considerations we find successively: $\Omega_1<0,\Omega_3>0,\bar{n}_{13}sA_3+\bar{n}_{33}zA_1>0\rightarrow B_3<0,\tilde{B}_4<0;A_1+A_3 = (\bar{n}_{13}-\bar{n}_{33})(\bar{n}_{14}-\bar{n}_{34})<0\rightarrow\Omega_4<0$ (due to $|\bar{n}_{34}/\bar{n}_{14}|>1$) and $B_4>0\rightarrow\tilde{B}_3>0$ and $\Omega_2>0$. Further $B_4<\tilde{B}_3$, due to $B_4-\tilde{B}_3 = A_1A_3(\bar{n}_{14}-\bar{n}_{34})^2/\Omega_1\Omega_3,<0$. The signs of Ω_i and the locations of the roots give the intersections of the n_{03} positive intervals.

Lemma 10: All Σ_i^{32} are positive if $n_{03}>\sup(n_{01}B_5,n_{02}\tilde{B}_5)$. From sign considerations we find: $\Omega_2>0,\Omega_3$ and $\tilde{B}_5$ are positive; $y^2A_3+A_1<A_3+A_1<0$, $n_{33}A_1+s\bar{n}_{14}A_3<0\rightarrow B_6<0,\Omega_1$ and Ω_4 are positive $\rightarrow B_5>0$; $z^2A_1+A_3 = (\bar{n}_{33}-z\bar{n}_{13})(\bar{n}_{34}-z\bar{n}_{14})>0$ (due to $\bar{n}_{34}+\bar{n}_{14}>0,|z|<1)\rightarrow\tilde{B}_6>0$. Further $\tilde{B}_6<B_5$, due to $\tilde{B}_6-B_5 = A_1A_3(z\bar{n}_{13}-\bar{n}_{34})^2/\Omega_2\Omega_4<0$. The signs of Ω_i and the locations of the roots give the n_{03} interval for which positivity is satisfied.

Theorem 2: For all k,j,i values the Σ_i^{kj} are positive if

$$\sup(n_{01}/B_1,n_{01}B_5;n_{02}/\tilde{B}_1,n_{02}\tilde{B}_5) < n_{03} < \inf(n_{01}B_1,n_{01}\tilde{B}_4,n_{02}\tilde{B}_1,n_{02}B_4). \quad \text{(A19)}$$

Lemma 11: $B_1<\tilde{B}_4$, $\tilde{B}_5<1/\tilde{B}_1$, $B_4<\tilde{B}_1$, $1/B_1<B_5$. We write down identities for which the sign of each factor is known:

$$B_1 - B_4 = A_1\bar{n}_{34}(-z+\bar{n}_{13}\bar{n}_{34})/\Omega_3^{31}<0,$$
$$\tilde{B}_5 - s/B_1 = A_1(\bar{n}_{33}\bar{n}_{13}-sz)\bar{n}_{34}/\Omega_2^{32}<0,$$
$$B_4 - \tilde{B}_1 = A_1\bar{n}_{33}(\bar{n}_{14}-\bar{n}_{34})/\Omega_1^{31}<0,$$
$$1/B_1 - B_5 = A_1(A_3\bar{n}_{13}+\bar{n}_{34}-z\bar{n}_{13})/\Omega_1^{32}<0.$$

Theorem 3: $\Sigma_i^{kj}>0$ for all k,j,i values if

$$\sup(n_{02}/\widetilde{B}_1,n_{01}B_5)<n_{03}<\inf(n_{01}B_1,n_{02}B_4)\ . \qquad (A20)$$

6. Positivity n_{02} intervals for Σ_i^{kj} with $s\in(0.9,1)$ and $n_{01}=1$

Lemma 12: (i) if $n_{03}\gtrless n_{02}X$ with $X>0$ then $n_{02}\lessgtr F(X)=(C+DX)/X(XC+D)$ where we define $C=\bar{n}_{13}-\bar{n}_{14}>0$ and $D=1-s>0$; (ii) if $n_{03}\gtrless\overline{X}n_{01}$ with $\overline{X}>0$, $C-\overline{X}D>0$, $n_{01}=1$ then $n_{02}\gtrless G(\overline{X})=(\overline{X}^2C-\overline{X}D)/(C-\overline{X}D)$; (iii) $H(X,\overline{X})=C(1-X\overline{X})+D(X-\overline{X})\gtrless 0$ then $F(X)\gtrless G(\overline{X})$. For the proofs we use the representation (A5): $2n_{03}=-\mu+\sqrt{\mu^2+4n_{02}}$. In (i) we get $1\gtrless n_{02}X^2+X\mu$ and $n_{02}\gtrless\overline{X}^2+\overline{X}\mu$ in (ii) while (iii) is trivial.

Lemma 13: (i) $C-DB_1>0$; (ii) $C-DB_5>0$ and we can apply Lemma 12 to $\overline{X}=B_1$ and B_5. For (i) we notice that $C-DB_1=s\bar{n}_{13}-\bar{n}_{14}>0$. For (ii) we remark that if $B_5<\mathrm{Sup}\,B_5$ and if $C>D\,\mathrm{Sup}\,B_5$ it follows that $C>DB_5$. First we seek an upper bound for $B_5=\Omega_2/\Omega_4$ (written down in Table II). From $\Omega_1>0$ we get the inequality $-\bar{n}_{34}A_1/s\bar{n}_{13}A_3<1$, using this result we find $\Omega_2<A_3\bar{n}_{13}(1-sz)$ and $\Omega_4>A_3\bar{n}_{13}(\bar{n}_{13}-s\bar{n}_{14})$ and $B_5<(1-sz)/(\bar{n}_{13}-s\bar{n}_{34})=\mathrm{Sup}\,B_5$. Second $C>D\,\mathrm{Sup}\,B_5$ is equivalent to

$$(\bar{n}_{13}-\bar{n}_{14})(\bar{n}_{13}-s\bar{n}_{34})>(s-1)(1-sz)\ ,$$

which is satisfied for the numerical bound values of subsection 4 of the Appendix. We apply Lemma 12 to Theorem 3:

Theorem 3′: Σ_i^{kj} are positive if

$$\sup(F(B_4),G(B_5))<n_{02}<\inf(G(B_1),F(s/B_1)). \quad (A20')$$

Lemma 14: (i) $H(s/B_1,B_1)>0$ and $F(s/B_1)>G(B_1)$, (ii) $H(B_4,B_5)>0$ and $F(B_4)>G(B_5)$. For (i) with the definition of Lemma 12 we find: $H=C(1-s)+D(-B_1+s/B_1)=(1-s)(s/B_1-\bar{n}_{14})>0$. For (ii) we have $H=C(1-B_4B_5)+D(B_4-B_5)$ or

$$H(B_4,B_5)=((1-s)A_1A_3/0.75sz\Omega_1^{32})\times(s-\bar{n}_{13}\bar{n}_{34})(1-\bar{n}_{13}\bar{n}_{33})>0.$$

The first factor is positive, the two others (using the numerical bounds of subsection 4) are also positive. Applying Lemma 14 to Theorem 3′ we find the last result:

Theorem 4: $\Sigma_i^{kj}>0$ for all k,j,i values if $s\in(0.9,1)$, $n_{01}=1$ and for n_{02}:

$$F(B_4)<n_{02}<G(B_1),$$

$$G(B_1)=\bar{n}_{13}(s-1+\bar{n}_{13}(\bar{n}_{13}-\bar{n}_{14}))/(s\bar{n}_{13}-\bar{n}_{14})\ ,$$

$$F(B_4)=(\bar{n}_{13}-\bar{n}_{14}+B_4(1-s))\times[B_4(B_4(\bar{n}_{13}-\bar{n}_{14})+1-s)]^{-1}, \qquad (A21)$$

which defines the (s,n_{02}) positivity domain of Fig. 1.

[1] J. E. Broadwell, Phys. Fluids 7, 1243 (1964); J. Fluid Mech. 9, 401 (1964); L. Tartar, Séminaire Goulauic-Schwartz 1 (1975–1976); J. Hardy and Y. Pomeau, J. Math. Phys. 13, 1042 (1972); R.Gatignol, *Lectures Notes in Physics*, Vol. 36 (Springer, Berlin, 1975); Trans. Theoret. Stat. Phys. 16, 837 (1987) with references therein; H. Cabannes, *Lectures Notes* (Berkeley U. P., Berkeley, 1980); Trans. Theoret. Stat. Phys. 16, 809 (1987) with references therein; S. Harris, Phys. Fluids 9, 1328 (1966); J. Hardy, and Y. Pomeau, J. Math. Phys. 13, 1042 (1972); U. Frisch, D. D'Humieres, B. Hasslacher, P. Lallemand, Y. Pomeau, and J. P. Rivet, Complex Systems 1, 649 (1987) and references therein; T. Platkowski, Mec. Res. Comm. 11, 201 (1987); Bull. Pol. Acad. Sci. 32, 247(1984); K. Hamdache, preprint ENSTA-GHN (1987).

[2] H. Cornille, J. Phys. A 20, L1063 (1987); Some Topics on Inverse Problems, edited by P. C. Sabatier (World Scientific, Singapore, 1988), p. 101; Taclay Pht 88-113, to appear in Trans. Theory Stat. Phys. *Proceedings of Sixteenth International Symposium on Rarefield Gas Dynamics* Pasadena Center, to be published in AIAA Progress in Astronautics and Aeronautics, edited by Weaver, Muntz, and Campbell.

[3] H. Cornille, J. Stat. Phys. 52, 897 (1988).

Billiard systems on quadric surfaces and the Poncelet theorem

Shau-Jin Chang
Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801

Kang-Jie Shi
CCAST (World Laboratory), Beijing, People's Republic of China and Northwest University, Xian, People's Republic of China

(Received 15 September 1988; accepted for publication 30 November 1988)

Billiard systems constrained to move on a quadric surface with confocal quadric boundaries were studied. The trajectory of a billiard is described as a collection of geodesic segments joining on the boundary. At the joining points, the trajectory obeys the law of reflection. It was found that the geodesic segments are all tangent to a common confocal quadric curve (a caustic) on this quadric surface. If a trajectory is closed, then all trajectories sharing the same caustic quadric are all closed and have the same period and length. Thus a generalization of Poncelet's theorem on a quadric surface is achieved. The elliptical billiard systems on spheres and pseudospheres were studied and results were obtained that are similar to those on a plane. The results can be extended to n dimensions.

I. INTRODUCTION

A billiard system describes the motion of a particle that travels as a free particle inside a closed region and reflects elastically at the boundary.[1] Billiard systems provide important insights in mathematics and physics and have been studied extensively.

For a two-dimensional billiard system with an ellipse as its boundary, the straight segments of a trajectory are always tangent to a caustic curve.[2,3] This caustic curve is a quadric confocal to the original ellipse. In addition, if a trajectory is closed after p bounces, then all trajectories sharing the same caustic quadric are also closed after p bounces. This fact is a special case of the Poncelet theorem in projective geometry.[4]

Chang and Friedberg gave one of the possible extensions of the Poncelet theorem to three and higher dimensions.[5,6] They studied billiard systems with elliptical boundaries and discovered that the trajectory of a particle inside a three-dimensional ellipsoid gives rise to two caustic quadric surfaces that are confocal to the original ellipsoid. Chang and Friedberg also established that if any trajectory in a three-dimensional ellipsoid is closed after p bounces, then all the trajectories that share the same caustic quadric surfaces are also closed after p bounces independent of the starting point.

In this paper, we study a billiard system where the particle is constrained to move on the surface of a quadric (ellipsoid or pseudoellipsoid, etc.) and is reflected elastically on boundaries defined by confocal quadrics on the surfaces. By taking appropriate limits, we extend the validity of our results to billiard systems on a sphere or pseudosphere. We can generalize our results to billiard systems constrained to move on an m-dimensional quadric surface in an n-dimensional space.

In Sec. II, we begin with a brief review of Hamilton–Jacobi equations in a three-dimensional elliptical coordinates system. We then introduce a billiard system with boundaries composed of several confocal quadrics. We prove results similar to those obtained in Ref. 5. In the limit when one of the caustics is identical to the boundary ellipsoid, we obtain a billiard system constrained to move on this ellipsoid. In Sec. III, we give an alternative proof based on a direct separation of variables for this constrained system. In Sec. IV, we study billiard systems on a sphere as the proper limit of an ellipsoid. In Secs. V and VI, we extend our results to billiard systems on a pseudosphere. We describe the projective definition of confocal quadrics in an Appendix.

II. BILLIARD SYSTEMS WITH QUADRICAL BOUNDARIES

Chang and Friedberg studied billiard systems with an ellipse as the boundary.[5] In this section, we shall extend our study to regions bounded on all sides by confocal quadrics.

In a three-dimensional space, we can express a family of confocal quadrics as

$$\frac{x^2}{A-\lambda}+\frac{y^2}{B-\lambda}+\frac{z^2}{C-\lambda}=1, \tag{2.1}$$

where we choose $A>B>C>0$. Depending on the value of λ relative to A, B, and C, we have different species of quadrics, as described in Ref. 5.

At any given point in space, there are three mutually orthogonal confocal quadric surfaces passing through it. The λ 's associated with these surfaces are Jacobi variables.[7] For real x, y, and z, the λ 's fall into the ranges

$$\infty<\lambda_1<C<\lambda_2<B<\lambda_3<A. \tag{2.2}$$

In the following, we follow the notation and method of Ref. 5.

The Cartesian and Jacobian coordinates are related by

$$x^2=\frac{(A-\lambda_1)(A-\lambda_2)(A-\lambda_3)}{(A-B)(A-C)}, \tag{2.3a}$$

$$y^2=\frac{(B-\lambda_1)(B-\lambda_2)(B-\lambda_3)}{(B-A)(B-C)}, \tag{2.3b}$$

$$z^2=\frac{(C-\lambda_1)(C-\lambda_2)(C-\lambda_3)}{(C-A)(C-B)}. \tag{2.3c}$$

The first fundamental form expressed in terms of the elliptical coordinates is

 0022-2488/89/040798-07$02.50

$$ds^2 = g_1\, d\lambda_1^2 + g_2\, d\lambda_2^2 + g_3\, d\lambda_3^2, \tag{2.4}$$

where

$$g_1 = \frac{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}{4(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}, \tag{2.5a}$$

$$g_2 = \frac{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}{4(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}, \tag{2.5b}$$

$$g_3 = \frac{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}{4(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)}. \tag{2.5c}$$

In the elliptical coordinates, the Lagrangian and Hamiltonian for a particle of unit mass are

$$L = \tfrac{1}{2}(g_1\dot\lambda_1^2 + g_2\dot\lambda_2^2 + g_3\dot\lambda_3^2) - V(\lambda_1,\lambda_2,\lambda_3) \tag{2.6}$$

and

$$\begin{aligned} H &= \tfrac{1}{2}\left(\frac{p_1^2}{g_1}+\frac{p_2^2}{g_2}+\frac{p_3^2}{g_3}\right) + V(\lambda_1,\lambda_2,\lambda_3) \\ &= 2\left[\frac{(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)\,p_1^2}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\right. \\ &\qquad \left. + \text{cyclic permutation of subscripts}\right] \\ &\qquad + V(\lambda_1,\lambda_2,\lambda_3), \end{aligned} \tag{2.7}$$

respectively, with

$$p_1 = g_1\dot\lambda_1,\quad p_2 = g_2\dot\lambda_2,\quad p_3 = g_3\dot\lambda_3. \tag{2.8}$$

We can generalize the result of Ref. 5 by considering the motion of a free particle inside a region specified by a more general boundary condition

$$\Lambda_1 < \lambda_1 < \Lambda_1',\quad \Lambda_2 < \lambda_2 < \Lambda_2',\quad \Lambda_3 < \lambda_3 < \Lambda_3'. \tag{2.9}$$

The particle bounces off the surfaces according to the law of reflections. We can achieve this in the framework of a Hamiltonian system by choosing the potential energy $V(\lambda_1, \lambda_2, \lambda_3)$ as

$$\begin{aligned} V(\lambda_1,\lambda_2,\lambda_3) &= V_1(\lambda_1)/[(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)] \\ &\quad + V_2(\lambda_2)/[(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)] \\ &\quad + V_3(\lambda_3)/[(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)], \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} V_i(\lambda_i) &= 0 \text{ for } \Lambda_i + \epsilon < \lambda < \Lambda_i' - \epsilon \\ &= (-1)^{i-1}V_0/\epsilon,\ V_0 > 0 \text{ for } \Lambda_i > \lambda_i \text{ or } \Lambda_i' < \lambda_i \end{aligned} \tag{2.11}$$

and $V_i(\lambda_i)$ have continuous first derivatives in the regions

$$\Lambda_i < \lambda_i < \Lambda_i + \epsilon,\quad \Lambda_i' - \epsilon < \lambda_i < \Lambda_i'.$$

As $\epsilon \to 0$, we recover the law of reflections on the boundaries. We can also impose these types of boundary conditions on only one or two of the λ's. The modification of (2.9)–(2.11) is straightforward.

With these boundary conditions, the Hamilton–Jacobian equation of the system,

$$\begin{aligned} &\frac{2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\left(\frac{\partial W}{\partial\lambda_1}\right)^2 \\ &\quad + \text{cyclic permutations of } \lambda\text{'s} + V(\lambda_1,\lambda_2,\lambda_3) \\ &\quad = \alpha, \text{ the energy,} \end{aligned} \tag{2.12}$$

is still separable. As described in Ref. 5, we can solve the Hamilton–Jacobian equation by choosing

$$W = W_1(\lambda_1) + W_2(\lambda_2) + W_3(\lambda_3) \tag{2.13}$$

and by introducing the separation constants α', α'' through

$$\begin{aligned} &2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)\left(\frac{dW_1(\lambda_1)}{d\lambda_1}\right)^2 \\ &\quad + V_1(\lambda_1) = \alpha(\lambda_1-\alpha')(\lambda_1-\alpha''), \end{aligned} \tag{2.14a}$$

$$\begin{aligned} &2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)\left(\frac{dW_2(\lambda_2)}{d\lambda_2}\right)^2 \\ &\quad + V_2(\lambda_2) = \alpha(\lambda_2-\alpha')(\lambda_2-\alpha''), \end{aligned} \tag{2.14b}$$

$$\begin{aligned} &2(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)\left(\frac{dW_3(\lambda_3)}{d\lambda_3}\right)^2 \\ &\quad + V_3(\lambda_3) = \alpha(\lambda_3-\alpha')(\lambda_3-\alpha''). \end{aligned} \tag{2.14c}$$

By the use of the identity

$$\begin{aligned} &\frac{(\lambda_1-\alpha')(\lambda_1-\alpha'')}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)} + \frac{(\lambda_2-\alpha')(\lambda_2-\alpha'')}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)} \\ &\quad + \frac{(\lambda_3-\alpha')(\lambda_3-\alpha'')}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)} = 1, \end{aligned} \tag{2.15}$$

we can show that the Hamilton–Jacobi equation (2.12) is indeed separated and satisfied. In Eqs. (2.12) and (2.14), α, α', and α'' are three independent constants of motion and are single-valued functions of coordinates and momenta. Since W in the Hamilton–Jacobi equation can be used as a generating function for a canonical transformation by which α, α', and α'' become new canonical momenta, the Poisson brackets of α, α', and α'' are identically zero. Thus this system is an integrable system. On the other hand, this system is also a bounded system. We can apply Arnold's theorem for a bounded integrable system[8] to our billiard system and obtain the following. The motion of the particle is equivalent to a quasiperiodic motion on a torus specified by α', α'', and α. The frequencies are also determined by α', α'', and α. If one orbit is closed, then all orbits for the same parameters α', α'', and α are also closed and have the same period. Since the speed of the particle is the same for these orbits ($v = \sqrt{2\alpha}$), we conclude that these closed trajectories have the same total length. Indeed, following the method described in Ref. 5, we can verify explicitly the above results in our billiard system.

In Eqs. (2.14), α is the energy. We now turn to the geometric meanings of α' α''. By the use of (2.14) and

$$p_i = \frac{dW_i}{d\lambda_i} = g_i\dot\lambda_1,\quad \dot\lambda_1 = \frac{1}{g_i}\left(\frac{dW_i}{d\lambda_i}\right) \tag{2.16}$$

we can prove that the straight segments are tangent to the quadrics $\lambda = \alpha'$ and $\lambda = \alpha''$, respectively. As we trace a particle with $v = \sqrt{2\alpha}$ along a straight-line segment and its extension from $-\infty$ to ∞, the λ's change continuously and return to their original values. Equations (2.14) are all satisfied if we set $V = 0$ in these equations. Setting $V = 0$ in (2.16), we have

$$\dot{\lambda}_1 = \frac{1}{g_1}\left(\frac{dW_1}{d\lambda_1}\right)$$
$$= \sqrt{\frac{\alpha(\lambda_1-\alpha')(\lambda_1-\alpha'')(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}} \tag{2.17}$$

and similar equations for $\dot{\lambda}_2$ and $\dot{\lambda}_3$. If the nominator is not zero, $\dot{\lambda}_i$ is a continuous function of λ. In addition, if $\dot{\lambda}_i \neq 0$, it must retain the same sign to guarantee the continuity of the motion. Therefore, λ_i can reach a turning point only at α' α'', A, B, or C. On the other hand, for $\dot{\lambda}_i$ to be real, $\dot{\lambda}_i$ must change sign at these points. Thus λ_i must reach α' and then bounce back. This implies that the straight line is tangent to $\lambda_i = \alpha'$. From the meanings of α, α', and α'', we see that they are really single-valued functions of coordinates and momenta.

In conclusion, we can summarize the motion of a billiard system subject to the boundary condition (2.9) by the following theorems.

Theorem 1: Different line segments of a trajectory are tangent to the same confocal quadrics $\lambda = \alpha'$ and $\lambda = \alpha''$.

Theorem 2 (generalized Poncelet theorem): If one of the trajectories is closed after p bounces, then all trajectories sharing the same caustics α' and α'' are closed after p bounces and have the same total path length.

III. BILLIARD SYSTEMS CONSTRAINED ON AN ELLIPSOID

We consider a billiard system moving on a quadric surface S. For the simplicity of presentation, we choose the surface to be an ellipsoid in three dimensions specified by the elliptical coordinates $\lambda_1 = \alpha'$. This system can be viewed as the limits of a particle moving between two confocal surfaces $\lambda_1 = \Lambda'$ and $\lambda_1 = \alpha^1$, with the trajectory tangent to the inner surface and bouncing elastically at the outer surface. In other words, the inner surface is chosen to be one of the caustics of the billiard system. At the limit $\Lambda' = \alpha'$, the two surfaces become one and the trajectory follows the geodesics on the surface. The other confocal surfaces $\lambda_2 = \text{const}$ and $\lambda_3 = \text{const}$ intersect the surface S at two families of mutually orthogonal curves. These define the confocal curves on the surface S. Using similar definitions, we can introduce confocal quadrics on an m-dimensional quadric surface in an n-dimensional space.

We can apply Theorems 1 and 2 obtained in Sec. II to the present system. The billiard is moving on S along geodesics inside a region defined by $\Lambda_2 < \lambda_2 < \Lambda_2'$ and $\Lambda_3 < \lambda_3 < \Lambda_3'$. At the boundary curves, the billiard bounces according to the law of reflection. The original system possesses two caustics, of which one is the surface S itself. The intersect of the other caustic and the surface S gives a caustic curve $\lambda = \alpha''$ on S. The different geodesic segment of the trajectory is tangent to this same caustic curve α''. We now have a generalized Poncelet theorem on S. Consider a billiard system on S with the boundary conditions described above. If one of the trajectories is closed after p bounces, then all trajectories sharing the same caustic are closed after p bounces. These closed polygons all have the same path length. (See Fig. 1.)

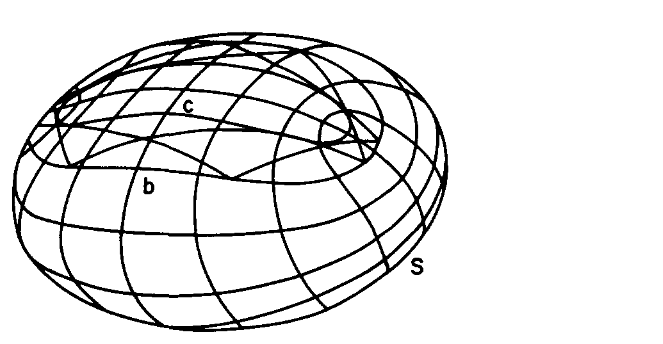


FIG. 1. Poncelet's theorem on an ellipsoid (S). The ellipsoid is parametrized by two families of confocal quadrics. The periodic orbits (polygons) are constructed by geodesic segments. The quadric b is the boundary of the billiard system and the quadric c is the caustic for these periodic trajectories.

Note that the result is still valid if we remove one or more of the boundary curves. Indeed, Arnold has studied the motion of a free particle constrained on such a surface without additional boundary curves and demonstrated the existence of caustic curves.[8]

We can also prove Theorems 1 and 2 for a billiard system on surface S directly by solving Hamilton–Jacobi equation. The advantage of this new derivation is that it can be extended to billiard systems on a pseudoquadric surface. The constraint $\lambda_1 = \alpha'$ is now built into the Hamiltonian:

$$\begin{aligned} H &= \frac{1}{2}\left(\frac{p_2^2}{g_2} + \frac{p_3^2}{g_3}\right) + V(\lambda_2,\lambda_3) \\ &= 2\left[\frac{(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)p_2^2}{(\lambda_3-\lambda_2)(\alpha'-\lambda_2)} \right. \\ &\quad \left. + \frac{(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)p_3^2}{(\alpha'-\lambda_3)(\lambda_2-\lambda_3)}\right] \\ &\quad + V(\lambda_2,\lambda_3)\,. \end{aligned} \tag{3.1}$$

The Hamilton–Jacobi equation is

$$\begin{aligned} &\frac{2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}{(\alpha'-\lambda_2)(\lambda_3-\lambda_2)}\left(\frac{\partial W}{\partial\lambda_2}\right)^2 \\ &\quad + \frac{2(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)}{(\alpha'-\lambda_3)(\lambda_2-\lambda_3)}\left(\frac{\partial W}{\partial\lambda_3}\right) \\ &\quad + V(\lambda_2,\lambda_3) = \alpha, \text{ the energy}, \end{aligned} \tag{3.2}$$

where $V(\lambda_2,\lambda_3)$ is chosen as

$$V(\lambda_2,\lambda_3) = \frac{V_2(\lambda_2)}{(\alpha'-\lambda_2)(\lambda_3-\lambda_2)} + \frac{V_3(\lambda_3)}{(\alpha'-\lambda_3)(\lambda_2-\lambda_3)}, \tag{3.3}$$

with $V_i(\lambda_i)$ the same as that in (2.11). We can solve (3.2) by introducing a separable W:

$$W = W_2(\lambda_2) + W_3(\lambda_3), \tag{3.4}$$

with the separation constants $\alpha' \equiv \lambda_1$, and α'' via

$$2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2) + V_2(\lambda_2) = \alpha(\lambda_2-\alpha')(\lambda_2-\alpha''), \tag{3.5a}$$

$$2(A-\lambda_3)(B-\lambda_3)(C-\lambda_3) + V_3(\lambda_3) = \alpha(\lambda_3-\alpha')(\lambda_3-\alpha''). \tag{3.5b}$$

Note that (3.5a) and (3.5b) are identical to (2.14b) and

(2.14c). Since this system obeys the same set of equations and boundary conditions in λ_2 and λ_3 variables as those in Sec. II, Theorems 1 and 2 established in Sec. II are also valid for the billiard system on surface S ($\lambda = \alpha'$).

It is straightforward to see that we can gereralize our results to n-dimensional space subjected to m constraints;

$$\lambda_i = \alpha^{(i)}, i = 1, 2, \ldots, m . \tag{3.6}$$

The billiard is now moving on (n-m) dimensional quadric hypersurface in an n-dimensional space.

IV. BILLIARD SYSTEM ON A SPHERE

In this section, we shall look at the billiard system on surfaces with additional symmetry. The cylindrical symmetrical cases $A = B \neq C$ or $A \neq B = C$ are not particularly interesting. The confocal quadric curves on these surfaces are circles and ellipses which are expected from the cylindrical symmetry. However, the spherical symmetrical cases $A = B = C$ are more interesting. We need to take the limit $C \to B \to A$ more carefully.

We begin with an ellipsoid

$$\frac{x^2}{A} + \frac{y^2}{B} + \frac{z^2}{C} = 1, \tag{4.1}$$

where we assume, as before, $A > B > C$. A quadric confocal to (4.1) is

$$\frac{x^2}{A-\lambda} + \frac{y^2}{B-\lambda} + \frac{z^2}{C-\lambda} = 1. \tag{4.2}$$

Only those quadrics with $A > \lambda > C$ can intersect with the original ellipsoid (4.1) in real space. Thus as B and C approach A, λ should also approach A. We shall consider the limit of B, C, and λ approaching A, with the fixed ratios

$$(\lambda - C)/(A - C) \equiv \xi, \quad (B - C)/(A - C) \equiv b, \quad 0 < b < 1. \tag{4.3}$$

Equations (4.3) imply

$$\begin{aligned} A - \lambda &= (A - C)(1 - \xi), \\ B - \lambda &= (A - C)(b - \xi), \\ c - \lambda &= -\xi(A - C). \end{aligned} \tag{4.4}$$

At the limit $A \to C$, we have

$$\frac{x^2}{1-\xi} + \frac{y^2}{b-\xi} - \frac{z^2}{\xi} = 0. \tag{4.5}$$

The intersection of (4.5) and the sphere S,

$$x^2 + y^2 + z^2 = A, \tag{4.6}$$

gives rise to the confocal quadrics on S. We have shown some of the typical confocal quadrics in Fig. 2. We can extend the results of Sec. II to regions bounded by these confocal quadric curves on S.

If we consider a region on S near $x = A$, $y = z = 0$ and if we choose small b and ξ, we then recover the original two-dimensional billiard results. In other words, we may also consider the results described in Sec. II as special cases of the results described in this section.

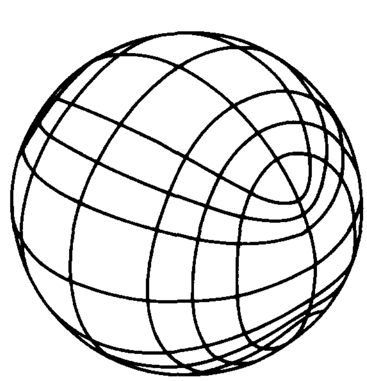


FIG. 2. Confocal quadrics on a sphere (S). We have obtained these confocal quadrics by treating the sphere as a proper limit of an ellipsoid. The parameter b used [see Eq. (4.5)] is $b = 0.25$.

V. PSEUDOELLIPSOIDS AND PSEUDOSPHERES

In this section, we shall generalize our results to the billiard systems defined on a pseudoellipsoidal surface. By taking the proper limit, we can extend our results to those defined on a pseudosphere. We shall restrict our discussions to three dimensions. Most of the results can be extended straightforwardly to n dimensions.

We shall first introduce elliptical coordinates on a pseudoellipsoid. We begin with a family of confocal quadrics, as described in Sec. II by Eq. (2.1):

$$\frac{x^2}{A-\lambda} + \frac{y^2}{B-\lambda} + \frac{z^2}{C-\lambda} = 1 \tag{5.1}$$

where we choose, as before, $A > B > C > 0$. For $\lambda < A$, (5.1) describes a real confocal quadric and for $\lambda > A$, (5.1) describes a pseudoellipsoid: The latter cannot be realized as real solutions to (5.1). However, as intrinsic surfaces, pseudoellipsoids and pseudospheres are well defined.

In Sec. II, we introduced the elliptical coordinates λ_1, λ_2, and λ_3. These coordinates are related to the Cartesian coordinates by (2.3). The first fundamental form is given in (2.4) and (2.5). To describe a pseudoellipsoid, we choose the ranges of the elliptical coordinates as

$$-\infty < \lambda_1 < C < \lambda_2 < B < A < \lambda_3. \tag{5.2}$$

Under restriction (5.2), we can show from (2.5) that $g_1 > 0$, $g_2 > 0$, and $g_3 < 0$. If we keep $\lambda_3 = \alpha'$ ($> A$), a constant, we obtain a pseudoellipsoidal surface S whose first fundamental form is

$$ds^2 = g_1\, d\lambda_1^2 + g_2\, d\lambda_2^2 , \tag{5.3}$$

where

$$g_1 = \frac{(\lambda_2 - \lambda_1)(\alpha' - \lambda_1)}{4(A - \lambda_1)(B - \lambda_1)(C - \lambda_1)} > 0, \tag{5.4a}$$

$$g_2 = \frac{(\lambda_1 - \lambda_2)(\alpha' - \lambda_2)}{4(A - \lambda_2)(B - \lambda_2)(C - \lambda_2)} > 0. \tag{5.4b}$$

The negative g_3 does not enter in (5.4) and consequently, the intrinsic surface S is well defined in $\lambda_1 \lambda_2$ space, as promised. The families of curves $\lambda_1 = \text{const}$ and $\lambda_2 = \text{const}$ are the confocal quadrics on S. Note that even though the pseudoellipsoid is well defined in $\lambda_1 \lambda_2$ space, not all the Cartesian coordinates given by (2.3) are real. One can check easily that the coordinate x is purely imaginary ($x^2<0$) and the coordinates y, z are real ($y^2>0, z^2>0$). This confirms the well-known fact that a pseudosphere cannot be embedded globally in a three-dimensional Euclidean space, but can be embedded in a Minkowskian space (see, e.g., Ref. 9).

The Hamiltonian for a billiard system constrained to move on the pseudoellipsoid S is

$$\begin{aligned} H &= \frac{1}{2}\left(\frac{p_1^2}{g_1}+\frac{p_2^2}{g_2}+V(\lambda_1,\lambda_2)\right) \\ &= 2\left[\frac{(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}{(\lambda_2-\lambda_1)(\alpha'-\lambda_1)}p_1^2 \right. \\ &\quad \left. +\frac{(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}{(\lambda_1-\lambda_2)(\alpha'-\lambda_2)}p_2^2\right] \\ &\quad + V(\lambda_1,\lambda_2), \end{aligned} \tag{5.5}$$

where $V(\lambda_1,\lambda_2)$ is chosen as

$$V(\lambda_1,\lambda_2)=\frac{V_1(\lambda_1)}{(\lambda_2-\lambda_1)(\alpha'-\lambda_1)}+\frac{V_2(\lambda_2)}{(\lambda_1-\lambda_2)(\alpha'-\lambda_2)}, \tag{5.6}$$

with $V_i(\lambda_i)$ the same as in (2.11). Following the same method described in Secs. II and III, we can show easily that the Hamilton–Jacobi equation for the system (5.5) is separable. We can therefore extend the results in Secs. II and III to the billiard system on a pseudoellipsoid.

To describe the motion on a pseudosphere, we take the limit $B, C\to A$. Following the method in Sec. IV, we let B, C, λ_1, and λ_2 approach A simultaneously by keeping the following ratios fixed:

$$(\lambda_i-C)/(A-C)=\xi_i, \quad i=1,2, \tag{5.7}$$

$$(B-C)/(A-C)=b, \tag{5.8}$$

with

$$\xi_1<0<\xi_2<b<1. \tag{5.9}$$

The ξ's are the new confocal coordinates on the pseudosphere and b is a new parameter. We introduce the (radius)^2 of the pseudosphere as $-R^2$, with

$$R^2\equiv\alpha'-A=\lambda_3-A>0. \tag{5.9$'$}$$

In terms of ξ_1 and ξ_2, the first fundamental form of the pseudosphere in elliptical coordinates is

$$\frac{ds^2}{R^2}=\frac{(\xi_2-\xi_1)}{4}\left[\frac{d\xi_1^2}{(1-\xi_1)(b-\xi_1)(-\xi_1)}+\frac{d\xi_2^2}{(1-\xi_2)(b-\xi_2)\xi_2}\right]. \tag{5.10}$$

In terms of ξ variables and with $\xi=$ constant boundaries, we can show that the Hamilton–Jacobi equation for the billiard system on the pseudosphere is again separable; we can extend the Poncelet theorems to here as well.

We shall conclude this section by giving the relations between the elliptical coordinates ξ and the polar coordinate $r\,\theta$ on a pseudopshere. The relations are

$$r\cos\theta=\sqrt{\xi_1\xi_2/b}, \tag{5.11a}$$

$$r\sin\theta=\sqrt{(b-\xi_1)(b-\xi_2)/b(1-b)}. \tag{5.11b}$$

The first fundamental form in terms of r and θ is the well-known expression

$$\frac{ds^2}{R^2}=\frac{dr^2}{1+r^2}+r^2\,d\theta^2, \tag{5.12a}$$

$$=d\rho^2+\sinh^2\rho\,d\theta^2, \tag{5.12b}$$

with

$$r=\sinh\rho. \tag{5.13}$$

Even though we cannot imbed globally a pseudosphere in Euclidean space, it is instructive to evaluate its Cartesian coordinates formally. Substituting (5.7)–(5.9) into (2.3), we have

$$x^2/R^2=-(1-\xi_1)(1-\xi_2)/(1-b)=-(1+r^2), \tag{5.14a}$$

$$y^2/R^2=(b-\xi_1)(b-\xi_2)/(1-b)b=r^2\sin^2\theta, \tag{5.14b}$$

$$z^2/R^2=(-\xi_1\xi_2)/b=r^2\cos^2\theta, \tag{5.14c}$$

and

$$x^2+y^2+z^2=-R^2, \tag{5.15}$$

which confirms that we have to introduce an imaginary x.

VI. BILLIARDS ON A PSEUDOSPHERE—PROJECTIVE METHOD

It is known that the geometry of a pseudosphere is the non-Euclidean geometry of constant negative curvature known as hyperbolic geometry.[9,10] Hyperbolic geometry can be represented projectively inside a quadric known as an "absolute configuration."[11] (This is known as Cayley's disk.) The absolute configuration consists of the images of points at infinity on the pseudosphere. For simplicty, we restrict our discussions to three dimensions only.

In order to have a projective geometry proof, we need to introduce projective invariant notions for "reflection" and "confocal quadric." After introducing an absolute configuration in the projective geometry, one can define "reflection" about a plane inside the absolute configuration as follows. Let us denote the absolute configuration by S_0. Let the reflecting plane be T and its "pole" about S_0 be Q. (The pole is defined as the vertex of the cone which is tangent to S_0 at the intersections of T.) Then the image of a given point P, called P', is its symmetric point about T and Q. (See Fig. 3.) Points

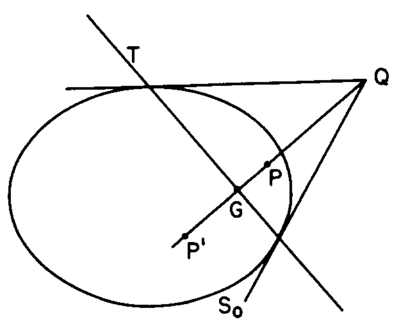


FIG. 3. Reflection on a pseudosphere in Cayley's representation. To construct the reflection of the point P with respect to the line (or plane) T, we first construct the pole Q. We then obtain the reflection image P′ by requiring that points Q, P, G, and P′ be linear and form a harmonic set. Point G is the intersection of line QP and T.

Q, P and P' are on a straight line. If this line intersects T at G, then these four points satisfy the relation

$$P'G/P'Q = -PG/PQ, \tag{6.1}$$

i. e., points $QPGP'$ form a harmonic set. One can show that the map of points to their images about a given plane is a projective transformation. This transformation leaves S_0 invariant, takes points inside of S_0 to points inside of S_0, and maps a straight line onto a straight line. If we identify the straight lines and reflections defined above inside an absolute configuration with the geodesics and ordinary reflections on a pseudosphere, we can show that these two geometries are equivalent.

To complete the projective proof, we need to introduce a projective definition for the confocal quadrics. We call three quadrics projectively confocal if for any plane the three poles with respect to these quadrics are on a straight line. We refer the details to the Appendix. In the Appendix, we have also proven the following lemma.

Lemma: Let P be a point on quadric S that is confocal to S_0 in the ordinary sense and let the tangent plane of S at P be T. Let the pole of T with respect to S_0 be Q. Then line QP is perpendicular to plane T.

Using this lemma, we can show easily that if S is confocal to S_0 in the ordinary sense, the law of reflection of S with respect to the absolute configuration S_0 is the ordinary law of reflection. (See Fig. 4.) The billiard system with boundary S and absolute configuration S_0 becomes a billiard system obeying the ordinary law of reflection. We can use our previous analysis to establish the existence of confocal caustics and the Poncelet theorem. Since we can show that a set of confocal quadrics on a pseudosphere can always be the image of a set of quadrics confocal to S_0, we thus establish the existence of confocal caustics and Poncelet theorems for a billiard system on a pseudosphere projectively. Note that we cannot use the projective method directly to prove that all closed trajectories sharing the same caustic surfaces have the same length. The result can be proven after we introduce the concepts of distance and angle in Cayley's representation. See Ref. 11 for discussions on the implementation of these measures.

VII. DISCUSSIONS

In an interesting article, Balazs and Voros studied the motion of a free particle on a pseudosphere[9] and imbedded

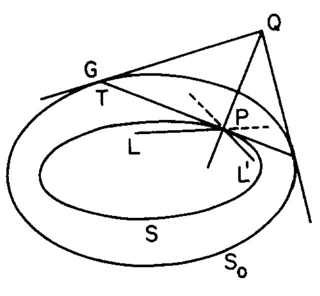


FIG. 4. When the quadric boundary S of a billiard system of a pseudosphere becomes confocal to the "absolute configuration" S_0, the reflection on S with respect to S_0 obeys the ordinary reflection law.

the pseudosphere as a quadric surface in a Minkowskian space:

$$x_1^2 + x_2^2 + x_3^2 = -R^2. \tag{7.1}$$

We can identify $-x_1^2$, x_2^2, and x_3^2 in (7.1) with x^2, y^2, and z^2 in (5.14), which leads to a realization of this imbedding. By projecting the hyperboloid (7.1) from the origin to the $x_1 = 1$ plane, we obtain projective representation of the pseudosphere, which is a special case of Cayley's disks.[11] All the points at infinity on the hyperboloid are mapped onto a circle, which is the absolute configuration of this representation.

Balazs and Voros studied the motion of a free particle on a compactified pseudosphere and used a procedure known to mathematicians as tesselation, which is the analog of filling the infinite plane with identical tiles.[12] After identifying all the tiles, we have a finite compact region with periodic boundary conditions. Balazs and Voros discovered that the motion on this compactified pseudosphere is always chaotic. In the present paper, among other things, we study the motions of a free particle on a pseudosphere, but with reflecting confocal quadric boundaries. We discover that these motions are always integrable, lead to confocal caustics, and obey a generalized Poncelet theorem. Since the present and Balazs–Voros systems share the same curved space, but with different boundary conditions, the latter must be the cause of the different behaviors. In the Balazs–Voros system, the compactified space preserves translational invariance. Since locally paralleled trajectories tend to separate exponentially on a pseudosphere, one can understand the chaotic behaviors exhibited in the Balazs–Voros system. Our quadric boundary conditions provide the necessary focusing effect to render the system integrable again. It is also interesting to know how these differences should affect their corresponding quantum systems.

In Ref. 5, Chang and Friedberg conjectured that the generalized Poncelet theorem is more general than the model studied in their paper and that it may depend only on the projective geometrical properties of the system. The findings in this paper certainly support the above conjecture.

ACKNOWLEDGMENTS

One of us (KJS) would like to thank Professor B. Y. Hou and Professor P. Wang for their stimulating discussions and acknowledge the support by CCAST (World Laboratory). KJS also wishes to thank Dr. Y. Zhang for calling his attention to Balazs and Voros' work. SJC wishes to thank Professor R. Friedberg for discussions and insights derived from their previous collaboration.

This work was supported in part by the US National Science Foundation under Grant No. PHY-87-01775.

APPENDIX: PROJECTIVE DEFINITION OF CONFOCAL QUADRICS

We call a family of quadrics *projectively confocal* if for any plane the poles with respect to these quadrics are on a straight line. We shall establish that (i) the projective mapping of a family of confocal quadrics is a family of projectively confocal quadrics and (ii) typically a family of projectively confocal quadrics can be mapped projectively into a family of confocal quadrics. Since our definition of projective confocal uses only the notions of tangency, straight line, and collinearity, it is obviously a projective invariant definition. To prove (i), we need only show that confocal quadrics are automatically projectively confocal.

We consider a set of confocal quadrics in its canonical coordinates:

$$\frac{x_1^2}{A_1-\lambda}+\frac{x_2^2}{A_2-\lambda}+\frac{x_3^2}{A_3-\lambda}=1. \tag{A1}$$

We choose an arbitrary plane T to be

$$n_1x_1+n_2x_2+n_3x_3=d, \tag{A2}$$

where $\mathbf{n}$ is a unit vector normal to the plane and d is the distance from the origin to the plane. The intersection of (A1) and (A2) gives a conic curve through which we can construct a cone tangent to (A1). The tip of this cone is the "pole" $\mathbf{Q}\,(\lambda)=(Q_1(\lambda),\,Q_2(\lambda),\,Q_3(\lambda))$. To obtain $\mathbf{Q}$, we make a scale transformation

$$x_i'=x_i/\sqrt{A_i-\lambda}, \tag{A3}$$

$$n_i'=n_i\sqrt{A_i-\lambda}/\sqrt{\Sigma n_j^2(A_j-\lambda)}. \tag{A4}$$

The scaled equations for the quadric and the plane are

$$x_1'^2+x_2'^2+x_3'^2=1, \tag{A5}$$

$$n_1'\,x_1'+n_2'\,x_2'+n_3'\,x_3'=d/\sqrt{\Sigma n_j^2(A_j-\lambda)}\equiv d', \tag{A6}$$

which represents a unit sphere and a plane (T') with a unit normal $\mathbf{n}'$. It is easy to compute the pole in the scaled variables as (see Fig. 5)

$$Q_i'=n_i'/d'=\sqrt{\Sigma n_j^2(A_j-\lambda)}\,n_i'/d=n_i\sqrt{A_i-\lambda}/d, \tag{A7}$$

which leads to

$$Q_i(\lambda)=\sqrt{A_i-\lambda}\,Q_i'(\lambda)=n_i(A_i-\lambda)/d. \tag{A8}$$

Note that $A_i(\lambda)$ is linear in λ and implies that $\mathbf{Q}_i\,(\lambda)$'s with different λ's all lie on the same straight line, with the direction given by $\mathbf{n}$.

To establish the inverse condition (ii), we note that we can always map two typical members of the family into a pair of confocal quadrics. We can now choose a coordinate system such that the two quadrics obey the canonical form (A1). By choosing plane T paralled to the x_1x_2, x_2x_3, and x_1x_3 planes, we can show that all quadrics in this familly can be written in the canonical form

$$x_1^2/A_1(\lambda)+x_2^2/A_2(\lambda)+x_3^2/A_3(\lambda)=1. \tag{A9}$$

The poles for the quadric (A9) with respect to the plane T of (A2) are

$$Q_i(\lambda)=n_iA_i(\lambda)/d. \tag{A10}$$

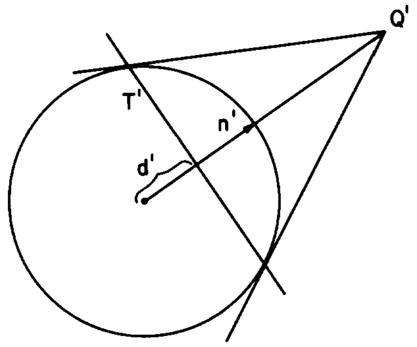


FIG. 5. In terms of the scaled variables, the quadric becomes a unit sphere. The pole $\mathbf{Q}'$ lies in the direction of the normal $\mathbf{n}'$ and its distance from the origin is $1/d'$.

For $Q_i(\lambda)$ to lie on a straight line, we can always parametrize $A_i(\lambda)$ to give

$$A_i(\lambda)=A_i-\lambda. \tag{A11}$$

This implies that the quadrics are confocal, as promised.

One of the consequences of (A8) is that the line of poles is always perpendicular to the cutting plane T. In the event that the cutting plane is tangent to a confocal quadric, this polar line becomes the normal of the tangent plane at the contact point. This is the lemma stated in Sec. VI.

[1] G. D. Birkhoff, "Dynamical system," Am. Math. Soc. Colloq. **IX** (1927); Ya. G. Sinai, Russ. Math. Surveys **25** (2), 137 (1970).
[2] J. B. Keller and Rubinow, Ann. Phys. **9**, 24 (1960).
[3] Ya. G. Sinai, *Introduction of Ergodic Theory* (Princeton U.P., Princeton, NJ, 1976).
[4] J. V. Poncelet, *Traité des Propriété Projectives des Figures* (Gauthier–Villars, Paris, 1866), 2nd ed., pp. 311–318. A summary of Poncelet's proof was given by A. Cayley in Q. J. Pure Appl. Math. **2**, 31 (1858).
[5] S. J. Chang and R. Friedberg, J. Math. Phys. **29**, 1537 (1988).
[6] A quite different three-dimensional version of Poncelet's theorem has been proven by Griffiths and Harris. See P. Griffiths and J. Harris, Comment. Math. Helvitici **52**, 145 (1977); L'Enseignement Math. **XXIV**, 31 (1978).
[7] For a survey of elliptical coordinates, see, e. g., E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1927), pp. 547–550; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1, pp. 511–513.
[8] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1984), pp. 264 and 271.
[9] N. L. Balazs and A. Voros, Phys. Rep. **143**, 109 (1986).
[10] There are many excellent books on non-Euclidean geometry. See, e.g., D. M. Y. Sommervile, *The Elements of Non-Euclidean Geometry* (Bell, London, 1914); H. E. Wolfe, *Introduction to Non-Euclidean Geometry* (Dryden, New York, 1945). See, also the references cited in Ref. 9.
[11] The concept of "absolute" was first introduced by A. Cayley in 1859. See discussions on Cayley's representation in Sommervile and Wolfe in Ref. 10.
[12] See, e.g., W. Magnus, *Non-Euclidean Tesselations and Their Groups* (Academic, New York, 1974).

Fractals and ultrasmooth microeffects

Robert A. Herrmann
Mathematics Department, U. S. Naval Academy, Annapolis, Maryland 21402

(Received 26 May 1988; accepted for publication 23 November 1988)

In this paper, a portion of the nonstandard methods first introduced by Nottale and Schneider [J. Math. Phys. **25**, 1296 (1984)] for the investigation of fractal behavior are replaced by methods employing polysaturated enlargements. This process yields internal functions that are ultrasmooth and have a standard part equal to the original fractal. These ultrasmooth nonstandard functions also possess a well-behaved * integral length concept. Additionally, a method is presented that shows that if certain behavior is modeled after a simple finitely discontinuous step function, then this function is also the standard part of an internal hypersmooth nonstandard approximation.

I. INTRODUCTION

The introduction of nonstandard analysis into the theory of fractals was accomplished by Nottale and Schneider,[1] where, in their basic paper, they discussed the many applications of fractals as physical models. Nottale and Schneider also introduced the concept of ϵ differentiability as an alternative to * differentiability (i.e., hyperdifferentiability): We recall their definition. Any function $f:D\rightarrow {}^*\mathbb{R}$, where $^*\mathbb{R}$ is the set of hyper-reals and $D\subset {}^*\mathbb{R}$, is ϵ differentiable at $p\in D$ if there exists a positive infinitesimal ϵ such that for every $x\in {}^*\mathbb{R}$ if $0<|x-p|<\epsilon$, then $(f(x)-f(p))/(x-p)$ is infinitely close to $r\in\mathbb{R}$. Nottale and Schneider then show that there is an ϵ differentiable function F' that is infinitely close to a fractal F in $\mathbb{R}^2$: Their stated motivation for introducing this restriction of * differentiation seems to imply that in $\mathbb{R}^2$ there may not exist a * differentiable function that is infinitely close to a given fractal; for this reason some such restriction appears necessary.

In a direct application of a major result of the present paper, it is shown that for nonempty compact $K\subset\mathbb{R}^n$ and any continuous function $f:K\rightarrow\mathbb{R}^m$ there exists an internal function $G:{}^*\mathbb{R}^n\rightarrow {}^*\mathbb{R}^m$ which is * continuously * differentiable: $^*f$, G are infinitely close on $^*X$; and $f=\mathrm{st}(G|{}^*K)$, among other properties. When applied to fractals this will considerably improve upon the concept of ϵ differentiability. As has now become customary, we assume $^*\mathscr{M}=({}^*\mathscr{H},\in,=)$ is a set-theoretic nonstandard model for a superstructure based upon the set $H=\mathbb{R}\cup X\cup Y$ (X,Y nonempty), with $^*\mathscr{H}$ the set of internal entities, and employ the usual definitions, conventions, and symbolism. Recall that the star notation "*" that appears before certain objects indicates that the starred objects are the nonstandard extensions of the unstarred objects. If a set X is a member of $\mathscr{H}$, then, conceptually, it may be assumed that $X\subset {}^*X$, even though this may not be the correct technical notation in many cases. Thus from the intuitive point of view, $^*X$ behaves as if it is, at the least, a set-theoretic extension. Further, assume that $^*\mathscr{M}$ is, at least, polysaturated.[2]

The entities $(X,\mathscr{T}_X)$ and $(Y,\mathscr{T}_Y)$ are topological spaces and, even though certain of our statements will hold under other criteria, we make the blanket assumption that X is compact. For any topological space $(Z,\mathscr{T})$ and for a point $p\in Z$, let $\mathscr{G}_p(Z)=\{G\mid p\in G\in\mathscr{T}\}$. Recall that for $p\in Z$ a *monad of* $p,\mu(p)$, is the set $\mu(p)=\cap\{{}^*G\mid G\in\mathscr{G}_p(Z)\}$ and since X is compact, then $^*X=\cup\{\mu(p)\mid p\in X\}$. Any $x,y\in {}^*Z$ are *infinitely close*, denoted by $x\approx y$, if there is some $p\in Z$ such that $x,y\in\mu(p)$. For any nonempty $D\subset {}^*X$ a function $f:D\rightarrow {}^*Y$ is *microcontinuous* at $p\in D$ if whenever $q\in D$ and $q\approx p$, then $f(q)\approx f(p)$. One of the major reasons that microcontinuous functions are considered interesting is that if X and Y are metric spaces, then $f:X\rightarrow Y$ is continuous if and only if $^*f$ is microcontinuous on $^*X$. Further, a sequence of functions $\{f_n|n\in\mathbb{N}\}$ on X into Y is equicontinuous if and only if $g\in {}^*\{f_n|n\in\mathbb{N}\}$ is microcontinuous.[3] Two functions f, G: $D\rightarrow {}^*Y$ *preserve nearness* if whenever p, $q\in D$ and $p\approx q$, then $f(p)\approx G(q)$. Two functions f, G:$D\rightarrow {}^*Y$ are *infinitely close on D* (denoted by $f\approx G$ or $f|D\approx G|D$) if for each $p\in D$, $f(p)\approx G(p)$.

In general, if a property is denoted by a term P in our metalanguage and expressible by the first-order language for the structure $\mathscr{M}$, then the * transfer of P's defining first-order characterization is denoted by "* P." Of course, any first-order sentence implied by P and expressible in the language of $\mathscr{M}$ will hold when interpreted, say by * transfer, in the structure $^*\mathscr{M}$. It is known that * continuity and microcontinuity are not equivalent concepts.[3] Moreover, it is significant to note that if Y is a metric space and $^*f$, $^*g$ map $^*X$ into $^*Y$, then, even if X is not compact, $^*f\approx {}^*g$ implies that $f=g$. Thus for many practical applications only nonstandard internal or external functions are infinitely close to a standard function.

II. BASIC APPROXIMATIONS

In establishing the following propositions, the polysaturated property of the model $^*\mathscr{M}$ is of major significance. First, we need an approximation theorem for microcontinuous functions that generalizes the Davis approximation theorem.[3] A point $q\in {}^*Z$ is *near standard* if there exists some $p\in Z$ such that $q\in\mu(p)$. If Z is Hausdorff, then the point p is unique and we write $\mathrm{st}(q)=p$. For the remainder of the present work, if the sets A, $B\in\mathscr{H}$, then let $\mathscr{F}(A,B)$ denote the set of all functions with domain A and codomain B.

Theorem 1.1: Suppose Y is a regular Hausdorff space. Let $f\in {}^*(\mathscr{F}(X,Y))$ be microcontinuous on $^*X$ and $f(p)$ be near standard for each $p\in X$. Define the function $F(p)=\mathrm{st}(f(p))$ for each $p\in X$. Then $F\in\mathscr{F}(X,Y)$ and F is

 0022-2488/89/040805-04$02.50

continuous on X. Moreover, f, ${}^*F$ preserve nearness and are infinitely close on ${}^*X$.

Proof: Suppose that arbitrary $p\in X$. Since $f(p)$ is near standard and Y is Hausdorff than there exists a unique $r\in Y$ such that $f(p)\in\mu(r)$ and $r=F(p)$. Let $p\approx x\in{}^*X$. Then $f(x)\approx f(p)\approx F(p)$ imples that $\mu(p)\subset f^{-1}[\mu(F(p))]$. Consider any $G_{F(p)}\in\mathscr{G}_{F(p)}$. Since Y is regular then there exists some $V\in\mathscr{G}_{F(p)}$ such that $F(p)\in V\subset\overline{V}\subset G_{F(p)}$. Hence $\mu(p)\subset f^{-1}[{}^*G_{F(p)}]$ and $\mu(p)\subset f^{-1}[{}^*V]$. Since $f\in{}^*(\mathscr{F}(X,Y))$ implies that f is internal then there exists some $G_p\in\mathscr{G}_p$ such that ${}^*G_p\subset f^{-1}[{}^*V]$. Let $x\in G_p$. Then $\mu(x)\subset{}^*G_p$ implies that $f(x)\in f[\mu(x)]\subset{}^*V$. However, there exists some $s\in Y$ such that $f(x)\approx s=F(x)$. Hence, $f(x)\in\mu(s)$ implies that $\mu(s)\cap{}^*V\neq 0$. Therefore, $s\in\overline{V}$ implies that $F(x)=s\in G_{F(p)}$. From this it follows that $x\in F^{-1}[G_{F(p)}]$ and $G_p\subset F^{-1}[G_{F(p)}]$. Thus F is continuous at each $p\in X$. Consequently, F is continuous on X in the general topological sense.

Let x, $y\in{}^*X$ and $x\approx y$. Then $x\approx y\approx r\in X$. Microcontinuity yields that $f(x)\approx f(y)\approx f(r)\approx F(r)$. Since F is continuous at r then $F(r)\approx{}^*F(y)\approx{}^*F(x)$. Therefore, f, ${}^*F$ preserve nearness. Finally, X being compact implies that for each $p\in{}^*X$ there exists some $r\in X$ such that $p\approx r$. Thus $f(p)\approx f(r)\approx F(r)\approx{}^*F(p)$ yields that $f|{}^*X\approx{}^*F|{}^*X$ and the proof is complete.

As usual, let $C(X,\mathbb{R})$ denote the set of all continuous real valued functions defined on X. For each $j\in\mathbb{N}$, $1\leqslant j\leqslant m$, $\mathscr{A}_j(X,\mathbb{R})$ is any subalgebra which separates points and contains some nonzero constant function. If $f:Z\rightarrow{}^*\mathbb{R}^m$, then $f=(f_1,\ldots,f_m)$ denotes f with its m component functions. Let $\mathscr{O}$ denote near standard members of ${}^*\mathbb{R}$. The set $\mathscr{O}$ is also termed the *limited* or *finite* members of ${}^*\mathbb{R}$. Further, $p\in{}^*\mathbb{R}^m$ is near standard if and only if $p\in\mathscr{O}^m$.

Theorem 1.2: Let X be Hausdorff. Suppose that $f\in{}^*(\mathscr{F}(X,\mathbb{R}^m))$ is microcontinuous on ${}^*X$ and that $f[{}^*X]\subset\mathscr{O}^m$. Then there is a function $G\in{}^*(\mathscr{F}(X,\mathbb{R}^m))$ that is $*$ continuous and microcontinuous on ${}^*X$ and for each $j\in\mathbb{N}$, $1\leqslant j\leqslant m$, each component function $G_j\in{}^*\mathscr{A}_j(X,\mathbb{R})$. Further, (i) f, G preserve nearness, (ii) f, G are infinitely close on ${}^*X$, and (iii) if for some standard g, $f={}^*g$, then $g=\mathrm{st}(G)$.

Proof: Consider $f=(f_1,\ldots,f_m)$. Then the function $f\in{}^*(\mathscr{F}(X,\mathbb{R}^m))$ is microcontinuous if and only if each component function $f_j\in{}^*(\mathscr{F}(X,\mathbb{R}))$ is microcontinuous and $f[{}^*X]\subset\mathscr{O}^m$ if and only if for each f_j, $f_j[{}^*X]\subset\mathscr{O}$. For each $j\in\mathbb{N}$, $1\leqslant j\leqslant m$, Theorem 1.1 implies that there exists an $F_j\in C(X,\mathbb{R})$ such that f_j, ${}^*F_j$ are infinitely close on ${}^*X$ and preserve nearness. For each $j\in\mathbb{N}$, $1\leqslant j\leqslant m$, consider the following internal binary relation:

$$B_j=\{(x,z)\,|\,(x\in{}^*\mathbb{R})\wedge(x>0)\wedge(z\in{}^*\mathscr{A}_j(X,\mathbb{R}))\wedge(\forall\omega(\omega\in{}^*X\rightarrow|{}^*F_j(\omega)-z(\omega)|<x))\}.$$

We now show that B_j is concurrent, at least, on the positive reals. Assume that

$$\{(x_1,z_1),\ldots,(x_k,z_k)\}\subset B_j,$$

where each x_i is a positive real number. Let $r=\min\{x_1,\ldots,x_k\}$. We know from the Stone–Weierstress theorem that there exists some $Q_j\in\mathscr{A}_j(X,\mathbb{R})$ such that for each $\omega\in X$, $|F_j(\omega)-Q_j(\omega)|<r$. By $*$ transfer, it follows that for each $\omega\in{}^*X$, $|{}^*F_j(\omega)-{}^*Q_j(\omega)|<r$. Consequently, $\{(x_i,{}^*Q_j)\,|\,1\leqslant i\leqslant k\}\subset B_j$ implies that B_j is concurrent, at least, on the positive reals. Saturation yields the existence of an internal $G_j\in{}^*\mathscr{A}_j(X,\mathbb{R})$ such that for arbitrary positive $r\in\mathbb{R}$, $|{}^*F_j(x)-G_j(x)|<r$ for each $x\in{}^*X$. Thus for each $x\in{}^*X$, ${}^*F_j(x)\approx G_j(x)$. Since every member of ${}^*\mathscr{A}_j(X,\mathbb{R})$ is $*$ continuous, then G_j is as well. Let x, $y\in{}^*X$ and $x\approx y$. The functions f_j, ${}^*F_j$ preserve nearness and are infinitely close on ${}^*X$. Consequently,

$${}^*F_j(x)\approx f_j(x)\approx f_j(y)\approx{}^*F_j(y).$$

Thus

$$G_j(x)\approx{}^*F_j(x)\approx f_j(x)\approx f_j(y)\approx{}^*F_j(y)\approx G_j(y)$$

yields that f_j, G_j preserve nearness; G_j is microcontinuous on ${}^*X$; and, since X is compact, f_j, G_j are infinitely close on ${}^*X$. Now simply define internal G: ${}^*X\rightarrow{}^*\mathbb{R}^m$ by setting $G=(G_1,\ldots,G_m)$ on ${}^*X$. Then $G\in{}^*(\mathscr{F}(X,\mathbb{R}^m))$ and (i) and (ii) hold.

Finally, assume that $f={}^*g$. Since X is compact then $\mathrm{st}({}^*X)=X$. Let $p\in X$ and $q\in\mu(p)$. Then, since ${}^*g$ and G preserve nearness on ${}^*X$, it follows that $G(q)\approx{}^*g(p)=g(p)$. Consequently, $\mathrm{st}(G(q))=g(p)=(\mathrm{st}(G))(p)$ and the proof is complete.

III. APPLICATIONS

There are many significant functions that satisfy the hypotheses of Theorem 1.2. We list a few examples. Assume that X is Hausdorff.

(i) For any $f\in C(X,\mathbb{R}^m)$ the function ${}^*f$ satisfies the hypotheses.

(ii) Assume that X is a metric space and that $Y\subset\mathbb{R}^m$ is compact. Consider any family $\mathscr{F}$ of equicontinuous functions defined on X into Y. Then each $f\in{}^*\mathscr{F}$ satisfies the hypotheses.

(iii) If X is a metric space, $Y=\mathbb{R}^m$, $\{f_n\}$ is a pointwise bounded sequence of continuous functions from X into $\mathbb{R}^m$, and $\{f_n\}$ converges uniformly to f; then each $g\in{}^*\{f_n\}$ satisfies the hypotheses.

When nonstandard mathematical structures are utilized to model natural system behavior certain rules of correspondence should be rigorously applied. In particular, a natural system process corresponds to standard mathematical entities or individuals. Nonstandard internal individuals or entities, or internal properties associated with nonstandard external entities, correspond to substratum objects or properties that may directly or indirectly effect natural system behavior, where the standard effects are either testable or yield observable data. The third category of properties corresponds directly to nonstandard external entities and, with certain well-known exceptions such as the standard part operator, are not assumed to effect directly natural system behavior, but, rather, are employed in discussions of pure substratum behavior, where their effects do not yield, directly or indirectly, either testable effects or observable data within a standard laboratory setting.[4] The usefulness of these external nonstandard entities as they correspond to a nonstandard type of physical model is conceptually the same as the usefulness of such physical language descriptions as the

Everett–Wheeler–Graham many-worlds interpretation.[5] These rules of correspondence tend to eliminate a considerable amount of *ad hoc* construction or definition, while yielding meaningful extensions of the standard natural system processes. Physical language models that follow these rules or correspondence are termed *nonstandard physical world models* or *NSP-world models*.

Under the NSP-world correspondence scheme, if an appropriate entity f possesses an internal nonstandard property and the standard object obtained by such processes as the standard part operator applied to f does not possess a corresponding standard property, then this is interpreted to mean that the internal nonstandard property is not detectable within the standard natural world after f is inserted into the natural system. This does not contradict the above rules of correspondence, for it is the direct or indirect effects that are either testable or yield observable data, while the property itself is not directly detectable.

In applications of nonstandard analysis to the behavior of a natural system, when a characterizing standard entity is infinitely close to an internal nonstandard entity, then the microeffects characterized by the nonstandard entity are conceived of as either a basic cause for the standard behavior or a process that sustains such behavior. This internal process is termed an *ultranatural process* and is hidden from direct observation within the standard laboratory environment. A major application of the procedures established within the present paper is relative to the concepts of the design and order that can be rationally assumed to influence the development of a natural system. The concept of ϵ differentiability, if corresponded to a nonstandard physical concept of smoothness, should probably not be considered as a hidden aspect of fractal behavior in lieu of other internal processes since ϵ differentiability is an external property. This situation is eliminated by application of the following theorem.

Theorem 3.1: Let $K \subset \mathbb{R}^n$ be compact. If $f \in {}^*(\mathscr{F}(K,\mathbb{R}^m))$ is microcontinuous on ${}^*K$ and $f[{}^*K] \subset \mathscr{O}^m$, then there exists a function $G \in {}^*(\mathscr{F}(\mathbb{R}^n,\mathbb{R}^m))$ such that the following holds.

(i) The function G is $*$ continuously $*$ differentiable on ${}^*\mathbb{R}^n$.

(ii) If $n = 1$, then G is $*$ continuously $*$ differentiable on ${}^*\mathbb{R}^n$ for any order $k \in {}^*\mathbb{N}$.

(iii) With respect to ${}^*K$, the function f and the restriction $G\,|{}^*K \in {}^*(\mathscr{F}(K,\mathbb{R}^n))$ preserve nearness.

(iv) The restriction $G\,|{}^*K$ is microcontinuous and uniformly S continuous on ${}^*K$.

(v) The function f and restriction $G\,|{}^*K$ are infinitely close on ${}^*K$.

(vi) If for standard g, $f = {}^*g$, then $g = \mathrm{st}(G\,|{}^*K)$.

Proof: Assume that K is compact, $f \in {}^*(\mathscr{F}(K,{}^*\mathbb{R}^m))$ is microcontinuous on ${}^*K$, and $f[{}^*K] \subset \mathscr{O}^m$. Let $\mathscr{P}(A,\mathbb{R})$ denote the algebra of all real valued polynomials in real coefficients in n variables considered as components and defined on $A \subset \mathbb{R}^n$. In general, by $*$ transfer, for each $j \in \mathbb{N}$, $1 \leqslant j \leqslant m$ if $P_j \in {}^*(\mathscr{P}(\mathbb{R}^n,\mathbb{R}))$; then the internal function defined on ${}^*\mathbb{R}^m$ by $P = (P_1,\ldots, P_m)$ satisfies the properties listed in (i) and (ii) of Theorem 3.1. Application of Theorem 1.2 implies that there exists a microcontinuous function $H \in {}^*(\mathscr{F}(K,\mathbb{R}^m))$ such that each component function $H_j \in {}^*(\mathscr{P}(K,\mathbb{R}))$ and f, H satisfy (i)–(iii) of that theorem. Moreover, since H is internal and microcontinuous on ${}^*K$, then H is uniformly S continuous on ${}^*K$.[3] However, by $*$ transform, each internal H_i may be extended to an internal $G_i \in {}^*(\mathscr{P}(\mathbb{R}^n,\mathbb{R}))$ by considering H_i defined on the internal domain ${}^*\mathbb{R}^n$. Thus $H_i = G_i\,|{}^*K$. Obviously, by letting $G = (G_1,\ldots, G_m)$ on ${}^*\mathbb{R}^m$ and $H = (H_1,\ldots, H_m)$ on ${}^*K$ it follows that G satisfies parts (i) and (ii) of Theorem 3.1 and that $H = G\,|{}^*K$ satisfies the remaining parts. This completes the proof.

When compared with the concept of ϵ differentiability the conclusions of Theorem 3.1 appear more significant for the following reasons. The function G is internal and thus has all of the first-order $*$ transfer properties which hold for the algebra $\mathscr{P}(\mathbb{R}^n,\mathbb{R})$. In the case that $n = 1$, the resulting m-dimensional curve $G[{}^*K] \subset \mathbb{R}^m$ has a well-behaved hyper-real length obtained by application of the $*$ integral operator and all of the $*$ transformed length properties. Moreover, in this case, since G is $*$ differentiable for any order $n \in {}^*\mathbb{N}$, then G is ultrasmooth. Since G is microcontinuous and uniformly S continuous on ${}^*K$ it will also satisfy all of the implied properties associated with these two significant concepts. From the NSP-world viewpoint, if the standard function is conceived of as representing a behavioral pattern produced by natural processes associated with a natural system, whether or not it is a fractal type, then G can be conceived of as a highly ordered, smooth, regular, and even a somewhat conventional ultranatural process that when applied to an associated ultranatural system, yields what may be perceived within the laboratory environment to be an irregular or even chaotic pattern of behavior.

However, it is obvious that G is not unique and that other distinctly different algebras would generate a function G_0 having all the properties listed in Theorem 3.1, as well others not shared by G. Although for this analysis it is clearly unnecessary, it may be that there are specific algebras that could be more closely associated with specific types of fractal behavior. The formal infinitesimal analysis that establishes that certain phenomena, for example, particle trajectories in quantum mechanics,[6] are fractal in character tends to utilize members of well-behaved algebras which also share the properties stated Theorem 3.1.

IV. STEP FUNCTIONS

As a preliminary to this section, we introduce the following definition. A function $f:[a,b] \to \mathbb{R}^m$ is *differentiable* C on $[a,b]$ if it is continuously differentiable on (a,b), except at finitely many removable discontinuities. This definition is extended to the end points $\{a,b\}$ by application of one-sided derivatives. For any $[a,b]$ consider a partition $P = \{a_0,a_1,\ldots,a_n,a_{n+1}\}$, $n \geqslant 1$, $a = a_0$, $b = a_{n+1}$, and $a_{j-1} < a_j$, $1 \leqslant j \leqslant n+1$. For any such partition P let the real valued function g be defined on the set $D = [a_0,a_1) \cup (a_1,a_2) \cup \cdots \cup (a_n,a_{n+1}]$ as follows: For each $x \in [a_0,a_1)$, let $g(x) = r_1 \in \mathbb{R}$; for each $x \in (a_{j-1},a_j)$, let $g(x) = r_j \in \mathbb{R}$, $1 < i \leqslant n$; and for each $x \in (a_n,b]$, let $g(x) = r_{n+1} \in \mathbb{R}$. It is obvious that g is a type of simple step function.

Theorem 4.1: There exists a function $G \in {}^*(\mathscr{F}([a,b],\mathbb{R}))$ with the following properties.

(i) The function G is $*$ continuously $*$ differentiable and $*$ uniformly $*$ continuous on ${}^*[a,b]$.

(ii) For each odd $n \in {}^*\mathbb{N}$, $(n \geqslant 3)$, G is $*$ differentiable C of order n on ${}^*[a,b]$.

(iii) For each even $n \in {}^*\mathbb{N}$, G is $*$ continuously $*$ differentiable in ${}^*[a,b]$, except at finitely many points.

(iv) If $c = \min\{r_1,\ldots, r_{n+1}\}$, $d = \max\{r_1,\ldots, r_{n+1}\}$, then the range of $G = {}^*[c,d]$, $\mathrm{st}(G)$ at least maps D into $[c,d]$ and $(\mathrm{st}\,(G))|D = g$.

Proof: First, for any real c,d, where $d \neq 0$, consider the finite set of functions

$h_j(x,c,d) = \frac{1}{2}(r_{j+1} - r_j)(\sin((x-c)\pi/(2d) + 1)) + r_j$,

$1 \leqslant j \leqslant n$. Each h_j is continuously differentiable for any order at each $x \in \mathbb{R}$. Observe that for each odd $m \in \mathbb{N}$, each mth derivative $h_j^{(m)}$ is continuous at $(c+d)$ and $(c-d)$ and $h_j^{(m)}(c+d) = h_j^{(m)}(c-d) = 0$ for each j.

Let positive $\delta \in \mu(0)$. Consider the finite set of internal intervals $\{[a_0,a_1-\delta),(a_1+\delta,a_2-\delta),\ldots, (a_n+\delta,b)\}$ obtained from the partition P. Denote these intervals in the expressed order by I_j, $1 \leqslant j \leqslant n+1$. Define the internal function

$$G_1 = \{(x,r_1)|x \in I_1\} \cup \cdots \cup \{(x,r_{n+1})|x \in I_{n+1}\}.$$

Let internal $I_j^\dagger = [a_j - \delta, a_j + \delta]$, $1 \leqslant j \leqslant n$ and for each $x \in I_j^\dagger$, let internal

$G_j(x) = \frac{1}{2}(r_{j+1} - r_j)({}^*\sin((x-c)\pi/(2\delta) + 1)) + r_j$.

Define the internal function

$$G_2 = \{(x,G_1(x))|x \in I_1^\dagger\} \cup \cdots \cup \{(x,G_n(x))|x \in I_n^\dagger\}.$$

The final step is to define $G = G_1 \cup G_2$. Then $G \in {}^*(\mathscr{F}([a,b],\mathbb{R}))$.

By $*$ transfer, the function G_1 has an internal $*$ continuous $*$ derivative $G_1^{(1)}$ such that $G_1^{(1)}(x) = 0$ for each $x \in I_1 \cup \cdots \cup I_{n+1}$. Applying $*$ transfer to the properties of the functions $h_j(x,c,d)$, it follows that G_2 has a unique internal $*$ derivative

$$G_2^{(1)} = (1/(4\delta))(r_{j+1} - r_j)\pi({}^*\cos((x-a_j)\pi/(2\delta)))$$

for each $x \in I_1^\dagger \cup \cdots \cup I_n^\dagger$. The results that the $*$ left limit for the internal $G_1^{(1)}$ and the $*$ right limit for internal $G_2^{(1)}$ at each $a_j - \delta$, as well as the $*$ left limit of $G_2^{(1)}$ and the $*$ right limit of $G_1^{(1)}$ at each $a_j + \delta$ are equal to zero and $0 = G_2^{(1)}(a_j - \delta) = G_2^{(1)}(a_j + \delta)$ imply that internal G has a $*$ continuous $*$ derivative $G^{(1)} = G_1^{(1)} \cup G_2^{(1)}$ defined on ${}^*[a,b]$.

A similar analysis and $*$ transfer yield that for each $m \in {}^*\mathbb{N}$, $m \geqslant 2$, G has an internal $*$ continuous $*$ derivative $G^{(m)}$ defined at each $x \in {}^*[a,b]$, except at the points $a_j \pm \delta$ whenever $r_{j+1} \neq r_j$. However, it is obvious from the definition of the functions h_j that for each odd $m \in {}^*\mathbb{N}$, $m \geqslant 3$, each internal $G^{(m)}$ can be made $*$ continuous at each $a_j \pm \delta$ by simply defining $G^{(m)}(a_j \pm \delta) = 0$; with this parts (i)–(iii) of Theorem 4.1 are established.

For part (iv) of Theorem 4.1 assume that $r_j \leqslant r_{j+1}$. From the definition of the functions h_j it follows that for each $x \in I_j \cup I_j^\dagger \cup I_{j+1}$, $r_j \leqslant G(x) \leqslant r_{j+1}$. The nonstandard intermediate value theorem implies that $G[{}^*[a_j,a_{j+1}]] = {}^*[r_j,r_{j+1}]$ and, in like manner, when $r_{j+1} < r_j$. Hence, $G[{}^*[a,b]] = {}^*[c,d]$. Clearly, $\mathrm{st}(D) = [a,b]$. If $p \in D$ and $x \in \mu(p)$, then $G(x) = r_j = g(p)$ for some j such that $1 \leqslant j \leqslant n+1$. This completes the proof.

The nonstandard approximation Theorem 4.1 can be extended easily to functions that map D into $\mathbb{R}^m$. For example, assume that $F: D \to \mathbb{R}^3$ and the component functions F_1, F_2 are continuously differentiable on $[a,b]$, but that F_3 is a g-type step function on D. Then letting $H = ({}^*F_1, {}^*F_2, G)$ on ${}^*[a,b]$, where G is defined in Theorem 4.1, we have an internal $*$ continuous $*$ differentiable function H: ${}^*[a,b] \to {}^*\mathbb{R}^3$, with the property that $\mathrm{st}(H)|D = F$.

Propositions such as Theorem 4.1 are being employed to model the behavior of natural objects which appear to alter suddenly some numerically expressed characterizing property.

V. CONCLUSION

Nonstandard methods utilizing polysaturated enlargements as applied to the study of fractals apparently have some significance relative to the design, order, and existence of possible microeffects. These methods greatly improve upon previous results that associated NSP-world smooth approximating functions with parametrized fractal curves. The selection of nonstandard smooth approximating functions from different algebras may also prove to be significant when analysis establishes that a pattern of behavior may be characterized by a fractal. Moreover, these nonstandard approximating functions have a very well-behaved length which satisfies all of the known first-order properties of the length of a curve as obtained by application of the standard Riemann integral. It is believed that these nonstandard methods under various generalizations and extensions will lead eventually to a better understanding of the underlying processes that generate the behavior of natural systems as they are perceived within the laboratory environment.

[1]L. Nottale and J. Schneider, J. Math. Phys. **25**, 1296 (1984).

[2]K. D. Stroyan and W. A. J. Luxemburg, *Introduction to the Theory of Infinitesimals* (Academic, New York, 1976).

[3]M. Davis, *Applied Nonstandard Analysis* (Wiley–Interscience, New York, 1977).

[4]R. A. Herrmann, *The NSP-world Model* (IMP, Annapolis, 1987).

[5]H. Everett, III, Rev. Mod. Phys. **29**, 454 (1957).

[6]L. F. Abbott and M. B. Wise, Am. J. Phys. **49**, 37 (1981).

The geometry of the space of null geodesics

R. J. Low
Mathematics Department, Coventry Polytechnic, Priory Street, Coventry, CV1 5FB, England

(Received 11 October 1988; accepted for publication 14 December 1988)

The topology and geometry of the space of null geodesics N of a space-time M are used to study the causal structure of the space-time itself. In particular, the question of whether the topology of N is Hausdorff or admits a compatible manifold structure carries information on the global structure of M, and the transversality properties of the intersections of skies of points tell whether the points are conjugate points on a null geodesic.

I. INTRODUCTION

The causal structure of space-time—which is modeled as a Lorentz manifold M—is of great importance, both physically (in the question of whether data at one point of M can affect what happens at another) and philosophically (owing to the paradoxes inherent in the existence of closed timelike curves).[1,2] Since the causal structure of a space-time is intimately connected with the null geodesics of that space-time, one might suspect that interesting information about the causal structure of M is coded naturally in the structure of the space of null geodesics, say N.

In Sec. II of this paper, we consider N both as a topological space, and, when appropriate, as a smooth manifold. There are two natural but slightly different (though intimately related) ways of inducing a topology on N, and we will see that these both actually induce the same topology on N. Furthermore, it is not necessarily the case that N admit any manifold structure compatible with its topology, as a simple example shows. A sufficient condition for N to have a natural manifold structure is that M be strongly causal, although this condition is not necessary. If this topology is non-Hausdorff, then M must have a naked singularity of a particularly naked type.

In Sec. III we restrict ourselves to the case where M is strongly causal, and N is therefore a manifold (though not necessarily Hausdorff). It is noted that a vector field on M induces a vector field on N if it is a conformal Killing vector field, and that a vector at a point of N gives a vector field on a null geodesic of M connecting it to a neighboring null geodesic. In this case, one can show that given any point $x \in M$, the corresponding subset X of N, called the sky of x, is a smoothly embedded S^2 in N.

Finally, in Sec. IV, the results of Sec. III are used to give an interpretation of the conjugacy of two points on a null geodesic in terms of N, and some related ideas.

II. THE TOPOLOGICAL STRUCTURE OF N

Let M be a space-time, i.e., a C^∞ manifold with a Lorentz metric. Denote its tangent bundle by TM, the reduced tangent bundle $TM \setminus \{\text{zero section}\}$ by $T'M$, and the fiber above a point x by T_xM. Then $T'M$ has the quotient space UM given by identifying proportional vectors at each point x in M, and UM is foliated by the natural lifts of the geodesics of M. Note that although a geodesic of M may fail to be a submanifold of M, its lift to UM is always a submanifold of UM. The set G, the space of geodesics of M, is given by identifying points in UM that lie on the lift of a common geodesic. Then G naturally splits up into the disjoint union of G^+, N, and G^-, which are the spaces of timelike, null, and spacelike geodesics, respectively; removal of N from G disconnects G, and N is the common boundary of G^+ and G^-. We can regard N as this subset of G, with the subspace topology induced by the quotient topology from UM.

Alternatively, we could restrict from UM, the bundle of directions in M, to NM, the bundle of null directions, then pass to the quotient space in the same way; denoting the quotient map from UM to G by p, it follows[3] from the facts that $p^{-1}(p(NM)) = NM$ and NM is closed in UM that we obtain the same topology in either case.

Any point in N, then, corresponds to a curve in M, namely the null geodetic curve whose lift it is the projection of; also any point in M gives a subset of N, namely the set of all null geodesics passing through it. A useful notational convention that will be adopted here is to use lower case roman letters to represent points of M, and the corresponding upper case letter to represent the subset of N determined by it, and to use lower case greek letters to represent points of N, and the corresponding upper case greek letters to represent the null geodesics in M that they determine.

First, then, we note that N need not have any manifold structure compatible with its topology.

Example 2.1: Let M be Minkowski space, with the usual coordinates (t,x,y,z), and let $\mathbb{Z}\times\mathbb{Z}$ act on M by

$$(m,n)\cdot(t,x,y,z) = (t + m\sqrt{2}, x + n, y, z).$$

Then the quotient space $M/\mathbb{Z}\times\mathbb{Z}$ has the topology $S^1\times S^1\times\mathbb{R}^2$, but the ratio of the lengths of the S^1's is $\sqrt{2}$. Thus any null geodesic in the (t,x) plane becomes a dense curve in the $y = z = 0$ torus of $M/\mathbb{Z}\times\mathbb{Z}$ and so any neighborhood of such a null geodesic will actually contain all the other geodesics in the $y = z = 0$ torus. Thus the space of null geodesics of this space-time fails to satisfy the separation axiom T_1, and so the topology cannot be induced by a differentiable structure.[4] □

On the other hand, if the space-time is strongly causal—as any space-time forming a reasonable model of the universe must be—then this cannot happen, as follows from Proposition 2.1

Proposition 2.1: Let M be strongly causal. Then N is naturally a smooth manifold with a C^∞ structure inherited from N^*M.

Proof (See Ref. 4 for notation)*:* Let $u \in N^*M$, and let (U,x) be a flat chart containing u. Then $\pi(U)$ is an open set

 0022-2488/89/040809-03$02.50

containing $\pi(u)$, so by strong causality there is an open subset V of $\pi(U)$ such that $\pi(u)$ lies in V, and any causal curve intersects V in a single connected component. Then the lift of any null geodesic in M to N^*M will intersect $\pi^{-1}(V)$ in a single connected component, and hence W, defined by $W=\pi^{-1}(V)\cap U$, gives a flat chart $(W,x|W)$ containing u, and such that any leaf of the foliation intersects W in a single connected component. Hence (Ref. 4, pp. 202ff) N inherits the structure of a C^∞ manifold from N^*M, and $p: N^*M\to N$ is a smooth submersion. □

As we can see from the proof, strong causality is not necessary. In fact the necessary condition is simply that no null geodetic curve Γ may be such that there is a sequence of points, each further along Γ than the preceding one, which has as a limit some earlier point of Γ, and is such that the tangents at these points tend to the tangent at that limit point. In other words, no null geodesic may approach itself arbitrarily closely and tangentially. Indeed it may even have self-intersections as long as they are transverse, without destroying the manifold structure of N. However, strong causality is such a natural condition that it is the one which will be used here to guarantee a manifold structure on N.

Even under these circumstances, N may fail to be Hausdorff as a topological space.

Example 2.2: Let M be Minkowski space minus the origin. Then N is given by taking two copies of $\mathbb{R}^3\times S^2$ and identifying (x,p) in one copy with (y,q) in the other whenever $x=y\neq(0,0,0)$ and $p=q$, for $x,y\in\mathbb{R}^3$, $p,q\in S^2$, which gives a non-Hausdorff manifold. □

In the case where M is strongly causal, we can use the technology of ideal points[5] to study the consequences of N being non-Hausdorff.

Proposition 2.2: Let M be a strongly causal space-time, and N its space of null geodesics. Then if N is non-Hausdorff, M must be nakedly singular.

Proof: Let $\gamma_1,\gamma_2\in N$ be such that whenever U_i is a neighborhood of γ_i for $i=1,2$, $U_1\cap U_2\neq\varnothing$. Then one can construct a sequence $\{\psi_n\}_{n>0}$. with the property that γ_1 and γ_2 are both limit points of $\{\psi_n\}_{n>0}$.

Then in M there is a sequence of null geodesic curves $\{\Psi_n\}_{n>0}$ which approaches both Γ_1 and Γ_2 as n increases. Now, let $x_i\in\Gamma_i$ and V_i be an open neighborhood of x_i (for $i=1,2$) such that $V_1\cap V_2=\varnothing$. Without loss of generality, it can be assumed that each Ψ_n intersects both V_1 and V_2 for $n>0$. Next, define sequences $\{c^i_n\}_{n>0}$ for $i=1,2$, such that $c^i_n\in V_i$ for $n>0$, and c^i_n approaches x_i as n approaches infinity. We can also assume that $c^2_n\in J^+(c^i_n)$ for $n>0$, by an appropriate choice of V_1 and V_2.

Now let $z\in I^+(x_2)$. Then $I^-(z)$ is an open set containing x_2 and so for n large enough, $c^2_n\in I^-(z)$, and therefore $c^1_n\in\overline{I^-(z)}$. But this implies that $x_1\in\overline{I^-(z)}$, since $\overline{I^-(z)}$ is closed.

Next, let $w\in I^-(x_1)$. Then $I^+(w)$ is a neighborhood of x_1, and so must intersect $I^-(x_2)$, which in turn implies that $w\in I^-(z)$. Finally, we observe that this is independent of the choice of the point x_1 on Γ_1, and so any point in $I^-(\Gamma_1)$ lies in $I^-(z)$, thus $I^-(\Gamma_1)\subset I^-(z)$.

But Γ_1 is a future endless null geodesic, and so $I^-(\Gamma_1)$ is a terminal indecomposable past set—abbreviated to TIP—lying inside a proper indecomposable past set, a PIP. (See Ref. 5 for the definitions and a discussion of the importance of these objects, and the dual future sets, TIF's and PIF's.) Hence M is nakedly singular. □

Note that by a dual argument, $I^+[\Gamma_2]\subset I^+(z')$ for some $z'\in M$, so we also have a TIF that is a subset of a PIF. The converse of this result is, however, false: for the subspace of Minkowski space with the standard coordinates given by $x^2+y^2+z^2<1$ is nakedly singular, but its space of null geodesics is Hausdorff (being a subspace of $\mathbb{R}^3\times S^2$, which is clearly Hausdorff).

Now, one can define an open neighborhood of the ideal point given by $I^-[\Gamma_1]$ as an open set in M that contains all the points of Γ_1 to the future of some point on Γ_1. Then in the notation used in the proof of Proposition 2.2, if U is a neighborhood of the ideal point given by $I^-[\Gamma_1]$, then there is some null geodesic Ψ_n which enters U, and eventually reaches c^2_n. Thus a geodesic observer can get arbitrarily close to the singularity, and get away again, without falling in. The point of this is not only that the singularity can be approached and then left again, but that this can actually be done with no acceleration.

One final point about the singularity that causes N to be non-Hausdorff is that if γ_1 is future complete, then $I^-[\Gamma_1]$ is an ∞-TIP, i.e., represents an ideal point at infinity, which the members of $\{\Psi_n\}_{n>0}$ approach arbitrarily closely before returning to a neighborhood of x_2. So in this sense, there are null geodesics in M that can go arbitrarily far away, and return to the "interior" of M. Thus one can get arbitrarily near infinity, and return to an arbitrarily small neighborhood of x_2, while remaining on a causal path (and, again, without undergoing any acceleration).

One particular example of a space-time for which this non-Hausdorff condition holds is that of a plane wave space-time.[6] The above considerations show that the plane wave space-time does not have a simple asymptotic structure, and that it is, therefore, difficult to use asymptotic techniques to try to study the mass energy carried by the wave.

Corollary 2.1: If M is strongly causal, then for any point x in M, $X\subset N$, the sky of x is a smooth S^2 in N.

Proof: NM is the bundle of null directions over M, and let N_xM—which is topologically S^2—be the fiber of NM over x. Then if p: $NM\to N$ is the projection, X is $p(N_xM)$, and since the fibers of NM lie inside those of UM, which are transverse to the geodesic flow, it follows that p is regular at all points of N_xM. Hence $p|N_xM$ is a smooth, regular bijection, and therefore a diffeomorphism. □

III. VECTORS AND VECTOR FIELDS ON *M* AND *N*

First, we observe that a vector field V on M will project to one on N precisely when its flow ϕ_t preserves null geodesic curves; but this is just the condition that V be a conformal Killing vector field, since the conformal motions of M are those preserving the causal structure and hence the null geodesics. On the other hand, a vector field on N will correspond to a vector field on M precisely when its flow preserves the skies of points.

More interestingly, one can consider the interpretation in M of a vector at a point of N. So let v be a vector in $T_\gamma N$, i.e., v is the tangent to some curve through γ. Then v is the projection of a vector field along Γ in M that connects points of Γ to points of a neighboring null geodesic; in other words it is the projection of a Jacobi field on Γ. Furthermore, any Jacobi field on Γ given by a one-parameter family of null geodesics containing Γ will project to a vector at γ. Call such a Jacobi field a null Jacobi field.

Now, fix some point x on Γ, and make a smooth choice of affine parameter on all the null geodesics through x. Then there is a two-parameter family of null Jacobi fields along Γ vanishing at x given this choice of affine parameter on the null geodesics passing through x, and this family precisely describes $T_\gamma N$. This relationship can be exploited to give an interpretation in N of the conjugacy of two points in M along a null geodesic.

First, we develop a little notation. Two points, x and y, lying on a common null geodesic Γ are said to be null conjugate of degree n if there are n linearly independent null Jacobi fields along Γ vanishing at both x and y. Note that two points in space-time can be null conjugate of degree at most two. (It is easy to see that in a strongly causal n-dimensional Lorentz manifold, two points can be null conjugate of degree at most $n-2$.)

Then we can give the following classification of points that lie on a common null geodesic. Let $x,y\in M$, and let $\gamma\in N$ such that $x,y\in\Gamma$. Then $\gamma\in X\cap Y$, and we have the following cases.

Classification 3.1:

(1) $T_\gamma X\cap T_\gamma Y=\{0\}$: x and y are not conjugate along Γ,

(2) $T_\gamma X\cap T_\gamma Y=\{tv\colon t\in\mathbb{R}\}$ for some $v\in T_\gamma N$: x and y are conjugate of degree one along Γ, and v is the projection of some Jacobi field that vanishes at both x and y.

(3) $T_\gamma X=T_\gamma Y$: x and y are conjugate of degree 2 along Γ, and any element of $T_\gamma X$ gives a null Jacobi field along Γ with a representative Jacobi field that vanishes at both x and y. □

One can also give a classification of vectors at a point of N in similar terms—let $\gamma\in M$: then we have the following classification.

Classification 3.2:

(1) $v\notin T_\gamma X$ for any $x\in\Gamma$: so v gives a null Jacobi field connecting Γ to neighboring null geodesics that never meet Γ.

(2) $v\in T_\gamma X$ for precisely one $x\in\Gamma$: v gives a null Jacobi field connecting Γ to neighboring null geodesics through x that never meet Γ again; i.e., $\Gamma\cap I(x)=\varnothing$.

(3) $v\in\bigcap_{i=1}^{n} T_\gamma X_i$: the points $x_i\cdots x_n$ on the null geodesic Γ are all conjugate to each other along Γ by a single Jacobi field. □

A point that is conjugate to a spacelike two-surface can be described in a similar way. Let $S\subset M$ be a spacelike two-surface; then S defines a two-dimensional surface in N, as follows. Let $x\in S$; then there are precisely two future pointing null directions that are orthogonal to S at x. As x varies over S, we obtain a subset Σ of N (which will be diffeomorphic to the disjoint union of two copies of S if S is orientable, or a double cover otherwise).

Now, similarly to the definition of null conjugacy, we say that x is conjugate to S along Γ if

(1) $x\in\Gamma$ and $\gamma\in\Sigma$,

(2) there is a nontrivial Jacobi field along Γ that vanishes at x, defined by a one-parameter family of null geodesics in Σ, the orthogonal congruence to S.

Then we obtain the following classification.

Classification 3.3: Let $x\in M$ lie on the null geodetic curve Γ, where $\gamma\in\Sigma$, the orthogonal congruence to S, a spacelike two-surface in M. Then x is conjugate to S along Γ if $T_\gamma X\cap T_\gamma\Sigma$ is nontrivial, and the degree of conjugacy is the dimension of $T_\gamma X\cap T_\gamma\Sigma$. □

Note that if $x\in S$, then $X\cap\Sigma=\{\gamma_1,\gamma_2\}$, where γ_1 and γ_2 are the two null geodesics through x which are orthogonal to S, and $T_{\gamma i}X=T_{\gamma i}\Sigma$, for $i=1,2$. However, the converse need not be true—it could be the case that two null geodesics in Σ will focus at some $y\in M\setminus S$, and y happens to be conjugate of degree 2 to S along each null geodesic. In fact, this can even happen in Minkowski space, by a careful choice of S.

IV. CONCLUSIONS

By examining the space of null geodesics N of a space-time M, one can obtain useful information about the causal structure of M; in particular, non-Hausdorffness of the topology of N tells us that M is nakedly singular in such a way that a singularity can be approached and then left without any acceleration, and the lack of a manifold topology tells us that M cannot be strongly causal. When M is strongly causal, and N is therefore a manifold, the skies of points in M are smooth S^2's in N, and two points on a common null geodesic are null conjugate along that geodesic precisely when the intersection of their skies is nontransversal.

ACKNOWLEDGMENTS

I would like to thank N. M. J. Woodhouse and G. B. Segal for helpful comments and for pointing out the inadequacy of an earlier version of the proof of Proposition 2.2. Part of this work was done with the aid of an SERC Research Studentship.

[1]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[2]R. Penrose, *Techniques of Differential Topology in Relativity*, CBMS Regional Conf. Ser. in Appl. Math., No. 7 (SIAM, Phaladelphia, 1972).
[3]R. Brown, *Elements of Modern Topology* (McGraw-Hill, London, 1968).
[4]F. Brickell and R. S. Clark, *Differentiable Manifolds an Introduction* (Van Nostrand Reinhold, London, 1970).
[5]R. Geroch, E. H. Kronheimer, and R. Penrose, Proc. R. Soc. London Ser. A **327**, 545 (1972).
[6]R. Penrose, Rev. Mod. Phys. **37**, 215 (1965).

A note on super Fock space

M. J. Bergvelt
Max-Planck-Institut-für-Mathematik, Gottfried-Claren-strasse 26, 5300 Bonn 03, Federal Republic of Germany

(Received 12 July 1988; accepted for publication 30 November 1988)

An explicit realization is given in terms of even and odd differential operators of certain highest weight representations for the affine Lie superalgebra $\widetilde{gl}_{1/1}$.

I. INTRODUCTION

Kac and van de Leur introduced in Ref. 1 a collection of highest weight modules V_m, $m\in\mathbb{Z}$, over the infinite-dimensional general linear Lie superalgebra $a_{\infty|\infty}$. These modules remain irreducible under the action of the principal subalgebra, isomorphic to the affine superalgebra $\widetilde{gl}_{1/1}$ ("super Heisenberg algebra").

In the nonsuper case, the irreducibility of a module, say V, under a Heisenberg algebra leads to a realization of V as a bosonic Fock space, i.e., V is realized as a polynomial algebra, and the action of the Lie algebra is given in terms of differential operators. This construction has many applications, especially in the theory of soliton equations and in string theory.

In this paper we show similarly that the modules V_m of Kac and van de Leur can be realized as "super Fock spaces", i.e., as a tensor product of a polynomial algebra and an exterior algebra. For representations of other Lie superalgebras this has been observed by Golitzin.[2] One might hope that this construction has similar applications as in the even case.

II. THE LIE SUPERALGEBRA $\widetilde{gl}_{1/1}$ AND THE MODULES V_m

In this section we recall some definitions and results from Ref. 1. We take as a basis for $gl_{1/1}$:

$$\lambda=\begin{pmatrix}-1 & 0\\ 0 & 0\end{pmatrix},\quad \mu=\begin{pmatrix}0 & 0\\ 0 & -1\end{pmatrix},$$
$$X_+=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},\quad X_-=\begin{pmatrix}0 & 0\\ -1 & 0\end{pmatrix}. \tag{2.1}$$

We define for $x\in gl_{1/1}$

$$x(n):=t^n x\in gl_{1/1}(\mathbb{C}[t,t^{-1}]). \tag{2.2}$$

The commutator in $\tilde{g}l_{1/1}=gl_{1/1}(\mathbb{C}[t,t^{-1}])\oplus\mathbb{C}c$ is given by

$$[x(n),y(n)]=[x,y](m+n)+m\delta_{m+n,0}\,\mathrm{Str}(xy)c, \tag{2.3}$$

where Str is the supertrace.

Explicitly we have

$$\begin{aligned}
[\lambda(m),X_\pm(n)]&=\mp X_\pm(m+n),\\
[\mu(m),X_\pm(n)]&=\pm X_\pm(m+n),\\
[\lambda(m),\lambda(n)]&=m\delta_{m+n,0}c,\\
[\mu(m),\mu(n)]&=-m\delta_{m+n,0}c,\\
[X_+(m),X_-(n)]&=(\lambda+\mu)(m+n)-m\delta_{m+n,0}c,\\
[X_\pm(m),X_\pm(n)]&=0,\\
[\lambda(m),\mu(n)]&=0.
\end{aligned} \tag{2.4}$$

Note that $\{\lambda(m),m\neq 0\}$ and $\{\mu(m),\neq 0\}$ generate infinite Heisenberg algebras.

We consider representations π_m on V_m with highest weight vector $|m\rangle$ that satisfy

$$\begin{aligned}
\pi_m(X_+(n))|m\rangle&=0,\quad n\geqslant 0,\\
\pi_m(X-(n+1))|m\rangle&=0,\quad n\geqslant 0,\\
\pi_m(\lambda(n))|m\rangle&=0,\quad n>0,\\
\pi_m(\mu(n))|m\rangle&=0,\quad n>0,\\
\pi_m(c)|m\rangle&=|m\rangle,\\
\pi_m(\lambda(0))|m\rangle&=\begin{cases}0, & m\geqslant 0,\\ -|m\rangle, & m<0,\end{cases}\\
\pi_m(\mu(0))|m\rangle&=\begin{cases}m|m\rangle, & m\geqslant 0,\\ (m+1)|m\rangle, & m\leqslant 0.\end{cases}
\end{aligned} \tag{2.5}$$

Introduce a gradation on $\tilde{g}l_{1/1}$ by

$$\begin{aligned}
\deg\lambda(n)&=2n,\\
\deg\mu(n)&=2n,\\
\deg X_\pm(n)&=2n\pm 1,\\
\deg c&=0.
\end{aligned} \tag{2.6}$$

Then the modules V_m have a q dimension given by

$$\dim_q V_m=\begin{cases}\dfrac{1}{1+q^{2m+1}}\displaystyle\prod_{k>1}\frac{(1+q^{2k-1})^2}{(1-q^{2k})^2}, & \text{if } m\geqslant 0,\\[2ex] \dfrac{1}{1+q^{-2m-1}}\displaystyle\prod_{k>1}\frac{(1+q^{2k-1})^2}{(1-q^{2k})^2}, & \text{if } m<0.\end{cases} \tag{2.7}$$

III. VACUUM SPACE AND Z OPERATORS

Define the vacuum space of V_m by

$$\Omega_{V_m}=\{v\in V_m\,|\,\pi_m(\lambda(n))v=\pi_m(\mu(n))v=0,\ \forall n>0\}. \tag{3.1}$$

By general results on the modules V_m we have

$$V_m\simeq\mathbb{C}[x_{2i},y_{2j},\,i,j>0]\otimes\Omega_{V_m}. \tag{3.2}$$

(see, e.g., Golitzin,[2] or in the nonsuper case Lepowsky and Wilson,[3]) where $\lambda(n)$ and $\mu(n)$ act as $\partial/\partial x_{2n},\partial/\partial y_{2n}$ for $n>0$, and as nx_{2n}, $-ny_{2n}$ for $n<0$, respectively. From Eq. (2.6) it follows that the q dimension of the vacuum space is given by

 0022-2488/89/040812-04\$02.50

$$\dim_q \Omega_{V_m} = \begin{cases} \dfrac{1}{1+q^{2m+1}} \prod_{k\geqslant 1} (1+q^{2k-1})^2, & m\geqslant 0, \\ \dfrac{1}{1+q^{-2m-1}} \prod_{k\geqslant 1} (1+q^{2k-1})^2, & m<0. \end{cases} \tag{3.3}$$

We will construct operators that map Ω_{V_m} into itself. To this end we introduce formal series

$$X_\pm(z) = \sum_{k\in Z} z^{-(2k\pm 1)} X_\pm(k). \tag{3.4}$$

Then we have

$$\begin{aligned} [\lambda(n),X_\pm(z)] &= \mp z^{2n} X_\pm(z), \\ [\mu(n),X_\pm(z)] &= \pm z^{2n} X_\pm(z). \end{aligned} \tag{3.5}$$

Define

$$E_\pm^\epsilon(z) := \exp\left\{\mp\epsilon \sum_{k>0} \frac{(\lambda+\mu)(\pm k)}{k} z^{\mp 2k}\right\},$$

$$\epsilon = \pm 1, \tag{3.6}$$

and

$$Z_\pm(z) := E_\pm^{\pm 1}(z) X_\pm(z) E_\mp^{\pm 1}(z). \tag{3.7}$$

Lemma 3.1: The homogeneous components of $Z_\pm(z)$ map Ω_{V_m} into itself.

Proof: Using

$$[A,B] = k\,1 \Rightarrow [A,\exp(B)] = k\exp(B), \tag{3.8}$$

we have

$$\begin{aligned} [\lambda(n),E_-^\epsilon(z)] &= \epsilon z^{2n} E_-^\epsilon(z), \\ [\mu(n),E_-^\epsilon(z)] &= -\epsilon z^{2n} E_-^\epsilon(z), \\ [\lambda(-n),E_+^\epsilon(z)] &= \epsilon z^{2n} E_+^\epsilon(z), \\ [\mu(-n),E_+^\epsilon(z)] &= -\epsilon z^{2n} E_+^\epsilon(z), \\ [\lambda(n),E_+^\epsilon(z)] &= 0, \\ [\mu(n),E_+^\epsilon(z)] &= 0, \\ [\lambda(-n),E_-^\epsilon(z)] &= 0, \\ [\mu(-n),E_-^\epsilon(z)] &= 0, \\ & n>0, \quad \epsilon = \pm 1. \end{aligned} \tag{3.9}$$

Combining (3.5), (3.7), and (3.9) we find

$$\begin{aligned} [\lambda(n),Z_\pm(z)] &= 0, \quad \forall n\neq 0. \\ [\mu(n),Z_\pm(z)] &= 0, \end{aligned} \tag{3.10}$$

Expanding $Z_\pm(z)$ in powers of z:

$$Z_\pm(z) = \sum_{k\in Z} Z_\pm(k) z^{-(2k\pm 1)}, \tag{3.11}$$

it is clear that the $Z_\pm(k)$ commute with the $\lambda(n),\mu(n)$, $\forall n\neq 0$. We leave it to the reader to show that the $Z_\pm(k)$ are well-defined operators on V_m. This proves the lemma. ∎

We will refer to the $Z_\pm(z)$ [or also to their components $Z_\pm(k)$] as Z operators.

IV. THE COMMUTATION RELATIONS OF THE Z OPERATORS

In this section we calculate the commutation relations of the Z operators in the modules V_m.

Theorem 4.1: In V_m we have

$$[Z_+(k),Z_-(l)] = (l+m)\delta_{k+l,0}. \tag{4.1}$$

Proof: We need a number of lemmas.

Lemma 4.2:

(a) $[E_\pm^\epsilon(z),E_\mp^{\hat\epsilon}(z)] = 0, \quad \epsilon,\hat\epsilon = \pm 1,$

(b) $[X_\pm(z),E_\pm^\epsilon(z)] = [X_\pm(z),E_+^\epsilon(z)] = 0.$

Proof of Lemma 4.2: (a) This follows from the fact that $\{(\lambda+\mu)(k),\ k\in\mathbf{Z}\}$ is an Abelian subalgebra [see Eq. (2.4)]. (b) Here one uses

$$[(\lambda+\mu)(k),X_\pm(n)] = 0, \quad \forall k,n\in\mathbf{Z}. \tag{4.2}$$

∎

Define

$$(\lambda+\mu)(z) := \sum_{k\in Z} z^{-2k}(\lambda+\mu)(k) \tag{4.3}$$

and let the "odd formal delta function" be

$$\delta^0(z) := \sum_{i\in Z} z^{2i-1} \tag{4.4}$$

and define

$$D^0\delta^0(z) := \sum_{i\in Z} i z^{2i-1}. \tag{4.5}$$

Lemma 4.3:

$$\begin{aligned} &[X_+(z_1),X_-(z_2)] \\ &\quad = (\lambda+\mu)(z_1)\delta^0(z_1/z_2) + D^0\delta^0(z_1/z_2)c. \end{aligned} \tag{4.6}$$

Proof of Lemma 4.3:

$$\begin{aligned} &[X_+(z_1),X_-(z_2)] \\ &\quad = \sum_{k,n\in Z} z_1^{-(2k+1)} z_2^{-(2n-1)} [X_+(k),X_-(n)] \\ &\quad = \sum_{k,n\in Z} z_1^{-(2k+1)} z_2^{-(2n-1)} \\ &\qquad \times\{(\lambda+\mu)(k+n) + n\delta_{k+n,0}c\} \\ &\quad = \sum_{k,n\in Z} z_1^{-2k}(\lambda+\mu)(k)\left(\frac{z_1}{z_2}\right)^{2n-1} + \sum_{n\in Z} n\left(\frac{z_1}{z_2}\right)^{2n-1} c. \end{aligned} \tag{4.7}$$

∎

Let

$$f^e(z_1,z_2) := \sum_{i,j\in Z} f_{ij} z_1^{2i} z_2^{2j}, \quad f_{ij}\in\mathbf{C} \tag{4.8}$$

be a formal power series in z_1,z_2 such that for every $p\in\mathbf{Z}$ the number of f_{ij} with $i+j=p$ is finite. [This means that it makes sense to consider $f^e(z_2,z_2)$.]

Lemma 4.4:

(a) $f^e(z_1,z_2)\delta^0(z_1/z_2) = f^e(z_2,z_2)\delta^0(z_1/z_2),$

(b) $f^e(z_1,z_2)D^0\delta^0(z_1/z_2)$

$$= f^e(z_2,z_2)D^0\delta^0(z_1/z_2) - \frac{1}{2} z_1 \frac{\partial}{\partial z_1} f^e(z_1,z_2)\Big|_{z_1=z_2} \delta^0\left(\frac{z_1}{z_2}\right).$$

Proof of Lemma 4.4: For (a) we have

$$f^e(z_1,z_2)\delta^0\left(\frac{z_1}{z_2}\right)$$

$$= \sum_{i,j,k\in Z} f_{ij} z_1^{2i} z_2^{2j} z_1^{2k-1} z_2^{-2k+1}$$

$$= \sum_{i,j,k\in Z} f_{ij} z_1^{2(i+k)-1} z_1^{2(j-k)+1}$$

$$= \sum_{i,j,l\in Z} f_{ij} z_1^{2l-1} z_2^{2(i+j-l)+1}$$

$$= \sum_{i,j,l\in Z} f_{ij} z_2^{2i} z_2^{2j} \left(\frac{z_1}{z_2}\right)^{2l-1} = f^e(z_2,z_2)\delta^0\left(\frac{z_1}{z_2}\right) \tag{4.9}$$

and for (b):

$$f^e(z_1,z_2)D^0\delta^0\left(\frac{z_1}{z_2}\right)$$

$$= \sum_{i,j,l\in Z} f_{ij} z_1^{2i} z_2^{2j} l \left(\frac{z_1}{z_2}\right)^{2l-1}$$

$$= \sum_{i,j,l\in Z} f_{ij} z_1^{2(i+l)-1} z_2^{2(j-l)+1} l$$

$$= \sum_{i,j,k\in Z} f_{ij} z_1^{2k-1} z_2^{2(j+i-k)+1}(k-i)$$

$$= \sum_{i,j,k\in Z} f_{ij} z_2^{2i} z_2^{2j} k \left(\frac{z_1}{z_2}\right)^{2k-1} - \sum_{i,j\in Z} (f_{ij} z_1^{2i} z_2^{2j} i) \sum_{k\in Z} \left(\frac{z_1}{z_2}\right)^{2k-1}$$

$$= f^e(z_2,z_2)D^0\delta^0\left(\frac{z_1}{z_2}\right) - \frac{1}{2} z_1 \frac{\partial}{\partial z_1} f^e(z_1,z_2)\big|_{z_1=z_2} \delta^0\left(\frac{z_1}{z_2}\right). \tag{4.10}$$

■

Lemma 4.6:

$[Z_+(z_1),Z_-(z_2)] = D^0\delta^0(z_1/z_2) + (\lambda+\mu)(0)\delta^0(z_1/z_2)$.

Proof of Lemma 4.6: We have, using definition (3.7) and Lemmas 4.2 and 4.3,

$$[Z_+(z_1),Z_-(z_2)]$$
$$= E_-(z_1)E_-(z_2)^{-1}[X_+(z_1),X_-(z_2)]$$
$$\times E_+(z_1)E_+(z_2)^{-1}$$
$$= E_-(z_1)E_-(z_2)^{-1}[(\lambda+\mu)(z_1)\delta^0(z_1/z_2)$$
$$+ D^0\delta^0(z_1/z_2)c]E_+(z_1)E_+(z_2)^{-1}. \tag{4.11}$$

Next, using Lemmas 4.4 and 4.5 and formula (4.6) we find

$$[Z_+(z_1),Z_-(z_2)]$$
$$= (\lambda+\mu)(z_2)\delta^0\left(\frac{z_1}{z_2}\right) + D^0\delta^0\left(\frac{z_1}{z_2}\right)c - \frac{1}{2} z_1 \frac{\partial}{\partial z_1}$$
$$\times [E_-(z_1)E_+(z_1)]_{z_1=z_2}$$
$$\times E_-(z_2)^{-1}E_+(z_2)^{-1}\delta^0\left(\frac{z_1}{z_2}\right). \tag{4.12}$$

Now

$$\frac{1}{2} z_1 \frac{\partial}{\partial z_1}[E_-(z_1)E_+(z_1)]$$
$$= \left\{\sum_{k>0} (\lambda+\mu)(-k) z_1^{2k} + (\lambda+\mu)(k) z_1^{-2k}\right\}$$
$$\times [E_-(z_1)E_+(z_1)] \tag{4.13}$$

using definition (3.5) and Lemma 4.2. Hence

$$[Z_+(z_1),Z_-(z_2)] = D^0\delta^0(z_1/z_2) + (\lambda+\mu)(0)\delta^0(z_1/z_2). \tag{4.14}$$

■

We continue the proof of Theorem 4.1. Using the expansion (3.11) we find from lemma 4.6:

$$\sum_{k,l\in Z} z_1^{-(2k+1)} z_2^{-(2l-1)}[Z_+(k),Z_-(l)]$$
$$= \sum_{n\in Z} z_1^{2n-1} z_2^{-(2n-1)}(n + (\lambda+\mu)(0)). \tag{4.15}$$

Hence

$$[Z_+(k),Z_-(l)] = (l + (\lambda+\mu)(0))\delta_{k+l,0}, \tag{4.16}$$

Now one easily checks that $(\lambda+\mu)(0)$ is a central element and since V_m is irreducible $(\lambda+\mu)(0)$ acts as the constant m [see Eqs. (2.4) and (2.5)]. This proves the theorem. ■

V. REALIZATION OF V_m

On an exterior algebra $\Lambda(v_i)$ with generators v_i we define operators of interior and exterior multiplication by

$$i(v_i)v_{i_0}\wedge v_{i_1}\wedge\cdots := \sum_{l=0} (-l)^l \delta_{ii_l} v_{i_0}\wedge v_{i_1}\wedge\cdots\wedge\hat{v}_{i_l}\wedge\cdots, \tag{5.1}$$

$$\epsilon(v_i)v_{i_0}\wedge v_{i_1}\wedge\cdots := v_i\wedge v_{i_0}\wedge v_{i_1}\wedge\cdots.$$

These operators satisfy

$$[i(v_i),\epsilon(v_j)]_+ = \delta_{ij},$$
$$[i(v_i),i(v_j)]_+ = [\epsilon(v_i),\epsilon(v_j)]_+ = 0. \tag{5.2}$$

The operators $Z_\pm(k)$ with commutators described by Theorem 4.1 act on the exterior algebras

$$\Lambda_m = \Lambda(v_1,w_1,v_3,w_3,\ldots,\hat{v}_{2m+1},w_{2m+1},\ldots), \quad m\geqslant 0,$$
$$= \Lambda(v_1,w_1,v_3,w_3,\ldots,v_{-2m-1},\hat{w}_{-2m-1},\ldots), \quad m<0, \tag{5.3}$$

by operators

$$Z_+(k) = i(v_{2k+1}),$$
$$Z_-(-k) = (m-k)\epsilon(v_{2k+1}), \quad k\geqslant 0,$$
$$Z_-(k) = i(w_{2k-1}),$$
$$Z_+(-k) = (m+k)\epsilon(\omega_{2k-1}), \quad k\geqslant 1. \tag{5.4}$$

Noting that the q dimension of Λ_m is precisely the q dimension (3.3) of the vacuum space Ω_{V_m} we find that we can realize V_m as

$$V_m = \mathbb{C}[x_{2i},y_{2j},i,j>0]\otimes\Lambda_m. \tag{5.5}$$

We will refer to the right-hand side of Eq. (5.5) as super Fock space. The action of $\widetilde{gl}_{1/1}$ on super Fock space can now explicitly be described:

$$\lambda(k)=\frac{\partial}{\partial x_{2k}},\quad \lambda(-k)=kx_{2k},$$
$$k>0. \qquad (5.6)$$
$$\mu(k)=\frac{\partial}{\partial y_{2k}},\quad \mu(-k)=-ky_{2k}.$$

Introduce Schur functions $p_k(x)$ by the generating function:

$$\exp\sum_{i>0}x_iz^{2i}=\sum_{k>0}p_k(x)z^{2k}. \qquad (5.7)$$

Then we have

$$X_\pm(z)=E_-(z)^{\pm1}E_+(z)^{\pm1}Z_\pm(z)$$
$$=\sum_{k>0}\hat{p}_k(-t)z^{2k}\sum_{l>0}\hat{p}_l(r)z^{-2l}\sum_{n\in Z}Z_\pm(n)z^{-(2n\pm1)}=\sum_{s\in Z}\left(\sum_{\substack{n\in Z,l,k>0,\\ l+n-k=s}}\hat{p}_{k+l+n}(-t)\hat{p}_l(r)Z_\pm(n)z^{-(2s\pm1)}\right), \qquad (5.8)$$

where

$$\hat{p}_k(t)=p_k(\mp(\lambda+\mu)(t)/t).$$

Hence

$$X_\pm(s)=\sum_{\left(\substack{n\in Z,l,k>0,\\ l+n-k=s}\right)}\hat{p}_{k+l+n}(-t)\hat{p}_l(r)Z_\pm(n). \qquad (5.9)$$

Using the formulas of Ref. 1 one can also explicitly describe the action of the complete superalgebra $a_{\infty|\infty}$ on super Fock space in terms of multiplication and differentiation operations with respect to the even and odd variables x_i,y_i,v_i,w_i. We leave this to the reader.

[1] V. G. Kac and J. W. van de Leur, Ann. Inst. Fourier 37, 99 (1987).
[2] G. Golitzin, "Representations of affine superalgebras," Ph.D. thesis, Yale University, 1985.
[3] J. Lepowsky and R. L. Wilson, Invent. Math. 15, 199 (1984).

On regularization ambiguities in fermionic models

D. Cabra and F. A. Schaposnik
Departamento de Fisica, Universidad Nacional de La Plata, 1900 La Plata, Argentina

(Received 30 June 1988; accepted for publication 30 November 1988)

Different regularization schemes based on the heat-kernel method using as samples the chiral Schwinger model and the Thirring model are discussed. As it happens with non-Abelian anomalies, it is shown that different physical contexts require different specifications of the fermionic generating functional. Also discussed is the introduction of an arbitrary parameter through the regularization and how this affects the resulting quantum theory.

I. INTRODUCTION

Quantization of gauge theories with Weyl fermions has recently received a lot of attention, particularly after the results of Jackiw and Rajaraman[1] on the consistency of the chiral Schwinger model and those of Faddeev and Shatashvili[2] on the modification of the canonical quantization by addition of new degrees of freedom through a Wess–Zumino action.

Prompted by these important developments, many investigations[3–11] analyzed the possibility of quantization of anomalous theories, usually considered as inconsistent—nonunitary and nonrenormalizable—if cancellation of anomalies was not carefully contrived.

Analysis of potentially anomalous theories requires the development of regularization procedures that take into account the peculiarities of chiral fermions. In the path-integral approach, where most of the above mentioned advances were made, calculation of anomalies depends on a precise understanding of the functional integration measure and on how it changes under the original classical symmetry of the theory.[12,13] In this approach one defines a Jacobian J associated to this symmetry transformation (known as the Fujikawa Jacobian[12]) as a ratio of fermion determinants. If this Jacobian is not trivial, the associated Noether current is anomalous.

The crucial point is that each determinant appearing in J needs a regularization (for example, for Dirac or Weyl fermions the Dirac operators appearing in J are unbounded and hence determinants are ill-defined quantities).

Typically, one regulates J by inserting $\exp(-R/M^2)$ (Refs. 12 and 13) with R some positive definite operator called the regularization operator (RO); then one performs all computations keeping M^2 fixed and finally one takes the limit $M^2 \to \infty$ at the end of the calculation. This method is usually known as the heat-kernel method. It is important to note that the choice of RO determines whether the Jacobian is trivial or not.

If one has more than one classical symmetry it may happen that the measure and the RO cannot be made invariant under all of them. It is then a matter of physical prejudice as to which symmetries one maintains at the quantum level and which symmetry becomes anomalous. Alternatively, it may happen that there is no choice of RO respecting the original symmetry of the theory (this last case corresponds to the case of chiral models). In any case, anomalies arise.

For gauge theories with Dirac fermions, the fermionic Lagrangian dictates the choice of RO. Indeed, taking $R = D^2$ (Refs. 12 and 13), with $\not{D} = i\not{\partial} + g\not{A}$, the (Hermitian) Dirac operator appearing in the Lagrangian, gauge invariance, the fundamental symmetry to be preserved, is ensured. With this choice classical chiral symmetry is sacrificed at the quantum level. The resulting anomaly in the axial current obtained using the heat-kernel approach, coincides with the perturbative results[14] as well as with the ones obtained using the ζ-function method. (In fact, for Hermitian $\not{D}$ one can show the equivalence between the heat-kernel method and ζ-function method.[13])

If A_μ does not represent a gauge field (for example, when it is introduced as an auxiliary field in purely fermionic models like the Thirring model) then it is not compulsory to make the choice $R = \not{D}^2$. More general RO can in fact be introduced[15] and in this form more general results obtained (as it happens in the Thirring model case, when a one-parameter family of solutions is known to exist[16]).

Gauge theories with Weyl fermions are another example in which regularization ambiguities arise. This is due to the following peculiarities of chiral models.[17,18]

Primum: The Dirac operator appearing in the Lagrangian, $D = \not{D}(1-\gamma_5)/2$ (we take left-handed fermions for definiteness), does not define an eigenvalue problem (it maps negative chirality spinors into positive chirality ones).

Secundum: A modified Dirac operator $\widetilde{D}$, constructed by addition of a free right-handed sector so as to define an eigenvalue problem, is neither gauge-covariant nor Hermitian. Either analytic continuation of $\widetilde{D}$ (Ref. 18) or use of $R = (\widetilde{D}\widetilde{D}^+ + \widetilde{D}^+\widetilde{D})/2$ (Ref. 10) breaks gauge invariance at the quantum level.

Tertium: A family of RO depending on certain parameters are in fact admissible.[19–26] This leads to regularization ambiguities that can be exploited to render the quantum theory consistent (at least in two space-time dimensions, as first shown in Ref. 1).

We address these regularization problems in the present work using the chiral-Schwinger model (CSM) and the Thirring model (TM) as examples. Solutions of these two-dimensional models are known and hence their path-integral treatment can be used to get insight into the more general problem of regulating Jacobians and fermion determinants.

The paper is organized as follows. In Sec. II, after briefly

 0022-2488/89/040816-08$02.50

defining the models (II A) we discuss the change in the fermonic measure under chiral and gauge transformations (Sec. II B). We then discuss the regularization dependence of Jacobians and fermion determinants for the CSM and TM in Sec. II C.

From these results we evaluate in Sec. III fermion currents, anomalies, and current–current commutators showing how different regularization schemes affect the results. We give in Sec. IV a summary of our results and conclusions leaving for an Appendix details of the calculations.

II. FERMION DETERMINANTS USING DIFFERENT REGULARIZATIONS

A. The models

We shall consider two two-dimensional fermionic models where regularization ambiguities may arise: the Thirring model and the chiral Schwinger model. The first one is defined by the (Euclidean) Lagrangian

$$\mathscr{L} = \bar{\psi} i \partial\!\!\!/ \psi - (g^2/2)(\bar{\psi}\gamma_\mu \psi)^2. \tag{2.1}$$

Our conventions for matrices are

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2\delta_{\mu\nu},$$
$$\gamma_5 = i\gamma_0\gamma_1,$$
$$\varepsilon_{01} = -\varepsilon_{10} = 1.$$

The generating functional for this model is

$$\mathscr{Z}_{\rm TM}[S_\mu] = \int \mathscr{D}\bar{\psi}\,\mathscr{D}\psi \exp\Big(-\int d^2x(\mathscr{L} + \bar{\psi} S\!\!\!/ \psi)\Big), \tag{2.2}$$

where $S_\mu(x)$ is an external source.

Using the identity[27]

$$\exp\Big(\frac{g^2}{2}\int(\bar{\psi}\gamma_\mu\psi)^2\, d^2x\Big) = \int \mathscr{D}A_\mu \exp\Big(-\int\Big[\frac{1}{2}A_\mu^2 + g\bar{\psi} A\!\!\!/ \psi\Big] d^2x\Big), \tag{2.3}$$

the generating functional can be rewritten in the form

$$\mathscr{Z}_{\rm TM}[S_\mu] = \int \mathscr{D}\bar{\psi}\,\mathscr{D}\psi\,\mathscr{D}A_\mu \exp\Big(-\int[\tfrac{1}{2}A_\mu^2 + \bar{\psi} S\!\!\!/ \psi + \bar{\psi} D\!\!\!\!/\,[A]\psi]d^2x\Big), \tag{2.4}$$

with

$$D\!\!\!\!/\,[A] = i\partial\!\!\!/ + gA\!\!\!/. \tag{2.5}$$

Note that the auxiliary field A_μ represents two degrees of freedom: due to the presence of the A_μ^2 term in (2.4), no gauge fixing has to be implemented. Then, when regularizing the fermionic path integral (i.e., the fermionic determinant), gauge invariance cannot be invoked to select the regulator.

A similar situation arises in the quantization of the chiral Schwinger model.[1] Indeed, the fermionic Lagrangian for this gauge theory coupled to left-handed fermions is

$$\mathscr{L}_{\rm CSM} = \bar{\psi}(i\partial\!\!\!/ + gA\!\!\!/)((1-\gamma_5)/2)\psi = \bar{\psi} D[A]\psi. \tag{2.6}$$

The fermionic part of the generating functional reads

$$\mathscr{Z}_{\rm CSM} = \int \mathscr{D}\bar{\psi}\,\mathscr{D}\psi \exp\Big(-\int \mathscr{L}_{\rm CSM}\, d^2x\Big) \doteq \det D_{\rm CSM}[A]. \tag{2.7}$$

Now, since the Dirac operator for chiral fermions ($D_{\rm CSM}[A]$) maps negative chirality spinors into positive chirality ones, it does not define a correct eigenvalue problem. Usually one overcomes this difficulty by introducing right-handed free fermions,[17,18]

$$\widetilde{D}[A] = D_{\rm CSM}[A] + i\partial\!\!\!/((1+\gamma_5)/2) = i\partial\!\!\!/ + gA\!\!\!/((1-\gamma_5)/2) \tag{2.8}$$

and defining

$$\det D_{\rm CSM}[A] = \det \widetilde{D}[A]\,|_{\rm Reg}. \tag{2.9}$$

We have indicated with the subscript "Reg" that some regularization has to be adopted since $\widetilde{D}[A]$ is an unbounded operator. Now, the addition of right-handed fermions (which can be justified by several arguments[17]) solves the chirality-flip problem but introduces an ambiguity. Indeed, $\widetilde{D}[A]$ is not a covariant derivative and hence there is no reason to expect

$$\det \widetilde{D}[A]\,|_{\rm Reg} = \det \widetilde{D}[A^g]\,|_{\rm Reg}$$

with A_μ^g the gauge transformed of A_μ. As in the Thirring model case, there is then no reason to use a gauge-invariant regularization prescription.

B. Jacobians

For notation brevity we shall call $D[A]$ either $D\!\!\!\!/\,[A]$ defined in (2.5) for the Thirring model or $\widetilde{D}[A]$, defined in (2.8) for the chiral Schwinger model.

We shall evaluate the change in the fermionic determinant (i.e., the Fujikawa Jacobian[12]) associated with the following change of variables:

$$\psi = u_t\psi', \quad \bar{\psi} = \bar{\psi}'\tilde{u}_t. \tag{2.10}$$

For the Thirring model u_t and $\tilde{u}_t$ are given by

$$u_t = e^{[\gamma_5\phi(x) + i\eta(x)]t}, \qquad \tilde{u}_t = e^{[\gamma_5\phi(x) - i\eta(x)]t}, \tag{2.11}$$

with η (ϕ) a scalar (pseudoscalar) field and t a real parameter.

For the CSM, since the fermions are left-handed, transformation[2–11] can be simplified to[10]

$$u_t = \exp[+((1-\gamma_5)/2)(\phi - i\eta)t], \qquad \tilde{u}_t = \exp[-((1+\gamma_5)/2)(\phi - i\eta)t]. \tag{2.12}$$

Transformations (2.10) and (2.12) are chosen so that for $t = 1$ the corresponding Lagrangians become free. Indeed, if we relate ϕ and η to the vector field A_μ through the identity

$$A_\mu = -(1/g)(\epsilon_{\mu\nu}\partial_\nu\phi - \partial_\mu\eta), \tag{2.13}$$

then, under transformations (2.10)–(2.12) the fermion Lagrangians become

$$\mathscr{L} \to \mathscr{L} = \bar{\psi}' D_t \psi', \tag{2.14}$$

with

$$D_t = D[(1-t)A] \tag{2.15}$$

and hence for $t=1$ one has $D_1 = i\partial\!\!\!/$.

The Fujikawa Jacobian $J(t)$ associated with transformations (2.10)–(2.12) relates the determinants of operators $D_t[A]$ and $D[A]$

$$\det D[A] = J(t) \det D_t[A]\,. \tag{2.16}$$

Since both determinants require a regularization, the result for $J(t)$ necessarily depends on this regularization. Here $J(t)$ can be evaluated from the following differential equation[13] [derived from (2.16)]:

$$-\frac{d}{dt}\log J(t) = \frac{d}{dt}\log\det D_t\Big|_{\text{Reg}} \equiv \omega'(t) \tag{2.17}$$

[a regularization is understood in (2.17)].

In particular, the fermion determinant in the lhs of (2.16) can be evaluated from the knowledge of $J(1)$,

$$\det D[A] = J(1)\det i\partial\!\!\!/\,, \tag{2.18}$$

with

$$J(1) = \exp\left(-\int_0^1 \omega'(t)dt\right). \tag{2.19}$$

We evaluate $\omega'(t)$ using

$$\omega'(t) = \lim_{\Delta t\to 0}\frac{\operatorname{tr}\log D_t^{-1}D_{t+\Delta t}}{\Delta t}. \tag{2.20}$$

For the TM ($D_t = D\!\!\!\!/_t$) one obtains

$$\omega'_{\text{TM}}(t) = \operatorname{tr} D\!\!\!\!/_t^{-1}[(\gamma_5\phi - i\eta)D\!\!\!\!/_t + D\!\!\!\!/_t(\gamma_5\phi + i\eta)]\,. \tag{2.21}$$

Analogously, for the CSM ($D_t = \widetilde{D}_t$) one has

$$\omega'_{\text{CSM}}(t) = \operatorname{tr}[D_t^{-1}(\widetilde{D}_t((1-\gamma_5)/2)(\phi - i\eta) - ((1+\gamma_5)/2)(\phi - i\eta)\widetilde{D}_t)]\,. \tag{2.22}$$

As we stated above, $w'(t)$ needs a regularization that can in principle destroy the cyclic property of the trace as first observed in Ref. 25. This is a very important point: if the cyclic property is assumed, (2.21) and (2.22) simplify to

$$\omega'_{\text{TM}} = 2\operatorname{tr}\gamma_5\phi\,, \tag{2.23}$$

$$\omega'_{\text{CSM}} = \operatorname{tr}\gamma_5(\phi - i\eta)\,. \tag{2.24}$$

We shall discuss this point in more details below.

C. Regularization dependence of Jacobians and determinants

For gauge invariant theories, the regularization prescription is chosen so as to respect gauge invariance. One can prove that the ζ-function method, the heat-kernel one [with $\exp(-D\!\!\!\!/_t^2/M^2)$ as regulator], etc., lead to the same gauge-invariant answer for the fermionic determinant. Moreover, the cyclic property of the trace holds. (This can be seen very simply in the heat-kernel approach, since $[\exp(-D_t^2/M^2), D_t] = 0$.)

We then have, for gauge-invariant models, for a transformation like (2.10),[13]

$$\omega'(t) = \lim_{M^2\to\infty} 2\operatorname{tr}\gamma_5\phi e^{-R/M^2} = -\frac{d}{dt}\frac{d\zeta}{ds}(s,D_t)\Big|_{s=0}, \tag{2.25}$$

provided the regulating operator (RO) R in the heat-kernel approach coincides with D_t. As we stated above gauge invariance cannot be invoked in regularizing the Thirring or the CS models. Then, there is no particular reason to use the ζ-function method or the heat-kernel one with $R = D_t$. In fact, more general prescriptions (compatible with Lorentz invariance) can be implemented and we address this point now.

We consider here the heat-kernel approach (for the ζ-function method, see Ref. 24). We shall use as regulating operator the following one[19–23]:

$$R = [D_t^a]^2 = [i\partial\!\!\!/ + g(1-t)A_+ + ag(1-t)A_-]^2, \tag{2.26}$$

with

$$A_\pm = A\!\!\!/((1\mp\gamma_5)/2)\,. \tag{2.27}$$

In (2.25) a is a parameter related to the one introduced by Jackiw and Rajaraman[1] to take care of regularization ambiguities in the CSM. In the present analysis, its presence gives different weights to the A_+ and A_- components which, for the TM and the CSM correspond to independent degrees of freedom. Another way of understanding the presence of this undetermined parameter is the following. In computing $\log\det(i\partial\!\!\!/_+ + gA_+)$ $[\log\det(i\partial\!\!\!/_- + gA_-)]$ one gets in principle a result $F[A_+]$ ($F[A_-]$) depending only on A_+ (A_-). The complete determinant $F[A_\mu] = \det(i\partial\!\!\!/ + gA\!\!\!/)$ including both right-handed and left-handed parts has necessarily an arbitrariness

$$F[A_\mu] = F[A_+] + F[A_-] + C\int A_\mu^2\, d^2x\,. \tag{2.28}$$

The arbitrary constant can be fixed only on gauge-invariance grounds. One can easily obtain a relation like (2.28) using D_t^a as RO.[19,23]

Now, since D_t^a is not Hermitian, two possibilities arise when using the heat-kernel approach in order to have a positive definite R (Refs. 13, 28, and 29).

$$\text{Analytic continuation } \gamma_5 \to i\gamma_5 \text{ in } R_A = D_t^{a^2} \tag{2.29a}$$

or

$$R_B = (D_t^a D_t^{a^+} + D_t^{a^+}D_t^a)/2\,. \tag{2.29b}$$

The alternative (2.29a) has been shown[13] to lead to the same result as the ζ function (which is defined even if D is not Hermitian). The alternative (2.29b) was proposed by Fujikawa[29] in his analysis of covariant and consistent anomalies.

Now, in both cases,

$$[e^{-R/M^2}; D_t] \neq 0\,,$$

and then, the regulated forms of (2.21) and (2.22) and (2.23) and (2.24) are not the same. This can be interpreted as if the cyclic property of traces in (2.21) and (2.22) does not hold.[25] We shall examine each possibility separately.

(I) One starts from (2.23) and (2.24) (RegI) obtained using cyclic property of the trace. Inserting the RO one gets

$$\omega'_{\text{TM}}(t)\big|_{\text{RegI}} = \lim_{M^2\to\infty} 2\operatorname{tr}\gamma_5\phi e^{-R/M^2}, \tag{2.30}$$

$$\omega'_{\text{CSM}}(t)\big|_{\text{RegI}} = \lim_{M^2\to\infty}\operatorname{tr}\gamma_5(\phi - i\eta)e^{-R/M^2}. \tag{2.31}$$

(II) One starts from (2.21) and (2.22) (RegII), getting

$$\omega'_{\rm TM}(t)\,|_{\rm RegII} = \lim_{M^2\to\infty} \operatorname{tr} \not{D}_t^{-1}[(\gamma_5\phi - i\eta)\not{D}_t + \not{D}_t(\gamma_5\phi + i\eta)]e^{-R/M^2}, \tag{2.32}$$

$$\omega'_{\rm CSM}(t)\,|_{\rm RegII} = \lim_{M^2\to\infty} \operatorname{tr} D_t^{-1}\Big[\widetilde{D}_t\Big(\frac{1-\gamma_5}{2}\Big)(\phi - i\eta) - \Big(\frac{1+\gamma_5}{2}\Big)(\phi - i\eta)\widetilde{D}_t\Big]e^{-R/M^2}. \tag{2.33}$$

We shall consider the two above-mentioned choices for R, namely,

$$R_{\rm A} = D_t^2, \qquad R_{\rm B} = (D_t D_t^+ + D_t^+ D_t)/2. \tag{2.34}$$

We have evaluated $\omega'_{\rm TM}(t)$ and $\omega'_{\rm CSM}(t)$ using regularization RegI and RegII, and regulating operator $R_{\rm A}$ and $R_{\rm B}$. We denote these choices in the form RegIA, RegIB, etc. Details of the calculations are given in the Appendix. The answer is

$$\omega'_{\rm TM}\,|_{\rm RegIA} = (1-t)\Big[\frac{g}{2\pi}(1+a)\int \phi\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x + \frac{i}{2\pi}g(a-1)\int \phi\partial_\mu A_\mu\, d^2x\Big],$$

$$\omega'_{\rm TM}\,|_{\rm RegIB} = (1-t)\frac{g}{2\pi}(1+a)\int \phi\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x,$$

$$\omega'_{\rm TM}\,|_{\rm RegIIA} = (1-t)\Big[\frac{g}{\pi}\int \phi\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x + \frac{g^2}{2\pi}(a-1)\times\int \phi\,\partial_\mu A_\nu(i\delta_{\mu\nu} + \epsilon_{\mu\nu})d^2x\Big],$$

$$\omega'_{\rm TM}\,|_{\rm RegIIB} = (1-t)\frac{g}{2\pi}(1+a)\int \phi\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x,$$

$$\omega'_{\rm CSM}\,|_{\rm RegIA} = (1-t)\Big[\frac{g}{2\pi}\int (\phi - i\eta)\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x + \frac{ig}{4\pi}(a-1)\int (\phi - i\eta)\partial_\mu A_\nu(\delta_{\mu\nu} - i\epsilon_{\mu\nu})d^2x\Big],$$

$$\omega'_{\rm CSM}\,|_{\rm RegIB} = (1-t)\frac{g}{2\pi}(1+a)\times\int (\phi - i\eta)\epsilon_{\mu\nu}\,\partial_\mu A_\nu\, d^2x,$$

$$\omega'_{\rm CSM}\,|_{\rm RegIIA} = (1-t)\frac{g}{4\pi}(1+a)\times\int (\phi - i\eta)\partial_\mu A_\nu(\epsilon_{\mu\nu} - i\delta_{\mu\nu})d^2x,$$

$$\omega'_{\rm CSM}\,|_{\rm RegIIB} = (1-t)\frac{g}{4\pi}(1+a)\times\int (\phi - i\eta)\partial_\mu A_\nu(\epsilon_{\mu\nu} - i\delta_{\mu\nu})d^2x. \tag{2.35}$$

We shall see in the next section how all these regularization-dependent results for ω' affect the anomaly in fermion currents, current–current commutators, etc. Note that for $a=1$ $R_{\rm A}$ and $R_{\rm B}$ coincide and $\omega'|_{\rm RegA} = \omega'|_{\rm RegB}$ as can be seen from (2.35).

Fermion determinants can be very easily obtained from (2.18)–(2.19) and (2.35). We list the results: For the TM we have

$$\log\det\Big[\frac{i\not\partial + g\not{A}}{i\not\partial}\Big]_{\rm RegIA} = \frac{1}{4\pi}(1+a)\int d^2x\,\phi\Box\phi + \frac{i}{4\pi}(a-1)\int d^2x\,\phi\Box\eta, \tag{2.36a}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}}{i\not\partial}\Big]_{\rm RegIB} = \frac{1}{4\pi}(1+a)\int d^2x\,\phi\Box\phi, \tag{2.36b}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}}{i\not\partial}\Big]_{\rm RegIIA} = \frac{1}{4\pi}(1+a)\int d^2x\,\phi\Box\phi + \frac{i}{4\pi}(a-1)\int d^2x\,\phi\Box\eta, \tag{2.36c}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}}{i\not\partial}\Big]_{\rm RegIIB} = \frac{1}{4\pi}(1+a)\int d^2x\,\phi\Box\phi. \tag{2.36d}$$

For the CSM

$$\log\det\Big[\frac{i\not\partial + g\not{A}((1-\gamma_5)/2)}{i\not\partial}\Big]_{\rm RegIA} = \frac{1}{8\pi}\int d^2x[(1+a)\phi\Box\phi + (a-1)\eta\Box\eta - 2i\eta\Box\phi], \tag{2.37a}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}((1-\gamma_5)/2)}{i\not\partial}\Big]_{\rm RegIB} = \frac{1}{8\pi}\int d^2x[(1+a)\phi\Box\phi - i(1+a)\eta\Box\phi], \tag{2.37b}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}((1-\gamma_5)/2)}{i\not\partial}\Big]_{\rm RegIIA} = \frac{(1+a)}{8\pi}\int d^2x[\phi\Box\phi - \eta\Box\eta - 2i\eta\Box\phi], \tag{2.37c}$$

$$\log\det\Big[\frac{i\not\partial + g\not{A}((1-\gamma_5)/2)}{i\not\partial}\Big]_{\rm RegIIB} = \frac{(1+a)}{8\pi}\int d^2x[\phi\Box\phi - \eta\Box\eta - 2i\eta\Box\phi]. \tag{2.37d}$$

Some comments are in order at this point. For the CSM, (2.36a) reproduces the usual result obtained following different approaches.[1,6,7,10,19–24] Indeed, using the identities

$$\partial_\mu A_\mu = (1/g)\Box\eta, \qquad \epsilon_{\mu\nu}\partial_\mu A_\nu = (1/g)\Box\phi, \tag{2.38}$$

Eq. (2.37a) can be rewritten as

$$\log\det\left[\frac{i\partial\!\!\!/ + gA\!\!\!/((1-\gamma_5)/2)}{i\partial\!\!\!/}\right]_{\mathrm{RegIA}}$$

$$= \frac{g^2}{8\pi}\int d^2x\, A_\mu(\delta_{\mu\alpha} - i\epsilon_{\mu\alpha})\partial_\alpha\Box^{-1}\partial_\beta(\delta_{\beta\nu} + i\epsilon_{\beta\nu})A_\nu$$

$$-\frac{ag^2}{8\pi}\int A_\mu^2\, d^2x\,. \tag{2.39}$$

This is, use of the heat-kernel method with $\exp(-D\!\!\!/_t^{a^2}/M^2)$ as regulator and use of the cyclic property of the trace before regularization leads to the result first discussed by Jackiw and Rajaraman.[1] The heat-kernel method with $(DD^+ + D^+D)/2$ as RO does not lead to (2.39). This difference in the results according to whether one uses R_A or R_B is not surprising. As it has been discussed in detail by Fujikawa[29,30] for non-Abelian anomalies, the choice of R_A or R_B leads to two different forms of the anomaly, namely, the consistent and covariant forms. The same difference occurs using the stochastic quantization method that is analogous to the heat-kernel method either with R_A or R_B (Ref. 36).

It is interesting to note that the extended version of the ζ-function approach presented in Ref. 24 so as to include an a dependence also leads to (2.39).

Introduction of the RO before employing the cyclic property of the trace yields a different result for the determinant

$$\log\det\left[\frac{i\partial\!\!\!/ + gA\!\!\!/((1-\gamma_5)/2)}{i\partial\!\!\!/}\right]_{\mathrm{RegII}}$$

$$\propto\int d^2x\, A_\mu(\delta_{\mu\alpha} - i\epsilon_{\mu\alpha})\partial_\alpha\Box^{-1}\partial_\beta(\delta_{\beta\nu} + i\epsilon_{\beta\nu})A_\nu\,. \tag{2.40}$$

Note that the addition of $\log\det(i\partial\!\!\!/ + gA_-)$ given by (2.39) and the corresponding result for $\log\det(i\partial\!\!\!/ + gA_+)$ leads to the Dirac fermion determinant (2.36a) which for $a=1$ coincides with the well-known gauge invariant Schwinger model determinant.

Concerning the Dirac fermion determinant evaluated for the Thirring model, note that for $a=1$ (i.e., $R = R_A = R_B = i\partial\!\!\!/ + gA\!\!\!/$), $[R,D]=0$, hence the four results coincide giving the Schwinger model determinant.

As we mentioned above, it is well known that different RO's lead to different forms of the non-Abelian anomaly. However, physical consequences of these different forms may be identical. In particular, the anomaly cancellation conditions are the same for the covariant and consistent anomalies.[31] We shall discuss this aspect of our results in the next section.

III. ANOMALIES AND CURRENT–CURRENT COMMUTATORS

From the generating functionals (2.4) and (2.6) we can easily find the vacuum expectation value of fermionic currents:

$$J_\mu^{\mathrm{TM}} = \langle j_\mu(x)\rangle_{\mathrm{TM}} = \frac{1}{\mathscr{Z}_{\mathrm{TM}}}\frac{\delta\mathscr{Z}_{\mathrm{TM}}}{\delta S^\mu(x)}\bigg|_{S^\mu=0} = \langle\bar\psi\gamma_\mu\psi\rangle\,, \tag{3.1}$$

$$J_\mu^{\mathrm{CSM}} = \langle j_\mu(x)\rangle_{\mathrm{CSM}} = \frac{1}{\mathscr{Z}_{\mathrm{CSM}}}\frac{\delta\mathscr{Z}_{\mathrm{CSM}}}{\delta A^\mu(x)}$$

$$= g\left\langle\bar\psi\gamma_\mu\left(\frac{1-\gamma_5}{2}\right)\psi\right\rangle. \tag{3.2}$$

[Note that in (3.2) the gauge field A_μ is considered as a background and then it can be used to find J_μ; for the TM being A_μ an auxiliary field, an external source S_μ has to be introduced in order to compute currents.]

Current–current correlation functions are given by

$$G_{\mu\nu}^{\mathrm{TM}}(x,y) = \langle T(j_\mu(x)j_\nu(y))\rangle_{\mathrm{TM}}$$

$$= \frac{1}{\mathscr{Z}_{\mathrm{TM}}}\cdot\frac{\delta^2\mathscr{Z}_{\mathrm{TM}}}{\delta S^\mu(x)\,\delta S^\nu(y)}\bigg|_{S^\mu=S^\nu=0}, \tag{3.3}$$

$$G_{\mu\nu}^{\mathrm{CSM}}(x,y) = \langle T(j_\mu(x)j_\nu(y))\rangle_{\mathrm{CSM}}$$

$$= \frac{1}{\mathscr{Z}_{\mathrm{CSM}}}\cdot\frac{\delta^2\mathscr{Z}_{\mathrm{CSM}}}{\delta A^\mu(x)\,\delta A^\nu(y)}, \tag{3.4}$$

and then current–current commutators can be evaluated from (3.3) and (3.4) by means of the Bjorken–Johnson–Low (BJL) limit.[32,33]

We start with the CSM currents. Relation (3.2) can be rewritten in the form

$$J_\mu^{\mathrm{CSM}} = \frac{\delta}{\delta A^\mu}\log\det D[A]\,, \tag{3.5}$$

and hence the results given in (2.37) for the CSM determinant, written in terms of A_μ can be used. For the regularization leading to the usual form of the fermion determinant [i.e., the regularization RegIA leading to (2.39)] we get

$$J_\mu^{\mathrm{CSM}} = (g/4\pi)$$

$$\times[(\delta_{\mu\nu} - i\epsilon_{\mu\nu})\partial_\nu(\phi + i\eta)] - (g^2/4\pi)aA_\mu\,, \tag{3.6}$$

and hence

$$\partial_\mu J_\mu^{\mathrm{CSM}} = (g^2/4\pi)[(1-a)\partial_\mu A_\mu + i\epsilon_{\mu\nu}\partial_\mu A_\nu]\,.$$

From this well-known result one infers that the theory is anomalous whatever the value of a is chosen. As proved by Jackiw and Rajaraman,[1] the effective action for A_μ resulting from the addition of the fermion determinant (2.39) and the $\frac{1}{4}F_{\mu\nu}^2$ terms defines a sensible unitary Lorentz invariant theory provided $a>1$ with a massive gauge meson (with mass $g^2[a^2/(a-1)]$ and massless excitations). It is important to stress at this point that any of the three other regularizations lead to this result,[1] confirmed by many other investigations.[19–24]

Concerning the other regularizations for the CSM, use of R_B instead to R_A leads to

$$J_\mu^{\mathrm{CSM}}\big|_{\mathrm{RegIB}} = (g^2/8\pi)(1+a)\epsilon_{\mu\nu}A_\nu \tag{3.7}$$

and

$$\partial_\mu J_\mu^{\mathrm{CSM}}\big|_{\mathrm{RegIB}} = (g^2/8\pi)(1+a)\epsilon_{\mu\nu}\,\partial_\mu A_\nu\,, \tag{3.8}$$

which coincides with (3.6) only for $a=1$. Neither this result nor those arising from the Type II regularization,

$$J_\mu\big|_{\mathrm{RegII}} = f(a)(A_\mu + \epsilon_{\mu\nu}A_\nu)\,, \tag{3.9}$$

[with $f(a) = (g^2/4\pi)(1+a)$ for types A, B of RO] lead to the results obtained by many authors using different approaches and reproduced by type IA regularization.

Current–current commutators for the CSM are, as it is well known,[34] anomalous (in particular $[j_0, j_0] \propto \delta'$). Only if gauge degrees of freedom are incorporated as dynamical ones as proposed in Refs. 2–5 one gets $[j_0, j_0] = 0$ as proved in Ref. 35.

Concerning the TM we shall compute current–current commutators in order to compare the results obtained using different regularization with the well-known Klaiber results.[16]

As we stated above we use the BJL method[32,33] that starts from the identity

$$\langle [j_\mu(x); j_\nu(y)]_{\text{e.t.}} \rangle = \lim_{\epsilon \to 0} [G_{\mu\nu}(x_1, x_0; y_1, x_0 + \epsilon) - G_{\mu\nu}(x_1, x_0; y_1, x_0 - \epsilon)] \quad (3.10)$$

and using Eq. (3.3) for evaluating $G^{\text{TM}}_{\mu\nu}$. If we write the fermion determinants listed in (2.36) in the form:

$$\log \det\left[\frac{i\not\partial + g\not A}{i\not\partial}\right] = -\frac{g^2}{2\pi} \int d^2x\, A_\mu D_{\mu\nu} A_\nu , \quad (3.11)$$

it is easy to see that

$$\langle j_\mu(x) j_\nu(y) \rangle = (1/g^2)[-\delta_{\mu\nu}\delta(x-y) + \mathscr{D}^{-1}_{\mu\nu}(x,y)] , \quad (3.12)$$

where

$$\mathscr{D}_{\mu\nu} = \delta_{\mu\nu} + (g^2/\pi) D_{\mu\nu} . \quad (3.13)$$

Only $\Box^{-1}$ terms in $\mathscr{D}_{\mu\nu}$ contribute to the $\epsilon \to 0$ limit in (3.10). Indeed being

$$\Box^{-1}_{xy} = (1/2\pi)\log \mu|x-y| \quad (3.14)$$

it appears in $G_{\mu\nu}(x_1, x_0; y_1, y_0 + \epsilon)$ through a term $G^{\Box^{-1}}_{\mu\nu}$,

$$G^{\Box^{-1}}_{\mu\nu} \propto \frac{\epsilon(x_1 - y_1)}{\epsilon^2 + (x_1 - y_1)^2} \underset{\epsilon \to 0}{\to} \delta'(x_1 - y_1) \quad (3.15)$$

responsible for a δ' behavior in $[j_0, j_1]$. All other contributions to $G_{\mu\nu}$ vanish when computing the rhs in (3.10).

It is important to stress that if one uses R_{B} (either with RegI or RegII) one then gets the same results obtained using the operator approach (see, for example, Ref. 16). In particular $[j_0, j_0]_{\text{e.t.}} = 0$ and

$$[j_0; j_1]_{\text{e.t.}} = (1/g^2) h(a) \delta'(x) , \quad (3.16)$$

with

$$h(a) = \begin{cases} (1 + \pi/g^2 a)^{-1}, & \text{for RegIB} , \\ (1 + 2\pi/g^2(1+a))^{-1}, & \text{for RegIIB} . \end{cases} \quad (3.17)$$

The a dependence of the coefficient corresponds to the existence of a one-parameter family of solutions in the TM. As stated by Klaiber, this corresponds to an undetermination of the coupling constant in the sense that any value can be produced with an appropriate current definition (or equivalently with an appropriate value of the regularization parameter a).

Concerning regularization A it leads to an anomalous result for $[j_0, j_0]$,

$$[j_0(x); j_0(y)] = k(a)\delta'(x) . \quad (3.18)$$

Only for $a = 1$, this choice leads to a sensible result.

IV. SUMMARY AND CONCLUSIONS

We have studied regularization ambiguities in fermionic theories using the path-integral approach. We have chosen as samples the Thirring model and the chiral Schwinger model, two well-known, two-dimensional theories in which gauge invariance cannot be used as a guiding principle when selecting a regularization prescription.

We have employed a heat-kernel regularization, which consists of inserting $\exp(-R/M^2)$ in ill-defined quantities (traces, fermionic Jacobians, etc.) with R some positive definite operator.

While in gauge-invariant theories identification of R with a square of the covariant derivative leads to gauge invariant and unique results (for fermion determinants, chiral anomalies, etc.) in the present models there is a wider choice of regulators, which yield a family of solutions depending (for the two-dimensional case) on an *a priori* arbitrary parameter.

A similar situation is encountered in the study of non-Abelian anomalies[30,31,36]: depending on the specification of the fermionic generating functional (i.e., the fermionic determinant) two different forms of Ward–Takahashi identities can be obtained. The corresponding anomalies are known as covariant and consistent anomalies and both have important applications in different physical contexts (see, for example, the discussion in Ref. 30). It has been shown by Bardeen and Zumino[31] that one can always pass from one form of the anomaly to the other one by changing the definition of fermion currents and hence the anomaly cancellation conditions are the same for either form.

Inspired in Fujikawa's discussion[30] of these non-Abelian anomalies using the path integral framework, we have tested two forms for RO [Eqs. (2.19a) and (2.19b)] (for non-Abelian anomalies R_{A} leads to the covariant result and R_{B} to the consistent one). Moreover, since neither R_{A} nor R_{B} commute with the Dirac operator appearing in the Lagrangian, regularization is not fully specified until one decides the step at which the RO is inserted in ill-defined functional traces. Different choices lead to the nonequivalent expressions (2.30) and (2.31) and (2.32) and (2.33).

We can conclude that, as it happens for non-Abelian anomalies, different physical contexts require different choices of regularization when studying chiral models or purely fermionic models (like the Thirring model). For the chiral Schwinger model, analytic continuation of the (originally non-Hermitian) Dirac operator and insertion of the RO after having used the cyclic property of the (ill-defined) trace leads to the definition of a consistent unitary, Lorentz invariant theory, as first discovered in Ref. 1 using a different approach. This result cannot be obtained if one adopts the other regularization schemes we have described.

For the Thirring model, a pure fermionic theory where A_μ is an auxiliary field, insertion of the RO before cyclic property of the trace is used, leads to a sensible answer (for R_{B}) coinciding with Klaiber operator result.[16]

In both cases there remains an ambiguity related to an undetermined parameter (a) introduced through the RO. For the CSM this leads to a whole family of consistent solutions provided $a > 1$ (the resulting theory consists of massless excitations plus a massive particle with mass $(g^2/4\pi)[a^2/(a-1)]$ and hence unitarity requires $a > 1$).

For the Thirring model the existence of a one-parameter family of solutions was already known from the operator analysis described in Ref. 16.

We then conclude that the physical situation must be carefully analyzed in order to select a regularization prescription. Moreover, ambiguities arising in this process of regularization can be utilized in order to define a physically sound theory. Although our discussion corresponds to the (simpler) two-dimensional world, we think that our conclusions also hold in other space-time dimensions. Indeed, different specification of fermion determinants and anomalies are also encountered in four-dimensional non-Abelian models, in gravitational theories in $4k+2$ dimensions, etc. In this context, it may be worth while to investigate the introduction of arbitrary parameters that can be used to define consistent quantum theories from potentially anomalous ones. We hope to address this point in a future work.

ACKNOWLEDGMENTS

We want to thank Luis Urrutia for helpful comments.

F.A.S. is partially supported by CIC, Buenos Aires, Argentina.

APPENDIX: DETAILS ON THE CALCULATION OF FERMION DETERMINANTS

In Sec. II C we obtained for the Jacobian associated to the fermionic change of variables (2.10)–(2.12) the result:

$$J = \exp\left(-\int_0^1 \omega'(t)dt\right), \tag{A1}$$

where $\omega'(t)$ takes different forms depending on which regularization is used.

In the first method we used (RegI) we inserted RO after using the cyclic property of the trace and in

$$\begin{aligned}\omega'_{\rm TM}(t)\big|_{\rm RegI} &= \lim_{M^2\to\infty} 2{\rm tr}\, \gamma_5 \phi e^{-R/M^2},\\ \omega'_{\rm CSM}(t)\big|_{\rm RegI} &= \lim_{M^2\to\infty} {\rm tr}\, \gamma_5(\phi - i\eta)e^{-R/M^2}.\end{aligned} \tag{A2}$$

In the second method (RegII) the cyclic property of the trace was not used. We then obtained

$$\begin{aligned}&\omega'_{\rm TM}(t)\big|_{\rm RegII}\\ &= \lim_{M^2\to\infty} {\rm tr}\, \not{D}_t^{-1}[(\gamma_5\phi - i\eta)\not{D}_t\\ &\quad + \not{D}_t(\gamma_5\phi + i\eta)]e^{-R/M^2},\\ &\omega'_{\rm CSM}(t)\big|_{\rm RegII}\\ &= \lim_{M^2\to\infty} {\rm tr}\, \not{D}_t^{-1}[\widetilde{D}_t((1-\gamma_5)/2)(\phi - i\eta)\\ &\quad - ((1+\gamma_5)/2)(\phi - i\eta)\widetilde{D}_t]e^{-R/M^2}.\end{aligned} \tag{A3}$$

In this Appendix we give some details on the calculation of $\omega'(t)$ leading to the results listed in Eqs. (2.35) and (2.36).

We begin with evaluation of $\omega'_{\rm TM}(t)|_{\rm RegI}$. Introducing a representation of δ function we have

$$\begin{aligned}\omega'_{\rm TM}(t)\big|_{\rm RegI} &= \lim_{\substack{M^2\to\infty\\ y\to x}} {\rm tr}\frac{2}{(2\pi)^2}\\ &\quad\times\int d^2x\, d^2k\, e^{iky}e^{-D_t^2/M^2}\gamma_5\phi(x)e^{-ikx}\end{aligned} \tag{A4}$$

and then

$$\begin{aligned}\omega'_{\rm TM}(t)\big|_{\rm RegI} &= \lim_{M^2\to\infty} {\rm tr}\frac{\gamma_5}{2\pi^2}\int \phi(x)\, dx\\ &\quad\times\int d^2k\, e^{-(D_t+k)^2/M^2}.\end{aligned} \tag{A5}$$

Expanding the exponential and making the change of variables $k = Mp$, it is easy to see that only the M^{-2} term in the expansion will contribute in the limit M^2.

Then

$$\omega'_{\rm TM}(t)\big|_{\rm RegI} = \frac{1}{2\pi^2}\int d^2x\, {\rm tr}\,\gamma_5\phi(x)D_t^2\int e^{-p^2}d^2p\,. \tag{A6}$$

Performing the Gaussian integral and using the corresponding RO we finally obtain the result given in (2.35a).

Concerning RegIB note that although $R_{\rm B}$ cannot be written as the square of some operator, one can show the fact that calculations are equivalent to those performed with an $R_{\rm B} = \not{\mathscr{D}}_t^2$, with

$$\not{\mathscr{D}}_t = i\not\partial + g((1+a)/2)\not{A}_t\,. \tag{A7}$$

Now let us consider the second method (RegII). In this case we have

$$\begin{aligned}&\omega'_{\rm TM}(t)\big|_{\rm RegII}\\ &= \lim_{M^2\to\infty} {\rm tr}\, D_t^{-1}[(\gamma_5\phi - i\eta)\not{D}_t\\ &\quad + \not{D}_t(\gamma_5\phi + i\eta)]e^{-R/M^2}.\end{aligned} \tag{A8}$$

The trace in (A8) implies one has to consider $\lim_{x\to y} D_t^{-1}(x,y)$. It is this divergent contribution that is the new ingredient in RegII that leads to a different result compared with RegI.

Inserting

$$\begin{aligned}D_t^{-1}(x,y) &= e^{[\gamma_5\phi(x) + i\eta(x)](1-t)}\\ &\quad\times G_0(x,y)e^{[\gamma_5\phi(y) - i\eta(y)](1-t)}\end{aligned} \tag{A9}$$

with G_0, the free fermion Green's function,

$$G_0(x) = (i/2\pi)(\gamma_\mu x_\mu/x^2) \tag{A10}$$

in (A8) one gets

$$\begin{aligned}\omega'_{\rm TM}(t)\big|_{\rm RegII} &= \lim_{M^2\to\infty}\int d^2x\, d^2k\frac{1}{2\pi}{\rm tr}\\ &\quad\times\left[\frac{\not{k}}{k^2}i\not\partial(\gamma_5\phi + i\eta)\right]e^{-ik(x-y)}.\end{aligned} \tag{A11}$$

Now, following the same steps for the RegI case we get Eqs. (2.36).

[1]R. Jackiw and R. Rajaraman, Phys. Rev. Lett. **54**, 1219, 2060(E) (1985); **55**, 224 (1985).
[2]L. D. Faddeev and L. S. Shatashvili, Phys. Lett. B **167**, 255 (1986).
[3]O. Babelon, F. A. Schaposnik, and C. M. Viallet, Phys. Lett. B **177**, 385 (1986).
[4]K. Harada and I. Tsusui, Phys. Lett. B **183**, 311 (1987).
[5]V. Kulikov, Serpukhov report, 1986.
[6]R. Rajaraman, Phys. Lett. B **162**, 148 (1985); J. LoHa and R. Rajaraman, *ibid.* **165**, 321 (1985).
[7]K. Harada and I. Tsutsui, Tokyo Int. Reports TIT HEP 101,102,112,113, 1987.
[8]S. Ryang, Phys. Rev. D **35**, 3158 (1987).
[9]N. K. Falck and G. Krammer, Phys. Lett. B **193**, 257 (1987); Ann. Phys. (NY) **176**, 330 (1987).
[10]F. A. Schaposnik and J. Webb, Z. Phys. C **34**, 367 (1987).
[11]Y. Zhang, Z. Phys. C **36**, 449 (1987).
[12]K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979); **44**, 2995 (1980); Phys. Rev. D **21**, 2848 (1980); **22**, 1499 (E) (1981); **23**, 2262 (1981).
[13]R. E. Gamboa Saravi, M. A. Muschietti, F. A. Schaposnik, and J. E. Solomin, Ann. Phys. (NY) **157**, 360 (1984).
[14]For a review and references to the original literature see S. Treiman, R. Jackiw, B. Zumino, and E. Witten, *Current Algebras and Anomalies* (Princeton U. P., Princeton, NJ, 1985).
[15]M. Rubin, J. Phys. A **19**, 2105 (1986).
[16]B. Klaiber, in *Lectures in Theoretical Physics*, edited A. O. Barut (Gordon and Breach, New York, 1968), and references therein.
[17]L. Alvarez Gaume and P. Ginsparg, Nucl. Phys. B **243**, 449 (1984).
[18]See, for example, R. Jackiw in *Quantum Mechanics of Fundamental Systems*, edited by C. Teitelboim (Plenum, New York, 1988).
[19]F. A. Schaposnik, Univ. Paris report 85/52, 1985.
[20]R. Banerjee, Phys. Rev. Lett. **56**, 1889 (1986).
[21]J. Webb, Z. Phys. C **31**, 301 (1986).
[22]T. Kubota and I. Tsutsui, Phys. Lett. B **173**, 77 (1986).
[23]S. H. Yi, D. K. Park, and B. H. Cho, Mod. Phys. Lett. A **2**, 579 (1987); Mod. Phys. Lett. A **3**, 201 (1988).
[24]R. E. Gamboa Saravi, M. A. Muschietti, F. A. Schaposnik, and J. E. Solomin, to be published in Lett. Math. Phys.
[25]J. Alfaro, L. F. Urrutia, and J. D. Vergara, Phys. Lett. B **202**, 121 (1988).
[26]N. K. Falch, DESY preprint, 1988.
[27]K. Furuya, R. E. Gamboa Saravi, and F. A. Schaposnik, Nucl. Phys. B **208**, 159 (1982).
[28]A. Andrianov and L. Bonora, Nucl. Phys. B **233**, 247 (1984).
[29]K. Fujikawa, Phys. Rev. D **31**, 341 (1985).
[30]K. Fujikawa, in *Proceedings of Kyoto Summer Institute*, edited by T. Inami (World Scientific, Singapore, 1986).
[31]W. A. Bardeen and B. Zumino, Nucl. Phys. B **244**, 421 (1984).
[32]J. D. Bjorken, Phys. Rev. **148**, 1467 (1966).
[33]K. Johnson and F. E. Low, Prog. Theor. Phys. (Kyoto), Suppl. **37**, and **38**, 74 (1966).
[34]S. Jo, Nucl. Phys. B **259**, 616 (1985).
[35]M. V. Manias, M. C. von Reichenbach, F. A. Schaposnik, and M. Trobo, J. Math. Phys. **28**, 1632 (1987).
[36]H. Montani, La Plata Univ. report, 1988.

Generalized Skyrme model on higher-dimensional Riemannian manifolds

Yisong Yang
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003

(Received 9 February 1988; accepted for publication 23 November 1988)

The Skyrme model is generalized to higher-dimensional Riemannian manifolds and the geometric conditions under which homothetic maps are absolute or local minimizers of the model are obtained.

I. INTRODUCTION

In the classical Skyrme model[1] the field configurations of baryons are minimizers among maps ρ: $\mathbf{R}^3 \rightarrow \mathrm{SU}(2)$, with a nontrivial "degree" of an energy functional $E(\rho)$ obtained by adding a higher-order nonlinear term to the σ-model functional; the degree of the minimizers is recognized as the baryon number of the baryons. This model lacks an exact analytic solution and mathematical existence results appeared only very recently.[2] (In Ref. 3 a family of extremals of the Skyrme energy are found under the radial ansätz.) However, on the other hand, since the model is only an approximation to the real world, it seems promising to have the model modified in various ways. It is not difficult to show that if ρ is a well-behaved finite energy field configuration, it must approach a constant matrix at infinity. Hence, intuitively, the space $\mathbf{R}^3$ in the Skyrme model can be topologically compactified to S^3 and one might as well study the modified model for field configurations ρ:$S^3 \rightarrow \mathrm{SU}(2) (=S^3)$ with the correspondingly corrected energy functional. For this reason, Manton and Ruback[4] established a compact version of the Skyrme model through the introduction of the energy

$$E(\rho) = \int_M \Omega_g \{\sigma_1(g^{-1}\rho^*h) + \sigma_2(g^{-1}\rho^*h)\},$$

where (M,g) and (N,h) are compact orientable three-dimensional Riemannian manifolds without boundary, $\rho \in C^\infty(M,N)$, ρ^*h is the pullback of the metric h under ρ, Ω_g denotes the canonical volume element of (M,g) and $\sigma_i(A)$ are coefficients of the characteristic polynomial of the $n \times n$ matrix A determined by the formula

$$\det(A - \lambda I) = \sum_{i=0}^{n} (-\lambda)^{n-i} \sigma_i(A).$$

In the classical limit, $M = S^3$, $N = S^3 (= \mathrm{SU}(2))$, Manton and Ruback[4] showed that up to isometries the identity map is the unique absolute minimizer of F among maps with nontrivial degrees; they conjectured further that this is true for the model $M = (1/\tau)S^3$ (the standard sphere of radius $1/\tau$) for $\tau \geqslant 1$, $N = S^3$, but not true for $\tau < 1$.

In the recent independent studies of Loss[5] and Manton,[6] it was shown that for maps ρ:$M \rightarrow N$ with degree k (in Ref. 6, $k=1$) between three-dimensional Riemannian manifolds,

$$E(\rho) \geqslant 3|M|(\mu^2 + \mu^4) \tag{1}$$

provided that $\mu^3 \equiv |k| |N| / |M| \geqslant 1$. Moreover, if (M,g) and (N,h) are homothetic,[7] i.e., there is a diffeomorphism ψ:$M \rightarrow N$ such that $\psi^*h = \tau^2 g$ for some constant $\tau^2 > 0$, then the lower bound estimate (1) for $k=1$ is uniquely saturated (up to isometries) by the homothetic map ψ provided that $\tau^2 \geqslant 1$. In particular, the first part of the conjecture of Manton and Ruback[4] was proved. Loss[5] and Manton[6] also showed that ψ is a local stable minimizer if $\tau^2 > \frac{1}{2}$, but it may fail to be so for $\tau^2 < \frac{1}{2}$. This picture confirms very well the opinion of Manton and Ruback[4] that the mathematical elegance of Skyrme's original formulation becomes clear if one allows the possibility that space is curved and that the skyrmion energy is a nice measure of the size and shape of space; the new version of the Skyrme model may be relevant to the quark confinement problem.

It appears that a higher-dimensional Skyrme model for field configurations ρ:$(M,g) \rightarrow (N,h)$ with $\dim M = \dim N = n \geqslant 3$ should be established through the introduction of the energy

$$E(\rho) = \int_M \Omega_g \{\sigma_1(g^{-1}\rho^*h) + \sigma_{n-1}(g^{-1}\rho^*h)\}. \tag{2}$$

Physically, this is a modified σ-model suitable for higher-dimensional target space; mathematically, in this model the energy contains the interesting term yielding harmonic maps.[7,8] Our study below will follow the main line of Loss[5].

II. GENERALIZED SKYRME MODEL AND ENERGY BOUND ESTIMATES

We assume throughout this paper that (M,g) and (N,h) are n-dimensional compact orientable Riemannian manifolds without boundary. For smooth maps ρ:$M \rightarrow N$ we might as well consider a more general model defined by the energy

$$E_m(\rho) = \int_M \Omega_g \{\sigma_m(g^{-1}\rho^*h) + \sigma_{n-m}(g^{-1}\rho^*h)\},$$

where $2 \leqslant 2m < n$. (The equality $2m = n$ will trivialize the model; see Sec. IV.)

We denote by C_n^k the usual binomial coefficients, with the convention

$$C_n^k = \frac{n!}{k!(n-k)!}, \quad k = 1,\ldots,n-1,$$

$$C_n^n = C_n^0 = 1, \quad C_n^k = 0 \text{ for other } k.$$

Lemma 2.1: Suppose A is an $n \times n$ semipositive definite symmetric matrix. Then

$$\sigma_k(A) \geqslant C_n^k [\sigma_n(A)]^{C_{n-1}^{k-1}/C_n^k}, \quad k = 1,\ldots,n-1;$$

the equality holds if and only if A is a scalar matrix.

Proof: Let $\det(A - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$; then $\lambda_j \geqslant 0$ and

 0022-2488/89/040824-05$02.50

$$\sigma_k(A) = \sum_{\{n_1,\ldots,n_k\}\subset\{1,\ldots,n\}} \lambda_{n_1}\cdots\lambda_{n_k}.$$

From the arithmetic mean–geometric mean inequality we have

$$\sigma_k(A)/C_n^k \geqslant (\lambda_1\cdots\lambda_n)^{C_{n-1}^{k-1}/C_n^k};$$

the equality holds if and only if $\lambda_1 = \cdots = \lambda_n$.

For a finite set S, let $\#S$ denote the number of points of S. Let R_ρ be the set of regular values of $\rho:M\to N$ and

$$k_\rho \equiv \inf\{\#\rho^{-1}(q)\,|\,q\in R_\rho\}.$$

We shall use the notation

$$|M| = \mathrm{vol}(M) \equiv \int_M \Omega_g, \quad |N| = \mathrm{vol}(N) \equiv \int_N \Omega_h.$$

Lemma 2.2:

$$\int_M \Omega_g \sigma_n^{1/2}(g^{-1}\rho^*h) \geqslant k_\rho |N|.$$

Proof: Take $q\in R_\rho$. Let $V\subset N$ be an open set such that $\phi:V\to\mathbf{R}^n$ is a positively oriented chart, with $\phi(y) = (y^1,\ldots,y^n)$ for $y\in V$. Then

$$(\phi^{-1})^*\Omega_h = (\det(h_{ij}))^{1/2}\, dy^1\wedge\cdots\wedge dy^n.$$

Assume $q\in V$ and V is sufficiently small such that $\rho^{-1}(V) = \cup_1^m U_j$; U_j are open in M; $\rho:U_j\to V$ is a diffeomorphism; and $U_i\cap U_j = \emptyset$, $i\neq j$. We suppose, also, that U_j is sufficiently small so that we can choose a positively oriented chart $\psi_j:U_j\to\mathbf{R}^n$ to make

$$(\psi_j^{-1})^*\sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g = \mathrm{sgn}(\rho|_{U_j})(\phi^{-1})^*\Omega_h,$$

where $\mathrm{sgn}(\rho|_{U_j}) = +1$ (respectively, -1) if ρ is orientation preserving (respectively, reversing) on U_j. Consequently,

$$\begin{aligned}\int_{U_j} &\sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g \\ &= \int_{\psi_j(U_j)} (\psi_j^{-1})^*\sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g \\ &= \int_{\mathrm{sgn}(\rho|_{U_j})\phi(V)} \mathrm{sgn}(\rho|_{U_j})(\phi^{-1})^*\Omega_h \\ &= \int_{\phi(V)} (\phi^{-1})^*\Omega_h = \int_V \Omega_h. \end{aligned} \tag{3}$$

Therefore,

$$\int_{\rho^{-1}(V)} \sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g = m\int_V \Omega_h \geqslant k_\rho \int_V \Omega_h.$$

Given $\epsilon>0$, there exist the above chosen $V_j\subset N$, $j=1,\ldots,l$, so that $V_i\cap V_j=\emptyset$ and

$$0\leqslant |N| - \int_{\cup_{j=1}^l V_j} \Omega_h < \epsilon.$$

But $\rho^{-1}(V_i)\cap\rho^{-1}(V_j) = \emptyset$, $i\neq j$, so

$$\begin{aligned}\int_M \sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g &\geqslant \sum_{j=1}^l \int_{\rho^{-1}(V_j)} \sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g \\ &\geqslant k_\rho \sum_{j=1}^l \int_{V_j}\Omega_h \geqslant k_\rho(|N|-\epsilon).\end{aligned}$$

The lemma is proved.

Lemma 2.3: If $\rho:M\to N$ is a k-fold covering, then

$$\int_M \sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g = k|N|.$$

Proof: Given $\epsilon>0$, choose a finite open covering $\{V_j\,|\,j=1,\ldots,l\}$ of N such that

$$0<\sum_{j=1}^l \int_{V_j}\Omega_h - |N| < \epsilon,$$

$\rho^{-1}(V_j) = \cup_{j=1}^k U_{ij}$; $U_{ij}\cap U_{i'j} = 0$, $i\neq i'$; $\rho:U_{ij}\to V_j$ is a diffeomorphism; $i=1,\ldots,k$, $j=1,\ldots,l$; and on every U_{ij} or V_j there is defined a positively oriented chart. Since $\{U_{ij}\,|\,i=1,\ldots,k, j=1,\ldots,l\}$ is a covering of M, we have from (3) that

$$\begin{aligned}\int_M &\sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g \\ &\leqslant \sum_{i=1}^k\sum_{j=1}^l \int_{U_{ij}} \sigma_n^{1/2}(g^{-1}\rho^*h)\Omega_g \\ &= \sum_{i=1}^k\sum_{j=1}^l \int_{V_j}\Omega_h < k(|N|+\epsilon).\end{aligned}$$

The desired equality follows from Lemma 2.2 and the above inequality.

We can now state the main result of this section.

Theorem 2.4: For $\rho:M\to N$, $k_\rho\geqslant k\geqslant 1$, $\mu\equiv(k|N|/|M|)^{1/n}$, and $2m<n$, we have

(i)

$$E_m(\rho)\geqslant C_n^m|M|(\mu^{2m}+\mu^{2(n-m)}), \quad \text{if } \mu\geqslant 1,$$

(ii)

$$E_m(\rho)\geqslant 2kC_n^m|N|, \quad \text{if } \mu<1.$$

In the case of (i), equality holds if and only if ρ is a k-fold covering satisfying $\rho^*h = \mu^2 g$.

Proof: From Lemma 2.1 we have

$$\sigma_m(g^{-1}\rho^*h)\geqslant C_n^m[\sigma_n(g^{-1}\rho^*h)]^{C_{n-1}^{m-1}/C_n^m}$$

and

$$\begin{aligned}\sigma_{n-m}(g^{-1}\rho^*h) &\geqslant C_n^{n-m}[\sigma_n(g^{-1}\rho^*h)]^{C_{n-1}^{n-m-1}/C_n^{n-m}} \\ &= C_n^m[\sigma_n(g^{-1}\rho^*h)]^{C_{n-1}^m/C_n^m}.\end{aligned}$$

Thus

$$E_m(\rho)\geqslant C_n^m(A+B), \tag{4}$$

where

$$A\equiv\int_M \Omega_g[\sigma_n(g^{-1}\rho^*h)]^{C_{n-1}^{m-1}/C_n^m},$$

$$B\equiv\int_M \Omega_g[\sigma_n(g^{-1}\rho^*h)]^{C_{n-1}^m/C_n^m}.$$

Using the formula $C_{n-1}^m + C_{n-1}^{m-1} = C_n^m$ we obtain, by Lemma 2.2 and the Schwarz inequality,

$$k|N|\leqslant\int_M \Omega_g[\sigma_n(g^{-1}\rho^*h)]^{(C_{n-1}^{m-1}+C_{n-1}^m)/2C_n^m}\leqslant A^{1/2}B^{1/2}. \tag{5}$$

Moreover, since $p\equiv(C_{n-1}^m/C_{n-1}^{m-1})>1$, it yields, from the Hölder inequality,

$$A \leqslant \left(\int_M \Omega_g [\sigma_n (g^{-1}\rho^* h)]^{(C_{n-1}^{m-1}/C_n^m)p} \right)^{1/p} |M|^{1/q}$$
$$= |M|^{1/q} B^{1/p}. \tag{6}$$

Combining (4)–(6) we obtain, by setting $\alpha \equiv A/|M|$ and $\beta \equiv B/|M|$, the inequality

$$E_m(\rho) \geqslant |M| C_n^m \min\{(\alpha+\beta) \mid \alpha\beta \geqslant \mu^{2n}, \alpha^p \leqslant \beta\}$$
$$= |M| C_n^m \min\{\alpha + \max(\mu^{2n}/\alpha, \alpha^p)\}$$
$$= |M| C_n^m \min\{f(\alpha) \mid \alpha > 0\},$$

where

$$f(\alpha) = \begin{cases} \alpha + \alpha^p, & \alpha \geqslant \alpha_0 \equiv \mu^{2n/(p+1)}, \\ \alpha + \mu^{2n}/\alpha, & \alpha < \alpha_0. \end{cases}$$

From this we easily conclude that

$$\min\{f(\alpha) \mid \alpha > 0\} = \begin{cases} f(\alpha_0) = \mu^{2m} + \mu^{2(n-m)}, & \mu \geqslant 1, \\ f(\mu^n) = 2\mu^n, & \mu < 1. \end{cases}$$

For $\rho: M \to N$, with $k_\rho \geqslant k$, if the equality in case (i) in Theorem 2.4 holds, then from Lemma 2.1 and the Schwarz inequality we reach (in local coordinates) $g^{-1}\rho^* h = \tau^2 I$ for some real constant τ. Hence $\tau^2 = \mu^2$. Now $\rho^* h = \mu^2 g$, $\rho: M \to N$ must be a local diffeomorphism and consequently, by the compactness of M, a covering map. From Lemma 2.2, $\rho: M \to N$ must be a k-fold covering.

Sufficiency is similarly proved by using Lemma 2.3.

Corollary 2.5: For $\rho: M \to N$, let $k = \deg(\rho)$ and assume $\mu \equiv (|k| |N|/|M|)^{1/n} \geqslant 1$. Then $E_m(\rho) \geqslant C_n^m |M| (\mu^{2m} + \mu^{2(n-m)})$; equality holds if and only if ρ is a k-fold covering which is orientation preserving or reversing and satisfies $\rho^* h = \mu^2 g$.

Proof: The result follows from Theorem 2.4 and $k_\rho \geqslant |k|$.

We shall call a smooth map $\psi: (M,g) \to (N,h)$ *k-homothetic* if ψ is a k-fold covering map and $\psi^* h = \tau^2 g$ for some constant τ. One-homothetic maps are conventionally called homothetic.

From Corollary 2.5 and the argument given in Ref. 5 the following result is immediate.

Corollary 2.6: Let $\psi: (M,g) \to (N,h)$ be homothetic, with $\rho^* h = \tau^2 g$. If $\tau^2 \geqslant 1$, then up to isometries ψ is the unique minimizer of E_m among all maps ρ of nontrivial degrees, and, hence,

$$\min\{E_m(\rho) \mid \rho: M \to N, \deg(\rho) \neq 0\}$$
$$= E_m(\psi) = C_n^m |M| (\tau^{2m} + \tau^{2(n-m)}).$$

Topology can often be used as a criterion for establishing some physically interesting nonexistence results.

For example, if N is simply connected then (M,g) and (N,h) can never be k-homothetic for $k > 1$ since for a covering map $\rho: M \to N$ there is an induced surjection $\pi_1(N,q) \to \rho^{-1}(q)$ for every $q \in N$. Moreover, if M is simply connected, up to isometries there can only exist, at most, one k-homothetic map $(M,g) \to (N,g)$ for $k = 1,2,\ldots$ because any two covering maps $M \to N$ must be equivalent. In particular, we can conclude that the energy lower bound obtained in Refs. 4 and 5 for the three-dimensional *spherical* Skyrme model can *never* be attained for $|k| > 1$.

III. STABILITY ANALYSIS OF THE HOMOTHETIC MAPS

If $\psi: (M,g) \to (N,h)$ is homothetic, with $\psi^* h = \tau^2 g$ and $\tau^2 \geqslant 1$, then ψ is an absolute minimizer among maps $M \to N$ with nontrivial degrees of the energy E_m. It is not difficult to show that ψ is always a critical point of E_m whatever the value of τ^2 (see the computation below); hence it will be interesting to know whether or not ψ can still be a stable local minimizer of E_m for $\tau^2 < 1$. For a three-dimensional model Loss[5] and Manton[6] have shown via the second-variation analysis that ψ is no longer an absolute minimizer for $\tau^2 < \frac{1}{2}$, but that for $\tau^2 > \frac{1}{2}$ the second variation of the energy functional is positive, which reveals that in this geometric range ψ is a stable local minimizer. Thus, the critical value of τ^2 for stability transition in the three-dimensional Skyrme model is $\tau_c^2 = \frac{1}{2}$. We believe this critical value can be reduced by making the space dimension larger.

Let $\rho: M \to N$ be a smooth map. Then $\rho = \psi \circ \phi$, where $\phi = \psi^{-1} \circ \rho$. Hence

$$F_m(\phi) \equiv E_m(\rho)$$
$$= \int_M \Omega_g \{\tau^{2m} \sigma_m (g^{-1}\phi^* g)$$
$$+ \tau^{2(n-m)} \sigma_{n-m} (g^{-1}\phi^* g)\}.$$

Now the stability problem of ψ is equivalent to the stability problem of $\phi = id: M \to M$.

Let $\phi_t: M \to M$ be a flow generated by an arbitrary vector field $X: M \to TM$. As usual, let $\mathscr{L}_X$ denote the Lie derivative with respect to X. Our computation below will follow a somewhat indirect path.

In local coordinate representations, let $\lambda_1(t),\ldots,\lambda_n(t)$ be the eigenvalues of matrix $g^{-1}\phi_t^* g$ at a fixed point $x \in M$. Then

$$\sigma_k (g^{-1}\phi_t^* g) = \sum_{\{n_1,\ldots,n_k\} \subset \{1,\ldots,n\}} \lambda_{n_1}(t) \cdots \lambda_{n_k}(t).$$

Hence

$$\frac{d}{dt}(\sigma_k (g^{-1}\phi_t^* g))|_{t=0}$$
$$= C_{n-1}^{k-1} (\lambda_1'(0) + \cdots + \lambda_n'(0))$$
$$= C_{n-1}^{k-1} \frac{d}{dt} (\sigma_1 (g^{-1}\phi_t^* g))|_{t=0} = C_{n-1}^{k-1} \sigma_1 (g^{-1} \mathscr{L}_X g)$$

because $\lambda_1(0) = \cdots = \lambda_n(0) = 1$.

Therefore, $\phi = id$ is always a critical point of F_m, as can be seen from

$$\frac{d}{dt} F_m(\phi_t)|_{t=0} = (\tau^{2m} C_{n-1}^{m-1} + \tau^{2(n-m)} C_{n-1}^{n-m-1})$$
$$\times \int_M \sigma_1 (g^{-1} \mathscr{L}_X g) \Omega_g$$
$$= (\tau^{2m} C_{n-1}^{m-1} + \tau^{2(n-m)} C_{n-1}^{m})$$
$$\times \int_M \operatorname{div}_g(X) \Omega_g = 0.$$

Furthermore,

$$
\begin{aligned}
\frac{d^2}{dt^2}&(\sigma_k(g^{-1}\phi_t^*g))|_{t=0}\\
&= C_{n-1}^{k-1}(\lambda_1''(0)+\cdots+\lambda_n''(0))\\
&\quad + C_{n-2}^{k-2}\,C_{n-1}^{k-1}\Big(\sum_{i<j}\lambda_i'(0)\lambda_j'(0)\Big)\\
&= C_{n-1}^{k-1}\,\frac{d^2}{dt^2}(\sigma_1(g^{-1}\phi_t^*g))|_{t=0}\\
&\quad + \frac{1}{2}\,C_{n-2}^{k-2}\,C_{n-1}^{k-1}(\sigma_1^2(g^{-1}\mathscr{L}_Xg)\\
&\quad - \sigma_1([g^{-1}\mathscr{L}_Xg]^2))
\end{aligned}
$$

and consequently,

$$
\begin{aligned}
\frac{d^2}{dt^2}&F_m(\phi_t)|_{t=0}\\
&= \frac{1}{2}\,(\tau^{2m}C_{n-2}^{m-2}\,C_{n-1}^{m-1} + \tau^{2(n-m)}C_{n-2}^{m}\,C_{n-1}^{m})\\
&\quad \times\int_M \Omega_g\{\sigma_1^2(g^{-1}\mathscr{L}_Xg) - \sigma_1([g^{-1}\mathscr{L}_Xg]^2)\}\\
&\quad + (\tau^{2m}\,C_{n-1}^{m-1} + \tau^{2(n-m)}C_{n-1}^{m})\\
&\quad \times\frac{d^2}{dt^2}\left\{\int_M \sigma_1(g^{-1}\phi_t^*g)\Omega_g\right\}\Big|_{t=0}.
\end{aligned}
\tag{7}
$$

One can derive from the semigroup property of the flow ϕ_t that

$$
\frac{d}{dt}\int_M \sigma_1(g^{-1}\phi_t^*g)\Omega_g = \int_M \sigma_1(g^{-1}\phi_t^*\mathscr{L}_Xg)\Omega_g
$$

and thus

$$
\begin{aligned}
\frac{d^2}{dt^2}&\left\{\int_M \sigma_1(g^{-1}\phi_t^*g)\Omega_g\right\}\Big|_{t=0}\\
&= \frac{d}{dt}\left\{\int_M \sigma_1((\phi_{-t}^*g)^{-1}\mathscr{L}_Xg)\Omega_{\phi_{-t}^*g}\right\}\Big|_{t=0}\\
&= \int_M \left\{\sigma_1([g^{-1}\mathscr{L}_Xg]^2) - \frac{1}{2}\,\sigma_1^2(g^{-1}\mathscr{L}_Xg)\right\}\Omega_g.
\end{aligned}
\tag{8}
$$

Substituting (8) into (7) we obtain

$$
\begin{aligned}
\frac{d^2}{dt^2}&F_m(\phi_t)|_{t=0}\\
&= l_1(n,m,\tau)\int_M \sigma_1(H^2)\Omega_g\\
&\quad + l_2(n,m,\tau)\int_M \sigma_1^2(g^{-1}\mathscr{L}_Xg)\Omega_g,
\end{aligned}
$$

where

$$
\begin{aligned}
&H \equiv g^{-1}\mathscr{L}_Xg - (1/n)\sigma_1(g^{-1}\mathscr{L}_Xg)I,\\
&l_1(n,m,\tau) = \tau^{2m}C_{n-1}^{m-1}(1-\tfrac{1}{2}C_{n-2}^{m-1})\\
&\qquad\qquad + \tau^{2(n-m)}C_{n-1}^{m}(1-\tfrac{1}{2}C_{n-2}^{m}),\\
&l_2(n,m,\tau)\\
&\quad = \tau^{2m}\,C_{n-1}^{m-1}\left(\frac{1}{n}-\frac{1}{2}+\frac{1}{2}\Big(1-\frac{1}{n}\Big)C_{n-2}^{m-2}\right)\\
&\qquad + \tau^{2(n-m)}\,C_{n-1}^{m}\left(\frac{1}{n}-\frac{1}{2}+\frac{1}{2}\Big(1-\frac{1}{n}\Big)\,C_{n-2}^{m}\right).
\end{aligned}
$$

Thus if $l_1(n,m,\tau)>0$ and $l_2(n,m,\tau)>0$, we must have $\{d^2F_m(\phi_t)/dt^2\}|_{t=0}\geqslant 0$; the equality holds if and only if $X:M\to TM$ is a Killing vector field. Therefore we can conclude with the following theorem.

Theorem 3.1: If $\psi:(M,g)\to(N,h)$ is a homothetic map with $\psi^*h=\tau^2g$ and $l_1(n,m,\tau)$, $l_2(n,m,\tau)>0$, then ψ is a stable local minimizer of the energy functional E_m in the sense that for any nonisometry flow $\phi_t: M\to M$, $\{d^2F_m(\phi_t)/dt^2\}|_{t=0}>0$.

Choosing $m=1$ [model (2)] as an important example, we have

$$
\begin{aligned}
&l_1(n,1,\tau) = \tau^2 + \tfrac{1}{2}\tau^{2(n-1)}(n-1)(4-n),\\
&l_2(n,1,\tau) = \tau^2(1/n - \tfrac{1}{2})\\
&\qquad\qquad + \tau^{2(n-1)}(2/n + n/2 - 2)(n-1).
\end{aligned}
$$

Accordingly, for $n=3$ we have $\tau^2>\frac{1}{2}$, as was obtained by Loss[5] and Manton[6]; for $n=4$ we have $\tau^4>\frac{1}{6}$, for $n=5$ we have $\frac{1}{2}>\tau^6>\frac{1}{12}$, and for $n=6$ we have $\frac{1}{5}>\tau^8>\frac{1}{20}$. In general we must require

$$
\frac{n-2}{(4+n(n-4))(n-1)} < \tau^{2(n-2)} < \frac{2}{(n-1)(n-4)},\quad n\geqslant 5
$$

to ensure that $l_1(n,1,\tau)$, $l_2(n,1,\tau)>0$.

In the parameter range

$$
2/(n-1)(n-4) < \tau^{2(n-2)} < 1,\quad \text{where } n\geqslant 5, \tag{9}
$$

$l_1<0$, $l_2>0$. If there is a non-Killing vector field $X:M\to TM$ so that $\sigma_1(g^{-1}\mathscr{L}_Xg)=0$ on M, then $H\neq 0$; hence $\{d^2F_m(\phi_t)/dt^2\}|_{t=0}<0$ for the one-parameter flow $\phi_t: M\to M$ generated from X. Therefore, the homothetic map $\psi:M\to N$ is no longer a local minimizer for τ in the range (9) when $n\geqslant 5$. In general we have nothing conclusive about this range.

IV. REMARKS

(i) If n is even and $2m=n$, let us consider another possible Skyrme model:

$$
E(\rho) = \int_M \sigma_m(g^{-1}\rho^*h)\Omega_g,
$$

where $\rho:M\to N$ is smooth.

Using Lemmas 2.1–2.3 and Theorem 2.4 we have

$$
E(\rho)\geqslant C_{2m}^{m}\int_M \sigma_{2m}^{1/2}(g^{-1}\rho^*h)\Omega_g \geqslant k_\rho C_{2m}^{m}|N|;
$$

the equality holds if and only if ρ is a k_ρ-homothetic map, with

$$
\tau^2 = (k_\rho|N|/|M|)^{1/m}.
$$

In particular there is no restriction on the range of $k_\rho|N|/|M|$. Consequently, the above model may be of little interest because not much geometry is captured. This example also shows that there is an effective compensation between the lower- and higher-order nonlinear terms σ_m and σ_{n-m}; when $m=n-m$, some interesting geometric restrictions will be absent.

(ii) An analogous investigation can be made for the following "full" Skyrme model:

$$E(\rho) = \sum_{1 < m < n-1} \epsilon_m \int_M \sigma_m (g^{-1}\rho^* h)\Omega_g,$$

where $\rho:(M,g) \to (N,h)$ is smooth and $\epsilon_m > 0$, $m = 1,2,\ldots,n-1$.

ACKNOWLEDGMENT

I wish to thank the referee for pointing out Refs. 3 and 6.

[1]T. H. R. Skyrme, Proc. R. Soc. London Ser. A **260**, 127 (1961); Nucl. Phys. **31**, 556 (1962).
[2]M. Estéban, Commun. Math. Phys. **105**, 571 (1986).
[3]V. N. Romanov, I. V. Frolov, and A. S. Schwarz, Theor. Math. Phys. **37**, 1038 (1978).
[4]N. S. Manton and P. J. Ruback, Phys. Lett. B **181**, 137 (1986).
[5]M. Loss, Lett. Math. Phys. **14**, 149 (1987).
[6]N. S. Manton, Commun. Math. Phys. **111**, 469 (1987).
[7]G. Tóth, *Harmonic and Minimal Maps* (Wiley, New York, 1984).
[8]J. Eells, Jr., and J. H. Sampson, Am. J. Math. **86**, 109 (1964).

Configurational interference in boundary-value problems governed by the Helmholtz equation

G. S. Singh
Department of Physics, University of Ranchi, Ranchi 834008, India
S. N. Singh
Department of Physics, Ranchi College, Ranchi 834008, India

(Received 23 June 1988; accepted for publication 2 November 1988)

A rigorous approach is discussed for solving the two-dimensional Helmholtz equation in a multiply-connected domain consisting of a ring of N circles distributed symmetrically within a closed space. The outer boundary has been taken to be such that the system as a whole has N-fold rotational symmetry. The Dirichlet boundary condition has been satisfied exactly at the outer as well as at each of the inner edges, using the addition theorems for the cylindrical Bessel functions in conjugation with the Fourier expansions. Numerical results, showing spatial configurational interference, are presented for the lowest cutoff value of the symmetric mode as a function of separation between centers of two inner circles, in the case $N = 2$ with circular outer boundary. The application of the method to various problems of physics and engineering is enunciated.

I. INTRODUCTION

Singh and co-workers[1,2] have obtained the solution of the two-dimensional Helmholtz equation

$$\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\left(\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\left(\frac{\partial^2}{\partial\theta^2}\right)+k^2\right]\psi(r,\theta)=0, \quad (1)$$

in a mathematically compact and computationally simple form for an eccentric annular circular region. The approach has been applied to study vibrations of membranes,[3] higher-order cutoff wavenumbers of TE and TM modes in electromagnetic waveguides,[4] and control rod problems in reactor physics.[5] These applications encompass both the Dirichlet and the Neumann boundary conditions. The method has been further extended to deal with the polarization characteristics of eccentric-core circular optical fibers.[6] However, a number of systems of technological importance coming within the purview of the Helmholtz equation may have the cross-sectional view of multiply-connected regions. The present paper considers the generalization of the solution of Eq. (1) to such domains.

Lin[7] presented an analytical scheme to calculate the admissible acoustic propagation modes of fluid in a circular duct containing an assembly of circular cylinders, as might occur in gas-cooled fast breeder reactors and advanced gas-cooled reactors. Murray[8] has discussed a method for calculating the effect of a number of control rods in thermal nuclear reactors requiring large reactivity control. But our comments,[3] where we have pointed out the errors in, and corrected the expressions of, works dealing with the vibrations of membranes,[9] would also affect the expressions obtained in Refs. 7 and 8.

Lamarsh[10] has done an approximate but analytical study of the effect of a ring of cylindrical control rods in bare cylindrical reactors. On the other hand, exact investigations by Saito and Nagaya[11] for thin plates with circular holes are limited to the infinite systems only. Yamashita *et al.*[12] have used the group-theoretic approach, as suggested by McIsaac,[13] for processing the symmetry of multiple dielectric waveguides, combined with the point-matching procedure for analyzing the optical fibers with symmetrically distributed multiple cores. Hence an exact analytical approach is needed for investigations of the systems characterized by finite domains of multiple connectivity.

The central issue in obtaining a rigorous solution for the multicentered problems under consideration is the search for an approach which could be utilized to satisfy the boundary conditions at all the edges exactly. In order to achieve this, we exploit Graf's addition theorems[14] for the cylindrical Bessel functions together with the Fourier expansions. It seems pertinent to mention that in contrast to the views expressed by Lin[7] as well as by Nagaya[15] (which were without any logical support), the use of the Fourier expansions is an exact method for satisfying the boundary conditions.[2]

Section II contains the detailed account of the solution. The numerical results, along with the discussions, are presented in Sec. III. Finally, Sec. IV gives the concluding remarks wherein the potentiality of our approach to various problems of physics and engineering are enumerated.

II. SOLUTION

Figure 1 shows the geometrical configurations discussed in this paper. The system consists of N circles, each of radius a, whose centers lie along the circumference of a circle

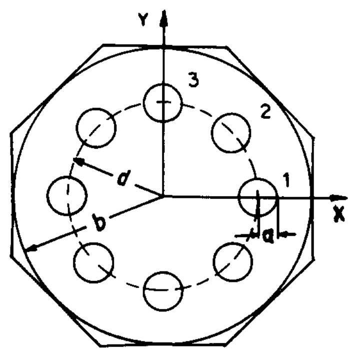


FIG. 1. Multiply-connected domain with N-fold rotational symmetry.

 0022-2488/89/040829-05$02.50

of radius d, enclosed concentrically within an N-sided regular polygon. A special case would correspond to the situation when the outer boundary would degenerate into a circle of radius b. We take the x-axis to be along the line joining points O and O_1, which are centers of the outer construction and one of the inner ones designated as circle 1, respectively. The centers of other inner circles, numbered 2,3,..., in sequence moving along the anticlockwise direction, are distributed symmetrically about this axis. The quantity d defines the eccentricity of the problem, and the system as a whole has N-fold rotational symmetry.

Let P be any point interior to the outer boundary but exterior to each of the inner ones. Figure 2 depicts the polar coordinates (r,θ) and (r_ℓ,θ_ℓ) with respect to O and O_ℓ, the latter point being the center of the ℓth circle. The separation between O_1 and O_ℓ is given by $d_{1\ell} = 2d \sin \frac{1}{2}\phi_\ell$, where $\phi_\ell = 2\pi(\ell - 1)/N$ is the angular separation of the ℓth point from the x-axis.

The solution of Eq. (1) with the origin at O is that corresponding to interior points of a simply connected domain as given by

$$\psi_0 = \sum_{m=0}^{\infty} \epsilon_m A_m f_m(\theta_0)\, J_m(kr)\sin(m\theta + \theta_0), \tag{2}$$

where A_m is an unknown constant, $J_m(kr)$ is the Bessel function of the first kind, and

$$\epsilon_m = 2 - \delta_{m,0}, \tag{3}$$

with $\delta_{m,0}$ as the Kronecker delta function.

The parameter θ_0 takes the values 0 and $\pi/2$ for the antisymmetric and the symmetric parts of the solutions, repectively. Moreover, we have introduced a function

$$f_m(\theta_0) = 1 - \cos\theta_0 \delta_{m,0}, \tag{4}$$

which would be needed later on to take care of the fact that the summation over m starts from $\cos\theta_0$, i.e., from 0 (1) in the symmetric (antisymmetric) case.

On the other hand, the solution of Eq. (1), with respect to the origin at (r_ℓ,θ_ℓ) and satisfying the Dirichlet boundary condition along the corresponding circular edge, can be written in the form

$$\psi_\ell = \sum_{m=0}^{\infty} \epsilon_m B_m f_m(\theta_0)\,[J_m(kr_\ell) - F_m(ka)\,Y_m(kr_\ell)]\sin(m\theta_\ell + \theta_0), \tag{5}$$

where unknown constant B_m has been taken ℓ independent due to symmetry consideration, $Y_m(kr_\ell)$ is the Bessel function of the second kind, and we have written

$$F_m(ka) = J_m(ka)/Y_m(ka). \tag{6}$$

On account of the linearity of Eq. (1), we invoke the principle of superposition to write the total solution as

$$\psi = \psi_0(r,\theta) + \sum_{\ell=1}^{N} \psi_\ell(r_\ell,\theta_\ell). \tag{7}$$

The boundary conditions can be conveniently applied if ψ given by the above equation can be expressed separately in terms of (r,θ) and any one of the set (r_ℓ,θ_ℓ), say (r_1,θ_1). This can be achieved through simple coordinate transformations. But then application of the boundary condition would not lead to appropriate eigenvalue equations for determining k and hence A_m's and B_m's. Therefore, we firstly obtain the latter shift (of origins) using the addition theorems in the form given in Ref. 1. Equation (2) then becomes

$$\psi_0 = \sum_{m=0}^{\infty} \epsilon_m A_m f_m(\theta_0) \sum_{p=-\infty}^{\infty} (-1)^{p+1} J_{m+p}(kd) \times J_p(kr_1)\sin(p\theta_1 - \theta_0). \tag{8}$$

Further, we have in Fig. 2 $\theta_\ell = \beta_\ell - \eta_\ell$, $\delta_\ell = \pi - \eta_\ell - \theta_1$, and $\eta_\ell = \frac{1}{2}(\pi - \phi_\ell)$. Hence for the points[16] where $r_1 < d_{1\ell}$ we employ the addition theorems to get, after manipulation,

$$\begin{aligned} Z_m(kr_\ell)\sin(m\theta_\ell + \theta_0) &= (-1)^m \sum_{p=-\infty}^{\infty} Z_{m+p}(kd_{1\ell}) J_p(kr_1) \\ &\quad \times \sin\{\tfrac{1}{2}(m+p)(\pi + \phi_\ell) - (p\theta_1 - \theta_0)\}, \end{aligned} \tag{9}$$

with $2 \leqslant \ell \leqslant N$ and $Z_m(x)$ representing either $J_m(x)$ or $Y_m(x)$.

We substitute Eqs. (5) and (8) in Eq. (7), and then use Eq. (9). We thus obtain an expression for ψ involving solely the variables r_1 *and* θ_1. The use of the Dirichlet boundary condition, i.e., $\psi = 0$ at $r_1 = a$ for all values of θ_1, then yields, after some rearrangements,

$$\begin{aligned} \sum_{m=0}^{\infty} \epsilon_m f_m(\theta_0) \sum_{p=-\infty}^{\infty} \Big[& A_m(-1)^{p+1} J_{m+p}(kd) \\ & \times \sin(p\theta_1 - \theta_0) + (-1)^m B_m \sum_{\ell=2}^{N} \{J_{m+p}(kd_{1\ell}) \\ & - F_m(ka)\, Y_{m+p}(kd_{1\ell})\}\sin\{\tfrac{1}{2}(m+p)(\pi + \phi_\ell) \\ & - (p\theta_1 - \theta_0)\}\Big] J_p(ka) = 0. \end{aligned} \tag{10}$$

We multiply throughout this equation by $\sin(n\theta_1 - \theta_0)$ and then integrate over θ_1 in the range $-\pi$ to π. We then use the orthogonality relations[17]

$$\int_{-\pi}^{\pi} \sin(p\theta_1 \pm \theta_0)\sin(n\theta_1 \pm \theta_0)\,d\theta_1 = \pi[\delta_{p,n} - \cos 2\theta_0 \delta_{p,-n}], \tag{11}$$

where the same sign is to be taken in the arguments of both the sine functions at a time.

We sum over p with the help of the Kronecker delta function and use the relation

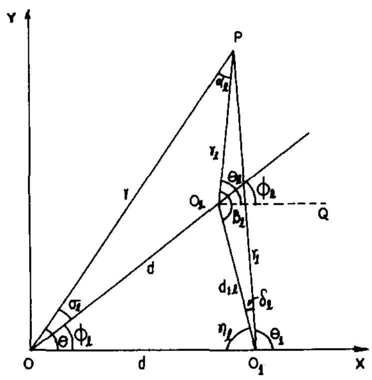


FIG. 2. Polar coordinates of a point P located in the interspatial region of various boundaries.

$$Z_{-m}(x) = (-1)^m Z_m(x), \tag{12}$$

together with the fact that $\cos^2 2\theta_0 = 1$. We finally obtain a set of linear homogeneous equations,

$$\sum_{m=0}^{\infty} \epsilon_m f_m(\theta_0)[A_m P_{mn} + B_m Q_{mn}] = 0, \tag{13}$$

where

$$P_{mn} = (-1)^{n+m}[J_{n-m}(kd) + (-1)^{m+1}\cos 2\theta_0 J_{n+m}(kd)] \tag{14}$$

and

$$Q_{mn} = \sum_{\ell=2}^{N} [U_{n-m}(kd_{1\ell}) + (-1)^{m+1}\cos 2\theta_0 U_{n+m}(kd_{1\ell})] \tag{15}$$

with

$$U_q(x) = [J_q(x) - F_m(ka)Y_q(x)]\cos \tfrac{1}{2}q(\pi + \phi_\ell). \tag{16}$$

We now concentrate on the outer boundary. If it is not circular (but may be polygonal as shown in Fig. 1), the use of the addition theorems at this stage would not give the solution of the problem. Hence in order to develop a generalized approach which could be applicable to noncircular edges as well, we firstly Fourier expand Eq. (5) term-by-term in θ to obtain

$$\psi_\ell = \frac{1}{2}\sum_{m=0}^{\infty} \epsilon_m B_m f_m(\theta_0) \times \sum_{q=0}^{\infty} \epsilon_q f_q(\theta_0) C^\ell_{mq} \sin(q\theta + \theta_0), \tag{17}$$

where the Fourier coefficients given by

$$C^\ell_{mq} = \frac{1}{\pi}\int_{-\pi}^{\pi} [J_m(kr_\ell) - F_m(ka)Y_m(kr_\ell)] \times \sin(m\theta_\ell + \theta_0)\sin(q\theta + \theta_0)d\theta, \tag{18}$$

are functionals of r.

With a view to express C^ℓ_{mq} explicitly in terms of r, we now make use of the addition theorems and shift the origin to the point (r,θ). Since $\theta_\ell = \alpha_\ell + \sigma_\ell + \phi_\ell = \alpha_\ell + \theta$ (Fig. 2), we write

$$Z_m(kr_\ell)\sin(m\theta_\ell + \theta_0) = \sum_{p=-\infty}^{\infty} Z_p(kr)J_{p-m}(kd)\sin\{(p\theta + \theta_0) - (p-m)\phi_\ell\}, \tag{19}$$

for those points which satisfy the inequality[16] $r > d$.

Substituting Eqs. (2) and (17) in Eq. (7), applying $\psi = 0$ at the outer boundary with $r = r_0(\theta)$, multiplying both sides of the resulting equation by $\sin(n\theta + \theta_0)$, and then integrating over θ in the range $-\pi$ to π, we obtain[18]

$$\sum_{m=0}^{\infty} \epsilon_m f_m(\theta_0)[A_m R_{mn} + B_m S_{mn}] = 0 \tag{20}$$

with

$$R_{mn} = \frac{1}{\pi}\int_{-\pi}^{\pi} J_m\{kr_0(\theta)\} \times \sin(m\theta + \theta_0)\sin(n\theta + \theta_0)d\theta \tag{21}$$

and

$$S_{mn} = \frac{1}{2\pi}\sum_{q=0}^{\infty} \epsilon_q f_q(\theta_0)\sum_{\ell=1}^{N} \times \int_{-\pi}^{\pi} C^\ell_{mq}\{r_0(\theta)\}\sin(q\theta + \theta_0)\sin(n\theta + \theta_0)d\theta. \tag{22}$$

The nontriviality of solution of Eqs. (13) and (20) for unknown A_m's and B_m's requires that

$$\det\begin{vmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{vmatrix} = 0, \tag{23}$$

where a quantity having the tilde represents the transpose of the corresponding matrix with the elements given by one of the Eqs. (14), (15), (21), and (22).

We observe that the elements given by Eqs. (14) and (15) are in closed analytical forms, whereas those in Eqs. (21) and (22) are to be evaluated numerically. We have the additional complication that C^ℓ_{mq} appearing in Eq. (22) involves numerical integration through Eq. (18), even after using Eq. (19), as the values of r for the points along the outer boundary are θ dependent. This is as far as we can go in a general way, i.e., for a polygonal outer boundary when the integrations can be broken into parts corresponding separately to each side of the polygon.

Now let us consider the situation when the outer boundary would be circular, as one shown in Fig. 1. Then we will have $r_0 = b$ for all values of θ, and hence using the orthogonality relation along with Eqs. (3) and (4), Eqs. (21) and (22) reduce, respectively, to

$$R_{mn} = J_m(kb)\delta_{mn}(1 + \delta_{m,0}) \tag{24}$$

and

$$S_{mn} = \sum_{\ell=1}^{N} C^\ell_{mn}(r_0 = b). \tag{25}$$

We combine Eqs. (18) and (19), make use of Eq. (11), and then follow the same steps which have been mentioned before Eq. (13) for getting that equation. We then obtain

$$C^\ell_{mn}(r_0 = b) = [J_n(kb) - F_m(ka)Y_n(kb)] \times [J_{n-m}(kd)\cos(n-m)\phi_\ell + (-1)^{m+1} \times \cos 2\theta_0 J_{n+m}(kd)\cos(n+m)\phi_\ell]. \tag{26}$$

Combining Eqs. (25) and (26), and then using the identity given by Eq. (A4), we finally obtain

$$S_{mn} = N[J_n(kb) - F_m(ka)Y_n(kb)] \times [W_{n-m}(kd) + (-1)^{m+1}\cos 2\theta_0 W_{n+m}(kd)], \tag{27}$$

where

$$W_q(kd) = J_q(kd)\sum_{p=-\infty}^{\infty} \delta_{q,pN}. \tag{28}$$

Each of m and n in the above equations takes integral values from $\cos\theta_0$ to ∞, the minimum value being dictated by the presence of the function $f_m(\theta_0)$ in Eqs. (13) and (20). We thus find that our basic solution [Eq. (23)] is in the form of a doubly infinite determinantal equation. The zeros of this determinant would determine the eigenvalues, called the cutoff wavenumbers, of a given physical problem. The values of A_m's and B_m's and hence the eigenfunctions (i.e., modal functions) can then be evaluated using Eq. (7) together with Eqs. (2) and (5). The series in these latter two equations are to be truncated for computational purposes, giving finite determinant in Eq. (23).

III. NUMERICAL RESULTS AND DISCUSSIONS

We consider the symmetric mode with $N=2$ corresponding to the configuration when a pair of identical circles would be located symmetrically within a circular region. Thus $\phi_1 = 0$ and $\phi_2 = \pi$, and hence $d_{12} = 2d$. We also have

$$W_{n\pm m}(kd) = J_{n\pm m}(kd)[1 + (-1)^{n+m}]. \tag{29}$$

The expressions in Eqs. (14) and (24) remain as such but the expressions (15) and 27) now reduce to

$$Q_{mn} = (-1)^{n+m}[G_{mn}(2kd) - F_m(ka)\{Y_{n-m}(2kd) + (-1)^{m+1}\cos 2\theta_0 Y_{n+m}(2kd)\}], \tag{30}$$

$$S_{mn} = 2[J_n(kb) - F_m(ka)Y_{n+m}(kb)] \times [1 + (-1)^{n+m}]G_{mn}(kd), \tag{31}$$

where

$$G_{mn}(x) = J_{n-m}(x) + (-1)^{m+1}\cos 2\theta_0 J_{n+m}(x). \tag{32}$$

The lowest zero k_0 of Eq. (23) has been calculated as a function of eccentricity d using Eqs. (14), (24), (30), and (31) for the radii ratio $\eta = a/b = 0.01$. Results normalized by taking the radius b of the outer circle equal unity are being depicted as the upper curve in Fig. 3. We also plot there the $k_0 - d$ curve for $N=1$ for the same value of η using the expressions taken from Ref. 1.

We observe that for $N=2$ the value of k_0 firstly increases, shows a maximum, and then decreases continuously as d increases. This behavior is in sharp contrast with that of the lower curve; in the latter case k_0 has its maximum value at $d=0$ (concentric annular situation) and shows continuous decrease with increase in d. The peak in the former case is the effect of the spatial configurational interference in the solutions contributed by the diametrically opposite points and supports the qualitative nature of the approximate studies done by Lamarsh[10] in the thermal nuclear reactors.

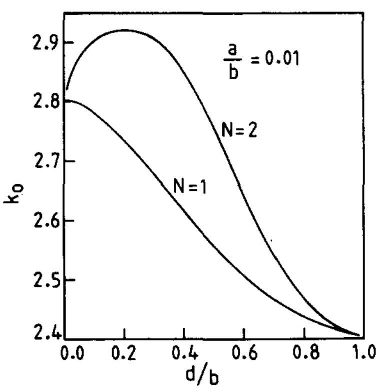


FIG. 3. The lowest cutoff value k_0 of the symmetric mode as a function of eccentricity d.

Obviously we cannot put $d=0$ directly in our expressions for checking the consistency of result in the limiting case. But we find that with $N=2$ we have $k_0 = 2.823$ and 2.406 for $d = 0.01$ and 0.99, respectively; the curve having a decreasing trend towards both ends. Hence it seems tempting to compare these values with the results for concentric annulus ($\eta = 0.01$) and simply connected circular domain. The latter set of values are 2.8009 and 2.4048, which correspond, respectively, to the lowest zeros of the functions

$$[J_0(kb)Y_0(ka) - J_0(ka)Y_0(kb)]$$

and $J_0(kb)$. Thus our results are tending towards the correct limiting behavior. We have further observed in the course of our calculation that for $N=2$ the lowest cutoff value converges at the third decimal place for 5×5 input matrices, i.e., for a 10×10 determinant in Eq. (23).

The restrictions imposed on the value of d for the general validity of Eqs. (9) and (19) require in our final analysis that $a/2 \leqslant d \leqslant b$. In our calculation we have considered the range $a \leqslant d \leqslant (b-a)$; the minimum and the maximum of the range correspond to the situations when the edges of the two inner circles would touch each other and the outer boundary, respectively. This is the range which covers most of the physical problems of interest and hence the restrictions under discussion do not undermine the applicability of our approach.

IV. CONCLUDING REMARKS

We have obtained the solution of the two-dimentional Helmholtz equation in a multiply-connected region, satisfying the Dirichlet boundary condition at each of the edges exactly. The elements of the singular matrix, arising in the course of our solution, have been obtained in closed analytical form for the case when the outer boundary is circular. The convergence of the solution, while computing zeros, has been found to be quite fast. The method therefore offers a great potentiality for application to boundary-value problems in many branches of applied sciences which we have come across in the Introduction. Some of the important problems corresponding to the case $N=2$ are, e.g., TE and TM modes in shielded wire-pairs (bifilar lines),[19] characteristics of two-core optical fibers,[20] so-called "shadowing" and "antishadowing" arising due to superposed effectiveness of a pair of control rods in thermal nuclear reactors,[10] etc.

It may also be noted that our solution is appropriate to a class of problems of even more general nature than discussed in this paper. We may take the outer boundary, consisting of M straight edges or arcs, with M/N an integer, such that the considered rotational symmetry is still preserved. Moreover, we conclude that the kind of results plotted in Fig. 3 for $N=2$ should be observable in all boundary-value problems coming within the purview of the Helmholtz equation.

ACKNOWLEDGMENTS

The authors are indebted to Professor Kosuke Nagaya and Professor F. E. Gardiol for illuminating discussions. The invaluable suggestions from Professor E. S. Raja Gopal and Professor S. K. Joshi are also gratefully acknowledged.

APPENDIX

Let us consider the sum

$$S=\sum_{\ell=1}^{N}\exp(in\phi_\ell), \quad (A1)$$

where $\phi_\ell=2\pi(\ell-1)/N$.

This geometric progression can easily be summed up to give

$$S=(1-e^{i2\pi n})/(1-e^{i2\pi n/N}). \quad (A2)$$

The numerator of this expression is zero for each integral value of n, whereas the denominator vanishes only if n/N is an integer. In the latter case, the ratio in Eq. (A2) becomes indeterminate and has the value N. Hence we can write from Eq. (A1),

$$\sum_{\ell=1}^{N}\sin(n\phi_\ell)=0 \quad (A3)$$

and

$$\sum_{\ell=1}^{N}\cos(n\phi_\ell)=N\sum_{p=-\infty}^{\infty}\delta_{n,pN}. \quad (A4)$$

[1]G. S. Singh and L. S. Kothari, J. Math. Phys. **25**, 810 (1984).

[2]G. S. Singh and A. Vishen, to be submitted to J. Math. Phys.

[3]G. S. Singh, A. Vishen, G. P. Srivastava, and L. S. Kothari, J. Sound Vib. **100**, 141 (1985).

[4]A. Vishen, G. P. Srivastava, G. S. Singh, and F. Gardiol, IEEE Trans. Microwave Theory Tech. **MTT-34**, 292 (1986); J. G. Fikioris, J. A. Roumeliotis, M. Davidovitz, A. Vishen, G. S. Singh, and F. E. Gardiol, *ibid.* **MTT-35**, 469 (1987).

[5]G. S. Singh, L. S. Kothari, and S. Garg, Ann. Nucl. Energy **12**, 391 (1985).

[6]G. S. Singh, A. Kapoor, and A. K. Ghatak (unpublished).

[7]W. H. Lin, J. Sound Vib. **79**, 463 (1981).

[8]R. L. Murray, in *Reactor Operational Problems*, edited by D. J. Hughes, S. McLain, and C. Williams (Pergamon, New York, 1957), pp. 264–268.

[9]W. H. Lin, J. Sound Vib. **81**, 425 (1982).

[10]J. R. Lamarsh, *Introduction to Nuclear Reactor Theory* (Addison–Wesley, Reading, MA, 1966), Chap. 14.

[11]H. Saito and K. Nagaya, Bull. JSME **16**, 1506 (1973).

[12]E. Yamashita, S. Ozeki and K. Atsuki, J. Lightwave Tech. **LT-3**, 341 (1985).

[13]P. R. McIsaac, IEEE Trans. Microwave Theory Tech. **MTT-23**, 421 (1975).

[14]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1958), Chap. XI.

[15]K. Nagaya, J. Sound Vib. **50**, 545 (1977).

[16]The restrictions are necessary for $Y_m(x)$, but not for $J_m(x)$.

[17]In Eq. (10), p runs through all positive and negative integral values, including zero. Hence the orthogonality relations must be taken in the form as given by Eq. (11). Serious errors had been committed earlier by some workers, leading to incorrect expressions as found in Refs. 7–9.

[18]The factor $1/\pi$ has been included in each of Eqs. (21) and (22) for later simplification.

[19]This has been pointed out to us by Professor F. E. Gardiol.

[20]K. Kitayama, N. Shibata and M. Ohashi, J. Opt. Soc. Am. A **2**, 84 (1985); K. Kitayama and Y. Ishida, *ibid.* **2**, 90 (1985).

Integrable dynamical systems with hierarchy. I. Formulation

Susumu Okubo
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 30 September 1988; accepted for publication 23 November 1988)

It is shown that any model with the zero (generalized) Nijenhuis tensor, under some additional assumptions, will automatically satisfy the hierarchy equation with an infinite number of conserved quantities in involution. Especially, if there is a dual symplectic structure with zero Nijenhuis tensor, then there exist infinite numbers of Poisson brackets and of Lagrangians giving the same equation of motion. Toda lattice is one such example.

I. INTRODUCTION

There are many classical dynamical systems that are completely integrable.[1–5] These are systems with a finite as well as infinite number of degrees of freedom, and these have been extensively studied by many authors. The integrability of such systems depends upon the existence of a sufficient number of conserved quantities that are in involution. One interesting subclass of such integrable models satisfies an additional property of having the so-called hierarchy equations among these conserved quantities with dual Poisson bracket structures; examples are Toda lattice and KdV equations. It has been known for some time[6–14] that identical vanishing of the Nijenhuis tensor associated with the dual symplectic structure leads to the hierarchy equation and the involution property of the conserved quantities. This fact has been rediscovered in our recent papers[15,16] (referred to as I and II, hereafter) but with additional information that the Toda lattice satisfies these conditions. The Toda lattice solution is so far the only nontrivial example of the dynamical system of finite degree that satisfies these conditions, although many infinite systems such as the KdV equation are known to obey them. Other finite integrable systems with such properties will be given in a subsequent paper.

The purpose of this paper is to generalize and study the problem in more detail. First, in Sec. II, we will introduce a notion of the generalized Nijenhuis tensor and show that conditions stated above can be, in reality, considerably relaxed. To obtain the hierarchy equation and the involution, it is really sufficient to assume a weaker ansatz of zero generalized Nijenhuis tensor (rather than vanishing Nijenhuis tensor itself) together with only one symplectic (instead of two) structure in theory. However, the stronger validity of the zero Nijenhuis tensor with dual symplectic structure is found to give physically (and perhaps mathematically) more interesting additional properties. For instance, the theory would then contain an infinite number of Poisson brackets as well as an infinite number of associated Lagrangians which describe, nevertheless, the same equation of motion. These facts will be shown in Sec. IV after some preliminary preparations in Sec. III, where some new identities involving Nijenhuis tensors are discussed. Section V is devoted to a study of the eigenvalue problem of a mixed tensor underlying our theory with the same weaker assumptions as in Sec. II, generalizing some earlier works. Finally, we will give a brief comment in Sec. VI on a possible connection between the present theory and the so-called P-matrix formalism of another important subclass of integrable models.

Although almost all of the results to be presented in this paper would be valid (at least formally) for any system with infinite number of freedoms by replacing summations by integrations, we restrict our discussion in this paper to cases of finite systems. Treatment of infinite systems requires subtle mathematical care for the convergence problem involving integrals. Also, some explicit solutions satisfying conditions of this paper will be presented in a subsequent paper.

II. GENERALIZED NIJENHUIS TENSOR AND HIERARCHY EQUATION

Let M be a differential manifold with finite dimension D. We assume that in any local coordinate frame with coordinate x^μ $(\mu = 1,2,...,D)$, there exists a mixed $1-1$ tensor $S_\mu^{\ \nu}$ $(\mu,\nu = 1,2...,D)$. We construct the Nijenhuis tensor out of $S_\mu^{\ \nu}$ by

$$N_{\mu\nu}^{\ \ \lambda} = -N_{\nu\mu}^{\ \ \lambda} = S_\mu^{\ \alpha}\,\partial_\alpha\, S_\nu^{\ \lambda} - S_\nu^{\ \alpha}\,\partial_\alpha\, S_\mu^{\ \lambda} - S_\alpha^{\ \lambda}\,(\partial_\mu\, S_\nu^{\ \alpha} - \partial_\nu\, S_\mu^{\ \alpha}). \tag{2.1}$$

Here, the repeated greek indices will be understood to imply automatical summations on D values $1,2,...,D$. The coordinate-free definition of the Nijenhuis tensor will be given in Sec. III. A simple way of verifying the fact that $N_{\mu\nu}^{\ \ \lambda}$ is a genuine tensor under coordinate transformations is to introduce[17] an affine connection $\Gamma_{\mu\nu}^{\lambda}$ in M with the covariant derivative

$$S_{\mu;\lambda}^{\ \nu} = \partial_\lambda\, S_\mu^{\ \nu} - \Gamma_{\lambda\mu}^{\alpha}\, S_\alpha^{\ \nu} + \Gamma_{\lambda\alpha}^{\nu}\, S_\mu^{\ \alpha}\,. \tag{2.2}$$

We now define the generalized Nijenhuis tensor $\widetilde{N}_{\mu\nu}^{\ \ \lambda}$ by

$$\widetilde{N}_{\mu\nu}^{\ \ \lambda} = -\widetilde{N}_{\nu\mu}^{\ \ \lambda} = S_\mu^{\ \alpha}\, S_{\nu;\alpha}^{\ \lambda} - S_\nu^{\ \alpha}\, S_{\mu;\alpha}^{\ \lambda} - S_\alpha^{\ \lambda}\,(S_{\nu;\mu}^{\ \alpha} - S_{\mu;\nu}^{\ \alpha}), \tag{2.3}$$

which clearly defines a genuine tensor under coordinate transformations. Moreover, from Eqs. (2.2) and (2.3), we find

$$\widetilde{N}_{\mu\nu}^{\ \ \lambda} = N_{\mu\nu}^{\ \ \lambda} + T_{\mu\nu}^{\beta}\, S_\alpha^{\ \lambda}\, S_\beta^{\ \alpha} + T_{\alpha\beta}^{\lambda}\, S_\mu^{\ \alpha}\, S_\nu^{\ \beta} - T_{\mu\alpha}^{\beta}\, S_\nu^{\ \alpha}\, S_\beta^{\ \lambda} - T_{\alpha\nu}^{\beta}\, S_\mu^{\ \alpha}\, S_\beta^{\ \lambda}, \tag{2.4}$$

where $T_{\mu\nu}^{\lambda}$ is the torsion tensor

$$T_{\mu\nu}^{\lambda} = -T_{\nu\mu}^{\lambda} = \Gamma_{\mu\nu}^{\lambda} - \Gamma_{\nu\mu}^{\lambda} \tag{2.5}$$

of the affine connection. Since the torsion tensor $T_{\mu\nu}^{\lambda}$ be-

 0022-2488/89/040834-10$02.50

haves covariantly[18] under coordinate transformations, Eq. (2.4) implies the same for $N^{\lambda}_{\mu\nu}$.

Before going into further discussion, we will first define powers of S^{ν}_{μ} inductively by

$$(S^0)^{\nu}_{\mu} = \delta^{\nu}_{\mu} \quad (n=0), \qquad (s^{n+1})^{\nu}_{\mu} = S^{\lambda}_{\mu}\,(S^n)^{\nu}_{\lambda} \quad (n\geqslant 0). \tag{2.6}$$

Moreover, we note that both $N^{\lambda}_{\mu\nu}$ and $\widetilde{N}^{\lambda}_{\mu\nu}$ are invariant under replacement

$$S^{\nu}_{\mu} \to (S')^{\nu}_{\mu} = S^{\nu}_{\mu} + C\delta^{\nu}_{\mu}$$

for any constant C. Therefore, we assume the existence of the inverse tensor $(S^{-1})^{\nu}_{\mu}$ satisfying

$$(S^{-1})^{\lambda}_{\mu}\, S^{\nu}_{\lambda} = \delta^{\nu}_{\mu} \tag{2.7}$$

by considering $(S')^{\nu}_{\mu}$ instead of S^{ν}_{μ} if necessary. Now, we define K_n for any integer n by

$$K_n = (1/2n)\operatorname{Tr} S^n = (1/2n)\,(S^n)^{\mu}_{\mu} \quad (n\neq 0), \tag{2.8a}$$

$$K_0 = \tfrac{1}{2}\log(\det S) = \tfrac{1}{2}\operatorname{Tr}(\log S) \quad (n=0). \tag{2.8b}$$

Note the presence of a factor of $\frac{1}{2}$ in Eqs. (2.8) in comparison to I and II. The reason we define K_0 as in Eq. (2.8b) will become clear soon. From Eqs. (2.2) and (2.4), we calculate

$$\tfrac{1}{2}(S^{n-1})^{\nu}_{\lambda}\, \widetilde{N}^{\lambda}_{\mu\nu} = \tfrac{1}{2}\,(S^{n-1})^{\nu}_{\lambda}\, N^{\lambda}_{\mu\nu} = S^{\lambda}_{\mu}\,\partial_{\lambda}K_n - \partial_{\mu}K_{n+1} \tag{2.9}$$

for any integer n, provided that we define K_n as in Eq. (2.8). Therefore, if we have $\widetilde{N}^{\lambda}_{\mu\nu} = 0$ or $N^{\lambda}_{\mu\nu} = 0$, we have the desired hierarchy equation

$$S^{\lambda}_{\mu}\,\partial_{\lambda}K_n = \partial_{\mu}K_{n+1} \tag{2.10}$$

for any integer n. Repeating this, we obtain

$$(S^m)^{\lambda}_{\mu}\,\partial_{\lambda}K_n = \partial_{\mu}K_{n+m} \tag{2.11}$$

further for any two integers n and m. A more careful inspection shows the following generalization. Define a tensor $M^{\lambda}_{\mu\nu}$ by

$$M^{\lambda}_{\mu\nu} = S^{\alpha}_{\mu}\, S^{\lambda}_{\nu;\alpha} - S^{\lambda}_{\alpha}\, S^{\alpha}_{\nu;\mu} + A\,(S^{\alpha}_{\nu}\, S^{\lambda}_{\mu;\alpha} - S^{\lambda}_{\alpha}\, S^{\alpha}_{\mu;\nu}) \tag{2.12}$$

for any function $A = A(x)$. We also find

$$\begin{aligned}(S^{n-1})^{\nu}_{\lambda}\, M^{\lambda}_{\mu\nu} &= (S^{n-1})^{\nu}_{\lambda}\, N^{\lambda}_{\mu\nu} \\ &= 2\{S^{\lambda}_{\mu}\,\partial_{\lambda}K_n - \partial_{\mu}K_{n+1}\},\end{aligned}$$

which is independent of the function $A(x)$. Therefore, we will have the same hierarchy equation (2.10) whenever we have $M^{\lambda}_{\mu\nu} = 0$. However, if we interchange the role of μ and ν, the condition is consistent only if we have either $A = -1$ or $A = 1$. Note that $M^{\lambda}_{\mu\nu}$ is the same as $\widetilde{N}^{\lambda}_{\mu\nu}$ for $A = -1$, while we find a symmetric tensor

$$\begin{aligned}\widetilde{M}^{\lambda}_{\mu\nu} = \widetilde{M}^{\lambda}_{\nu\mu} &= S^{\alpha}_{\mu}\, S^{\lambda}_{\nu;\alpha} + S^{\alpha}_{\nu}\, S^{\lambda}_{\mu;\alpha} \\ &\quad - S^{\lambda}_{\alpha}\,(S^{\alpha}_{\nu;\mu} + S^{\alpha}_{\mu;\nu}),\end{aligned} \tag{2.13}$$

for $A = 1$ in contrast to the antisymmetric tensor $\widetilde{N}^{\lambda}_{\mu\nu} = -\widetilde{N}^{\lambda}_{\nu\mu}$.

Summarizing, we have proved the following proposition.

Proposition 1: Suppose that we have either $N^{\lambda}_{\mu\nu} = 0$ or $\widetilde{N}^{\lambda}_{\mu\nu} = 0$ or $\widetilde{M}^{\lambda}_{\mu\nu} = 0$. Then, K_n's constructed in Eq. (2.8) satisfy the hierachy equations (2.10) and (2.11) for any two integers n and m.

Before proceeding further, the following remark is perhaps in order.

Remark 1: The proposition has already been noted in the previous papers (I and II) as well as Ref. 14 for the case of $N^{\lambda}_{\mu\nu} = 0$. The requirement $\widetilde{N}^{\lambda}_{\mu\nu} = 0$ is weaker, however, since it demands only

$$\begin{aligned}N^{\lambda}_{\mu\nu} &= T^{\beta}_{\mu\alpha}\, S^{\alpha}_{\nu}\, S^{\lambda}_{\beta} + T^{\beta}_{\alpha\nu}\, S^{\alpha}_{\mu}\, S^{\lambda}_{\beta} \\ &\quad - T^{\beta}_{\mu\nu}\, S^{\lambda}_{\alpha}\, S^{\alpha}_{\beta} - T^{\lambda}_{\alpha\beta}\, S^{\alpha}_{\mu}\, S^{\beta}_{\nu}\end{aligned} \tag{2.14}$$

for any tensor $T^{\lambda}_{\mu\nu}$ satisfying $T^{\lambda}_{\mu\nu} = -T^{\lambda}_{\nu\mu}$. Also, we note that we have $N^{\lambda}_{\mu\nu} = 0$ identically if $S^{\nu}_{\mu} = f(x)\,\delta^{\nu}_{\mu}$ for arbitrary function $f(x)$. However, this case is rather trivial since K_n $(n\neq 0)$ is a constant multiple of $(K_1)^n$ and $K_0 = \frac{1}{2} D\log K_1 + \text{const}$.

An interesting case for physics occurs when the dimension of M is even with

$$D = 2N, \tag{2.15}$$

and M is moreover a symplectic manifold.[19–21] Let

$$f = \tfrac{1}{2} f_{\mu\nu}(x)\,dx^{\mu}\wedge dx^{\nu}, \tag{2.16a}$$

$$f_{\mu\nu} = -f_{\nu\mu}, \tag{2.16b}$$

be the symplectic form with the inverse $f^{\mu\nu} = -f^{\nu\mu}$, so that

$$f_{\mu\lambda}\, f^{\lambda\nu} = \delta^{\nu}_{\mu}. \tag{2.17}$$

Then, the condition $df = 0$ is equivalent to

$$\partial_{\lambda} f_{\mu\nu} + \partial_{\mu} f_{\nu\lambda} + \partial_{\nu} f_{\lambda\mu} = 0, \tag{2.18}$$

which, in turn, is equivalent to [see Eq. (4.15)]

$$f^{\lambda\alpha}\,\partial_{\alpha} f^{\mu\nu} + f^{\mu\alpha}\,\partial_{\alpha} f^{\nu\lambda} + f^{\nu\alpha}\,\partial_{\alpha} f^{\lambda\mu} = 0. \tag{2.19}$$

Let $h = h(x)$ and $g = g(x)$ be two functions in M and define the Poisson bracket by

$$\{h,g\}_0 = f^{\mu\nu}\,\partial_{\mu}h\,\partial_{\nu}g. \tag{2.20}$$

It is known[22,23] that the conditions $f^{\mu\nu} = -f^{\nu\mu}$ and (2.19) are equivalent to the validity of Jacobi identity

$$\{h,g\}_0 = -\{g,h\}_0, \tag{2.21a}$$

$$\{\{h,g\}_0,k\}_0 + \{\{g,k\}_0,h\}_0 + \{\{k,h\}_0,g\}_0 = 0 \tag{2.21b}$$

among three functions $h(x)$, $g(x)$, and $k(x)$.

Now, suppose that there exists another antisymmetric tensor $F_{\mu\nu}$, i.e.,

$$F_{\mu\nu} = -F_{\nu\mu}, \tag{2.22}$$

such that S^{ν}_{μ} has a form of

$$S^{\nu}_{\mu} = F_{\mu\lambda}\, f^{\lambda\nu}. \tag{2.23}$$

Then, it is easy to verify the validity of

$$(S^n)^{\nu}_{\lambda}\, f^{\lambda\mu} = -(S^n)^{\mu}_{\lambda}\, f^{\lambda\nu} = \text{antisymmetric in } \mu\leftrightarrow\nu \tag{2.24}$$

for any integer n. In that case, Eqs. (2.24) and (2.11) give

$$f^{\mu\lambda}\,\partial_{\lambda}K_{n+m} = (S^m)^{\mu}_{\lambda}\, f^{\lambda\nu}\,\partial_{\nu}K_n,$$

so that we calculate

$$\{K_{n+l},K_{m-l}\}_0 = f^{\mu\nu}\,\partial_\mu K_{n+l}\,\partial_\nu K_{m-l}$$
$$= -(S^l)^\nu_\lambda f^{\lambda\mu}\,\partial_\mu K_n\,\partial_\nu K_{m-l}$$
$$= f^{\mu\lambda}\,\partial_\mu K_n\,(S^l)^\nu_\lambda\,\partial_\nu K_{m-l}$$
$$= f^{\mu\lambda}\,\partial_\mu K_n\,\partial_\lambda K_m$$
$$= \{K_n,K_m\}_0,$$

for any integer l, n, and m. Especially, choosing $l = m - n$, this gives

$$\{K_m,K_n\}_0 = \{K_n,K_m\}_0 = -\{K_m,K_n\}_0 = 0.$$

Therefore, we have the following proposition as in (I) and (II).

Proposition 2: Let S^ν_μ have the form of Eq. (2.23) with the condition Eq. (2.22), satisfying the hierarchy equation (2.10). Then, for any two integer n and m, we have $\{K_n, K_m\}_0 = 0$. In other words, K_n's are in involution with respect to the Poisson bracket defined by Eq. (2.20).

Remark 2: If we have assumed the symmetric condition

$$F_{\mu\nu} = F_{\nu\mu} \tag{2.22'}$$

instead of the antisymmetric condition Eq. (2.22), then we can prove similarly $\{K_n, K_m\}_0 = 0$, provided that the difference $n - m$ is an even integer. Therefore for such a case, all K_n's with even n only or with odd n are separately in involution with each other. However, in Sec. V we will show that all K_n with odd n are identically zero. Then, the hierarchy equation (2.10) requires that $\partial_\lambda K_n = 0$ also for even integer n. Therefore, this case is rather uninteresting.

The next question we must ask is whether we could find a Lagrangian and Hamiltonian H such that K_n are also conserved quantities of the Hamiltonian dynamics. The answer turns out to always be yes, if we use the linearized Hamilton's formulation of the Lagrangian mechanics. Since the two-form f is closed, there exists a function $\theta_\mu(x)$ satisfying

$$f_{\mu\nu} = \partial_\mu\theta_\nu - \partial_\nu\theta_\mu \tag{2.25}$$

at least locally by Poincare's lemma. Moreover, for any fixed integer value p, we identify

$$H = K_p, \tag{2.26}$$

and consider Lagrangian

$$L = \theta_\mu(x)\dot{x}^\mu - H(x), \tag{2.27}$$

which is linear in the time derivative $\dot{x}^\mu$. Now, the Euler–Lagrange equation of motion based upon this Lagrangian gives the Hamilton's equation of motion

$$f_{\mu\nu}\dot{x}^\nu = \partial_\mu H = \partial_\mu K_p. \tag{2.28}$$

Then, any function $g = g(x)$ of x^μ satisfies

$$\dot{g} = \{g,H\}_0 = \{g,K_p\}_0. \tag{2.29}$$

Especially, choosing $g = K_n$, we find

$$\dot{K}_n = \{K_n,K_p\}_0 = 0,$$

so that all K_n are constants of motion of the Hamiltonian and in involution with each other. Therefore, if we can find N algebraically independent conserved quantities among K_n's, then the system is completely integrable by Liouville's theorem.[20,21] In summary, we conclude that, given tensor S^ν_μ of form Eqs. (2.23) with Eq. (2.22) satisfying either $N^\lambda_{\mu\nu} = 0$, or $\tilde{N}^\lambda_{\mu\nu} = 0$ or $\tilde{M}^\lambda_{\mu\nu} = 0$, there exists a Hamiltonian dynamical system with an infinite number of conserved quantities K_n in involution, satisfying the hierarchy equation.

Remark 3: In the previous paper (II), we have shown that the Toda lattice corresponds to the choice of $p = 2$ and $N^\lambda_{\mu\nu} = 0$. Also, in Sec. V, we will show that the maximal number of algebraically independent components among K_n's is at most N, as it should be.

According to the celebrated Darboux theorem,[19,20] there exists a coordinte system in which the symplectic tensor $f_{\mu\nu}$ is constant. We call it the canonical coordinate frame since Eq. (2.28) reproduces then the standard Hamilton's equation of motion.

If we assume moreover as in Sec. IV that the antisymmetric tensor $F_{\mu\nu}$ satisfies an additional symplectic condition

$$\Delta_{\lambda\mu\nu}(F) = \partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} = 0, \tag{2.30}$$

we can say more. When we note the identity,

$$(S^{n-1})^\lambda_\alpha f^{\alpha\nu}\Delta_{\mu\nu\lambda}(F)$$
$$= 2\partial_\mu K_n - 2\partial_\lambda\,(S^n)^\lambda_\mu + 2S^\lambda_\mu\,\partial_\nu\,(S^{n-1})^\nu_\lambda$$
$$+ (S^n)^\alpha_\lambda f_{\alpha\beta}\,\partial_\mu f^{\lambda\beta} - 2(S^{n-1})^\alpha_\lambda\,F_{\mu\beta}\,\partial_\alpha f^{\lambda\beta},$$

then the condition Eq. (2.30) implies the validity of

$$\partial_\mu K_n = \partial_\lambda\,(S^n)^\lambda_\mu - S^\lambda_\mu\,\partial_\nu\,(S^{n-1})^\nu_\lambda \tag{2.31}$$

for any integer n in the canonical coordinate frame in which $f_{\mu\nu}$ and hence $f^{\mu\nu}$ are constants. From Eq. (2.31), we can conclude

$$\partial_\lambda\,(S^n)^\lambda_\mu = n\,\partial_\mu K_n, \tag{2.32}$$

as follows. For $n = 0$, Eq. (2.32) is trivially satisfied since $(S^0)^\lambda_\mu = \delta^\lambda_\mu$. Similarly, setting $n = 1$ in Eq. (2.31), we see the validity of Eq. (2.32) for $n = 1$. Now by induction together with the hierarchy equation (2.10), we can prove the validity of Eq. (2.32) from Eq. (2.31). We note that Eq. (2.32) is rewritten as a conservation law

$$\partial_\lambda\{(S^n)^\lambda_\mu - \tfrac{1}{2}\delta^\lambda_\mu\,\mathrm{Tr}\,S^n\} = 0. \tag{2.33}$$

Furthermore, set

$$(V_{n,m})^\nu_\mu = \{nmK_nK_m + (n+m)K_{n+m}\}\,\delta^\nu_\mu$$
$$- nK_n\,(S^m)^\nu_\mu - mK_m\,(S^n)^\nu_\mu, \tag{2.34}$$

for any integer n and m. Then, it is easy to see that it satisfies also the conservation law

$$\partial_\lambda\,(V_{n,m})^\lambda_\mu = 0 \tag{2.35}$$

in the canonical coordinate frame. However, the physical or geometrical significance of these identities is at the moment obscure.

In ending this section, we would like to make the following comments. First of all, we note that if $S^\nu_{\mu;\lambda} = 0$ identically, this leads to a trivial result that all K_n's are constants. This is due to the validity of the following identity:

$$(S^{n-1})^\mu_\nu\,S^\nu_{\mu;\lambda} = 2\,\partial_\lambda K_n.$$

Therefore, unless we are interested in trivial K_n's, we cannot impose the constraint $S^\nu_{\mu;\lambda} = 0$. Second, the hierarchy equation is essentially unique in the following sense. Suppose that we have a modified hierarchy equation

$$S_\mu^\lambda\, \partial_\lambda K_n = B\, \partial_\mu K_{n+1} \tag{2.36}$$

for some function $B = B(x)$. In Sec. V, we will present a plausibility argument that Eq. (2.36) will lead in general to physically uninteresting cases unless we have $B(x) = 1$ identically. Then, we cannot have the validity of

$$S_\mu^\alpha\, (S_\nu^\lambda)_{;\alpha} = BS_\alpha^\lambda\, (S_\nu^\alpha)_{;\mu} \tag{2.37}$$

for $B \neq 1$ since this will give Eq. (2.36) automatically when we multiply $(S^{n-1})_\lambda^\gamma$ to both sides. The exception is the case of $B = 1$ with

$$S_\mu^\alpha\, (S_\nu^\lambda)_{;\alpha} = S_\alpha^\lambda\, (S_\nu^\alpha)_{;\mu}. \tag{2.38}$$

This equation may be of some interest since it gives $\widetilde{N}_{\mu\nu}^\lambda = \widetilde{M}_{\mu\nu}^\lambda = 0$ identically. Also, if we have

$$S_\mu^\alpha\, (S_\nu^\lambda)_{;\alpha} = \pm S_\alpha^\lambda\, (S_\mu^\alpha)_{;\nu} \tag{2.39}$$

instead of Eq. (2.38), we will obtain $\widetilde{M}_{\mu\nu}^\lambda = 0$ for the plus sign and $\widetilde{N}_{\mu\nu}^\lambda = 0$ for the negative sign. These possibilities will be studied elsewhere.

III. NIJENHUIS TENSOR

In the previous section, we have seen that a weaker condition $\widetilde{N}_{\mu\nu}^\lambda = 0$ will be sufficient for the underlying dynamical system to be a candidate for complete integrability. However, the stronger condition $N_{\mu\nu}^\lambda = 0$ is physically more interesting, as we will see in Sec. IV. Because of this, we will study some properties of the Nijenhuis tensor. It is often more convenient to use the coordinate-free formulation[24,25] as follows. Let T_M be the tangent space of the manifold M. Let

$$\mathrm{gl}(T_M)\colon T_M \to T_M \tag{3.1}$$

be a set consisting of all pointwise general linear transformations of T_M. In other words, $S \in \mathrm{gl}(T_M)$ satisfies

$$S(gX + hY) = gSX + hSY \tag{3.2}$$

for any tangent vectors $X, Y \in T_M$ and for any two functions $g = g(x)$ and $h = h(x)$ at any point $x \in M$. For a given coordinate system, the components S_μ^ν of S is defined by

$$S\partial_\mu = S_\mu^\nu\, (x) \partial_\nu \tag{3.3}$$

for holonomic basis $X_\mu = \partial_\mu$.

For S, $T \in \mathrm{gl}(T_M)$, we define

$$N = N(S,T)\colon T_M \times T_M \to T_M \tag{3.4}$$

by

$$\begin{aligned} 2N(S,T\,|X,Y) = {} & (ST + TS)\,[X,Y] \\ & + [SX,TY] + [TX,SY] \\ & - S\,[TX,Y] - S\,[X,TY] \\ & - T\,[SX,Y] - T\,[X,SY], \end{aligned} \tag{3.5}$$

for any $X, Y \in T_M$. Clearly, $N(S,T\,|X,Y)$ is symmetric in S and T but antisymmetric in X and Y,

$$N(S,T\,|X,Y) = N(T,S\,|X,Y) = -N(S,T\,|Y,X). \tag{3.6}$$

Moreover, it can be shown to be g linear in the sense that it satisfies

$$N(S,T\,|gX,Y) = N(S,T\,|X,gY) = gN(S,T\,|X,Y) \tag{3.7}$$

for any function $g = g(x)$. The components of $N(S,T)$ is defined now by

$$N(S,T\,|\partial_\mu,\partial_\nu) = N_{\mu\nu}^\lambda\, (S,T)\partial_\lambda, \tag{3.8}$$

which gives

$$\begin{aligned} 2N_{\mu\nu}^\lambda\, (S,T) = {} & S_\mu^\alpha\, \partial_\alpha\, T_\nu^\lambda - S_\nu^\alpha\, \partial_\alpha\, T_\mu^\lambda \\ & - S_\alpha^\lambda\, (\partial_\mu\, T_\nu^\alpha - \partial_\nu\, T_\mu^\alpha) \\ & + T_\mu^\alpha\, \partial_\alpha\, S_\nu^\lambda - T_\nu^\alpha\, \partial_\alpha\, S_\mu^\lambda \\ & - T_\alpha^\lambda\, (\partial_\mu\, S_\nu^\alpha - \partial_\nu\, S_\mu^\alpha). \end{aligned} \tag{3.9}$$

Especially, if we set $T = S$, the expression for $N_{\mu\nu}^\lambda\, (S,S)$ reproduces that of Eq. (2.1) of Sec. II. Further, the property given in Eq. (3.7) guarantees the covariant tensor character of $N_{\mu\nu}^\lambda\, (S,T)$ under the general coordinate transformation.

Now, as in Sec. II, we assume hereafter that the inverse S^{-1} of S exists, by using $S' = S + CI$ instead of S for a suitable constant C if necessary. Here, $I \in \mathrm{gl}(M)$ is the identity transformation and we note $N(S',T) = N(S,T)$. We shall then first prove the following proposition.

Proposition 3: Suppose that we have $N(S,S) = 0$. Then, for any two integers n and m, we have also $N(S^n,S^m) = 0$.

The proof consists of studying the following four cases, separately; (i) $n = 0$, or $m = 0$, (ii) $n \geqslant 1$, $m \geqslant 1$, (iii) $n \leqslant -1$, $m \leqslant -1$, (iv) $n \geqslant 1$, and $m \leqslant -1$. The case (i) is trivial since $S_0 = I$. For the case (ii), we note the validity of the following identity:

$$\begin{aligned} & 2N(S^n,S^m|X,Y) \\ & = \sum_{l_1=1}^{n} \sum_{l_2=1}^{m} S^{l_1+l_2-2}\{N(S,S\,|S^{n-l_1}X,S^{m-l_2}Y) \\ & \quad + N(S,S\,|S^{m-l_2}X,S^{n-l_1}Y)\}, \end{aligned} \tag{3.10}$$

for $n \geqslant 1$ and $m \geqslant 1$, which can be proved from Eq. (3.5). Case (iii) is reduced to case (ii) by considering S^{-1} instead of S and noting another identity

$$N(S^{-1},S^{-1}|X,Y) = S^{-1}S^{-1}N(S,S\,|S^{-1}X,S^{-1}Y). \tag{3.11}$$

Finally, for the last case of (iv), we utilize

$$\begin{aligned} & N(S^n,S^{-l}|X,Y) \\ & = S^{n-l}N(S^l,S^l|S^{-l}X,S^{-l}Y) \\ & \quad - S^{-l}N(S^{n+l},S^l|S^{-l}X,S^{-l}Y) \end{aligned} \tag{3.12}$$

for any integer n and l, and reduce the problem to the previous cases. This completes the proof of Proposition 3.

We also note the validity of

$$\begin{aligned} & N(S^n,S^{m+l}|X,Y) + N(S^m,S^{n+l}|X,Y) \\ & = N(S^n,S^m|X,S^lY) + N(S^n,S^m|S^lX,Y) \\ & \quad + S^mN(S^n,S^l|X,Y) \\ & \quad + S^nN(S^m,S^l|X,Y) \end{aligned} \tag{3.13}$$

for any three integers n, m, and l. We could have also utilized this identity for the proof of Proposition 3.

Remark 4: Actually, we can generalize Proposition 3 as follows. We first define $\widetilde{N}_{\mu\nu}^\lambda\, (S,T)$ by replacing all ordinary derivatives in Eq. (3.9) by covariant derivatives as in Eq. (2.2). Similary, we can define $\widetilde{M}_{\mu\nu}^\lambda\, (S,T)$ by changing some signs in $\widetilde{N}^{\lambda}_{\mu\nu}(S,T)$ as in Sec. II. A basis-independent definition of $\widetilde{N}(S,T)$ and $\widetilde{M}(S,T)$ can also be given by redefining the commutor $[X,Y]$ suitably. Then, we can prove that all relations discussed so far in this section hold valid also for $\widetilde{N}(S,T)$ and $\widetilde{M}(S,T)$, instead of $N(S,T)$. Especially, we find that all of $\widetilde{N}^{\lambda}_{\mu\nu}(S^n,S^m)$ or $\widetilde{M}^{\lambda}_{\mu\nu}(S^n,S^m)$ are identically zero, provided that $\widetilde{N}^{\lambda}_{\mu\nu}(S,S)$ or $\widetilde{M}^{\lambda}_{\mu\nu}(S,S)$ vanishes, respectively. However, since this generalization does not appear to give any useful result in contrast to the case of $N(S,T)$ in Sec. IV, we will not elaborate them here.

Next, let $L(X)=L_X$ be the Lie derivative[25] with respect to a tangent vector X; its actions upon $Y\in T_M$ and $S\in \mathrm{gl}(T_M)$ are defined by

$$L(X)Y=[X,Y], \tag{3.14a}$$

$$(L(X)S)Y=[X,SY]-S[X,Y]. \tag{3.14b}$$

Then, it is not difficult to show the validity of

$$L(X)N(S,T)=N(L(X)S,T)+N(S,L(X)T), \tag{3.15}$$

and

$$N(S,S\,|X,Y)=(L(SX)S)Y-S(L(X)S)Y. \tag{3.16}$$

To be complete, we will prove the following proposition, which is essentially due to Filippo *et al.*[10]

Proposition 4: Suppose that we have a tangent vector X_0 satisfying $L(X_0)S=0$. Then, we have

$$L(X_0)N(S,S)=0.$$

Moreover define X_n for any integer n by

$$X_n=S^nX_0,$$

and assume that we have $N(S,S)=0$ in addition to $L(X_0)S=0$, then for any two integers n and m, we have

(i) $L(X_n)S=0$,

(ii) $S^m[X_n,Y]=[X_n,S^mY]$ for any tangent vector Y,

(iii) $[X_n,X_m]=0$.

Now we will give the proof of the proposition. First of all, $L(X_0)N(S,S)=0$ follows immediately from Eq. (3.15) if we have $L(X_0)S=0$. Now, suppose that we have $N(S,S)=0$, in addition. Setting $X=X_n$ in Eq. (3.16), it gives then

$$(L(X_{n+1})S)Y-S(L(X_n)S)Y=0.$$

By induction, this leads to $L(X_n)S=0$ for both positive and negative values of n, starting with $n=0$, or $n=-1$. Next, we set $X=X_n$ in Eq. (3.14) and note $L(X_n)S=0$. This immediately gives $S[X_n,Y]=[X_n,SY]$. Repeating this, we find relation (ii). Then, we calculate

$$\begin{aligned}[X_n,X_m]&=[X_n,S^mX_0]\\ &=S^m[X_n,X_0]=-S^m[X_0,S^nX_0]\\ &=-S^mS^n[X_0,X_0]=0,\end{aligned}$$

which proves relation (iii).

IV. THEORY WITH DUAL SYMPLECTIC STRUCTURE

Let $f^{\mu\nu}=-f^{\nu\mu}$ be a fixed bivector field. For any two antisymmetric tensors $F_{\mu\nu}=-F_{\nu\mu}$ and $G_{\mu\nu}=-G_{\nu\mu}$, we define the third antisymmetric tensor $(F\cdot G)_{\mu\nu}$ by

$$(F\cdot G)_{\mu\nu}=-(F\cdot G)_{\nu\mu}=\tfrac{1}{2}\{F_{\mu\alpha}f^{\alpha\beta}G_{\beta\nu}-F_{\nu\alpha}f^{\alpha\beta}G_{\beta\mu}\}, \tag{4.1}$$

which is clearly commutative, i.e.,

$$(F\cdot G)_{\mu\nu}=(G\cdot F)_{\mu\nu}. \tag{4.2}$$

The product defines a nonassociative Jordan product[26] which is special. Moreover, we assume hereafter that $f^{\mu\nu}$ possesses the inverse $f_{\mu\nu}=-f_{\nu\mu}$,

$$f_{\mu\lambda}f^{\lambda\nu}=\delta^{\nu}_{\mu}. \tag{4.3}$$

Then, the tensor $f_{\mu\nu}$ acts the role of the unit element of the Jordan product, since

$$f\cdot F=F\cdot f=F. \tag{4.4}$$

Hereafter, we assume also that $F_{\mu\nu}$ has its inverse $F^{\mu\nu}=-F^{\nu\mu}$,

$$F_{\mu\lambda}F^{\lambda\nu}=\delta^{\nu}_{\mu}. \tag{4.5}$$

Moreover, we define tensors F^0 and F^{-1} by

$$(F^0)_{\mu\nu}=f_{\mu\nu},\quad (F^0)^{\mu\nu}=f^{\mu\nu} \tag{4.6}$$

and

$$(F^{-1})_{\mu\nu}=f_{\mu\alpha}F^{\alpha\beta}f_{\beta\nu}, \tag{4.7a}$$

$$(F^{-1})^{\mu\nu}=f^{\mu\alpha}F_{\alpha\beta}f^{\beta\nu}. \tag{4.7b}$$

We can now construct the nth power tensor $(F^n)_{\mu\nu}$, and $(F^n)^{\mu\nu}$ inductively as usual. Then, it is easy to verify the power associative law

$$F^n\cdot F^m=F^{n+m}, \tag{4.8}$$

as well as

$$(F^n)_{\mu\lambda}(F^n)^{\lambda\nu}=\delta^{\nu}_{\mu}, \tag{4.9a}$$

$$(F^{-n})_{\mu\nu}=f_{\mu\alpha}(F^n)^{\alpha\beta}f_{\beta\nu}, \tag{4.9b}$$

for any integers n and m. Note that the validity of the power-associative law [Eq. (4.8)] is of course a special case of a more general property[26] of any Jordan algebra.

Next, let $F_{\mu\nu}=-F_{\nu\mu}$ and $G_{\mu\nu}=-G_{\nu\mu}$ be any two antisymmetric tensors. We set for later conveniences

$$\Delta_{\lambda\mu\nu}(G)=\partial_\lambda G_{\mu\nu}+\partial_\mu G_{\nu\lambda}+\partial_\nu G_{\lambda\mu}, \tag{4.10}$$

$$\begin{aligned}2[F|G]^{\lambda\mu\nu}=2[G|F]^{\lambda\mu\nu}&=F^{\lambda\alpha}\partial_\alpha G^{\mu\nu}\\ &+F^{\mu\alpha}\partial_\alpha G^{\nu\lambda}\\ &+F^{\nu\alpha}\partial_\alpha G^{\lambda\mu}+G^{\lambda\alpha}\partial_\alpha F^{\mu\nu}+G^{\mu\alpha}\partial_\alpha F^{\nu\lambda}\\ &+G^{\nu\alpha}\partial_\alpha F^{\lambda\mu},\end{aligned} \tag{4.11}$$

both of which are totally antisymmetric in λ, μ, and ν. Then, it is straightforward to find the following identity:

$$\begin{aligned}&2F_{\alpha\mu}F_{\beta\nu}F_{\gamma\lambda}[F|G]^{\mu\nu\lambda}\\ &\quad=-\Delta_{\alpha\beta\gamma}(FGF)+G^{\mu\lambda}\{F_{\alpha\mu}\Delta_{\lambda\beta\gamma}(F)\\ &\qquad+F_{\beta\mu}\Delta_{\lambda\gamma\alpha}(F)+F_{\gamma\mu}\Delta_{\lambda\alpha\beta}(F)\},\end{aligned} \tag{4.12}$$

where we have set

$$(FGF)_{\mu\nu}=F_{\mu\alpha}G^{\alpha\beta}F_{\beta\nu}, \tag{4.13}$$

so that

$$F\cdot G=\tfrac{1}{2}(FfG+GfF), \tag{4.14}$$

and especially for $G=f$, this gives

$$FfF = F\cdot F. \tag{4.14'}$$

Especially, if we set $F_{\mu\nu} = f_{\mu\nu}$ and note $f\cdot f = f$, then it gives the relation

$$f_{\alpha\mu} f_{\beta\nu} f_{\gamma\lambda} [f|f]^{\mu\nu\lambda} = \Delta_{\alpha\beta\gamma}(f). \tag{4.15}$$

Since $f_{\mu\nu}$ is actually arbitrary, this relation is also valid for replacing $f_{\mu\nu}$ by $F_{\mu\nu}$, i.e.,

$$F_{\alpha\mu} F_{\beta\nu} F_{\gamma\lambda} [F|F]^{\mu\nu\lambda} = \Delta_{\alpha\beta\gamma}(F). \tag{4.16}$$

Next, as in Sec. II we set

$$S_\mu^{\ \nu} = F_{\mu\lambda} f^{\lambda\nu}. \tag{4.17}$$

Then we find

$$(S^n)_\mu^{\ \nu} = (F^n)_{\mu\lambda} f^{\lambda\nu}, \tag{4.18a}$$

$$(S^{-n})_\mu^{\ \nu} = (F^{-n})_{\mu\lambda} f^{\lambda\nu} = f_{\mu\lambda} (F^n)^{\lambda\nu}. \tag{4.18b}$$

Moreover, the following identity can be verified:

$$\begin{aligned} 2N_{\mu\nu}^{\tau}(S,T) &= 2 f^{\tau\lambda} \Delta_{\lambda\mu\nu}(F\cdot G) \\ &\quad + f^{\tau\lambda} f^{\alpha\beta} \{ F_{\nu\beta} \Delta_{\mu\alpha\lambda}(G) \\ &\quad + G_{\nu\beta} \Delta_{\mu\alpha\lambda}(F) - F_{\mu\beta} \Delta_{\nu\alpha\lambda}(G) \\ &\quad - G_{\mu\beta} \Delta_{\nu\alpha\lambda}(F) \} + (F_{\mu\alpha} G_{\nu\beta} + G_{\mu\alpha} F_{\nu\beta}) [f|f]^{\tau\alpha\beta} \end{aligned} \tag{4.19}$$

or

$$\begin{aligned} 2 f_{\lambda\tau} N_{\mu\nu}^{\tau}(S,T) &= 2\Delta_{\lambda\mu\nu}(F\cdot G) + S_\mu^{\ \alpha} \Delta_{\alpha\lambda\nu}(G) + T_\mu^{\ \alpha} \Delta_{\alpha\lambda\nu}(F) \\ &\quad - S_\nu^{\ \alpha} \Delta_{\alpha\lambda\mu}(G) - T_\nu^{\ \alpha} \Delta_{\alpha\lambda\mu}(F) \\ &\quad + (S_\mu^{\ \alpha} T_\nu^{\ \beta} + T_\mu^{\ \alpha} S_\nu^{\ \beta}) \Delta_{\lambda\alpha\beta}(f), \end{aligned} \tag{4.19'}$$

where we have set

$$T_\mu^{\ \nu} = G_{\mu\lambda} f^{\lambda\nu}. \tag{4.20}$$

Now, we assume hereafter that

$$f = \tfrac{1}{2} f_{\mu\nu}\, dx^\mu \wedge dx^\nu \tag{4.21}$$

is a symplectic form, so that we have

$$[f|f]^{\lambda\mu\nu} = \Delta_{\lambda\mu\nu}(f) = 0. \tag{4.22}$$

Then, replacing $F_{\mu\nu}$ and $G_{\mu\nu}$ in Eq. (4.19) by $(F^n)_{\mu\nu}$ and $(F^m)_{\mu\nu}$, respectively, it gives

$$\begin{aligned} 2N_{\mu\nu}^{\tau}(S^n,S^m) &= 2 f^{\tau\lambda} \Delta_{\lambda\mu\nu}(F^{n+m}) + f^{\tau\lambda} f^{\alpha\beta} \{ (F^n)_{\nu\beta} \Delta_{\mu\alpha\lambda}(F^m) \\ &\quad + (F^m)_{\nu\beta} \Delta_{\mu\alpha\lambda}(F^n) - (F^n)_{\mu\beta} \Delta_{\nu\alpha\lambda}(F^m) \\ &\quad - (F^m)_{\mu\beta} \Delta_{\nu\alpha\lambda}(F^n) \}, \end{aligned} \tag{4.23}$$

for any two integers n and m. Further, we assume hereafter that

$$F = \tfrac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu \tag{4.24}$$

is also a symplectic form and hence

$$[F|F]^{\lambda\mu\nu} = \Delta_{\lambda\mu\nu}(F) = 0. \tag{4.25}$$

Especially, Eq. (4.23) for $n = m = 1$ gives then

$$N_{\mu\nu}^{\tau}(S,S) = f^{\tau\lambda} \Delta_{\lambda\mu\nu}(F\cdot F). \tag{4.26}$$

Finally, suppose that the Nijenhuis tensor $N(S,S)$ is identically zero also, i.e.,

$$N(S,S) = 0. \tag{4.27}$$

In that case, Proposition 3 in Sec. III tells us that we have $N(S^n,S^m) = 0$ also for any integers n and m. Then, by induction on Eq. (4.23), we find $\Delta_{\lambda\mu\nu}(F^n) = 0$ for any nonnegative integer n. Next, we will prove shortly that $\Delta_{\lambda\mu\nu}(F^{-1}) = 0$. Accepting this, Eq. (4.23) leads also to $\Delta_{\lambda\mu\nu}(F^n) = 0$ for any negative integers by induction, again. Now, we will proceed to its proof. Setting $G = f$ in Eq. (4.12) and noting Eqs. (4.14) and (4.26), we see first $[f|F]^{\mu\nu\lambda} = [F|f]^{\mu\nu\lambda} = 0$ if $N_{\mu\nu}^{\lambda}(S,S) = 0$ and $\Delta_{\lambda\mu\nu}(F) = 0$. Then, we make the following substitution $G \to F \to f$ in Eq. (4.12) and note $FGF \to fFf = F^{-1}$ because of Eq. (4.7). This gives the desired result $\Delta_{\lambda\mu\nu}(F^{-1}) = 0$ when we use $[f|F]^{\mu\nu\lambda} = 0$ and $\Delta_{\lambda\mu\nu}(f) = 0$. Therefore, we have proved $\Delta_{\lambda\mu\nu}(F^n) = 0$ for any integer n. Finally, we substitute $F \to F^n$ and $G \to F^m$ in Eq. (4.12) to find $[F^n|F^m]^{\mu\nu\lambda} = 0$. Hence we have proved the following proposition.

Proposition 5: Let us assume that both two-forms f and F are symplectic and that we have $N(S,S) = 0$ for the tensor S defined by Eq. (4.17). Then, we have

$$\begin{aligned} \Delta_{\lambda\mu\nu}(F^n) &= \partial_\lambda (F^n)_{\mu\nu} + \partial_\mu (F^n)_{\nu\lambda} \\ &\quad + \partial_\nu (F^n)_{\lambda\mu} = 0 \end{aligned} \tag{4.28}$$

for any integer n. Moreover, we also find

$$[F^n|F^m]^{\lambda\mu\nu} = 0 \tag{4.29}$$

for any two integers n and m.

Because of Eq. (4.28) or Eq. (4.29) with $n = m$, we can construct an infinite series of Poisson brackets by

$$\{g,h\}_n = (F^n)^{\mu\nu} \partial_\mu g\, \partial_\nu h, \tag{4.30}$$

for any integer n. Especially for $n = 0$, and ± 1, this gives

$$\{g,h\}_0 = f^{\mu\nu} \partial_\mu g\, \partial_\nu h, \tag{4.31a}$$

$$\{g,h\}_1 = F^{\mu\nu} \partial_\mu g\, \partial_\nu h, \tag{4.31b}$$

$$\{g,h\}_{-1} = -F_{\mu\nu} f^{\mu\alpha} \partial_\alpha g\, f^{\nu\beta} \partial_\beta h. \tag{4.31c}$$

Moreover, any K_n's are also in involution with respect to any of these Poisson brackets, i.e.,

$$\{K_n, K_m\}_p = 0, \tag{4.32}$$

for all integers n, m, and p because we calculate

$$\{K_n, K_m\}_p = \{K_{n-p}, K_m\}_0 = 0$$

by the hierarchy relation Eq. (2.11). Further, if we set

$$Q^{\mu\nu} = \sum_n C_n (F^n)^{\mu\nu} = -Q^{\nu\mu} \tag{4.33}$$

for arbitrary constants C_n, Eq. (4.29) gives

$$[Q|Q]^{\lambda\mu\nu} = 0, \tag{4.34}$$

so that the bracket

$$\{g,h\}_Q = Q^{\mu\nu} \partial_\mu g\, \partial_\nu h \tag{4.35}$$

defines a Lie algebra with $\{K_n, K_m\}_Q = 0$ again. This generalizes the result of earlier references for the case of $Q^{\mu\nu} = f^{\mu\nu} + F^{\mu\nu}$. If the inverse $Q_{\mu\nu}$ of $Q^{\mu\nu}$ exists, then $\{g,h\}_Q$ defines a Poisson bracket, of course.

We can improve on the result of Sec. II for existence of a Lagrangian in our cases as follows. In order to avoid possible confusion, we write the time variable t as t_p for the Hamiltonian $H = K_p$ as in the KdV theory. Because of Eq. (4.28), the Poincare lemma assures existence of functions $\theta_\mu^{(n)}(x)$ such that

$$(F^n)_{\mu\nu} = \partial_\mu \theta_\nu^{(n)} - \partial_\nu \theta_\mu^{(n)}, \tag{4.36}$$

at least locally. Now for a given integer value of p, we construct a infinite series of Lagrangians by

$$L_p^{(n)} = \theta_\mu^{(n)}(x)\dot{x}^\mu - K_{n+p}(x), \tag{4.37a}$$

$$\dot{x}^\mu = \frac{d}{dt_p} x^\mu \tag{4.37b}$$

for $n = 0, \pm 1, \pm 2,...$.

The Hamilton–Lagrange equation of motion based upon Eq. (4.37) is evidently

$$(F^n)_{\mu\lambda}\dot{x}^\lambda = \partial_\mu K_{p+n}. \tag{4.38}$$

However, because of the hierarchy relation Eq. (2.11) with Eq. (2.24), Eq. (4.38) is equivalent to a single equation

$$f_{\mu\lambda}\dot{x}^\lambda = \partial_\mu K_p, \tag{4.39}$$

irrespective of the value of n. In other words, all Lagrangians $L_p^{(n)}$ ($n = 0, \pm 1, \pm 2,...$) given by Eq. (4.37) are dynamically equivalent. Although existence, in principle, of infinite number of equivalent Lagrangians is known,[23,27] it is in general not possible to construct them explicitly as here. Moreover, our Lagrangians $L_p^{(n)}$ is associated with the Poisson bracket $\{g,h\}_n$ defined by Eq. (4.30). Further, let us set

$$P_{\mu\nu} = \sum_n b_n (F^n)_{\mu\nu} \tag{4.40}$$

for arbitrary constants b_n. Then, by Proposition 5, we have $\Delta_{\lambda\mu\nu}(P) = 0$ so that there exists functions $\theta_\mu(x)$ satisfying

$$P_{\mu\nu} = \partial_\mu\theta_\nu - \partial_\nu\theta_\mu. \tag{4.41}$$

Now, the Lagrangian defined by

$$L_P = \theta_\mu(x)\dot{x}^\mu - H(x), \tag{4.42a}$$

$$H(x) = \sum_n b_n\, \partial_\mu K_{p+n} \tag{4.42b}$$

gives an equation of motion compatible with Eq. (4.39). Especially, if the inverse $P^{\mu\nu}$ of $P_{\mu\nu}$ exists, then L_P is equivalent to Lagrangians $L_p^{(n)}$ given by Eq. (4.37). Similarly, if the inverse $Q_{\mu\nu}$ of $Q^{\mu\nu}$ defined by Eq. (4.33) exists, we can construct another equivalent Lagrangian.

Next, let us introduce tangent vectors by

$$X_n = f^{\mu\nu}\partial_\nu K_n\, \partial_\mu, \tag{4.43}$$

for any integer n. In view of the hierarchy equation, we can rewrite it also as

$$X_n = S^n X_0 = F^{\mu\nu}\partial_\nu K_{n+1}\, \partial_\mu. \tag{4.44}$$

Since two forms f and F are symplectic, Eqs. (4.43) and (4.44) lead to $L(X_n)f = L(X_n)F = 0$ so that

$$L(X_n)S = 0, \tag{4.45}$$

as in (II). Therefore, all results of Proposition 4 of Sec. III hold valid. Especially, we have

$$[X_n, X_m] = 0.$$

However, as we will show in Sec. V, this relation can be really derived under assumptions much weaker than those assumed in this section. Also, Eq. (4.45) with $n = p$ implies the validity of the Lax equation[28]

$$\frac{d}{dt_p} S_\mu^\nu = S_\mu^\lambda (U_p)_\lambda^\nu - (U_p)_\mu^\lambda S_\lambda^\nu, \tag{4.46a}$$

$$(U_p)_\mu^\nu = \partial_\mu (f^{\nu\lambda}\partial_\lambda K_p), \tag{4.46b}$$

as in (II), so that all K_n's are conserved quantities of the theory also from this.

In ending this section, we would like to make the following remark.

Remark 5: We introduce the antisymmetric tensor $(K_{n,m})_{\mu\nu}$ by

$$(K_{n,m})_{\mu\nu} = -(K_{m,n})_{\mu\nu} = \partial_\mu K_n\, \partial_\nu K_m - \partial_\nu K_n\, \partial_\mu K_m, \tag{4.47}$$

for any two integers n and m. Then, the hierarchy relation together with $\{K_n, K_m\}_0 = 0$ leads to

$$F^l \cdot K_{n,m} = \tfrac{1}{2}(K_{n+l,m} + K_{n,m+l}), \tag{4.48}$$

$$K_{n,m} \cdot K_{l,p} = 0. \tag{4.49}$$

Together with

$$F^n \cdot F^m = F^{n+m}, \tag{4.50}$$

these define the multiplication table of a special Jordan algebra which is not associative. Moreover, a set consisting of all $K_{n,m}$'s generates a nilpotent ideal of the algebra so that the Jordan algebra is not simple.

Setting

$$(\widetilde{K}_{n,m})_\mu^\nu = (K_{n,m})_{\mu\lambda} f^{\lambda\nu}, \tag{4.51}$$

then we can show readily from Eqs. (4.19), (4.48), and (4.49) that we have

$$N(S^l, \widetilde{K}_{n,m}) = N(\widetilde{K}_{n,m}, \widetilde{K}_{p,q}) = 0. \tag{4.52}$$

Next, for arbitrary constants $a_{n,m} = -a_{m,n}$, we define

$$\tilde{f}_{\mu\nu} = f_{\mu\nu} + \sum_{m,n} a_{n,m} (K_{n,m})_{\mu\nu}, \tag{4.53}$$

It evidently satisfies

$$\Delta_{\lambda\mu\nu}(\tilde{f}) = 0. \tag{4.54}$$

Moreover, Eq. (4.49) can be used to show the invertibility of $\tilde{f}_{\mu\nu}$ with its inverse $\tilde{f}_{\mu\nu}$ given by

$$\tilde{f}^{\mu\nu} = f^{\mu\nu} - \sum_{n,m} a_{nm} f^{\mu\alpha}(K_{nm})_{\alpha\beta} f^{\beta\nu}. \tag{4.55}$$

Therefore,

$$\tilde{f} = \tfrac{1}{2}\tilde{f}_{\mu\nu}\, dx^\mu \wedge dx^\nu \tag{4.56}$$

defines another symplectic form, which we could use instead of f.

V. GENERATING FUNCTION AND EIGENVALUES OF S

In this section, we assume only the minimum conditions specified in Sec. II, i.e., the validity of the hierarchy equation

$$(S^n)_\mu^\lambda\, \partial_\lambda K_m = \partial_\mu K_{n+m}, \tag{5.1}$$

and a particular form Eq. (2.23), i.e.,

$$S_\mu^\nu = F_{\mu\lambda} f^{\lambda\nu}, \tag{5.2}$$

for antisymmetric tensors $F_{\mu\nu} = -F_{\nu\mu}$ and symplectic $f_{\mu\nu}$. However, we need not assume the two-form F to be symplectic. Now, Eq. (5.1) is rewritten as

$$f^{\mu\lambda}\,\partial_\lambda K_n = (F^m)^{\mu\lambda}\,\partial_\lambda K_{n+m}. \tag{5.3}$$

Let z be a complex variable, and define a generating function

$$\phi_p(x,z) = \sum_{n=0}^{\infty} K_{n+p}(x) z^n, \tag{5.4}$$

which will define an analytic function of z as will be shown shortly. Then, it is easy to see the validity of

$$\frac{\partial}{\partial x^\mu}\phi_p(x,z) = S_\mu^\nu(x)\frac{\partial}{\partial x^\nu}\phi_{p-1}(x,z), \tag{5.5}$$

which is essentially equivalent to the hierarchy equation (5.1). Moreover, if we introduce the Lagrangian L as in Eq. (2.24) or $L_p^{(0)}$ as in Eq. (4.35), and write the time variable explicitly as t_p, then we have

$$\dot{x}^\mu = \frac{d}{dt_p}x^\mu = (F^{\mu\lambda} - zf^{\mu\lambda})\frac{\partial}{\partial x^\lambda}\phi_{p+1}(x,z). \tag{5.6}$$

After these preparations, let us now discuss eigenvalues of S_μ^ν. We regard S_μ^ν, $f_{\mu\nu}$, and $F_{\mu\nu}$ to be $2N \times 2N$ matrices, and consider the secular equation

$$\det(S - \lambda I) = 0, \tag{5.7}$$

where $I_\mu^\nu = \delta_\mu^\nu$ is the $2N \times 2N$ identity matrix. Because of Eq. (5.2) and since $f^{\mu\nu}$ is nonsingular, Eq. (5.7) is equivalent to

$$\det(F - \lambda f) = 0. \tag{5.8}$$

However, since $F_{\mu\nu} - \lambda f_{\mu\nu}$ is antisymmetric, its determinant is a square of its associated Pfaffian. Therefore, we conclude that the solution for λ of Eq. (5.8) appear always in pairs and we write them as $(\lambda_1,\lambda_1),(\lambda_2,\lambda_2),\ldots,(\lambda_N,\lambda_N)$. Then, we find

$$K_n = \frac{1}{n}\sum_{j=1}^{N}(\lambda_j)^n \quad (n \neq 0), \tag{5.9a}$$

$$K_0 = \sum_{j=1}^{N}\log\lambda_j \quad (n=0). \tag{5.9b}$$

Note that the matrix S need not be fully diagonalizable for this. Because of our assumption of the existence of $F^{\mu\nu}$, $\det F \neq 0$ and hence none of λ_j can be zero. Inserting these values of K_n's into Eq. (5.4), we calculate

$$\phi_0(x,z) = \sum_{j=1}^{N}\log\lambda_j - \sum_{j=1}^{N}\log(1-\lambda_j z), \tag{5.10a}$$

$$\phi_1(x,z) = -\frac{1}{z}\sum_{j=1}^{N}\log(1-\lambda_j z). \tag{5.10b}$$

Then, Eq. (5.5) for $p = 1$ is rewritten as

$$\sum_{j=1}^{N}\frac{1}{1-\lambda_j z}\partial_\mu\lambda_j = \sum_{j=1}^{N}\frac{1}{\lambda_j(1-\lambda_j z)}S_\mu^\nu\,\partial_\nu\lambda_j. \tag{5.11}$$

Note that both λ_j and S_μ^ν are functions of the coordinate x^μ but not of the complex parameter z.

We now assume hereafter that $\lambda_j \neq \lambda_k$ whenever $j \neq k$, except for some accidental points. This would be the case if $K_1,K_2,\ldots,K_n$ are functionally independent so that the system is completely integrable. As we noted in (II), such a situation occurs for example in the case of the Toda lattice. Then, taking the residue of both sides of Eq. (5.11) at $z = 1/\lambda_j$, this gives

$$S_\mu^\nu\,\partial_\nu\lambda_j = \lambda_j\,\partial_\mu\lambda_j \quad (j = 1,2,\ldots,N). \tag{5.12}$$

In other words, $\partial_\mu\lambda_j$ is one of two eigenvectors belonging to the eigenvalue λ_j, provided that $\partial_\mu\lambda_j$ is not identically zero. Note that our ansatz are weaker than those assumed in Ref. 10, since our results are valid under weaker assumptions of $\widetilde{N}_{\mu\nu}^\lambda = 0$ or $\widetilde{M}_{\mu\nu}^\lambda = 0$ instead of $N_{\mu\nu}^\lambda = 0$ and also without assuming the tensor $F_{\mu\nu}$ being symplectic.

We rewrite Eq. (5.12) also as

$$((F^{-1})^{\mu\nu} - \lambda_j f^{\mu\nu})\partial_\nu\lambda_j = 0, \tag{5.13}$$

which gives the orthogonality relation

$$f^{\mu\nu}\,\partial_\mu\lambda_j\,\partial_\nu\lambda_k = 0, \tag{5.14}$$

which will lead again to the involution law $\{K_n,K_m\}_0 = 0$. Now, we introduce N tangent vectors Y_j by

$$Y_j = f^{\mu\nu}\,\partial_\nu\lambda_j\,\partial_\mu \quad (j = 1,2,\ldots,N). \tag{5.15}$$

Then, Eq. (5.13) can be rewritten in a more compact form

$$SY_j = \lambda_j Y_j \quad (j = 1,2,\ldots,N), \tag{5.16}$$

where we regard S now as an element of $\mathrm{gl}(T_M)$ by Eq. (3.3). We remark that Y_j can be a complex (rather than real) vector since the eigenvalue λ_j is generally complex.

Next, we can solve, in principle, N unknown quantities $\lambda_1,\lambda_2,\ldots,\lambda_N$ in terms of $K_1,K_2,\ldots,K_N$ from Eq. (5.9a). Therefore, we regard λ_j $(j = 1,2,\ldots,N)$ as algebraic functions of $K_1,K_2,\ldots,K_N$, hereafter. Then, setting,

$$X_n = f^{\mu\nu}\,\partial_\nu K_n\,\partial_\mu, \tag{5.17}$$

we can rewrite Y_j as

$$Y_j = \sum_{n=1}^{N}\theta_j^{(n)} X_n, \tag{5.18a}$$

$$\theta_j^{(n)} = \frac{\partial\lambda_j}{\partial K_n}. \tag{5.18b}$$

Moreover, we find

$$X_m(\theta_j^{(n)}) = 0, \tag{5.19}$$

since we calculate

$$X_m(\theta_j^{(n)}) = -\{K_m,\theta_j^{(n)}\}_0 = -\sum_{l=1}^{N}\{K_m,K_l\}_0\frac{\partial}{\partial K_l}\theta_j^{(n)} = 0$$

when we note $\theta_j^{(n)}$ to be also functions of $K_1,K_2,\ldots,K_N$. Therefore, we obtain

$$[X_m,Y_j] = \sum_{n=1}^{N}\theta_j^{(n)}[X_m,X_n], \tag{5.20a}$$

$$[Y_j,Y_k] = \sum_{n,m=1}^{N}\theta_j^{(n)}\theta_k^{(m)}[X_n,X_m]. \tag{5.20b}$$

However, we show shortly

$$[X_n,X_m] = 0 \tag{5.21}$$

identically, so that we find

$$[X_m,Y_j] = [Y_j,Y_k] = 0. \tag{5.22}$$

Now, we will prove the validity of Eq. (5.21). We calculate in a straightforward way

$$[X_n,X_m] = C^{\mu}_{nm}\,\partial_\mu \tag{5.23a}$$

with

$$C^{\mu}_{nm} = -f^{\mu\nu}\partial_\nu\{K_n,K_m\}_0 + [f|f]^{\mu\alpha\beta}\partial_\alpha K_n\,\partial_\beta K_m$$
$$= 0, \tag{5.23b}$$

since $\{K_n,K_m\}_0 = [f|f]^{\mu\alpha\beta} = 0$ by our assumptions. It is rather remarkable that we need not assume the validity of stronger ansatz such as $N^{\lambda}_{\mu\nu} = 0$ and $L(X_n)S = 0$ as in Sec. IV in order to obtain the relations Eqs. (5.21) and (5.22) with Eq. (5.16).

Remark 6: Let $h(x)$ and $g(x)$ be two differentiable functions of x, and set

$$X_h = f^{\mu\nu}\partial_\mu h\,\partial_\nu, \tag{5.24a}$$

$$X_g = f^{\mu\nu}\partial_\mu g\,\partial_\nu, \tag{5.24b}$$

then, Eqs. (5.23) by replacing K_n and K_m by h and g, respectively, give

$$[X_h,X_g] = X_{\{h,g\}}, \tag{5.25}$$

provided that $[f|f]^{\lambda\mu\nu} = 0$, i.e., f is a symplectic form. Equation (5.25), of course, is the well-known homomorphism between Lie algebras of Poisson bracket and tangent vectors.

Remark 7: If we have assumed $F_{\mu\nu} = F_{\nu\mu}$ instead of $F_{\mu\nu} = -F_{\nu\mu}$, then Eq. (5.8) requires that if λ_j is an eigenvalue of S, then so will be $-\lambda_j$. Therefore, for any odd integer n, K_n is identically zero. Then, because of the hierarchy equation, we also have $\partial_\mu K_n = 0$ for any even integer n, as we stated in Remark 2 of Sec. II.

Also, returning to the case of $F_{\mu\nu} = -F_{\nu\mu}$, Eq. (5.9) implies that at most only N quantities among K_n's can be algebraically independent.

In ending this section, we will comment on the unlikeliness of the validity of the modified hierarchy equation (2.36) of Sec. II. In that case, Eq. (5.12) will be replaced by

$$B^{-1}S^{\nu}_{\mu}\,\partial_\nu\lambda_j = \lambda_j\,\partial_\mu\lambda_j.$$

Therefore, assuming that none of $\partial_\mu\lambda_j$ is identically zero, both S^{ν}_{μ} and $B^{-1}S^{\nu}_{\mu}$ must have the same sets of eigen-values. In other words, both sets $\{\lambda_1,\lambda_2,...,\lambda_N\}$ and $\{B\lambda_1,B\lambda_2,...,B\lambda_N\}$ must coincide. This is possible only if we have $B = 1$ or $B = -1$. Note that $B = 1$ corresponds to the standard hierarchy equation. For the latter case of $B = -1$, eigenvalues should occur in pairs λ_j and $-\lambda_j$, leading to the uninteresting case of $\partial_\mu K_n = 0$ by the same reasoning we already used. If we have $\partial_\mu\lambda_j = 0$ for some indices j, the situation allows other possibilities to say something definite. However, the case is physically rather unlikely, not to mention the fact that $K_1,K_2,...,K_N$ are no longer algebraically independent. We believe that the validity of the modified hierarchy equation (2.36) with $B \neq 1$ is physically not interesting at best.

VI. FINAL COMMENTS

We have assumed through this paper that the inverse $(S^{-1})^{\nu}_{\mu}$ of S^{ν}_{μ} exists. However, in case that it does not exist, most of our results would hold if we remove all cases dependent on the assumptions. Then for example, only K_n's for positive integers are to be used in the hierarchy equation. Similarly the involution law $\{K_n,K_m\}_0 = 0$ is still valid if we restrict ourselves to positive integer values of n and m.

In ending, it may be of some interest to make a remark on a possible connection between the present approach and that based upon the P matrix.[29] Let lower-case latin indices $a,b,...,$ assume M values $1,2,...,M$, which correspond to some indices of an internal space. If S^{b}_{a}'s satisfies

$$\{S^{b}_{a}, S^{c}_{d}\}_0 = S^{j}_{a}\,P^{bc}_{jd} - S^{b}_{j}\,P^{jc}_{ad} - S^{j}_{d}\,P^{cb}_{ja}$$
$$+ S^{c}_{j}\,P^{jb}_{da} \tag{6.1}$$

for some functions P^{bc}_{ad} of x^{μ}, and if we define

$$K_n = (1/2n)\mathrm{Tr}\,S^n = (1/2n)(S^n)^b_b \quad (n\neq 0), \tag{6.2}$$

then it is easy to show the validity of the involution law

$$\{K_n,K_m\}_0 = 0. \tag{6.3}$$

It is known[29] that many completely integrable models satisfy the condition Eq. (6.1). However, this approach cannot tell us anything about the hierarchy equation and possible double symplectic structure. A possible connection exists if the internal indices $a,b,c,...,$ coincides with the greek indices $\mu,\nu,...,$ of our symplectic space with $M = 2N$. In that case, we may ask the question whether our S^{ν}_{μ}'s studied in the previous section do satisfy

$$\{S^{\nu}_{\mu},S^{\beta}_{\alpha}\}_0 = S^{\lambda}_{\mu}\,P^{\nu\beta}_{\lambda\alpha} - S^{\nu}_{\lambda}\,P^{\lambda\beta}_{\mu\alpha} - S^{\lambda}_{\alpha}\,P^{\beta\nu}_{\lambda\mu} + S^{\beta}_{\lambda}\,P^{\lambda\nu}_{\alpha\mu}. \tag{6.4}$$

Let $\Gamma^{\lambda}_{\mu\nu}$ be an affine connection as in Sec. II with the covariant derivative $S^{\nu}_{\mu;\lambda}$ defined by Eq. (2.2). Suppose that we have a covariant generalization of Eq. (6.4) with

$$f^{\lambda\tau}S^{\nu}_{\mu;\lambda}\,S^{\beta}_{\alpha;\tau} = S^{\lambda}_{\mu}\,Q^{\nu\beta}_{\lambda\alpha} - S^{\nu}_{\lambda}\,Q^{\lambda\beta}_{\mu\alpha} - S^{\nu}_{\lambda}\,Q^{\beta\nu}_{\lambda\mu} + S^{\beta}_{\lambda}\,Q^{\lambda\nu}_{\alpha\mu} \tag{6.5}$$

for some $Q^{\nu\beta}_{\mu\alpha}$. Then, we can readily see that Eq. (6.5) implies the validity of Eq. (6.4) with

$$P^{\nu\beta}_{\mu\alpha} = Q^{\nu\beta}_{\mu\alpha} - f^{\lambda\tau}\Gamma^{\nu}_{\lambda\mu}\,\{S^{\beta}_{\alpha;\tau} + \tfrac{1}{2}(\Gamma^{\gamma}_{\tau\alpha}\,S^{\beta}_{\gamma} - \Gamma^{\beta}_{\tau\gamma}\,S^{\gamma}_{\alpha})\}. \tag{6.6}$$

If $Q^{\nu\beta}_{\mu\alpha}$ is written as a covariant linear function of S^{ν}_{μ}'s, then a simple choice would be of the form

$$Q^{\nu\beta}_{\mu\alpha} = A\,(\delta^{\nu}_{\alpha}\,S^{\beta}_{\mu} + \delta^{\beta}_{\mu}\,S^{\nu}_{\alpha}) + B\,(\delta^{\nu}_{\alpha}\,S^{\beta}_{\mu} - \delta^{\beta}_{\mu}\,S^{\nu}_{\alpha})$$
$$+ C\,\delta^{\nu}_{\mu}\,S^{\beta}_{\alpha} + D\,\delta^{\beta}_{\alpha}\,S^{\nu}_{\mu} \tag{6.7}$$

for some functions, A, B, C, and D. Then, Eq. (6.5) can be rewritten as

$$f^{\lambda\tau}S^{\nu}_{\mu;\lambda}\,S^{\beta}_{\alpha;\tau} = 2A\,\{\delta^{\nu}_{\alpha}\,(SS)^{\beta}_{\mu} - \delta^{\beta}_{\mu}\,(SS)^{\nu}_{\alpha}\}. \tag{6.8}$$

Unfortunately however, the relation between the validity of, say, Eq. (6.8) and the condition $\tilde{N}^{\lambda}_{\mu\nu}(S,S) = 0$ is still not clear. Presumably, they are independent of each other and we could perhaps impose both conditions simultaneously. Such a possibility will be studied in the future.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Professor A. Das for many useful conversations, especially on the

P-matrix formulation of Sec. VI. Also, he is grateful to Professor G. Marmo for sending him reprints relevant to the present subject matter.

This work is supported in part by the U.S. Department of Energy Contract No. DE-AC02-76ER13065.

[1]G. B. Whitham, *Linear and Non-linear Waves* (Wiley, New York, 1974).
[2]M. A. Olshansky and A. M. Perelomov, Phys. Rep. **71**, 315 (1981).
[3]M. Toda, *Theory of Non-linear Lattices* (Springer, Berlin, 1981).
[4]A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[5]L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, Berlin, 1987).
[6]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[7]B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[8]I. M. Gel'fand and I. Y. Dorfmann, Funkcional. Anal. Prilozen **13**, 4, (1979); **14**, 3 (1980).
[9]S. DeFilippo, M. Salerno, G. Vilasi, and G. Marmo, Lett. Nuovo Cimento **37**, 105 (1983).
[10]S. DeFilippo, G. Vilasi, G. Marmo, and M. Salerno, Nuovo Cimento B **83**, 97 (1984).
[11]C. Ferrario, G. LoVecchio, G. Marmo, and G. Morandi, Lett. Math. Phys. **9**, 140 (1985).
[12]P. Antonini, G. Marmo, and C. Rubano, Nuovo Cimento B **86**, 17 (1985).
[13]F. Magri, C. Morosi, and O. Ragnisco, Commun. Math. Phys. **99**, 115 (1988).
[14]A. P. Stone, Ann. Inst. Fourier (Grenoble) **15**, 319 (1966).
[15]S. Okubo and A. Das, Phys. Lett. B **209**, 311 (1988).
[16]A. Das and S. Okubo, "A systematic study of the Toda lattice," Ann. Phys. (NY) (in press).
[17]M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U. P., Cambridge, 1987), Vol. 2, p. 421.
[18]E.g. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1970).
[19]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[20]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).
[21]W. D. Curtis and F. R. Miller, *Differential Manifolds and Theoretical Physics* (Academic, New York, 1985).
[22]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics; A Modern Approach* (Wiley, New York, 1974).
[23]R. M. Santilli, *Foundations of Theoretical Mechanics, (I) and (II)* (Springer, Berlin, 1978).
[24]A. Nijenhuis, Indag. Math. **13**, 200 (1951).
[25]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I, pp. 37 and 38.
[26]R. D. Shafer, *An Introduction to Nonassociative Algebras* (Academic, New York, 1966).
[27]P. Havas, Nuovo Cimento Suppl. **5**, 363 (1957).
[28]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); **28**, 141 (1978).
[29]D. L. Olive and N. Turok, Nucl. Phys. B **220**, 491 (1983), and references quoted therein.

On the recurrence relations, summations, and integrations of finite rotation matrix elements[a)]

Shan-Tao Lai[b)] and Ying-Nan Chiu[c)]
Center for Molecular Dynamics and Energy Transfer, Department of Chemistry, The Catholic University of America, Washington, DC 20064

(Received 30 August 1988; accepted for publication 30 November 1988)

In this work a number of recurrence relations of rotation matrix elements are derived and a method is proposed to evaluate some useful summations and integrations involving rotation matrix elements. Examples of applications of these in molecular spectroscopy and in chemical physics are also presented.

I. INTRODUCTION

The matrix representations of finite rotations are very useful in a variety of chemical and physical problems and their many properties and derivations have been investigated.[1–4] Because the Wigner rotation matrix elements involve the sum of two binomial coefficients, the calculation involving them is tedious. Although the evaluation and symmetries of finite rotations have been well known,[1–4,5] their summations and integrations have been scarce in literature and textbooks. In this paper, using the differential properties of rotation matrix elements and recurrence relations of Jacobi polynomials,[6] we obtained a number of recurrence relations of rotation matrix elements and using those recurrence relations we are able to easily calculate useful sum formulas and integrals involving the rotation matrix elements.[7,8] In so doing we are able to avoid tedious double summations and obtain the results efficiently.

In Sec. II, we simply give the definition and some useful formulas of $d^j_{m'm}(\beta)$ that are available in standard textbooks on angular momentum.[1,2,5] In Sec. III, we give some differential formulas of $d^j_{m'm}(\beta)$ and several new recurrence relations of the rotation matrix elements. In Sec. IV, we give the relations between the Jacobi polynomials and rotation matrix elements and present our results of recurrence relations. In Sec. V, we give examples of applications of those recurrence relations in chemical physics. Detailed derivations will be found in the Appendices.

II. DEFINITION AND SYMMETRIES OF $d^j_{m'm}(\beta)$

We shall follow the notation of Edmonds.[2] The matrix elements of finite rotations are defined

$$(jm'|D(\alpha\beta\gamma)|jm)\equiv D^j_{m'm}(\alpha\beta\gamma), \tag{1}$$

where the α,β,γ are the three Euler angles. Under the diagonal representations of the matrices of J_z, Eq. (1) becomes

$$D^j_{m'm}(\alpha\beta\gamma)=e^{im'\alpha}d^j_{m'm}(\beta)e^{im\gamma} \tag{2}$$

and[2]

$$d^j_{m'm}(\beta)=\left[\frac{(j+m')!(j-m')!}{(j+m)!(j-m)!}\right]^{1/2}\times\sum_\sigma(-)^{j-m'-\sigma}\binom{j+m}{j-m'-\sigma}\binom{j-m}{\sigma}\times\left(\cos\frac{\beta}{2}\right)^{2\sigma+m'+m}\left(\sin\frac{\beta}{2}\right)^{2j-2\sigma-m'-m}, \tag{3}$$

where $\binom{j}{\sigma}$ are binomial coefficients. The summation over σ is restricted to those values for which the argument of any factorial is non-negative. The $d^j_{m'm}(\beta)$'s have the following symmetries[2]:

$$\begin{aligned}
d^j_{m'm}(\beta)&=(-)^{m'-m}d^j_{mm'}(\beta)\\
&=(-)^{m'-m}d^j_{-m'-m}(\beta)\\
&=(-)^{m'-m}d^j_{m'm}(-\beta),\\
d^j_{m'm}(\pi+\beta)&=(-)^{j-m'}d^j_{-m'm}(\beta),\\
d^j_{m'm}(\pi-\beta)&=(-)^{j-m'}d^j_{m-m'}(\beta),\\
d^j_{m'm}(\pi)&=(-)^{j+m}\delta_{m'+m,0},\\
d^j_{m'm}(-\pi)&=(-)^{j-m}\delta_{m'+m,0},\\
d^j_{m'm}(0)&=\delta_{m'm}.
\end{aligned} \tag{4}$$

III. NEW DIFFERENTIAL PROPERTIES OF $d^j_{m'm}(\beta)$ AND NEW RECURRENCE RELATIONS

Fano and Racah[1] have obtained the following relations (the convention of Fano and Racah[1] is the same as Edmonds,[2] but different from that of Brink and Satchler[5]):

$$[j(j+1)-m'(m'\mp1)]^{1/2}d^j_{m'\mp1m}(\beta)=\frac{m'\cos\beta-m}{\sin\beta}d^j_{m'm}(\beta)\pm\frac{d}{d\beta}d^j_{m'm}(\beta), \tag{5a}$$

$$[j(j+1)-m(m\pm1)]^{1/2}d^j_{m'm\pm1}(\beta)=\frac{m'-m\cos\beta}{\sin\beta}d^j_{m'm}(\beta)\pm\frac{d}{d\beta}d^j_{m'm}(\beta). \tag{5b}$$

Using Eq. (4.1.14) of Edmonds,[2] we obtained

$$\frac{d}{d\beta}d^j_{m'm}(\beta)=\tfrac12[(j-m)(j+m+1)]^{1/2}d^j_{m',m+1}(\beta)-\tfrac12[(j+m)(j-m+1)]^{1/2}d^j_{m'm-1}(\beta) \tag{6}$$

[a)] Taken in part from a Ph.D. dissertation to be submitted by Shan-Tao Lai in partial fulfillment of the Ph.D. degree at the Catholic University of America.

[b)] Permanent address: Department of Chemistry, Xiamen University, Xiamen, People's Republic of China.

[c)] To whom requests for reprints should be addressed.

 0022-2488/89/040844-10$02.50

and

$$\frac{d^2}{d\beta^2} d^j_{m'm}(\beta)$$
$$= -\tfrac{1}{2}(j^2+j-m^2)d^j_{m'm}(\beta)+\tfrac{1}{4}$$
$$\times[(j+m)(j+m-1)(j-m+1)$$
$$\times(j-m+2)]^{1/2}d^j_{m'm-2}(\beta)$$
$$+\tfrac{1}{4}[(j-m)(j-m-1)(j+m+1)$$
$$\times(j+m+2)]^{1/2}d^j_{m'm+2}(\beta). \tag{7}$$

Using Eq. (22.8.1) of the differential property of the Jacobi polynomials,[6] we obtained

$$\frac{d}{d\beta} d^j_{m'm}(\beta) - j\cot\beta d^j_{m'm}(\beta)$$
$$= -\frac{[(j^2-m^2)(j^2-m'^2)]^{1/2}}{j\sin\beta} d^{j-1}_{m'm}(\beta). \tag{8}$$

From Eqs. (5a) and (5b), we obtained the following recurrence relations:

$$[j(j+1)-m'(m'-1)]^{1/2}d^j_{m'-1,m}(\beta)$$
$$+[j(j+1)-m'(m'+1)]^{1/2}d^j_{m'+1,m}(\beta)$$
$$=2[(m'\cos\beta-m)/\sin\beta]d^j_{m'm}(\beta), \tag{9}$$

$$[j(j+1)-m'(m'-1)]^{1/2}d^j_{m'-1,m}(\beta)$$
$$+[j(j+1)-m(m-1)]^{1/2}d^j_{m'm-1}(\beta)$$
$$=(m'-m)\cot(\beta/2)d^j_{m'm}(\beta), \tag{10a}$$

$$[j(j+1)-m'(m'+1)]^{1/2}d^j_{m'+1,m}(\beta)$$
$$+[j(j+1)-m(m+1)]^{1/2}d^j_{m'm+1}(\beta)$$
$$=(m'-m)\cot(\beta/2)d^j_{m'm}(\beta), \tag{10b}$$

and

$$\tfrac{1}{2}[j(j+1)-m'(m'-1)]^{1/2}\{[j(j+1)$$
$$-(m'-1)(m'-2)]^{1/2}d^j_{m'-2,m}(\beta)$$
$$-[j(j+1)-m'(m'-1)]^{1/2}d^j_{m'm}(\beta)\}$$
$$+\tfrac{1}{2}[j(j+1)-m'(m'+1)]^{1/2}\{[j(j+1)$$
$$-m'(m'+1)]^{1/2}d^j_{m'm}(\beta)-[j(j+1)$$
$$-(m'+1)(m'+2)]^{1/2}d^j_{m'+2,m}(\beta)\}$$
$$=2\frac{m\cos\beta-m'}{\sin^2\beta}d^j_{m'm}(\beta)+\frac{m'\cos\beta-m}{\sin\beta}$$
$$\times\{[j(j+1)-m'(m'-1)]^{1/2}d^j_{m'-1,m}(\beta)$$
$$-[j(j+1)-m'(m'+1)]^{1/2}d^j_{m'+1,m}(\beta)\}. \tag{10c}$$

If $m'=0$, then

$$\tfrac{1}{2}[j(j+1)(j+2)(j-1)]^{1/2}[d^j_{\bar{2}m}(\beta)-d^j_{2m}(\beta)]$$
$$=\frac{2m\cos\beta}{\sin^2\beta}d^j_{0m}(\beta)-\frac{m[j(j+1)]^{1/2}}{\sin\beta}$$
$$\times[d^j_{\bar{1}m}(\beta)-d^j_{1m}(\beta)]. \tag{10d}$$

Since $d^j_{m'm}(\beta)$ satisfies the following differential equation[2]:

$$\left\{\frac{d^2}{d\beta^2}+\cot\beta\frac{d}{d\beta}-\frac{m^2+m'^2-2mm'\cos\beta}{\sin^2\beta}+j(j+1)\right\}$$
$$\times d^j_{m'm}(\beta)=0, \tag{10e}$$

substituting Eqs. (7) and (8) into Eq. (10e) we obtained

$$\tfrac{1}{4}[(j+m)(j+m-1)(j-m+1)$$
$$\times(j-m+2)]^{1/2}d^j_{m'm-2}(\beta)$$
$$+\tfrac{1}{4}[(j-m)(j-m-1)(j+m+1)$$
$$\times(j+m+2)]^{1/2}d^j_{m'm+2}(\beta)$$
$$=\left[\frac{m^2+m'^2-2mm'\cos^2\beta-j\cos^2\beta}{\sin^2\beta}\right.$$
$$\left.+\frac{1}{2}(j^2+j-m'^2)-j(j+1)\right]d^j_{m'm}(\beta)$$
$$+\frac{\cos\beta}{j\sin^2\beta}[(j^2-m^2)(j^2-m'^2)]^{1/2}d^{j-1}_{m'm}(\beta). \tag{10f}$$

Using Eqs. (6), (7), and (10e) we obtained

$$\tfrac{1}{4}[(j+m)(j+m-1)(j-m+1)(j-m+2)]^{1/2}$$
$$\times d^j_{m'm-2}(\beta)$$
$$+\tfrac{1}{4}[(j-m)(j-m-1)(j+m+1)$$
$$\times(j+m+2)]^{1/2}d^j_{m'm+2}(\beta)$$
$$=\left[\frac{m^2+m'^2-2mm'\cos\beta}{\sin^2\beta}\right.$$
$$\left.-\frac{1}{2}(j^2+j+m)\right]d^j_{m'm}(\beta)$$
$$+\tfrac{1}{2}\cot\beta[\{(j+m)(j-m+1)\}^{1/2}d^j_{m'm-1}(\beta)$$
$$-\{(j-m)(j+m+1)\}^{1/2}d^j_{m'm+1}(\beta)]. \tag{10g}$$

We will give examples of the application of those new recurrence relations later (Sec. V).

IV. RECURRENCE RELATIONS FROM JACOBI POLYNOMIALS

Edmonds[2] has given the Jacobi polynomial in terms of the rotation matrix elements

$$P^{(m'-m,m'+m)}_{j-m}(\cos\beta)$$
$$=\left[\frac{(j+m)!(j-m)!}{(j+m')!(j-m')!}\right]^{1/2}\left(\cos\frac{\beta}{2}\right)^{-(m'+m)}$$
$$\times\left(\sin\frac{\beta}{2}\right)^{-(m'-m)}d^j_{m'm}(\beta). \tag{11}$$

Several recurrence relations of Jacobi polynomials are available.[6] By use of Eq. (11) and those recurrence relations of Jacobi polynomials, we obtained

$$\left(\frac{j-m}{j-m'}\right)^{1/2}\cos\frac{\beta}{2}d^{j-1/2}_{m'-1/2,m-1/2}(\beta)-\left(\frac{j+m}{j-m'}\right)^{1/2}$$
$$\times\sin\frac{\beta}{2}d^{j-1/2}_{m'-1/2,m+1/2}(\beta)=d^{j-1}_{m'm}(\beta)\quad(j\neq m'), \tag{12a}$$

$$\left(\frac{j-m+1}{j+m'+1}\right)^{1/2}\sin\frac{\beta}{2}d^{j+1/2}_{m'+1/2,m-1/2}(\beta)$$
$$+\left(\frac{j+m+1}{j+m'+1}\right)^{1/2}\cos\frac{\beta}{2}d^{j+1/2}_{m'+1/2,m+1/2}(\beta)$$
$$=d^j_{m'm}(\beta), \tag{12b}$$

$$\frac{2(j+1)}{(j+m'+1)^{1/2}}\sin\frac{\beta}{2}\,d^{j+1/2}_{m'+1/2,m-1/2}(\beta)$$
$$=(j-m+1)^{1/2}d^{j}_{m'm}(\beta)$$
$$-\left[\frac{(j-m'+1)(j+m+1)}{j+m'+1}\right]^{1/2}d^{j+1}_{m'm}(\beta),\tag{12c}$$

$$\frac{2(j+1)}{(j+m'+1)^{1/2}}\cos\frac{\beta}{2}\,d^{j+1/2}_{m'+1/2,m+1/2}(\beta)$$
$$=(j+m+1)^{1/2}d^{j}_{m'm}(\beta)$$
$$+\left[\frac{(j-m'+1)(j-m+1)}{j+m'+1}\right]^{1/2}d^{j+1}_{m'm}(\beta),\tag{12d}$$

$$\frac{2j}{(j-m')^{1/2}}\sin\frac{\beta}{2}\,d^{j-1/2}_{m'-1/2,m+1/2}(\beta)$$
$$=\left[\frac{(j-m)(j+m')}{j-m'}\right]^{1/2}d^{j}_{m'm}(\beta)$$
$$-(j+m)^{1/2}d^{j-1}_{m'm}(\beta)\quad(j\neq m'),\tag{12e}$$

$$\frac{2j}{(j-m')^{1/2}}\cos\frac{\beta}{2}\,d^{j-1/2}_{m'-1/2,m-1/2}(\beta)$$
$$=\left[\frac{(j+m')(j+m)}{j-m'}\right]^{1/2}d^{j}_{m'm}(\beta)$$
$$+(j-m)^{1/2}d^{j-1}_{m'm}(\beta)\quad(j\neq m'),\tag{12f}$$

and

$$j\{[(j+1)^2-m'^2][(j+1)^2-m^2]\}^{1/2}d^{j+1}_{m'm}(\beta)$$
$$=[-m'm(2j+1)$$
$$+j(j+1)(2j+1)\cos\beta\,]d^{j}_{m'm}(\beta)$$
$$-(j+1)[(j^2-m^2)(j^2-m'^2)]^{1/2}d^{j-1}_{m'm}(\beta).\tag{12g}$$

From Eq. (3), it is easy to obtain the following:

$$d^{j+m}_{mm}(\beta)=\sum_\sigma(-)^{j-\sigma}\frac{j!(j+2m)!}{\sigma!(j-\sigma)!(2m+\sigma)!(j-\sigma)!}\times\left(\cos\frac{\beta}{2}\right)^{2\sigma+2m}\left(\sin\frac{\beta}{2}\right)^{2j-2\sigma},$$

$$d^{j}_{j-1,m}(\beta)=\left[\frac{(2j-1)!}{(j+m)!(j-m)!}\right]^{1/2}\times\left[(j-m)\cos^2\frac{\beta}{2}-(j+m)\sin^2\frac{\beta}{2}\right]\times\left(\cos\frac{\beta}{2}\right)^{j+m-1}\left(\sin\frac{\beta}{2}\right)^{j-m-1},$$

and

$$d^{j}_{j-2,m}(\beta)=\left[\frac{(2j-2)!}{(j+m)!(j-m)!}\right]^{1/2}\times\frac{1}{2}\left[(j+m)(j+m-1)\left(\sin\frac{\beta}{2}\right)^4-2(j+m)(j-m)\left(\cos\frac{\beta}{2}\right)^2\left(\sin\frac{\beta}{2}\right)^2+(j-m)(j-m-1)\left(\cos\frac{\beta}{2}\right)^4\right]\times\left(\cos\frac{\beta}{2}\right)^{j+m-2}\left(\sin\frac{\beta}{2}\right)^{j-m-2},$$

and

$$d^{0}_{00}(\beta)=1\quad\text{and}\quad d^{1}_{00}(\beta)=\cos\beta,$$

using the above two values of the rotation matrices as a starting point and Eqs. (12a) and (12c), we obtained as a check

$$d^{1/2}_{m'm}(\beta)=\begin{pmatrix}\cos(\beta/2) & \sin(\beta/2)\\ -\sin(\beta/2) & \cos(\beta/2)\end{pmatrix},$$

which agrees with standard textbook results.[1,2] In principle, we can compute $d^{j}_{m'm}(\beta)$ of any order by using the above recurrence relations.

V. EXAMPLES OF EVALUATIONS OF INTEGRALS AND SUMMATIONS INVOLVING $d^{j}_{m'm}(\beta)$

A. Example (1)

In discussing mechanisms for collision-induced transitions between Λ doublets in $^1\Pi$ molecules, Green and Zare[8a] for the first time have computed the following integral but omitted detailed proof:

$$S_{1-1}(J,M)=\frac{2J+1}{4\pi}\int_0^{2\pi}d\alpha\int_0^{\pi}D^{J}_{M1}(\alpha\beta 0)\times D^{J*}_{M-1}(\alpha\beta 0)\sin\beta\,d\beta$$
$$=\left[\frac{(J+M)!(J-M)!}{(J+1)!(J-1)!}\right]\sum_S\sum_T(-)^{S+T}\times\binom{J+1}{J-M-S}\binom{J-1}{S}\times\binom{J-1}{J-M-T}\binom{J+1}{T}\left[\binom{2J}{M+S+T}\right]^{-1}$$
$$=-1+\frac{|M|(2J+1)}{J(J+1)}.$$

The summation in the last equality is not easy to visualize. We shall rederive it from the recurrence relations and show that it is a special case of a general derivation. We use Eq. (9) and let $m'=0$, then we have

$$[j(j+1)]^{1/2}[d^{j}_{\bar{1}m}(\beta)+d^{j}_{1m}(\beta)]=-\frac{2m}{\sin\beta}d^{j}_{0m}(\beta).$$

Introducing Eq. (4), squaring the above equation on both sides, and integrating both sides, we obtained

$$j(j+1)\left[\int_0^{\pi}(d^{j}_{m\bar{1}}(\beta))^2\sin\beta\,d\beta+\int_0^{\pi}(d^{j}_{m1}(\beta))^2\sin\beta\,d\beta+2\int_0^{\pi}d^{j}_{m1}(\beta)d^{j}_{m\bar{1}}(\beta)\sin\beta\,d\beta\right]$$
$$=4m^2\int_0^{\pi}\left(\frac{d^{j}_{m0}}{\sin\beta}\right)^2\sin\beta\,d\beta.\tag{13}$$

Since the orthonormality of $D^{j}_{m'm}(\alpha\beta\gamma)$ with respect to integrations over the Euler angles is[2]

$$\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} D^{j_1^*}_{m_1'm_1}(\alpha\beta\gamma)D^{j_2}_{m_2'm_2}(\alpha\beta\gamma)d\alpha\,\sin\beta\,d\beta\,d\gamma$$

$$=\delta_{m_1'm_2'}\delta_{m_1m_2}\delta_{j_1j_2}\frac{1}{2j_1+1}, \tag{14}$$

the orthonormality of the associated Legendre function is[2]

$$\int_{-1}^{1} P_j^m(x)\,P_j^n(x)\,\frac{dx}{1-x^2}=\frac{\delta_{mn}(j+m)!}{m(j-m)!}, \tag{15}$$

and since[2]

$$d^j_{m0}(\beta)=\left[\frac{(j-m)!}{(j+m)!}\right]^{1/2}P_j^m(\cos\beta), \tag{16}$$

using Eqs. (14), (15), and (16), we have

$$\begin{aligned}j(j+1)&\left[\frac{2}{2j+1}+\frac{2}{2j+1}\right.\\&\left.+2\int_0^{\pi} d^j_{m1}(\beta)d^j_{m\bar{1}}(\beta)\sin\beta\,d\beta\right]\\&=\frac{4m^2(j-m)!}{(j+m)!}\int_{-1}^{1}\frac{(P_j^m(x))^2}{1-x^2}dx\\&=4|m|.\end{aligned} \tag{17}$$

Upon rearrangement, Eq. (17) becomes

$$\int_0^{\pi} d^j_{m1}(\beta)d^j_{m\bar{1}}(\beta)\sin\beta\,d\beta=\frac{2|m|}{j(j+1)}-\frac{1}{2j+1}. \tag{18}$$

Changing j into J and m into M, we obtain Zare and Green's[8a] formulas

$$\begin{aligned}S_{1\bar{1}}(J,M)&=\frac{2J+1}{2}\int_0^{\pi} d^J_{M1}(\beta)d^J_{M\bar{1}}(\beta)\sin\beta\,d\beta\\&=\frac{|M|(2J+1)}{J(J+1)}-1.\end{aligned}$$

The generalized derivation of integrations similar to the above are given in the Appendices.

B. Example (2): Derivation of the summation $\Sigma_m m^n|Y_{jm}(\beta\alpha)|^2$

Case of n=1. We multiply $d^j_{m'm}(\beta)$ into both sides of Eq. (9), then sum over m

$$\begin{aligned}&[j(j+1)-m'(m'-1)]^{1/2}\sum_m d^j_{m'-1,m}(\beta)d^j_{m'm}(\beta)\\&\quad+[j(j+1)-m'(m'+1)]^{1/2}\\&\quad\times\sum_m d^j_{m'+1m}(\beta)d^j_{m'm}(\beta)\\&=2\sum_m\frac{m'\cos\beta-m}{\sin\beta}|d^j_{m'm}(\beta)|^2.\end{aligned} \tag{19}$$

Since the orthonormality of the unitary matrix elements is

$$\sum_m d^j_{m'm}(\beta)d^j_{m''m}(\beta)=\delta_{m'm''}, \tag{20}$$

Eq. (19) becomes

$$0=2\sum_m\frac{m'\cos\beta-m}{\sin\beta}|d^j_{m'm}(\beta)|^2,$$

that is,[6]

$$\sum_m m|d^j_{m'm}(\beta)|^2=m'\cos\beta \tag{21}$$

When $m'=0$, Eq. (21) reduces to

$$\sum_m m|Y_{jm}(\beta\alpha)|^2=0. \tag{21a}$$

Case of n=2. Squaring Eq. (9) on both sides yields

$$\begin{aligned}&[j(j+1)-m'(m'-1)]|d^j_{m'-1m}(\beta)|^2\\&\quad+[j(j+1)-m'(m'+1)]|d^j_{m'+1m}(\beta)|^2\\&\quad+2[j(j+1)-m'(m'-1)]^{1/2}\\&\quad\times[j(j+1)-m'(m'+1)]^{1/2}\\&\quad\times d^j_{m'-1m}(\beta)d^j_{m'+1m}(\beta)\\&=4\frac{|d^j_{m'm}(\beta)|^2}{\sin^2\beta}(m'\cos\beta-m)^2.\end{aligned} \tag{22}$$

Summing over m on both sides of Eq. (22) and using Eqs. (20) and (21), we obtained[7]

$$\sum_m m^2|d^j_{m'm}(\beta)|^2=\frac{j(j+1)}{2}\sin^2\beta+\frac{m'^2}{2}(3\cos^2\beta-1). \tag{23}$$

Since[2]

$$\begin{aligned}D^j_{m0}(\alpha\beta 0)&=(-)^m\left(\frac{4\pi}{2j+1}\right)^{1/2}Y_{jm}(\beta\alpha)\\&=e^{im\alpha}d^j_{m0}(\beta),\end{aligned}$$

when $m'=0$ then Eq. (23) can be rewritten as[8b]

$$\sum_m m^2|Y_{jm}(\beta\alpha)|^2=\frac{1}{8\pi}j(j+1)(2j+1)\sin^2\beta. \tag{24}$$

Using Eqs. (12g) and (23), we obtained

$$\begin{aligned}&\sum_m[(j+m+1)(j+m)(j-m+1)(j-m)]^{1/2}d^{j+1}_{m'm}(\beta)d^{j-1}_{m'm}(\beta)\\&=\frac{[j(j+1)-m'(m'-1)][j(j+1)-m'(m'+1)]}{2[(j+m'+1)(j+m')(j-m'+1)(j-m')]^{1/2}}(3\cos^2\beta-1).\end{aligned} \tag{24a}$$

Let $m'=0$, we have

$$\begin{aligned}&\sum_m[(j+m+1)(j+m)(j-m+1)(j-m)]^{1/2}\\&\quad\times d^{j+1}_{m0}(\beta)d^{j-1}_{m0}(\beta)\\&=\frac{j(j+1)}{2}(3\cos^2\beta-1).\end{aligned} \tag{24b}$$

Similarly, Eq. (23a) can be written as[8b]

$$\begin{aligned}&\sum_m\left[\frac{(j+m+1)(j+m)(j-m+1)(j-m)}{(2j+1)(2j+3)(2j-1)(2j+1)}\right]^{1/2}\\&\quad\times Y^*_{j+1m}(\beta\alpha)Y_{j-1m}(\beta\alpha)\\&=\frac{j(j+1)}{8\pi(2j+1)}(3\cos^2\beta-1).\end{aligned} \tag{24c}$$

C. Example (3). General derivation of summations $\Sigma_m m^n d^j_{m'm}(\beta) d^j_{m'm}(\beta)$

We have also obtained the following general summation formulas over $d^j_{m'm}(\beta)$, some of which were used in Jacobs and Zare's work[9]:

$$\sum_m m d^j_{m'm} d^j_{m'+1m} = -\frac{\sin\beta}{2}[j(j+1)-m'(m'+1)]^{1/2}, \quad (25)$$

$$\sum_m m d^j_{m'm} d^j_{m'-1m} = -\frac{\sin\beta}{2}[j(j+1)-m'(m'-1)]^{1/2}, \quad (26)$$

$$\sum_m m d^j_{m'-1m} d^j_{m'+1m} = 0, \quad (27)$$

$$\sum_m m^2 d^j_{m'+1m} d^j_{m'm} = -\cos\beta\left[m' + \frac{\sin\beta}{2}(m'+1)\right] \times [j(j+1)-m'(m'+1)]^{1/2}, \quad (28)$$

$$\sum_m m^2 d^j_{m'+1m} d^j_{m'-1m} = \frac{\sin^2\beta}{4}[j(j+1)-m'(m'+1)]^{1/2} \times [j(j+1)-m'(m'-1)]^{1/2}, \quad (29)$$

$$\sum_m m d^j_{m'-2,m} d^j_{m'+1,m} = 0, \quad (30)$$

$$\sum_m m^3 |d^j_{m'm}|^2 = \frac{m'}{2}[3j(j+1)-5m'^2-1] \times \sin^2\beta\cos\beta + m'^3\cos\beta, \quad (31)$$

$$\sum_m m^4|d^j_{m'm}|^2 = \frac{3}{8}\sin^4\beta[j(j+1)]^2 + \frac{\sin^2\beta}{4}j(j+1)(3\cos^2\beta-1) + \frac{m'^2}{8}\sin^2\beta(5\cos^2\beta-1)(6j^2+6j-5) + \frac{m'^4}{8}(-30\cos^2\beta+35\cos^4\beta+3). \quad (32)$$

Here we have abbreviated $d^j_{m'm}(\beta)$ as $d^j_{m'm}$. All values of j (integer or half-integer) satisfy all of the above formulas.

D. Example (4): Computation of ground state alignment distribution

Jacobs and Zare[9] have given the quantum mechanical treatment of ground-state population and alignment. They obtained the following ground-state alignment distribution in terms of monopole $A^{(0)}_0$, quadrupole $A^{(2)}_0$, hexadecapole $A^{(4)}_0$ moments, and in terms of summation over $M' = -J$ to $+J$:

$$N_0(J,M,\theta',t=0) = A^{(0)}_0 + A^{(2)}_0\left[-1+\frac{3}{J(J+1)}\sum_{M'}|d^J_{M'M}(\theta')|^2 M'^2\right] + A^{(4)}_0\left\{\frac{3}{8}-\frac{3}{4J(J+1)}-\frac{5}{8J^2(J+1)^2}\sum_{M'}|d^J_{M'M}(\theta')|^2[6J(J+1)M'^2+5M'^2+7M'^4]\right\}. \quad (33a)$$

Using Eqs. (23) and (32), we obtained the following closed form that achieves the actual summation over M':

$$\begin{aligned}N_0(J,M,\theta',t=0) = {} & A^{(0)}_0 + A^{(2)}_0\left[-1+\frac{3}{2}\sin^2\theta' + \frac{3M^2}{2J(J+1)}(3\cos^2\theta'-1)\right] \\ & + A^{(4)}_0\left\{\frac{3}{8}-\frac{3}{4J(J+1)}-\frac{15}{8}\sin^2\theta' - \frac{25\sin^2\theta'}{16J(J+1)} - \frac{105\sin^4\theta'}{64} - \frac{35\sin^2\theta'}{32J(J+1)}(3\cos^2\theta'-1)\right. \\ & + \frac{5}{64J^2(J+1)^2}[35M^2\sin^2\theta'(5\cos^2\theta'-1) - 24M^2J(J+1)(3\cos^2\theta'-1) \\ & - 20M^2(3\cos^2\theta'-1) - 42M^2J(J+1)\sin^2\theta'(5\cos^2\theta'-1) \\ & \left. - 7M^4(35\cos^4\theta'-30\cos^2\theta'+3)]\right\}. \end{aligned} \quad (33b)$$

Considering the hyperfine depolarization and magnetic precession effect and the actually randomize ground-state alignment, Jacobs and Zare[9] obtained

$$\begin{aligned}N_0(J,M,\theta',\theta'',t=0) = {} & A^{(0)}_0 + A^{(2)}_0\bar{g}^{(2)}\left[-1+\frac{3}{J(J+1)}\sum_{M''}\sum_{M'}|d^J_{M''M}(\theta')d^J_{M'M''}(\theta'')|^2 M'^2\right] \\ & + A^{(4)}_0\bar{g}^{(4)}\left\{\frac{3}{8}-\frac{3}{4J(J+1)}-\frac{5}{8J^2(J+1)^2}\right. \\ & \left. \times\sum_{M''}\sum_{M'}|d^J_{M''M}(\theta')d^J_{M'M''}(\theta'')|^2[6J(J+1)M'^2+5M'^2+7M'^4]\right\}, \end{aligned} \quad (33c)$$

where θ'' is the angle between the cylindrical symmetry axis and the magnetic field vector and the plane of polarization, and[9]

$$\bar{g}^{(K)} = \sum_i \frac{(2F_i+1)^2}{2I+1} \begin{Bmatrix} F_i & F_i & K \\ J & J & I \end{Bmatrix}^2 .$$

Using Eqs. (23), (32), and (33b), we have obtained the following closed form that achieves the actual summations over M' and M'' not given by Jacobs and Zare[9]:

$$\begin{aligned}
&N_0(JM,\theta',\theta'',t=0)\\
&= A_0^{(0)} + A_0^{(2)}\bar{g}^{(2)} \left\{ -1 + \frac{3}{2}\sin^2\theta'' + \frac{3}{4}\sin^2\theta'(3\cos^2\theta''-1) + \frac{3M^2}{4J(J+1)}(3\cos^2\theta'-1)(3\cos^2\theta''-1) \right\}\\
&\quad + A_0^{(4)}\bar{g}^{(4)} \left\{ \frac{3}{8} - \frac{3}{4J(J+1)} - \frac{15}{8}\sin^2\theta'' - \frac{25\sin^2\theta''}{16J(J+1)} - \frac{105\sin^4\theta''}{64} - \frac{35\sin^2\theta''}{32J(J+1)}(3\cos^2\theta''-1) \right.\\
&\quad + \frac{5}{64J^2(J+1)^2}\left[\frac{\sin^2\theta'}{2}J(J+1) + \frac{M^2}{2}(3\cos^2\theta'-1)\right]\cdot[35\sin^2\theta''(5\cos^2\theta''-1)\\
&\quad - 24J(J+1)(3\cos^2\theta''-1) - 20(3\cos^2\theta''-1) - 42J(J+1)\sin^2\theta''(5\cos^2\theta''-1)]\\
&\quad - \frac{105\sin^4\theta'}{512}(35\cos^4\theta''-30\cos^2\theta''+3) - \frac{35\sin^2\theta'}{256J(J+1)}(3\cos^2\theta'-1)(35\cos^4\theta''-30\cos^2\theta''+3)\\
&\quad \left. + \frac{5}{512J^2(J+1)^2}[M^2\sin^2\theta'(5\cos^2\theta'-1)(6J^2+6J-5) + M^4(35\cos^4\theta'-30\cos^2\theta'+3)] \right\}. \qquad (34)
\end{aligned}$$

E. Example (5): Integration over trigonometric functions, the associated Legendre polynomials, and rotation matrices

We shall first give some preliminaries. Using recurrence formulas of Eq. (9) and Eqs. (12a)–(12g), we obtained some special products involving rotation matrix elements expressed in terms of the associated Legendre polynomials,

$$\sin(\beta/2)\, d^j_{(1/2)(\bar{1}/2)} = \tfrac{1}{2}(P_{j-1/2}(\cos\beta) - P_{j+1/2}(\cos\beta)), \qquad (35a)$$

$$\cos(\beta/2)\, d^j_{(1/2)(1/2)} = \tfrac{1}{2}(P_{j-1/2}(\cos\beta) + P_{j+1/2}(\cos\beta)), \qquad (35b)$$

$$\sin\beta\, d^j_{01} = \frac{[j(j+1)]^{1/2}}{2j+1}(P_{j+1}(\cos\beta) - P_{j-1}(\cos\beta)), \qquad (35c)$$

$$\begin{aligned}
&(1+\cos\beta)\, d^j_{11}\\
&\quad = (P_{j-1}(\cos\beta) + P_j(\cos\beta))\\
&\qquad + \frac{j}{2j+1}(P_{j+1}(\cos\beta) - P_{j-1}(\cos\beta)), \qquad (35d)
\end{aligned}$$

$$\begin{aligned}
&(1-\cos\beta)\, d^j_{1\bar{1}}\\
&\quad = (P_{j-1}(\cos\beta) - P_j(\cos\beta))\\
&\qquad + \frac{j}{2j+1}(P_{j+1}(\cos\beta) - P_{j-1}(\cos\beta)), \qquad (35e)
\end{aligned}$$

$$\begin{aligned}
&\sin\beta\, d^j_{m(1/2)}\\
&\quad = \left[\frac{(j-m)!}{(j+\frac{1}{2})(j+m)!}\right]^{1/2}\\
&\qquad \times\cos\frac{\beta}{2}(P^{m+1/2}_{j+1/2}(\cos\beta) - P^{m+1/2}_{j-1/2}(\cos\beta)), \qquad (35f)
\end{aligned}$$

$$\begin{aligned}
&\sin\beta\, d^j_{m(\bar{1}/2)}\\
&\quad = \left[\frac{(j-m)!}{(j+\frac{1}{2})(j+m)!}\right]^{1/2}\\
&\qquad \times\cos\frac{\beta}{2}[(j+m)P^{m-1/2}_{j-1/2}(\cos\beta)\\
&\qquad - (j-m+1)P^{m-1/2}_{j+1/2}(\cos\beta)], \qquad (35g)
\end{aligned}$$

$$\begin{aligned}
&\sin\beta\, d^j_{m(3/2)}\\
&\quad = \left[\frac{(j-m)!}{(j+\frac{1}{2})(j^2+j-\frac{3}{4})(j+m)!}\right]^{1/2}\\
&\qquad \times\cos\frac{\beta}{2}\left\{\frac{(2m-\cos\beta)}{\sin\beta}(P^{m+1/2}_{j+1/2}(\cos\beta)\right.\\
&\qquad - P^{m+1/2}_{j-1/2}(\cos\beta)) - (j+\tfrac{1}{2})\\
&\qquad \times[(j+m)P^{m-1/2}_{j-1/2}(\cos\beta)\\
&\qquad \left. - (j-m+1)P^{m-1/2}_{j+1/2}(\cos\beta)]\right\}, \qquad (35h)
\end{aligned}$$

$$\begin{aligned}
&\sin\beta\, d^j_{m(\bar{3}/2)}\\
&\quad = \left[\frac{(j-m)!}{(j+\frac{1}{2})(j^2+j-\frac{3}{4})(j+m)!}\right]^{1/2}\\
&\qquad \times\cos\frac{\beta}{2}\left\{\frac{(2m+\cos\beta)}{\sin\beta}\right.\\
&\qquad \times[(j+m)P^{m-1/2}_{j-1/2}(\cos\beta)\\
&\qquad - (j-m+1)P^{m-1/2}_{j+1/2}(\cos\beta)]\\
&\qquad \left. - (j+\tfrac{1}{2})[P^{m+1/2}_{j+1/2}(\cos\beta) - P^{m+1/2}_{j-1/2}(\cos\beta)]\right\}, \qquad (35i)
\end{aligned}$$

$$\begin{aligned}
\sin\beta\, d^j_{m1} &= \left[\frac{(j-m)!}{j(j+1)(j+m)!}\right]^{1/2}\\
&\quad \times\{[m-(j+1)\cos\beta]P^m_j(\cos\beta)\\
&\quad + (j-m+1)P^m_{j+1}(\cos\beta)\}, \qquad (35j)
\end{aligned}$$

$$\begin{aligned}
\sin\beta\, d^j_{m\bar{1}} &= \left[\frac{(j-m)!}{j(j+1)(j+m)!}\right]^{1/2}\\
&\quad \times\{[m+(j+1)\cos\beta]P^m_j(\cos\beta)\\
&\quad - (j-m+1)P^m_{j+1}(\cos\beta)\}, \qquad (35k)
\end{aligned}$$

$$d^j_{m2}=\left[\frac{(j-m)!}{j(j+1)(j^2+j-2)(j+m)!}\right]^{1/2}$$
$$\times\left\{\left[2\,\frac{m-\cos\beta}{\sin^2\beta}\,(m-\ (j+1)\cos\beta)-j(j+1)\right]\right.$$
$$\times P_j^m(\cos\beta)$$
$$\left.+\frac{2(m-\cos\beta)(j-m+1)}{\sin^2\beta}\,P^m_{j+1}(\cos\beta)\right\}. \tag{35l}$$

$$d^j_{m\bar{2}}=\left[\frac{(j-m)!}{j(j+1)(j^2+j-2)(j+m)!}\right]^{1/2}$$
$$\times\left\{\left[\frac{2(m+\cos\beta)}{\sin^2\beta}\,(m+(j+1)\cos\beta)\right.\right.$$
$$\left.-j(j+1)\right]P_j^m(\cos\beta)$$
$$\left.-\frac{2(m+\cos\beta)}{\sin^2\beta}\,(j-m+1)P^m_{j+1}(\cos\beta)\right\}. \tag{35m}$$

Similarly we obtained the following recurrence relations of the associated Legendre polynomials:

$$\cos\beta P_j^m(\cos\beta)$$
$$=\frac{\sin\beta}{2m}\left[(j+m)(j-m+1)P_j^{m-1}(\cos\beta)\right.$$
$$\left.+P_j^{m+1}(\cos\beta)\right], \tag{36}$$

$$P^m_{j-1}(\cos\beta)$$
$$=\frac{\sin\beta}{2m}\left[(j-m)(j-m+1)P_j^{m-1}(\cos\beta)\right.$$
$$\left.+P_j^{m+1}(\cos\beta)\right]. \tag{37}$$

Since the orthonormality of the associated Legendre polynomials is

$$\int_0^\pi P_j^m(\cos\beta)\,P_k^m(\cos\beta)\sin\beta\,d\beta=\frac{2\delta_{jk}(j+m)!}{(2j+1)\cdot(j-m)!}, \tag{38}$$

using Eqs. (15), (36), (37), (35j), and (35k), we obtained the following results of integrations:

$$\int_0^\pi\frac{(P_j^m(\cos\beta))^2}{\sin^4\beta}\sin\beta\,d\beta$$
$$=\frac{(j+m)!(j^2+m^2+j-1)}{2m(j-m)!(m^2-1)}\quad(m\neq\pm1), \tag{39a}$$

$$\int_0^\pi\frac{\cos\beta}{\sin^4\beta}\,(P_j^m(\cos\beta))^2\sin\beta\,d\beta=0, \tag{39b}$$

$$\int_0^\pi\frac{\cos^2\beta}{\sin^4\beta}\,(P_j^m)^2\sin\beta\,d\beta$$
$$=\frac{(j+m)!(j^2-m^2+j+1)}{2m(j-m)!(m^2-1)}\quad(m\neq\pm1), \tag{39c}$$

$$\int_0^\pi\frac{\cos\beta}{\sin^4\beta}\,P_j^mP^m_{j+1}\sin\beta\,d\beta$$
$$=\frac{(j+m+1)!(j+1)}{2m(j-m)!(m^2-1)}\quad(m\neq0,\pm1), \tag{39d}$$

$$\int_0^\pi\frac{\cos^2\beta}{\sin^4\beta}\,P_j^mP^m_{j+1}\sin\beta\,d\beta=0, \tag{39e}$$

$$\int_0^\pi P^{m-1}_{j+1}P^{m+1}_{j+1}\sin\beta\,d\beta$$
$$=\frac{2(j+m)!}{(j-m+2)!}\left\{m-\frac{j^2+m^2+3j+2}{2j+3}\right\}, \tag{39f}$$

$$\int_0^\pi\frac{1}{\sin^4\beta}\,P^{m-1}_{j+1}P^{m+1}_{j+1}\sin\beta\,d\beta$$
$$=\frac{(j+m+2)!}{4m(m^2-1)(j-m)!}\quad(m\neq0,\pm1), \tag{39g}$$

$$\int_0^\pi\frac{\cos^3\beta}{\sin^4\beta}\,P_j^mP^m_{j+1}\sin\beta\,d\beta$$
$$=\frac{1}{2m}\,\frac{(j+m)!}{(j-m)!}\left[\frac{(j+1)(j+m+1)}{m^2-1}\right.$$
$$\left.-\frac{2(2j^2-2jm+5j-3m+3)}{(2j+3)(j-m+1)}\right], \tag{39h}$$

$$\int_0^\pi\frac{(d^j_{\bar{1}m})^2}{\sin^2\beta}\sin\beta\,d\beta$$
$$=\int_0^\pi\frac{(d^j_{1m})^2}{\sin^2\beta}\sin\beta\,d\beta=\frac{m}{m^2-1}\quad(m\neq\pm1), \tag{39i}$$

$$\int_0^\pi\frac{(P_j^m)^2}{\sin^6\beta}\sin\beta\,d\beta=\frac{1}{8m^3}\,\frac{(j+m)!}{(j-m)!}\left\{\frac{(j-m+2)(j-m+1)}{(m-2)(m-1)}\,(j^2+3j+m^2-2m+2)\right.$$
$$+\frac{(j+m+2)(j+m+1)(j^2+3j+m^2+2m+2)}{(m+2)(m+1)}$$
$$\left.+\frac{(j-m+2)(j-m+1)(j+m+2)(j+m+1)}{m^2-1}\right\}\quad(m\neq\pm1), \tag{39j}$$

$$\int_0^\pi\frac{\cos\beta}{\sin^6\beta}\,P_j^mP^m_{j+1}\sin\beta\,d\beta=\frac{1}{8m^3}\,\frac{(j+m+1)!}{(j-m)!}\left\{\frac{(j-m+2)(j^2+3j+m^2-2m+2)}{(m-1)(m-2)}\right.$$
$$+\frac{(j+m+2)(j^2+3j+m^2+2m+2)}{(m+1)(m+2)}$$
$$\left.+\frac{(j-m+2)(j+m+2)(j+1)}{m^2-1}\right\}\quad(m\neq\pm1), \tag{39k}$$

$$\int_0^\pi \frac{\cos^2\beta}{\sin^6\beta}(P_{j+1}^m)^2 \sin\beta\, d\beta = \frac{1}{8m^3}\frac{(j+m+1)!}{(j-m+1)!}\left\{\frac{(j+m+1)(j-m+2)(j^2+3j+m^2-2m+2)}{(m-1)(m-2)} + \frac{(j+m+2)(j-m+1)(j^2+3j+m^2+2m+2)}{(m+1)(m+2)} + \frac{(j+m+1)(j+m+2)(j-m+2)(j-m+1)}{m^2-1}\right\}. \tag{391}$$

F. Example (6): General formula for

$$S_{\Omega\Omega'}(J,M) = \frac{2J+1}{4\pi}\int_0^{2\pi} d\alpha \int_0^\pi D^J_{M\Omega}(\alpha\beta 0) \times D^{J*}_{M\Omega'}(\alpha\beta 0)\sin\beta\, d\beta.$$

Special formulas are useful in diatomic molecular spectroscopy and molecular dynamics. For example, Alexander and Dagdigian and their co-workers[10,11] have obtained

$$S_{(1/2)(\bar{3}/2)}(J,J) = \left[\frac{2J-1}{2J+3}\right]^{1/2} \tag{40}$$

and

$$S_{2\bar{2}}(J,0) = +1. \tag{41}$$

We shall show that these can be obtained from the general formula that we shall derive as follows.

First we consider

$$S_{(1/2)(\bar{3}/2)}(J,M) = \frac{2J+1}{4\pi}\int_0^{2\pi} d\alpha \int_0^\pi D^J_{M(1/2)}(\alpha\beta 0) \times D^{J*}_{M(\bar{3}/2)}(\alpha\beta 0)\sin\beta\, d\beta,$$

which we obtained (see proof in Appendix A) as

$$S_{(1/2)(\bar{3}/2)}(J,M) = \frac{4M-(2J+1)}{[(2J+3)(2J-1)]^{1/2}}. \tag{42}$$

For $M = J$, Eq. (42) reduced to Eq. (40).[10]

G. Example (7). Specific formula $S_{2\bar{2}}(J,M)$ and $S_{3\bar{3}}(J,M)$ for arbitrary M

We have obtained (see proof in Appendix B)

$$S_{2\bar{2}}(J,M) = \frac{2J+1}{4\pi}\int_0^{2\pi} d\alpha \int_0^\pi D^J_{M2}(\alpha\beta 0) D^{J*}_{M\bar{2}}(\alpha\beta 0)\sin\beta\, d\beta = 1 - \frac{2M(2J+1)}{J(J+1)} + \frac{2M(2J+1)(M^2-1)}{J(J+1)(J^2+J-2)}. \tag{43}$$

For $M = 0$, the above reduces to Alexander's formula Eq. (41). Similarly, we have obtained (see proof in Appendix C)

$$\begin{aligned} S_{3\bar{3}}(J,M) &= \frac{2J+1}{4\pi}\int_0^{2\pi} d\alpha \int_0^\pi D^J_{M3}(\alpha\beta 0) D^{J*}_{M\bar{3}}(\alpha\beta 0)\sin\beta\, d\beta \\ &= -1 - \frac{3M(2J+1)(J^2+13J+4M^2+14)}{J(J+1)(J-1)(J-2)} \\ &\quad + \frac{2M(2J+1)}{J(J+1)(J+2)(J+3)(J-1)(J-2)} \\ &\quad \times [2(2J+5)(3J^2+13J+3M^2+12) \\ &\quad + 3(J^2+3J+M^2+2)^2 \\ &\quad - M^2(2J+3)(2J-1)]. \end{aligned} \tag{44}$$

H. Example (8). Angular distribution of photodissociation

The cross section for angular distribution of photofragments has been given[12,13] as

$$\sigma(iJ'' \to eJ' \to \hat{k}) \sim \sum_\Omega A_{eJ'}(\Omega) I_{J'J''\Omega}(\theta), \tag{45a}$$

where

$$I_{J'J''\Omega}(\theta) = \sum_{M''}(2J'+1) \times \begin{pmatrix} J'' & 1 & J' \\ M'' & 0 & \bar{M}'' \end{pmatrix}^2 |D^J_{M''\Omega}(\phi,\theta,0)|^2. \tag{45b}$$

Here $\left(\begin{smallmatrix}\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot\end{smallmatrix}\right)$ is Wigner 3-j symbol. Using Eq. (23), the summation in this formula can be evaluated easily as follows. For R lines $(J' = J'' + 1)$,

$$\begin{aligned} I_{J',J'-1\Omega}(\theta) &= \sum_{M''}(2J'+1)\frac{J'^2 - M''^2}{J'(2J'-1)(2J'+1)}|d^{J'}_{M''\Omega}(\theta)|^2 \\ &= \frac{J'^2}{J'(2J'-1)} - \frac{1}{J'(2J'-1)} \times \left[\frac{\sin^2\theta}{2}J'(J'+1) + \frac{\Omega^2}{2}(3\cos^2\theta - 1)\right] \\ &= \frac{1}{J'(2J'-1)}\left\{(J'^2-\Omega^2)\cos^2\theta + \frac{1}{2}[J'(J'-1)+\Omega^2]\sin^2\theta\right\}, \end{aligned} \tag{46}$$

for Q lines $(J' = J'')$,

$$\begin{aligned} I_{J',J'\Omega}(\theta) &= \sum_{M''}(2J'+1)\frac{M''^2}{J'(J'+1)(2J'+1)}|d^{J'}_{M''\Omega}(\theta)|^2 \\ &= \frac{1}{J'(J'+1)}\left\{\Omega^2\cos^2\theta + \frac{1}{2}[J'(J'+1)-\Omega^2]\sin^2\theta\right\}, \end{aligned} \tag{47}$$

and for P lines $(J' = J'' - 1)$,

$$\begin{aligned} I_{J'J'+1\Omega}(\theta) &= \sum_{M''}\frac{(J'+1)^2 - M''^2}{(J'+1)(2J'+3)}|d^{J'}_{M''\Omega}(\theta)|^2 \\ &= \frac{1}{(J'+1)(2J'+3)}\left\{[(J'+1)^2-\Omega^2]\cos^2\theta + \frac{1}{2}[(J'+1)(J'+2)-\Omega^2]\sin^2\theta\right\}. \end{aligned} \tag{48}$$

Those three formulas are exactly the same as Zare's[12,13] except for the constant factor $3\pi/4$.

VI. SUMMARY

In this paper, we have given many recurrence relations of the rotation matrix elements and presented general methods as to how to derive the summations and integrals involving the finite rotation matrix elements, and as to how to expand specific $d^j_{m'm}(\beta)$ into the associated Legendre polynomials. Since $D^j_{m'm}(\alpha\beta\gamma)$ and $d^j_{m'm}(\beta)$ have a relationship as in Eq. (2) and are related to spherical harmonics, our recursion relations may be applied to solid spherical harmonics[14] and other special rotation matrices.[15,16]

APPENDIX A: THE PROOF OF EQ. (42)

From Eq. (9) with $m' = -1/2$, we get

$$\left[j(j+1)-\frac{3}{4}\right]^{1/2} d^j_{(\bar{3}/2)m} + \left[j(j+1)+\frac{1}{4}\right]^{1/2} d^j_{(1/2)m}$$
$$= -\frac{\cos\beta + 2m}{\sin\beta} d^j_{(\bar{1}/2)m}. \tag{A1}$$

Squaring this equation on both sides and integrating with respect to β, we obtain

$$\left[j(j+1)-\frac{3}{4}\right]\int_0^\pi (d^j_{(\bar{3}/2)m})^2 \sin\beta\, d\beta$$
$$+\left[j(j+1)+\frac{1}{4}\right]\int_0^\pi (d^j_{(1/2)m})^2 \sin\beta\, d\beta$$
$$+2\left[j(j+1)-\frac{3}{4}\right]^{1/2}\left[j(j+1)+\frac{1}{4}\right]^{1/2}$$
$$\times\int_0^\pi d^j_{(\bar{3}/2)m} d^j_{(1/2)m} \sin\beta\, d\beta$$
$$=\int_0^\pi \frac{\cos^2\beta + 4m\cos\beta + 4m^2}{\sin^2\beta}(d^j_{(\bar{1}/2)m})^2 \sin\beta\, d\beta. \tag{A2}$$

Using Eqs. (4), (14), (35g), and (39a)–(39d), and with some algebraic manipulation, we obtained (after changing j to J and m to M)

$$\int_0^\pi d^J_{M(1/2)} d^J_{M(\bar{3}/2)} \sin\beta\, d\beta$$
$$=\frac{2}{2J+1}\frac{4M-(2J+1)}{[(2J+3)(2J-1)]^{1/2}}. \tag{A3}$$

Therefore

$$S_{(1/2)(\bar{3}/2)}(J,M) = \frac{2J+1}{4\pi}\int_0^{2\pi} d\alpha \int_0^\pi D^J_{M(1/2)}(\alpha\beta 0)$$
$$\times D^{J*}_{M(\bar{3}/2)}(\alpha\beta 0)\sin\beta\, d\beta$$
$$=\frac{2J+1}{2}\int_0^\pi d^J_{M(1/2)} d^J_{M(\bar{3}/2)} \sin\beta\, d\beta$$
$$=\frac{4M-(2J+1)}{[(2J+3)(2J-1)]^{1/2}}. \tag{A4}$$

APPENDIX B: THE PROOF OF EQ. (43)

From Eq. (10d) and using Eq. (4), we have

$$\frac{1}{2}[j(j+1)(j+2)(j-1)]^{1/2}(d^j_{m\bar{2}} - d^j_{m2})$$
$$=\frac{2m\cos\beta}{\sin^2\beta} d^j_{m0} + \frac{m[j(j+1)]^{1/2}}{\sin\beta}[d^j_{m\bar{1}} - d^j_{m1}]. \tag{B1}$$

Using Eqs. (35j) and (35k), we obtained

$$\frac{1}{4m}[j(j+1)(j+2)(j-1)]^{1/2}[d^j_{m\bar{2}} - d^j_{m2}]$$
$$=\left[\frac{(j-m)!}{(j+m)!}\right]^{1/2}\frac{1}{\sin^2\beta}$$
$$\times[(j+2)\cos\beta P^m_j - (j-m+1)P^m_{j+1}]. \tag{B2}$$

Squaring Eq. (B2) on both sides, integrating over $\sin\beta\, d\beta$, and using Eqs. (14), (39a), (39c), and (39d), finally we obtained (after changing j to J and m to M)

$$\int_0^\pi d^J_{M2} d^J_{M\bar{2}} \sin\beta\, d\beta$$
$$=\frac{2}{2J+1}\left[1-\frac{2M(2J+1)}{J(J+1)}\right.$$
$$\left.+\frac{2M(2J+1)(M^2-1)}{J(J+1)(J+2)(J-1)}\right]. \tag{B3}$$

Therefore

$$S_{2\bar{2}}(J,M) = \frac{2J+1}{2}\int_0^\pi d^J_{M2} d^J_{M\bar{2}} \sin\beta\, d\beta$$
$$=1-\frac{2M(2J+1)}{J(J+1)}+\frac{2M(2J+1)(M^2-1)}{J(J+1)(J+2)(J-1)}. \tag{B4}$$

APPENDIX C: THE PROOF OF EQ. (44)

From Eq. (9) with $m' = -2$ and $m' = 2$, respectively, after adding them together, and after using Eq. (4), we have

$$[j(j+1)-6]^{1/2}[d^j_{m\bar{3}} + d^j_{m3}]$$
$$= -[j(j+1)-2)]^{1/2}[d^j_{m\bar{1}} + d^j_{m1}]$$
$$-\frac{4\cos\beta - 2m}{\sin\beta} d^j_{m2} + \frac{4\cos\beta + 2m}{\sin\beta} d^j_{m\bar{2}}. \tag{C1}$$

Using Eqs. (35j) and (35k), we have

$$d^j_{m\bar{1}} + d^j_{m1} = \frac{2m}{\sin\beta}\left[\frac{(j-m)!}{j(j+1)(j+m)!}\right]^{1/2} P^m_j. \tag{C2}$$

And using Eqs. (35l) and (35m), we have

$$-\frac{4\cos\beta - 2m}{\sin\beta} d^j_{m2} + \frac{4\cos\beta + 2m}{\sin\beta} d^j_{m\bar{2}}$$
$$=\frac{4\cos\beta}{\sin\beta}[d^j_{m\bar{2}} - d^j_{m2}] + \frac{2m}{\sin\beta}[d^j_{m\bar{2}} + d^j_{m2}]$$
$$=\left[\frac{(j-m)!}{j(j+1)(j+2)(j-1)(j+m)!}\right]^{1/2}\frac{4m}{\sin^3\beta}$$
$$\times\{[2(3j+5)\cos^2\beta + 2m^2 - j(j+1)\sin^2\beta]$$
$$\times P^m_j - 6(j-m+1)\cos\beta P^m_{j+1}\}. \tag{C3}$$

Substituting Eqs. (C2) and (C3) into (C1), we obtained

$$[j(j+1)-6]^{1/2}[d^j_{m\bar{3}} + d^j_{m3}]$$
$$=\frac{2m}{\sin\beta}\left[\frac{(j-m)!}{j(j+1)(j+2)(j-1)(j+m)!}\right]^{1/2}$$

$$\times\left\{\left[-3(j^2+5j+6)+\frac{4}{\sin^2\beta}(3j+5+m^2)\right]P_J^m - \frac{12(j-m+1)}{\sin^2\beta}\cos\beta\, P_{j+1}^m\right\}. \quad \text{(C4)}$$

Squaring Eq. (C4) on both sides, integrating over $\sin\beta\, d\beta$, and using Eqs. (14), (15), (39a), (39d), (39j), (39l), and (39k), then we obtained

$$\int_0^\pi d^j_{m3} d^j_{m\bar{3}} \sin\beta\, d\beta = -\frac{2}{2j+1} - \frac{6m(j^2+13j+4m^2+14)}{j(j+1)(j-1)(j-2)} + \frac{4m}{j(j+1)(j+2)(j+3)(j-1)(j-2)} \times\{2(2j+5)(3j^2+13j+3m^2+12) + 3(j^2+3j+m^2+2)^2 - m^2(2j+3)(2j-1)\}. \quad \text{(C5)}$$

Therefore, after changing j to J and m to M, we obtained

$$S_{3\bar{3}}(J,M) = -1 - \frac{3M(2J+1)(J^2+13J+4M^2+14)}{J(J+1)(J-1)(J-2)} + \frac{2M(2J+1)}{J(J+1)(J+2)(J+3)(J-1)(J-2)} \times\{2(2J+5)(3J^2+13J+3M^2+12) + 3(J^2+3J+M^2+2)^2 - M^2(2J+3)(2J-1)\}. \quad \text{(C6)}$$

[1]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1957).
[2]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960).
[3]S. L. Altmann and C. J. Bradley, Philos. Trans. R. Soc. London. Ser. A **255**, 193 (1961).
[4]H. A. Buckmaster, Can. J. Phys. **42**, 386 (1964); Can. J. Phys. **44**, 2525 (1966).
[5]D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U.P., Oxford, 1962).
[6]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U.S. Government Printing Office, Washington, DC, 1970).
[7]J. J. Sakurai, *Modern Quantum Mechanics* (Benjamin–Cummings, Reading, MA, 1985).
[8](a) S. Green and R. N. Zare, Chem. Phys. **7**, 62 (1975); (b) R. N. Zare, J. Chem. Phys. **47**, 204 (1967).
[9]D. C. Jacobs and R. N. Zare, J. Chem. Phys. **85**, 5457 (1986).
[10]M. H. Alexander and P. J. Dagdigian, J. Chem. Phys. **80**, 4325 (1984).
[11]M. H. Alexander, H. J. Werner, and P. J. Dagdigian, J. Chem. Phys. **89**, 1388 (1988).
[12]R. N. Zare, Ph.D. thesis, Harvard University, 1964.
[13]C. Pernot, J. Durup, J.-B. Ozenne, J. A. Beswick, P. C. Cosby, and J. T. Moseley, J. Chem. Phys. **71**, 2387 (1979).
[14]Y.-N. Chiu, J. Math. Phys. **5**, 283 (1964).
[15]Q.-E. Zhang, L.-T. Lin, N.-Q. Wang, and S.-T. Lai, Acta Sci. Natur. Univ. Amoiensis **20**, 209, 221 (1981).
[16]Y.-N. Chiu, J. Chem. Phys. **45**, 2969 (1966).

Observation versus evolution: A short time theorem for positive definite functions, unitary groups, and quantum processes

Marek Kanter
1216 Monterey Ave., Berkeley, California 94707

(Received 17 February 1988; accepted for publication 7 December 1988)

The change of state of a quantum process is studied in the limit when the time t between observations tends to 0. General quantum processes are treated, as represented by a strongly continuous group G of unitary operators on a Hilbert space. Observation is defined as the effect of a self-adjoint operator A on the state vector. When A has countable spectrum and nondegenerate eigenvalues, the limiting change of state (given there is a change) is completely characterized. The most interesting special case occurs when the initial state is an eigenvector of A not in the domain of the Hamiltonian generating G: the whereabouts of the next observed state become infinitely diffuse as $t \to 0$. The results involve some analytic and geometric facts about complex valued positive definite functions.

I. INTRODUCTION

A. Motivation

The development in time of a nonrelativistic quantum process can be modeled as a strongly continuous group of unitary operators $(\mathrm{U}(t): t \in R)$, where t is time, R is the real line, and the unitary operators $\mathrm{U}(t)$ act on some Hilbert space L. The state of the process at time t is represented by the vector $\phi(t) = \mathrm{U}(t)\phi$, where ϕ is in L. Many results in quantum mechanics depend on getting a good approximation to $\mathrm{U}(t)$ for t near 0. Our goal is to develop such approximations in order to lay the groundwork for a study[1] of the limiting effect of infinitely frequent observations on a quantum process.

Our work involves a relatively simple application of von Neumann's mathematical formulation of the concept of "measurement" or "observation" in quantum mechanics. (We use the two words interchangeably.) We postulate that an orthonormal basis for L, denoted by $\{\phi_j : j \in J\}$, completely determines the result of an observation of the quantum process: if the quantum process is observed at time t, then its state $\phi(t)$ is abruptly transformed into ϕ_j with probability $|\langle \phi_j | \phi(t) \rangle|^2$. (We shall suppose L is separable to ensure that J is countable and we will adopt the usual convention that state vectors have norm 1 to force the probabilities to add to 1.) Von Neumann[2] identified the notion of measurement with the effect of a self-adjoint operator A on the state vector. We treat only the case when A has countable spectrum and nondegenerate eigenvalues in order to get concrete results. (In our work the observable A corresponding to the basis $\{\phi_j\}$ does not change with time, i.e., we adopt the "Schrödinger picture.")

From a probabilistic point of view, it is natural to define the quantities $Q_{jk}(t) = |\langle \phi_k | \phi_j(t) \rangle|^2$ for j,k in J. Since $\mathrm{U}(t)$ is unitary, the matrix $Q(t) = (Q_{jk}(t): j,k \in J)$ is doubly stochastic, i.e.,

$$\textstyle\sum_j Q_{jk_0}(t) = \sum_k Q_{j_0 k}(t) = 1 \tag{1.1}$$

for all j_0, k_0 in J. Various quantities derived from $Q(t)$ are pertinent to understanding how a quantum process evolves when repeatedly subjected to observation. For instance, if $\phi(0) = \phi_j$, then the number $(Q_{jj}(t/n))^n$ represents the probability that the observed quantum state was ϕ_j at all the times $s_1 = t/n, s_2 = 2t/n, \ldots, s_n = t$. (This is called "selective observation" of the state ϕ_j.) Another quantity of interest is formed by letting $Q^{(n)}(s)$ stand for the matrix product of $Q(s)$ with itself n times. The term $Q_{jj}^{(n)}(t/n)$ represents the probability that the observed quantum state is ϕ_j at time t, given than it was ϕ_j at time 0 and that observations were carried out at the intermediate points $s_1, s_2, \ldots, s_{n-1}$. The outcome of the observations at the intermediate times are not restricted in any way, hence these observations are called "nonselective." Is is of fundamental significance that these nonselective observations still affect the evolution of the quantum process; in fact

$$\lim_{n\to\infty} Q_{jj}^{(n)}(t/n) = \lim_{n\to\infty} (Q_{jj}(t/n))^n, \tag{1.2}$$

under an extra, possibly nonessential, regularity hypothesis.[1] The limit relation in (1.2) indicates that there is no distinction between selective and nonselective measurement under a regime of infinitely frequent observations. This highlights a fundamental difference between quantum processes and ordinary stochastic processes.

Previous work on the effect of infinitely frequent observation of a quantum process has been restricted to unitary groups $\mathrm{U}(t)$ with the property that $(Q_{jj}(t/n))^n \to 1$ as $n \to \infty$. (This property is known as "Zeno's paradox."[3]) The restrictions made on $\mathrm{U}(t)$ [e.g., non-negativity of H and continuity in t of the second limit in (1.2)] have been justified on physical grounds. In fact $(Q_{jj}(t/n))^n$ does not always tend to 1, within the context of a *general* strongly continuous unitary group on L.[1] Although we do not yet have any physical justification for working in this general context, we believe that it is worthwhile to explore all the interesting mathematical properties of this beautiful model which has been physically fruitful in the past.

 0022-2488/89/040854-04$02.50

B. Summary of main results

The rest of this paper is organized into three sections. In Sec. II we will prove our main result, of which we now give a special case. We let H stand for the infinitesimal generator (Hamiltonian) of $\mathrm{U}(t)$, and we let $D(H)$ stand for the domain of H. We shall prove, for ϕ_j not in $D(H)$, that

$$\lim_{s\to 0} Q_{jk}^{[1]}(s) = 0 \quad (\text{for } k \neq j). \tag{1.3}$$

Here $Q_{jk}^{[1]}(s)$ stands for $Q_{jk}(s)/(1-Q_{jj}(s))$, the conditional probability that the observed quantum state is ϕ_k at time s, given that the observed quantum state was ϕ_j at time 0 *and that the observed quantum state at time s differs from* ϕ_j. Consequently, when a quantum process is observed to leave ϕ_j, its immediate whereabouts are infinitely diffuse. [This provides strong intuitive justification for (1.2) in the case when $\phi_j \notin D(H)$.] In the course of establishing (1.3) we use the key Lemma 2.1, which establishes for $\phi \notin D(H)$ that $[\phi(t)-\phi]/\|\phi(t)-\phi\|$ converges weakly to 0 in L as $t\to 0$. (Remember that, in an infinite-dimensional Hilbert space, a sequence of norm 1 vectors *can* converge weakly to 0.)

Section III is devoted to establishing some facts about the short time behavior of any continuous positive definite function $\Phi(t)$ with $\Phi(0)=1$. These facts will be necessary for the proof of (1.3) and are of some interest in their own right. Our findings about Φ are most conveniently stated in terms of the well-known spectral representation

$$\Phi(t) = \int e^{itx}\,d\nu(x) = E(e^{itX}) \quad (\text{for } t\in R), \tag{1.4}$$

where ν is a probability measure on R and X is a random variable with distribution ν. For convenience we define the parameters $c(\Phi) = (E(X^2))^{1/2}$ and

$$v(\Phi) = \begin{cases} (E(X))^2 c(\Phi)^{-2}, & \text{if } 0 < c(\Phi) < \infty, \\ 0, & \text{if } c(\Phi) = 0 \text{ or } \infty. \end{cases} \tag{1.5}$$

[$v(\Phi)$ is just the square of the reciprocal of the coefficient of variation for X.] If $\Phi(t)$ is not identically 1, then we will show that

$$\lim_{t\to 0}(1-|\Phi(t)|^2)/2(1-\mathrm{Re}\ \Phi(t)) = 1 - v(\Phi), \tag{1.6}$$

and that

$$\lim_{t\to 0}|1-\Phi(t)|^2/(1-|\Phi(t)|^2) = v(\Phi)/(1-v(\Phi)). \tag{1.7}$$

We will need (1.6) and (1.7) in the case when $c(\Phi)$ is infinite, which is precisely the case when these limits do not follow from a straightforward Taylor expansion of Φ about $t=0$. We shall also give an interesting geometric reformulation of (1.6) and (1.7).

Section IV contains a brief discussion regarding the physical consequences of our mathematical results.

II. OBSERVATION VERSUS EVOLUTION

In this section, we shall demonstrate the limiting relation (1.3) as a consequence of a more general result about the change of state of a quantum process a short time after observation. In the course of the demonstration we will make use of (1.6), to be proved in the next section. We start with a lemma which is basic to all that follows.

Lemma 2.1: Let $\{\mathrm{U}(t)\colon t\in R\}$ be a strongly continuous group of unitary operators on L with infinitesimal generator H. For ϕ in L and $t\neq 0$ define $\phi^*(t) = 0$ if $\mathrm{U}(t)\phi = \phi$; otherwise let

$$\phi^*(t) = (\phi(t)-\phi))/\|\phi(t)-\phi\|. \tag{2.1}$$

At 0 define $\phi^*(0) = H\phi/\|H\phi\|$ if $\phi\in D(H)$ and $H\phi\neq 0$; otherwise let $\phi^*(0)=0$. Then $\phi^*(t)$ is weakly continuous at 0, i.e., for ψ in L

$$\lim_{t\to 0}\langle\psi|\phi^*(t)\rangle = \langle\psi|\phi^*(0)\rangle. \tag{2.2}$$

Proof: If $\phi\in D(H)$ and $H\phi\neq 0$, then plainly $\phi^*(t)$ is strongly continuous at 0. If $\phi\in D(H)$ and $H\phi=0$, then $\phi(t)=\phi$ for all t, so (2.2) holds in that case too. It remains only to consider the case $\phi\notin D(H)$. We can further simplify by assuming that $\psi\in D(H)$. [If (2.2) is valid for $\psi\in D(H)$ then it holds for all ψ_0, because $D(H)$ is dense in L and

$$|\langle(\psi-\psi_0)|\phi^*(t)\rangle| \leqslant \|\psi-\psi_0\| \tag{2.3}$$

for all t in R and ψ_0 in L.]

We can write

$$\langle\psi|\phi^*(t)\rangle = \langle(\psi(-t)-\psi)|\phi\rangle/\|\phi(t)-\phi\|, \tag{2.4}$$

where $\psi(t) = \mathrm{U}(t)\psi$. We now observe that the function $\Phi(t) = \langle\phi|\phi(t)\rangle$ is positive definite; thus it has the "spectral" representation (1.4) with

$$\|\phi(t)-\phi\|^2 = 2(1-\mathrm{Re}\ \Phi(t)) = \int 4\sin^2\!\left(\frac{tx}{2}\right)d\nu(x). \tag{2.5}$$

The hypothesis that $\phi\notin D(H)$ is thereby translated into the condition

$$\int x^2\,d\nu(x) = \infty, \tag{2.6}$$

and an application of Fatou's lemma[4] yields

$$\lim_{t\to 0}\inf\|\phi(t)-\phi\|/t = \infty. \tag{2.7}$$

Furthermore, since $\psi\in D(H)$, there exist $\psi' = H\psi$ in L such that

$$\lim_{t\to 0} t^{-1}(\psi(t)-\psi) = \psi', \tag{2.8}$$

where the limit in (2.8) is taken in the norm topology on L. Combining (2.4), (2.7), and (2.8), we conclude that $\langle\psi|\phi^*(t)\rangle$ tends to 0 with t. □

In order to study the limiting behavior of $Q_{jk}^{[1]}(t)$ near $t=0$ we will multiply the denominator in (2.1) by the square root of the quotient in (1.6). Division by 0 will not occur because of the inequality

$$v(\Phi) < 1, \tag{2.9}$$

a special case of the Cauchy–Schwartz inequality which holds unless the spectral measure ν in (1.4) assigns all mass to one point. [In the exceptional case $v(\Phi)=1$ and $|\Phi(t)|=1$ for all t.]

Theorem 2.1: Let $(\mathrm{U}(t)\colon t\in R)$ be a strongly continuous group of unitary operators on L with infinitesimal generator H, and let $\{\phi_j\colon j\in J\}$ be a complete orthonormal basis for L. Let M_j^* stand for the projection operator onto the subspace spanned by vectors orthogonal to ϕ_j. Define the vector $\hat{\phi}_j(0)$ via

$$\hat{\phi}_j(0) = \begin{cases} H\phi_j/\|M_j^* H\phi_j\|, & \text{if } \phi_j\in D(H) \\ & \text{and } M_j^* H\phi_j \neq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (2.10)$$

For $t \neq 0$, define $\hat{\phi}_j(t)$ via

$$\hat{\phi}_j(t) = \begin{cases} 0, & \text{if } \phi_j\in D(H) \text{ and } M_j^* H\phi_j = 0, \\ (\phi_j(t) - \phi_j)/(1 - Q_{jj}(t))^{1/2}, & \text{otherwise.} \end{cases} \quad (2.11)$$

Then the function $\hat{\phi}_j(t)$ is weakly continuous from R to L, and

$$\lim_{t\to 0} Q_{jk}^{[1]}(t) = |\langle \phi_k | \hat{\phi}_j(0)\rangle|^2 \quad (\text{for } k \neq j). \quad (2.12)$$

Proof: We shall first deal with the case $\phi_j \notin D(H)$. Let $\Phi_{jj}(t)$ stand for the positive definite function $\langle \phi_j | \phi_j(t)\rangle$ and note that $Q_{jj}(t) = |\Phi_{jj}(t)|^2$. The assumption on ϕ_j entails the $v(\Phi_{jj}) = 0$ in (1.5). We now apply (1.6) and Lemma 2.1 to conclude that $\hat{\phi}_j(t)$ is weakly continuous at $t = 0$. Furthermore,

$$Q_{jk}^{[1]}(t) = |\langle \phi_k | \hat{\phi}_j(t)\rangle|^2 \quad (\text{for } k \neq j), \quad (2.13)$$

whence the limit in (2.12) is 0 when $\phi_j \notin D(H)$.

If $\phi_j \in D(H)$, note

$$1 - |\Phi_{jj}(t)|^2 = \|M_j^*(\phi_j(t) - \phi_j)\|^2. \quad (2.14)$$

Thus the assertions of the theorem follow in the case $M_j^* H\phi_j \neq 0$ by the same reasoning as used in the previous case, except that now we can only assert that $v(\Phi_{jj}) < 1$ in accordance with (2.9).

The case that $\phi_j \in D(H)$, with $M_j^* H\phi_j = 0$, entails that $\phi_j(t) = e^{it\alpha_j}\phi_j$ for some α_j in R. The proof of the theorem is then a triviality. □

III. SHORT TIME PROPERTIES FOR POSITIVE DEFINITE FUNCTIONS

In this section we will demonstrate the "short time" properties (1.6) and (1.7) for any continuous positive definite function $\Phi(t)$ with $\Phi(0) = 1$. As an interesting geometric consequence we will determine the largest β in $[0,1]$ such that the complex numbers $\Phi(t)$ lie within the complex disk K_β for t small, where

$$K_\beta = \{z\colon\ |z - \beta| \leqslant 1 - \beta\}. \quad (3.1)$$

We shall maintain the convention that the spectral measure associated with Φ is ν, as given by (1.4).

Theorem 3.1: Let Φ be a continuous positive definite function normalized by the condition $\Phi(0) = 1$, and let the parameter $v(\Phi)$ be defined by (1.5). Given that $\Phi(t)$ is not identically 1, we claim (1.6) and (1.7) hold. Furthermore

(a) if $0 \leqslant \beta < 1 - v(\Phi)$, then there exist $\delta > 0$ such that $\Phi(t)\in K_\beta$ for $|t| < \delta$.

(b) If there exists $\delta > 0$ such that $\Phi(t)\in K_\beta$ for $|t| < \delta$, then $0 \leqslant \beta \leqslant 1 - v(\Phi)$.

Proof: The trivial case when $\Phi(t)$ is identically 1 is equivalent to the case when $c(\Phi) = 0$ and can be eliminated from discussion. If $c(\Phi) > 0$ we claim that

$$\lim_{t\to 0} |\Phi(t) - 1|^2/2(1 - \mathrm{Re}\,\Phi(t)) = v(\Phi). \quad (3.2)$$

To prove (3.2) we will use Lemma 2.1 in a special context. Let $L = L^2(\nu)$, the Hilbert space of complex valued functions $\phi(x)$ defined on R, with inner product

$$\langle \psi | \phi \rangle = \textstyle\int \bar{\psi}(x)\phi(x)\,d\nu(x). \quad (3.3)$$

For $\phi \in L^2(\nu)$, define $(\mathrm{U}(t)\phi)(x) = e^{itx}\phi(x)$. It is clear that we can represent $\Phi(t)$ as $\langle \phi_0 | \mathrm{U}(t)\phi_0\rangle$, where $\phi_0(x) = 1$ for all x in R; a direct application of (2.2) with $\phi = \phi_0$ and $\psi = \phi_0$ then yields (3.2).

We now use the identity

$$2\beta(1 - \mathrm{Re}\,\Phi(t)) + (1-\beta)^2 = |\beta - \Phi(t)|^2 + 1 - |\Phi(t)|^2. \quad (3.4)$$

Letting $\beta = 1$ in (3.4) yields (1.6) and (1.7) as a consequence of (3.2). Furthermore, it is clear from (3.4) that $\Phi(t)\in K_\beta$ for $|t| < \delta$ if and only if

$$(1 - |\Phi(t)|^2)/2(1 - \mathrm{Re}\,\Phi(t)) \geqslant \beta \quad (\text{for} |t| < \delta). \quad (3.5)$$

We thus immediately derive (a) and (b) from (1.6).

Remark: The limit relation (1.7) will be used in Ref. 1 in the course of proving (1.2). The same relation is also useful for a Fourier analytic treatement of the convergence of scaled sums of independent random variables with common distribution.[5]

IV. DISCUSSION

It is of interest to consider the possible physical significance of our results. There are two cases: (1) when the initial state ϕ_0 is in $D(H)$, and (2) when $\phi_0 \notin D(H)$.

The task of experimentally verifying (2.12) in the first case reduces to preparing a quantum system in state ϕ_0 and then, a short time t later, applying an observational procedure with possible outcomes $\{\phi_j\}$. This procedure could be repeated many times and the frequency of outcomes in state ϕ_j could be compared with the predicted frequency implied by (2.12). An obvious problem is that we do not know how small t should be so that $Q_{jk}^{[1]}(t)$ is close to its limit. In special cases involving quantum mechanical systems with finite-dimensional state space L (e.g., spin systems) further calculations might clarify the situation.

The case when $\phi_0 \notin D(H)$ is harder to implement experimentally because there is no direct way of preparing a quantum system in such a state. For instance, if H were the Hamiltonian for a free particle, then it would be necessary to put the quantum system into a state with infinite energy. One approach might be to let the initial state ϕ_0 of the system have finite energy E_0 and then change ϕ_0 in such a way that $E_0 \to \infty$. Then the verification of (2.12) would involve two limits: $E_0 \to \infty$ and $t \to 0$. The physical results might very well depend on the rate at which $E_0 \to \infty$ versus the rate at which $t \to 0$, and further mathematical work would be needed to clarify this dependence.

ACKNOWLEDGMENTS

The original impetus for this research grew out of many discussions with Dr. Jack Grodsky, by dint of which he finally succeeded in imparting to me his fascination with the connections between probability theory and quantum mechanics. His untimely death is a great personal and professional loss.

[1]M. Kanter, "A probabilistic framework for the continuous observation of a quantum process," preprint.

[2]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, 1955).

[3]B. Misra and E. C. G. Sudarshan, "The Zeno's paradox in quantum theory," J. Math. Phys. **18**, 756 (1977).

[4]M. Loève, Probability Theory (Springer, New York, 1977), Vol. I.

[5]M. Kanter, "A trichotomy theorem for scaled powers of probability characteristic functions," preprint.

An operator analysis for the Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential

O. L. de Lange
Physics Department, University of Natal, P. O. Box 375, Pietermaritzburg 3200, South Africa

(Received 17 June 1988; accepted for publication 28 September 1988; revised 12 December 1988)

An operator analysis is presented that provides a unified treatment of the Schrödinger (S), Klein–Gordon (KG), and Dirac (D) equations with a Coulomb potential. The analysis uses energy shift operators that factorize an appropriate radial equation. This radial equation is based on standard results and a recent formulation of the Dirac–Coulomb problem [J. Y. Su, Phys. Rev. A **32**, 3251 (1985)]. The shift operators yield energy eigenvalues and a formula that contains normalized, radial coordinate-space wave functions for the S, KG, and D equations. Formulas that contain expectation values for the S, KG, and D equations are obtained by applying the hypervirial theorem and the Hellmann–Feynman theorem to the radial equation.

I. INTRODUCTION

The quantum-mechanical solution for the nonrelativistic hydrogen atom was first given by Pauli,[1] soon after Heisenberg's discovery of the new mechanics.[2] Pauli's treatment is essentially an abstract operator analysis that uses the canonical commutation relations for the position and momentum operators, **r** and **p**, and is independent of any representation of these operators.[3] His analysis is based on an operator (the Pauli–Lenz vector) that alters the angular momentum quantum numbers l and m in the kets $|Elm\rangle$.[3] An alternative operator solution, which is contained in Pauli's method,[4] uses factorization of a radial equation for the nonrelativistic hydrogen atom.[5,6] This factorization yields shift operators for l in the eigenkets $|El\rangle$ of the radial equation. The shift operators are linear in either **p** or **r**.[5,6] The connection with wave mechanics is made by considering either the coordinate representation or the momentum representation.[6] For example, in the coordinate representation the shift operators linear in **p** are the first-order differential operators derived by Schrödinger[7] in his study of the factorization method of solving wave-mechanical problems. These differential operators have been discussed by several authors.[8–12]

The purpose of this paper is to present an operator analysis that provides a unified treatment of the Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential. The operator calculations mentioned above involve transformations between eigenkets with different angular momentum, and there does not appear to be any straightforward way of extending them to a relativistic Coulomb problem.[13] An alternative is to consider shift operators for energy eigenvalues. In the nonrelativistic case such operators are well known. Schrödinger[14] deduced recurrence relations for the radial coordinate-space wave functions in spherical coordinates. Musto[15] introduced scaling operators for wave functions and generalized Schrödinger's differential operators to obtain abstract shift operators for energy in the nonrelativistic hydrogen atom. A similar treatment, based on recurrence relations for coordinate-space wave functions in parabolic coordinates, has been given by Pratt and Jordan.[16] These operators have been widely used.[17]

The starting point of the operator method presented here is a radial equation that contains, as particular cases, the Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential (Sec. II). Shift operators for energy in the bound-state eigenkets of this radial equation are derived in Sec. III. In Sec. IV the hypervirial theorem and the Hellmann–Feynman theorem are applied to the radial equation, and in Sec. V the shift operators are used in an algebraic method for calculating energy eigenvalues.[18] The coordinate representation is considered in Sec. VI, and the shift operators are used to obtain a formula [Eq. (83)] for normalized, radial wave functions of the Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential. In Sec. VII the operator method is compared with wave-mechanical properties of the Coulomb problem and anomalous states are discussed.

The standard wave-mechanical solution to the Dirac–Coulomb equation is rather involved[19] and there has been considerable interest in simplifying this solution.[20–23] The results obtained here are based on an abstract operator analysis that allows a unified treatment of the Schrödinger, Klein–Gordon, and Dirac equations. This operator method is straightforward and self-contained; thus, for example, expectation values and normalized wave functions are calculated without using any properties of special functions.

II. BACKGROUND

The radial Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential can be written in the form

$$[\hbar^{-2}a_\lambda^2 p_r^2 + \tfrac{1}{4} - \lambda a_\lambda r^{-1} + q(q+\delta)a_\lambda^2 r^{-2}]|\lambda q\rangle = 0 . \tag{1}$$

Here $r = (\mathbf{r}\cdot\mathbf{r})^{1/2}$ and $p_r = \frac{1}{2}(\hat{\mathbf{r}}\cdot\mathbf{p} + \mathbf{p}\cdot\hat{\mathbf{r}})$ satisfy the commutation relation

$$[p_r, f(r)] = -i\hbar \frac{df}{dr} . \tag{2}$$

In the coordinate representation

$$p_r = -i\hbar\frac{\partial}{\partial r} - i\hbar r^{-1} , \tag{3}$$

and Eq. (1) is a second-order differential equation. The parameters λ, a_λ, q, and δ are defined in Table I, where

$$\alpha = Ze^2(4\pi\epsilon_0\hbar c)^{-1} \tag{4}$$

 0022-2488/89/040858-09$02.50

TABLE I. The parameters λ, a_λ, q, and δ in Eq. (2). The terms α and a are defined in Eqs. (4) and (5). The terms $|\lambda q\rangle_U$ and $|\lambda q\rangle_L$ are defined in Eq. (10).

	Schrödinger	Klein–Gordon	Dirac
λ	$\left[\frac{\alpha^2 mc^2}{2(mc^2-E)}\right]^{1/2}$	$\left[\frac{\alpha^2 E^2}{m^2c^4-E^2}\right]^{1/2}$	
a_λ	$\frac{1}{2}\lvert\lambda\rvert a$	$\frac{1}{2}(\lambda^2+\alpha^2)^{1/2}a$	
q	l	$[(l+\frac{1}{2})^2-\alpha^2]^{1/2}-\frac{1}{2}$	$[(j+\frac{1}{2})^2-\alpha^2]^{1/2}$
δ	1	1	for $j=l\pm\frac{1}{2}$,[a] $\delta=\mp 1$ if $\lvert\lambda q\rangle=\lvert\lambda q\rangle_U$; $\delta=\pm 1$ if $\lvert\lambda q\rangle=\lvert\lambda q\rangle_L$.

[a] If $j=l-\frac{1}{2}$, $l\neq 0$.

is the fine-structure constant times Z, and

$$a=4\pi\epsilon_0\hbar^2(Zme^2)^{-1} \tag{5}$$

is the Bohr radius divided by Z.

To express the Schrödinger equation in the form (1), the operator identity $\mathbf{L}^2=\mathbf{r}^2(\mathbf{p}^2-p_r^2)$ is used to eliminate $\mathbf{p}^2$ in favor of p_r^2 in the Hamiltonian

$$H=(2m)^{-1}\mathbf{p}^2-\hbar^2(ma)^{-1}r^{-1}+mc^2\,. \tag{6}$$

Then $\mathbf{L}^2$ is replaced with its eigenvalues $\hbar^2 l(l+1)$, to obtain the radial Hamiltonian

$$H_l=(2m)^{-1}[\,p_r^2+\hbar^2 l(l+1)r^{-2} - 2\hbar^2a^{-1}r^{-1}+2m^2c^2]\,. \tag{7}$$

Let $|El\rangle$ denote an eigenket of H_l with eigenvalue E,

$$(H_l-E)|E_l\rangle=0\,. \tag{8}$$

Equations (7) and (8) yield Eq. (1) for the Schrödinger equation where $|\lambda q\rangle$ denotes $|El\rangle$. Similarly, the radial Klein–Gordon equation corresponding to Eq. (8) is

$$(H_l^2-[E+\hbar^2(ma)^{-1}r^{-1}]^2)|El\rangle=0\,, \tag{9}$$

where $H_l^2=c^2p_r^2+c^2\hbar^2l(l+1)r^{-2}+m^2c^4$. Hence we obtain Eq. (1) for the Klein–Gordon equation, where $|\lambda q\rangle$ denotes $|El\rangle$.

If the Dirac equation is multiplied on the left with a suitable operator, it yields a radial equation that is quadratic in p_r.[24] However, not all the eigenkets of this radial equation correspond to eigenkets of the original Dirac equation, and one has the additional task of projecting out the appropriate kets.[24,25] Recently, Su[23] has used a simple similarity transformation of the Dirac equation to derive a radial equation that is quadratic in p_r and can be expressed in the form Eq. (1). An advantage of this method is that the kets $|\lambda q\rangle$ in this quadratic equation are also kets of a linear, radial Dirac equation. Specifically, $|\lambda q\rangle$ is either $|\lambda q\rangle_U$ or $|\lambda q\rangle_L$, which are the kets in the transformed, linear, radial equation[23]

$$H_r'\begin{pmatrix}|\lambda q\rangle_U\\ |\lambda q\rangle_L\end{pmatrix}=E\begin{pmatrix}|\lambda q\rangle_U\\ |\lambda q\rangle_L\end{pmatrix}. \tag{10}$$

This considerable simplification provides part of the motivation for the present paper, where we use Su's formulation of the Dirac equation. Equation (10) consists of the coupled equations[23]

$$(icp_r+\eta\hbar cqr^{-1}-\eta\alpha Eq^{-1})|\lambda q\rangle_U = (mc^2+q^{-1}\sqrt{q^2+\alpha^2}E)|\lambda q\rangle_L \tag{11}$$

and

$$(icp_r-\eta\hbar cqr^{-1}+\eta\alpha Eq^{-1})|\lambda q\rangle_L = (mc^2-q^{-1}\sqrt{q^2+\alpha^2}E)|\lambda q\rangle_U\,, \tag{12}$$

where q and λ are defined in Table I, and $\eta=\mp 1$ for $j=l\pm\frac{1}{2}$. The radial Dirac equation (1) follows from Eqs. (11) and (12). It is clear that the ket $|\lambda q\rangle$ denotes $|Ej\rangle$ in the case of the Dirac equation.

In what follows, bound states ($|E|<mc^2$) are considered. Then λ and a_λ are real. We can, without loss of generality, define a_λ to be positive, as in Table I. Then for an attractive (repulsive) potential, $\lambda>0$ (<0). In Eq. (1), a_λ and $\hbar a_\lambda^{-1}$ are, respectively, scale factors for r and p_r. The energy is given in terms of λ by

$$E=mc^2(1-\tfrac{1}{2}\alpha^2\lambda^{-2}) \tag{13}$$

for the Schrödinger equation, and

$$E=mc^2(1+\alpha^2\lambda^{-2})^{-1/2} \tag{14}$$

for the Klein–Gordon and Dirac equations.

III. SHIFT OPERATORS FOR ENERGY

Equation (1) can be rewritten as an eigenvalue equation in q,

$$O_\lambda|\lambda q\rangle=-q(q+\delta)|\lambda q\rangle\,, \tag{15}$$

where

$$O_\lambda=\hbar^{-2}r^2p_r^2+\tfrac{1}{4}r^2a_\lambda^{-2}-\lambda ra_\lambda^{-1}\,. \tag{16}$$

We wish to derive shift operators $A_\lambda^\pm$ that transform $|\lambda q\rangle$ into $|\lambda\pm 1,q\rangle$. Such operators must satisfy

$$O_{\lambda\pm 1}A_\lambda^\pm=A_\lambda^\pm O_\lambda\,, \tag{17}$$

where $O_{\lambda\pm 1}$ is given by Eq. (16) with a_λ replaced by $a_{\lambda\pm 1}$ and λ by $\lambda\pm 1$. The former change is a scale transformation and we consider it separately.

Thus we first consider a related but simpler problem. Define the operators

$$\tilde{O}_{\lambda\pm 1}=O_\lambda\mp ra_\lambda^{-1}\,, \tag{18}$$

which are obtained from O_λ by leaving a_λ unchanged and changing λ to $\lambda\pm 1$ in Eq. (16). For these we derive operators $T_\lambda^\pm$ such that

$$\tilde{O}_{\lambda\pm 1}T_\lambda^\pm=T_\lambda^\pm O_\lambda\,. \tag{19}$$

Because O_λ is quadratic in p_r, and because a_λ is the same on both sides of Eq. (19), we expect that $T_\lambda^\pm$ are linear in p_r. Let

$$T=f(r)p_r+g(r)\,. \tag{20}$$

From Eq. (18),

$$\tilde{O}_{\lambda\pm1}T=TO_\lambda+[O_\lambda,T]\mp ra_\lambda^{-1}T\,. \tag{21}$$

The commutator in Eq. (21) is evaluated by substituting Eqs. (16) and (20), and using Eq. (2),

$$\tilde{O}_{\lambda\pm1}T=TO_\lambda+F_0+F_1p_r+F_2p_r^2\,, \tag{22}$$

where

$$F_0=i\hbar a_\lambda^{-1}(\tfrac{1}{2}ra_\lambda^{-1}-\lambda)f-r^2g''\mp ra_\lambda^{-1}g\,, \tag{23}$$

$$F_1=-r^2f''\mp ra_\lambda^{-1}f-2i\hbar^{-1}r^2g'\,, \tag{24}$$

$$F_2=2i\hbar^{-1}r(f-rf')\,, \tag{25}$$

and $f'=df/dr$, etc. Set $F_i=0$ $(i=0,1,2)$ and solve for f and g. Then $F_2=0$ gives

$$f=i\hbar^{-1}r\,, \tag{26}$$

where the factor $i\hbar^{-1}$ has been included for convenience. Substituting Eq. (26) in $F_1=0$ and integrating, we find

$$g=\mp\tfrac{1}{2}ra_\lambda^{-1}+b\,, \tag{27}$$

where b is constant. Equations (26) and (27) in $F_0=0$ yield $b=\pm\lambda$. Thus the operators $T_\lambda^\pm$ in Eq. (19) are

$$T_\lambda^\pm=i\hbar^{-1}rp_r\mp(\tfrac{1}{2}ra_\lambda^{-1}-\lambda)\,. \tag{28}$$

If there exist operators $\Delta_\lambda^\pm$ that change $\tilde{O}_{\lambda\pm1}$ into $O_{\lambda\pm1}$; that is, if

$$\Delta_\lambda^\pm\tilde{O}_{\lambda\pm1}=O_{\lambda\pm1}\Delta_\lambda^\pm\,, \tag{29}$$

then Eq. (19), when multiplied on the left by $\Delta_\lambda^\pm$, becomes Eq. (17) with

$$A_\lambda^\pm=\Delta_\lambda^\pm T_\lambda^\pm\,. \tag{30}$$

To determine the operators $\Delta_\lambda^\pm$, substitute Eqs. (18) and (16) in (29). This shows that the condition (29) is satisfied if

$$\Delta_\lambda^\pm(r/a_\lambda)=(r/a_{\lambda\pm1})\Delta_\lambda^\pm \tag{31}$$

and

$$\Delta_\lambda^\pm a_\lambda p_r=a_{\lambda\pm1}p_r\Delta_\lambda^\pm\,. \tag{32}$$

Thus $\Delta_\lambda^\pm$ are scaling operators for the conjugate operators r and p_r. Explicit expressions for these scaling operators are derived in Appendix A,

$$\Delta_\lambda^\pm=c_\lambda^\pm\exp[i\hbar^{-1}rp_r\ln(a_\lambda/a_{\lambda\pm1})]\,, \tag{33}$$

where $c_\lambda^\pm$ are arbitrary constants. We take

$$c_\lambda^\pm=\exp[\tfrac{1}{2}\ln(a_\lambda/a_{\lambda\pm1})]\,,$$

then $(\Delta_\lambda^\pm)^\dagger\Delta_\lambda^\pm=1$ if r and p_r are Hermitian. Using Eqs. (28), (30)–(33), and (2), it is straightforward to show that the shift operators factorize O_λ,

$$-A_{\lambda\pm1}^\mp A_\lambda^\pm=O_\lambda+\lambda(\lambda\pm1)\,. \tag{34}$$

Of particular interest are the special cases in which a shift operator annihilates a ket. That such cases must exist can be seen by allowing both sides of Eq. (34), with $\lambda=\lambda'$, to act on a ket $|\lambda'q\rangle$. Then if λ' is a solution to

$$\lambda'(\lambda'-1)=q(q+\delta)\,, \tag{35}$$

either

$$A_{\lambda'}^-|\lambda'q\rangle=0\,, \tag{36}$$

or

$$A_{\lambda'-1}^+|\lambda'-1,q\rangle=0, \tag{37}$$

or both of these must be true. If the operator

$$\Omega=\hbar^{-2}p_r^2+\tfrac{1}{4}a_\lambda^{-2}-\lambda a_\lambda^{-1}r^{-1}+q(q+\delta)r^{-2} \tag{38}$$

in Eq. (1) is Hermitian, then

$$\langle\lambda q|(A_\lambda^\pm)^\dagger A_\lambda^\pm|\lambda q\rangle=\lambda^{-1}(\lambda\pm1)[\lambda(\lambda\pm1)-q(q+\delta)]\langle\lambda q|\lambda q\rangle\,. \tag{39}$$

(See Appendix B.) Then both Eqs. (36) and (37) are valid, and they can be written [see Eq. (B4)]

$$T_{\lambda'}^-|\lambda'q\rangle=0 \tag{40}$$

and

$$T_{\lambda'-1}^+|\lambda'-1,q\rangle=0\,, \tag{41}$$

where the possible values of λ' are

$$\lambda_\pm=\tfrac{1}{2}\pm[q(q+\delta)+\tfrac{1}{2}]^{1/2}\,. \tag{42}$$

In fact, these are the only values of λ' for which Eqs. (40) and (41) are consistent with Eq. (1). [This can be proved by multiplying Eqs. (40) and (41) on the left with p_rr^{-1} and substituting Eq. (28).] Thus the possibility $\lambda'=1$ that is contained in Eq. (39) for the lowering operation applies only if $q=0$ or $-\delta$; that is, in the nonrelativistic case (see Table I).

From Eqs. (17), (15), and (39),

$$A_\lambda^\pm|\lambda q\rangle=\alpha_{\lambda q}^\pm|\lambda\pm1,q\rangle\,, \tag{43}$$

where

$$\alpha_{\lambda q}^\pm=e^{\pm i\theta}[\lambda^{-1}(\lambda\pm1)\{\lambda(\lambda\pm1)-q(q+\delta)\}]^{1/2}K^\pm\,, \tag{44}$$

$$K^\pm=[\langle\lambda q|\lambda q\rangle\langle\lambda\pm1,q|\lambda\pm1,q\rangle^{-1}]^{1/2}\,, \tag{45}$$

and θ is a constant. For normalized kets of the Schrödinger and Klein–Gordon equations, $K^\pm=1$. For the Dirac equation, $|\lambda q\rangle$ in the above denotes $|\lambda q\rangle_U$ or $|\lambda q\rangle_L$ (see Sec. II). In terms of these, the Dirac ket is[23]

$$|\psi\rangle_D=\begin{pmatrix} i[a|\lambda q\rangle_U-b|\lambda q\rangle_L]|jm,l\rangle \\ [-b|\lambda q\rangle_U+a|\lambda q\rangle_L]\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}|jm,l\rangle \end{pmatrix},$$

where $a=\cosh\tfrac{1}{2}\epsilon$, $b=\sinh\tfrac{1}{2}\epsilon$, and

$$\epsilon=\tanh^{-1}[-\eta\alpha(q^2+\alpha^2)^{-1/2}]\,.$$

Thus the normalization ${}_D\langle\psi|\psi\rangle_D=1$ requires

$$(a^2+b^2)[{}_U\langle\lambda q|\lambda q\rangle_U+{}_L\langle\lambda q|\lambda q\rangle_L]-4ab\,{}_U\langle\lambda q|\lambda q\rangle_L=1\,. \tag{46}$$

Relationships between the scalar products in Eq. (46) can be obtained by multiplying Eqs. (11) and (12) on the left by ${}_L\langle\lambda q|$ and ${}_U\langle\lambda q|$,

$$(mc^2+\sqrt{1+\alpha^2q^{-2}}E)\,{}_L\langle\lambda q|\lambda q\rangle_L=(\sqrt{1+\alpha^2q^{-2}}E-mc^2)\,{}_U\langle\lambda q|\lambda q\rangle_U\,, \tag{47}$$

$$(mc^2+\sqrt{1+\alpha^2q^{-2}}E)\,{}_U\langle\lambda q|\lambda q\rangle_L=\eta\alpha Eq(\lambda^{-2}-q^{-2})\,{}_U\langle\lambda q|\lambda q\rangle_U\,. \tag{47'}$$

In Eq. (47′) we have used Eqs. (54) and (56). Equations (46)–(47′), and (14) give

$$
{}_U\langle\lambda q|\lambda q\rangle_U=\frac{1}{2(1+\alpha^2\lambda^{-2})^{1/2}}\left[\left(\frac{1+\alpha^2q^{-2}}{1+\alpha^2\lambda^{-2}}\right)^{1/2}+1\right],
$$

$$
{}_L\langle\lambda q|\lambda q\rangle_L=\frac{1}{2(1+\alpha^2\lambda^{-2})^{1/2}}\left[\left(\frac{1+\alpha^2q^{-2}}{1+\alpha^2\lambda^{-2}}\right)^{1/2}-1\right],
\tag{48}
$$

and

$$
{}_U\langle\lambda q|\lambda q\rangle_L=-\tfrac{1}{2}\eta\alpha\frac{\lambda^2-q^2}{q(\lambda^2+\alpha^2)}.
\tag{49}
$$

For Dirac kets the factors $K^{\pm}$ are given by Eqs. (45) and (48), except that for $|q+1,q\rangle_L$ the lowering operation (43) is [see Eqs. (39) and (48)]

$$
A^-_{q+1}|q+1,q\rangle_L=0.
\tag{50}
$$

This completes the discussion of shift operators for energy in the kets of Eq. (1). We now consider some applications of these operators and the radial equation (1). In these applications we assume that Ω is Hermitian. The effect of relaxing the Hermitian requirement on Ω is discussed in Sec. VII.

IV. EXPECTATION VALUES

The algebraic method of calculating expectation values is based on the hypervirial theorem[26–28]: If $|\omega\rangle$ is an eigenket of Ω, and if Ω is Hermitian with respect to $|\omega\rangle$ and $W|\omega\rangle$, then

$$
\langle\omega|[\Omega,W]|\omega\rangle=0.
\tag{51}
$$

To apply this theorem to the eigenkets in Eq. (1), let Ω be given by Eq. (38). With $W=G(r)$, Eqs. (51) and (2) yield

$$
\langle\lambda q|G'p_r-\tfrac{1}{2}i\hbar G''|\lambda q\rangle=0,
\tag{52}
$$

where $G'=dG/dr$, etc. Similarly, with $W=G(r)p_r+\tfrac{1}{2}i\hbar G'(r)$, we find

$$
\begin{aligned}
&\langle\lambda q|\tfrac{1}{4}G'''+[-\tfrac{1}{4}a_\lambda^{-2}+\lambda a_\lambda^{-1}r^{-1}-q(q+\delta)r^{-2}]G'\\
&\quad+[q(q+\delta)r^{-3}-\tfrac{1}{2}\lambda a_\lambda^{-1}r^{-2}]G|\lambda q\rangle=0.
\end{aligned}
\tag{53}
$$

Several useful expectation values can be obtained from Eqs. (52) and (53). For example, if $G=r^{s+1}$ we have

$$
\langle r^sp_r\rangle=\tfrac{1}{2}i\hbar s\langle r^{s-1}\rangle\quad(s\neq-1)
\tag{54}
$$

and

$$
\begin{aligned}
&(s+1)\langle r^s\rangle-2(2s+1)\lambda a_\lambda\langle r^{s-1}\rangle\\
&\quad-sa_\lambda^2[s^2-1-4q(q+\delta)]\langle r^{s-2}\rangle=0,
\end{aligned}
\tag{55}
$$

where $\langle\ \rangle$ means $\langle\lambda q|\ \ |\lambda q\rangle$. The expectation value of r^s for all integer $s\geqslant-1$ follows from the recursion relation (55),

$$
\langle r^{-1}\rangle=(2\lambda a_\lambda)^{-1}\langle\lambda q|\lambda q\rangle,
\tag{56}
$$

$$
\langle r\rangle=a_\lambda\lambda^{-1}[3\lambda^2-q(q+\delta)]\langle\lambda q|\lambda q\rangle,
\tag{57}
$$

etc. For integer $s\leqslant-2$ one needs $\langle r^{-2}\rangle$. If λ is a known function of q, $\langle r^{-2}\rangle$ can be calculated in terms of $\langle r^{-1}\rangle$ by treating λ as a continuous variable and applying the Hellmann–Feynman theorem[29] to Ω

$$
\left\langle\frac{\partial\Omega}{\partial\lambda}\right\rangle=0.
\tag{58}
$$

Consider, for example, $|\lambda'q\rangle$ defined by Eq. (40), and the kets obtained from $|\lambda'q\rangle$ by repeated application of raising operators. That is, kets with $\lambda=\lambda_\pm+N$, where $N=0,1,2,\ldots$. Then $q(q+\delta)=(\lambda-N)(\lambda-N-1)$, and Eqs. (38) and (58) give

$$
\begin{aligned}
(2\lambda_\pm-1)a_\lambda^2\langle r^{-2}\rangle&=(2a_\lambda)^{-1}\frac{\partial a_\lambda}{\partial\lambda}\langle\lambda q|\lambda q\rangle\\
&\quad-\left(\lambda\frac{\partial a_\lambda}{\partial\lambda}-a_\lambda\right)\langle r^{-1}\rangle.
\end{aligned}
\tag{59}
$$

From Eqs. (59) and (56),

$$
\langle r^{-2}\rangle=a_\lambda^{-2}[2\lambda(2\lambda_\pm-1)]^{-1}\langle\lambda q|\lambda q\rangle.
\tag{60}
$$

From Eqs. (56), (60), and (1),

$$
\begin{aligned}
\langle p_r^2\rangle&=\tfrac{1}{4}\hbar^2a_\lambda^{-2}[1-2\lambda_\pm(\lambda_\pm-1)\\
&\quad\times\{\lambda(2\lambda_\pm-1)\}^{-1}]\langle\lambda q|\lambda q\rangle.
\end{aligned}
\tag{61}
$$

We can also treat the fine-structure constant α as a continuous variable and apply the Hellmann–Feynman theorem to Ω:

$$
0=\left\langle\frac{\partial\Omega}{\partial\alpha}\right\rangle=\left\langle\frac{1}{4}\frac{\partial a_\lambda^{-2}}{\partial\alpha}-r^{-1}\frac{\partial(\lambda a_\lambda^{-1})}{\partial\alpha}\right\rangle.
$$

This yields Eq. (56). It is clear that Eqs. (56), (60), and (61) depend only on the assumption that Ω is Hermitian.

For the Schrödinger and Klein–Gordon equations, the results (54), (55), and (59) have been obtained before[27,28]: we see that they apply also to the transformed Dirac equation and that the calculations can be performed in a unified manner. For the Dirac equation, the above expectation values are with respect to the kets $|\lambda q\rangle_U$ and $|\lambda q\rangle_L$ in Eq. (10). It is also useful to calculate ${}_U\langle\lambda q|f(r)|\lambda q\rangle_L$. This can be done using Eqs. (11), (12), and the above expectation values. For example, multiplying Eq. (11) on the left by ${}_U\langle\lambda q|r$, and using Eqs. (54) and (57), we have

$$
\begin{aligned}
{}_U\langle\lambda q|r|\lambda q\rangle_L&=\tfrac{1}{2}a\alpha(1+\alpha^2\lambda^{-2})^{-1/2}[-\tfrac{1}{2}+\eta q\\
&\quad-\tfrac{1}{2}\eta q^{-1}\{3\lambda^2-q(q+\delta)\}]\,{}_U\langle\lambda q|\lambda q\rangle_U.
\end{aligned}
\tag{57$'$}
$$

Expectation values with respect to the Dirac kets $|\psi\rangle_D$ can be obtained from the above results (see Sec. V).

Because $\langle r\rangle$ is real, it follows from Eq. (57) and Table I that q must be real; thus we have the restrictions $\alpha\leqslant\frac{1}{2}$ and $\alpha\leqslant1$, respectively, for the Klein–Gordon and Dirac equations.

V. ENERGY EIGENVALUES

A normalizable eigenket $|\lambda q\rangle$ of the operator Ω is a bound state of an attractive Coulomb potential if $\lambda>0$. Suppose there exists such a ket with $\lambda\geqslant\lambda_++1$. For the sequence

$$
A^-_\lambda|\lambda q\rangle,A^-_{\lambda-1}A^-_\lambda|\lambda q\rangle,\ldots,
\tag{62}
$$

there are three possibilities; either it terminates at $\lambda'=\lambda_+$, or at $\lambda'=\lambda_-$, or it does not terminate at all (see Sec. III). We now examine which of these possibilities ensures non-negativity of the norm in the sequence (62). The kets in (62) will have non-negative norm provided

$$
F(\lambda)=\lambda^{-1}(\lambda-1)(\lambda-\lambda_-)(\lambda-\lambda_+)
\tag{63}
$$

does not become negative [see Eqs. (39) and (42)].

(i) For the Schrödinger equation, $\lambda_+ = l+1$ and $\lambda_- = -l$. If $l \geqslant 1$, $F(\lambda) < 0$ in $(1, l+1)$. Therefore (62) must terminate at λ_+, and the possible values of λ are $\lambda_+ + N$ $(N = 0,1,...)$. If $l = 0$, $F(\lambda) = (\lambda - 1)^2$. Now a normalizable s state with nonintegral λ cannot exist because from any such state one could obtain a normalizable p state with nonintegral λ.[30] Thus for s states (62) terminates at $\lambda = 1$.

(ii) For the Klein–Gordon equation, $\lambda_\pm = \frac{1}{2} \pm [(l+\frac{1}{2})^2 - \alpha^2]^{1/2}$. If $l \geqslant 1$, $F(\lambda) < 0$ in $(\lambda_-, 0)$ and $(1, \lambda_+)$. Therefore all sequences (62) that do not terminate at λ_+ will produce a ket for which $F(\lambda) < 0$. If $l = 0$, $F(\lambda) < 0$ in $(0, \frac{1}{2} - \frac{1}{2}[1 - 4\alpha^2]^{1/2})$ and $(\frac{1}{2} + \frac{1}{2}[1 - 4\alpha^2]^{1/2}, 1)$. These intervals are too small to conclude that (62) must terminate at λ_+. However, if (62) does not terminate it will contain kets with $\lambda < 0$; that is, kets for which the expectation value of the positive-definite operator r^{-1} is negative [see Eq. (56)]. Therefore, for these kets, one or both of the assumptions made in Sec. IV must be wrong: either Ω is not Hermitian or the kets are not normalizable. Similarly, if (62) terminates at λ_-, from Eq. (61),

$$\langle \lambda_- q | p_r^2 | \lambda_- q \rangle = -\tfrac{1}{4}(1 - 4\alpha^2)^{-1/2} \hbar^2 a_\lambda^{-2} \langle \lambda_- q | \lambda_- q \rangle \tag{64}$$

is negative, which contradicts the assumption that p_r is Hermitian. The properties of s states for which (62) does not terminate at λ_+ are discussed in more detail in Sec. VII.

(iii) For the Dirac equation, $\lambda_+ = q+1$ if $\delta = 1$ and $\lambda_+ = q$ if $\delta = -1$. For the kets $|\lambda q\rangle_L$, the sequence (62) must terminate at $\lambda' = q+1$ [see Eq. (50)] and therefore $\lambda = q+1, q+2, \ldots$. This result and Eqs. (11) and (39) show that for the kets $|\lambda q\rangle_U$, $\lambda = N + q + 1$ if $\delta = 1$, and $\lambda = N + q$ if $\delta = -1$ $(N = 0,1,...)$. These results and the various transformations of the Dirac kets are depicted in Fig. 1.

To summarize: the normalizable eigenkets $|\lambda q\rangle$ of the Hermitian operator Ω are those with

$$\lambda = N + \tfrac{1}{2}(1+\delta) + q \tag{65}$$

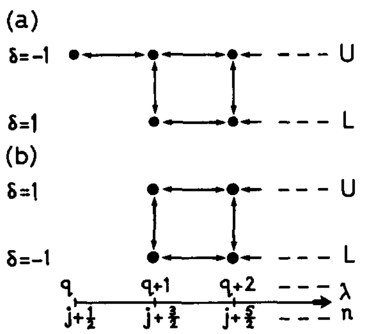


FIG. 1. Schematic illustration of transformations for Dirac kets: (a) for $j = l + \frac{1}{2}$, (b) for $j = l - \frac{1}{2}$. The transformations depicted by vertical arrows are between $|\lambda q\rangle_U$ and $|\lambda q\rangle_L$; they are effected by the operators in Eqs. (11) and (12). The transformations depicted by horizontal arrows are between $|\lambda q\rangle_U$ and $|\lambda \pm 1, q\rangle_U$, etc.; they are effected by the shift operators $A_\lambda^{\pm}$ in Eq. (43). The values of λ and n are from Eqs. (65) and (70). The parameter δ is defined in Table I. In (b), the ket $|qq\rangle_L$ does not exist because of the normalization condition (49).

$(N = 0,1,...)$, except that for $N = 0$ and $j = l - \frac{1}{2}$ the Dirac ket $|qq\rangle_L$ does not exist. The latter restriction is a consequence of the normalization condition (48).

We can now write down the energy eigenvalues. Equations (13), (14), (65), and Table I yield the standard formulas

$$E = mc^2(1 - \tfrac{1}{2}\alpha^2 n^{-2}) \tag{66}$$

for the Schrödinger equation, and

$$E = mc^2(1 + \alpha^2[n - l - \tfrac{1}{2} + \sqrt{(l+\tfrac{1}{2})^2 - \alpha^2}\,]^{-2})^{-1/2} \tag{67}$$

for the Klein–Gordon equation, where

$$n = N + l + 1 \quad (= 1,2,\ldots)\,. \tag{68}$$

For the Dirac equation

$$E = mc^2(1 + \alpha^2[n - j - \tfrac{1}{2} + \sqrt{(j+\tfrac{1}{2})^2 - \alpha^2}\,]^{-2})^{-1/2}, \tag{69}$$

where

$$n = N + \tfrac{1}{2}(1+\delta) + j + \tfrac{1}{2} \quad (= 1,2\ldots) \tag{70}$$

and $n \neq j + \frac{1}{2}$ if $j = l - \frac{1}{2}$.

It is interesting to compare the quantization condition Eq. (65) with that given by the WKB approximation.[31] This approximation, as refined by Kemble[32] and Langer,[33] provides the quantization rule

$$\int_{r_1}^{r_2} (p_r^2 + \tfrac{1}{4}\hbar^2 r^{-2})^{1/2}\, dr = \pi(N + \tfrac{1}{2})\hbar, \tag{71}$$

where r_1 and r_2 are the zeros of the integrand, and $N = 0,1,\ldots$. In our case, p_r^2 is found by setting $\Omega = 0$ in Eq. (38). Then Eq. (71) yields Eq. (65),[34] but without the restriction on the Dirac ket $|qq\rangle_L$.

For the Dirac equation, expectation values with respect to the kets $|\lambda q\rangle_U$ and $|\lambda q\rangle_L$ were obtained in Sec. IV. The expectation value with respect to the Dirac ket $|\psi\rangle_D$ (see Sec. III) can be obtained from these results. For example

$$\begin{aligned} {}_D\langle \psi | r | \psi \rangle_D &= (a^2 + b^2)({}_U\langle \lambda q | r | \lambda q \rangle_U + {}_L\langle \lambda q | r | \lambda q \rangle_L) \\ &\quad - 4ab\,{}_U\langle \lambda q | r | \lambda q \rangle_L\,. \end{aligned} \tag{72}$$

Substituting Eqs. (57), (57'), and the values of a and b given in Sec. III, this yields

$$\begin{aligned} \langle r \rangle_D &= \tfrac{1}{2} a (1 + \alpha^2 \lambda^{-2})^{-1/2} [3\lambda^2 - q^2 \\ &\quad - \eta q (1 + \alpha^2 q^{-2})^{1/2} (1 + \alpha^2 \lambda^{-2})^{1/2} + 2\alpha^2]\,, \end{aligned} \tag{73}$$

where $\lambda = n - j - \frac{1}{2} + q$, and $\eta = \mp 1$ for $j = l \pm \frac{1}{2}$. Similarly for the other expectation values in Sec. IV.

VI. COORDINATE-SPACE WAVE FUNCTIONS

The shift operators $\Delta_\lambda^{\pm} T_\lambda^{\pm}$ obtained in Sec. III are used to calculate the radial coordinate-space wave functions, $\varphi_{\lambda q}(r)$. For this purpose, the effect of the scaling operators $\Delta_\lambda^{\pm}$ on these wave functions must be determined. Let

$$\Delta_\lambda^{\pm} \varphi_{\lambda q}(r) = \delta_\lambda^{\pm} \varphi_{\lambda q}\left(\frac{a_\lambda}{a_{\lambda \pm 1}}\, r\right), \tag{74}$$

where $\delta_\lambda^{\pm}$ are real and independent of r and p_r. The norm,

$$\int_0^\infty \varphi^\dagger_{\lambda q}\varphi_{\lambda q} r^2\, dr\,,$$

must be invariant under the transformation (74). Thus

$$\delta^\pm_\lambda = (a_\lambda / a_{\lambda\pm 1})^{3/2}\,. \tag{75}$$

In the coordinate representation Eq. (40) is

$$\left(r\frac{d}{dr} + \frac{1}{2} r a^{-1}_{\lambda'} - \lambda' + 1\right)\varphi_{\lambda' q}(r) = 0\,, \tag{76}$$

and the solution normalized to unity is

$$\varphi_{\lambda' q}(r) = a_{\lambda'}^{-3/2} \times [\Gamma(2\lambda' + 1)]^{-1/2}(r/a_{\lambda'})^{\lambda' - 1} e^{-r/2a_{\lambda'}}\,. \tag{77}$$

It is convenient in the following derivation of $\varphi_{\lambda q}(r)$ to use wave functions normalized to unity; for the Dirac wave functions, the normalization is corrected to (48) by including an appropriate factor in the final result [the factor M in Eq. (84)]. From Eq. (43),

$$A^+_{\lambda' + N - 1} \cdots A^+_{\lambda' + 1} A^+_{\lambda'} \varphi_{\lambda' q}(r) = \alpha^+_{\lambda' + N - 1, q} \cdots \alpha^+_{\lambda' + 1, q}\alpha^+_{\lambda' q}\varphi_{\lambda' + N, q}(r)\,. \tag{78}$$

To evaluate the coefficient on the right-hand side of Eq. (78), use Eqs. (44) (with $K^\pm = 1$) and (35),

$$\alpha^+_{\lambda' + N - 1, q} = e^{i\theta}\left[\frac{N(\lambda' + N)(2\lambda' + N - 1)}{\lambda' + N - 1}\right]^{1/2}. \tag{79}$$

Thus

$$\alpha^+_{\lambda' + N - 1, q} \cdots \alpha^+_{\lambda' + 1, q}\alpha^+_{\lambda' q} = e^{iN\theta}[N!(\lambda' + N)(2\lambda')(2\lambda' + 1) \cdots (2\lambda' + N - 1)/\lambda']^{1/2}\,. \tag{80}$$

The left-hand side of Eq. (78) contains N scaling operators. These can be removed by first commuting them to the right of the operators $T^+_{\lambda' + N - 1} \cdots T^+_{\lambda'}$, using Eqs. (31) and (32), and then using Eqs. (74), (75), and (77): the left-hand side of Eq. (78) simplifies to

$$a^{-3/2}_{\lambda' + N}[\Gamma(2\lambda' + 1)]^{-1/2} D_N \cdots D_2 D_1 x^{\lambda' - 1} e^{-x/2}\,, \tag{81}$$

where $D_k = x d/dx - \frac{1}{2}x + \lambda' + k$ and $x = r/a_{\lambda' + N}$. The differentiations in (81) are readily performed (Appendix C),

$$D_N \cdots D_2 D_1 x^{\lambda' - 1} e^{-x/2} = (2\lambda')(2\lambda' + 1)\cdots(2\lambda' + N - 1) \times \left[1 + \frac{(-N)}{2\lambda'}x + \cdots + \frac{(-N)(-N+1)\cdots(-1)}{2\lambda'(2\lambda' + 1)\cdots(2\lambda' + N - 1)}\frac{x^N}{N!}\right] \times x^{\lambda' - 1} e^{-x/2}\,. \tag{82}$$

The polynomial in square brackets in Eq. (82) is the confluent hypergeometric function, ${}_1F_1(-N; 2\lambda'; x)$.[35] Collecting the above results we have

$$\varphi_{\lambda q}(r) = C_{\lambda q}(r/a_\lambda)^{\lambda' - 1} e^{-r/2a_\lambda}\, {}_1F_1(-N; 2\lambda'; r/a_\lambda)\,, \tag{83}$$

where $\lambda = \lambda' + N$, $\lambda' = q + \frac{1}{2}(1 + \delta)$, $N = 0, 1, \ldots$,

$$C_{\lambda q} = e^{-iN\theta} a_\lambda^{-3/2}[\Gamma(2\lambda' + N)]^{1/2} \times [N! 2\lambda\{\Gamma(2\lambda')\}^2]^{-1/2} M\,. \tag{84}$$

For Schrödinger and Klein–Gordon wave functions that are normalized to unity, $M = 1$. For Dirac wave functions the normalizations (48) require

$$M = [2(1 + \alpha^2\lambda^{-2})^{1/2}]^{-1/2}\left[\left(\frac{1 + \alpha^2 q^{-2}}{1 + \alpha^2\lambda^{-2}}\right)^{1/2} \pm 1\right]^{1/2}, \tag{85}$$

where the upper (lower) sign applies to the upper (lower) wave function in Eq. (10).

The quantities N, λ', and a_λ in Eq. (83) can readily be written in terms of quantum numbers (see Sec. V). In the Schrödinger and Klein–Gordon wave functions, $N = n - l - 1$. In the Schrödinger wave functions, $\lambda' = l + 1$ and $a_\lambda = \frac{1}{2}na$. In the Klein–Gordon wave functions,

$$\lambda' = \tfrac{1}{2} + [(l + \tfrac{1}{2})^2 - \alpha^2]^{1/2}$$

and

$$a_\lambda = \tfrac{1}{2}([n - l - \tfrac{1}{2} + \sqrt{(l + \tfrac{1}{2})^2 - \alpha^2}]^2 + \alpha^2)^{1/2} a\,.$$

The upper and lower components of Dirac wave functions yield sequences (62) that terminate at different values of λ (see Sec. V). Thus if $j = l + \frac{1}{2}$, for the upper component $\lambda' = q$ and $N = n - j - \frac{1}{2}$ in Eq. (83); for the lower component $\lambda' = q + 1$ and $N = n - j - \frac{3}{2}$ [see Eq. (70)]. Similarly, if $j = l - \frac{1}{2}$. Thus for $j = l \pm \frac{1}{2}$, the bound-state wave functions in Eq. (10) can be written

$$|nj\rangle = \begin{pmatrix}\varphi^\pm_{nj}\\ \varphi^\mp_{nj}\end{pmatrix}, \tag{86}$$

where

$$\varphi^\pm_{nj} = C^\pm_{nj}\rho^{q - 1/2 \mp 1/2} e^{-(1/2)\rho} \times {}_1F_1(-N + \tfrac{1}{2} \mp \tfrac{1}{2};\ 2q + 1 \mp 1; \rho)\,, \tag{87}$$

$$C^\pm_{nj} = e^{-i(N - 1/2 \pm 1/2)\theta} a_{nj}^{-3/2} \times \left[\frac{\pi^\pm\Gamma(N + \frac{1}{2} \mp \frac{1}{2} + 2q)}{N! 2(N + q)\{\Gamma(2q + 1 \mp 1)\}^2}\right]^{1/2} M^\pm_{nj}\,, \tag{88}$$

$$M^\pm_{nj} = [2(1 + \alpha^2\{q + N\}^{-2})^{1/2}]^{-1/2} \times \left[\left(\frac{1 + \alpha^2 q^{-2}}{1 + \alpha^2\{q + N\}^{-2}}\right)^{1/2} \mp \eta\right]^{1/2}, \tag{89}$$

$$a_{nj} = \tfrac{1}{2}[(N + q)^2 + \alpha^2]^{1/2} a\,, \tag{90}$$

$\rho = r/a_{nj}$, $N = n - j - \frac{1}{2}$, $\pi^+ = 1$, $\pi^- = N$, and $\eta = \mp 1$ for $j = l \pm \frac{1}{2}$. The phase θ in Eq. (88) is not arbitrary: it is fixed by the choice of phase in Eqs. (11) and (12), and one can readily show that $\theta = \frac{1}{2}(1 - \eta)\pi$.

Su[23] has used properties of the confluent hypergeometric function to obtain similar expressions for the radial wave functions of the transformed Dirac–Coulomb equation. The wave functions of this transformed equation are[23]

$$\psi' = \begin{pmatrix} i\varphi^{\pm}_{nj}\chi^{l}_{jm} \\ \varphi^{\mp}_{nj}\boldsymbol{\sigma}\cdot\hat{\mathbf{r}}\chi^{l}_{jm} \end{pmatrix}, \tag{91}$$

where σ_i are Pauli matrices. The solutions of the original Dirac–Coulomb equation are more complicated,[19] and are related to ψ' by the similarity transformation[23]

$$\psi = (\cosh(\epsilon/2) + i\beta\boldsymbol{\alpha}\cdot\hat{\mathbf{r}}\sinh(\epsilon/2))^{-1}\psi', \tag{92}$$

where α_i and β are Dirac matrices, and $\epsilon = \tanh^{-1}[-\alpha\eta^{-1}(j+\frac{1}{2})^{-1}]$.

VII. DISCUSSION

It is interesting to compare the operator analysis presented here with wave-mechanical properties of the Coulomb problem. In the coordinate representation of wave mechanics, Eq. (1) is a second-order differential equation. For the Klein–Gordon and Dirac equations, the (unnormalized) solutions that vanish at $r=\infty$ are[36]

$$\begin{aligned}\varphi_{\lambda q}\left(\frac{r}{a_\lambda}\right) &= \frac{\Gamma(1-2\lambda_-)}{\Gamma(\lambda_+-\lambda)}\,{}_1F_1\left(\lambda_--\lambda;2\lambda_-;\frac{r}{a_\lambda}\right) \\ &\quad\times\left(\frac{r}{a_\lambda}\right)^{\lambda_--1}e^{-r/2a_\lambda} \\ &\quad+\frac{\Gamma(1-2\lambda_+)}{\Gamma(\lambda_--\lambda)}\,{}_1F_1\left(\lambda_+-\lambda;2\lambda_+;\frac{r}{a_\lambda}\right) \\ &\quad\times\left(\frac{r}{a_\lambda}\right)^{\lambda_+-1}e^{-r/2a_\lambda}.\end{aligned} \tag{93}$$

It is straightforward to show that[37]

$$A_\lambda^-\varphi_{\lambda q}(r/a_\lambda) = (\lambda-\lambda_-)(\lambda-\lambda_+)\varphi_{\lambda-1,q}(r/a_{\lambda-1}) \tag{94}$$

and

$$A_{\lambda-1}^+\varphi_{\lambda-1,q}(r/a_{\lambda-1}) = (\lambda-\lambda_-)(\lambda-\lambda_+)\varphi_{\lambda q}(r/a_\lambda), \tag{95}$$

where

$$A_\lambda^\pm = \Delta_\lambda^\pm\left(r\frac{d}{dr}\mp\frac{1}{2}ra_\lambda^{-1}\pm\lambda+1\right), \tag{96}$$

and $\Delta_\lambda^\pm$ are given by Eqs. (33) and (3). When $\lambda=\lambda_\pm$, Eqs. (94) and (95) are the differential forms of Eqs. (40) and (41).

The quadratic integrability of the wave functions (93) depends on their behavior as $r\to 0$. The confluent hypergeometric functions in Eq. (93) are well-behaved, for all values of λ, as $r\to 0$.[36] Because $\lambda_- < \lambda_+$, the first term in Eq. (93) dominates as $r\to 0$, and $\varphi_{\lambda q}\sim r^{\lambda_--1}$. Thus if $\lambda_-\leqslant -\frac{1}{2}$, $\varphi_{\lambda q}$ is quadratically integrable only if the first term in Eq. (93) is zero. This restricts the values of $\lambda_+-\lambda$ to the poles of $\Gamma(\lambda_+-\lambda)$; that is,

$$\lambda = N+\lambda_+ \tag{97}$$

($N=0,1,\ldots$). With these values of λ the wave functions (93) are, when normalized, identical to the Klein–Gordon and Dirac wave functions given by Eq. (83) with $\lambda'=\lambda_+$. For s states of the Klein–Gordon equation, $\lambda_- > -\frac{1}{2}$ (Sec. V) and the wave functions (93) are quadratically integrable for all values of λ, including $\lambda<0$. The existence of these "anomalous states" (quadratically integrable states with $\lambda\neq N+\lambda_+$) has been noticed before,[38] and bound s states of the Klein–Gordon equation with a repulsive Coulomb potential (s states with $\lambda<0$) have been used in a model of the electron.[39] The anomalous s states with $\lambda=N+\lambda_-$ in Eq. (93) are, when normalized, identical to the solutions (83) for s states of the Klein–Gordon equation with $\lambda'=\lambda_-$. Anomalous states are not confined to the Klein–Gordon equation: if half-integral values of orbital angular momentum are admitted, there are anomalous states of the Dirac equation with $j=0$.[39]

Next we consider whether the operator Ω [Eq. (38)] is Hermitian with respect to the wave functions (93). The operator Ω is Hermitian on a subspace of $L^2(0,\infty)$. This subspace, $L_{1/2}{}^2$, consists of quadratically integrable functions $\varphi(r)$ that satisfy the condition $r^{1/2}\varphi(r)\to 0$ as $r\to 0$.[40] The wave functions (93), with λ given by Eq. (97), belong to $L_{1/2}{}^2$, but the anomalous s states of the Klein–Gordon equation do not. Another feature of these anomalous states is that they do not have a satisfactory nonrelativistic limit. For example, for the anomalous 1s state, $\lambda=\lambda_-$: in the limit $\alpha\to 0$, $\lambda_-\approx\alpha^2$ (Sec. V), and therefore $E\approx\alpha mc^2$ [Eq. (14)], and $\langle r\rangle\approx\frac{1}{2}\alpha a$ [Eq. (57)]. This behavior occurs because the nonrelativistic anomalous states are not solutions of the Schrödinger–Coulomb equation.[38]

For the Schrödinger–Coulomb equation, the solutions that vanish at $r=\infty$ are different from Eq. (93).[36,41] The quadratically integrable wave functions are those with λ given by Eq. (97): they are identical to the Schrödinger wave functions given by Eq. (83) with $\lambda'=\lambda_+$. The operator Ω is Hermitian with respect to these wave functions.

Thus the quadratically integrable wave functions of the Hermitian operator Ω are, when normalized, identical to those obtained in Sec. VI by the operator method. If the condition that Ω be Hermitian with respect to these wave functions is relaxed, one has, in addition, anomalous s states for the Klein–Gordon equation.

APPENDIX A: DERIVATION OF EQ. (33)

We derive explicit expressions for operators $\Delta_\lambda^\pm$ that perform the scaling operations (31) and (32). Because $\Delta_\lambda^\pm$ commute with rp_r, we expect that they are functions of rp_r. Consider the operator

$$\Delta(rp_r) = \sum_{n=0}^{\infty} b_n(rp_r)^n, \tag{A1}$$

where b_n are independent of r and p_r. From Eq. (2),

$$r(rp_r)^n = (rp_r+i\hbar)^n r \tag{A2}$$

and

$$p_r(rp_r)^n = (rp_r-i\hbar)^n p_r. \tag{A3}$$

Therefore

$$r\Delta(rp_r) = \Delta(rp_r+i\hbar)r \tag{A4}$$

and

$$p_r\Delta(rp_r) = \Delta(rp_r-i\hbar)p_r. \tag{A5}$$

Equation (A4) will be the scaling operation (31) if

$$\Delta(rp_r + i\hbar) = (a_{\lambda\pm 1}/a_\lambda)\Delta(rp_r)\,. \tag{A6}$$

The solutions to (A6) are

$$\Delta_\lambda^\pm(rp_r) = c_\lambda^\pm (a_\lambda/a_{\lambda\pm 1})^{i\hbar^{-1}rp_r}, \tag{A7}$$

where $c_\lambda^\pm$ are arbitrary constants. Similarly, (A5) and (32) yield (A7). Hence we obtain Eq. (33).

APPENDIX B: PROOF OF EQ. (39)

From Eq. (30) and $(\Delta_\lambda^\pm)^\dagger\Delta_\lambda^\pm = 1$,

$$\langle (A_\lambda^\pm)^\dagger A_\lambda^\pm\rangle = \langle (T_\lambda^\pm)^\dagger T_\lambda^\pm\rangle\,. \tag{B1}$$

The adjoint of Eq. (28) is

$$(T_\lambda^\pm)^\dagger = -T_\lambda^\mp - 1\,. \tag{B2}$$

Substituting Eqs. (28) and (B2) in (B1), and using Eqs. (2) and (15),

$$\langle (T_\lambda^\pm)^\dagger T_\lambda^\pm\rangle = \langle -2i\hbar^{-1}rp_r \pm ra_\lambda^{-1} + \lambda^2 \mp \lambda - q(q+\delta)\rangle\,. \tag{B3}$$

Equations (B3), (54), and (57) give

$$\langle (T_\lambda^\pm)^\dagger T_\lambda^\pm\rangle = \lambda^{-1}(\lambda\pm 1) \times [\lambda(\lambda\pm 1) - q(q+\delta)]\langle \lambda q|\lambda q\rangle\,. \tag{B4}$$

Equations (B1) and (B4) yield Eq. (39).

APPENDIX C: PROOF OF EQ. (82)

We use induction to prove Eq. (82). The result is obvious if $N=1$. In Eq. (82) the term in $x^{n+\lambda'-1}$ ($n \nless N$) is

$$a_n(N) = 2\lambda'(2\lambda'+1)\cdots(2\lambda'+N-1) \times\left[\frac{(-N)(-N+1)\cdots(-N+n-1)}{2\lambda'(2\lambda'+1)\cdots(2\lambda'+n-1)}\frac{x^n}{n!}\right] \times x^{\lambda'-1}e^{-(1/2)x},$$

$$a_0(N) = 2\lambda'(2\lambda'+1)\cdots(2\lambda'+N-1)x^{\lambda'-1}e^{-(1/2)x}\,.$$

If Eq. (82) is multiplied on the left by $D_{N+1} = xd/dx - \frac{1}{2}x + \lambda' + N + 1$, the term in $x^{n+\lambda'-1}$ is (for $0\neq n \nless N$)

$$(n+\lambda'-1)a_n(N) - xa_{n-1}(N) + (\lambda'+N+1)a_n(N)$$
$$= 2\lambda'(2\lambda'+1)\cdots(2\lambda'+N-1) \times\left[(2\lambda'+N)(-N-1) \times\frac{(-N)(-N+1)\cdots(-N+n-1)}{2\lambda'(2\lambda'+1)\cdots(2\lambda'+n-1)}\frac{x^n}{n!}\right] \times x^{\lambda'-1}e^{-(1/2)x}$$
$$= a_n(N+1)\,.$$

If $n=0$, the result is

$$(\lambda'-1)a_0(N) + (\lambda'+N+1)a_0(N) = a_0(N+1)\,.$$

Finally, D_{N+1} on Eq. (82) also produces a term $-xa_N(N) = a_{N+1}(N+1)$.

[1] W. Pauli, Z. Phys. **36**, 336 (1926) [translated in *Sources of Quantum Mechanics*, edited by B. L. van der Waerden (Dover, New York, 1968), pp. 387–415].

[2] See, for example, J. Mehra and H. Rechenberg, *The Historical Development of Quantum Theory* (Springer, New York, 1982), Vol. 3, Chap. 4.

[3] See, for example, A. Böhm, *Quantum Mechanics* (Springer, New York, 1979), Chap. 6.

[4] O. L. de Lange and R. E. Raab, Phys. Rev. A **35**, 951 (1987).

[5] J. D. Newmarch and R. M. Golding, Am. J. Phys. **46**, 658 (1978).

[6] O. L. de Lange and R. E. Raab, Am. J. Phys. **55**, 913 (1987).

[7] E. Schrödinger, Proc. R. Irish Acad. Sect. A **46**, 9 (1940).

[8] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).

[9] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Pt. 1, Chap. 6.

[10] L. C. Biedenharn and P. J. Brussaard, *Coulomb Excitation* (Clarendon, Oxford, 1965), Chap. 3.

[11] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981), p. 357.

[12] L. C. Biedenharn, Found. Phys. **13**, 13 (1983).

[13] There is, however, a useful analogy: the radial Dirac equations can be formally identified with angular momentum shift operations $l=0$ to $l=\pm 1$ in the radial wave functions of a second-order equation. This result yields the energy eigenvalues of the Dirac equation (see Ref. 8, Sec. 8.4). A similar method for calculating energy eigenvalues is discussed in H. Green, *Matrix Mechanics* (Noordhoff, Groningen, 1965), Chap. 7.

[14] E. Schrödinger, Proc. R. Irish Acad. Sect. A **46**, 183 (1941). An alternative treatment of Schrödinger's recurrence relations, and their extension to the Dirac–Coulomb equation, is given in Ref. 8, Secs. 8.2 and 8.3. In this method, quantum numbers in wave functions are replaced by variables and operators are defined that involve differentiation with respect to these variables. These "O operators" are different from the energy shift operators derived above: the former cannot be expressed in terms of the radial operators r and p_r, and they do not factorize any radial equation.

[15] R. Musto, Phys. Rev. **148**, 1274 (1966). Note that the scaling operators in this paper, and in Ref. 16, differ by constant factors from the nonrelativistic limit of our Eq. (33).

[16] R. H. Pratt and T. F. Jordan, Phys. Rev. **148**, 1276 (1966).

[17] See, for example, B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), Chap. 21, and the references therein.

[18] An alternative algebraic method of calculating these eigenvalues is the group-theoretical method based on the SO(2,1) spectrum-generating algebra that is common to the radial Schrödinger, Klein–Gordon, and Dirac equations with a Coulomb potential: see Ref. 17, Chap. 18, and the references therein.

[19] See, for example, H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957), Sec. 14. The complexity of the Dirac–Coulomb problem, and the desirability of alternative solutions to this problem, have been stressed by Biedenharn (Ref. 12).

[20] P. R. Auvil and L. M. Brown, Am. J. Phys. **46**, 679 (1978).

[21] S. Waldenstrøm, Am. J. Phys. **47**, 1098 (1979).

[22] E. H. de Groot, Am. J. Phys. **50**, 1141 (1982).

[23] J. Y. Su, Phys. Rev. A **32**, 3251 (1985). Our Eqs. (11) and (12) correspond to Eqs. (2.11) and (2.12) of this reference. The latter are expressed in the coordinate representation. This is an inessential feature, and it is straightforward to derive Eqs. (2.11) and (2.12) in terms of the radial operator p_r.

[24] See, for example, G. Baym, *Lectures on Quantum Mechanics* (Benjamin, New York, 1969), Chap. 23.

[25] P. C. Martin and R. J. Glauber, Phys. Rev. **109**, 1307 (1958); L. C. Biedenharn, *ibid.* **126**, 845 (1962); M. K. F. Wong and H. Y. Yeh, Phys. Rev. D **25**, 3396 (1982).

[26] J. O. Hirschfelder, J. Chem. Phys. **33**, 1762 (1960).

[27] J. H. Epstein and S. T. Epstein, Am. J. Phys. **30**, 266 (1962) and references therein.

[28] S. T. Epstein, Am. J. Phys. **44**, 251 (1976). There is a misprint in Eq. (7) of this reference: the coefficient of $\langle r^{-1}\rangle$ should be $-2Z\,\partial E/\partial l$.

[29] H. Hellmann, Acta Physicochim. URSS **1**, 913 (1935); **4**, 225 (1936); *Einführung in die Quantechemie* (Deuticke, Leipzig, 1937); R. P. Feynman, Phys. Rev. **56**, 340 (1939). See also E. C. Kemble, Ref. 31, Chap. 13.

[30] The shift operation that raises l in the nonrelativistic kets is

$$[p_r + i\hbar(l+1)r^{-1} - i\hbar a^{-1}(l+1)^{-1}]|El\rangle = [2mE + \hbar^2a^{-2}(l+1)^{-2}]^{1/2}|E,l+1\rangle.$$

See, for example, Refs. 5 and 6.
[31]This is a standard application for the Schrödinger–Coulomb equation: see, for example, E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (Dover, New York, 1958), Chap. 5. For the Dirac–Coulomb equation, see Ref. 23.
[32]E. C. Kemble, Phys. Rev. **48**, 549 (1935). See also Ref. 31, Chap. 3.
[33]R. E. Langer, Phys. Rev. **51**, 669 (1937).
[34]A similar coincidence exists between the Bohr–Wilson–Sommerfeld quantization rules of the old quantum theory applied to the relativistic Coulomb problem, and the exact solution to the Dirac–Coulomb equation. This "Sommerfeld puzzle" has been discussed in detail by Biedenharn (Ref. 12).
[35]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), p. 504.
[36]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, Chap. 3.
[37]Equations (10) on p. 258 of Ref. 36 are helpful in obtaining these results.
[38]B. H. Armstrong and E. A. Power, Am. J. Phys. **31**, 262 (1963).
[39]B. H. Armstrong, Phys. Rev. **130**, 2506 (1963).
[40]$L^2_{1/2}$ is the subspace in which the kinetic energy is finite. See, for example, R. L. Liboff, I. Nebenzahl, and H. H. Fleischmann, Am. J. Phys. **41**, 976 (1973), and H. A. Bethe and R. Jackiw, *Intermediate Quantum Mechanics*, (Benjamin, New York, 1968) 2nd. ed., Chap. 21.
[41]T. Tietz, Sov. Phys. JETP **3**, 777 (1956).

Viscous fluid spheres in general relativity

A. Banerjee and S. B. Dutta Choudhury
Department of Physics, Jadavpur University, Calcutta-700032, India

(Received 31 March 1988; accepted for publication 2 November 1988)

The spherically symmetric distribution of a dissipative fluid with nonvanishing bulk as well as shear viscosity is discussed. The cosmological constant is included in the field equations for more generality. The junction conditions are applied at the boundary of the sphere to match with the exterior Schwarzschild metric to yield the condition $T^1_1 = 0$. In the presence of the bulk viscosity alone spatial homogeneity demands the vanishing of shear as well as the conformal flatness of the space-time. On the other hand, nonzero shear viscosity along with the uniform density assumption result in a unique temporal behavior of the matter density increasing always in the course of time.

I. INTRODUCTION

A widely accepted cosmological model at present is Friedmann's homogeneous and isotropic expanding universe, which assumes a continuous distribution of perfect fluid. Recently, however, the idea of introducing dissipative mechanisms in the fluid at a certain stage has motivated many authors to study the effects of viscosity, heat conduction, etc., in the dynamics of cosmological models. In 1968, Misner[1] suggested that the neutrino viscosity could be an efficient mechanism by means of which arbitrary initial anisotropy dies away rapidly as the universe expands. The bulk viscosity coefficient has been used by Klimek[2] and later by Murphy[3] in order to construct a singularity-free isotropic and homogeneous cosmological model. Belinsky and Khalatnikov[4] studied qualitatively the role of both bulk and shear viscosities in the singularities and the behavior in asymptotic limits. One of their interesting results is the possibility of the creation of matter in the course of evolution by the gravitational field. The dissipative mechanisms not only modify the nature of singularities but also can successfully account for the large entropy per baryon in the present universe.[5] The dissipative phenomena may play a quite significant role in the process of galaxy formation and gravitational collapse as well; more extensive investigations in this field are needed. It may therefore be useful in the above context to extend at the first stage the study of the dynamics of a spherically symmetric distribution of perfect fluid as discussed by Misner and Sharp[6] and that of Glass[7] in the presence of dissipative effects due to shear and bulk viscosities. In the energy momentum tensor, therefore, some additional terms due to dissipative phenomena[8] are incorporated. The conditions of fit at the boundary are also explicitly derived, which may be relevant in the problems of gravitational collapse for a bounded distribution.

In Sec. II the stress energy tensor is explicitly written and the corresponding field equations are given in a comoving coordinate system for the spherically symmetric metric. The total energy scalar $m(r,t)$ and also the free gravitational energy within a comoving radius r are computed. These are generalizations of the expressions given in the paper of Glass for a perfect fluid. The main result stressed in this section is arrived at by applying the junction conditions of Israel[9] at the boundary of the sphere, where the interior fits with the external Schwarzschild metric. The continuity of the first and second fundamental forms demand that at the boundary $T^1_1 = 0$ or, in other words $(p - \xi\theta)^2 = \frac{16}{3}\eta^2\sigma^2$, where ξ and η stand for the bulk and shear viscosity coefficients, and θ and σ^2 represent the expansion and shear scalars, respectively. For a perfect fluid the result reduces simply to the vanishing of pressure at the boundary.

In Sec. III some general results are derived for a viscous fluid sphere and are compared with those of a perfect fluid. One of these results is that if the fluid has only the bulk viscosity and it is spatially homogeneous the sphere will have isotropic motion. This is an interesting generalization of the theorem of Misra and Srivastava[10] for a perfect fluid case. The additional information we obtain from the present treatment is that the metric in this case is also conformally flat.

In Sec. IV we start with the assumptions of a spatially uniform fluid density but nonzero shear and bulk viscosity coefficients. It is found that in this case a positive shear viscosity coefficient restricts the matter density increase monotonically with time, which in turn leads to the result that the spherical surface at each comoving radius can only decrease in time, exhibiting the irreversible character of the dynamics of a viscous fluid.

II. FIELD EQUATIONS AND THE MATCHING AT THE BOUNDARY

The general time-dependent spherically symmetric metric can be expressed in the form

$$ds^2 = e^{\nu}\, dt^2 - e^{\lambda}\, dr^2 - e^{\mu}(d\theta^2 + \sin^2\theta\, d\phi^2), \qquad (2.1)$$

where ν, λ, and μ are functions of both the radial coordinate and time. The energy momentum tensor for a viscous fluid is

$$T^{\mu}_{\nu} = (\rho + \bar{p})v^{\mu}v_{\nu} - \bar{p}\delta^{\mu}_{\nu} + 2\eta\sigma^{\mu}_{\nu} + \Lambda\delta^{\mu}_{\nu}, \qquad (2.2a)$$

with

$$\bar{p} = p - \xi\theta \qquad (2.2b)$$

and

$$\sigma_{\mu\nu} = \tfrac{1}{2}(v_{\mu;\nu} + v_{\nu;\mu}) - \tfrac{1}{3}\theta(g_{\mu\nu} - v_{\mu}v_{\nu}) - \tfrac{1}{2}(\dot{v}_{\mu}v_{\nu} + \dot{v}_{\nu}v_{\mu}), \qquad (2.3)$$

where $\dot{v}^{\mu} = v^{\mu}_{;\lambda}v\lambda$ is the acceleration. In the above ρ and p stand for the matter density and the thermodynamic pressure, respectively, η and ξ for the shear and bulk viscosity

 0022-2488/89/040867-05$02.50

coefficients, respectively, and θ for the expansion scalar. In comoving coordinates, $v^\mu = e^{-\nu/2}\delta^\mu_0$, and one thus obtains, from (2.3),

$$\sigma^1_1 = -2\sigma^2_2 = -2\sigma^3_3 = \tfrac{1}{3}e^{-\nu/2}(\dot\lambda - \dot\mu). \tag{2.4}$$

The expansion and shear scalars are therefore given by

$$\theta = e^{-\nu/2}(\dot\lambda/2 + \dot\mu) \tag{2.5}$$

and

$$\sigma^2 = \tfrac{1}{2}\sigma^{\mu\nu}\sigma_{\mu\nu} = \tfrac{1}{12}e^{-\nu}(\dot\lambda - \dot\mu)^2, \tag{2.6}$$

respectively. The Bianchi identity $T^{\mu\nu}{}_{;\nu} = 0$ leads to the following equations:

$$e^{-\nu/2}\dot\rho + (\rho + \bar p)\theta - 4\eta\sigma^2 = 0, \tag{2.7}$$

$$\bar p' + \frac{\nu'}{2}(\rho + \bar p) - (2\eta\sigma^1_1)' - 2\eta\sigma^1_1\left(3\frac{\mu'}{2} + \frac{\nu'}{2}\right)0. \tag{2.8}$$

Now Einstein's field equations can be explicitly written in the form

$$8\pi\rho = e^{-\mu} + e^{-\lambda}\left(-\mu'' - \tfrac{3}{4}\mu'^2 + \frac{\lambda'\mu'}{2}\right) + e^{-\nu}\left(\frac{\dot\mu^2}{4} + \frac{\dot\lambda\dot\mu}{2}\right) - \Lambda, \tag{2.9}$$

$$8\pi(\bar p - 2\eta\sigma^1_1) = -e^{-\mu} + e^{-\lambda}\left(\frac{\mu'^2}{4} + \frac{\mu'\nu'}{2}\right) - e^{-\nu}\left(\ddot\mu + \tfrac{3}{4}\dot\mu^2 - \frac{\dot\mu\dot\nu}{2}\right) + \Lambda, \tag{2.10}$$

$$8\pi(\bar p - 2\eta\sigma^2_2) = e^{-\lambda}\left(\frac{\mu''}{2} + \frac{\mu'^2}{4} + \frac{\nu''}{2} + \frac{\nu'^2}{4} - \frac{\lambda'\mu'}{4} - \frac{\lambda'\nu'}{4} + \frac{\mu'\nu'}{4}\right) + e^{-\nu}\left(-\frac{\ddot\mu}{2} - \frac{\dot\mu^2}{4} - \frac{\ddot\lambda}{2} - \frac{\dot\lambda^2}{4} - \frac{\dot\lambda\dot\mu}{4} + \frac{\dot\lambda\dot\nu}{4} + \frac{\dot\mu\dot\nu}{4}\right) + \Lambda, \tag{2.11}$$

$$0 = 2\dot\mu' - \mu'(\dot\lambda - \dot\mu) - \nu'\dot\mu \tag{2.12}$$

In the field equations given above the cosmological constant Λ has been included for generality. The two scalars, namely, the energy scalar within a definite comoving r at any instant and the conformal curvature scalar, are invariantly defined as (see Glass)

$$m(r,t) = \tfrac{1}{2}e^{3\mu/2}R_{\alpha\beta\gamma\rho}\xi^\alpha\xi^\beta\eta^\lambda\eta^\rho(\xi^\delta\xi_\delta\eta^\sigma\eta_\sigma)^{-1} \tag{2.13}$$

and

$$\psi_2 = \tfrac{1}{2}C_{\alpha\beta\gamma\rho}\xi^\alpha\xi^\beta\eta^\gamma\eta^\rho(\xi^\delta\xi_\delta\eta^\sigma\eta_\sigma)^{-1}, \tag{2.14}$$

respectively, with respect to the two vectors

$$\xi^\alpha = (0,0,1,0), \quad \eta^\alpha = (0,0,0,1).$$

One can easily compute the Riemann curvature tensor component R_{2323} and the only independent component of the Weyl curvature tensor C_{2323}:

$$R_{2323} = e^\mu \sin^2\theta\,[1 - \tfrac{1}{4}e^{(\mu-\lambda)}\mu'^2 + \tfrac{1}{4}e^{(\mu-\nu)}\dot\mu^2] \tag{2.15}$$

and

$$C_{2323} = -\tfrac{1}{3}e^\mu \sin^2\theta\left[1 + e^{(\mu-\lambda)}\left(\frac{\mu''}{2} - \frac{\nu''}{2} - \frac{\nu'^2}{4} - \frac{\lambda'\mu'}{4} + \frac{\lambda'\nu'}{4} + \frac{\mu'\nu'}{4}\right) + e^{(\mu-\nu)}\left(-\frac{\ddot\mu}{2} + \frac{\ddot\lambda}{2} + \frac{\dot\lambda^2}{4} - \frac{\dot\lambda\dot\mu}{4} + \frac{\dot\mu\dot\nu}{4} - \frac{\dot\lambda\dot\nu}{4}\right)\right], \tag{2.16}$$

so that, in view of (2.13) and (2.14), m and ψ_2 are explicitly written as

$$m(r,t) = \tfrac{1}{2}e^{\mu/2} - \tfrac{1}{8}e^{(3\mu/2-\lambda)}\mu'^2 + \tfrac{1}{8}e^{(3\mu/2-\nu)}\dot\mu^2, \tag{2.17}$$

and

$$\psi_2 = -\tfrac{1}{6}\left[e^{-\mu} + e^{-\lambda}\left(\frac{\mu''}{2} - \frac{\nu''}{2} - \frac{\nu'^2}{4} - \frac{\lambda'\mu'}{4} + \frac{\lambda'\nu'}{4} + \frac{\mu'\nu'}{4}\right) + e^{-\nu}\left(-\frac{\ddot\mu}{2} - \frac{\dot\lambda\dot\mu}{4} - \frac{\dot\lambda\dot\nu}{4} + \frac{\dot\mu\dot\nu}{4} + \frac{\ddot\lambda}{2} + \frac{\dot\lambda^2}{4}\right)\right]. \tag{2.18}$$

Differentiating (2.17) with respect to t and r, one obtains

$$\dot m = \tfrac{1}{6}(e^{3\mu/2})^{\cdot}\,[e^{-\mu} + e^{-\lambda}\{-\tfrac{3}{4}\mu'^2 - \tfrac{1}{2}(\mu'/\dot\mu)(2\dot\mu' - \dot\lambda\mu')\} + e^{-\nu}(\ddot\mu + \tfrac{3}{4}\dot\mu^2 - \dot\mu\dot\nu/2)], \tag{2.19}$$

$$m' = \tfrac{1}{6}(e^{3\mu/2})\,[e^{-\mu} + e^{-\lambda}(-\mu'' - \tfrac{3}{4}\mu'^2 + \tfrac{1}{2}\lambda'\mu') + e^{-\nu}\{\tfrac{3}{4}\dot\mu^2 + \tfrac{1}{2}(\dot\mu/\mu')(2\dot\mu' - \nu'\dot\mu)\}], \tag{2.20}$$

respectively. Now using the field equations on the right-hand sides of (2.19) and (2.20) it is not much more difficult to show that time and space derivatives of m are

$$\dot m = -(4\pi/3)(R^3)^{\cdot}\,(\bar p - 2\eta\sigma^1_1 - \Lambda/8\pi) \tag{2.21}$$

and

$$m' = (4\pi/3)(R^3)'(\rho + \Lambda/8\pi), \tag{2.22}$$

where R stands for $e^{\mu/2}$. The above two expressions may be said to be generalizations, for a viscous fluid in the presence of the cosmological parameter, of what has been done for a perfect fluid by Glass. Similarly we can now attempt to generalize the equation obtained by Glass for a perfect fluid connecting the matter density ρ with the two scalars m and ψ_2, viz., $m + \psi_2 R^3 = (4\pi/3)\rho R^3$, in the more general situation under consideration. Combining (9), (17), and (18) one obtains

$$m(r,t) = \frac{4\pi}{3}\left(\rho + \frac{\Lambda}{8\pi} - \frac{3\psi_2}{4\pi}\right)e^{3\mu/2} + \frac{1}{6}e^{3\mu/2}\left\{e^{-\mu} + e^{-\lambda}\left(\frac{\mu''}{2} + \frac{\nu''}{2} + \frac{\nu'^2}{4} - \frac{\lambda'\nu'}{4} - \frac{\lambda'\nu'}{4} - \frac{\mu'\nu'}{4}\right) + e^{-\nu}\left(\frac{\ddot\mu}{2} + \frac{\dot\mu^2}{4} - \frac{\ddot\lambda}{2} - \frac{\dot\lambda^2}{4} - \frac{\dot\lambda\dot\mu}{4} + \frac{\dot\lambda\dot\nu}{4} - \frac{\dot\mu\dot\nu}{4}\right)\right\}. \tag{2.23}$$

Again subtracting (2.10) from (2.11) one has

$$16\pi\eta(\sigma_1^1-\sigma_2^2)$$

$$=e^{-\mu}+e^{-\lambda}\left[\frac{\mu''}{2}+\frac{\nu''}{2}+\frac{\nu'^2}{4}-\frac{\lambda'\mu'}{4}-\frac{\lambda'\nu'}{4}-\frac{\mu'\nu'}{4}\right]$$
$$+e^{-\nu}\left[\frac{\ddot\mu}{2}+\frac{\dot\mu^2}{2}-\frac{\ddot\lambda}{2}-\frac{\dot\lambda^2}{4}-\frac{\dot\lambda\dot\mu}{4}+\frac{\dot\lambda\dot\nu}{4}-\frac{\dot\mu\dot\nu}{4}\right] \tag{2.24}$$

and finally substituting (2.24) in (2.23) one obtains

$$m(r,t)=\frac{4\pi}{3}\left(\rho+\frac{\Lambda}{8\pi}-\frac{3\psi_2}{4\pi}+3\eta\sigma_1^1\right)R^3. \tag{2.25}$$

The quantity $-\psi_2R^3$ is said to be the free gravitational energy within the spherical surface of radius r and will henceforth be denoted by E (see Glass). The space-time derivatives of E can now be expressed as

$$\dot E=-(4\pi/3)\left[\dot\rho R^3+(\rho+\bar p-2\eta\sigma_1^1)(R^3)^{\cdot}+(3\eta\sigma_1^1R^3)^{\cdot}\right] \tag{2.26}$$

and

$$E'=-(4\pi/3)\left[\rho'R^3+(3\eta\sigma_1^1R^3)'\right]. \tag{2.27}$$

In what has been said before we have generalized some of the results of Glass introducing additional properties of the fluid such as viscosity and the presence of a cosmological parameter. Now we proceed to derive conditions of fit at the boundary of the sphere with the exterior Schwarzschild metric. In doing this we apply the junction conditions of Israel that utilize the matching of the first and second fundamental forms at the boundary.

The hypersurface Σ divides the space-time into two regions V^- the interior and V^+ the exterior of the sphere, both of which contain Σ as their boundary surface. The intrinsic metric Σ is given by

$$ds_\Sigma^2=d\tau^2-x^2(\tau)(d\theta^2+\sin^2\theta\,d\varphi^2). \tag{2.28}$$

When approaching Σ in V^- and V^+ the continuity of the first and second fundamental forms demands

$$ds_\Sigma^2=(ds_-)_\Sigma^2=(ds_+)_\Sigma^2 \tag{2.29a}$$

and

$$[K_{ij}]=K_{ij}^+-K_{ij}^-=0, \tag{2.29b}$$

respectively, where $K_{ij}^\pm$ is the extrinsic curvature at the boundary Σ in $V^\pm$. It is given by (Eisenhart[11])

$$K_{ij}^\pm=n_\alpha^\pm\frac{\partial^2x_\pm^\alpha}{\partial\psi^i\,\partial\psi^j}+n_\alpha^\pm(\Gamma_{\mu\nu}^\alpha)_\pm\left(\frac{\partial x_\pm^\mu}{\partial\psi^i}\right)\left(\frac{\partial x_\pm^\nu}{\partial\psi^j}\right), \tag{2.30}$$

where the ψ's stand for intrinsic coordinates at the boundary surface and $n^\pm$ for the normal vectors to Σ in the coordinates $x_\pm^\alpha$, respectively. The comoving boundary of the interior region described by the line element (1) is given by

$$f(r)=r-r_\Sigma=0, \tag{2.31}$$

where r_Σ is a constant and the unit normal vector to Σ is

$$n_\alpha^-=(0,e^{(1/2)\lambda(r_\Sigma,t)},0,0).$$

The junction condition (2.29a) yields, in view of (2.1) and (2.28),

$$R(r_\Sigma,t)+X(\tau)\quad\text{and}\quad e^{(1/2)\nu(r_\Sigma,t)}\left(\frac{dt}{d\tau}\right)=1. \tag{2.32}$$

Now the extrinsic curvature component $K_{\theta\theta}^-$ can be computed with the help of (2.30) and (2.32) yielding the result

$$K_{\theta\theta}^-=-(e^{-\lambda/2}RR')_\Sigma. \tag{2.33}$$

Following Misner, as mentioned earlier, we now write $U=e^{-\nu/2}\dot R$, so that we have, from (2.17),

$$e^{-\lambda}=(1+U^2-2m/R)(R')^{-2}$$

and thus (2.33) can be written as

$$K_{\theta\theta}^-=-R_\Sigma(1-2m(r_\Sigma,t)/R_\Sigma+U_\Sigma^2)^{1/2}. \tag{2.34}$$

The exterior space-time is described by the Schwarzschild metric:

$$ds_+^2=(1-2M/\bar r)d\bar t^2-(1-2M/\bar r)^{-1}d\bar r^2-\bar r^2(d\theta^2+\sin\theta\,d\varphi^2), \tag{2.35}$$

where $\bar r$ and $\bar t$ are the radial and time coordinate, respectively, and M is the usual Schwarzschild constant mass parameter. The equation of the boundary surface Σ is now given by

$$f(\bar r,\bar t)=\bar r-\bar r_\Sigma(\bar t)=0. \tag{2.36}$$

The unit normal vector on Σ described in terms of the exterior metric is now

$$n_\alpha^+=\left[\left(1-\frac{2M}{\bar r_\Sigma}\right)-\left(1-\frac{2M}{\bar r_\Sigma}\right)^{-1}\left(\frac{d\bar r_\Sigma}{dt}\right)^2\right]^{-1/2}\times\left[-\left(\frac{d\bar r_\Sigma}{dt}\right),1,0,0\right]. \tag{2.37}$$

It should be noted that the Schwarzschild radial coordinate $\bar r$ is different from the comoving radial coordinate in the interior. The junction condition (2.29a) yields

$$\bar r_\Sigma(\bar t)=X(\tau) \tag{2.38a}$$

and

$$\left[\left(1-\frac{2M}{\bar r_\Sigma}\right)-\left(1-\frac{2M}{\bar r_\Sigma}\right)^{-1}\left(\frac{d\bar r_\Sigma}{d\bar t}\right)^2\right]^{1/2}\left(\frac{d\bar t}{d\tau}\right)_\Sigma=1. \tag{2.38b}$$

Using (2.38b) in (2.37) we obtain

$$n_\alpha^+=\left[-\left(\frac{d\bar r_\Sigma}{d\tau}\right),\left(\frac{d\bar t}{d\tau}\right),0,0\right]. \tag{2.39}$$

One can now compute $K_{\theta\theta}^+$ from (2.30), (2.38a), and (2.39) and obtains the following relation:

$$K_{\theta\theta}^+=-\bar r_\Sigma\left(1-\frac{2M}{\bar r_\Sigma}\right)\left(\frac{d\bar t}{d\tau}\right)_\Sigma. \tag{2.40}$$

From the definition of U it is now evident that one can write

$$U_\Sigma=\left(\frac{dR_\Sigma}{d\tau}\right)=\left(\frac{d\bar r_\Sigma}{d\tau}\right)=\left(\frac{dx(\tau)}{d\tau}\right). \tag{2.41}$$

Since

$$U_\Sigma=\frac{d\bar r_\Sigma}{d\tau}=\frac{d\bar r_\Sigma}{d\bar t}\left(\frac{d\bar t}{d\tau}\right),$$

one can write, in view of the relation (2.38b),

$$\left(\frac{d\bar t}{d\tau}\right)_\Sigma=\left(1-\frac{2M}{\bar r_\Sigma}\right)^{-1}\left[1-\frac{2M}{\bar r_\Sigma}+U_\Sigma^2\right]^{1/2} \tag{2.42}$$

and the result (2.40) can thus be expressed as

$$K^{+}_{\theta\theta} = -\bar{r}_\Sigma [1 - 2M/\bar{r}_\Sigma + U^2_\Sigma]^{1/2}. \tag{2.43}$$

So the matching condition of (2.34) and (2.43) at the boundary immediately yields the result

$$m(r_\Sigma,t) = M, \tag{2.44}$$

which in turn leads to the relation

$$\dot{m}(r_\Sigma,t) = 0. \tag{2.45}$$

This result, again in view of (2.21) and ignoring the cosmological constant Λ locally, is in effect at the boundary,

$$\bar{p} - 2\eta\sigma^1_1 = 0, \tag{2.46}$$

which can be written in an invariant form

$$\bar{p}^2 = \tfrac{16}{3}\eta^2\sigma^2. \tag{2.47}$$

Again for a perfect fluid this is equivalent to the vanishing of pressure at the boundary surface.

We now proceed to see what the other matching condition, namely, the continuity of $K_{\tau\tau}$, leads to. In the way shown above one can calculate $K^{-}_{\tau\tau}$ in the form

$$K^{-}_{\tau\tau} = (e^{-\lambda/2}\nu'/2)_\Sigma. \tag{2.48}$$

In view of (2.8), this again yields

$$K^{-}_{\tau\tau} = \left(1 + U^2(r_\Sigma,t) - \frac{2m(r_\Sigma,t)}{R_\Sigma}\right)^{1/2} \times \left[\frac{\{(\partial/\partial R)(2\eta\sigma^1_1 R^3)\}_{,t} - (\partial\bar{p}/\partial R)_{,t}R^3}{(\rho + \bar{p} - 2\eta\sigma^1_1)R^3}\right]_{r=r_\Sigma} \tag{2.49}$$

where, as previously mentioned, R stands for $e^{\mu/2}$ in the interior metric. Further using the notation $DU = e^{-\nu/2}\,\partial U/\partial t$, one gets, in view of (2.10) and (2.17),

$$DU = e^{-\lambda}R'\nu'/2 - [4\pi R(\bar{p} - 2\eta\sigma^1_1) + m/R^2]. \tag{2.50}$$

Finally substituting $e^{-\lambda} = (1 + U^2 - 2m/R)(R')^{-2}$ as mentioned earlier in (2.50) we obtain

$$DU = \left(1 + U^2 - \frac{2m}{R}\right) \times \left[\frac{\{(\partial/\partial R)(2\eta\sigma^1_1 R^3)\}_{,t} - (\partial\bar{p}/\partial R)_{,t}R^3}{(\rho + \bar{p} - 2\eta\sigma^1_1)R^3}\right] - \left[4\pi R(\bar{p} - 2\eta\sigma^1_1) + \frac{m}{R^2}\right], \tag{2.51}$$

and thus $K^{-}_{\tau\tau}$, as given in (2.49), can finally be written in the form

$$K^{-}_{\tau\tau} = (1 + U^2(r_\Sigma,t) - 2m(r_\Sigma,t)/R_\Sigma)^{-1/2} \times [DU + m/R^2 + 4\pi(\bar{p} - 2\eta\sigma^1_1)R\,]_{r=r_\Sigma}. \tag{2.52}$$

Calculating from (2.30) in the exterior

$$K^{+}_{\tau\tau} = \left(\frac{d\bar{t}}{d\tau}\right)\left(\frac{d^2\bar{r}_\Sigma}{d\tau^2}\right) - \left(\frac{d\bar{r}_\Sigma}{d\tau}\right)\left(\frac{d^2\bar{t}}{d\tau^2}\right) - \frac{3M}{\bar{r}^2_\Sigma}\left(1 - \frac{2M}{\bar{r}_\Sigma}\right)\left(\frac{d\bar{t}}{d\tau}\right)\left(\frac{d\bar{r}_\Sigma}{d\tau}\right)^2 + \frac{M}{\bar{r}^2_\Sigma}\left(1 - \frac{2M}{\bar{r}_\Sigma}\right)\left(\frac{d\bar{t}}{d\tau}\right)^3. \tag{2.53}$$

Since at the boundary

$$DU = \frac{dU}{d\tau} = \frac{d^2R_\Sigma}{d\tau^2} = \frac{d^2\bar{r}_\Sigma}{d\tau^2}$$

Eq. (2.42) being substituted in (2.53) yields

$$K^{+}_{\tau\tau} = [1 - 2M/\bar{r}_\Sigma + U^2(r_\Sigma,t)]^{-1/2}(DU + M/\bar{r}^2_\Sigma). \tag{2.54}$$

Matching of (2.52) and (2.54) thus immediately leads to the conclusion that at the boundary $\bar{p} - 2\eta\sigma^1_1 = 0$ and so we do not get any further constraint at the boundary from the conditions of fit of $K_{\tau\tau}$ component of the extrinsic curvature.

III. THE FLUID WITH $\eta = 0$

In this section we consider a fluid in which the bulk viscosity is nonvanishing whereas the shear viscosity is negligibly small. The relations (2.26) and (2.27) in this case reduce to

$$\dot{E} = 2\pi e^{\nu/2}R^3\sigma^1_1(\rho + \bar{p}) \tag{3.1}$$

and

$$E' = (4\pi/3)\rho' R^3, \tag{3.2}$$

respectively. Remembering that the shear tensor has only one independent nonzero component, namely, $\sigma^1_1 = -2\sigma^2_2 = -2\sigma^3_3$, it follows from (3.1) that in the isotropic case σ^1_1 vanishes, so that we have $\dot{E} = 0$, i.e., the free gravitational energy becomes time independent. On the other hand, if the space-time described by the metric (2.1) is conformally flat, i.e., if $E = 0$, it follows from (3.1) and (3.2) that the matter density is spatially uniform and the motion is shear-free. The third case is that when the matter density is spatially uniform we have $\rho' = 0$ and this leads, in view of (3.2), to the result $E' = 0$. Therefore E is a function of time alone and the condition of regularity at the center of the distribution demands that E must vanish. It follows from the fact that for regularity at $r \to 0$, we must have $R \to 0$, which in turn requires $E = -\psi_2 R^3$ vanish at the center, being finite there. Since $E = 0$ at $r = 0$ always and $E' = 0$ for uniform density distribution we have $E = 0$ throughout the interior. This leads to the vanishing of shear also. Thus we get the result that a spherically symmetric uniform density fluid with nonzero bulk viscosity but negligible shear viscosity must be isotropic as well as conformally flat. This is an interesting result that generalizes the same theorem proved by Misra and Srivastava for a perfect fluid. The present procedure giving the proof reveals another feature that the spherically symmetric distribution in this case is also conformally flat, which is not apparent in the work of Misra and Srivastava.

IV. UNIFORM DENSITY IMPERFECT FLUID WITH $\eta \neq 0$

The integration of Eq. (2.27) in general yields

$$E = -4\pi\eta\sigma^1_1 R^3 + E_0(t), \tag{4.1}$$

where $E_0(t)$ is an arbitrary function of time. From the regularity conditions at the center the first term in (4.1) on the right-hand side vanishes at $r = 0$. Again $E = 0$ at $r = 0$ by the arguments given in Sec. III.

Since E_0 is a function of time alone it is zero throughout the interior, so that one obtains in general the relation

$$E= -4\pi\eta\sigma_1^1 R^3. \tag{4.2}$$

Remembering the definition of E we get immediately from (4.2) for the Weyl invariant

$$\psi_2 = 4\pi\eta\sigma_1^1. \tag{4.3}$$

Again from (2.26) we have

$$(E + 4\pi\eta\sigma_1^1 R^3)^{\cdot} = -(4\pi/3)[\dot\rho + (\rho+\bar p)3\dot\mu/2 - 3\eta\sigma_1^1\dot\mu]R^3. \tag{4.4}$$

Now rearranging the right-hand side of (4.4), multiplying by $e^{-\nu/2}$, and using (2.4)–(2.6), the above equation may be written as

$$\begin{aligned} e^{-\nu/2}&(E + 4\pi\eta\sigma_1^1 R^3)^{\cdot} \\ &= -(4\pi/3)\big[\dot\rho e^{-\nu/2} + (\rho+\bar p)\theta - \tfrac{3}{2}(\rho+\bar p)\sigma_1^1 \\ &\quad - 2\eta\sigma_1^1\theta + 4\eta\sigma^2\big]R^3. \end{aligned} \tag{4.5}$$

Now substituting (4.2) in the left-hand side and using the Bianchi identity (2.7), we obtain

$$\tfrac{16}{3}\eta\sigma^2 = (\rho + \bar p + \tfrac{4}{3}\eta\theta)\sigma_1^1. \tag{4.6}$$

For nonvanishing shear the relation (4.6) can also be written as

$$4\eta\sigma_1^1 = (\rho+\bar p) + \tfrac{4}{3}\eta\theta. \tag{4.7}$$

The relation (4.7) gives us the information that, at least near the center, the expansion scalar θ is negative or, in other words, the core contracts. This is because at $r\to 0$, the local flatness demands $\sigma_1^1 = 0$ (see Ref. 9) and thus for positive $(\rho+\bar p)$ and η, only the contraction is possible in this region.

Next we eliminate the term like $(\rho+\bar p)$ between (2.7) and (4.6) and arrive at the relation

$$\dot\rho\sigma_1^1 = 3\eta e^{-\nu/2}\dot\mu^2\sigma_1^1 \tag{4.8}$$

when the shear vanishes, i.e., for isotropic motion both sides in (4.8) identically vanish. But, for $\sigma_1^1 \neq 0$, (4.8) leads to the nontrivial relation

$$\dot\rho = 3\eta e^{-\nu/2}\dot\mu^2, \tag{4.9}$$

which indicates that for anisotropic motion of the spherically symmetric viscous fluid and for the viscosity coefficient having a physically reasonable behavior ($\eta > 0$) the spatially uniform matter density can only increase with time. This monotonic character of the density variation with time in turn leads to a monotonic time behavior of another geometric quantity such as the surface area of the sphere. This can be explicitly shown as follows.

From (2.26) and (4.2) one obtains

$$\dot\rho/(\rho+\bar p - 2\eta\sigma_1^1) = -3\dot R/R. \tag{4.10}$$

So in order that the strong energy condition is satisfied, that is, $(\rho+\bar p - 2\eta\sigma_1^1) > 0$, and since $\dot\rho > 0$ one arrives at another conclusion—in this case the surface area of the sphere at each value of the radial coordinate monotonically decreases with time. This exhibits the irreversible nature of the motion in case the fluid is dissipative.

Lastly it may be of some interest to investigate what happens to the expansion scalar for a uniform density viscous fluid. We may recall the theorem (Raychaudhuri[12]) for a perfect fluid in general that uniform density fluid with irrotational motion must have spatially uniform expansion scalar. When the fluid is dissipative the above theorem is not valid. Taking the derivative of Eq. (2.7) with respect to the radial coordinate one obtains

$$\begin{aligned}(\rho+\bar p)\theta' &+ (\rho+\bar p')\theta \\ &= e^{-\nu/2}\dot\rho' + e^{-\nu/2}(\nu'/2)\dot\rho + (4\eta\sigma^2)',\end{aligned}$$

which, for uniform density ($\rho' = 0$) in view of the field equation (2.8), leads to the relation

$$\begin{aligned}(\rho+\bar p)\theta' &= \Big[-(2\eta\sigma_1^1)' - 2\eta\sigma_1^1\Big(\frac{3\mu'}{2} + \frac{\nu'}{2}\Big)\Big]\theta \\ &\quad + 4\eta\sigma^2\frac{\nu'}{2} + (4\eta\sigma^2)'.\end{aligned} \tag{4.11}$$

So, for $\eta\neq 0$, $\theta'\neq 0$, in general. On the other hand, when $\eta = 0$, $\theta' = 0$, even if the bulk viscosity is present. So, in fact, Raychaudhuri's result for a perfect fluid can be extended for a dissipative fluid with bulk viscosity only, at least in the spherically symmetric case.

V. CONCLUSION

In this paper we have discussed the matching conditions at the boundary of a sphere consisting of a viscous fluid and have also derived a few general relations in a dissipative medium following the line shown by Glass. A few exact solutions in different cases will be given in a subsequent paper, where the cosmological constant may be interpreted as the vacuum energy in the inflationary scenario.

[1]C. W. Misner, Astrophy. J. **151**, 431 (1968).
[2]Z. Klimek, Acta Cosmologica **2**, 49 (1973).
[3]G. Murphy, Phys. Rev. D **8**, 4231 (1973).
[4]U. A. Belinsky and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976).
[5]J. D. Nightingale, Astrophys. J. **185**, 105 (1973).
[6]C. W. Misner and D. H. Sharp, Phys. Rev. B **136**, 571 (1964).
[7]E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[8]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[9]W. Israel, Nuovo Cimento B **44**, 1 (1966); **48**, 463 (1967).
[10]R. M. Misra and D. C. Srivastava, Phys. Rev. D **8**, 1653 (1973).
[11]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, NJ, 1949).
[12]A. K. Raychaudhuri, *Theoretical Cosmology* (Clarendon, Oxford, 1979).

Initial value problem for colliding gravitational plane waves. I

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 10 August 1988; accepted for publication 7 December 1988)

A new form of the general solution of the initial value problem for colliding gravitational plane waves with collinear polarization is obtained. The solution of the linear hyperbolic field equation for $\psi(u,v)$ is expressed as a linear superposition $\int d\sigma\, g(\sigma)\omega(u,v,\sigma)$ of a one-parameter family of basic solutions of the form $\omega(u,v,\sigma) = \omega_1(u,\sigma)\omega_2(v,\sigma)$, where u and v are arbitrary null coordinates and σ is the parameter. The coefficients $g(\sigma)$ in this superposition are expressed in terms of the initial data by using a generalization of an integral transform obtained by Abel in his solution of a tautochrone problem of classical particle mechanics.

I. INTRODUCTION

A. Objective

This is the first of several papers on the initial value problem for two colliding gravitational plane waves in the Einstein theory, i.e., on the search for a systematic method of solving the Einstein field equations for the outcoming scattered wave when the incoming plane waves are prescribed.

In our first paper, we shall obtain a complete solution of the initial value problem for the general case of two colliding gravitational plane waves with collinear polarizations. This only requires that one find the solution of a certain linear hyperbolic partial differential equation of the second order. A general solution of that equation has already been given by Szekeres[1] and, in a slightly different form, by Xanthopoulos,[2] who both use a method of Riemann to obtain their results. However, we have not been able to employ their solution as a starting point for creating a viable method of handling the initial value problem for colliding plane waves with noncollinear polarizations.

That is why we hunted for another approach. The structure of our solution is different from that found by the method of Riemann and is designed to pave the way for a formalism that we shall present in the third paper of this series and that is applicable to noncollinear polarizations.

The method that we shall employ here to construct our solution of the collinear problem is the classical one of first finding a one-parameter family of basic solutions by means of a suitable separation of variables in the partial differential equation and then expressing the final solution as a linear superposition of these basic solutions. A new and interesting feature of the method is that the parameter-dependent coefficients in this linear superposition are obtained in terms of the prescribed initial data by solving generalized versions of Abel's integral equation[3] for the tautochrone problem of classical particle mechanics.

B. The solution

Let us become more specific. The line element in the space-time region occupied by the scattered wave is, in the collinear case,

$$ds^2 = \rho[e^{-2\psi}(dx^1)^2 + e^{2\psi}(dx^2)^2] - (2/\sqrt{\rho})e^{2\Gamma}\, du\, dv\,, \tag{1.1}$$

where the metrical fields, ρ, ψ, and Γ depend only on the coordinates u and v over a simply connected planar domain consisting of all (u,v) such that

$$0\leqslant u < u_0,\quad 0\leqslant v < v_0,\quad 0 < \rho(u,v) \leqslant 1.$$

Here, as we shall show in Sec. II,

$$\rho(u,v) = \tfrac{1}{2}[s(v) - r(u)]\,, \tag{1.2}$$

where $r(u)$ is a monotonic increasing function of u over the interval $0\leqslant u < u_0$, $s(v)$ is a monotonic decreasing function of v over $0 \leqslant v < v_0$, and

$$r(0) = -1,\quad s(0) = 1,\quad r(u) < s(v). \tag{1.3}$$

The field $\Gamma(u,v)$ is simply expressed in terms of definite integrals once $\psi(u,v)$ is known; the expression for $\Gamma(u,v)$ will be given in Sec. II.

The key problem is to find the solution ψ of the linear hyperbolic field equation ($\psi_u := \partial\psi/\partial u$, etc.),

$$2\rho\psi_{uv} + \rho_u\psi_v + \rho_v\psi_u = 0\,, \tag{1.4}$$

corresponding to the prescribed initial data $r(u)$, $s(v)$, $\psi(u,0)$, and $\psi(0,v)$. The ignorable coordinates x^1,x^2 are scaled so that $\psi(0,0) = 0$. Our solution will be discussed in Sec. IV and is given by

$$\psi = {}_3\psi + {}_2\psi\,, \tag{1.5}$$

where

$${}_3\psi(u,v) = \frac{1}{\pi}\int_{-1}^{r(u)} d\sigma \frac{g_3(\sigma)\sqrt{1-\sigma}}{\sqrt{[r(u)-\sigma][s(v)-\sigma]}} \tag{1.6}$$

and

$${}_2\psi(u,v) = \frac{1}{\pi}\int_{1}^{s(v)} d\sigma \frac{g_2(\sigma)\sqrt{1+\sigma}}{\sqrt{[\sigma-s(v)][\sigma-r(u)]}} \tag{1.7}$$

and where the parameter-dependent coefficients $g_3(\sigma)$ and $g_2(\sigma)$ in the above linear superposition are to be determined from the prescribed initial data and the requirements [see Eqs. (1.3)]

 0022-2488/89/040872-16$02.50

$$
{}_3\psi(0,v) := \lim_{u\to 0} {}_3\psi(u,v) = 0\,,
$$
$$
{}_2\psi(u,0) := \lim_{v\to 0} {}_2\psi(u,v) = 0\,. \tag{1.8}
$$

In fact, upon setting $v=0$ in Eq. (1.6) and $u=0$ in Eq. (1.7), one obtains, with the aid of Eqs. (1.3), (1.5), and (1.8),

$$
\psi(u,0) = \frac{1}{\pi}\int_{-1}^{r(u)} d\sigma \frac{g_3(\sigma)}{\sqrt{r(u)-\sigma}}\,,
$$
$$
\psi(0,v) = \frac{1}{\pi}\int_{1}^{s(v)} d\sigma \frac{g_2(\sigma)}{\sqrt{\sigma - s(v)}}\,. \tag{1.9}
$$

The above Eqs. (1.9) are the generalized[4] Abel integral equations that we mentioned previously. For C^1-differentiable $r(u)$, $s(v)$, $\psi(u,0)$, and $\psi(0,v)$, we shall prove in Sec. III that the solutions of these integral equations are

$$
g_3(\sigma) = \int_0^{u_\sigma} du \frac{\psi_u(u,0)}{\sqrt{\sigma - r(u)}}\,,
$$
$$
g_2(u) = -\int_0^{v_\sigma} dv \frac{\psi_v(0,v)}{\sqrt{s(v)-\sigma}}\,, \tag{1.10}
$$

where u_σ and v_σ are defined by $r(u_\sigma)=\sigma$ and $s(v_\sigma)=\sigma$, respectively. Thus $g_j(\sigma)$ $(j=3,2)$ are determined from the initial data by Eqs. (1.10) and the scattered wave potential $\psi(u,v)$ is then given by Eqs. (1.6) and (1.7). The existence of all integrals appearing above will be proven in Secs. III and IV.

If the first derivatives $\dot r(u)$ and $\dot s(v)$ are not zero for all $0\leqslant u<u_0$ and $0\leqslant v<v_0$, respectively, then Eqs. (1.9) and (1.10) become conventional Abel transforms after the coordinate transformations $u\to r$ and $v\to s$. However, we shall prove in Sec. II that the vacuum field equations for colliding plane wave metrics imply that $\dot r(0)=\dot s(0)=0$ [though $\dot r(u)>0$ and $\dot s(v)<0$ for other u and v] and that $|\psi_u(u,0)/\dot r(u)|$ is unbounded as $u\to 0$ and $|\psi_v(0,v)/\dot s(v)|$ is unbounded as $v\to 0$. Therefore, for colliding gravitational plane waves, Eqs. (1.9) and (1.10) are not covered by the usual theorems on Abel transforms, and it will be necessary for us to prove that Eqs. (1.9) imply Eqs. (1.10) and vice versa. This will be done in Sec. III.

C. Colliding wave conditions

Section II will cover some salient points concerning the field equations and, since it can be done without additional complications and with benefit to later papers of this series, we shall not assume that the polarizations are collinear in Sec. II. When the polarizations are not collinear, the key field Eq. (1.4) is replaced by the nonlinear Ernst equation[5] whose solution is a complex potential $E(u,v)$. The idea is to seek $E(u,v)$ corresponding to the prescribed initial data $r(u)$, $s(v)$, $E(u,0)$, and $E(0,v)$. When E is real, one lets $\psi := -\frac{1}{2}\ln E$ and we are back to the collinear polarization Eq. (1.4).

We must caution, however, that one cannot arbitrarily choose $r(u)$, $s(v)$, $E(u,0)$, and $E(0,v)$ and expect the corresponding solution of the Ernst equation to yield a colliding gravitational plane wave metric. In Sec. II, we obtain simple criteria that must be satisfied by $r(u)$, $s(v)$, $E(u,0)$, and $E(0,v)$ if they are to determine a colliding gravitational plane wave metric. These criteria are generalizations of colliding wave conditions previously given by Ernst, García Diaz, and Hauser[6] for the special case when $r(u)=2u^2-1$ and $s(v)=1-2v^2$. Various errors in the derivation of the colliding wave condition of the aforementioned paper[6] will be corrected in Sec. II of this paper.

II. THE FIELD EQUATIONS AND THE COLLIDING WAVE CONDITIONS

A. The chart

For any space-time to be considered in this series of papers, there exist coordinates x^1,x^2,u,v such that the line element has the form

$$
ds^2 = \hat\rho \hat F^{-1}|\hat E\,dx^1 + i\,dx^2|^2 - \frac{2}{\sqrt{\hat\rho}} e^{2\hat\Gamma}\,du\,dv\,, \tag{2.1}
$$

where

$$
\hat F := \operatorname{Re}\hat E > 0\,,\quad \hat\rho > 0\,, \tag{2.2}
$$

and where the real fields $\hat\rho$ and $\hat\Gamma$ and the complex field $\hat E$ depend at most on u and v. The domain of the chart (which should not be assumed to be global) is the set of all (x^1,x^2,u,v) such that $(x^1,x^2)\in R^2$ and such that

$$
(u,v)\in(\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV})\,,
$$

where I, II, III, and IV are the contiguous planar regions:

$$
\begin{aligned}
&\mathrm{I} := \{(u,v)\in R^2 : u\leqslant 0,\ v\leqslant 0\}\,,\\
&\mathrm{II} := \{(u,v)\in R^2 : u\leqslant 0,\ 0\leqslant v<v_0\}\,,\\
&\mathrm{III} := \{(u,v)\in R^2 : v\leqslant 0,\ 0\leqslant u<u_0\}\,,\\
&\mathrm{IV} := \{(u,v)\in R^2 : 0\leqslant u<u_0,\ 0\leqslant v<v_0,\ 0<\hat\rho(u,v)\}\,.
\end{aligned} \tag{2.3}
$$

Here, u_0 and v_0 are positive real numbers. They are the right-hand end points of those open intervals of the u axis and v axis, respectively, on which $\hat\rho(u,0)$, $\hat E(u,0)$, and $\hat\rho(0,v)$, $\hat E(0,v)$ are defined and satisfy the inequalities

$$
\hat\rho(u,0)>0\,,\quad \hat F(u,0)>0\,,\quad \hat\rho(0,v)>0\,,\quad \hat F(0,v)>0\,,
$$

and other requisite conditions that will be given one at a time as we proceed in Sec. II.[7] Also, further details respecting region IV will be given later in Sec. II.

We shall always scale the Killing vectors $\partial/\partial x^1$ and $\partial/\partial x^2$ so as to make

$$
\hat\rho(0,0) = 1\,. \tag{2.4}
$$

That still enables us to subject the Killing vectors to any $SL(2,R)$ transformation

$$
\begin{pmatrix} \dfrac{\partial}{\partial x^1} \\[2mm] \dfrac{\partial}{\partial x^2}\end{pmatrix} \to \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\begin{pmatrix} \dfrac{\partial}{\partial x^1} \\[2mm] \dfrac{\partial}{\partial x^2}\end{pmatrix}, \tag{2.5}
$$
$$
\alpha\delta - \beta\gamma = 1\,.
$$

This leaves $\hat\rho(u,v)$ and $\hat\Gamma(u,v)$ invariant but transforms $\hat E(u,v)\to\hat E'(u,v)$, where

$$
\hat E' = i[(\alpha\hat E + i\beta)/(\gamma\hat E + i\delta)]\,. \tag{2.6}
$$

We can use this tranformation to make

$$
\hat E(0,0) = 1\,. \tag{2.7}
$$

The option (2.7) will not be adopted until Sec. III of this paper.

B. Restrictions of the metrical fields to regions I, II, III, and IV

It is assumed, of course, that $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$ are continuous throughout $\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV}$. It is further assumed *as part of the definition of colliding gravitational plane waves* that u and v can be chosen so that the restrictions of $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$ to region I are independent of both u and v, while the restrictions of these functions to region II are independent of u and their restrictions to III are independent of v. These assumptions are concisely expressed[8] by the following equations in which (u,v) is any point in $\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV}$:

$$\begin{aligned}\hat{\rho}(u,v) &= \hat{\rho}(u\theta(u),v\theta(v)),\\ \hat{E}(u,v) &= \hat{E}(u\theta(u),v\theta(v)),\\ \hat{\Gamma}(u,v) &= \hat{\Gamma}(u\theta(u),v\theta(v)).\end{aligned} \tag{2.8}$$

Here, $\theta(x)$ is the Heavyside unit symbol, which is here defined for all real x by the equations[8]

$$\theta(x) = \begin{cases} 0, & \text{if } x\leqslant 0,\\ 1, & \text{if } x\geqslant 0.\end{cases} \tag{2.9}$$

In view of Eqs. (2.3), (2.8), and (2.9), the restriction of the line element (2.1) to that subset of the domain of the chart for which $(u,v)\in\mathrm{I}$ is Minkowskian. This becomes clear if one uses Eq. (2.4) and the option (2.7). It follows that the conform tensor as well as the Ricci tensor vanishes at all spacetime points such that (u,v) lies in the interior of I.

On the other hand, the restrictions of (2.1) to $(u,v)\in\mathrm{II}$ and to $(u,v)\in\mathrm{III}$ are both Rosen[9] forms of line elements for plane gravitational waves. Note that

$$\mathrm{I}\cap\mathrm{II} = \{(u,v): v=0,\ u\leqslant 0\}$$

and

$$\mathrm{I}\cap\mathrm{III} = \{(u,v): u=0,\ v\leqslant 0\}.$$

It is assumed *as part of the definition of colliding gravitational plane waves* that the origin of the (u,v) plane can be chosen so that the two null hypersurfaces $v=0$, $u<0$ and $u=0, v<0$ are wave fronts. We shall not attempt to define the general concept of a wave front but we shall need the following specialized definitions.

Definitions: The hypersurface $v=0$, $u<0$ will be called a *wave front* if no $\epsilon>0$ exists such that the conform tensor exists and vanishes at all (u,v) for which $-\infty<v<\epsilon$ and $u<0$. The hypersurface $u=0, v<0$ will be called a *wave front* if no $\epsilon>0$ exists such that the conform tensor exists and vanishes at all (u,v) for which $-\infty<u<\epsilon$ and $v<0$.

Note that the above definitions cover those colliding gravitational plane wave metrics[6,8,10–14] for which the plane waves associated with regions II and III have conform tensor terms which are nonzero multiples of the Dirac delta symbols $\delta(v)$ and $\delta(u)$, respectively. For these metrics, the conform tensor does not exist on the hypersurfaces $v=0$, $u<0$ and $u=0$, $v<0$.

Henceforth, it will be assumed that the null hypersurfaces $v=0$, $u<0$ and $u=0$, $v<0$ are wave fronts as defined above. In terms descriptive of temporal order, they are regarded as the fronts of the plane waves *prior to the collision*. The set of all events for which $u<0$ and $v<0$, i.e., for which (u,v) lies in the interior of I, is regarded as a flat space-time region that lies between the plane waves *prior to the collision*. The spacelike two-surface $u=v=0$ is identified as the set of all events *at the instant of collision*.

The *line element of the scattered wave* is, by definition, the restriction of (2.1) to $(u,v)\in\mathrm{IV}$.

Definitions: ρ, E, and Γ will denote the restrictions of $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$, respectively, to IV.

Note that Eqs. (2.8) can be expressed[8] as follows:

$$\begin{aligned}\hat{\rho}(u,v) &= \rho(u\theta(u),v\theta(v)),\\ \hat{E}(u,v) &= E(u\theta(u),v\theta(v)),\\ \hat{\Gamma}(u,v) &= \Gamma(u\theta(u),v\theta(v)),\end{aligned} \tag{2.10}$$

for all (u,v) in $\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV}$.

C. Derivatives of the metrical fields

We recall that the continuity and the derivatives of a function are defined using only such values of its arguments as do not take one outside the domain of the function. For example, let E_{III} denote the restriction of $\hat{E}$ to III. Then

$$(E_{\mathrm{III}})_v(u,0) := \left(\frac{\partial E_{\mathrm{III}}(u,v)}{\partial v}\right)_{v=0} \quad \text{where } 0\leqslant u<u_0$$

is defined as the limit of

$$\Delta(u,h) := [\hat{E}(u,h) - \hat{E}(u,0)]h^{-1} \quad (h\neq 0)$$

as $h\to 0$ through negative values, i.e., as $h\to 0$ while (u,h) remains a member of III (other than $(u,0)$). Note that Eq. (2.8) implies that $(E_{\mathrm{III}})_v(u,v)=0$ for all (u,v) in III.

In contrast, recall that E denotes the restriction of $\hat{E}$ to IV. For $0\leqslant u<u_0$ $E_v(u,0)$ is defined as the limit of the above $\Delta(u,h)$ as $h\to 0$ through positive values, i.e., (u,h) remains a member of IV. Thus $E_v(u,0)$ may differ from 0 if it exists.

Finally, $\hat{E}_v(u,0)$ for $0\leqslant u<u_0$ is defined as the limit of the above $\Delta(u,h)$ as $h\to 0$ without constraint on the sign of h. From the statements in the preceding two paragraphs, *$\hat{E}_v(u,0)$ exists if and only if $E_v(u,0)$ exists and equals* 0. If $E_v(u,0)$ exists but does not equal 0, then $\hat{E}_v(u,v)$ has a finite step discontinuity at $v=0$. This possibility is allowed by the following premises that we shall adopt throughout the remainder of Sec. II.

Premises: ρ and E are C^2 (i.e., belong to the differentiability class C^2)[15] and Γ is C^1.

The premises that ρ and E are C^2 will be relaxed in our treatment of the initial value problem in Secs. III and IV.

From Eqs. (2.3) and (2.10), the restrictions of $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$ to II have the values

$$\hat{\rho}(u,v) = \rho(0,v),\quad \hat{E}(u,v) = E(0,v),\quad \hat{\Gamma}(u,v) = \Gamma(0,v),$$

for $u\leqslant 0$ and $0\leqslant v<v_0$, i.e., $(u,v)\in\mathrm{II}$.

Therefore, the restrictions of $\hat{\rho}$ and $\hat{E}$ to II have continuous first and second derivatives with respect to v, and the restriction of $\hat{\Gamma}$ to II has a continuous first derivative with respect to v. Likewise, the restrictions of $\hat{\rho}$ and $\hat{E}$ to III have continuous first and second derivatives with respect to u, and the restriction of $\hat{\Gamma}$ to III has a continuous first derivative with respect to u.

We next discuss the derivatives of $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$. As we have already indicated, some of the first and/or second derivatives of $\hat{\rho}$, $\hat{E}$, and $\hat{\Gamma}$ may not exist at points on the interfaces of I, II, III, and IV and it can happen that one or more of these derivatives occur in the expressions for the Ricci tensor components. Nevertheless, we shall want to apply that vacuum condition which asserts that the Ricci tensor must vanish throughout I∪II∪III∪IV including at all interface points. There is clearly a problem here.

One standard way of handling this problem is to broaden the definitions (for our particular class of metrics) of "connection form" and "Riemann tensor" by permitting the former to contain terms that are linear combinations of the Heavyside unit symbols $\theta(u)$ and $\theta(v)$ and by permitting the latter to contain terms that are linear combinations of the Dirac delta symbols $\delta(u)$ and $\delta(v)$, other terms that are linear combinations of $\theta(u)$ and $\theta(v)$, and still another term that is a multiple of the product $\theta(v)\theta(v)$. The new "connection form" and "Riemann tensor" are thus symbolic (generalized) functions and will fittingly be called *symbolic connection form* and *symbolic Riemann tensor*. We shall now specify how these objects are constructed.

An acquaintance with the theory of distributions of Schwartz[16] is most helpful for understanding what we are about to do. However, a working knowledge[17] of the Dirac delta symbol and of a few symbolic equalities (an equivalence relation for symbolic functions) such as

$$\delta(u) = \frac{d\theta(u)}{du}, \quad u\delta(u) = 0, \tag{2.11}$$

will suffice. In fact, all we need at present are the following (symbolic) derivatives that are derived from Eq. (2.10) with the aid of Eqs. (2.11) and a known generalization of the chain rule of differentiation:

$$\begin{aligned}
\hat{E}_u(u,v) &= \theta(u)E_u(u\theta(u),v\theta(v)),\\
\hat{E}_v(u,v) &= \theta(v)E_v(u\theta(u),v\theta(v)),\\
\hat{E}_{uv}(u,v) &= \theta(u)\theta(v)E_{uv}(u\theta(u),v\theta(v)),\\
\hat{E}_{uu}(u,v) &= \delta(u)E_u(u\theta(u),v\theta(v))\\
&\quad + \theta(u)E_{uu}(u\theta(u),v\theta(v)),\\
\hat{E}_{vv}(u,v) &= \delta(v)E_v(u\theta(u),v\theta(v))\\
&\quad + \theta(v)E_{vv}(u\theta(u),v\theta(v)),
\end{aligned} \tag{2.12}$$

and likewise for $\hat{\rho}$ and $\hat{\Gamma}$.

The *symbolic connection forms* and the *symbolic Riemann tensor* can now be defined as the results that are obtained by formally entering $\hat{\rho}$, $\hat{E}$, $\hat{\Gamma}$, and their derivatives as exemplified by Eqs. (2.12) into any of the usual expressions for the connection forms and the Riemann tensor.[18] A procedure that is used by the authors is to introduce the null tetrad of one-forms

$$\begin{aligned}
&k = -e^{\hat{\Gamma}}\rho^{-1/4}\,dv, \quad m = e^{\hat{\Gamma}}\rho^{-1/4}\,du,\\
&t = \sqrt{\hat{\rho}/2\hat{F}}\,(\hat{E}\,dx^1 + i\,dx^2), \quad t^* = \text{c.c. of } t,
\end{aligned} \tag{2.13}$$

and then employ exterior algebra and differentiation to compute the corresponding symbolic connection forms and null tetrad components of the symbolic connection forms and null tetrad components of the symbolic Riemann tensor directly in terms of $\hat{\rho}$, $\hat{E}$, $\hat{\Gamma}$, and their derivatives.[19–21]

Suppose one of the components of the symbolic Ricci tensor is formally set equal to zero for all (u,v) in I∪II∪III∪IV. This is understood to mean that the component is symbolically equal to a (true) function whose domain is I∪II∪III∪IV and whose value at every point of the domain is zero. In this way one can give meaning to the vanishing of the matter tensor even when the metrical functions have finite step discontinuities at the interfaces. Let us now see how this works out in practice.

D. The vacuum field equations

By using the procedure[19–21] mentioned above, one obtains the following complete family of vacuum field equations for the metric (2.1):

$$\hat{\rho}_{uv} = 0, \tag{2.14}$$

the Ernst equation,

$$\hat{F}(2\hat{\rho}\hat{E}_{uv} + \hat{\rho}_u\hat{E}_v + \hat{\rho}_v\hat{E}_u) = 2\hat{\rho}\hat{E}_u\hat{E}_v, \tag{2.15}$$

the pair of equations,

$$\begin{aligned}
\hat{\rho}_{uu} - 2\hat{\rho}_u\hat{\Gamma}_u &= -2\hat{\rho}\left|\frac{\hat{E}_u}{2\hat{F}}\right|^2,\\
\hat{\rho}_{vv} - 2\hat{\rho}_v\hat{\Gamma}_v &= -2\hat{\rho}\left|\frac{\hat{E}_v}{2\hat{F}}\right|^2;
\end{aligned} \tag{2.16}$$

and, granted the existence and continuity of Γ_{uv} throughout IV,

$$\hat{\Gamma}_{uv} = -\mathrm{Re}\left[\left(\frac{\hat{E}_u}{2\hat{F}}\right)\left(\frac{\hat{E}_v}{2\hat{F}}\right)^*\right]. \tag{2.17}$$

As we shall prove later in Sec. II, the existence and continuity of Γ_{uv} throughout IV and, also, Eq. (2.17) are implied by Eqs. (2.14)–(2.16).

We shall now consider the above symbolic Eqs. (2.14)–(2.16) one at a time with the objective of reducing them to an equivalent family of equations and conditions involving only (true) functions.

E. Solution of $\hat{\rho}_{uv} = 0$ and definitions of r and s

From Eqs. (2.12), one sees that Eq. (2.14) is equivalent to the equation

$$\theta(u)\theta(v)\rho_{uv}(u\theta(u),v\theta(v)) = 0$$

for all (u,v) in I∪II∪III∪IV. This is in turn, as can be seen from Eqs. (2.9), equivalent to the equation

$$\rho_{uv}(u,v) = 0 \text{ for all } (u,v)\in\text{IV}.$$

The solution of the above equation subject to the condition (2.4) is easily shown to be

$$\rho(u,v) = \tfrac{1}{2}[s(v) - r(u)], \tag{2.18}$$

where

$$\begin{aligned}
r(u) &:= 1 - 2\rho(u,0) \text{ for all } 0 \leqslant u < u_0,\\
s(v) &:= 2\rho(0,v) - 1 \text{ for all } 0 \leqslant v < v_0,
\end{aligned} \tag{2.19}$$

and

$$r(0) = -1, \quad s(0) = 1. \tag{2.20}$$

For given $r(u)$ and $s(v)$, the definition of IV given in Eqs. (2.3) becomes

$$\mathrm{IV}:=\{(u,v)\in R^2:0\leqslant u<u_0,\ 0\leqslant v<v_0,\ r(u)<s(v)\}. \tag{2.21}$$

We shall soon have more to say about the properties of $r(u)$ and $s(v)$.

F. The Ernst equation

By reasoning which employs Eqs. (2.12) and is exactly analogous to that used above for $\hat{\rho}_{uv}=0$, one shows that the Ernst symbolic Eq. (2.15) is equivalent to the Ernst equation

$$F(2\rho E_{uv}+\rho_u E_v+\rho_v E_u)=2\rho E_u E_v \tag{2.22}$$

over region IV.

G. Definitions of K, L and the field equations for Γ

Let K and L denote functions that each have the domain IV and are defined by

$$K:=\rho\left|\frac{E_v}{2F}\right|^2,\quad L:=\rho\left|\frac{E_u}{2F}\right|^2. \tag{2.23}$$

With the aid of Eqs. (2.9), (2.10), (2.12), and (2.18), a straightforward calculation reveals that the pair of Eqs. (2.16) is equivalent to the pair of symbolic equations

$$\begin{aligned}\tfrac{1}{2}\delta(u)\dot r(0)=\theta(u)[&-\tfrac{1}{2}\ddot r(u\theta(u))\\&+\dot r(u\theta(u))\Gamma_u(u\theta(u),v\theta))\\&+2L(u\theta(u),v\theta(v))],\\ \tfrac{1}{2}\delta(v)\dot s(0)=\theta(v)[&-\tfrac{1}{2}\ddot s(v\theta(v))\\&+\dot s(v\theta(v))\Gamma_v(u\theta(u),v\theta(v))\\&-2K(u\theta(u),v\theta(v))],\end{aligned}$$

for all (u,v) in I∪II∪III∪IV. It is then a simple exercise to prove that the above pair of symbolic equations is equivalent to the family of equations:

$$\dot r(0)=0,\quad \dot s(0)=0, \tag{2.24}$$

and, for all (u,v) in IV,

$$\begin{aligned}&-\dot r(u)\Gamma_u(u,v)+\tfrac{1}{2}\ddot r(u)=2L(u,v),\\&-\dot s(v)\Gamma_v(u,v)+\tfrac{1}{2}\ddot s(v)=-2K(u,v).\end{aligned} \tag{2.25}$$

We shall next prove that Eqs. (2.25) for $v=0$ and $u=0$, respectively, and the condition that the null hypersurfaces $v=0$, $u<0$ and $u=0$, $v<0$ are wave fronts imply that $r(u)$ is monotonic increasing and $s(v)$ is monotonic decreasing.

H. The wave front conditions

Lemma:

$$\int_0^v db\,K(0,b)>0,\quad \text{for all } 0<v<v_0 \tag{2.26}$$

and

$$\int_0^u da\,L(a,0)>0,\quad \text{for all } 0<u<u_0. \tag{2.27}$$

Proof: It is sufficient to prove (2.27) since the proof of (2.26) is similar. It is to be understood in the following proof that *no symbolic functions are admitted,* i.e., *all derivatives which occur in the proof are conventionally defined.*

Suppose that statement (2.27) is false. Then $\epsilon>0$ exists such that the integral in (2.27) vanishes when $u=\epsilon$. From the definition (2.23) of L and from the continuity of ρ, E, and E_u, it follows that $E_u(u,0)=0$ over $0\leqslant u<\epsilon$. Therefore, $\hat{E}_u(u,0)=0$ over $-\infty<u<\epsilon$. Now, the only component [relative to the null tetrad given by Eqs. (2.13)] of the conform tensor that does not necessarily exist and identically vanish in the interior of I∪III is[19–22]

$$C_2=\hat\rho^{-1}\left\{\frac{\partial}{\partial u}\left[\frac{\hat\rho^{3/2}\hat E_u\exp(-2\hat\Gamma)}{2\hat F}\right]-\frac{\hat\rho^{3/2}(\hat E_u-\hat E_u^*)\hat E_u\exp(-2\hat\Gamma)}{(2\hat F)^2}\right\}.$$

Note that the points (u,v) such that $-\infty<u<\epsilon$ and $v<0$ lie in the interior of I∪III and that, for these points,

$$\hat\rho(u,v)=\hat\rho(u,0),\quad \hat\Gamma(u,v)=\hat\Gamma(u,0),$$
$$\hat E(u,v)=\hat E(u,0).$$

It follows that $C_2(u,v)=0$ for $-\infty<u<\epsilon$ and $v<0$, i.e., the hypersurface $u=0$, $v<0$ is not a wave front. This contradicts one of our premises.

Therefore, statement (2.27) is true.

Theorem:

$$\dot r(u)>0,\quad \text{for all } 0<u<u_0 \tag{2.28}$$

and

$$\dot s(v)<0,\quad \text{for all } 0<v<v_0. \tag{2.29}$$

Proof: It is sufficient to prove (2.28) since the proof of (2.29) is similar. We set $v=0$ and multiply through by $\exp[-2\Gamma(u,0)]$ in the first of Eqs. (2.25), whereupon

$$(\tfrac{1}{2}\dot r(u)e^{-2\Gamma(u,0)})_u=2L(u,0)e^{-2\Gamma(u,0)}.$$

Therefore, from Eq. (2.24)

$$\frac{1}{2}\dot r(u)e^{-2\Gamma(u,0)}=2\int_0^u da\,L(a,0)e^{-2\Gamma(a,0)},$$

which, together with Eq. (2.27) and the mean value theorem for integrals, implies Eq. (2.28).

Note: We have proven that (2.28) and (2.29) are sufficient as well as necessary for the null hypersurfaces $u=0, v<0$ and $v=0, u<0$ to be wave fronts. However, that proof will not be given here since it is not relevant to the objective of our current series of papers on the initial value problem.

Corollary: For any given $(u,v)\in$IV, the closed rectangular region

$$\{(a,b)\in R^2:0\leqslant a\leqslant u, 0\leqslant b\leqslant v\} \tag{2.30}$$

is a subset of IV. (The rectangle degenerates to a line segment or to a singlet set when $u=0$ or $v=0$.)

Proof: Since $(u,v)\in$IV, Eq. (2.21) implies

$$0\leqslant u<u_0,\quad 0\leqslant v<v_0,\quad r(u)<s(v).$$

Since r is monotonic increasing and s is monotonic decreasing, we then have for all (a,b) such that $0\leqslant a\leqslant u$ and $0\leqslant b\leqslant v$,

$$0\leqslant a<u_0,\quad 0\leqslant b<v_0,\quad r(a)<s(b),$$

whereupon Eq. (2.21) implies $(a,b)\in$IV. That completes the proof.

I. $\Lambda(u,v)$, k, l and the colliding wave conditions

The preceding corollary implies the existence of the real invariant

$$\Lambda(u,v) := \mathrm{Re} \int_0^u da \int_0^v db \frac{E_a(a,b)E_b^*(a,b)}{[2F(a,b)]^2}, \quad (2.31)$$

for all $(u,v) \in \mathrm{IV}$. Note that Λ_{uv} exists and is continuous throughout IV and that

$$\Lambda(u,0) = 0, \quad \Lambda(0,v) = 0, \quad \Lambda_u(u,0) = 0, \quad \Lambda_v(0,v) = 0. \quad (2.32)$$

From Eqs. (2.15), (2.18), and (2.23), one obtains

$$K_u = -\tfrac{1}{2}\dot{s}\Lambda_{uv}, \quad L_v = \tfrac{1}{2}\dot{r}\Lambda_{uv}. \quad (2.33)$$

So, with the aid of Eqs. (2.32),

$$\begin{aligned} K(u,v) &= K(0,v) - \tfrac{1}{2}\dot{s}(v)\Lambda_v(u,v), \\ L(u,v) &= L(u,0) + \tfrac{1}{2}\dot{r}(u)\Lambda_u(u,v). \end{aligned} \quad (2.34)$$

The insertions of the above expressions into the field Eqs. (2.25) for Γ yield

$$\begin{aligned} \dot{r}(u)[\Gamma(u,v) + \Lambda(u,v)]_u &= \tfrac{1}{2}\ddot{r}(u) - 2L(u,0), \\ \dot{s}(v)[\Gamma(u,v) + \Lambda(u,v)]_v &= \tfrac{1}{2}\ddot{s}(v) + 2K(0,v). \end{aligned} \quad (2.35)$$

In our formulation of the initial value problem for colliding gravitational plane waves, one regards $r(u)$, $s(v)$, $E(u,0)$, and $E(0,v)$ as prescribed functions from which $E(u,v)$ and $\Gamma(u,v)$ are to be determined throughout region IV. It is understood, of course, that $r(u)$ and $s(v)$ satisfy the conditions (2.20), (2.24), (2.28), and (2.29).

Now Eqs. (2.35) reveal that not all choices of $r(u)$, $s(v)$, $E(u,0)$, and $E(0,v)$ which satisfy the conditions (2.20), (2.24), (2.28), and (2.29) are admissible. It is clear that necessary and sufficient conditions for continuously differentiable $\Gamma(u,0)$ and $\Gamma(0,v)$ to exist in region III and II, respectively, are that the limits

$$\lim_{u\to 0}\left[\frac{\tfrac{1}{2}\ddot{r}(u) - 2L(u,0)}{\dot{r}(u)}\right](u>0) \quad (2.36)$$

and

$$\lim_{v\to 0}\left[\frac{\tfrac{1}{2}\ddot{s}(v) + 2K(0,v)}{\dot{s}(v)}\right](v>0) \quad (2.37)$$

exist. One can readily prove that the *existence* of the above limits (but not necessarily the values of the limits) are invariant under any null coordinate transformation $u\to u'$, $v\to v'$, under any $SL(2,R)$ transformation of the Killing vectors $\partial/\partial x^i$ [i.e., any transformation $E\to E'$ as given by Eq. (2.6)] and under any Kramer–Neugebauer transformation as defined and discussed, for example, by Ernst, García Diaz, and Hauser.[6] Obvious implications of Eqs. (2.24), (2.36), and (2.37) are

$$l := L(0,0) = \tfrac{1}{4}\ddot{r}(0), \quad k := K(0,0) = -\tfrac{1}{4}\ddot{s}(0). \quad (2.38)$$

However, Eqs. (2.38) do not generally imply that the limits (2.36) and (2.37) exist.

Equations (2.24), (2.36), and (2.37) also imply that, if a,b,c,d are real numbers such that $c>a>0$ and $d>b>0$, then

$$\begin{aligned} &\int_a^c du \frac{L(u,0)}{\dot{r}(u)} \to \infty \text{ as } a\to 0, \\ &\int_b^d dv \frac{K(0,v)}{[-\dot{s}(v)]} \to \infty \text{ as } b\to 0. \end{aligned} \quad (2.39)$$

Therefore, from the definitions (2.23) of K and L, the ratios $|E_v(u,0)/\dot{r}(u)|$ and $|E_v(0,v)/\dot{s}(v)|$ are unbounded as $u\to 0$ and $v\to 0$, respectively. (This does not mean that the reciprocals of these ratios converge to 0 as $u\to 0$ and $v\to 0$.)

J. Solution for Γ in terms of E

Now, let us grant that the solution $E(u,v)$ of the Ernst equation corresponding to the prescribed $E(u,0)$ and $E(0,v)$ and to the prescribed $r(u)$ and $s(v)$ satisfying the conditions (2.20), (2.24), (2.28), (2.29), (2.36), and (2.37) has been found. Then Eqs. (2.32) and (2.35) yield the following solution for $\Gamma(u,v)$ throughout region IV:

$$\Gamma(u,v) = -\Lambda(u,v) + \Gamma(u,0) + \Gamma(0,v) - \Gamma(0,0), \quad (2.40)$$

where, if we wish, the value of $\Gamma(0,0)$ can always be made zero by an appropriate scaling of uv and where

$$\begin{aligned} \Gamma(u,0) &= \Gamma(0,0) + \int_0^u da \frac{\tfrac{1}{2}\ddot{r}(a) - 2L(a,0)}{\dot{r}(a)}, \\ \Gamma(0,v) &= \Gamma(0,0) + \int_0^v db \frac{\tfrac{1}{2}\ddot{s}(b) + 2K(0,b)}{\dot{s}(v)}. \end{aligned} \quad (2.41)$$

From Eqs. (2.31), (2.40), and (2.41), it is apparent as we stated early in Sec. II that Γ_{uv} exists and is continuous throughout IV. Furthermore, $\Gamma_{uv} = -\Lambda_{uv}$ throughout IV which implies [with the aid of Eqs. (2.12)] that Eq. (2.17) holds throughout $\mathrm{I}\cup\mathrm{II}\cup\mathrm{III}\cup\mathrm{IV}$.

K. The case $\rho(u,v) = 1-u^2-v^2$

For almost all of the colliding gravitational plane wave solutions[2,8,10–14] that have been explicitly given so far, null coordinates u and v exist such that

$$r(u) = 2u^2 - 1, \quad s(v) = 1 - 2v^2. \quad (2.42)$$

The class of colliding gravitational plane wave metrics for which Eqs. (2.42) hold was denoted by CW_1 in a paper by Ernst, García Diaz, and Hauser.[6] For this class, it is readily proven that the colliding wave conditions (2.36) and (2.37) are equivalent to the pair of conditions (2.38) and that the conditions (2.38) become

$$k = l = 1, \quad (2.43)$$

which are precisely the conditions given in the aforementioned paper on CW_1.

If one does *not* assume that the colliding wave conditions (2.43) hold, then Eqs. (2.40), (2.41), and (2.42) yield

$$\begin{aligned} \exp[2\Gamma(u,v)] = (\text{const.})u^{1-l}v^{1-k}\exp[&-2\Lambda(u,v) \\ &+ 2\Gamma_3(u) + 2\Gamma_2(v)], \end{aligned} \quad (2.44)$$

where

$$\begin{aligned} \Gamma_3(u) &:= 2\int_0^u da \frac{l - L(a,0)}{a}, \\ \Gamma_2(v) &:= 2\int_0^v db \frac{k - K(0,b)}{b}. \end{aligned} \quad (2.45)$$

Equation (2.45) supplies insight into what happens when the colliding wave conditions (2.43) do not hold.

In the aforementioned paper[6] on CW_1, errors occurred in some equations used to derive the above colliding wave conditions (2.43). To correct these errors, replace $\Lambda(u,v)$ by $2\Lambda(u,v)$ in Eq. (2.28) of that paper, replace Λ_v by $2\Lambda_v$, Λ_u by $2\Lambda_u$, l by $K(0,v)$, k by $L(u,0)$, Γ by 2Γ, and Λ by 2Λ in the three equations preceding Eq. (2.29) of that paper, and replace Eq. (2.29) of that paper by the above Eq. (2.44) of this paper.

L. The shape and boundary of IV

Note that, since $r(u)$ and $s(v)$ are monotonic, the limits[7]

$$r_0 = r(u_0) := \lim_{u \to u_0} [r(u)] ,\qquad s_0 = s(v_0) := \lim_{v \to v_0} [s(v)] , \tag{2.46}$$

exist. Furthermore, from Eqs. (2.19) and (2.20) and the inequality $\rho(u,v) > 0$ for all (u,v) in the domain of the chart,

$$-1 \leqslant r(u) < r_0 \leqslant 1 , \quad -1 \leqslant s_0 < s(v) \leqslant 1 . \tag{2.47}$$

With the aid of the above Eqs. (2.46) and (2.47), as well as Eqs. (2.18), (2.20), (2.21), (2.24), (2.28), (2.29), and (2.30), we can now make definitive statements about the connectedness, shape, and boundary of IV. The proofs of these statements will not be given since they are fairly obvious.

IV is always a subset of the rectangular region

$$R(u_0,v_0) := \{(u,v) \in R^2 : 0 \leqslant u < u_0, 0 \leqslant v < v_0\} . \tag{2.48}$$

In particular,

$$\mathrm{IV} = R(u_0,v_0), \quad \text{if } s(v_0) \geqslant r(u_0) . \tag{2.49}$$

Note that, if we extend Eq. (2.18) to all (u,v) in the closure of $R(u_0,v_0)$, then the condition $s(v_0) \geqslant r(u_0)$ is equivalent to the condition $\rho(u_0,v_0) \geqslant 0$.

Next, we want to discuss the case $s(v_0) < r(u_0)$ [which is equivalent to $\rho(u_0,v_0) < 0$]. Let us introduce the following open line segment on which $\rho(u,v) = 0$:

$$L(u_0,v_0) := \{(u,v) \in R(u_0,v_0) : s(v) = r(u)\} . \tag{2.50}$$

This line is simple (does not intersect itself), has a negative slope at each of its points, and has the end points (u_0,v_1) and (u_1,v_0) where v_1 and u_1 are defined by

$$s(v_1) = r(u_0) , \quad s(v_0) = r(u_1) , \tag{2.51}$$

and satisfy

$$0 \leqslant v_1 < v_0 , \quad 0 \leqslant u_1 < u_0 . \tag{2.52}$$

Moreover, $L(u_0,v_0)$ divides $R(u_0,v_0)$ into two simply connected parts and IV is the lower left part as illustrated in Fig. 1. The boundary of IV is the union of

$$\{(u,0) : 0 \leqslant u < u_0\} \cup \{(0,v) : 0 \leqslant v < v_0\} ,$$

which is a subset of IV, and

$$\{(u_0,v) : 0 \leqslant v \leqslant v_1\} \cup \{(u,v_0) : 0 \leqslant u \leqslant u_1\} \cup L(u_0,v_0) ,$$

which has no points in common with IV.

M. Use of r and s as null coordinates

Recall that $\dot{r}(u) > 0$ for $0 < u < u_0$ and $\dot{s}(v) < 0$ for $0 < v < v_0$. Therefore, one can replace u,v as coordinate variables by r,s in the domains $0 < u < u_0$ and $0 < v < v_0$, respectively. We shall often do that. Note that we are violating notational consistency by employing "r" and "s" to designate both functions and coordinate variables. However, the context should always indicate which use is intended.

In spite of the facts that $\dot{r}(0) = \dot{s}(0) = 0$, we shall also frequently employ the transformations $u \to r$ and $v \to s$ over the domains $0 \leqslant u < u_0$ and $0 \leqslant v < v_0$, respectively. These transformations are one–one and bicontinuous. The corresponding ranges of r and s are $-1 \leqslant r < r_0$ and $s_0 < s \leqslant 1$.

The transformation $(u,v) \to (r,s)$ over the domain $0 \leqslant u < u_0, 0 \leqslant v < v_0$ maps region IV in the (u,v) plane onto a region D_{IV} in the (r,s) plane. It is easier to visualize D_{IV} than it is to visualize IV (in the general case) because the lines of constant ρ in the (r,s) plane are the straight lines of slope $+1$. Always D_{IV} is a subset of the rectangular region

$$\{(r,s) : -1 \leqslant r < r_0, s_0 < s \leqslant 1\}$$

whose vertices are $(-1,1)$, $(-1,s_0)$, $(r_0,1)$, and (r_0,s_0) and which equals D_{IV} when $s_0 \geqslant r_0$, i.e., when the line $\rho = 0 (s = r)$ passes through no point of the rectangular region. If $s_0 < r_0$, then the line $\rho = 0$ intercepts the rectangular region and thus divides it into two parts as illustrated in Fig. 2. The part which contains the vertex $(-1,1)$ is D_{IV}.

N. The collinear case

In Secs. III and IV of this paper, E is real and ψ is defined by

$$E = \exp(-2\psi) \tag{2.53}$$

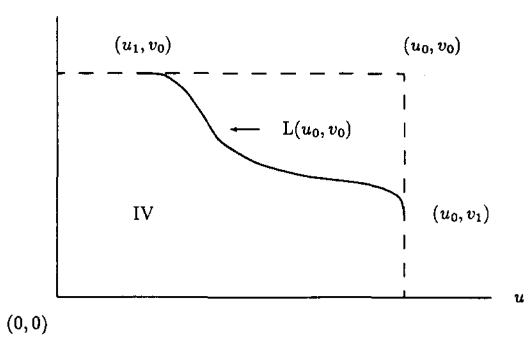


FIG. 1. Region IV in a typical case when $r(u_0) > s(v_0)$. Here, $\rho(u,v) = 0$ on $L(u_0,v_0)$, and $\rho(u,v) > 0$ on IV.

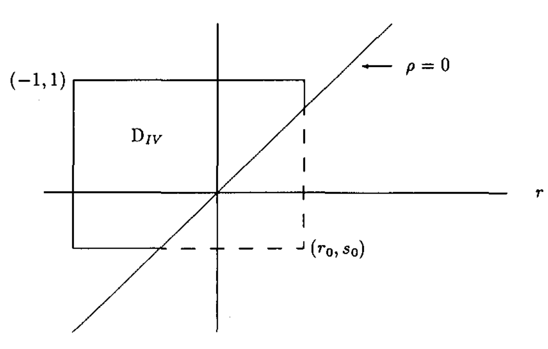


FIG. 2. Region D_{IV} in a typical case when $r_0 > s_0$.

whereupon the Ernst equation (2.22) reduces to the linear equation (1.4) for the field ψ. The problem is to find $\psi(u,v)$ for given $r(u)$, $s(v)$, $\psi(u,0)$, and $\psi(0,v)$. We adopt the option (2.7) which implies

$$\psi(0,0) = 0 . \tag{2.54}$$

III. GENERALIZED ABEL TRANSFORMS

A. Definitions of $\psi_3, \psi_2, u_\sigma, v_\sigma, \gamma_3, \gamma_2$

In Sec. III, we focus attention on $r(u)$ and $\psi(u,0)$ for $0\leqslant u<u_0$ and on $s(v)$ and $\psi(0,v)$ for $0\leqslant v<v_0$. These functions were assumed to be C^2 in Sec. II. However, in this section and in Sec. IV, the only assumptions that we shall make (apart from explicitly stated premises in theorems and corollaries) respecting $r(u)$, $s(v)$, $\psi(u,0)$, and $\psi(0,v)$ are that they exist, that they are C^1, that the conditions (2.28) and (2.29) hold and that $r(0) = -1$ and $s(0) = 1$. We shall *not* impose the conditions (2.24) and the colliding wave conditions (2.36) and (2.37).

Definitions: Let

$$\psi_3(u) := \psi(u,0) , \quad \psi_2(v) := \psi(0,v) . \tag{3.1}$$

Definitions: For any given real number σ such that $-1\leqslant\sigma<r_0$ [recall that r_0 and s_0 are defined by Eqs. (2.46)], let u_σ be that real number such that

$$0\leqslant u_\sigma<u_0 , \quad r(u_\sigma) = \sigma . \tag{3.2}$$

For any given real number σ such that $s_0<\sigma\leqslant 1$, let v_σ be that real number such that

$$0\leqslant v_\sigma<v_0 , \quad s(v_\sigma) = \sigma . \tag{3.3}$$

Definitions: Let

$$\begin{aligned} &\gamma_3(r) := \psi_2(u_r) \text{ , where } -1\leqslant r<r_0 , \\ &\gamma_2(s) := \psi_2(v_s) \text{ , where } s_0<s\leqslant 1 . \end{aligned} \tag{3.4}$$

Overdots will be used to designate ordinary derivatives like, for example, $\dot\psi_3(u) := d\psi_3(u)/du$ and $\dot\gamma_3(r) := d\gamma_3(r)/dr$.

From our assumptions, $\gamma_3(r)$ and $\gamma_2(s)$ are continuous over the semiclosed intervals $-1\leqslant r<r_0$ and $s_0<s\leqslant 1$ and are C^1 over the open intervals $-1<r<r_0$ and $s_0<s<1$, respectively. However, $\dot\gamma_3(r)$ and $\dot\gamma_2(s)$ may not exist or may not be continuous at $r = -1$ and $s = 1$, respectively. For example, if conditions (2.24), (2.36), and (2.37) hold, then Eqs. (2.39) and (2.53) imply that $\dot\gamma_3(r)$ is unbounded as $r\to -1$ and $\dot\gamma_2(s)$ is unbounded as $s\to 1$.

B. Definitions and properties of g_3 and g_2

We employ the Lebesgue definition of *integral* throughout this paper and we use term "integrable" as a synonym for "summable." The reader will observe that all integrals in this paper are also Riemann integrals or absolutely convergent improper Riemann integrals.

Theorem: The integral (Abel transform)[4]

$$g_3(\sigma) := \int_{-1}^{\sigma} dr \frac{\dot\gamma_3(r)}{\sqrt{\sigma-r}}, \quad \text{where } -1<\sigma<r_0 , \tag{3.5}$$

exists and is a continuous function of σ on the open interval $]-1,r_0[$. The integral

$$g_2(\sigma) := -\int_{1}^{\sigma} ds \frac{\dot\gamma_2(s)}{\sqrt{s-\sigma}}, \quad \text{where } s_0<\sigma<1, \tag{3.6}$$

exists and is a continuous function of σ on $]s_0,1[$.

Proof: The proofs for g_3 and g_2 are alike, so we give a proof only for g_3.

Let α and σ be any real numbers such that $-1<\alpha<\sigma<r_0$ and let

$$\begin{aligned} &I_1(\sigma,\alpha) := \int_0^{\mu_\alpha} du \frac{\dot\psi_3(u)}{\sqrt{\sigma-r(u)}} , \\ &I_2(\sigma,\alpha) := 2\sqrt{\sigma-\alpha} \int_0^1 dx\, \dot\gamma_3[\sigma-(\sigma-\alpha)x^2] . \end{aligned} \tag{3.7}$$

For fixed α and for any fixed ϵ such that $0<2\epsilon<r_0-\alpha$, the integrand of $I_1(\sigma,\alpha)$ is a continuous function of (σ,u) on the set

$$\{(\sigma,u):\alpha+\epsilon\leqslant\sigma\leqslant r_0-\epsilon,\ 0\leqslant u\leqslant u_\alpha\}$$

and the integrand of $I_2(\sigma,\alpha)$ is a continuous function of (σ,x) on

$$\{(\sigma,x):\alpha+\epsilon\leqslant\sigma\leqslant r_0-\epsilon,\ 0\leqslant x\leqslant 1\}.$$

Hence, $I_1(\sigma,\alpha)$ and $I_2(\sigma,\alpha)$ exist and are continuous functions of σ over $\alpha+\epsilon\leqslant\sigma\leqslant r_0-\epsilon$, and, therefore, over $\alpha<\sigma<r_0$.

In the next phase of the proof, we use the theorem on changing the integration variable in a Lebesgue integral.[23] Let us make the changes

$$u\to u_r \quad\text{and}\quad x\to\sqrt{\frac{\sigma-r}{\sigma-\alpha}}$$

in the integrands of $I_1(\sigma,\alpha)$ and $I_2(\sigma,\alpha)$, respectively. Note that u_r and $\sqrt{(\sigma-r)/(\sigma-\alpha)}$ are monotonic and absolutely continuous functions of r over the intervals $0\leqslant r\leqslant\alpha$ and $\alpha\leqslant r\leqslant\sigma$, respectively. Therefore, the theorem on changing the integration variable is applicable to both of the integrals in Eqs. (3.7) and we obtain

$$I_1(\sigma,\alpha) = \int_{-1}^{\alpha} dr \frac{\dot\gamma_3(r)}{\sqrt{\sigma-r}}, \quad I_2(\sigma,\alpha) = \int_{\alpha}^{\sigma} dr \frac{\dot\gamma_3(r)}{\sqrt{\sigma-r}} .$$

So

$$g_3(\sigma) = I_1(\sigma,\alpha) + I_2(\sigma,\alpha) \tag{3.8}$$

exists and is a continuous function of σ over $\alpha<\sigma<r_0$. However, $g_3(\sigma)$ is independent of α. Hence, $g_3(\sigma)$ exists and is a continuous function of σ over $-1<\sigma<r_0$. Q.E.D.

Corollary: (1) If $\psi_3(u)$ and $r(u)$ are C^{n+1} over $0<u<r_0$ then $g_3(\sigma)$ is C^n over $-1<\sigma<r_0$. (2) If $\psi_2(v)$ and $s(v)$ are C^{n+1} over $0<v<v_0$, then $g_2(\sigma)$ is C^n over $s_0<\sigma<1$.

Proof: For example, consider the proof of (1). One uses a procedure similar to that which we employed to prove continuity. The first step is to prove that $I_1(\sigma,\alpha)$ and $I_2(\sigma,\alpha)$ as defined by Eqs. (3.7) are (for fixed α) C^n over $\alpha<\sigma<r_0$, whereupon Eq. (3.8) implies that $g_3(\sigma)$ is C^n over $-1<\sigma<r_0$.

Definition: Consider any real or complex valued function f whose domain contains a closed interval $[a,b]$ of the real axis. Suppose that there exists a real number v such that $0\leqslant v\leqslant 1$ and a positive real number $M(a,b,v)$ such that for all $\sigma,\sigma'\in[a,b]$ for which $\sigma'>\sigma$:

$$|f(\sigma')-f(\sigma)|\leqslant(\sigma'-\sigma)^v M(a,b,v).$$

We shall then say that f is $\mathscr{H}(a,b,\nu)$. In particular, if $\nu>0$, one says that f obeys a Hölder condition of index ν on $[a,b]$. Note that f is $\mathscr{H}(a,b,0)$ if f is continuous on $[a,b]$ and f is $\mathscr{H}(a,b,1)$ if f is C^1 on $[a,b]$.

Theorem: (1) If $\dot\gamma_3$ is $\mathscr{H}(p,q,\nu)$ for every closed subinterval $[p,q]$ of the open interval $]-1,r_0[$, then g_3 is $\mathscr{H}(a,b,(1+3\nu)/4)$ for every closed subinterval $[a,b]$ of $]-1,r_0[$. If $\dot\gamma_2$ is $\mathscr{H}(p,q,\nu)$ for every closed subinterval $[p,q]$ of $]s_0,1[$, then g_2 is $\mathscr{H}(a,b,(1+3\nu)/4)$ for every closed subinterval $[a,b]$ of $]s_0,1[$.

Proof: We prove only statement (1) of this theorem since the proof of statement (2) is similar. Let $[a,b]$ denote any closed subinterval of $]-1,r_0[$ and let σ and σ' be any points of $[a,b]$ such that $a\leqslant\sigma<\sigma'\leqslant b$. Let

$$c:=a-\tfrac{1}{2}(a+1) \tag{3.9}$$

and let α be any real number such that $c\leqslant\alpha<\sigma$. Thus

$$-1<c<a\leqslant\sigma<\sigma'\leqslant b<r_0,\quad c\leqslant\alpha<\sigma. \tag{3.10}$$

Observe that $\sigma-(\sigma-\alpha)x^2$ and $\sigma'-(\sigma'-\alpha)x^2$ for $0\leqslant x\leqslant 1$ are both members of $[c,b]$, which is a closed subinterval of $]-1,r_0[$. Therefore, from the premise of the theorem and from the definition of $\mathscr{H}(c,b,\nu)$,

$$\begin{aligned}&|\dot\gamma_2(\sigma'-(\sigma'-\alpha)x^2)-\dot\gamma_3(\sigma-(\sigma-\alpha)x^2)|\\&\quad\leqslant(\sigma'-\sigma)^\nu(1-x^2)^\nu M_1,\end{aligned} \tag{3.11}$$

where M_1 is independent of σ, σ', α, and x. We shall also be using the following easily proven inequalities:

$$\begin{aligned}&\frac{1}{\sqrt{\sigma-\alpha}}-\frac{1}{\sqrt{\sigma'-\alpha}}\leqslant\frac{\sigma'-\sigma}{2(\sigma-\alpha)^{3/2}},\\&\sqrt{\sigma'-\alpha}-\sqrt{\sigma-\alpha}\leqslant\frac{\sigma'-\sigma}{2(\sigma-\alpha)^{1/2}}\\&\quad\leqslant\frac{(\sigma'-\sigma)^\kappa(b-a)^{1-\kappa}}{2(\sigma-\alpha)^{1/2}},\end{aligned} \tag{3.12}$$

where $\kappa:=\tfrac{1}{2}(1+\nu)$.

As the next phase of our proof, note that σ and α satisfy the inequalities $-1<\alpha<\sigma<r_0$ that are sufficient for the definitions (3.7) of $I_1(\sigma,\alpha)$ and $I_2(\sigma,\alpha)$ to be applicable. From the conclusion in the proof of the preceding theorem, $I_1(\sigma,\alpha)$ and $I_2(\sigma,\alpha)$ exist and Eq. (3.8) holds. Therefore,

$$\begin{aligned}&|g_3(\sigma')-g_3(\sigma)|\\&\quad\leqslant\int_0^{u_\alpha}du|\dot\psi_3(u)|\left[\frac{1}{\sqrt{\sigma-r(u)}}-\frac{1}{\sqrt{\sigma'-r(u)}}\right]\\&\quad+2(\sqrt{\sigma'-\alpha}-\sqrt{\sigma-\alpha})\int_0^1dx|\dot\gamma_3(\sigma'-(\sigma'-\alpha)x^2)|\\&\quad+2\sqrt{\sigma-\alpha}\int_0^1dx|\dot\gamma_3(\sigma'-(\sigma'-\alpha)x^2)\\&\quad-\dot\gamma_3(\sigma-(\sigma-\alpha)x^2)|\leqslant2\left(\frac{1}{\sqrt{\sigma-\alpha}}-\frac{1}{\sqrt{\sigma'-\alpha}}\right)M_2\\&\quad+2(\sqrt{\sigma'-\alpha}-\sqrt{\sigma-\alpha})M_3+2\sqrt{\sigma-\alpha}\\&\quad\times\int_0^1dx|\dot\gamma_3(\sigma'-(\sigma'-\lambda)x^2)-\dot\gamma_3(\sigma-(\sigma-\lambda)x^2)|,\end{aligned}$$

where

$$M_2:=\frac{1}{2}\int_0^{u_b}du|\dot\psi_3(u)|,$$

M_3: = maximum value of $|\dot\gamma_3(r)|$ in the interval $[c,b]$.

So, upon using the inequalities (3.11) and (3.12), we obtain

$$\begin{aligned}|g_3(\sigma')-g_3(\sigma)|&\leqslant\left[\frac{\sigma'-\sigma}{(\sigma-\alpha)^{3/2}}\right]M_2\\&+\left[\frac{(\sigma'-\sigma)^\kappa}{(\sigma-\alpha)^{1/2}}\right](b-a)^{1-\kappa}M_3\\&+2(\sigma-\alpha)^{1/2}(\sigma'-\sigma)^\nu M_1,\end{aligned}$$

where it is important to note that M_1, M_2, M_3 are independent of σ, σ', and α. Now, into the above inequality, substitute

$$\alpha=\sigma-\left(\frac{\sigma'-\sigma}{b-a}\right)^{(1-\nu)/2}\left(\frac{a+1}{2}\right),$$

which satisfies the defining condition for α as can be seen from (3.9) and (3.10). There results:

$$|g_3(\sigma')-g_3(\sigma)|\leqslant(\sigma'-\sigma)^{(1+3\nu)/4}M,$$

where M is independent of σ and σ'. Q.E.D.

The above theorem will be useful in our next paper on the initial value problem.

C. Inversion of the Abel transform

In the proof of the next theorem, we shall be using the well-known relation[4]

$$\frac{1}{\pi}\int_{r'}^{r}d\sigma\frac{1}{\sqrt{(r-\sigma)(\sigma-r')}}=1,\quad\text{for } r>r', \tag{3.13}$$

which can be proven by a contour integration method.

Theorem: The following integrals exist and are equal to $\gamma_3(r)$ and $\gamma_2(s)$, respectively, as stated below:[4]

$$\gamma_3(r)=\frac{1}{\pi}\int_{-1}^{r}d\sigma\frac{g_3(\sigma)}{\sqrt{r-\sigma}},\quad\text{for } -1\leqslant r<r_0 \tag{3.14}$$

and

$$\gamma_2(s)=\frac{1}{\pi}\int_1^s d\sigma\frac{g_2(\sigma)}{\sqrt{\sigma-s}},\quad\text{for } s_0<s\leqslant1, \tag{3.15}$$

where the values of the integrals at $r=-1$ and $s=1$ are defined as their limits as $r\to-1$ from above and $s\to1$ from below, respectively.

Proof: It is sufficient to supply the proof of the theorem for Eq. (3.14) since the proof for Eq. (3.15) is similar.

In the first phase of the proof, let $-1<r<r_0$. For fixed r, we introduce the function

$$f(r',\sigma):=\frac{\dot\gamma_3(r')}{\pi\sqrt{(r-\sigma)(\sigma-r')}} \tag{3.16}$$

whose domain is the open triangle

$$D_r:=\{(r',\sigma)\in R^2:-1<\sigma<r,-1<r'<\sigma\},$$

which is illustrated by Fig. 3. Note that f is continuous throughout this domain.

Now, $|\dot\psi_3|$ is continuous and is consequently integrable over $[0,u]$. It follows, by employing the theorem[23] on change of the integration variable in a Lebesgue integral, that $|\dot\gamma_3|$ is integrable over $[-1,r]$. Therefore, from Eqs. (3.13) and (3.16), the following iterated integral exists for given r such that $-1<r<r_0$:

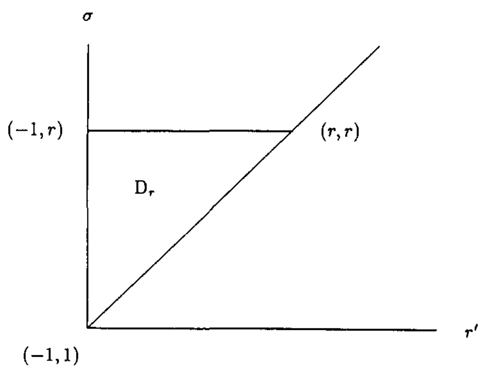


FIG. 3. The triangular domain D_r.

$$\int_{-1}^{r} dr' \int_{r'}^{r} d\sigma |f(r',\sigma)| = \int_{-1}^{r} dr' |\dot\gamma_3(r')|.$$

From theorems of Fubini,[24] it follows that the double integral of f over D_r exists and is equal to both of the following iterated integrals which also exist:

$$\int_{-1}^{r} dr' \int_{r'}^{r} d\sigma f(r',\sigma) = \int_{-1}^{r} d\sigma \int_{-1}^{\sigma} dr' f(r',\sigma). \quad (3.17)$$

Substitution from Eq. (3.16) into Eq. (3.17) and the use of Eqs. (3.5) and (3.13) yield the existences and the equality of the following two integrals:

$$\int_{-1}^{r} dr' \dot\gamma_3(r') = \frac{1}{\pi}\int_{-1}^{r} d\sigma \frac{g_3(\sigma)}{\sqrt{r-\sigma}}$$

However, $\dot\gamma_3(r')$ is continuous over $-1<r'<r_0$ and $\gamma_3(-1) = \psi_3(0) = \psi(0,0) = 0$. Therefore, we obtain Eq. (3.14) for $-1<r<r_0$.

To complete the proof, note that the continuity of $\gamma_3(r)$ over $-1\leqslant r<r_0$ guarantees that the integral in Eq. (3.14) has a limit as $r\to -1$ from the above and that this limit is $\gamma_3(-1) = 0$.

D. Theorem on uniqueness

An obvious corollary of the above theorem is the uniqueness of γ_3 for a given g_3 such that Eq. (3.5) holds and the uniqueness of γ_2 for a given g_2 such that Eq. (3.6) holds. The following theorem is on the uniqueness of g_3 for a given γ_3 such that Eq. (3.14) holds and on the uniqueness of g_2 for a given γ_2 such that Eq. (3.15) holds.

Theorem: (1) let g_3' be any real valued and continuous function with the domain $]-1,r_0[$ such that, for all r satisfying $-1\leqslant r<r_0$, $g_3'(\sigma)/\sqrt{r-\sigma}$ is integrable over $[-1,r]$ and

$$\gamma_3(r) = \frac{1}{\pi}\int_{-1}^{r} d\sigma \frac{g_3'(\sigma)}{\sqrt{r-\sigma}}, \quad (3.18)$$

where the value of the integral at $r=-1$ is defined as its limit as $r\to -1$ from above. Then $g_3' = g_3$. (2) Let g_2' be any real valued and continuous function with the domain $]s_0,1[$ such that, for all s satisfying $s_0<s\leqslant 1$, $g_2'(\sigma)/\sqrt{\sigma-s}$ is integrable over $[s,1]$ and

$$\gamma_2(s) = \frac{1}{\pi}\int_{1}^{s} d\sigma \frac{g_2'(\sigma)}{\sqrt{\sigma-s}}, \quad (3.19)$$

where the value of the integral at $s=1$ is defined as its limit as $s\to 1$ from below. Then $g_2' = g_2$.

Proof: We shall only supply the proof of part (1) of the theorem. From Eqs. (3.14) and (3.18),

$$\frac{1}{\pi}\int_{-1}^{r} d\sigma \frac{\Delta(\sigma)}{\sqrt{r-\sigma}} = 0, \quad \Delta := g_3' - g_3. \quad (3.20)$$

For fixed $r>-1$ we introduce the function

$$\phi(\sigma,r') := \frac{\Delta(\sigma)}{\pi\sqrt{(r-r')(r'-\sigma)}} \quad (3.21)$$

defined over the open triangular domain

$$\tilde D_r := \{(\sigma,r') : -1<r'<r, -1<\sigma<r'\},$$

which is obtained from the triangular domain D_r (illustrated in Fig. 3) by transposing the r' and σ axes. Here, ϕ is clearly continuous on $\tilde D_r$.

Also, since $\Delta(\sigma)/\sqrt{r-\sigma}$ is integrable over $[-1,r]$, so is $|\Delta(\sigma)|$.[25] Therefore, from Eq. (3.13), the following iterated integral exists:

$$\int_{-1}^{r} d\sigma \int_{\sigma}^{r} dr' |\phi(\sigma,r')| = \int_{-1}^{r} d\sigma |\Delta(\sigma)|.$$

Therefore, from the same theorems[24] of Fubini that were used to prove the preceding theorem, the double integral of ϕ over $\tilde D_r$ exists and is equal to both of the following iterated integrals that also exist:

$$\int_{-1}^{r} d\sigma \int_{\sigma}^{r} dr' \phi(\sigma,r') = \int_{-1}^{r} dr' \int_{-1}^{r'} d\sigma \phi(\sigma,r'),$$

Therefore, from Eqs. (3.13), (3.20), and (3.21),

$$\int_{-1}^{r} d\sigma\, \Delta(\sigma) = 0. \quad (3.22)$$

However, from the second of Eqs. (3.20) and from the continuity of g_3' (a premise of this theorem) and the continuity of g_3 [line after Eq. (3.5)], $\Delta(\sigma)$ is continuous over $-1<\sigma<r_0$. Therefore, upon differentiating[26] (3.22) with respect to r, one obtains $\Delta(r) = 0$ for all r such that $-1<r<r_0$. Therefore, $g_3'(\sigma) = g_3(\sigma)$ for all σ such that $-1<\sigma<r_0$. Q.E.D.

E. Replacement of (r,s) by (u,v) in the Abel transforms

Upon introducing the changes

$$r\to r(u), \quad s\to s(v)$$

in the variables of integration in Eqs. (3.5) and (3.6), respectively, and upon using Eqs. (3.1) to (3.4), one obtains that form of the Abel transforms that was given in Sec. I by Eqs. (1.10). The same substitutions for the free variables r and s in Eqs. (3.14) and (3.15), respectively, yields Eqs. (1.9) of Sec. I.

IV. SOLUTION OF THE INITIAL VALUE PROBLEM

A. Some preliminary definitions

In the following definitions, w and z denote any given finite complex numbers, C denotes the extended complex plane and τ denotes any member of C. All curves that we consider in the complex plane are understood (though we do not explicitly say so) to be continuous, piecewise smooth, rectifiable, and simple. An oriented closed curve is called a *contour*.

Definition: $|w,z| = |z,w|$ will denote that straight line segment whose end points are w and z and which includes w and z.

Definition: $\chi_3(w,\tau)$ and $\chi_2(z,\tau)$ will denote the holomorphic functions of τ given by those branches of

$$\chi_3(w,\tau) = \left(\frac{\tau+1}{\tau-w}\right)^{1/2} \text{ and } \chi_2(z,\tau) = \left(\frac{\tau-1}{\tau-z}\right)^{1/2} \tag{4.1}$$

whose cuts are $|-1,w|$ and $|1,z|$, respectively, and which have the values $+1$ at $\tau = \infty$. It is to be understood that

$$\chi_3(-1,\tau) := 1 \text{ and } \chi_2(1,\tau) := 1,$$

for all $\tau \in C$.

Definition: $\mu(w,z,\tau)$ will denote the holomorphic function of τ which is given by that branch of

$$\mu(w,z,\tau) = [(\tau-w)(\tau-z)]^{1/2} \tag{4.2}$$

whose cut is $|w,z|$ and which satisfies

$$\tau^{-1}\mu(w,z,\tau) \to 1 \text{ as } \tau \to \infty.$$

Note that $\mu(w,w,\tau) = \tau - w$.

Definitions of Γ_+ *and* Γ_- *for a given contour* Γ*:* For any given positively oriented contour Γ in the complex plane, Γ_+ and Γ_- will denote those bounded and unbounded open subsets of C, respectively, which have Γ as their mutual boundary. (In descriptive terms, Γ_+ is the open region inside Γ and Γ_- is the open region outside Γ.)

B. Motivation

We shall first explain how we arrived at the form of the solutions given by Eqs. (1.6) and (1.7). This explanation has a strictly motivational purpose and will, therefore, be scant on computational details and proofs.

We follow a classical pattern and seek a solution of Eq. (1.4) which is a function of u times a function of v. Equivalently, recalling that $\rho = (s-r)/2$, we seek a solution

$$\gamma(r,s) := \psi(u_r,v_s) \tag{4.3}$$

of

$$2(s-r)\gamma_{rs} - \gamma_s + \gamma_r = 0 \tag{4.4}$$

such that $\gamma(r,s)$ is a function of r times a function of s. This particular solution will be denoted by $\chi(r,s,\tau)$ and is expressible in the form

$$\chi(r,s,\tau) = \chi_3(r,\tau)\chi_2(s,\tau), \tag{4.5}$$

where τ is the separation parameter and where we have normalized the factors so that they are given by Eqs. (4.1). One obtains the functions of u and v by the substitutions $r \to r(u)$ and $s \to s(v)$.

Real-valued solutions γ of Eq. (4.4) with the domain (see Sec. II M and Fig. 2)

$$D_{\rm IV} = \{(r,s) \in R^2: -1 \leqslant r < r_0,\ s_0 < s \leqslant 1,\ s < r\}$$

can now be constructed by taking suitable superpositions of $\chi(r,s,\tau)$ over the complex parameter τ. Contour integrals immediately come to mind and, in particular, one thinks of contours Γ_3 and Γ_2 which surround the cuts of $\chi_3(r,\tau)$ and $\chi_2(s,\tau)$, respectively. Thus, we construct the superposition

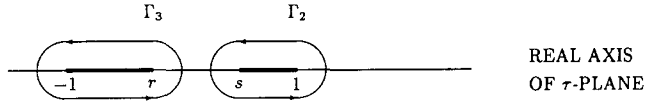


FIG. 4. Illustrative choices for the contours Γ_3 and Γ_2.

$$\gamma(r,s) = {}_3\gamma(r,s) + {}_2\gamma(r,s), \tag{4.6}$$

where $(j = 3,2)$

$${}_j\gamma(r,s) = \frac{1}{2\pi i}\int_{\Gamma_j} d\tau\, \chi(r,s,\tau) f_j(\tau) \tag{4.7}$$

and where $f_3(\tau)$ and $f_2(\tau)$ are any holomorphic functions of τ whose domains satisfy

$$[-1,r_0[:= \{r: -1 \leqslant r < r_0\} \subset \operatorname{dom} f_3,$$

$$]s_0,1] := \{s: s_0 < s \leqslant 1\} \subset \operatorname{dom} f_2.$$

In addition, f_j satisfies the reality condition

$$[f_j(\tau^*)]^* = f_j(\tau).$$

For given f_3, f_2 and (r,s), Γ_3 and Γ_2 are any positively oriented contours which are disjoint and which satisfy

$$[-1,r] \subset \Gamma_{3+}, \quad [s,1] \subset \Gamma_{2+},$$

$$(\Gamma_j \cup \Gamma_{j+}) \subset \operatorname{dom} f_j \quad (j = 3,2).$$

Possible choices of Γ_3 and Γ_2 are shown in Fig. 4. It is apparent from Eqs. (4.1), (4.5), and (4.7) that

$${}_3\gamma(-1,s) = 0, \quad {}_2\gamma(r,1) = 0.$$

Therefore, from Eqs. (4.6),

$$\gamma(r,1) = {}_3\gamma(r,1), \quad \gamma(-1,s) = {}_2\gamma(-1,s).$$

Now let the contours of integration be contracted so that the points on Γ_3 and Γ_2 approach points on $[-1,r]$ and $[s,1]$, respectively. Specifically, choose Γ_3 as we can always do so that it is made up of two straight line segments parallel to $[-1,r]$ and two semicircular arcs with centers at -1 and r (see Fig. 4). Likewise, choose Γ_2 so that it is made up of two straight line segments parallel to $[s,1]$ and two semicircular arcs with centers at 1 and s. Then let the distances of Γ_3 and Γ_2 from $[-1,r]$ and $[s,1]$, respectively, go to zero. The corresponding limits of the right side of Eq. (4.7) for $j = 3$ and $j = 2$ exist and are given by

$${}_3\gamma(r,s) = \frac{1}{\pi}\int_{-1}^{r} d\sigma \sqrt{\frac{1-\sigma}{(r-\sigma)(s-\sigma)}}\, g_3(\sigma) \tag{4.8}$$

and

$${}_2\gamma(r,s) = \frac{1}{\pi}\int_{1}^{s} d\sigma \sqrt{\frac{1+\sigma}{(r-\sigma)(s-\sigma)}}\, g_2(\sigma), \tag{4.9}$$

where

$$g_3(\sigma) = \sqrt{1+\sigma}\, f_3(\sigma), \quad g_2(\sigma) = \sqrt{1-\sigma}\, f_2(\sigma). \tag{4.10}$$

Upon replacing r by $r(u)$ and s by $s(v)$ in Eqs. (4.8) and (4.9), one obtains precisely the same forms as the expressions for ${}_3\psi(u,v)$ and ${}_2\psi(u,v)$ in Eqs. (1.6) and (1.7).

However, the $g_3(\sigma)$ and $g_2(\sigma)$ given by Eqs. (4.10) are analytic functions of σ and satisfy $g_3(-1) = g_2(1) = 0$. What we shall now do in Sec. IV is to retain the forms given by Eqs. (4.8) and (4.9) but drop any analyticity or even smoothness conditions on the functions g_j. Nor shall we require that $g_3(-1)$ and $g_2(1)$ vanish or even exist. Instead, $g_3(\sigma)$ and $g_2(\sigma)$ are to be defined in terms of the prescribed initial data by Eqs. (3.5) and (3.6).

C. Definitions and existences of ${}_3\gamma$ and ${}_2\gamma$

It will be observed that we remove former constraints on the range of (r,s) in the following definitions and at various later stages of Sec. IV. This is not merely done because we can do it. We have found such extensions useful.

Definitions: Let the functions ${}_3\gamma$ and ${}_2\gamma$ be defined by Eqs. (4.8) and (4.9) over the domains

$$\mathrm{dom}({}_3\gamma) := \{(r,s)\in R^2: -1\leqslant r<r_0,\ r<s<\infty\}, \tag{4.11}$$

and

$$\mathrm{dom}({}_2\gamma) := \{(r,s)\in R^2: s_0<s\leqslant 1,\ -\infty<r<s\}, \tag{4.12}$$

respectively, where it is to be understood that

$$\begin{aligned} {}_3\gamma(-1,s) &:= \lim_{r\to -1}[{}_3\gamma(r,s)],\\ {}_2\gamma(r,1) &:= \lim_{s\to 1}[{}_2\gamma(r,s)], \end{aligned} \tag{4.13}$$

and where $g_3(\sigma)$ and $g_2(\sigma)$ are defined by Eqs. (3.5) and (3.6).

Theorem: ${}_3\gamma$ and ${}_2\gamma$ exist.

Proof: Consider ${}_3\gamma$, for example. The proof of its existence will be given in three stages labeled (1), (2), and (3).

(1) Let us first restrict ourselves to those (r,s) in the domain (4.11) of ${}_3\gamma$ such that $r>-1$. Note that the integrand in Eq. (4.8) is expressible as the product

$$\left[\frac{g_3(\sigma)}{\pi\sqrt{r-\sigma}}\right]\left[\sqrt{\frac{1-\sigma}{s-\sigma}}\right].$$

For fixed (r,s), we already know from a theorem in Sec. III C that the first factor in the above product is integrable over $[-1,r]$. The second factor is continuous over $[-1,r]$. Therefore, from a theorem[25] on Lebesgue integrals, the product is integrable over $[-1,r]$. That proves the existence of ${}_3\gamma(r,s)$ when $r>-1$.

(2) Before we go on with the proof, we shall introduce a few definitions. For $0\leqslant u<u_0$, let $\dot\psi_3(u,+)$ and $\dot\psi_3(u,-)$ be defined by

$$\dot\psi_3(u,+) := \begin{cases}\dot\psi_3(u), & \text{if } \dot\psi_3(u)\geqslant 0,\\ 0, & \text{otherwise,}\end{cases}$$

$$\dot\psi_3(u,-) := \begin{cases}-\dot\psi_3(u), & \text{if } \dot\psi_3(u)\leqslant 0,\\ 0, & \text{otherwise,}\end{cases}$$

$$\psi_3(0,+) = \psi_3(0,-) = 0.$$

Note that $\psi_3(u,\pm)$ are non-negative nondecreasing functions of u and are C^1. Also,

$$\dot\psi_3(u,\pm)\geqslant 0,\quad \psi_3(u) = \psi_3(u,+) - \psi_3(u,-).$$

All of the preceding proofs and theorems clearly remain valid if we replace $\psi_3(u)$ by $\psi_3(u,\pm)$ and, correspondingly, replace the definitions (3.4), (3.5), and (4.8) by

$$\gamma_3(r,\pm) := \psi_3(u_r,\pm),\quad g_3(\sigma,\pm) := \int_{-1}^{\sigma} dr\,\frac{\dot\gamma_3(r,\pm)}{\sqrt{\sigma-r}},$$

$${}_3\gamma(r,s,\pm) := \frac{1}{\pi}\int_{-1}^{r} d\sigma\,\frac{g_3(\sigma,\pm)\sqrt{1-\sigma}}{\sqrt{(r-\sigma)(s-\sigma)}}. \tag{4.14}$$

From Eqs. (2.28) and (4.14), $\dot\gamma_3(r,\pm)\geqslant 0$ and

$$\begin{aligned} &g_3(\sigma,\pm)\geqslant 0\quad {}_3\gamma(r,s,\pm)\geqslant 0,\\ &g_3(\sigma) = g_3(\sigma,+) - g_3(\sigma,-),\\ &{}_3\gamma(r,s) = {}_3\gamma(r,s,+) - {}_3\gamma(r,s,-). \end{aligned} \tag{4.15}$$

(3) Now still assuming that $r>-1$, note that for given r and s,

$$M(r,s) := \left[\sqrt{\frac{1-\sigma}{s-\sigma}}\right]_{\max} = \begin{cases}\sqrt{\dfrac{1-r}{s-r}}, & \text{if } s\leqslant 1,\\ \sqrt{\dfrac{2}{s+1}}, & \text{if } s\geqslant 1.\end{cases} \tag{4.16}$$

From Eqs. (3.14) [with $\gamma_3(r)$ and $g_3(\sigma)$ replaced by $\gamma_3(r,\pm)$ and $g_3(\sigma,\pm)$, respectively] and (4.14), one obtains[27]

$${}_3\gamma(r,s,\pm)\leqslant M(r,s)\gamma_3(r,\pm).$$

Since $\gamma_3(r,\pm)$ is a continuous function of r over $-1\leqslant r<r_0$ and $\gamma_3(-1,\pm)=0$, it follows from Eqs. (4.15) and (4.16) that the limit of ${}_3\gamma(r,s)$ as $r\to -1$ exists and equals 0. Therefore, $\gamma_3(-1,s)$ exists. That completes the proof. The following corollary is obvious.

Corollary: For all $-1<s<\infty$ and $-\infty<r<1$, respectively,

$${}_3\gamma(-1,s)=0,\quad {}_2\gamma(r,1)=0. \tag{4.17}$$

D. The Green's functions G_3 and G_2

Definitions: Let G_3 and G_2 denote those functions whose domains are

$$\mathrm{dom}\ G_3 := \{(r',r,s)\in R^3: -\infty<r'<r<1, r<s<\infty\}$$

and

$$\mathrm{dom}\ G_2 := \{(s',s,r)\in R^3: -1<s<s'<\infty, -\infty<r<s\}$$

and whose values are defined by

$$G_3(r',r,s) := \frac{1}{\pi}\int_{r'}^{r} d\sigma\sqrt{\frac{1-\sigma}{(\sigma-r')(r-\sigma)(s-\sigma)}} \tag{4.18}$$

and

$$G_2(s',s,r) := -\frac{1}{\pi}\int_{s'}^{s} d\sigma\sqrt{\frac{1+\sigma}{(s'-\sigma)(\sigma-s)(\sigma-r)}}. \tag{4.19}$$

Theorem: G_3 and G_2 exist.

Proof: Consider G_3, for example. For given (r',r,s) in dom G_3, it can be seen from Eq. (3.13) that the integrand in Eq. (4.18) is the product of a function of σ which is integrable over $[r',r]$ and of a function of σ, viz., $\sqrt{(1-\sigma)/(s-\sigma)}$, which is continuous on $[r',r]$. The existence of $G_3(r',r,s)$ then follows from a well-known theorem.[25]

In Sec. IV E, we shall obtain contour integrals for G_3 and G_2. The contour integrals will permit us in an obvious way to extend G_3 and G_2 to the domains

$$\begin{aligned}&\operatorname{dom} G_3 = \{(r',r,s)\in R^3 : r<s,\ r'<s,\ r<1,\ r'<1\},\\ &\operatorname{dom} G_2 = \{(s',s,r)\in R^3 : r<s,\ r<s',\ -1<s,\ -1<s'\}\end{aligned} \tag{4.20}$$

such that the extended G_3 and G_2 are continuous over these domains and satisfy

$$G_3(r,r',s') = G_3(r',r,s), \quad G_2(s,s',r) = G_2(s',s,r). \tag{4.21}$$

Given the existence of these continuous extensions and Eqs. (4.17), the two sentences in the following theorem are manifest for $r=-1$ and $s=1$, respectively. They need be proven only for $r>-1$ and $s<1$, respectively.

Theorem: For any given (r,s) in dom ${}_3\gamma$, the product $\dot\gamma_3(r')G_3(r',r,s)$ is integrable over $-1\leqslant r'\leqslant r$ and

$${}_3\gamma(r,s) = \int_{-1}^{r} dr'\,\dot\gamma_3(r')G_3(r',r,s). \tag{4.22}$$

For any given (r,s) in dom ${}_2\gamma$, the product $\dot\gamma_2(s')G_2(s',s,r)$ is integrable over $s\leqslant s'\leqslant 1$ and

$${}_2\gamma(r,s) = \int_{1}^{s} ds'\,\dot\gamma_2(s')G_2(s',s,r). \tag{4.23}$$

Proof: Consider, for example, ${}_3\gamma$. For fixed (r,s) such that $-1<r<r_0$ and $r<s<\infty$, let F denote that function whose domain is D_r and which has the values

$$F(r',\sigma) := f(r',\sigma)\sqrt{\frac{1-\sigma}{s-\sigma}}, \tag{4.24}$$

where $f(r',\sigma)$ and the open triangular region D_r were defined by Eq. (3.16) and Fig. 3. We have already proven in Sec. III C that f is integrable over D_r. Also, $\sqrt{(1-\sigma)/(s-\sigma)}$ is a continuous function of (r',σ) over the closure of D_r. Therefore,[25] F is integrable over D_r, whereupon it follows from a theorem of Fubini[24] that the following two iterated integrals exist and are equal:

$$\int_{-1}^{r} d\sigma \int_{-1}^{\sigma} dr'\, F(r',\sigma) = \int_{-1}^{r} dr' \int_{r'}^{r} d\sigma\, F(r',\sigma). \tag{4.25}$$

Note, however, from Eqs. (3.5), (3.16), and (4.24) that the left side of the above Eq. (4.25) is ${}_3\gamma(r,s)$ as defined by Eq. (4.8). The right side of Eq. (4.25) is precisely the right side of Eq. (4.22) with G_3 given by Eq. (4.18). Q.E.D.

E. Contour integrals for G_3 and G_2

Theorem: Recalling the definitions (4.1) and (4.2) of χ_j and μ,

$$G_3(r',r,s) = \frac{1}{2\pi i}\int_{\mathscr{C}_3} d\tau\,\frac{\chi_2(s,\tau)}{\mu(r',r,\tau)} \tag{4.26}$$

and

$$G_2(s',s,r) = \frac{1}{2\pi i}\int_{\mathscr{C}_2} d\tau\,\frac{\chi_3(r,\tau)}{\mu(s',s,\tau)}, \tag{4.27}$$

where $\mathscr{C}_3$ and $\mathscr{C}_2$ are any positively oriented contours such that

$$\begin{aligned}&|r',r| \subset \mathscr{C}_{3+}, \quad |s,1| \subset \mathscr{C}_{3-},\\ &|s',s| \subset \mathscr{C}_{2+}, \quad |r,-1| \subset \mathscr{C}_{2-}.\end{aligned} \tag{4.28}$$

The topological relations of the contours $\mathscr{C}_j$ to the cuts of the integrands are depicted in Figs. 5 and 6.

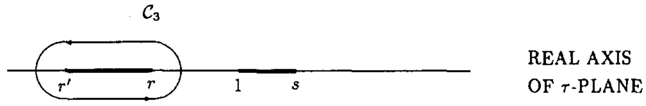


FIG. 5. Sample choice of the contour $\mathscr{C}_3$.

Proof: Consider, for example, the proof of Eq. (4.26). We start with the integral on the right side of Eq. (4.26) and select a $\mathscr{C}_3$ that is dumbbell shaped and is composed of two straight line segments parallel to $|r',r|$ and of two circular arcs with centers at r' and r. Specifically, let δ and ϵ be any positive real numbers such that $\epsilon<\delta<(r-r')/2$. Then the straight line segments are

$$L_{\pm} := \{\sigma \pm i\epsilon : r' + \sqrt{\delta^2-\epsilon^2} \leqslant \sigma \leqslant r - \sqrt{\delta^2-\epsilon^2}\}$$

and the circular arcs are

$$\begin{aligned}&\mathscr{A}' := \{r' + \delta e^{i\theta} : \theta_0 \leqslant \theta \leqslant 2\pi - \theta_0\},\\ &\mathscr{A} := \{r + \delta e^{i\theta} : -\pi + \theta_0 \leqslant \theta \leqslant \pi - \theta_0\},\end{aligned}$$

where $0<\theta_0 := \arcsin(\epsilon/\delta) < \pi/2$.

Now, let $\epsilon\to 0$ while holding δ fixed. Since the value of the integral on the right side of Eq. (4.26) is independent of ϵ and since

$$\begin{aligned}&\mu(\gamma',r,\sigma\pm i\epsilon) \to \pm i\sqrt{(\sigma-r')(r-\sigma)},\\ &\chi_2(s,\sigma\pm i\epsilon) \to \sqrt{(1-\sigma)/(s-\sigma)} \quad (r'<\sigma<r<s),\end{aligned}$$

as $\epsilon\to 0$, it follows that the right side of Eq. (4.26) equals

$$\frac{1}{2\pi i}\int_{\mathscr{C}} d\tau\,\frac{\chi_2(s,\tau)}{\mu(r',r,\tau)} + \frac{1}{\pi}\int_{r'+\delta}^{r-\delta} d\sigma \sqrt{\frac{1-\sigma}{(\sigma-r')(r-\sigma)(s-\sigma)}}, \tag{4.29}$$

where $\mathscr{C}$ is the following union of two positively oriented circles:

$$\mathscr{C} := \{r' + \delta e^{i\theta} : 0 \leqslant \theta < 2\pi\} \cup \{r + \delta e^{i\theta} : -\pi \leqslant \theta < \pi\}.$$

Next, let $\delta\to 0$ in (4.29) where it is to be noted that the right side of Eq. (4.26) is also independent of δ. The first term in (4.26) clearly has the limit 0 as $\delta\to 0$ and, by a well-known theorem on Lebesgue integrals, the second term has the limit given by the defining expression for $G_3(r',r,s)$ in Eq. (4.18). Q.E.D.

One can deform $\mathscr{C}_3$ and $\mathscr{C}_2$ subject only to the conditions (4.28). In this way, one can obtain the following alternative expressions for G_3 and G_2:

$$\begin{aligned}&G_3(r',r,s) = 1 - \frac{1}{2\pi i}\int_{\mathscr{K}_3} d\tau\,\frac{\chi_2(s,\tau)}{\mu(r',r,\tau)},\\ &G_2(s',s,r) = 1 - \frac{1}{2\pi i}\int_{\mathscr{K}_2} d\tau\,\frac{\chi_3(r,\tau)}{\mu(s',s,\tau)},\end{aligned} \tag{4.30}$$

where $\mathscr{K}_3$ and $\mathscr{K}_2$ are any positively oriented contours such that

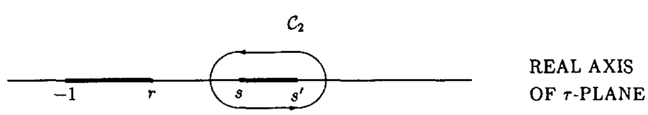


FIG. 6. Sample choice of the contour $\mathscr{C}_2$.

$$|1,s|\subset\mathscr{K}_{3+},\quad |r',r|\subset\mathscr{K}_{3-},$$
$$|-1,r|\subset\mathscr{K}_{2+},\quad |s',s|\subset\mathscr{K}_{2-}. \tag{4.31}$$

F. A few properties of G_3 and G_2

Below, C^3 denotes the Cartesian product $C\times C\times C$.

Lemma: For any given $(a',a,b)\in\operatorname{dom} G_3$ and $(d',d,c)\in\operatorname{dom} G_2$ and any given positively oriented circles $\mathscr{C}_3$, $\mathscr{K}_3$, $\mathscr{C}_2$, and $\mathscr{K}_2$ such that

$$|a',a|\subset\mathscr{C}_{3+},\quad |1,b|\subset\mathscr{K}_{3+},\quad \mathscr{K}_3\subset\mathscr{C}_{3-},$$
$$|d',d|\subset\mathscr{C}_{2+},\quad |-1,c|\subset\mathscr{K}_{2+},\quad \mathscr{K}_2\subset\mathscr{C}_{2-},$$

let $\mathscr{G}_3$ and $\mathscr{G}_2$ denote those functions whose domains are

$$\operatorname{dom}\mathscr{G}_3:=\{(r',r,s)\in C^3:r',r\in\mathscr{C}_{3+}\text{ and } s\in\mathscr{K}_{3+}\},$$
$$\operatorname{dom}\mathscr{G}_2:=\{(s',s,r)\in C^3:s',s\in\mathscr{C}_{2+}\text{ and } r\in\mathscr{K}_{2+}\},$$

and whose values are defined by the right sides of Eqs. (4.26) and (4.27), respectively. Then dom $\mathscr{G}_3$ and dom $\mathscr{G}_2$ are open spheroids of C^3 and are neighborhoods of (a',a,b) and (d',d,c), respectively. Moreover, $\mathscr{G}_3$ and $\mathscr{G}_2$ are holomorphic. (Note that $\mathscr{G}_3$ and $\mathscr{G}_2$ are extensions of the restrictions of G_3 and G_2 to the real sections of dom $\mathscr{G}_3$ and dom $\mathscr{G}_2$, respectively.)

Outline of Proof: Consider, for example, $\mathscr{G}_3$. The statements that dom $\mathscr{G}_3$ is an open spheroid in C^3 and is a neighborhood of (a',a,b) are evident. The first major step is to prove that the integrand in Eq. (4.26) is continuous over the set of all (r',r,s,τ) such that $(r',r,s)\in\operatorname{dom}\mathscr{G}_3$ and $\tau\in\mathscr{C}_3$. Then one proves that, for any fixed $\tau\in\mathscr{C}_3$, the integrand is a holomorphic function of each one of the complex variables r', r, and s when the other two variables are held fixed in value. (For example, for any given $\tau\in\mathscr{C}_3$, $r\in\mathscr{C}_{3+}$, and $s\in\mathscr{K}_{3+}$, the integrand is a holomorphic function of r' throughout $\mathscr{C}_{3+}$.) It follows from a theorem[28] on path integrals of functions of two complex variables that $\mathscr{G}_3(r',r,s)$ exists and is a holomorphic function of each one of the variables r', r, and s for any given values of the other two variables. Hence, from a theorem[29] of Hartogs, $\mathscr{G}_3(r',r,s)$ is a holomorphic function of (r',r,s) throughout dom $\mathscr{G}_3$. Q.E.D.

The above lemma taken together with well-known theorems[29] on holomorphic functions of many complex variables directly yields the following two theorems.

Theorem: G_3 and G_2 are real analytic functions throughout their domains. Also, there exist holomorphic extensions of G_3 and G_2 to open subsets of C^3.

Theorem: For any given $\tau\in\mathscr{C}_3$, the integrand in Eq. (4.26) is a real analytic function of (r',r,s) and

$$\frac{\partial^{k+m+n}G_3(r',r,s)}{\partial r'^k\,\partial r^m\,\partial s^n}=\frac{1}{2\pi i}\int_{\mathscr{C}_3}d\tau\,\frac{\partial^{k+m+n}}{\partial r'^k\,\partial r^m\,\partial s^n}\left[\frac{\chi_2(s,\tau)}{\mu(r',r,\tau)}\right]$$

for any non-negative integers k, m, and n (where, of course, the above integral exists). A like statement holds for partial differentiation in Eq. (4.27).

Theorem:

$$\left[2(s-r)\frac{\partial^2}{\partial r\partial s}+\frac{\partial}{\partial r}-\frac{\partial}{\partial s}\right]G_3(r',r,s)=0 \tag{4.32}$$

and

$$\left[2(s-r)\frac{\partial^2}{\partial r\partial s}+\frac{\partial}{\partial r}-\frac{\partial}{\partial s}\right]G_2(s',s,r)=0. \tag{4.33}$$

Moreover, the above Eq. (4.32) remains valid if one replaces r and s in the operator in brackets by r' and s or by r' and r. Likewise, Eq. (4.33) remains valid if one replaces r and s in the operator in brackets by s' and r or by s' and s.

Proof: Equation (4.32) follows from the preceding theorem and from Eqs. (4.1), (4.2) and the equation

$$\left[2(s-r)\frac{\partial^2}{\partial r\partial s}+\frac{\partial}{\partial r}-\frac{\partial}{\partial s}\right]\{[(r-\tau)(s-\tau)]^{-1/2}\}=0.$$

The other parts of the theorem are proven in the same way.

Other properties of G_3 and G_2 that we shall need are

$$G_3(r',r,1)=1,\quad G_2(s',s,1)=1, \tag{4.34}$$

and

$$G_3(r,r,s)=\sqrt{\frac{1-r}{s-r}},\quad G_2(s,s,r)=\sqrt{\frac{s+1}{s-r}}. \tag{4.35}$$

Equations (4.34) are implied by Eqs. (4.1), (4.2), and (4.30). Equations (4.35) are implied by Eqs. (4.1), (4.2), (4.26), and (4.27).

Interested readers are invited to explore other properties of these functions including the relation between G_3 and G_2, their reduction to elliptic integrals and their relation to the Green's function employed by Szekeres.[1] Our only current reason for introducing G_3 and G_2 is that they furnish a neat way of proving that Eqs. (1.5) to (1.7) and Eq. (1.10) furnish a solution of the general initial value problem for the collinear polarization case. For this purpose, all of the properties that we shall need have been given above in our theorems and in Eqs. (4.34) and (4.35).

G. The final theorems

Definitions: Let ${}_3\psi$ and ${}_2\psi$ denote those functions which each have the domain IV and whose values are given by

$${}_j\psi(u,v):={}_j\gamma(r(u),s(v))\quad (j=3,2). \tag{4.36}$$

From Eqs. (4.8) and (4.9) and from the definitions and the theorem in Sec. IV C we see that ${}_3\psi$ and ${}_2\psi$ exist and are given by Eqs. (1.6) and (1.7). From Eqs. (2.20) and (4.17),

$${}_3\psi(0,v)=0,\quad {}_2\psi(u,0)=0 \tag{4.37}$$

for all $0\leqslant v<v_0$ and $0\leqslant u<u_0$, respectively. From Eqs. (3.14), (3.15), (4.8), and (4.9)

$${}_3\gamma(r,1)=\gamma_3(r),\quad {}_2\gamma(-1,s)=\gamma_2(s). \tag{4.38}$$

Therefore, Eqs. (2.20), (3.4), and (4.36) yield

$${}_3\psi(u,0)=\psi_3(u),\quad {}_2\psi(0,v)=\psi_2(v), \tag{4.39}$$

where we recall that $\psi_3(u):=\psi(u,0)$ and $\psi_2(v):=\psi(0,v)$. From Eqs. (3.4), (4.22), (4.23), and (4.36), we further obtain

$${}_3\psi(u,v)=\int_0^u du'\,\dot{\psi}_3(u')G_3[r(u'),r(u),s(v)] \tag{4.40}$$

and

$${}_2\psi(u,v)=\int_0^v dv'\,\dot{\psi}_2(v')G_2[s(v'),s(v),r(u)]. \tag{4.41}$$

We are now in a position to prove a key theorem.

Theorem: (a) $\partial_j\psi(u,v)/\partial u, \partial_j\psi(u,v)/\partial v$ and

$\partial^2{}_j\psi(u,v)/\partial u\partial v$ exist and are continuous throughout region IV. (b) ${}_j\psi$ satisfies the linear hyperbolic partial differential Eq. (1.4) throughout IV.

Proof: Consider ${}_3\psi$, for example. Part (a) of the theorem follows easily from the corollary in Sec. VI F concerning the analyticity of G_3 and from our premise that $r(u)$, $s(v)$ and $\psi_3(u)$ are C^1. To prove part (b), we first note that Eq. (4.32) and the relation $\rho(u,v) = [s(v) - r(u)]/2$ imply

$$\left[2\rho \frac{\partial^2}{\partial u\, \partial v} + \rho_u \frac{\partial}{\partial v} + \rho_v \frac{\partial}{\partial u}\right] G_3[r(u'),r(u),s(v)] = 0.$$

Therefore, from Eq. (4.40)

$$\left[2\rho \frac{\partial^2}{\partial u\, \partial v} + \rho_u \frac{\partial}{\partial v} + \rho_v \frac{\partial}{\partial u}\right]{}_3\psi(u,v)$$
$$= [s(v) - r(u)]\frac{\partial G_3[r(u),r(u),s(v)]}{\partial v}$$
$$+ \frac{1}{2}\dot{s}(v)G_3[r(u),r(u),s(v)].$$

The above equation and the first of Eqs. (4.35) imply that the right side of the above equation vanishes for all (u,v) in IV. Q.E.D.

The focus of this entire paper is the following corollary which is directly implied by the preceding theorem and Eqs. (4.37) and (4.39).

Corollary: ${}_3\psi + {}_2\psi$ satisfies Eq. (1.4) throughout IV and, moreover,

$$({}_3\psi + {}_2\psi)(u,0) = \psi(u,0), \quad ({}_3\psi + {}_2\psi)(0,v) = \psi(0,v)$$

for all $0 \leqslant u < u_0$ and $0 \leqslant v < v_0$, respectively.

V. PERSPECTIVES

The method of linear superposition that was used in this paper to obtain a general solution of colliding gravitational plane waves with collinear polarizations is not applicable to noncollinear polarizations. A key ingredient of this method, viz., the fact that the solutions ψ of Eq. (1.4) over region IV is a linear space, is lacking in the noncollinear case.

However, we have recently found an alternative method that yields our form of the solution of the collinear case but which is applicable to noncollinear as well as collinear polarizations. This method employs a 2×2 matrix homogeneous Hilbert problem (HHP).[30] It was suggested to us by an older 2×2 matrix HHP that the authors employ to compute stationary axisymmetric gravitational fields when the values of these fields on the symmetry axis are prescribed.[31–34]

To avoid any misunderstanding we stress that there is no known solution in a finite closed form of the general HHP for colliding plane waves. (This is also true of the HHP for stationary axisymmetric fields.) However, the HHP can be solved in a finite closed form for many interesting particular cases. Moreover, the HHP can be an effective starting point for reducing the initial value problem to certain classical linear equations (e.g., Fredholm equations of the second kind) which are throughly understood.

Although the HHP for colliding plane waves involves mathematical objects which are similar to those used in the HHP for stationary axisymmetric fields, we must caution that it is different in significant ways and that it is far more complicated as regards questions like those of existence and uniqueness of its solutions. [The mathematical difficulties are not eased by our present policy of not restricting our choices of $r(u)$ and $s(v)$ and of not making any highly restrictive assumptions concerning the differentiability class of the initial data functions.] The fact is that we are treading on new mathematical terrain.[35] For this reason, we shall postpone the exposition of the HHP until the third paper (III) of this series and first lay some groundwork.

This groundwork will be given in our next paper (II), which will chiefly be devoted to a one-dimensional Hilbert problem that is equivalent to the restriction of the HHP to the collinear case. In this one-dimensional Hilbert problem, $g_3(\sigma)$ and $g_2(\sigma)$ appear as given functions which have already been determined from the initial data by employing Eqs. (1.10). The solution of the one-dimensional Hilbert problem is a complex-valued function $\phi(r,s,\tau)$ of a complex parameters τ as well as of r and s. A generalization of a theorem of Plemelj supplies this solution, whereupon the potential ψ corresponding to the prescribed initial data is

$$\psi(u,v) = [-\tau\phi(r(u),s(v),\tau)]_{\tau=\infty},$$

which, as we have already noted, turns out to be the expression given by Eqs. (1.5)–(1.7).

Note that the class of ψ potentials given by Eqs. (1.5)–(1.7) and Eqs. (1.10) is actually broader than the class of ψ potentials corresponding to colliding gravitational plane waves with collinear polarizations. For the proof of these equations in Secs. III and IV we assumed only that $r(u)$, $s(v)$, $\psi(u,0)$, and $\psi(0,v)$ are C^1 and that $\dot{r}(u) > 0$ if $u > 0$ and $\dot{s}(v) < 0$ if $v > 0$. We did not assume that $\dot{r}(0) = \dot{s}(0) = 0$ as is required by the vacuum field equations. Nor did we assume that the colliding wave conditions (2.36) and (2.37) hold.

In our next paper (II), we shall continue this policy of not assuming that $\dot{r}(0) = \dot{s}(0) = 0$ and of not assuming the colliding wave conditions when we give the one-dimensional Hilbert problem and its general solution. However, we shall also devote some space to an investigation of the colliding wave conditions [understood to include $\dot{r}(0) = \dot{s}(0) = 0$] for the collinear case and of corresponding conditions that are satisfied by $g_3(\sigma)$ and $g_2(\sigma)$. This investigation will yield especially detailed results for those metrics for which $\rho(u,v) = 1 - u^2 - v^2$ since $g_3(\sigma)$ and $g_2(\sigma)$ are then expressible in terms of $\psi_u(u,0)$ and $\psi_v(0,v)$ by means of simple (and proper) Riemann integrals.

Note added in proof: The conclusions of a theorem in Sec. III B on the Hölder conditions obeyed by g_j can be strengthened. In our next paper, we shall prove that (for $\nu = 0$) g_j obeys a Hölder condition of index 1/2 on any closed subinterval of $]-1, r_0[$ if $j = 3$ and $]s_0, 1[$ if $j = 2$.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. PHY-8605958 from the National Science Foundation.

[1]P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[2]B. C. Xanthopoulos, J. Math. Phys. **27**, 2129 (1986). The space-times covered in this reference include a Maxwell field to which the method of Riemann is also applied.
[3]See, e.g., M. Bôcher, *An Introduction to the Study of Integral Equations*

(Cambridge U.P., London, 1914), 2nd ed., p. 6.

[4]Along lines originally taken by Abel (see e.g., Ref. 3, pp. 8–11), a further generalization is obtained if one replaces $\pi(r-\sigma)^{1/2}$ in Eqs. (1.9), (3.14), and (3.15) by $(\pi/\sin\lambda\pi)\ (r-\sigma)^{\lambda}$, where λ is any real constant such that $0<\lambda<1$. Correspondingly, one replaces $(\sigma-r)^{1/2}$ in Eqs. (1.10), (3.5), and (3.6) by $(\sigma-r)^{1-\lambda}$. With obvious modifications, the proofs of these equations in Sec. III are applicable to this further generalization if one replaces $\pi(r-\sigma)^{1/2}\ (\sigma-r')^{1/2}$ in Eq. (3.13) by $(\pi/\sin\lambda\pi)\ (r-\sigma)^{1-\lambda}\ (\sigma-r')^{\lambda}$.

[5]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); **168**, 1415 (1968); J. Math. Phys. **15**, 1409 (1974).

[6]F. J. Ernst, A. García-Diaz, and I. Hauser, J. Math. Phys. **28**, 2951 (1987).

[7]We assume that the prescribed initial values $\rho(u,0)$, $E(u,0)$ and $\rho(0,v)$, $E(0,v)$ are maximally extended on the u and v axes, respectively, while subject to the requisite conditions which are given in Sec. II. Note that, unlike others, we do not scale u and v to make $u_0=v_0=1$. Nor do we assume that u and v can be chosen so that $\rho(u_0,0)=\rho(0,v_0)=0$ since this may not be possible for all colliding gravitational plane wave metrics.

[8]We owe this use of the Heavyside symbol to K. A. Kahn and R. Penrose, Nature **229**, 185 (1971), which contains the first published example of an impulsive colliding gravitational plane wave metric. The value of $\theta(0)$ is arbitrary and some authors do not even assign a value. We choose to make $\theta(0)=0,1$ (double valued) since that is convenient for our derivations.

[9]N. Rosen, Phys. Z. Sowjet. **12**, 366 (1937).

[10]Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).

[11]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985); A **408**, 175 (1986); A **410**, 311 (1987).

[12]F. J. Ernst, A. García-Diaz, and I. Hauser, J. Math. Phys. **28**, 2155 (1987).

[13]V. Ferrari, J. Ibañez, and M. Bruni, Phys. Rev. D **36**, 1053 (1987).

[14]F. J. Ernst, A. García-Diaz, and I. Hauser, J. Math. Phys. **29**, 681 (1988).

[15]As regards E, it is sufficient in Sec. II (except in parts of Sec II K) to assume that E is C^1 and that E_{uv} exists and is continuous throughout IV. It is also possible to introduce other relaxations of our premises [such as finite number of finite step discontinuities in $E(u,0)$ and $E(0,v)$]. However, we prefer not to do that here.

[16]L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1957, 1959), Parts I and II.

[17]See, e.g., B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1957), pp. 137–143.

[18]In particular, Appendix A of Ref. 11 supplies conform tensor components directly in terms of E and its derivatives.

[19]F. J. Ernst, J. Math. Phys. **15**, 1409 (1974).

[20]I. Hauser, J. Math. Phys. **19**, 661 (1978).

[21]I. Hauser and F. J. Ernst, J. Math. Phys. **20**, 1041 (1979).

[22]One can obtain this expression for C_2 (apart from an irrelevant phase factor) from Sec. V of Ref. 12. Reference 12 employs specific null coordinates for which $\rho(u,v)=1-u^2-v^2$ but a formal transformation to general null coordinates is easily effected.

[23]F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955), p. 55.

[24]J. H. Williamson, *Lebesgue Integration* (Holt, Rinehart and Winston, New York, 1962), Theorems 4.2b and 4.2c on pp. 64–65.

[25]See Theorem 3.3b on p. 48 and Corollary 4 on p. 51 of Ref. 24.

[26]Lemma 5.2a on p. 83 of Ref. 24.

[27]Theorem 3.2b on p. 43 of Ref. 24.

[28]See, e.g., M. A. Evgrafov, *Analytic Functions* (Dover, New York, 1978), Theorem 4.3 on p. 39.

[29]See, e.g., S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, NJ, 1948), Chap. II.

[30]The standard treatise on homogeneous Hilbert problems is N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).

[31]I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980).

[32]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-Verlag, Berlin, 1984), pp. 128–175.

[33]I. Hauser and F. J. Ernst, "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).

[34]The HHP that we employed to generate stationary axisymmetric gravitational fields can also be used to generate some colliding gravitational plane waves. This was done, e.g., in Ref. 14.

[35]For example, consider that our HHP for colliding plane waves is an HHP on arcs (as opposed to our HHP for stationary axisymmetric fields which is defined on a contour). The one-dimensional HHP on arcs is covered in Chap. 10 of Ref. 30 under premises that are a bit too restrictive for our needs. The matrix HHP on arcs is not even mentioned in Ref. 30.

Separable coordinates and particle creation. II: Two new vacua related to accelerating observers

Isaias Costa[a)]
Institut für Theoretische Physik, Universität Wien, Wien, Austria and Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro 22290, Brazil

(Received 13 November 1986; accepted for publication 9 November 1988)

An exactly solvable example of quantum field theory in a nonstationary system is presented, which has an inertial and a uniform accelerated asymptotic region. Two sets of solutions are constructed that are quasiclassical in each of these regions and they are compared. The Bogoliubov coefficients have thermal character and show a temperature of $a_\infty/2\pi$, where a_∞ is the asymptotic acceleration. This result is indeed what one would expect on the grounds of the Hawking effect.

I. INTRODUCTION

This work has its place at the interface of quantum field theory and gravitation. There we attack the very first problem of writing quantum field theory in general coordinates without leaving flat space-time. This is known, after the work of Fulling,[1] to be anything but trivial: the very concept of particle is not well defined,[2] so that what appears to be the vacuum to an inertial observer will look like a thermal state for uniformly accelerated observer with a temperature proportional to his acceleration. In the technically very similar Hawking effect,[3] the acceleration will be replaced by the surface gravity of a black hole showing the first well-established physical effect of quantum gravity.[4]

In this series of papers we investigate the two-dimensional Minkowski space with the help of the separable orthogonal coordinates with which the massive Klein–Gordon equation separates.[5] They play the same important role as their Euclidean equivalents—polar, elliptic, and parabolic coordinates—play in all physics. These coordinates are adapted to very interesting physical situations like a global boost and an observer that is inertial in the past and uniformly accelerated in the future. They are also useful to study compact regions of the Minkowski plane. They are often not static and, because of their simplicity, give us hope to master this difficult situation. They are also easily generalized to more dimensions.[6] And, above all, they are such that the Klein–Gordon equation (and most other interesting quantities) is exactly solvable, distilling the physical understanding from the mathematical difficulties of the problem.

We solve the special case of an accelerating observer in Sec. II. This example provides us with a clear interpretation of the effect of particle creation, thus giving an alternative to the detector analysis,[7] which has been so deeply criticized by Grove and Ottewill and by Hinton.[8] In Sec. III there is a discussion.

II. TWO NEW VACUA

We now proceed to construct quantum field theory in the special case of an accelerating coordinate system.[5] We choose it for three reasons: its well-known asymptotes; its techical simplicity; and because it gives a unique opportunity to exactly study a system that describes inertial and accelerated movements asymptotically, allowing for the interpretation of particle creation in the framework of only one system of observers. This has an obvious episthemological superiority to the Rindler systems.[9] The straightforward application of the recipe given here to other systems will be developed in forthcoming papers.

Like Letaw and Pfautsch[10] we merely calculate the solutions of the Klein–Gordon equation in two systems of coordinates with the special boundary conditions that we call quasiclassical. After normalizing them, we calculate the Bogoliubov coefficients that give the expectation value of the number operator in this basis over the vacuum of the Cartesian plane waves. We use the conventions and notation of Birrel and Davies.[11]

We consider the following transformation of coordinates:

$$\begin{aligned} t+x &= (2/w)\sinh w(T+X), \\ t-x &= (-1/w)\exp -w(T-X), \end{aligned} \tag{1}$$

where (t,x) are Cartesian and (T,X) are accelerating coordinates (see Fig. 1). In the latter coordinates, the Minkowski metric is written as

$$ds^2=[e^{-2wT}+e^{2wX}]\,[dT^2-dX^2]\,. \tag{2}$$

The proper time along $X=X_0$, which is everywhere timelike, is then

$$s=(1/w)\sqrt{e^{-2wT}+e^{2wX_0}}+(1/w)e^{wX_0}\operatorname{arcsinh}e^{w(T+X_0)}\,. \tag{3}$$

The acceleration along $X=X_0$ is

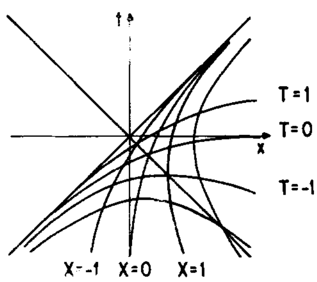


FIG. 1. The accelerating coordinate system, with $w=1$.

[a)] Permanent address: Centro Brasileiro de Pesquisas Fisicas, Rua Dr. Xavier Sigaud, 150, Rio de Janeiro 22290, Brazil (LC54).

 0022-2488/89/040888-04$02.50

$$a = [e^{-2wT} + e^{2wX_0}]^{-3/2} we^{2wX_0}, \tag{4}$$

so that

$$\lim_{T\to -\infty} a(T,X_0) = w \exp(2wX_0 + 3wT) = 0 \tag{5}$$

and

$$\lim_{T\to -\infty} a(T,X_0) = w \exp(-wX_0) = :a_\infty . \tag{6}$$

The curve $T = T_0$ is a Cauchy surface for the semiplane $t - x < 0$, since it is spacelike and has either a spacelike curve or the boundary $t + x = 0$ as an asymptote. On the problem of analytic continuation over the horizon see Ref. 12. For more details see Ref. 13.

Here, the Klein–Gordon equation is

$$(\partial_T^2 - \partial_X^2) U(T,X) = -m^2(e^{-2wT} + e^{2wX}) U(T,X) \tag{7}$$

and separates in the following equations:

$$\frac{d^2F}{dT^2} + (m^2w^2Y^2 + K^2)F = 0, \tag{8a}$$

$$\frac{d^2G}{dX^2} - (m^2w^2Z^2 - K^2)G = 0, \tag{8b}$$

with

$$Y = e^{-wT}/w,$$
$$Z = e^{wX}/w, \tag{9}$$

and

$$U(T,X) = F(T)\cdot G(X). \tag{10}$$

Equations (8) are two Bessel equations with imaginary order $i\nu$, where we write

$$\nu = K/w \tag{11}$$

for simplicity.

To state the orthonormality conditions

$$\langle U_K | U_L \rangle = \delta(K - L), \quad \text{etc.}, \tag{12}$$

we define the scalar product as

$$\langle U,\overline{U} \rangle : = -i \int_S dS^a\, U \overleftrightarrow{\partial}_a \overline{U}^*, \tag{13}$$

where S^a is a Cauchy surface and dS^a is a future pointing unit vector:

$$dS^a = (g)^{1/2} g^{ab} \epsilon_{bc}\, dx^c, \tag{14}$$

that is,

$$dS^a = \eta^{ab} \epsilon_{bc}\, dX^c. \tag{15}$$

As the Cauchy surface can be chosen to be the coordinate line $T = T_0$, we have $dT = 0$ and

$$dS^a = \delta_0^a\, dx, \tag{16}$$

so that

$$\langle U | \overline{U} \rangle = -iF \overleftrightarrow{d} F^* \int dx\, G\overline{G}^*. \tag{17}$$

The use of Eq. (8b) and partial integration give

$$\langle U_K | U_L \rangle = [-iF_K \overleftrightarrow{d}_T F_L^*/(K^2 - L^2)]|_{T_0}\, G_K \overleftrightarrow{d}_X G_L^* |_{X_{\min}}^{X_{\max}}. \tag{18}$$

We now need the boundary conditions. For that we substitute

$$U(T,X) = A(T,X) \exp[iS(T,X)] \tag{19}$$

in the Klein–Gordon equation and obtain

$$-m^2 U = A_{;a}{}^a e^{iS} + i(2A_{;a} S^{;a} e^{iS} + S_{;a}{}^a U) - S_{;a} S^{;a} U. \tag{20}$$

In the limit of geometrical optics, or quasiclassical, where A varies slower than S, that is, where

$$\frac{\Box A}{A} \ll S_{;a} S^{;a} - m^2, \tag{21}$$

the real part of Eq. (20) becomes the Hamilton–Jacobi equation,

$$S_{;a} S^{;a} = m^2. \tag{22}$$

In the conformal metric it is written as

$$\left(\frac{\partial S}{\partial T}\right)^2 - \left(\frac{\partial S}{\partial X}\right)^2 = w^2 m^2 (Y^2 + Z^2) \tag{23}$$

and then

$$S(T,X) = \pm \int dT \sqrt{K^2 + m^2 w^2 Y^2(T)} \pm \int dX \sqrt{K^2 - m^2 w^2 Z^2(X)}, \tag{24}$$

so that $A \exp[iS(T,X)]$ gives the quasiclassical asymptote of the solution.

We illustrate this in the special case of accelerating coordinates. The boundary conditions select the Bessel functions that have as asymptotes

$$\begin{aligned} &\lim_{Y\to 0} F_0 \sim e^{\pm iKT} \sim Y^{\pm i\nu}, \\ &\lim_{Y\to\infty} F_\infty \sim e^{\pm imY}, \\ &\lim_{Z\to 0} G_0 \sim e^{\pm iKX} \sim Z^{\pm i\nu}, \\ &\lim_{Z\to\infty} G_\infty \sim e^{\pm mZ}. \end{aligned} \tag{25}$$

These are

$$\begin{aligned} &F_0 = J_{\pm i\nu}(mY), \\ &F_\infty = H_{i\nu}^2(mY) \text{ or } H_{i\nu}^1(mY), \\ &G_0 = I_{\pm i\nu}(mZ), \\ &G_\infty = K_{i\nu}(mZ) \text{ or } I_{i\nu}(mZ). \end{aligned} \tag{26}$$

We have to disregard the solutions with $I_{i\nu}$ since they diverge at large arguments. We remain with two complete sets of orthonormal exact solutions:

$$U_\nu^{\text{inert}} = \tfrac{1}{2} \sqrt{\nu(1 - e^{-2\pi\nu})/\pi w}\, H_{i\nu}^1(mY) K_{i\nu}(mZ), \tag{27}$$

$$U_\nu^{\text{acc}} = \sqrt{(\nu/\pi w)}\, J_{i\nu}(mY) K_{i\nu}(mZ). \tag{28}$$

We call them inertial and accelerated solutions because they are adapted to observers that are asymptotically inertial and accelerated, respectively, as they are quasiclassical at $Y\to\infty$, $Z\to\infty$ and at $Y\to 0$, $Z\to\infty$. Note that the inertial modes are not included in the Sanchez[14] class of vaccua as they are not eigenstates of the operator ∂_T even asymptotically. For details, see Ref. 13.

We may, in general, compare two bases U_K and V_L through the Bogoliubov coefficients A and B:

$$U_K = A_{KL} V_L + B_{KL} V_L^* . \tag{29}$$

The vacuum states defined with the help of annihilation operators are also related by means of these coefficients. In particular, the expectation value of the number operator of V particles in the U vacuum is given by

$$\langle 0_K | N_L | 0_K \rangle = \int dK \, |B_{KL}|^2 . \tag{30}$$

We now compare the two bases $U^{\rm acc}$ and $U^{\rm inert}$. We will indulge the reader with some detail in the calculations to stress the importance of the boundary conditions, as they determine the Bogoliubov coefficients and particle creation. This is in accordance with Sanchez.[14] The Bogoliubov coefficient,

$$B^{\rm acc\ inert}_{\nu\mu} = -\langle U^{\rm acc}_\nu, U^{\rm inert*}_\mu \rangle \tag{31}$$

is, much like Eq. (17),

$$B^{\rm acc\ inert}_{\nu\mu} = \frac{i}{\nu^2 - \mu^2} \left(\frac{1}{w^2} \frac{\sqrt{\nu}}{\pi w} \frac{\sqrt{\mu(1 - e - 2\pi u)/\pi w}}{2} \right) \times \left[J_{i\nu} \overleftrightarrow{wY\partial_Y} H^1_{i\mu} \right|_\infty \left[K_{i\nu} \overleftrightarrow{wZ\partial_Z} K_{i\mu} \right|_0^\infty , \tag{32}$$

so that

$$B^{\rm acc}_{\nu\mu}\ \text{inert} = (w\sqrt{e^{2\pi\nu} - 1})^{-1} \delta(\nu - \mu) . \tag{33}$$

For these calculations see Ref. 13.

Next we compare these bases with the Cartesian one, that is, with plane waves in the original coordinates,

$$u_k = \tfrac{1}{2} (\pi\epsilon)^{-1/2} \exp[-i(\epsilon t - kx)] . \tag{34}$$

In the variables (Y,Z), it holds that

$$\epsilon t - kx = -\frac{1}{2} \left[\left(\frac{\epsilon - k}{w} \right) \left(\frac{Y}{Z} - \frac{Z}{Y} \right) + w(\epsilon + k) YZ \right] \tag{35}$$

and

$$\partial_T u_k = -\frac{i}{2} \left[(\epsilon - k) \left(\frac{Y}{Z} + \frac{Z}{Y} \right) + w^2 (\epsilon + k) YZ \right] u_k . \tag{36}$$

Their limits when $Y \to 0$ are

$$\epsilon t - kx = [(\epsilon - k)/2wY] Z , \tag{37}$$

$$\partial_T u_k = -(i/2)[(\epsilon - k)/Y] Z u_k , \tag{38}$$

so that

$$u_k = \tfrac{1}{2} (\pi\epsilon)^{-1/2} \exp\{ -i[(\epsilon - k)/2wY] Z\} \tag{39}$$

and

$$\partial_T u_k = -\frac{i}{4} \frac{\epsilon - k}{Y} Z (\pi\epsilon)^{-1/2} \exp\left(-i \frac{\epsilon - k}{2wY} Z \right) . \tag{40}$$

The Bogoliubov coefficient

$$B_{\nu k} = -\langle U_\nu | u_k^* \rangle \tag{41}$$

is

$$B_{\nu k} = -iF'_\nu \int dX \, G_\nu u_k + iF_\nu \int dX \, G_\nu \partial_T u_k . \tag{42}$$

We calculate the two integrals

$$I_1 := \int dX \, G_\nu (X) u_k (t(T,X), x(T,X)) \tag{43}$$

and

$$I_0 := \int dX \, G(X) \partial_T u_k (t(T,X), x(T,X)) . \tag{44}$$

If we put the expression (39) in Eq. (44) and use

$$G_\nu = K_{i\nu} (mZ) \tag{45}$$

we get the Fourier transformation

$$I_0 = \frac{1}{2\sqrt{\pi\epsilon}} \left[-i \left(\frac{\epsilon - k}{2Y} \right) \right] \times \int \frac{dZ}{w} K_{i\nu}(mZ) \exp\left[-i \left(\frac{\epsilon - k}{2wY} \right) Z \right] . \tag{46}$$

These two integrals are easy to evaluate[15] in the limit $Y \to 0$:

$$I_0 = \frac{i}{4 \sinh \pi\nu} \left(\frac{\pi}{\epsilon} \right)^{1/2} \left[\left(\frac{\epsilon - k}{wmY} \right)^{i\nu} e^{-\pi\nu/2} - \left(\frac{\epsilon - k}{wmY} \right)^{-i\nu} e^{\pi\nu/2} \right] \tag{47}$$

and

$$I_1 = \frac{-1}{4\nu \sinh \pi\nu} \left(\frac{\pi}{\epsilon} \right)^{1/2} \left[\left(\frac{\epsilon - k}{wmY} \right)^{i\nu} e^{-\pi\nu/2} + \left(\frac{\epsilon - k}{wmY} \right)^{-i\nu} e^{\pi\nu/2} \right] . \tag{48}$$

After the inclusion of F_ν we get the Bogoliubov coefficients

$$B^{\rm acc\ cart}_{\nu k} = \left(\frac{\nu}{\epsilon w} \right)^{1/2} \frac{\exp(-\pi\nu/2)}{(i\nu)! 2 \ \sinh(\pi\nu)} \left(\frac{\epsilon - k}{2w} \right)^{i\nu} \tag{49}$$

and

$$B^{\rm inert\ cart}_{\nu k} = (\pi\sqrt{2\nu\epsilon w} \sinh(\pi\nu))^{-1} \times {\rm Re}\{ (-i\nu)! [(\epsilon - k)/2w]^{i\nu} \} . \tag{50}$$

Their absolute squares are

$$|B^{\rm acc\ cart}_{\nu k}|^2 = (2\pi\epsilon\omega)^{-1} (e^{2\pi\nu} - 1)^{-1} \tag{51}$$

and

$$|\beta^{\rm inert\ cart}_{\nu k}|^2 = [2\pi^2 \nu\epsilon w \sinh(\pi\nu)]^{-1} \times {\rm Re}^2 \{ (-i\nu)! \, [(\epsilon - k) 2w]^{i\nu} \} . \tag{52}$$

We discuss these results in the following section.

III. CONCLUSION

Before we proceed to see what can be said with the results we have at hand, we want to remark that one should be careful in coming to conclusions using only the expectation value of the number operator, as shown by many works.[16] We need to have the expectation value of the energy-momentum tensor, the Green's functions, the density matrix, and so on, to have a complete understanding of the problem.

The Planckian form of Eq. (51) in the last section seems to indicate that the asymptotically accelerated vacuum is, compared to the Cartesian one, a thermal state with temperature

$$\Theta_0 = w/2\pi k_B. \tag{53}$$

After Tolmann[17] this would mean that in the proper frame the temperature is

$$\Theta = \Theta_0 (g_{00})^{-1/2} \tag{54}$$

or, with

$$\lim_{Y \to 0} g_{00} = \exp(2wX), \tag{55}$$

$$\Theta = w \exp(-wX)/2\pi k_B = a_\infty /2\pi k_B. \tag{56}$$

Also, compared to the asymptotically inertial vacuum, the asymptotically accelerated vacuum would have the same Fulling temperature. The asymptotically inertial vacuum compared to the Cartesian one cannot be a thermal state.

We illustrate the situation in the following diagram:

$$\begin{tikzcd}
\boxed{\text{asymptotically inertial}} \arrow[rr, leftrightarrow, "\Theta = a_\infty/2\pi k_B"] \arrow[dr, leftrightarrow, "\text{nonthermal}"'] & & \boxed{\text{asymptotically accelerated}} \arrow[dl, leftrightarrow, "\Theta = a_\infty/2\pi k_B."] \\
& \boxed{\text{Cartesian}} &
\end{tikzcd}$$

The Fulling effect make this result expected. The observer defines its "natural"—that is, in accordance with our boundary conditions—vacuum as long as it is inertial; then it accelerates and reaches a uniform acceleration a_∞, thus seeing a thermal sea of particles around it, with temperature $\Theta = a_\infty/2\pi k_B$. For this observer the "natural" particle number is not conserved.

The temperature between accelerated and Cartesian modes is in accordance with the result of Sanchez for massless particles.[18] It is not surprising, however, that the asymptotically inertial vacuum is not a perfect vacuum as compared to the Cartesian one: the exactness of the approach forces the asymptotically inertial states to contain vestiges of the nonstationarity. This is because ∂_T is not a Killing-vector field in this region and you cannot choose a solution that is an eigenfunction of ∂_T there.

In a following paper we intend to extend this analysis to the calculation of other physical magnitudes, like the expectation value of the energy-momentum tensor. Likewise we will consider the other separable orthogonal systems of coordinates. We hope then to be able to attack the problem at a more abstract level, guided by our new knowledge of nonstatic systems of coordinates. The generalization to three dimensions allows the immediate verification of a speculation of Grove and Otewill,[8] that says that rigidity is the criterium to choose the "good" detector by handling a system of coordinates that is rigid but nonstatic.

ACKNOWLEDGMENTS

I want to thank Professor H. Rumpf and Professor H. Urbantke for suggesting the subject of this work and for their orientation, Professor R. Beig for his support and many helpful conversations, and the whole Institut für Theoretische Physik of the University of Vienna for its hospitality during my stay there. I am also very glad with the cooperation and friendship found here in Centro Brasileiro de Pesquisas Fisicas.

This work was supported in part by CAPES—Education Ministery of Brazil, and CNPq—Science Ministry of Brazil.

[1] S. A. Fulling, Phys. Rev. D 7, 2850 (1973).

[2] P. C. W. Davies, in *Quantum Theory of Gravity: Essays in honor of the 60th birthday of Bryce De Witt*, edited by C. M. Christensen (Hilger, London, 1984).

[3] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[4] J. S. Bell, R. J. Hughes, and J. M. Leinaas, "Electrons as accelerated thermometers," Preprint CERN, TH 3948 (1984).

[5] I. Costa, Rev. Brasil. Fis. **17**, 585 (1987).

[6] E. G. Kalnins, SIAM J. Math. Anal. **6**, 340 (1975); E. G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1025 (1974); **16**, 1531 (E) (1975); E. G. Kalnins and W. Miller, Jr., J. Math. Phys. **19**, 1233 (1978).

[7] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).

[8] P. G. Grove and A. C. Ottewill, J. Phys. A **16**, 3905 (1983); K. J. Hinton, Class. Quantum Gravit. **1**, 27 (1984).

[9] W. Rindler, Am. J. Phys. **34**, 1174 (1966).

[10] J. R. Letaw, Phys. Rev. D **23**, 1709 (1981); J. R. Letaw, and J. D. Pfautsch, *ibid.* **22**, 1345 (1980); **24**, 1491 (1981); J. Math. Phys. **23**, 425 (1982).

[11] N. D. Birrel and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U. P., Cambridge, 1984).

[12] H. Rumpf, Phys. Rev. D **24**, 275 (1982).

[13] I. Costa, dissertation, Universität Wien, 1985.

[14] N. Sanchez, These d'Etat, Universite de Paris VII, 1979; Phys. Rev. D **24**, 2100 (1981); "The relativity of vacuum," in *Quantum Gravity*, edited by M. A. Markov and P. C. West (Plenum, London, 1984).

[15] *Tables of Integral Transforms I*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Eq. 4.16.29.

[16] We thank the referee for pointing this point out and for the reference: Freese *et al.* Nucl. Phys. B **255**, 693 (1985); see also Refs. 2, 8, and 11.

[17] R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Clarendon, Oxford, 1934).

[18] N. Sanchez, Ref. 14, especially Appendix II.1, Eq. 4.7.

Parastatistics and conformal field theories in two dimensions

Freydoon Mansouri and Xizeng Wu
Physics Department, University of Cincinnati, Cincinnati, Ohio 45221

(Received 25 May 1988; accepted for publication 2 November 1988)

The relation between parafermion field theories of order Q and the corresponding fermion field theories with SO(Q) symmetry is studied. It is shown that these theories are related but not identical. The explicit relation between the states and the observables of the two classes of theories are given without using the Klein transformations. The formalism is applied to the free conformally invariant parafermion theories in two dimensions. Their Virasoro algebra and SO(N) Kac–Moody algebra are given. The equivalence of their canonical form of the energy-momentum tensor with the Sugawara–Sommerfield form is also elucidated.

I. INTRODUCTION

The quantum field theory of particles that obey parastatistics of some order Q was formulated by Green[1,2] some time ago. Since such particles do not exist in nature as far as we know, the applications of these generalized statistics to physical problems have been few. Notable among these are the paraquark model[3] and the parastring models.[4] Recently, there has been renewed interest on this subject in connection with the possibility of putting bounds on the violation of Pauli's exclusion principle.[5,6]

The general properties of the quantum field theories obeying parastatistics have been established, in a model-independent way, in a number of theorems.[2,7] The most important consequence of these theorems is that there is a relation between a parafield theory of a given order Q and a field theory obeying standard statistics but having an internal symmetry such as SO(Q) or SU(Q). This relation is often referred to as "equivalence."[2,7] The proof of such an equivalence depends on a number of restrictions that are imposed on the two theories under consideration. For example, in the comparison of the spectrum of the original paraquark model[8] with an SU(3) colored quark theory, the word equivalence does not mean a unique one-to-one correspondence between all aspects of the two theories. It only refers to the relation between the bosonic and fermionic bound states of the paraquark theory, on the one hand, and the color singlet states of the SU(3) theory, on the other. In fact, to extend the equivalence to the corresponding gauged theories, the original form of the Green's ansatz[1] had to be replaced by a new one.[9]

The main purpose of this work is to shed light on the nature of the equivalences mentioned above in the context of model canonical quantum field theories. One distinct advantage of our approach to the more traditional axiomatic approach is that the various properties of the corresponding theories can be scrutinized in detail. Although much of the formalism we develop is applicable to parafield theories in any number of space-time dimensions, we choose our specific examples from those in two dimensions. One reason for this is that two-dimensional theories possess a number of properties not found in any other dimension. For example, the conformal group in two dimensions is infinite dimensional, and this leads to an intimate relation between the Virasoro and the Kac–Moody algebras of a conformally invariant field theory.[10,11] These algebras are characterized by their central charges, which are thus among the "observables" of a given theory. As pointed out by Witten,[12] these observables can be used to study the equivalence, or lack thereof, between two theories.

The explicit two-dimensional models that we use to illustrate most of our points are free chiral O(N) parafermion theories of order Q and free chiral O(N) $\times$ O(Q) fermion theories. We find that these theories are related but not identical. The Hilbert space A of the parafermion theory is in one-to-one correspondence with a subspace H of the Hilbert B of the fermion theory, in the sense that for every state (ray) in A, there is a state in H, and vice versa.[13] Similarly, the observables in the Hilbert space B can be divided into two parts, those that leave the subspace H invariant and those that do not. The observables of the Hilbert space A are in one-to-one correspondence with the subset that leave the subspace H invariant. In particular, we find that the Virasoro and the O(N) Kac–Moody algebras of the theory in B belong to the subset leaving H invariant. In many respects, the relation between the Hilbert space B and its subspace H is similar to that between the original Ramond–Neveu–Schwarz form of superstrings[14] and the more recent Green–Schwarz version.

The plan of this work is as follows: In Sec. II, we review the structure of free chiral O(N) $\times$ O(Q) fermion theories and construct both the canonical and the Sugawara–Sommerfield forms of the energy-momentum tensor. We also obtain the Virasoro and the Kac–Moody algebras of these theories. In Sec. III, we present in some detail the properties of the parafermionic Hilbert space A and its relation to the fermionic Hilbert space B. We do this without introducing Klein transformations that are nonlinear and nonlocal. Instead, we use an ansatz suggested by Domokos and Kovesi-Domokos and extended by Greenberg and Macrae,[9] which is linear. In Sec. IV, we discuss the properties of free chiral O(N) parafermion theories of order Q in two dimensions and compare their properties with the properties of the corresponding O(N) $\times$ O(Q) fermion theories.

II. O(N)$\times$O(Q) FREE FERMIONS IN TWO DIMENSIONS

We consider a free Majorana fermion theory with chiral O(N) $\times$ O(Q) symmetry in two space-time dimensions. The action of the theory is

 0022-2488/89/040892-10$02.50

$$S_1 = \frac{i}{2}\int d^2x\, \overline{\Psi}^{i\alpha}\gamma^a \partial_a \Psi^{i\alpha}, \tag{2.1}$$

where

$$a = 0,1, \quad i = 1,...,N, \quad \alpha = 1,...,Q,$$

and, e.g.,

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^5 = \gamma^0\gamma^1. \tag{2.2}$$

Following Witten,[12] we use a light cone basis for both the coordinates and the spinors:

$$x_\pm = 1/\sqrt{2}(x^0 \pm x^1), \tag{2.3}$$

$$\Psi^{i\alpha} = \begin{pmatrix} \Psi^{i\alpha}_- \\ \Psi^{i\alpha}_+ \end{pmatrix}. \tag{2.4}$$

In this basis, the action (2.1) takes the form

$$S_1 = \frac{i}{2}\int dx_+\, dx_- [\Psi^{i\alpha}_-\, \partial_+ \Psi^{i\alpha}_- + \Psi^{i\alpha}_+\, \partial_- \Psi^{i\alpha}_+]. \tag{2.5}$$

The nontrivial anticommutators for the chiral fields are

$$\{\Psi^{i\alpha}_\pm(x_\pm), \Psi^{j\beta}_\pm(x'_\pm)\} = \delta^{ij}\delta^{\alpha\beta}\delta(x_\pm - x'_\pm) \tag{2.6}$$

The chiral O(N) and O(Q) currents are given, respectively, by

$$J^{ij}_\pm(x^\pm) = \frac{i}{2\sqrt{2}} {:}\Psi^{i\alpha}_\pm \Psi^{j\alpha}_\pm{:}, \tag{2.7}$$

and

$$J^{\alpha\beta}_\pm(x^\pm) = \frac{i}{2\sqrt{2}} {:}\Psi^{i\alpha}_\pm \Psi^{i\beta}_\pm{:}. \tag{2.8}$$

In these expressions, normal orderings are with respect to the creation and annihilation operators of the chiral fields. The currents satisfy the conservation laws

$$\partial_\mp J^{ij}_\pm = \partial_\mp J^{\alpha\beta}_\pm = 0. \tag{2.9}$$

It follows that, for both O(N) and O(Q) currents,

$$J_\pm = J_\pm(x_\pm) \tag{2.10}$$

and

$$\Box J_\pm = 0. \tag{2.11}$$

Since they satisfy free-field equations, we are justified in treating them as free bosonic fields and in expanding them in normal modes.[15] To simplify notation, we frequently replace the pair of indices (ij) by a single index (a) and ($\alpha\beta$) by (m). In this notation, a general expression for the currents will have the form

$$J^a_\pm = (i/2\sqrt{2}) T^a_{ij} {:}\Psi^i_\pm \Psi^j_\pm{:}, \tag{2.12}$$

where T^a are the representation matrices of the generators of the symmetry group. If the fermions transform according to the fundamental representation of SO(N), say, then

$$T^a_{ij} = T^{kl}_{ij} = \tfrac{1}{2}(\delta_{ki}\delta_{lj} - \delta_{kj}\delta_{li}) \tag{2.13}$$

and (2.12) reduces to (2.7) or (2.8). The current algebra commutation relations are given by

$$\begin{aligned}[][J^a_\pm(x_\pm), J^b_\pm(x'_\pm)] &= 2if^{abc}J^c_\pm(x_\pm)\delta(x_\pm - x'_\pm) \\ &\quad \mp i\pi k\delta^{ab}\delta'(x_\pm - x'_\pm),\end{aligned} \tag{2.14}$$

and

$$[J^a_\pm, J^m_\pm] = 0,$$

where f^{abc} are the structure constants of the group, i.e.,

$$[T^a, T^b] = if^{abc}T^c. \tag{2.15}$$

The quantity k is a real constant known as central element (charge) and characterizes the Schwiger term or the central extension of the Kac–Moody algebra. It can be written as

$$k = \tfrac{1}{2}k_\lambda, \tag{2.16}$$

where k_λ is the Dynkin index given by

$$\mathrm{Tr}(T^aT^b) = -k_\lambda\delta_{ab}. \tag{2.17}$$

One way to obtain (2.16) is to use the following relation:

$$\begin{aligned}&\partial_-\langle 0|T(J^a_+(x_+)J^b_+(x'_+))|0\rangle \\ &\quad = (1/\sqrt{2})\delta(x_0 - x'_0)\langle 0|[J^a_+(x_+), J^b_+(x'_+)]|0\rangle,\end{aligned} \tag{2.18}$$

where T denotes time ordering. From (2.14) it is clear that the vacuum expectation value on the right-hand side of this equation is just the Schwinger term. So the computation of k reduces to the evaluation of the left-hand side of (2.18). This can be done by noting that

$$\langle 0|T(\Psi_+(x_+)\Psi_+(x'_+))|0\rangle = (x_+ - x'_+)^{-1}, \tag{2.19}$$

and then using the definition of the currents to get

$$\begin{aligned}&\langle 0|T(J^a_+(x_+)J^b_+(x'_+))|0\rangle \\ &\quad = -\tfrac{1}{4}(x_+ - x'_+)^{-2}\,\mathrm{Tr}(T^aT^b).\end{aligned} \tag{2.20}$$

Moreover, since

$$\partial_-(1/x_+) = 2i\pi\delta^2(x), \tag{2.21}$$

we find

$$\langle 0|[J^a_+(x_+), J^b_+(x'_+)]|0\rangle = -i\pi k\delta^{ab}\delta'(x_\pm - x'_\pm) \tag{2.22}$$

as anticipated. The value of k can be computed from (2.17):

$$k = \begin{cases} Q, & \text{for SO}(N)\text{ currents}, \\ N, & \text{for SO}(Q)\text{ currents}. \end{cases} \tag{2.23}$$

Consider next the conformal symmetry of the action (2.1). Under the transformations

$$x_+ \to u(x_+), \quad x_- \to v(x_-),$$

the fields $\Psi_\pm(x_\pm)$ transform as follows:

$$\begin{aligned}\Psi_+(x_+) &\to \left(\frac{du}{dx_+}\right)^{-1/2}\Psi_+(x_+) \\ \Psi_-(x_-) &\to \left(\frac{dv}{dx_-}\right)^{-1/2}\Psi_-(x_-).\end{aligned} \tag{2.24}$$

These transformations leave the action (2.1) invariant. The invariance under translations lead to the local conservation of the energy-momentum tensor. As in any scale invariant theory, the energy-momentum tensor, $\Theta_{\mu\nu}$ is traceless. We take its two independent components to be

$$\Theta_{\pm\pm} = (1/2\sqrt{2})[\Theta_{00} \pm 2\Theta_{01} + \Theta_{11}]. \tag{2.25}$$

For the action (1), these are given by

$$\Theta_{++} = \frac{i}{4} \sum_{i=1}^{N} \sum_{\alpha=1}^{Q} \Psi_+^{i\alpha} \frac{d}{dx_+} \Psi_+^{i\alpha},$$
$$\Theta_{--} = \frac{i}{4} \sum_{i=1}^{N} \sum_{\alpha=1}^{Q} \Psi_-^{i\alpha} \frac{d}{dx_-} \Psi_-^{i\alpha}. \tag{2.26}$$

Clearly,

$$\partial_- \Theta_{++} = \partial_+ \Theta_{--} = 0. \tag{2.27}$$

Each of the components Θ_{++} and Θ_{--} satisfies its own Kac–Moody algebra. The central charge of the algebra can be computed in the same way as that for the current algebra outlined above. It is straightforward to show that

$$\langle 0| [\Theta_{++}(x,t), \Theta_{++}(x',t)] |0\rangle = -\frac{i\pi NQ}{6} \delta'''(x-x'), \tag{2.28}$$

where

$$\delta'''(x-x') = \frac{d^3}{dx^3} \delta(x-x').$$

We can obtain the Virasoro generators by Fourier transforming $\Theta_{++}(x,t)$

$$L_m = \frac{1}{2\pi} \int_0^{\pi} dx\, e^{2imx} \Theta_{++}(x,t). \tag{2.29}$$

These operators satisfy the algebra

$$[L_m, L_n] = (m-n) L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{n+m,0}. \tag{2.30}$$

From (2.28), it follows that

$$c = \tfrac{1}{2} NQ. \tag{2.31}$$

In two-dimensional field theories, it is often convenient to work with complex coordinates[10]

$$z = x^0 + ix^1, \quad \bar{z} = x^0 - ix^1. \tag{2.32}$$

In terms of these coordinates, the line element can be written as

$$ds^2 = dz\, d\bar{z}. \tag{2.33}$$

Under conformal transformations

$$z \to f(z), \quad z \to \bar{f}(\bar{z}),$$

where $f(\bar{f})$ is an arbitrary (anti)holomorphic function, and the pairs $(z,\bar{z})$ are regarded as two independent coordinates in the complex space C^2. Then, the line element transforms as

$$ds'^2 = dz'\, d\bar{z}' = \frac{dz'}{dz} \frac{d\bar{z}'}{d\bar{z}} ds^2.$$

Under these transformations, a primary field $\Psi(z,\bar{z})$ transforms as

$$\Psi'(z',\bar{z}') = \left(\frac{dz'}{dz}\right)^{-k'} \left(\frac{d\bar{z}'}{d\bar{z}}\right)^{-k''} \Psi(z,\bar{z}). \tag{2.34}$$

In particular, for a holomorphic fermion field we have

$$\Psi'(z') = \left(\frac{dz'}{dz}\right)^{-1/2} \Psi(z). \tag{2.35}$$

One advantage of the complex formulation is that the infinitesimal conformal transformations can be written in the form

$$z' = z + \epsilon(z), \tag{2.36}$$

where $\epsilon(z)$ is an infinitesimal holomorphic function that has the Laurent series expansion

$$\epsilon(z) = \sum_{n=-\infty}^{\infty} \epsilon_n z^{n+1}. \tag{2.37}$$

The mode expansion in Laurent series also holds for observables of the theory. As an example, let us consider the Sugawara–Sommerfield form of the energy-momentum tensor.[15] As we have noted above, since currents satisfy standard bosonic field equations, it is reasonable to expect that the energy-momentum tensor can be constructed directly in terms of the currents. For simple groups, a properlv regularized expression for this quantity is given by[16]

$$\mathscr{L}(z) = \frac{1}{2\beta_\lambda} \sum_{A=1}^{d_G} {}^{\times}_{\times} J^A(z) J^A(z) {}^{\times}_{\times}, \tag{2.38}$$

where

$$d_G = \text{dimension of the group} \tag{2.39}$$

and

$$\beta_\lambda = \tfrac{1}{2}(k_\lambda + C_\psi). \tag{2.40}$$

Here, C_ψ is the eigenvalue of the quadratic Casimir operator of G in the adjoint representation, and k_λ is the Dynkin index of the representation λ of G given by (2.17). The double crosses in (2.38) denote normal ordering with respect to the positive and negative frequencies not of fermion fields but of current operators. Let us consider the expansion of $J^A(z)$ in Laurent series

$$J^A(z) = \sum_{-\infty}^{\infty} J_n^A z^{-(n+1)}, \tag{2.41}$$

where the coefficients J_m^A satisfy the commutation relations

$$[J_m^A, J_n^B] = \delta^{AB} \delta_{m+n,o}. \tag{2.42}$$

Then the normal ordering of the currents can be stated as follows:

$${}^{\times}_{\times} J_m^A J_n^A {}^{\times}_{\times} = \begin{cases} J_m^A J_n^A, & m<0, \\ J_n^A J_m^A, & m \geqslant 0. \end{cases} \tag{2.43}$$

The expression for the energy-momentum tensor given by (2.38) is in general different from the canonical one given by (2.26) or its complex version $\Theta(z)$. For one thing, one is quartic in the fermion field operators, and the other is quadratic. But it turns out that for certain representations of some groups, the two expressions are equivalent up to multiplicative constants. This is the case, e.g., when the fermions transform as the fundamental representation of the group $SO(N)$. Using Wick's theorem it has been shown that[11]

$${}^{\times}_{\times} J^A(z) J^A(z) {}^{\times}_{\times} = :J^A(z) J^A(z): + 2C_\lambda \Theta(z), \tag{2.44}$$

where on the right-hand side the normal ordering is with respect to the fermion field operators, and where

$$\Theta(z) = \frac{1}{4} z{:}\left[\frac{d\Psi^i}{dz}, \Psi^i\right]{:}. \tag{2.45}$$

With the fermions $\Psi^i(z)$ in fundamental representation of $SO(N)$,

$$:J^{ij} J^{ij}: = 0. \tag{2.46}$$

Therefore the two realizations of the energy-momentum tensor become proportional.

The above results can be readily generalized to the semisimple groups $G = G_1 \times G_2$, where $G_1 = \mathrm{SO}(N)$ and $G_2 = \mathrm{SO}(Q)$. With fermions in the fundamental representations of both $\mathrm{SO}(N)$ and $\mathrm{SO}(Q)$, the currents have the forms given by (2.7) and (2.8):

$$\begin{aligned} J^{ij}(z) &= (i/2) : \Psi^{i\alpha}(z)\Psi^{j\alpha}(z): , \\ J^{\alpha\beta}(z) &= (i/2) : \Psi^{i\alpha}(z)\Psi^{i\beta}(z): . \end{aligned} \tag{2.47}$$

In this case, the Sugawara–Sommerfield form of the energy-momentum tensor is given by

$$\mathscr{L}(z) = \frac{1}{2\beta'_{\lambda_N}} \overset{\times}{\times} J^{ij}(z) J^{ij}(z) \overset{\times}{\times} + \frac{1}{2\beta'_{\lambda_Q}} \overset{\times}{\times} J^{\alpha\beta}(z) J^{\alpha\beta}(z) \overset{\times}{\times} . \tag{2.48}$$

It can be seen that the expression for $\mathscr{L}(z)$ can be broken up into two parts, one coming from G_1 and one from G_2. However, the coefficients $\beta'_{\lambda N}$ and $\beta'_{\lambda Q}$ are not given by the same expressions as when the symmetry group is, respectively, G_1 alone or G_2 alone. In fact, instead of (2.40) we get

$$\beta'_{\lambda_1} = \tfrac{1}{2}(d_{\lambda_2} k_{\lambda_1} + C_{\psi_1}), \tag{2.49}$$

where d_{λ_2} is the dimension of the fermion representation, λ_2, of the group G_2. A similar expression holds for β'_{λ_2}. For $G_1 = \mathrm{SO}(N)$ and $G_2 = \mathrm{SO}(Q)$, we have, with fermions in the fundamental representation of both groups

$$\begin{aligned} C_{\psi_1} &= 2(N-1), \quad k_{\lambda_1} = 2, \quad d_{\lambda_1} = N, \\ C_{\psi_2} &= 2(Q-1), \quad k_{\lambda_1} = 2, \quad d_{\lambda_1} = Q. \end{aligned} \tag{2.50}$$

It follows that

$$\beta'_{\lambda N} = \beta'_{\lambda Q} = N + Q - 2. \tag{2.51}$$

We therefore have

$$\mathscr{L}(z) = [2(N+Q-2)]^{-1} [\overset{\times}{\times} J^{ij}(z) J^{ij}(z) \overset{\times}{\times} + \overset{\times}{\times} J^{\alpha\beta}(z) J^{\alpha\beta}(z) \overset{\times}{\times}]. \tag{2.52}$$

Moreover, as in the case of simple groups, one can use Wick's theorem to show that

$$\overset{\times}{\times} J^{ij}(z) J^{ij}(z) \overset{\times}{\times} = : J^{ij}(z) J^{ij}(z): + 2C_{\lambda N} \Theta(z), \tag{2.53}$$

$$\overset{\times}{\times} J^{\alpha\beta}(z) J^{\alpha\beta}(z) \overset{\times}{\times} = : J^{\alpha\beta}(z) J^{\alpha\beta}(z): + 2C_{\lambda Q} \Theta(z). \tag{2.54}$$

In these expressions C_{λ_N} (C_{λ_Q}) is the value of the Casimir operator $\mathrm{SO}(N)$ [$\mathrm{SO}(Q)$] in the λ representation. For $G = \mathrm{SO}(N) \times \mathrm{SO}(Q)$, the quantities $:J^{ij} J^{ij}:$ and $:J^{\alpha\beta} J^{\alpha B}:$ no longer vanish. Instead,

$$:J^{\alpha\beta} J^{\alpha\beta}: = - :J^{ij} J^{ij}: . \tag{2.55}$$

As a result,

$$\mathscr{L}(z) = [2(N+Q-2)]^{-1} [2(C_{\lambda_N} + C_{\lambda_Q}) \Theta(z)], \tag{2.56}$$

where

$$C_{\lambda_N} = N - 1, \quad C_{\lambda_Q} = Q - 1; \tag{2.57}$$

hence

$$C_{\lambda_N} + C_{\lambda_Q} = N + Q - 2 \tag{2.58}$$

and

$$\Theta(z) = \mathscr{L}(z). \tag{2.59}$$

We shall make use of this result in establishing the structure of parafermion field theories.

III. GENERAL FEATURES OF PARAFERMIONS

A. The trilinear relations

To discuss parafermion theories in two dimensions, it is convenient to describe some of the general features of parafield theories in a model-independent manner. Parafield theories differ from standard field theories in that the dynamical variables satisfy not bilinear but trilinear relations.[1] To see how this comes about, let us recall that for a single Bose or Fermi field in momentum space, we have the commutator

$$[N_k, a_j^+]_- = \delta_{kj} a_j^+, \tag{3.1}$$

where, modulo a constant,

$$N_k = \tfrac{1}{2}[a_k^+, a_k]_\pm . \tag{3.2}$$

Here, the plus (minus) sign corresponds to bosons (fermions). By substituting (3.2) into (3.1) and allowing all three indices to vary independently, one obtains Green's trilinear relations[11]

$$[[a_k^+, a_l]_\pm, a_m]_- = -2\delta_{km} a_l, \tag{3.3}$$

$$[[a_k, a_l]_\pm, a_m]_- = 0. \tag{3.4}$$

Standard Bose and Fermi quantization schemes can be obtained from these relations if it is further assumed that the commutators (anticommutators) of any canonical operators are C numbers. Otherwise, we are dealing with paraquantization, and the corresponding operators are known as parabosons and parafermions, respectively.

Of the many inequivalent unitary irreducible representations of the paracommutation relations (3.3) and (3.4), the ones relevant to our work are the Fock-type irreducible representations[1,2] in which there exists a unique vacuum state annihilated by all a_k,

$$a_k |0\rangle = 0. \tag{3.5}$$

The Hilbert space A associated with a parafermion theory is fairly restricted, and the observables of the theory are determined by the condition of locality. Although it is, in principle, possible to study various features of these theories within the Hilbert space A, it is often convenient to put the Hilbert space A in correspondence with a larger Hilbert space B in which the operators satisfy bilinear relations. This can be done by means of the Green ansatz

$$a_k = \sum_{\alpha=1}^{Q} a_k^{(\alpha)}, \tag{3.6}$$

where Q is the order of the parafield, α is the Green index, and the $a_k^{(\alpha)}$ are the Green components. They satisfy bilinear but anomalous commutation relations

$$\begin{aligned} &[a_n^\alpha, a_m^{\alpha+}]_{-\epsilon} = \delta_{nm}, \\ &[a_n^\alpha, a_m^\alpha]_{-\epsilon} = 0, \\ &[a_n^\alpha, a_m^\beta]_\epsilon = [a_n^\alpha, a_m^{\beta+}]_\epsilon = 0, \quad \alpha \neq \beta \end{aligned} \tag{3.7}$$

where

$$\epsilon = \begin{cases} +1, & \text{for parabosons,} \\ -1, & \text{for parafermions.} \end{cases}$$

All the Fock-type irreducible representations of Eqs. (3.3) and (3.4) are given by the Green's ansatz.

One can remove this anomaly by means of a Klein transformation. Then the Green's components $a_k^{(\alpha)}$ will transform as a representation of the group SO(Q). The main disadvantage of Klein transformations is that they are nonlinear and nonlocal. Although it is possible to show[7] that for a large class of parafield theories the locality requirement makes the observables independent of these nonlocal operators, some ambiguities remain. For one thing, normal ordering in space B does not necessarily imply normal ordering in space A, and vice versa. For these reasons, it turns out to be more convenient to write the Green's ansatz in the form[9]

$$\psi(x) = \sum_{\alpha=1}^{Q} e^\alpha \psi^\alpha(x) \tag{3.8}$$

where ψ^α, $\alpha = 1,...,Q$, are ordinary Bose (Fermi) fields if $\psi(x)$ is a paraboson (parafermion). The quantities e^α are elements of the real (or complex, see below) Clifford algebra

$$\{e^\alpha, e^\beta\} = 2\delta^{\alpha\beta}, \quad \alpha,\beta = 1,...,Q. \tag{3.9}$$

Moreover, $[e^\alpha, \psi^\beta] = 0$, $\alpha,\beta = 1,...,Q$. With the expression (3.8) for the parafields, the trilinear relation (3.3) is altered slightly, but (3.4) remains intact, so we have at equal times

$$[[a^+(\mathbf{p}),a(\mathbf{q})]_\pm - \langle [a^+(\mathbf{p})]_\pm \rangle_0,$$
$$a(\mathbf{k})]_- = -2\delta^3(\mathbf{p}-\mathbf{k})a(\mathbf{q}), \tag{3.10}$$
$$[[a(\mathbf{p}),a(\mathbf{q})]_\pm, a(\mathbf{k})]_- = 0, \tag{3.11}$$

where the symbol $\langle\ \rangle_0$ stands for the vacuum expectation value.

The ansatz (3.8) relates the parafield operators and states to a set of field operators and states in the Fermi–Bose Hilbert state B. The linearity of this relation, and the fact the $\psi^\alpha(x)$ are standard fermions or boson fields with an SO(Q) internal symmetry, facilitates the comparison of the parafield theory with corresponding theories satisfying standard statistics. The linearity also allows us to regard the ansatz (3.8) as an expansion of $\psi(x)$ in the basis $\{e^\alpha\}$. The expansion coefficients can then be obtained in the usual way. For example, for parafermions

$$\psi^\alpha(x) = \tfrac{1}{2}\{e^\alpha, \psi(x)\}.$$

It is important to note that a parafield theory is formulated in the Hilbert space A. One can put the states and the observables of this space into correspondence with a *subset* of the states and observables of the Hilbert space B which has a larger number of states and a larger class of observables. But not every one of the operators and states in space B can be represented in space A. In particular, observables in space B that carry uncontracted SO(Q) indices have no equivalents in the original space A. Since the parafield theory is defined in the Hilbert space A, the Hilbert space B is useful, but not necessary, to the extent that it facilitates making deductions about the space A.

B. Structure of observables

In any field theory, an observable can be expressed as a functional of the field operators. Let $F(V)$ and $F'(V')$ be any two such observables defined in regions V and V', respectively. When V and V' are spacelike separated, then the principle of locality (microscopic causality) requires that $F(V)$ and $F'(V')$ commute with each other:

$$[F(V),F'(V')] = 0, \quad V \text{ and } V' \text{ spacelike separated.} \tag{3.12}$$

Thus the observables behave like bosonic field operators. For bosonic field theories, the field operators themselves satisfy this condition. This means that the observables can be any functional of the bosonic field operators, and locality imposes no additional restriction on them. On the other hand, fermion field operators anticommute at spacelike separations, so that the observables must be functionals of even products of fermion field operators. For observables involving parafield operators, the situation is more complicated[7] because they do not simply commute or anticommute at spacelike separations. A general method of constructing the observables of a parafield theory is discussed in Ref. 7. Then, using Klein transformations, they are compared with the corresponding quantities in a fermionic and/or bosonic theory with appropriate internal symmetry.

As pointed out above, our objective is to elucidate the relation between the states and observables of parafermion theories of order Q with the corresponding fermion theories using a canonical approach. In establishing the relation between any two such theories, it is essential that we compare their *regularized* observables. Since Klein transformations are nonlinear and nonlocal, they make the connection between the regularization schemes in the two Hilbert spaces at best ambiguous. To overcome this difficulty, we present below an alternative method of constructing the observables and states of the Hilbert space A using the ansatz (3.8), which does not involve Klein transformations.[13] We illustrate the method for parafermions, but it will be clear from the context that it is equally applicable to parabosons.

Consider a set of parafermion fields $\{\Psi_i\}$ of order Q, where the subscript i represents a space-time label, an internal symmetry label, or both. From these, we construct the normal ordered antisymmetric product

$$:[\Psi_i,\Psi_j]: \;=\; :(\Psi_i\Psi_j - \Psi_j\Psi_i):. \tag{3.13}$$

Using the ansatz (3.8), this can be written as

$$\begin{aligned} :[\Psi_i,\Psi_j]: &= \sum_{\alpha=1}^{Q}\sum_{\beta=1}^{Q} :(e^\alpha\Psi_i^\alpha e^\beta\Psi_j^\beta - e^\beta\Psi_j^\beta e^\alpha\Psi_i^\alpha): \\ &= \sum_{\alpha=1}^{Q}\sum_{\beta=1}^{Q} :(e^\alpha e^\beta\Psi_i^\alpha\Psi_j^\beta - e^\beta e^\alpha\Psi_j^\beta\Psi_i^\alpha):, \end{aligned}$$

where the last step follows since $[e^\alpha, \psi^\beta] = 0$, for any α and β. Now we note that e^α's can be taken out of the normal ordering sign because they are not quantum fields, i.e.,

$$\begin{aligned} :[\Psi_i,\Psi_j]: &= \sum_{\alpha,\beta=1}^{Q} \{e^\alpha,e^\beta\} : \Psi_i^\alpha\Psi_j^\beta: \\ &= 2\sum_{\alpha=1}^{Q} :\Psi_i^\alpha\Psi_j^\alpha:. \end{aligned} \tag{3.14}$$

It is clear from the right-hand side that the left-hand side is a bosonic operator. In this way, normal ordered bilinear observables in the Hilbert space A become related to the bilinear SO(Q) invariant observables in the Hilbert space B.[13]

Another set of observables can be constructed from a symmetric product of Q parafermion field operators. To study their structure, it is convenient to define an analog of the Dirac γ_5 matrix for the algebra (3.9). It is given by

$$\Gamma_Q = (1/Q!)\epsilon^{\alpha_1\alpha_2\cdots\alpha_Q}e^{\alpha_1}e^{\alpha_2}\cdots e^{\alpha_Q}, \tag{3.15}$$

where ϵ is the totally antisymmetric tensor. Then using the ansatz (3.8), we can express a normal ordered symmetric product of parafermions as

$$:[\Psi_1\Psi_2\cdots\Psi_c]_+: \;= c! \sum_{\alpha_1<\alpha_2\cdots<\alpha_c}^{Q} e^{\alpha_1}\cdots e^{\alpha_c} \times\left[\sum_{\sigma\in S_c}(-1)^{\rho(\sigma)} : \Psi_1^{\alpha_{\sigma_1}}\cdots\Psi_c^{\alpha_{\sigma(c)}}:\right], \tag{3.16}$$

where $1\leqslant c\leqslant Q$, σ is a permutation of $1,2,...,c$, and

$$\rho(\sigma) = \begin{cases} 0, & \text{if } \sigma \text{ is even,} \\ 1, & \text{if } \sigma \text{ is odd.} \end{cases}$$

In particular, when $c = Q$, the symmetric product can be written as

$$:[\Psi_1,\Psi_2\cdots\Psi_c]_+: \;= \Gamma_Q\epsilon^{\alpha_1\cdots\alpha_Q} : \Psi_1^{\alpha_1}\cdots\Psi_Q^{\alpha_Q}:. \tag{3.17}$$

From the right-hand side of this expression, it is clear that this symmetric product is a bosonic operator when Q is even and a fermionic operator when Q is odd. Since observables must be bosonic operators, we conclude that for even-order parafermion theories, there are two classes of observables, or functionals thereof, given by (3.14) and (3.17), respectively. For odd-order theories, there is only one class given by (3.14). Note that the expression (3.17) is invertible. Multiplying by Γ_Q and taking the trace, we find

$$\epsilon^{\alpha_1\cdots\alpha_Q} : \Psi_1^{\alpha_1}\cdots\Psi_Q^{\alpha_Q}: \;= (1/Q)\mathrm{Tr}(\Gamma_Q :[\Psi_1\cdots\Psi_Q]_+:). \tag{3.18}$$

It follows that there is a unique correspondence between the observables in the Hilbert space A and a subset of observables in Hilbert space B. The fact that this subset does not exhaust all the observables in the Hilbert space B can be seen by noting that the SO(Q) currents (2.8) are perfectly legitimate operators in B but have no equivalents in A.

Finally, we wish to emphasize that in the above discussion the only restriction on the structure of observables has come from the requirement of locality. Further restrictions often arise from the invariances of a given theory or its renormalizability. For example, chiral invariance and conformal invariance rule out the symmetric product (3.17) as a term in the action for most parafermion theories.

C. Classification of states

The states in the Hilbert space A have a more complicated structure than those in the Hilbert space B. Since the parafield operators do not simply commute or anticommute, the structure of a given n-particle state depends on the order in which the n creation operators are applied to the vacuum. Moreover, it turns out that the n-particle states that can be constructed in this way are not all linearly independent. It thus becomes necessary to develop a systematic method of constructing a linearly independent set of basis vectors in terms of which every state of the Hilbert space A can be described. The solution to this problem is the content of a general decomposition theorem. For parafermions, it states that any n-particle state, $|n\rangle$, can be expressed as a superposition of the basis states of the form

$$|n,c\rangle = [\Psi_{i_1},\Psi_{j_1}]\cdots[\Psi_{i_b},\Psi_{j_b}][\Psi_{k_1}\cdots\Psi_{k_c}]_+|0\rangle, \tag{3.19}$$

where

$$c\leqslant Q \quad \text{and} \quad 2b+c=n.$$

For a given n, states with different values of c are orthogonal to each other.[7] Thus the Hilbert space A breaks up into orthogonal subspaces, $A^{(n)}$, of different n, which, in turn, break up into orthogonal subspaces, $A^{(n,c)}$, of different c:

$$A = \sum_{n=1}^{\infty} A^{(n)} = \sum_{n=1}^{Q}\sum_{c=1}^{Q} A^{(n,c)}. \tag{3.20}$$

The states in $A^{(n,c)}$ organize themselves into irreducible representations of the symmetric group S_n and can be classified by their Young diagrams. Each Young diagram consists of n squares with c odd columns and no more than Q squares in the first row. In general, there can be several inequivalent irreducible representations of S_n corresponding to the same tableau. But it can be shown[7] that for states constructed from parafermions one and only irreducible representation of S_n corresponds to each tableau. To see how this comes about, let us consider a simple example. Let $Q\geqslant 3$ and consider three particle states $a_i^+ a_j^+ a_k^+|0\rangle$, with $i,j,k = 1, 2, 3$. Since the creation operators do not simply commute or anticommute, there may be as many as 3! states which can be classified into the irreducible representations of the symmetric group S_3. The four irreducible representations of this group are described according to the Young diagrams: (i) one symmetric one-dimensional representation,

$$[a_1^+ a_2^+ a_3^+]_+|0\rangle; \tag{3.21}$$

(ii) two two-dimensional representation of mixed symmetry,

$$\begin{cases} [a_3^+,[a_1^+,a_2^+]]|0\rangle, \\ 1/\sqrt{3}([a_1^+,[a_2^+,a_3^+]] - [a_2^+,[a_3^+,a_1^+]])|0\rangle, \end{cases} \tag{3.22}$$

and

$$\begin{cases} [a_3^+,[a_1^+,a_2^+]_+]|0\rangle, \\ -1/\sqrt{3}([a_1^+,[a_2^+,a_3^+]_+] - [a_2^+,[a_3^+,a_1^+]_+])|0\rangle; \end{cases} \tag{3.23}$$

(iii) one antisymmetric one-dimensional representation,

$$[a_1^+,a_2^+,a_3^+]_-|0\rangle. \tag{3.24}$$

For the parafermions of order $Q\geqslant 3$ we are considering, not all of these possibilities can be realized. In fact, it follows from the trilinear relation $[a_1^+,[a_2^+,a_3^+]] = 0$ that the states for the first two-dimensional representation vanish identically. Thus each irreducible representation of S_3 is realized only once for parafermions. As pointed out above, this is a general feature of the states in the parafermionic Hilbert space A. One of the remaining four nontrivial states given by (3.21) already conforms to the structure of the basis states given by Eq. (3.19). The other three given by (3.23) and (3.24) can be cast into that form by using the trilinear relations (3.11). For example,

$$[a_3^+,[a_1^+,a_2^+]_+] = 2[a_1^+,a_3^+,]a_2^+ - [a_2^+,a_3^+]a_1^+ \tag{3.25}$$

and

$$\begin{aligned}&[a_1^+,a_2^+,a_3^+]_- \\ &\quad = [a_1^+,a_2^+]a_3^+ + [a_2^+,a_3^+]a_1^+ + [a_3^+,a_1^+]a_2^+ .\end{aligned} \tag{3.26}$$

So, there are altogether three $c=1$ states and one $c=3$ state. This is as expected, since c corresponds to the number of odd columns of the Young diagram. For a given n, since there is more than one state for a given c, it is clear that the labels n and c do not completely specify a state in the Hilbert space A. In general, the total number of linearly independent state vectors of the form (3.19) with a given n and c is equal to the sum of the dimensions of all the realizable irreducible representations of S_n for which the corresponding Young diagrams have c odd columns and no more than Q squares in the first row. Therefore, a system of n parafermions possesses additional degrees of freedom and requires additional quantum numbers to completely specify a state. The labels of the symmetric group S_n supply these additional quantum numbers. It will be shown below that in the Hilbert space B, the same quantum numbers can be specified by the Casimir operators of the group SO(Q).

D. The relation to states in space B

We now discuss the precise connection between the states of Hilbert space A and those of Hilbert space B. To avoid notational complications, we will suppress the SO(N) symmetry which is the same in both Hilbert spaces. Thus, given a parafermion theory of order Q, we want to describe how the states in such a theory are related to the states of a fermion theory with SO(Q) symmetry and with the same structure. The connection between the two theories can be worked out by means of the ansatz (3.8). Using this ansatz, we can write the basis states (3.19) in the form

$$\begin{aligned}|n,c\rangle &= c!\Big(\sum_{\alpha=1}^{Q} [a_{i_1}^{\alpha+},a_{j_1}^{\alpha+}]\Big)\cdots\Big(\sum_{\beta=1}^{Q} [a_{i_b}^{\beta+},a_{j_b}^{\beta+}]\Big) \\ &\quad\times \sum_{\alpha_1<\alpha_2\cdots<\alpha_c}^{Q} e^{\alpha_1}\cdots e^{\alpha_c} \\ &\quad\times \sum_{\sigma\in S_c} (-1)^{p(\sigma)} a_1^{\alpha_{\sigma(1)}+}\cdots a_c^{\alpha_{\sigma(c)}+}|0\rangle,\end{aligned} \tag{3.27}$$

where we have used the notation of Eq. (3.16). If we now identify the in indices, α,β, etc., with the labels of the irreducible representations of SO(Q), it follows that the quantity

$$\sum_{\alpha=1}^{Q} [a_i^{\alpha^+},a_j^{\alpha^+}]$$

is SO(Q) singlet, and the product

$$\sum_{\sigma\in S_c} (-1)^{p(\sigma)} a_1^{\alpha_{\sigma(1)}+}\cdots a_c^{\alpha_{\sigma(c)}+}$$

transforms, as a rank c antisymmetric tensor under SO(Q). Therefore the right-hand side of equation (3.27) corresponds to a unique state in the Hilbert space B. This correspondence is quite general, and states that *any* state in the Hilbert space A is uniquely related to a state in the Hilbert space B. It is important to note, however, that states in the Hilbert space A do not exhaust those of the Hilbert space B. This can be seen from the right-hand side of Eq. (3.27). In that expression, if we replace $[a^{\alpha}{}_i{}^+,a^{\alpha}{}_j{}^+]$ by $[a^{\alpha}{}_i{}^+,a^{\beta}{}_j{}^+]$, with $\alpha\neq\beta$, the corresponding state is a perfectly legitimate state in the Hilbert space B but has no analog in the space A. In fact, from every irreducible representation of SO(Q) in the space B, there is one and only one state that has an equivalent in the space A. The subspace of the Hilbert space B, in which the states are in one-to-one correspondence with those of the Hilbert space A, will be called the Hilbert space H.

Starting from the irreducible representations of SO(Q) in the Hilbert space B, it is easy to see that one can reverse the steps indicated above and construct the states of the Hilbert space A. Here, we illustrate this for the three parafermion states described above. Assigning the fermions to the fundamental representation of SO(Q), it is clear that we can construct the states of three distinct irreducible Q-dimensional representations of SO(Q) as follows:

$$\begin{aligned}&\Big(\sum_{\alpha=1}^{Q} a_1^{\alpha+}a_2^{\alpha+}\Big)a_3^{\beta+}|0\rangle,\quad \beta=1,2,\dots,Q,\\ &\Big(\sum_{\alpha=1}^{Q} a_1^{\alpha+}a_3^{\alpha+}\Big)a_2^{\beta+}|0\rangle,\\ &\Big(\sum_{\alpha=1}^{Q} a_2^{\alpha+}a_3^{\alpha+}\Big)a_1^{\beta+}|0\rangle,\end{aligned} \tag{3.28}$$

Using the ansatz (3.8), we can uniquely select a single state from each of these irreducible representations. The resulting states are, in obvious notation

$$\begin{aligned}&[a_1^+,a_2^+]a_3^+|0\rangle,\\ &[a_1^+,a_3^+]a_2^+|0\rangle,\\ &[a_2^+,a_3^+]a_1^+|0\rangle.\end{aligned} \tag{3.29}$$

Comparing these to (3.25) and (3.26), it is easy to see that their linear combinations form the states of the representations (3.23) and (3.24) of S_3. Similarly, the states of the one-dimensional symmetric representation can be constructed from the rank-3 antisymmetric representation

$$\sum_{\sigma\in S_c} (-1)^{p(\sigma)} a_1^{\alpha_{\sigma(1)}+}\cdots a_3^{\alpha_{\sigma(3)}+}|0\rangle$$

of SO(Q).

IV. APPLICATION TO PARAFERMION THEORIES IN TWO DIMENSIONS

Parafermion theories in two dimensions provide excellent examples of the general formalism developed in the previous section. We consider a free two-dimensional Majorana parafermion theory of order Q with chiral $O(N)$ symmetry.[13] The action for the theory in the Hilbert space A is given by

$$S_2 = \frac{i}{2} \int d^2x [\overline{\Psi}^i \gamma^a, \partial_a \Psi^i], \tag{4.1}$$

where the γ^a's are given by (2.2). The antisymmetrization is necessary because S_2 is an observable, and, as discussed in the previous section, the structure of observables are restricted by the requirement of locality. Using (2.3) and (2.4) we write S_2 in the light cone coordinates:

$$S_2 = \frac{i}{4} \int dx_+ \, dx_- ([\Psi^i_-, \partial_+ \Psi^i_-] + [\Psi^i_+, \partial_- \Psi^i_+]). \tag{4.2}$$

Following the procedure outlined in Sec. III, we now use the ansatz (3.8) and transform this action to one in the Hilbert space H. We get

$$S_2 = \frac{i}{2} \int dx_+ \, dx_- [\Psi^{i\alpha}_- \, \partial_+ \Psi^{i\alpha}_- + \Psi^{i\alpha}_+ \, \partial_- \Psi^{i\alpha}_+]. \tag{4.3}$$

This action is indistinguishable from the action S_1 of the fermion theory with $O(N) \times O(Q)$ symmetry given by (2.5), which is defined over the entire Hilbert space B. This is, of course, consistent with the relation between the observables of the Hilbert space A and those of space B established in Sec. III. We emphasize again that the equality of the action S_1 with S_2 does not mean that the corresponding theories are identical. This is because the theory based on S_1 has other observables that are not $SO(Q)$ invariant and that have no equivalents in the theory based on S_2.

O(N) currents: From the class of observables in the paraspace A, which have counterparts in space B, consider next the $O(N)$ currents $J^{ij}_\pm$ given in terms of the parafields by

$$\begin{aligned} J^{ij}_\pm &= (i/4) : (\Psi^i_\pm \Psi^j_\pm - \Psi^j_\pm \Psi^i_\pm) : \\ &= (i/4) : [\Psi^i_\pm, \Psi^j_\pm] :, \end{aligned} \tag{4.4}$$

where the normal ordering is defined with respect to the parafields $\psi^i_\pm$. From (3.14), the corresponding currents in the Hilbert space B are

$$J^{ij}_\pm = (i/2) : \Psi^{i\alpha}_\pm \Psi^{j\alpha}_\pm :, \tag{4.5}$$

where the normal ordering is now with respect to the fermion fields $\psi^{i\alpha}_\pm$. This expression has precisely the same form as the $O(N)$ currents of the $O(N) \cdot \times O(Q)$ fermion theory. Therefore these currents satisfy the same commutation relations as that given by (3.14). In particular, the central charge is given by $k = Q$.

Virasoro algebra: As another example of the observables in the Hilbert space A, we consider the canonical energy momentum tensor of the parafermion theory. The two nonvanishing components are given by

$$\Theta_{\pm\pm} = \frac{i}{8} \sum_{j=1}^{N} : \left[\Psi^j_\pm, \frac{d}{dx_\pm} \Psi^j_\pm \right] :, \tag{4.6}$$

where the sum from 1 to N indicates that, like fermion fields, the parafermion fields are assigned to the fundamental representation of $O(N)$. Using the ansatz (3.8) as before, we get

$$\Theta_{\pm\pm} = \frac{i}{4} \sum_{i=1}^{N} \sum_{\alpha=1}^{Q} : \Psi^{i\alpha}_\pm \frac{d}{dx_\pm} \Psi^{i\alpha}_\pm : . \tag{4.7}$$

Again, not surprisingly, these are the same expressions as the energy momentum tensors of the $O(N) \times O(Q)$ fermion theory given by (2.26). It follows that the Virasoro operators of the parafermion theory satisfy the same algebra with the same central charge $C = NQ/2$.

In obtaining the values of the central charge for the $O(N)$ current algebra and Virasoro algebra of the parafermion theory, we have used the general arguments of Sec. III. Since these arguments rely on the use of the ansatz (3.8), the reader might wonder whether the values of the central charge depend on the choice of this ansatz. To show that they do not, let us obtain the expression for the Virasoro operators using the Green's ansatz (3.6). Thus we have

$$\Psi^i(x) = \sum_{\alpha=1}^{Q} \Psi^{i(\alpha)}(x), \tag{4.8}$$

where $\Psi^{i(\alpha)}$ ($\neq \Psi^{i\alpha}$) satisfy the *anomalous anticommutation relations*

$$\begin{aligned} [\Psi^{i(\alpha)}(x), \Psi^{j(\alpha)}(y)]_+ &= \delta^{ij}\delta(x-y), \\ [\Psi^{i(\alpha)}(x), \Psi^{j(\beta)}(y)]_- &= 0, \quad \alpha \neq \beta. \end{aligned} \tag{4.9}$$

Fourier expanding the fields $\psi^i(x)$ and substituting (4.8) in (4.7), we can write the Virasoro operators of the parafermion theory in the form

$$L_m = \frac{1}{8} \sum_{i=1}^{N} \sum_{k=-\infty}^{\infty} (m+2k) : [a^i_{-k}, a^i_{m+k}]_- :, \tag{4.10}$$

where a^i_k, $k = -\infty, ..., \infty$, satisfy the trilinear relations (3.3) and (3.4). They can be written in terms of the Green's ansatz as in (3.6). Then it is easy to see that we can write

$$L_m = \sum_{\alpha=1}^{Q} L^{(\alpha)}_m, \tag{4.11}$$

where, for fixed α,

$$L^{(\alpha)}_m = \frac{1}{8} \sum_{i=1}^{N} \sum_{k=-\infty}^{\infty} (m+2k) \times : [a^{i(\alpha)}_{-k}, a^{i(\alpha)}_{m+k}]_- : . \tag{4.12}$$

It follows that

$$\begin{aligned} [L^{(\alpha)}_m, L^{(\alpha)}_n] &= (m-n) L^{(\alpha)}_{m+n} + (N/24)(m^3 - m)\delta_{n+m,0}, \\ [L^{(\alpha)}_m, L^{(\beta)}_n] &= 0, \quad \alpha \neq \beta. \end{aligned} \tag{4.13}$$

Therefore we end up with the same value of central charge:

$$[L_m, L_n] = (m-n) L_{m+n} + (NQ/24)(m^3 - m)\delta_{n+m,0}. \tag{4.14}$$

The Sugawara–Sommerfield construction: We have seen that in the $SO(N) \times SO(Q)$ free fermion theory it is possible to construct the Sugawara–Sommerfield form of the energy-momentum tensor, and that this was equal to the canonical expression $\Theta(z)$. Given that this equality was established in the Hilbert space B, and given that its subspace H is in one-to-one correspondence with the para-Hilbert space A, we would expect that the analog of $\mathscr{L}(z)$ should exist for the parafermion theory in space A. In the Hilbert space B, $\mathscr{L}(z)$ received contributions from both the SO(N) currents J^{ij} and the SO(Q) currents $J^{\alpha\beta}$. This is because in this space J^{ij} and $J^{\alpha\beta}$ satisfy the requirement of locality and are observables. But in the space A, or the space H, the $J^{\alpha\beta}$ are not observables, and this might be taken as indication that the Sugawara–Sommerfield construction is not possible in these spaces. We note, however, that it is not $J^{\alpha\beta}$ but its square $J^{\alpha\beta}J^{\alpha\beta}$ that appears in the expression for $\mathscr{L}(z)$. Since $J^{\alpha\beta}J^{\alpha\beta}$ is an observable in the space A, or the subspace H, the construction of $\mathscr{L}(z)$ in these spaces is indeed possible.

The relation between the Sugawara–Sommerfield energy-momentum tensor, $\mathscr{L}^B$, of the Hilbert space B and $\mathscr{L}^A$ of the Hilbert space A can be used to obtain an explicit expression for $J^{\alpha\beta}J^{\alpha\beta}$ in the space A. We have already noted the relation between the canonical energy-momentum tensor $\Theta^A_\pm$ given by (4.6) in the space A, and $\Theta^B_\pm$, given by (2.26) in the space B. By using (2.59) we have also seen that for our theory in space B, $\mathscr{L}^B = \Theta^B$. Then, by solving the expression (2.52) for $J^{\alpha\beta}J^{\alpha\beta}$, we get, in space B (or H)

$$\overset{\times}{\times} J^{\alpha\beta}_B J^{\alpha\beta}_B \overset{\times}{\times} = 2(N+Q-2)\Theta^B - \overset{\times}{\times} J^{ji}_B J^{ji}_B \overset{\times}{\times} . \quad (4.15)$$

Since the right-hand side of this expression is constructed from observables that have counterparts in the Hilbert space A, it follows from the general arguments of Sec. III that $\overset{\times}{\times} J^{\alpha\beta}_\pm J^{\alpha\beta}_\pm \overset{\times}{\times}$ must have the same structure in space A also. Therefore, from (4.4) we get

$$\overset{\times}{\times} J^{\alpha\beta}_\pm J^{\alpha\beta}_\pm \overset{\times}{\times} = (i/4)(N+Q-2) : \left[\Psi^j_\pm , \frac{d}{dx_\pm} \Psi^j_\pm \right] : - \overset{\times}{\times} J^{ij}_\pm J^{ij}_\pm \overset{\times}{\times}, \quad (4.16)$$

where the $J^{ij}_\pm$ satisfy the field equations (2.11). The normal ordering in terms of currents can be expressed in terms of normal ordering of the parafermion fields by means of Wick's theorem. The result is

$$\begin{aligned} \overset{\times}{\times} J^{ij}_\pm J^{ij}_\pm \overset{\times}{\times} &= :J^{ij}_\pm J^{ij}_\pm: + 2C_{\lambda N}\Theta_{\pm\pm} \\ &= \tfrac{1}{16} :(:[\Psi^i_\pm, \Psi^j_\pm]::[\Psi^i_\pm, \Psi^j_\pm]:): \\ &\quad + 2(N-1)\Theta_{\pm\pm} . \end{aligned} \quad (4.17)$$

We thus have, entirely in terms of the parafermion fields,

$$\overset{\times}{\times} J^{\alpha\beta}_\pm J^{\alpha\beta}_\pm \overset{\times}{\times} = i/4(Q-1) : \left[\Psi^i_\pm , \frac{d}{dx_\pm} \Psi^i_\pm \right] : + \tfrac{1}{16} :(:[\Psi^i_\pm, \Psi^j_\pm]::[\Psi^i_\pm, \Psi^j_\pm]:): . \quad (4.18)$$

In quantum field theory, one often encounters situations in which a field operator is well defined but its "square" either does not exist at all or would have to be defined using some prescription. Our parafermion field theories of order Q provide us with prototypes of an operator $(J^{\alpha\beta})^2$ that is well defined, but its "square root" does not exist in the Hilbert space A.

Remarks: Our main objective in this section has been to show what a free parafermion theory of order Q is equivalent to in the context of canonical quantum field theory. We have found that the Hilbert space A of such a theory is in one-to-one correspondence with a subspace H of the Hilbert space B of a free fermion theory with an SO(Q) symmetry. The observables of the parafermion theory are also in one-to-one correspondence with those of the subspace H. These are a subset of the observables of the Hilbert space B, which are SO(Q) invariant and which exist in the subspace H. The remaining observables of the Hilbert space B, which are not SO(Q) invariant, connect the states in the subspace H to the other states of the B. Therefore, these operators are not defined within the subspace H alone. An important example of such observables are the SO(Q) currents $J^{\alpha\beta}$, which exist in B but not in H or A. On the other hand, the SO(Q) invariant operator $\overset{\times}{\times} J^{\alpha\beta}_\pm J^{\alpha\beta}_\pm \overset{\times}{\times}$ exists in B, H, and A. It has the same eigenvalues in any one of these three spaces.

Our results are to be contrasted with the work of Antoniadis and Bachas.[16] These authors construct a constrained (not free) fermion theory by gauging the SO(Q) symmetry and then setting $J^{\alpha\beta} = 0$ as a constraint. Then they argue that the theory so obtained is equivalent to a free parafermion theory of order Q. For $Q = 2$, it has been shown that[17] such an equivalence is regularization dependent and therefore cannot hold. Our general arguments show not only what a parafermion theory is equivalent to but also what happens when the condition $J^{\alpha\beta} = 0$ is imposed. Among other things, it also implies $J^{\alpha\beta}J^{\alpha\beta} = 0$. Since this operator is SO(Q) invariant, it is a function of the Casimir operators of SO(Q), so that its eigenvalues vary from one irreducible presentation to another. So, the $J^{\alpha\beta}J^{\alpha\beta} = 0$ condition on states can at most be consistent with some representations of SO(Q). But the parafermionic Hilbert space, as we have seen, has states related to all representation of SO(Q). It follows that there cannot be an equivalence between such a constrained fermion theory and the corresponding parafermion theory.

Finally, we note that the word "parafermion" has also been used in two-dimensional field theories to denote fermions with fractional spin and anomalous dimensions.[18] The connection (if any) between such theories and standard parafield theories has not been made clear.

ACKNOWLEDGMENTS

This work was supported, in part, by the Department of Energy under Contract No. DOE-FG02-84ER4015.

[1]H. S. Green, Phys. Rev. **90**, 270 (1953).
[2]O. W. Greenberg and A. M. L. Messiah, Phys. Rev. **138**, 1155 (1965).
[3]O. W. Greenberg, Phys. Rev. Lett. **13**, 598 (1964).
[4]F. Ardalan and F. Mansouri, Phys. Rev. D **9**, 3341 (1974); Phys. Rev. Lett. **56**, 2456 (1986); Phys. Lett. B **176**, 99 (1986); F. Mansouri and X. Wu, Mod. Phys. Lett. A **2**, 215 (1987).
[5]O. W. Greenberg and R. N. Mohapatra, Phys. Rev. Lett. **59**, 2507 (1987).
[6]L. B. Okun, preprint, 1987.
[7]Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (Springer, Berlin, 1982).
[8]O. W. Greenberg and C. A. Nelson, Phys. Rep. C **32**, 69 (1977); P. G. O. Freund, Phys. Rev. D **13**, 2322 (1976).
[9]G. Domokos and S. Kovesi-Domokos, J. Math. Phys. **19**, 1477 (1978); Phys. Rev. D **19**, 2984 (1979); O. W. Greenberg and K. I. Macrae, Nucl. Phys. B **219**, 358 (1983).
[10]A. A. Belavin, A. M. Polyakov, and A. B. Zomolodchikov, Nucl. Phys. B **241**, 333 (1984).
[11]P. Goddard and D. Olive, Mod. Phys. A **1**, 303 (1986).

[12] E. Witten, Commun. Math. Phys. **92**, 455 (1984).
[13] F. Mansouri and X. Wu, Phys. Lett. B **203**, 417 (1988).
[14] P. Rumond, Phys. Rev. D **3**, 2415 (1971); A. Neven and J. H. Schwarz, Nucl. Phys. B **31**, 1109 (1971).
[15] H. Sugawara, Phys. Rev. **170**, 1659 (1968); C. M. Sommerfield, *ibid.* **176**, 2019 (1968).
[16] I. Antoniadis and C. Backas, Nucl. Phys. B **278**, 343 (1986).
[17] D. Chang, A. Kumar, and R. Mohapatra, Z. Phys. C **32**, 417 (1986); A. N. Redick and H. J. Schnitzer, Phys. Lett. B **167**, 315 (1985).
[18] A. B. Zomolodchikov and V. A. Fateev, Sov. Phys. JETP **62**, 215 (1985); **63**, 913 (1986); D. Gepner and Z. Qiu, Princeton preprint PUPT-1038, 1986.

Equivalence of higher-order Lagrangians. II. The Cartan form for particle Lagrangians

Lucas Hsu and Niky Kamran
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
Peter J. Olver
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 18 November 1988; accepted for publication 14 December 1988)

It is shown how Cartan's method of equivalence may be used to obtain the Cartan form for an r th-order particle Lagrangian on the line by solving the standard equivalence problem under contact transformations on the jet bundle J^{r+k} for $k \geqslant r-1$.

I. INTRODUCTION

This is the second in a series of papers investigating different aspects of the Cartan equivalence problem for higher-order variational problems. In Pt. I,[1] it was shown how each of the basic Lagrangian equivalence problems, in any number of independent and dependent variables, could be formulated as a Cartan equivalence problem, and a fundamental reduction theorem, demonstrating that the equivalence problem for an r th-order Lagrangian could always be reduced to the minimal-order jet bundle J^r, was proved. In this paper, we will be exclusively concerned with r th-order variational problems in one independent and one dependent variable,

$$\mathscr{L}[u] = \int_\Omega L(x,u^{(r)})dx. \tag{1.1}$$

The Lagrangian L depends analytically on the coordinates $(x,u^{(r)}) = (x,u,u_1,...,u_r)$ on the jet bundle $J^r = J^r(\mathbb{R},\mathbb{R})$. Here, the coordinate u_j represents the jth-order derivative of the single dependent variable u with respect to the single independent variable x, so $u_j = D_x^j\, u$, where D_x denotes the total derivative operator. We will be interested in properties of the functional (1.1) that are preserved under change of variables, which we take to mean general contact transformations. In the language of Pt. I,[1] we are dealing with the standard equivalence problem for the particle Lagrangian (1.1) under the pseudogroup of contact transformations. The *contact ideal* on J^r, denoted $\mathscr{I}^{(r)}$, plays a key role; it is generated by the contact forms

$$\theta_j = du_j - u_{j+1}\, dx, \quad 0 \leqslant j < r. \tag{1.2}$$

According to Bäcklund's theorem,[1,2] a transformation $\Psi: J^r \to J^r$ will preserve the contact ideal $\mathscr{I}^{(r)}$ if and only if it is the prolongation of a contact transformation $\Psi_0: J^1 \to J^1$ of the first-order jet bundle, a fact that will play an important role in our discussion.

An important invariant one-form associated to the functional (1.1) is the so-called *Cartan form*,[3]

$$\Theta_C = L\, dx + \sum_{i=1}^{r} \sum_{j=0}^{r-i} (-D_x)^j \left(\frac{\partial L}{\partial u_{i+j}}\right) \cdot \theta_{i-1}. \tag{1.3}$$

It is well known that the Cartan form encodes both the Euler–Lagrange equations for (1.1), and that it plays an important role in the formulation of Noether's theorem relating symmetries and conservation laws. It also figures prominently in the implementation of field theory via the Hamilton–Jacobi equation, which is used to deduce the existence of strong minimizers.[4,5] Note that Θ_C lives on the jet bundle J^{2r-1}, which reflects the fact that the Euler–Lagrange equations for a nondegenerate r th-order Lagrangian are of order $2r$. We will see that the Cartan form remains invariant under contact transformations of the Lagrangian (1.1), i.e., if one Lagrangian is mapped to another via a contact transformation, then the corresponding Cartan forms are mapped to each other by the appropriate prolongation of the same contact transformation. (We remark that the Cartan form is not invariant under the more general operations of transforming and adding a total divergence to the Lagrangian. This explains why we consider the standard equivalence problem and not the divergence equivalence problem in this paper.)

A powerful construction that produces the invariants (functions and differential forms) associated with such a variational problem is Cartan's *Method of Equivalence*,[6–8] a general method for determining if two exterior differential systems generated by one-forms are equivalent under a change of variables belonging to a prescribed pseudogroup. It has been observed by Gardner[7] that, in the first-order case ($r=1$), the Cartan form is part of an invariant adapted coframe obtained by formulating and solving the equivalence problem for (1.1) as a Cartan equivalence problem on J^1. In the higher-order case ($r>1$), it is *not* true that the Cartan form can be recovered by solving the equivalence problem on J^r. This had led some researchers, such as Shadwick,[9] to suggest that one should study the equivalence problem for r th-order variational problems on jet bundles J^{r+k}, where k is sufficiently large so as to yield the Cartan form (i.e., $k \geqslant r-1$), but otherwise arbitrary.

In the first paper in this series,[1] it was shown how to formulate the equivalence problem for the Lagrangian (1.1) as a Cartan equivalence problem on the space of $(r+k)$-jets for any $k \geqslant 0$. Moreover, we found that each of these potentially different equivalence problems, on the different bundles J^{r+k}, $k \geqslant 0$, are really the same problem, in that they all encode the same equivalence problem, and hence must have isomorphic solutions. Let us begin by recalling the basic definition and theorem on the equivalence of Lagrangians under point transformations.[1]

 0022-2488/89/040902-05$02.50

Definition 1: Two r th-order Lagrangians are said to be $(r+k)$*-standard equivalent*, $k \geqslant 0$, if and only if there is a contact map $\Psi: J^{r+k} \to J^{r+k}$ such that

$$\Psi^*\{\bar{L}\, d\bar{x}\} = L\, dx \bmod \mathscr{I}^{(r+k)}. \tag{1.4}$$

Theorem 1: Two r th-order Lagrangians are $(r+k)$-standard equivalent if and only if they are r-standard equivalent.

From this point of view, one does not gain anything as far as the ultimate solution to the equivalence problem is concerned by increasing the order of the jet bundle to serve as the base space, and, for simplicity, may as well solve the problem on the minimal-order jet bundle, viz. J^r. On the other hand, since the Cartan form (1.3) clearly involves $(2r-1)$st-order derivatives of u, it cannot arise as an invariant one-form if one solves the equivalence problem on a jet bundle of order $r+k$ for any $k<r-1$. For a first-order Lagrangian, this does not present any difficulties, as $1=r=2r-1$; however, for higher-order Lagrangians, difficulties arise since $r<2r-1$. For example, Cartan's solution to the second-order particle Lagrangian equivalence problem[10] does not lead to the Cartan form, as he implements the solution to this problem on the jet bundle J^2, while the relevant Cartan form lives on the bundle J^3.

A resolution of this apparent contradiction has been proposed in Ref. 1, where it was argued that the solution to the r th-order equivalence problem will lead to a purely r th-order differential form, which can be obtained from the Cartan form Θ_C by replacing all derivatives of order greater than r by the associated "derivative covariants," which are certain universal r th-order functions that can be constructed from the Lagrangian and its derivatives, with the remarkable property that they transform precisely like the higher-order derivatives of u. In a subsequent paper in this series, we hope to illustrate explicitly this point in the case of a second-order Lagrangian, but for now we will content ourselves with this rather general statement, and refer the reader to Ref. 1 for the details on this point. See also remarks in Sec. III.

II. THE CARTAN FORM

Our goal now is to prove the main result, that by setting up the equivalence problem for the variational problem (1.1) as a Cartan equivalence problem on J^{r+k}, where $k \geqslant r-1$, one obtains, after several iterations of Cartan's reduction procedure, the Cartan form Θ_C given by (1.3) as part of an adapted coframe. We will not attempt to make the complete reduction here (this is too hard to do for general r and k), but will discuss the second-order case in more detail in a future publication.

We begin by recalling how the standard equivalence problem for r th-order Lagrangians was encoded in terms of certain differential one-forms on the jet bundle J^{r+k}. The base coframe is given by the one-forms

$$\theta_0, \theta_1, \ldots, \theta_{r+k-1}, \quad \omega_0 = L\, dx, \quad \pi_0 = du_{r+k}, \tag{2.1}$$

where the θ_j are the contact forms given in (1.2). We assemble these into a column vector $\boldsymbol{\theta}_0 = (\theta_0, \theta_1, \ldots, \theta_{r+k-1})^T$, and use $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0, \omega_0, \pi_0)^T = (\theta_0, \theta_1, \ldots, \theta_{r+k-1}, \omega_0, \pi_0)^T$ to denote the complete column vector of coframe elements.

Given any non-negative integer $m \leqslant r+k$, we define a $\{\frac{1}{2}(r+k+3)(r+k)+m+2\}$-dimensional matrix Lie group $G^{(m)}$. It consists of all lower triangle matrices of the form

$$g = \begin{pmatrix} A & 0 & 0 \\ B & 1 & 0 \\ C & D & E \end{pmatrix}, \tag{2.2}$$

where $A = (A^i_j)$ is an invertible $(r+k)\times(r+k)$ lower triangular matrix, D and E are scalars, $E \neq 0$, and $B = (B_1, B_2, \ldots, B_{r+k})$ and $C = (C_1, C_2, \ldots, C_{r+k})$ are row vectors, with

$$B = (B_1, B_2, \ldots, B_m, 0, \ldots, 0). \tag{2.3}$$

Note that $G^{(l)} \subset G^{(m)}$ for $l<m$. In Pt. I,[1] we showed how the structure groups $G^{(m)}$ can be used to encode our equivalence problem in Cartan form.

Theorem 2: Let $r>1$, $k \geqslant 0$, and let $2 \leqslant m \leqslant r+k$. Two rth-order Langrangians L and $\bar{L}$ are $(r+k)$-standard equivalent under the pseudogroup of contact transformations if and only if there is a diffeomorphism $\Psi: J^{r+k} \to J^{r+k}$ that satisfies

$$\Psi^*(\bar{\boldsymbol{\eta}}_0) = g \cdot \boldsymbol{\eta}_0, \tag{2.4}$$

where $\bar{\boldsymbol{\eta}}_0$ and $\boldsymbol{\eta}$ are the respective coframes associated with the two Lagrangians and g is a $G^{(m)}$-valued function on J^{r+k}.

According to Bäcklund's theorem,[1,2] since any transformation preserving the contact ideal on J^{r+k} is the prolongation of a contact transformation on J^1, we could take the minimal value of $m=2$ to encode the equivalence problem; however, as we shall see, this would not lead us directly to the Cartan form. (The cases $m=1$ and $m=0$ will further restrict the allowable change of variables to the pseudogroups of point and fiber-preserving transformations, respectively.) The case $r=1$ is special, since it can be shown that equivalence of first-order Lagrangians under contact transformations automatically reduces to equivalence under point transformations,[7,11,12] so $m=1$ anyway. From here on, we will leave this case aside as it is already well understood.

The main result to be proved in this paper can now be stated as follows:

Theorem 3: Let L be an r th-order Lagrangian for $r \geqslant 2$. The Cartan form Θ_C, given by (1.3), appears naturally among the invariant adapted coframe elements resulting from an application of the Cartan method of equivalence to the equivalence problem (2.4) on the jet bundle J^{r+k} under the structure group $G^{(m)}$ provided $k \geqslant r-1$ and $m \geqslant r$.

Proof: The restrictions on k and m both follow from the ultimate form (1.3) for the Cartan form Θ_C; more on this later. To prove the general result, it suffices to start with the largest of the possible structure groups, so we assume that we are working in J^{r+k}, $k \geqslant r-1$, and using the structure group G^{r+k}. In accordance with the Cartan algorithm, we begin by lifting the problem to $J^{r+k} \times G^{r+k}$, and use $\boldsymbol{\eta} = g^{-1} \cdot \boldsymbol{\eta}_0$, i.e.,

$$\begin{pmatrix}\theta\\ \omega\\ \pi\end{pmatrix}=\begin{pmatrix}A & 0 & 0\\ B & 1 & 0\\ C & D & E\end{pmatrix}^{-1}\begin{pmatrix}\theta_0\\ \omega_0\\ \pi_0\end{pmatrix}, \tag{2.5}$$

as our lifted coframe. (The exponent -1 in the group element parametrization is introduced solely for computational convenience.) In particular,

$$\omega = L\,dx + \sum_{j=0}^{r+k-1} Z_j\theta_j, \tag{2.6}$$

where the coefficients Z_j are the entries of the row vector

$$Z = -BA^{-1}. \tag{2.7}$$

In the first loop of the algorithm implementing Cartan's method of equivalence,[6,7,8] we are supposed to compute the exterior derivatives of the lifted coframe and rewrite the result in terms of the right-invariant one-forms on the structure group $G^{(r+k)}$, i.e., the entries of the matrix differential $g^{-1}\cdot dg$, and the lifted coframe elements $\boldsymbol{\eta}$. It turns out that, for the purposes of recovering the Cartan form, we need only look at the formulas for the differential $d\omega$, and so we will concentrate on this single component of the structure equations throughout. Using (2.5) and (2.6), we find that

$$d\omega = \sum_{i=0}^{r+k-1} (\beta_{i+1}\wedge\theta_i + T_i\omega\wedge\theta_i + T_i'\pi\wedge\theta_i) + T^*\pi\wedge\omega,$$

where T_i, T_i', T^*, are certain *torsion coefficients*, depending on the group parameters and the base coframe, and where the β_j are the right-invariant one-forms on the structure group $G^{(r+k)}$ corresponding to the group parameters B_j. After performing an obvious Lie algebra compatible absorption of torsion,[7,8] we are left with the structure equation

$$d\omega = \sum_{i=0}^{r+k-1} \tilde\beta_{i+1}^{(r+k)}\wedge\theta_i + T^*\pi\wedge\omega,$$

where the $\tilde\beta_j^{(r+k)}$ are congruent modulo the lifted coframe to the right-invariant one-forms β_j. Thus we readily deduce that only the coefficient

$$T^* = -(E/L)Z_{r+k},$$

is essential torsion. Clearly, $G^{(r+k)}$ acts on the essential torsion coefficient T^* by translation, and we can normalize this torsion coefficient to 0 by setting $Z_{r+k}=0$, or, equivalently, by setting $B_{r+k}=0$. Thus, at this stage, the algorithm automatically tells us to reduce the structure group to the subgroup $G^{(r+k-1)}$.

Thanks to the reduction theorem for the Cartan equivalence problem,[7,8] we know that the reduced problem with the same base coframe (2.1) and reduced structure group $G^{(r+k-1)}$ has the *same* set of solutions as the original equivalence problem. We proceed to analyze this reduced equivalence problem. Since $k\geqslant r-1\geqslant 1$, a second Lie algebra compatible absorption of torsion in the recomputed structure equation for $d\omega$ will yield

$$d\omega = \sum_{i=0}^{r+k-2} \tilde\beta_{i+1}^{(r+k-1)}\wedge\theta_i + T_{r+k-1}^*\,\theta_{r+k-1}\wedge\omega,$$

where the $\tilde\beta_j^{(r+k-1)}$ are congruent modulo the lifted coframe to the right-invariant one-forms β_j. Again, $G^{(r+k-1)}$ acts on the essential torsion coefficient

$$T_{r+k-1}^* = -(A_{r+k}^{r+k}/L)Z_{r+k-1}$$

by translation, and we can normalize the torsion coefficient to 0 by setting $Z_{r+k-1}=0$, or, equivalently, $B_{r+k-1}=0$, further reducing to the structure group $G^{(r+k-2)}$.

Clearly, this procedure continues until the derivatives of the Lagrangian L start contributing to the essential torsion in $d\omega$. This will occur when we have reduced our original problem to an equivalence problem with the same base coframe (2.1), and reduced structure group $G^{(r)}$. We now indicate how the above analysis changes at this point. After Lie algebra compatible absorption of torsion, we find

$$d\omega = \sum_{i=0}^{r-1} \tilde\beta_{i+1}^{(r)}\wedge\theta_i + T_r^*\,\theta_r\wedge\omega$$

as before, but where the essential torsion is now given by

$$T_r^* = -\frac{A_{r+1}^{r+1}}{L}\left(Z_r - \frac{\partial L}{\partial u_r}\right).$$

The structure group $G^{(r)}$ still acts on the essential torsion by translation, but there is an additional inhomogeneous term. Consequently, we can normalize this torsion coefficient to 0 by setting

$$Z_r = \frac{\partial L}{\partial u_r}. \tag{2.8}$$

Now, plugging (2.8) into the formula (2.6) for ω (with earlier normalizations $Z_{r+1}=\cdots=Z_{r+k}=0$ also taken into account) has the effect of (a) reducing the structure group to $G^{(r-1)}$, just as before, and (b) to incorporate the inhomogeneity, changing the base coframe so as to replace our original one-form $\omega_0 = L\,dx$ by the new one-form

$$\omega_0^{(r-1)} = L\,dx + \frac{\partial L}{\partial u_r}(du_{r-1} - u_r\,dx).$$

This new base coframe element constitutes our first "approximation" to the Cartan form. The corresponding lifted one-form coincides with (2.6) taking (2.8) into account, i.e.,

$$\omega = L\,dx + \frac{\partial L}{\partial u_r}\theta_r + \sum_{i=1}^{r-1} Z_i\theta_{i-1}.$$

We continue our reduction procedure by again recomputing the basic structure equation and reabsorbing. Now we find

$$d\omega = \sum_{i=0}^{r-2} \tilde\beta_{i+1}^{(r-1)}\wedge\theta_i + T_{r-1}^*\,\theta_{r-1}\wedge\omega + \cdots,$$

where the dots stand for other essential torsion terms that we will try not to deal with here. As usual, we normalize the torsion coefficient

$$T_{r-1}^* = -\frac{A_r^r}{L}\left(-Z_{r-1} + \frac{\partial L}{\partial u_{r-1}} - D_x\frac{\partial L}{\partial u_r}\right)$$

to 0 by setting

$$Z_{r-1} = \frac{\partial L}{\partial u_{r-1}} - D_x\frac{\partial L}{\partial u_r}.$$

Now we have reduced the structure group to $G^{(r-2)}$, and also modified the base coframe so as to replace $\omega_0^{(r-1)}$ by

$$\omega_0^{(r-2)} = L\,dx + \frac{\partial L}{\partial u_r}\theta_r + \left(\frac{\partial L}{\partial u_{r-1}} - D_x\frac{\partial L}{\partial u_r}\right)\theta_{r-1},$$

giving the next approximation to the Cartan form.

Clearly, we can continue in this manner, and a simple inductive argument will show that we end up normalizing all the entries of the vector Z, cf. (2.7), as

$$Z_i = \sum_{j=0}^{r-1} (-D_x)^j\left(\frac{\partial L}{\partial u_{i+j}}\right), \quad i = 1,...,r. \tag{2.9}$$

The structure group has finally been reduced to $G^{(0)}$, which consists of all invertible matrices of the form (2.2) with $B = 0$. Substituting all the normalizations (2.9) into (2.6), we see that the base coframe element replacing ω_0 is now the Cartan form (1.3). Moreover, since the corresponding row of the structure group matrix consists of all 0s save for a 1 in the diagonal position, the Cartan form Θ_C is invariant under contact transformations (for the standard equivalence problem), and will be part of the invariant adapted coframe resulting from the full implementation of the Cartan algorithm. The general reduction theorem completes the proof of Theorem 3.

III. DISCUSSION

We now return to a more detailed discussion of our initial formulation of the equivalence problem. What we have shown is that, if we formulate the basic Lagrangian equivalence problem on the jet bundle J^{r+k} for $k \geqslant r-1$, *and* use the group $G^{(m)}$ for $m \geqslant r$ as our structure group, then the Cartan reduction procedure will naturally lead us to the Cartan form as discussed in Sec. II. There are two obvious objections to this formulation: First, according to the reduction theorem of Ref. 1, we are really working on too high an order jet bundle, and second, according to Bäcklund's theorem, we are using too large a structure group. Let us discuss the latter difficulty first.

As was presented in Pt. I,[1] Bäcklund's theorem[2] tells us that any contact transformation on J^{r+k} is just the prolongation of a contact transformation on the first jet bundle J^1. In particular, the base transformation of the independent variable x depends on at most first-order derivatives of u, $\bar{x} = \varphi(x,u,u_1)$, so the pull back $\Psi^*(d\bar{x}) = d\varphi$ will only involve the form dx and the first two contact forms θ,θ_1. This means that the structure group will naturally reduce to a subgroup of the group $G^{(2)}$, and we could have begun our reduction procedure with $G^{(2)}$ as the starting structure group without losing anything as far as the final solution to our equivalence problem is concerned. However, it is easy to see that, for $r \geqslant 3$, the $G^{(2)}$ equivalence problem can never lead to the Cartan form Θ_C as an adapted invariant coframe element. Indeed, in this case the lifted coframe element corresponding to the base form ω_0 just depends on the first two contact forms:

$$\omega = \omega_0 + B_1\theta + B_2\theta_1. \tag{3.1}$$

Barring prolongation, the Cartan reduction algorithm will eventually normalize the group parameters B_1, B_2, to be certain combinations of the Lagrangian and its derivatives, leading to an adapted coframe element of the same form (3.1). If $r \geqslant 3$, this cannot be the Cartan form (1.3) since it does not involve enough contact forms!

We seem to be left with a paradox: If we reduce the equivalence problem using the larger structure group $G^{(m)}$ for $m \geqslant r$, we are naturally led to the Cartan form, whereas if we reduce using the more reasonable structure group $G^{(2)}$, which mathematically encodes the *same* equivalence problem, we cannot obtain the Cartan form directly. This state of affairs appears to be contradictory, especially considering that all these problems are the same, and must therefore lead to the same necessary and sufficient conditions for equivalence of the two variational problems. The resolution of the difficulty is to realize that the Cartan solution to the $G^{(2)}$ equivalence problem will lead to additional adapted invariant coframe elements that will be certain particular linear combinations of the contact forms alone. Since any linear combination of invariant one-forms, whose coefficients are scalar invariants, is itself an invariant one-form, we conclude that the Cartan form (1.3) *must* appear in this version of the equivalence problem, but in disguised form. Namely, we deduce that there *is* an invariant one-form of the form

$$\omega^* = \omega_0 + B_1^*\theta + B_2^*\theta_1,$$

where B_1^* and B_2^* will be certain combinations of L and its derivatives. Moreover, there exist *additional* invariant one-forms that are certain combinations of contact forms

$$\theta_j^* = \sum_{i=0}^{r} A_{ij}^*\,\theta_i,$$

with the property that Θ_C is the sum of these component pieces,

$$\Theta_C = \omega^* + \sum I_j\,\theta_j^*, \tag{3.2}$$

where the I_j are either constants, or, perhaps, invariants of the problem. Thus the Cartan form does appear as an invariant one-form for the $G^{(2)}$, but in the disguised form (3.2), *not* directly as an adapted coframe element. (In a subsequent paper, we will illustrate this point in some special cases.) It would be interesting to find the formulas for the "reduced" invariant one-form ω^* and determine its geometric or analytic significance for the original Lagrangian.

However, as we have demonstrated, the Cartan form appears much more directly if we "artificially" expand the original structure group to be $G^{(r+k)}$ (or even just $G^{(r)}$) even though we know that this is ultimately not necessary for the solution of the equivalence problem. A key lesson of this exercise appears to be that the use of different (larger) structure groups to encode the self-same equivalence problem can lead to different adapted coframe elements, even though all the different possible invariant coframes must be related to each other according to a formula like (3.2).

There is another way to interpret our results. We could begin the entire reduction procedure by using the reduced structure group $G^{(2)}$ initially, as would be warranted by the form of the contact transformations. However, we would need to compensate by replacing our original base coframe element $\omega_0 = L\,dx$ by a slightly different one-form having the form

$$\tilde{\omega}_0 = L\,dx + \sum_{j=0}^{r+k-1} \lambda_j \theta_j,$$

where the coefficients $\lambda_j: J^{r+k} \to \mathbb{R}$ are arbitrary, to be determined during the course of the application of Cartan's method. However, as the reader can verify, these two approaches are essentially the same and lead to the same conclusion.

The second difficulty with our original formulation is that we were forced to use a higher-order jet bundle, namely, J^{2r-1}, than is really necessary for solving the equivalence problem. Indeed, if we do solve the Cartan equivalence problem on the minimal-order jet bundle J^r, then, barring prolongation, we will be led to a complete set of r th-order invariants and invariant one-forms. How does the Cartan form arise here? The answer is provided by the "derivative covariants," which are certain combinations of the Lagrangian and its derivatives of u. (See Ref. 1 for the details.) If we replace all the derivatives of u of order higher than r that appear in the Cartan form (1.2) by their corresponding derivative covariants, we will be led to a purely r th-order one-form, which incorporates all the transformation properties of the Cartan form, even though the explicit formula for it will be quite a bit more complicated than (1.3). (For instance, it will depend nonlinearly on the Lagrangian.) Thus there is purely r th-order invariant one-form that corresponds to the Cartan form, and hence will appear in the equivalence problem on J^r, either directly as an adapted coframe element, or more probably, in disguised form similar to (3.2).

In a subsequent paper in this series, we will illustrate all these matters with a concrete problem—the equivalence problem for a second-order particle Lagrangian. Also, we hope to extend these techniques to higher dimensional Lagrangians, where the nonuniqueness of the Cartan form becomes an issue.[13,14]

ACKNOWLEDGMENTS

NK's research supported in part by a grant from NSERC. PJO's research supported in part by NSF grant DMS 86-02004.

[1] N. Kamran and P. J. Olver, "Equivalence of higher-order Lagrangians I. Formulation and reduction" (preprint, 1988).

[2] R. L. Anderson and N. H. Ibragimov, *Lie-Bäcklund Transformations in Applications* (SIAM Studies in Appl. Math. #1, Philadelphia, 1979).

[3] W. F. Shadwick, "The Hamiltonian formulation of regular r th-order Lagrangian field theories," Lett. Math. Phys. **6**, 409 (1982).

[4] H. Goldschmidt and S. Sternberg, "The Hamiltonian–Cartan formalism in the calculus of variations," Ann. Inst. Fourier Grenoble **23**, 203 (1973).

[5] P. L. Garcia, "The Poincaré–Cartan invariant in the calculus of variations," in *Symposia Mathematica*, Vol. 14, Instituto Nazionale di Alta Matematica (Academic, New York, 1974), pp. 219–246.

[6] E. Cartan, "Les problèmes d'équivalence," in *Oeuvres Complètes* (Gauthiers–Villars, Paris, 1953), Pt. II, Vol. 2, pp. 1311–1334.

[7] R. B. Gardner, "Differential geometric methods interfacing control theory," in *Differential Geometric Control Theory*, edited by R. W. Brockett *et al.* (Birkhauser, Boston, 1983), pp. 117–180.

[8] N. Kamran, "Contributions to the study of the equivalence problem of Elie Cartan and its applications to partial and ordinary differential equations" (preprint, 1988).

[9] W. F. Shadwick (private communication).

[10] E. Cartan, "La géométrie de l'intégrale $\int F(x,y,y',y'')\,dx$," in *Oeuvres Complètes* (Gauthiers–Villars, Paris, 1955), Pt. III, Vol. 1, pp. 1341–1368.

[11] R. L. Bryant, "On notions of equivalence of variational problems with one independent variable," Contemp. Math. **68**, 65 (1987).

[12] N. Kamran and P. J. Olver, "Equivalence problems for first order Lagrangians on the line," to be published in J. Differential Equations.

[13] M. Horák and I. Kolář, "On the higher order Poincaré–Cartan forms," Czech. Math J. **33**, 467 (1983).

[14] D. J. Saunders, "An alternative approach to the Cartan form in Lagrangian field theories," J. Phys. A: Math. Gen. **20**, 339 (1987).

The Φ^4 equations of motion. III. The sweeping factors

Marietta Manolessou
Institut des Hautes Etudes Scientifiques, 35, route de Chartres, 91440 Bures-sur-Yvette, France

(Received 15 December 1987; accepted for publication 27 July 1988)

This is the third in a series of three papers concerning the study of the infinite system of Φ^4 equations of motion for the Schwinger functions by a fixed-point method. In Paper II [J. Math. Phys. **30**, 175 (1989)] the solution of the zero- (one-) and two-dimensional systems was constructed inside particular subsets of the corresponding Banach spaces characterized by the "signs" and "splitting" properties of the Green's functions at zero external momenta. In the present article the recursive formulas and the absolute bounds of the *sweeping factors*, presented without proof in II, are proved. These sequences of sweeping factors for the global terms of the equations have been the most important combinatorial tools throughout the arguments, and their properties have been used in the proofs of the basic theorems in II. In Appendix A a first confirmation of the method, provided by the numerical identification of the author's zero-dimensional solution, with the one obtained directly by the standard functional integral formalism, is presented.

I. INTRODUCTION

This is the third paper (III) in a series concerning the solution of the zero-, one-, and two-dimensional Φ^4 equations of motion for the Schwinger functions. The results of I and II[1] and of III constitute an intermediate step in our general program towards a fixed-point method for the construction of a nontrivial Φ_4^4 Wightman theory.

In I we presented the general outlook of the method, together with the results revealed by the convergent iterative procedure of the Φ_2^4 system of equations of motion, namely, the Φ iteration (cf. Sec. II of I), which motivated us to the present approach. Precisely, we have explained how the conservation of signs, the splitting properties in terms of bounded sequences $\{\delta_n\}_n$, and the precise norms in the space $\mathscr{B}$ of the Schwinger sequences $H=\{H^{n+1}\}_{n=2k+1,\,k\in\mathbb{N}}$ have been crucial for the convergence of the Φ iteration to the solution.

In II the "nice" properties of the Φ iteration have been reformulated in terms of self-consistent conditions, on the H^{n+1} functions and the splitting sequences, which characterize the subsets $\Phi_{0\Lambda}\subset\mathscr{B}_0$ (for the zero-dimensional case) or $\Phi_\Lambda\subset\mathscr{B}$ (for the one- and two-dimensional cases), at fixed values of the real and positive coupling constant Λ (cf. Sec. II of II). Using appropriate norms for the corresponding spaces $\mathscr{B}_0$, $\mathscr{B}$ a fixed point method of contractive mappings on Banach spaces has been applied. We found a unique nontrivial solution of the zero-dimensional problem inside $\Phi_{0\Lambda}$ (cf. Sec. III of II) and, respectively, a unique nontrivial solution of the two- (or one-) dimensional problem inside Φ_Λ (cf. Sec. IV of II).

As we found in I (cf. Sec. III) and II (cf. Secs. II–IV), the fundamental combinatorial tools for the method (revealed by the study of the zero-dimensional problem) are the so-called "sweeping factors" β_n, ${}^\Phi\beta_n$ and α_n, ${}^\Phi\alpha_n$ ($\forall n=3,5,\dots$), respectively. These sequences play a double role in the construction of the technique [cf. Definitions 3(a), 3(c), and 3(e) of I and 2(f) of II].

(a) They replace the sums C_0^{n+1} and ${}^\Phi C_0^{n+1}$ (resp. B_0^{n+1} and ${}^\Phi B_0^{n+1}$),

$$C_0^{n+1}=-6\Lambda\sum_{w(I)}\theta^n_{i_1i_2i_3}\prod_{l=1}^3 H_0^{i_l+1},$$

$$\theta^n_{i_1i_2i_3}=\frac{n!}{i_1!i_2!i_3!\sigma_{\text{sym}}(i_1i_2i_3)},$$

$$\left(\text{resp. } B_0^{n+1}=-3\Lambda\sum_{w(J)}\theta^n_{j_1j_2}H_0^{j_2+2}H_0^{j_1+1}\right),$$

$$\theta^n_{j_1j_2}=\frac{n!}{j_1!j_2!}\tag{1.1}$$

[here $w(I)\equiv(i_1i_2i_3)$, $w(J)\equiv(j_1j_2)$ are partitions of n], by only one term proportional to the dominant contribution: $[n(n-1)/2]H_0^{n-1}H_0^2H_0^2$ and $nH_0^{n+1}H_0^2$, respectively. This is the sweeping procedure, which also can be written as a recursive formula in terms of $\beta_{\bar n}$, ${}^\Phi\beta_{\bar n}$ [and the splitting constants $\delta_{\bar n}$ and ${}^\Phi\delta_{\bar n}$ with $\bar n\leqslant n-2$, $\forall H_0\in\Phi_{0\Lambda}$ (or $H\in\Phi_\Lambda$)]. Moreover the splitting reformulated in Sec. III of I is simpler in terms of β_n:

$$|H_0^{n+1}|=\delta_n\beta_n|H_0^{n-1}|\,|H_0^2|.\tag{1.2}$$

(b) They carry precise combinatorial information about all the terms of the corresponding sums they sweep. This information (explicit dependence on n) is formulated in terms of bounds they satisfy, namely,

$$\begin{aligned}&\beta_n\sim\overline{\mathscr{T}}_n\quad(\text{resp. }\alpha_n\sim\mathscr{T}_n),\\&{}^\Phi\beta_n\sim\overline{\mathscr{T}}_n\quad(\text{resp. }{}^\Phi\alpha_n\sim\mathscr{T}_n).\end{aligned}\tag{1.3}$$

Here $\overline{\mathscr{T}}_n$ (resp. $\mathscr{T}_n$) is the number of different possible configurations $(i_1i_2i_3)$ [resp. (j_1j_2)] of n inside C_0^{n+1} (resp. B_0^{n+1}) and is calculated precisely in Appendix B of the present paper. We find $\overline{\mathscr{T}}_n\sim[(n-3)^2/48]$ [resp. $\mathscr{T}_n=(n-1)/2$]. The proportionality constants in (1.3) are positive and smaller than unity.

This twofold aspect of β_n (resp. α_n) and ${}^\Phi\beta_n$ (resp. ${}^\Phi\alpha_n$) that we explained in (a) and (b) has been presented without proof in II in terms of Lemmas 2.2, 2.3, and 2.6 (resp. 2.5 and 2.7) and it has been extensively used there for

 0022-2488/89/040907-23$02.50

all the proofs concerning the zero- and two- (or one-) dimensional problems mentioned above.

The purpose of the present paper is the proof of these lemmas. So, in Sec. II we first present the demonstration of the recursive sweeping procedure of β_n's and ${}^{\Phi}\beta_n$'s, respectively, in terms of the splitting sequences (cf. Propositions 2.1 and 2.2, respectively). Then we show the absolute bounds mentioned above, together with relative bounds, stated by Propositions 2.3 and 2.4, respectively.

In Sec. III we prove the corresponding sweeping procedure and absolute and relative bounds satisfied by the sweeping sequences $\{\alpha_n\}$, $\{{}^{\Phi}\alpha_n\}$, respectively, stated by Propositions 3.1 and 3.2, respectively. Moreover, in this section we show the corresponding absolute and relative bounds of the functionals,

$$\Delta_n \equiv \alpha_n/(n-1) - \delta_{n+2}\beta_{n+2}/3n(n-1)$$

(resp. ${}^{\Phi}\Delta_n$) and in particular we find the interval in $\mathbb{R}^+$ where the precise value of the $\delta_\infty^\Lambda \equiv \lim_{n\to\infty} \delta_n(\Lambda)$ (resp. $\tilde{\delta}_\infty^\Lambda$) is allowed to belong. These results have been crucial in order to ensure the stability of the subset $\Phi_{0\Lambda}$ (resp. Φ_Λ) (cf. Secs. III and IV of II).

In Appendix B, we evaluate exactly the numbers $\mathcal{T}_n$ and in Appendix C (as promised in I) we give explicitly our original "experimental" results, in the first three orders of the Φ iteration, concerning the signs, splitting, bounds, and convergence properties of the two-dimensional H^{n+1} functions and their nontrivial Φ convolutions. Finally, in Appendix A we present the numerical verification of the method in the zero-dimensional case mentioned in I and II. The bases for this verification are the numerical results of Voros[2] concerning the identification of the solution of the zero-dimensional problem obtained by our method, with the corresponding solution of the equivalent zero-dimensional system obtained directly by the standard functional integral method.[3]

II. RECURSIVE PROCEDURE AND BOUNDS FOR β_n's

In I the sweeping procedure through the sum C_0^{n+1} has been defined in terms of the Green's functions H_0^{n+1}. We start this section by proving that when $H_0 \in \Phi_{0\Lambda} \subset B_0$ [cf. Defintion 2(b) of II], then it also can be explicitly written as a recursion in terms of the ratios of the splitting constants $\delta_{\bar{n}}$, $\forall \bar{n} \leqslant n-2$. We state this property by the following proposition, which is exactly the same as Lemma 2.1 of II. Notation: always $n = 2k+1$, $k \in \mathbb{N}$.

Proposition 2.1 (sweeping procedure): Let $H_0 \in \Phi_{0\Lambda}$, $\forall n \geqslant 3$. The sweeping factors β_n (resp. $\beta^n_{i_1 i_2 i_3}$), $\forall$ partition $(i_1 i_2 i_3)$ are given recurrently as follows:

$\beta_3 = \beta_5 = 1,$

$$\forall n \geqslant 7, \quad \beta_n \equiv \beta^n_{n-2,1,1} \quad \text{and} \quad \forall (i_1 i_2 i_3) \text{ with } i_1 \geqslant i_2 \geqslant i_3,$$

$$\begin{aligned}\beta^n_{i_1 i_2 i_3} = 1 &+ \zeta(i_1 i_2) \frac{\theta^n_{i_1-2, i_2+2, i_3}}{\theta^n_{i_1 i_2 i_3}} \frac{\delta_{i_2+2}}{\delta_{i_1}} \frac{z(\beta_{i_2+2})}{z(\beta_{i_1})} \beta^n_{i_1-2, i_2+2, i_3} \\ &+ \zeta(i_2 i_3) \frac{\theta^n_{i_1, i_2-2, i_3+2}}{\theta^n_{i_1 i_2 i_3}} \frac{\delta_{i_3+2}}{\delta_{i_2}} \frac{z(\beta_{i_3+2})}{z(\beta_{i_2})} \beta^n_{i_1, i_2-2, i_3+2} \\ &+ \zeta(i_1 i_3) \frac{\theta^n_{i_1-2, i_2, i_3+2}}{\theta^n_{i_1 i_2 i_3}} \frac{\delta_{i_3+2}}{\delta_{i_1}} \frac{z(\beta_{i_3+2})}{z(\beta_{i_1})}. \end{aligned} \tag{2.1}$$

Here

$$z(\beta_i) = \begin{cases} \beta_i, & \forall i \neq 5, \\ 4, & \text{for } i = 5; \end{cases} \tag{2.1a}$$

$$\zeta(i_1 i_2) = \begin{cases} 1, & \text{if } \begin{cases} i_1 > (n+1)/2 \text{ and } i_2 = n - i_1 - 1, \\ i_1 \leqslant (n+1)/2 \text{ and } i_2 = i_1 - 4, \end{cases} \\ 0, & \text{otherwise;} \end{cases} \tag{2.1b}$$

$$\zeta(i_2 i_3) = \begin{cases} 1, & \text{if } i_3 \leqslant i_2 - 4, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and } \zeta(i_1 i_3)$$

$$= \begin{cases} 1, & \text{if } i_1 - 2 = i_2 = n/3, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1c}$$

Proof of Proposition 2.1: For the proof, Definition 3(c) of I and formulas (3.5a) and (3.5b) are essentially applied. We shall use a recursion. The starting point of this is the initial equal values of (2.1), $\beta_3 = \beta_5 = 1$, which are trivially obtained from the definitions [cf. (1.1)]

$$C_0^4 \equiv -6\Lambda[H_0^2]^3, \quad C_0^6 = -60\Lambda H_0^4 [H_0^2]^2. \tag{2.2}$$

We then suppose that (2.1) holds $\forall \bar{n}$ with $5 \leqslant \bar{n} \leqslant n-2$ and call this hypothesis H.1. To prove it for $\bar{n} = n$, we first note that for the minimal value of i_1, $\hat{i}_1 \equiv i_{1\,\min}(n) = [n/3]$ (where [ ] means the integral part), we have the three possible particular cases of partitions $(\hat{i}_1, \hat{i}_2, \hat{i}_3)$ when $\hat{i}_2 = i_{2\,\min}(i_1)$ (i.e., the minimal value of i_2 corresponding to $\hat{i}_1$), namely, (2.3a)

either

$$\hat{i}_1 = \hat{i}_2 = \hat{i}_3 = n/3 \tag{2.3a}$$

or

$$\hat{i}_1 = \hat{i}_2 + 2 = \hat{i}_3 + 2 \tag{2.3b}$$

or

$$\hat{i}_1 = \hat{i}_2 = \hat{i}_3 + 2. \tag{2.3c}$$

For all of these cases we trivially obtain, in view of formulas (3.5a) and (3.5b) of I (only one term is contained in the partial sum of the rhs),

$$\beta^n_{\hat{i}_1, \hat{i}_2, \hat{i}_3} = 1. \tag{2.4}$$

This is the starting point of the following secondary recurrence hypothesis called H.2: For an arbitrary fixed triplet, $w(\bar{I}) = (\bar{i}_1, \bar{i}_2, \bar{i}_3)$. We suppose that property (2.1) holds for all partitions $w(I) = (i_1 i_2 i_3)$ such that $i_1 \leqslant \bar{i}_1 - 2$ or $i_1 \leqslant \bar{i}_1$, $i_2 \leqslant \bar{i}_2 - 2$, $i_3 \geqslant \bar{i}_3 + 2$, and prove it for the partition $w(\bar{I})$. We first show the following auxiliary equality:

$$H_0^{j-1} H_0^{j'+3} = \frac{\delta_{j'+2}}{\delta_j} \frac{z(\beta_{j'+2})}{z(\beta_j)} H_0^{j+1} H_0^{j'+1},$$
$$\forall 5 \leqslant j \leqslant n-2, \quad 1 \leqslant j' \leqslant n-4. \tag{2.5}$$

Here the $z(\beta)$ are defined by (2.1a). By combining the assumption $H_0 \in \Phi_{0\Lambda}$ (signs) and the splitting property (1.2) we write

$$H_0^{j'+3} = -\delta_{j'+2} z(\beta_{j'+2}) H^{j'+1} H_0^2 \tag{2.5a}$$

and

$$H_0^{j-1} = -\frac{H_0^{j+1}}{H_0^2}\frac{1}{\delta_j z(\beta_j)}. \quad (2.5b)$$

Insertion of (2.5a) and (2.5b) in the lhs of (2.5) yields directly the rhs of it. Q.E.D.

Notice that in the above equations, the sweep factors $\beta_{j'+2}, \beta_j$ are well defined recurrently by the corresponding formula (2.1) in view of the recurrence hypothesis H.1.

Now we establish a double ordering on the terms inside the sum C_0^{n+1}: On one hand, following the values of i_1 inside $[n/3] \leqslant i_1 \leqslant n-2$ (fixed n), and, on the other hand, following the possible values of i_2: $[(n-1)/2] \leqslant i_2 \leqslant i_1$ (fixed i_1). Of course, the recurrence hypothesis H.2, and the assumption $H_0 \in \Phi_{0\Lambda}$ (signs) have to be taken into account in this ordered sum. The basic ingredient for the proof is the application of formula (2.5) on the original definition of the sweeping procedure [(3.5a) and (3.5b) in I (Sec. III)] in terms of the Green's functions $H_0^{i_l+1}$. We shall present the argument for only one of the cases (2.1b) or (2.1c), which seems to us the most intricate, all the remaining being simpler to verify by analogy.

Let $\bar{i}_1 > (n+1)/2$, $\bar{i}_2 = n - \bar{i}_1 - 1$, and $\bar{i}_3 \leqslant \bar{i}_2 - 4$. The second condition implies that $\bar{i}_3 = 1$ so that the partition of n $(\bar{i}_1 - 2, \hat{i}_2, \hat{i}_3)$, with $\hat{i}_3 \equiv i_{3\,\min}(\bar{i}_1 - 2)$ (i.e., the one that corresponds to the minimal value of i_3 at fixed $i_1 = \bar{i}_1 - 2$) satisfies $\hat{i}_3 = 1$. It follows that $\bar{i}_3 = \hat{i}_3$. This remark means that we can use formula (3.5b) of I for the definition of $\beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3}$, i.e.,

$$\left| \sum_{\substack{w(I) \\ i_1 \leqslant \bar{i}_1,\ i_2 \leqslant \bar{i}_2(\bar{i}_1)}} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1} \right| = \beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \prod_{l=1}^{3} |H_0^{\bar{i}_l+1}|. \quad (2.6)$$

By application of the sign properties (in view of the hypothesis $H_0 \in \Phi_{0\Lambda}$) in (2.6), we can rewrite it in a more expanded form by separating the partial sums of the lhs:

$$\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \prod_{l=1}^{3} H_0^{\bar{i}_l+1} + \sum_{\substack{w(I) \\ i=\bar{i}_1,\ i_2 \leqslant \bar{i}_2 - 2}} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1}$$
$$+ \sum_{\substack{w(I) \\ i_1 \leqslant \bar{i}_1 - 2}} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1} = \beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \prod_{l=1}^{3} H_0^{\bar{i}_l+1}. \quad (2.7)$$

Application of H.2 on the second and third term of the lhs of (2.7) yields

$$\beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} = 1 + \beta^n_{\bar{i}_1,\bar{i}_2-2,\bar{i}_3+2} \frac{\theta^n_{\bar{i}_1,\bar{i}_2-2,\bar{i}_3+2} H^{\bar{i}_2-1} H^{\bar{i}_3+3}}{\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} H^{\bar{i}_2+1} H^{\bar{i}_3+1}}$$
$$+ \beta^n_{\bar{i}_1-2,\hat{i}_2,\hat{i}_3} \frac{\theta^n_{\bar{i}_1-2,\hat{i}_2\hat{i}_3} H_0^{\bar{i}_1-1} H_0^{\hat{i}_2+1} H_0^{\hat{i}_3+1}}{\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \Pi_{l=1}^3 H_0^{\bar{i}_l+1}}. \quad (2.8)$$

We notice that when $\hat{i}_3 = 1 = \bar{i}_3$, then $\hat{i}_2 = \bar{i}_2 + 2$. So the third term on the rhs of (2.8) contains also only ratios of couples of Green's functions, namely, $H_0^{\bar{i}_1-1} H_0^{\bar{i}_2+3}/H_0^{\bar{i}_1+1} H_0^{\bar{i}_2+1}$. This last remark allows us to apply the equality (2.5) on both the 2^d and 3^d terms of (2.8). We finally obtain formula (2.1) with $\zeta(i_1 i_2) = 1 = \zeta(i_1 i_3)$ and $\zeta(i_1 i_3) = 0$ as expected. Q.E.D.

All other situations given by conditions (2.1b) and (2.1c) can be treated following similar considerations and arguments.

So, we have proved H.2 for the arbitrarily chosen $w(\bar{I})$. To obtain the proof of H.1 we apply H.2 to the particular case $w(\bar{I}) = (n-2,1,1)$ and obtain

$$\beta_n \equiv \beta^n_{n-2,1,1} = 1 + \frac{\theta^n_{n-4,3,1}\delta_3}{\theta^n_{n-2,1,1}\delta_{n-2} z(\beta_{n-2})} \beta^n_{n-4,3,1}. \quad (2.9)$$

Q.E.D.

This closes the recursion H.1 and the proof of Proposition 2.1.

Using exactly the analogous arguments of that presented above we can prove, when $H \in \Phi_\Lambda$, the sweeping procedure in $\mathscr{B}$ (cf. Lemma 2.6 of II) which gives the corresponding recursive definition of the sweeping factors $^\Phi\beta_n$ in terms of the sequences of parameters $^\Phi C^n_{i_1 i_2 i_3}$ associated to every coherent sequence of ΦC's and the corresponding splitting sequences $\{\delta_n\}$. The key equality for the proof is the analog of (2.5) [obtained again by application of the corresponding sign-splitting properties (2.31a) of II for Φ_Λ], namely,

$$^\Phi H_0^{j-1} H_0^{j'+3}\, {^\Phi C}^n_{j'+2,j-2,n-j-j'}$$
$$= \frac{\delta_{j'+2}}{\delta_j} \frac{z(^\Phi\beta_{j'+2})}{z(^\Phi\beta_j)}\, {^\Phi C}^n_{j',j,n-j-j'}\, {^\Phi H}_0^{j+1}\, {^\Phi H}_0^{j'+1}. \quad (2.10)$$

Again one uses a double recursion: a principal (the analog of H.1) over $\bar{n}$ with $5 \leqslant \bar{n} \leqslant n$, and a secondary (the analog of H.2) over i_1 with $[n/3] \leqslant i_1 \leqslant n-2$. The basic step is again the application of (2.10) inside the expanded form of an equality analogous to (2.6), namely,

$$\left| \sum_{\substack{w_n(I) \\ i_1 \leqslant \bar{i}_1,\ i_2 \leqslant i_2(\bar{i}_1)}} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} {^\Phi H}_0^{i_l+1} \left[\Phi^{(\bar{n},n)}(H) \prod_{l=1}^{3} N_1^{(i_l)} \right]_0 \right|$$
$$= {^\Phi\beta}^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \left| \prod_{l=1}^{3} {^\Phi H}_0^{\bar{i}_l+1} \right| \left| \left[\Phi^{(\bar{n},n)} \prod_{l=1}^{3} N_1^{(\bar{i}_l)} \right] \right|. \quad (2.11)$$

These considerations allow one to state without proof the following proposition.

Proposition 2.2 (sweeping procedure in $\mathscr{B}$): Let $H \in \Phi_\Lambda$. For every coherent sequence of ΦC's, the corresponding sweeping factors $^\Phi\beta_n$ and $^\Phi\beta^n_{i_1 i_2 i_3}$ introduced by definitions (1.8a) and (1.8b) of II, satisfy the recurrence relations presented by Eq. (2.33) of Lemma 2.5 in Sec. II of II, which are the analogs of (2.1) (Proposition 2.1 above).

Using the above explicit expressions of β_n's (resp. $^\Phi\beta$'s) or $\beta^n_{i_1 i_2 i_3}$'s (resp. $^\Phi\beta^n_{i_1 i_2 i_3}$'s) we derive through a double nested recurrence procedure absolute and relative bounds of the sweeping factors, in terms of purely combinatorial quantities $\overline{\mathscr{T}}_n$ or $\overline{\mathscr{T}}^n_{i_1 i_2 i_3}$. As we already mentioned in I, II, and the Introduction to the present paper, these integers represent the total number of possible different configurations $(i_1 i_2 i_3)$ inside the "tree"-like sum C_0^{n+1} (resp. $^\Phi C_0^{n+1}$) [the third global term of the equations of motion (H.0)] or inside the corresponding partial sums of it. They are explicitly calculated in Appendix B, and for sufficiently large n we could say that $\overline{\mathscr{T}}_n \sim n^2$.

We state these bounds by the following Propositions 2.3 and 2.4, which are statements equivalent to Lemma 2.3 of II or the second part of Lemma 2.6 of II, respectively.

Proposition 2.3: Let $\delta\in\overline{\mathscr{C}}_\Lambda$ and $\delta_{(1)}>\delta_{(2)}\in\overline{\mathscr{C}}^{\,c}_\Lambda$. Then, the sweeping factors β_n and $\beta^n_{i_1i_2i_3}$, for all partitions $w(I)$ of n, satisfy the following bounds: $\forall n\geqslant 7$ and $\forall\Lambda\in\mathbb{R}$ fixed in $0<\Lambda\leqslant 0.1$ (resp. $0<\Lambda\lesssim 0.01$),

$$\mathscr{T}_n\overline{Y}_n(\Lambda)\leqslant\beta_n\leqslant\mathscr{T}_nY_n(\Lambda),\tag{2.12}$$

$$\begin{aligned}&(\text{resp. } \beta_{n(2)}-\beta_{n(1)}\geqslant 0,\\ &\beta^n_{n-4,3,1(1)}/\beta^n_{n-4,3,1(2)}\geqslant\beta_{n-2(1)}/\beta_{n-2(2)},\\ &\delta_{n(1)}\beta_{n(1)}-\delta_{n(2)}\beta_{n(2)}\geqslant\delta_{n-2(1)}\beta_{n-2(1)}\\ &\qquad-\delta_{n-2(2)}\beta_{n-2(2)}\geqslant 0).\end{aligned}\tag{2.12a1}$$

The quantities $Y_n(\Lambda),\overline{Y}_n(\Lambda)$ are defined recurrently in the proof that follows and satisfy

(a) $0<\overline{Y}_n(\Lambda)\leqslant Y_n(\Lambda)\leqslant 1,$ (2.12a)

(b) $\exists Y_n^{(0)}>0,\ \lim_{\Lambda\to 0}Y_n(\Lambda)=\lim_{\Lambda\to 0}\overline{Y}_n(\Lambda)=Y_n^{(0)},$ (2.12b)

(c) $\exists Y_\infty(\Lambda)>0,\ \lim_{n\to\infty}Y_n(\Lambda)\geqslant Y_\infty(\Lambda),$ (2.12c)

(d) $\forall n\geqslant 9,$

$$\frac{Y_{n-2}}{\overline{Y}_n}\leqslant\begin{cases}[\mathscr{T}_n/\mathscr{T}_{n-2}](1-\Psi_n(\Lambda)), & \text{if } n\leqslant 4/3\Lambda+1\\ 1+\xi_n(\Lambda)/\mathscr{T}_n, & \text{if } n\geqslant 4/3\Lambda+1\end{cases}\tag{2.12d1}$$

and

$$\frac{\overline{Y}_{n-2}}{Y_n}\geqslant\begin{cases}(1-\overline{\Psi}_n(\Lambda))\mathscr{T}_n/\mathscr{T}_{n-2}, & \text{if } n\leqslant 4/3\Lambda+1,\\ 1+\overline{\xi}_n(\Lambda)/\mathscr{T}_n, & \text{if } n\geqslant 4/3\Lambda+1.\end{cases}\tag{2.12d2}$$

Here, $\quad 0\leqslant\Psi_n(\Lambda)\leqslant\overline{\Psi}_n(\Lambda)\leqslant 4/n,\quad 0\leqslant\overline{\xi}_n(\Lambda)\leqslant\xi_n(\Lambda)\leqslant 1.$ (2.12e)

Proof of Proposition 2.3: We show all the above properties except (2.12d1), which will be proved in Appendix D. We use two "nested" recursions: One principal recurrence (called H.β.1) that proceeds through increasing n and another secondary one for fixed n and increasing i_1, $[n/3]\leqslant i_1\leqslant n-2$ (called H.β.2). For the latter we shall, of course, always suppose the double ordering of the terms inside the sum C_0^{n+1}, $\forall n\geqslant 7$, following the relation $i_1\geqslant i_2\geqslant i_3$ for all triplets $(i_1i_2i_3)$ with $\Sigma_{l=1}^3 i_l=n$. Moreover, H.β.2 contains as a particular case H.β.2.b characterized by fixed $i_1(n)$ and increasing i_2: $[(n-i_1)/2]\leqslant i_2\leqslant n-i_1-1$.

Hypothesis H.β.1: We state H.β.1 as follows: $\forall\bar{n}\leqslant n-2$, the absolute bounds (2.12) hold together with the particular properties (2.12a)–(2.12d) for $Y_{\bar{n}}(\Lambda),\overline{Y}_n(\Lambda)$. The starting points of H.β.1 are given by the explicit bounds one obtains for

$$\beta_7=1+\tfrac{5}{6}\delta_3/\delta_5\tag{2.13}$$

and

$$\beta_9=1+14(1+(\delta_3/\delta_5)\tfrac{5}{36})\,\delta_3/\delta_7\beta_7.\tag{2.14}$$

(Notice that these formulas result from Proposition 2.1.) By application of the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$ one then verifies that

$$2\overline{Y}_7(\Lambda)\leqslant\beta_7(\Lambda)\leqslant 2Y_7(\Lambda),\tag{2.13a}$$

$$3\overline{Y}_9(\Lambda)\leqslant\beta_9(\Lambda)\leqslant 3Y_9(\Lambda),\tag{2.14a}$$

where

$$0<\overline{Y}_7(\Lambda)\leqslant Y_7(\Lambda)<1,\quad 0<\overline{Y}_9(\Lambda)\leqslant Y_9(\Lambda)<1,$$

with

$$\begin{aligned}&\lim_{\Lambda\to 0}\overline{Y}_7(\Lambda)=\lim_{\Lambda\to 0}Y_7(\Lambda)=\tfrac{2}{3},\\ &\lim_{\Lambda\to 0}\overline{Y}_9(\Lambda)=\lim_{\Lambda\to 0}Y_9(\Lambda)=0.4533,\end{aligned}\tag{2.15}$$

and

$$Y_7/\overline{Y}_9\leqslant\tfrac{3}{2}(1-\Psi_9(\Lambda)),\quad \overline{Y}_7/Y_9\geqslant\tfrac{3}{2}(1-\overline{\Psi}_9(\Lambda))\tag{2.15a}$$

with

$$\Psi_9(\Lambda)\leqslant\overline{\Psi}_9(\Lambda)\leqslant\tfrac{4}{9},\quad \lim_{\Lambda\to 0}\Psi_9(\Lambda)=\lim_{\Lambda\to 0}\overline{\Psi}_9(\Lambda)\approx 0.13.\tag{2.15b}$$

(In Ref. 4, Appendix β, we have presented in detail all these estimations and verifications, together with the explicit expressions of $Y_{\bar{n}},\overline{Y}_{\bar{n}}$, $\bar{n}=7,9$.) The above relations [(2.13a), (2.14a), (2.15a), and (2.15b)] ensure the validity of the recurrence hypothesis H.β.1 at the first step.

To prove H.β.1 for $\bar{n}=n$ we apply the secondary recurrence hypothesis H.β.2 (for $[n/3]\leqslant i_1\leqslant n-2$) as defined below [cf. (2.24)–(2.30)].

We put $\hat{i}_1\equiv i_{1\,\min}(n)=[n/3]$ for the minimal value of i_1 among the triplets $w(I)$ of the ordered sum C_0^{n+1}. The starting point of H.β.2 is obtained either for $i_1=\hat{i}_1$ and $i_2=\hat{i}_2=\hat{i}_1$ or for $i_1=\hat{i}_1+2$ and $i_2=\hat{i}_1-2$ or $i_2=\hat{i}_1$. Following formula (2.1) we have for the two cases either

$$\beta^n_{\hat{i}_1\hat{i}_2\hat{i}_2-4}=1+\frac{\delta_{\hat{i}_2-2}}{\delta_{\hat{i}_2}}\frac{\beta_{\hat{i}_2-2}}{\beta_{\hat{i}_2}}\frac{\hat{i}_2(\hat{i}_2-1)}{(\hat{i}_2-2)(\hat{i}_2-3)},\quad \text{if } i_2=\hat{i}_2=\hat{i}_1,\tag{2.16a}$$

or

$$\beta^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1}=1+\frac{\delta_{\hat{i}_1}}{\delta_{\hat{i}_1+2}}\frac{\beta_{\hat{i}_1}}{\beta_{\hat{i}_1+2}}\frac{(\hat{i}_1+1)(\hat{i}_1+2)}{\hat{i}_1(\hat{i}_1-1)},\quad \text{if } i_1=\hat{i}_1+2,\ i_2=\hat{i}_1-2,\ \text{or } i_2=\hat{i}_1.\tag{2.16b}$$

From the hypothesis $H_0\in\Phi_{0\Lambda}$ (or $\delta\in\overline{\mathscr{C}}_\Lambda$) we are allowed to use the corresponding relative bounds (cf. Definitions 2.a and 2.a.1 of II). Let $N_\Lambda\equiv 4/3\Lambda+1$. Then

$$\begin{aligned}&(1-\gamma_{\hat{i}_1+2}(\Lambda))^{-1}\frac{\hat{i}_1(\hat{i}_1-1)}{(\hat{i}_1+2)(\hat{i}_1+1)}\\ &\quad\leqslant\frac{\delta_{\hat{i}_1}}{\delta_{\hat{i}_1+2}}\leqslant\frac{\hat{i}_1(\hat{i}_1-1)}{(\hat{i}_1+2)(\hat{i}_1+1)}(1-\overline{\gamma}_{\hat{i}_1+2}(\Lambda))^{-1},\\ &\qquad\text{if } \hat{i}_1+2\leqslant N_\Lambda,\end{aligned}\tag{2.17a}$$

$$\left[1+\frac{\mu_{\hat{i}_1+2}}{\hat{i}_1(\hat{i}_1-1)}\right]^{-1}\leqslant\frac{\delta_{\hat{i}_1}}{\delta_{\hat{i}_1+2}}\leqslant\left[1+\frac{\overline{\mu}_{\hat{i}_1+2}}{\hat{i}_1(\hat{i}_1-1)}\right]^{-1},\quad \text{if } \hat{i}_1+2\geqslant N_\Lambda.\tag{2.17b}$$

Moreover, from the recurrence hypothesis H.β.1 we obtain the corresponding bounds for the ratios $\beta_{\hat{i}_1}/\beta_{\hat{i}_1+2}$. Insertion of the latter together with (2.17a) and (2.17b) into (2.16b) yields the following bounds for the second term on the rhs of (2.16b):

$$\bar{y}_{\hat{i}_1+2}(\Lambda) \leqslant \frac{\delta_{\hat{i}_1}}{\delta_{\hat{i}_1+2}} \frac{\beta_{\hat{i}_1}}{\beta_{\hat{i}_1+2}} \frac{(\hat{i}_1+1)(\hat{i}_1+2)}{\hat{i}_1(\hat{i}_1-1)} \leqslant y_{\hat{i}_1+2}(\Lambda), \tag{2.18}$$

where

$$\left.\begin{aligned} y_{\hat{i}_1+2} &\equiv (1-\bar{\gamma}_{\hat{i}_1+2})^{-1}(1-\Psi_{\hat{i}_1+2}),\\ \bar{y}_{\hat{i}_1+2} &\equiv (1-\gamma_{\hat{i}_1+2})^{-1}(1-\bar{\Psi}_{\hat{i}_1+2}), \end{aligned}\right\} \text{ if } \hat{i}_1+2 \leqslant N_\Lambda, \tag{2.18a}$$

or

$$\left.\begin{aligned} y_{\hat{i}_1+2} &\equiv \frac{(\hat{i}_1+2)(\hat{i}_1+1)}{\hat{i}_1(\hat{i}_1-1)} \frac{\bar{\mathscr{T}}_{\hat{i}_1}}{\bar{\mathscr{T}}_{\hat{i}_1+2}} \left(1+\frac{\xi_{\hat{i}_1+2}}{\bar{\mathscr{T}}_{\hat{i}_1+2}}\right)\left(1+\frac{\bar{\mu}_{\hat{i}_1+2}}{\hat{i}_1(\hat{i}_1-1)}\right)^{-1},\\ \bar{y}_{\hat{i}_1+2} &\equiv \frac{(\hat{i}_1+2)(\hat{i}_1+1)}{\hat{i}_1(\hat{i}_1-1)} \frac{\bar{\mathscr{T}}_{\hat{i}_1}}{\bar{\mathscr{T}}_{\hat{i}_1+2}} \left(1+\frac{\bar{\xi}_{\hat{i}_1+2}}{\bar{\mathscr{T}}_{\hat{i}_1+2}}\right)\left(1+\frac{\mu_{\hat{i}_1+2}}{\hat{i}_1(\hat{i}_1-1)}\right)^{-1}, \end{aligned}\right\} \text{ if } \hat{i}_1+2 \geqslant N_\Lambda. \tag{2.18b}$$

In both cases,

$$\lim_{\Lambda\to 0} y_{\hat{i}_1+2} = \lim_{\Lambda\to 0} \bar{y}_{\hat{i}_1+2} \equiv y^{(0)}_{\hat{i}_1+2} = 1+\Psi^{(0)}_{\hat{i}_1+2}. \tag{2.19}$$

Insertion of (2.18) into (2.16b) finally yields

$$2\bar{Y}^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1} \leqslant \beta^n_{\hat{i}_1+2,\hat{i}_1-2,\hat{i}_1} \leqslant 2Y^n_{\hat{i}_1+2,\hat{i}_1-2,\hat{i}_1}. \tag{2.20}$$

Here we have identified

$$Y^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1}(\Lambda) \equiv [1+y_{\hat{i}_1+2}(\Lambda)]/2 \tag{2.20a}$$

$$(\text{resp.}\ \bar{Y}^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1}(\Lambda) \equiv [1+\bar{y}_{\hat{i}_1+2}(\Lambda)]/2) \tag{2.20b}$$

with $y_{\hat{i}_1+2}(\Lambda), \bar{y}_{i_1+2}(\Lambda)$ defined by (2.18a) and (2.18b). One then verifies that, in view of (2.19), (2.20a), and (2.20b),

$$\lim_{\Lambda\to 0} Y^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1} = \lim_{\Lambda\to 0} \bar{Y}^n_{\hat{i}_1+2,\hat{i}_1-2,\hat{i}_1} = Y^{n(0)}_{\hat{i}_1+2,\hat{i}_1-2,\hat{i}_1} \equiv (1+y^{(0)}_{\hat{i}_1+2})/2. \tag{2.21}$$

Inequalities analogous to (2.20) can be obtained for the case (2.16a). We also note that bounds similar to those above hold for the corresponding "minimal" sweeping factor of $\bar{n} = n-2$, because trivially we have either

$$\beta^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1-1} = 1 \quad\text{or}\quad \beta^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1} = 1$$

so

$$Y^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1-2} = 1 \quad\text{or}\quad Y^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1} = 1. \tag{2.22}$$

One then finds

$$\frac{Y^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1-2}}{\bar{Y}^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1}} \leqslant \frac{2}{1+\bar{y}_{\hat{i}_1+2}} \stackrel{\text{def}}{\equiv} \begin{cases} 2(1-\Psi^n_{\hat{i}_1+2,\hat{i}_2-2}) \\ \text{or} \\ 1+\xi^n_{\hat{i}_1+2,\hat{i}_2-2}/2, \end{cases}$$

$$\frac{\bar{Y}^{n-2}_{\hat{i}_1,\hat{i}_1,\hat{i}_1-2}}{Y^n_{\hat{i}_1+2,\hat{i}_2-2,\hat{i}_1}} \geqslant \frac{2}{1+y_{\hat{i}_1+2}} \stackrel{\text{def}}{\equiv} \begin{cases} 2(1-\bar{\Psi}^n_{\hat{i}_1+2,\hat{i}_2-2}) \\ \text{or} \\ 1+\bar{\xi}^n_{\hat{i}_1+2,\hat{i}_2-2}/2, \end{cases} \tag{2.23}$$

with

$$\Psi^n_{\hat{i}_1+2,\hat{i}_2-2} = \frac{\bar{y}_{\hat{i}_1+2}}{1+\bar{y}_{\hat{i}_1+2}}, \quad \bar{\Psi}^n_{\hat{i}_1+2,\hat{i}_2-2} = \frac{y_{\hat{i}_1+2}}{1+y_{\hat{i}_1+2}},$$

$$\lim_{\Lambda\to 0} \Psi^n_{\hat{i}_1+2,\hat{i}_2-2} = \lim_{\Lambda\to 0} \bar{\Psi}^n_{\hat{i}_1+2,\hat{i}_2-2} = \frac{y^{(0)}_{\hat{i}_1+2}}{1+y^{(0)}_{\hat{i}_1+2}} \tag{2.23a}$$

or

$$\xi^n_{\hat{i}_1+2,\hat{i}_2-2} = 2\frac{1-y_{\hat{i}_1+2}}{1+y_{\hat{i}_1+2}},$$

$$\bar{\xi}^n_{\hat{i}_1+2,\hat{i}_2-2} = 2\frac{1-y_{\hat{i}_1+2}}{1+\bar{y}_{\hat{i}_1+2}}. \tag{2.23b}$$

These results allow us to state the following secondary (n fixed) recurrence hypothesis.

Hypothesis H.β.2: (i) For every $\bar{i}_1 \leqslant i_1-2$ (n fixed) and $\bar{i}_2$ ($\bar{i}_1$ fixed) $\leqslant i_2-2$, we suppose

$$\text{H.}\beta\text{.2a}\quad \bar{Y}^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\mathscr{T}^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \leqslant \beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \leqslant \mathscr{T}^n_{\bar{i}_1\bar{i}_2\bar{i}_3} Y^n_{\bar{i}_1\bar{i}_2\bar{i}_3},$$

$$\text{if } \bar{i}_3 \leqslant i_{3\min}(\bar{i}_1-2), \tag{2.24a}$$

and

$$\text{H.}\beta\text{.2.b}\quad \bar{Y}^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\mathscr{T}^{(n\bar{i}_1)}_{\bar{i}_2\bar{i}_3} \leqslant \beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3} \leqslant \mathscr{T}^{(n\bar{i}_1)}_{\bar{i}_2\bar{i}_3} Y^n_{\bar{i}_1\bar{i}_2\bar{i}_3},$$

$$\text{if } \bar{i}_3 > \bar{i}_{3\min}(\bar{i}_1-2). \tag{2.24b}$$

(Notice that H.β.2b corresponds to the case $\zeta(i_2i_3)=1$, $\zeta(i_1i_2)=0$ [cf. Proposition 2.1, definition (2.1c)], i.e., $\bar{i}_3 \leqslant \bar{i}_2-4$ and $\bar{i}_2 \neq n-i_1-1, \bar{i}_2 \neq i_1-4$.)

Here $\mathscr{T}^{(n\bar{i}_1)}_{\bar{i}_2\bar{i}_3}$ means the number of different terms (partitions) of the ordered sum defining C^{n+1}_0 for n and $\bar{i}_1$ fixed when i_2 varies inside the interval

$$\bar{i}_2 \leqslant i_2 \leqslant \bar{\bar{i}}_2, \quad \text{with } \bar{\bar{i}}_2 \equiv n - i_{3\max}(\bar{i}_1) - \bar{i}_1 = [(n-\bar{i}_1)/2] \tag{2.25}$$

i.e.,

$$\mathscr{T}^{(n\bar{i}_1)}_{\bar{i}_2\bar{i}_3} = (2\bar{i}_2+\bar{i}_1-n+2)/2, \tag{2.26}$$

and $\mathscr{T}^n_{\bar{i}_1\bar{i}_2\bar{i}_3}$ means the number of different terms (partitions) that correspond to the part of the ordered sum defining C^{n+1}_0 contained between the partition $w_N(\widehat{I}) = [\hat{i}_1,\hat{i}_2,i_{3\max}(\hat{i}_1)]$ and the partition $w_n(\bar{I}) = (\bar{i}_1\bar{i}_2\bar{i}_3)$. This number is given exactly by formulas (B19) and (B20) of Appendix B [for the case $\bar{i}_2 = \bar{i}_{2\max}(i_1)$]. [The reader can repeat the analogous arguments of Appendix B for every other case with $\bar{i}_2 < i_{2\max}(\bar{i}_1)$.]

The quantities $Y^n_{\bar{i}_1\bar{i}_2\bar{i}_3}(\Lambda)$ and $\bar{Y}^n_{\bar{i}_1\bar{i}_2\bar{i}_3}(\Lambda)$ are defined recurrently by the following relations:

$$Y^{n}_{\bar i_1\bar i_2\bar i_3}(\Lambda)=\frac{\mathcal{T}^{n-2}_{i_1-2,i_2,i_3}}{\mathcal{T}^{n}_{i_1i_2i_3}}y_{\bar i_1}\bar I^{n}_{\bar i_1-2,\bar i_2+2}\,Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}+\frac{(1-y_{\bar i_1}\bar I^{n}_{\bar i_1-2,\bar i_2+2})}{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}+\frac{\mathcal{T}^{(n-2,\bar i_1)}_{\bar i_2-2,\bar i_3+2}}{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}y_{\bar i_2}\bar I^{(n\bar i_1)}_{i_2-2,i_3+2}\,Y^{n-2}_{\bar i_1,\bar i_2-2,\bar i_3}-y_{\bar i_2}\frac{\bar I^{(ni_1)}_{\bar i_2-2,\bar i_3+2}}{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}},\quad \text{if } \bar i_3\leqslant \bar i_{3\,\min}(\bar i_1-2), \tag{2.27}$$

and

$$Y^{n}_{\bar i_1\bar i_2\bar i_3}(\Lambda)=\frac{\mathcal{T}^{(n-2,\bar i_1)}_{\bar i_2-2,\bar i_3+2}}{\mathcal{T}^{(n\bar i_1)}_{\bar i_2\bar i_3}}y_{\bar i_2}\bar I^{(n\bar i_1)}_{\bar i_2-2,\bar i_3+2}\,Y^{n-2}_{\bar i_1,\bar i_2-2,\bar i_3}+\frac{(1-y_{\bar i_2}\bar I^{(ni_1)}_{\bar i_2-2,\bar i_3+2})}{\mathcal{T}^{(ni_1)}_{i_2i_3}},\quad \text{if } \bar i_3> i_{3\,\min}(\bar i_1-2), \tag{2.28}$$

with

$$\bar I^{n}_{\bar i_1-2,\bar i_2+2}\ (\text{resp. } \bar I^{(n\bar i_1)}_{i_2-2,i_3+2})=\begin{cases}(1-\bar\Psi^{n}_{i_1-2,i_2+2})^{-1}[\text{resp. } (1-\bar\Psi^{(n\bar i_1)}_{i_2-2,i_3+2})^{-1}], & \text{if } i_1-2\leqslant N_\Lambda\ (\text{resp. } i_2-2\leqslant N_\Lambda),\\[4pt] \dfrac{\mathcal{T}^{n}_{\bar i_1-2,\bar i_2+2,\bar i_3}}{\mathcal{T}^{n-2}_{\bar i_1-4,\bar i_2+2,\bar i_3}}\left(1+\dfrac{\bar\xi^{n}_{\bar i_1-2,\bar i_2+2}}{\mathcal{T}^{n}_{\bar i_1-2,\bar i_2+2,i_3}}\right)^{-1}\left[\text{resp. } \dfrac{\mathcal{T}^{(n\bar i_1)}_{\bar i_2-2,\bar i_3+2}}{\mathcal{T}^{(n-2,\bar i_1)}_{\bar i_2-4,\bar i_3+2}}\left(1+\dfrac{\xi^{n\bar i_1}_{\bar i_2-2,\bar i_3+2}}{\mathcal{T}^{(ni_1)}_{\bar i_2-2,i_3+2}}\right)^{-1}\right] & i_1-2>N_\Lambda\ (\text{resp. } i_2-2>N_\Lambda).\end{cases} \tag{2.28a}$$

Here $y_{\bar i_1}$ (resp. $y_{\bar i_2}$) are defined by the corresponding relations (2.18a) and (2.18b). In both cases the quantities $y_{\bar i_1}, y_{\bar i_2}, Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}, \bar I^{n}_{\bar i_1-2,\bar i_2+2}$, etc., are replaced by the corresponding $\bar y_{\bar i_1}, \bar y_{\bar i_2}, \bar Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}, I^{n}_{\bar i_1-2,i_2+2}$, etc., to yield the analogous expressions for the factors $\bar y^{n}_{i_1\bar i_2\bar i_3}$ of the lower bounds in (2.24a) and (2.24b), respectively.

(ii) The following properties are satisfied by the $Y^{n}_{\bar i_1\bar i_2\bar i_3}$'s and their "nearest neighbors," the $Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}$'s:

$$0<\bar Y^{n}_{\bar i_1,\bar i_2,\bar i_3}\leqslant Y^{n}_{\bar i_1,\bar i_2,\bar i_3}\leqslant 1, \tag{2.29a}$$

$$\exists\, Y^{n(0)}_{\bar i_1\bar i_2\bar i_3}>0:\quad \lim_{\Lambda\to 0} Y^{n}_{\bar i_1\bar i_2\bar i_3}(\Lambda)=\lim_{\Lambda\to 0}\bar Y^{n}_{\bar i_1\bar i_2\bar i_3}(\Lambda)=Y^{n(0)}_{\bar i_1\bar i_2\bar i_3}. \tag{2.29b}$$

Moreover there exist positive finite constants

$$0<\Psi^{n}_{\bar i_1\bar i_2}(\Lambda)\leqslant\bar\Psi^{n}_{\bar i_1\bar i_2}\leqslant\frac{4}{\bar i_1},\quad 0<\Psi^{(ni_1)}_{\bar i_2\bar i_3}\leqslant\bar\Psi^{(n\bar i_1)}_{\bar i_2\bar i_3}\leqslant\frac{4}{i_2},\quad 0<\bar\xi^{n}_{\bar i_1\bar i_2}(\Lambda)\leqslant\xi^{n}_{\bar i_1\bar i_2}\leqslant\xi^{n}_{\bar i_1-2,\bar i_2+2}\leqslant 1,\quad 0<\bar\xi^{(ni_1)}_{\bar i_2\bar i_3}\leqslant\xi^{(ni_1)}_{\bar i_2 i_3}\leqslant\xi^{(ni_1)}_{\bar i_2-2,\bar i_3+2}\leqslant 1, \tag{2.30}$$

such that

$$\lim_{\Lambda\to 0}\Psi^{n}_{\bar i_1\bar i_2}(\Lambda)=\lim_{\Lambda\to 0}\bar\Psi_{i_1i_2}(\Lambda)\equiv\Psi^{n(0)}_{\bar i_1\bar i_2}>0,\quad \lim_{\Lambda\to 0}\Psi^{(ni_1)}_{i_2i_3}=\lim_{\Lambda\to 0}\bar\Psi^{(ni_1)}_{i_2i_3}\equiv\Psi^{(n_1i_1)(0)}_{i_2i_3}>0, \tag{2.30a}$$

and

$$\frac{Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}}{\bar Y^{n}_{\bar i_1\bar i_2\bar i_3}}\leqslant\begin{cases}\dfrac{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}{\mathcal{T}^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}}(1-\Psi^{n}_{\bar i_1\bar i_2}), & \text{if } \bar i_1\leqslant N_\Lambda \quad \left(\begin{array}{l}\text{resp. [according to the two cases (2.24a) and (2.24b)]}\\ \leqslant\dfrac{\mathcal{T}^{(n\bar i_1)}_{\bar i_2\bar i_3}}{\mathcal{T}^{(n-2,\bar i-2)}_{\bar i_2\bar i_3}}(1-\Psi^{(n\bar i_1)}_{\bar i_2\bar i_3}), \text{ if } \bar i_2\leqslant N_\Lambda\end{array}\right.,\\[6pt] 1+\dfrac{\xi^{n}_{\bar i_1\bar i_2}}{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}, & \text{if } \bar i_1\geqslant N_\Lambda\quad \left(\text{resp.}\leqslant 1+\dfrac{\xi^{(ni_1)}_{\bar i_2\bar i_3}}{\mathcal{T}^{(ni_1)}_{\bar i_2\bar i_3}},\ \text{if } \bar i_2\geqslant N_\Lambda\right),\end{cases} \tag{2.30b}$$

$$\frac{\bar Y^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}}{Y^{n}_{\bar i_1,\bar i_2,\bar i_3}}\geqslant\begin{cases}\dfrac{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}{\mathcal{T}^{n-2}_{\bar i_1-2,\bar i_2,\bar i_3}}(1-\bar\Psi^{n}_{\bar i_1\bar i_2}), & \text{if } \bar i_1\leqslant N_\Lambda\quad \left(\text{resp.}\geqslant\dfrac{\mathcal{T}^{(n\bar i_1)}_{\bar i_2\bar i_3}}{\mathcal{T}^{(n-2,\bar i_1-2)}_{\bar i_2\bar i_3}}(1-\bar\Psi^{(n\bar i_1)}_{\bar i_2\bar i_3}),\ \text{if } \bar i_2\leqslant N_\Lambda\right),\\[6pt] 1+\dfrac{\bar\xi^{n}_{i_1i_2}}{\mathcal{T}^{n}_{\bar i_1\bar i_2\bar i_3}}, & \text{if } \bar i_1\geqslant N_\Lambda\quad \left(\text{resp.}\geqslant 1+\dfrac{\bar\xi^{(ni_1)}_{\bar i_2\bar i_3}}{\mathcal{T}^{(n-2,\bar i_1-2)}_{\bar i_2\bar i_3}},\ \text{if } \bar i_1\geqslant N_\Lambda\right).\end{cases} \tag{2.30c}$$

The set of relations and bounds (2.27)–(2.30) complete the definition (2.24a) and (2.24b) of the recurrence hypothesis (2.24a) and (2.24b).

Let us prove all the above properties of H.β.2 for the partition $w_n(I) = (i_1 i_2 i_3)$.

Following definition (2.1) of Proposition 2.1 we have in the most general case, when $\zeta(i_1 i_2) = 1 = \zeta(i_2 i_3)$ and $\sigma(i_1 i_2 i_3) = 1$,

$$\beta^{n}_{i_1 i_2 i_3} = 1 + \frac{i_2(i_2-1)}{(i_3+1)(i_3+2)} \frac{\delta_{i_3+2}}{\delta_{i_2}} \frac{\beta_{i_3+2}}{\beta_{i_2}} \beta^{n}_{i_1, i_2-2, i_3+2} + \frac{i_1(i_1-1)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1}} \frac{\beta_{i_2+2}}{\beta_{i_1}} \beta^{n}_{i_1-2, i_2+2, i_3}. \tag{2.31}$$

We write

$$\frac{i_1(i_1-1)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1}} \frac{\beta_{i_2+2}}{\beta_{i_1}} \beta^{n}_{i_1-2, i_2+2, i_3} = \frac{i_1(i_1-1)}{(i_1-2)(i_1-3)} \frac{\delta_{i_1-2}}{\delta_{i_1}} \frac{\beta_{i_1-2}}{\beta_{i_1}} \frac{\beta^{n}_{i_1-2, i_2+2, i_3}}{\beta^{n-2}_{i_1-4, i_2+2, i_3}} \times \beta^{n-2}_{i_1-4, i_2+2, i_3} \frac{(i_1-2)(i_1-3)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1-2}} \frac{\beta_{i_2+2}}{\beta_{i_1-2}}, \tag{2.32a}$$

with

$$\beta^{n-2}_{i_1-4, i_2+2, i_3} \frac{(i_1-2)(i_1-3)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1-2}} \frac{\beta_{i_2+2}}{\beta_{i_1-2}} \overset{\text{def}}{\equiv} \beta^{n-2}_{i_1-2, i_2, i_3} - 1, \tag{2.32a'}$$

or, respectively,

$$\frac{i_2(i_2-1)}{(i_3+1)(i_3+2)} \frac{\delta_{i_3+2}}{\delta_{i_2}} \frac{\beta_{i_3+2}}{\beta_{i_2}} \beta^{n}_{i_1, i_2-2, i_3+2} = \frac{i_2(i_2-1)}{(i_2-2)(i_2-3)} \frac{\delta_{i_2-2}}{\delta_{i_2}} \frac{\beta_{i_2-2}}{\beta_{i_2}} \frac{\beta^{n}_{i_1, i_2-2, i_3+2}}{\beta^{n-2}_{i_1, i_2-4, i_3+2}} \times \beta^{n-2}_{i_1, i_2-4, i_3+2} \frac{(i_2-2)(i_2-3)}{(i_3+1)(i_3+2)} \frac{\delta_{i_3+2}}{\delta_{i_2-2}} \frac{\beta_{i_3+2}}{\beta_{i_2-2}}, \tag{2.32b}$$

with

$$\beta^{n-2}_{i_1, i_2-4, i_3+2} \frac{(i_2-2)(i_2-3)\delta_{i_3+2}\beta_{i_3+2}}{(i_3+1)(i_3+2)\delta_{i_2-2}\beta_{i_2-2}} \overset{\text{def}}{\equiv} \beta^{n-2}_{i_1, i_2-2, i_3} - 1. \tag{2.32b'}$$

Then by an analogous argument as for (2.18), we obtain

$$\bar{y}_{i_1}(\Lambda) \leqslant \frac{i_1(i_1-1)}{(i_1-2)(i_1-3)} \frac{\delta_{i_1-2}}{\delta_{i_1}} \frac{\beta_{i_1-2}}{\beta_{i_1}} \leqslant y_{i_1}(\Lambda) \tag{2.33a}$$

and, respectively,

$$\bar{y}_{i_2}(\Lambda) \leqslant \frac{i_2(i_2-1)}{(i_2-2)(i_2-3)} \frac{\delta_{i_2-2}}{\delta_{i_2}} \frac{\beta_{i_2-2}}{\beta_{i_2}} \leqslant y_{i_2}(\Lambda), \tag{2.33b}$$

where $y_{i_1}(\Lambda)$, $y_{i_2}(\Lambda)$, $\bar{y}_{i_1}(\Lambda)$, and $\bar{y}_{i_2}(\Lambda)$ are defined by relations analogous to (2.18a) and (2.18b). Application of the above bounds (2.33a) into (2.32a) (rhs) together with the recurrence hypothesis (2.24a) and (2.24b) (for $\beta^{n}_{i_1-2, i_2+2, i_3}/\beta^{n-2}_{i_1-4, i_2+2, i_3}$ and $\beta^{n-2}_{i_1-2, i_2, i_3}$) yields

$$\begin{aligned} \frac{i_1(i_1-1)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1}} \frac{\beta_{i_2+2}}{\beta_{i_1}} \beta^{n}_{i_1-2, i_2+2, i_3} &\leqslant y_{i_1} \frac{\bar{\mathcal{T}}^{n}_{i_1-2, i_2+2, i_3}}{\bar{\mathcal{T}}^{n-2}_{i_1-4, i_2+2, i_3}} \frac{Y^{n}_{i_1-2, i_2+2, i_3}}{\bar{Y}^{n-2}_{i_1-4, i_2+2, i_3}} \times (\bar{\mathcal{T}}^{n-2}_{i_1-2, i_2, i_3} Y^{n-2}_{i_1-2, i_2, i_3} - 1) \\ &\geqslant \bar{y}_1 \frac{\bar{\mathcal{T}}^{n}_{i_1-2, i_2+2, i_3}}{\bar{\mathcal{T}}^{n-2}_{i_1-4, i_2+2, i_3}} \frac{\bar{Y}^{n}_{i_1-2, i_2+2, i_3}}{Y^{n-2}_{i_1-4, i_2+2, i_3}} \times (\bar{\mathcal{T}}^{n-2}_{i_1-2, i_2, i_3} \bar{Y}^{n-2}_{i_1-2, i_2, i_3} - 1) \end{aligned} \tag{2.34}$$

and an expression analogous to the rhs of (2.32b) by applying (2.33b). Taking into account these results, inside the rhs of (2.31) we finally obtain, by (2.24a) and (2.24b),

$$\begin{aligned} \beta^{n}_{i_1 i_2 i_3} \leqslant{} & \bar{\mathcal{T}}^{n-2}_{i_1-2, i_2, i_3} Y^{n-2}_{i_1-2, i_2, i_3} y_{i_1} \bar{I}^{n}_{i_1-2, i_2+2} + (1 - y_{i_1} \bar{I}^{n}_{i_1-2, i_2+2}) \\ & + \bar{\mathcal{T}}^{(n-2, i_1)}_{i_2-2, i_3+2} y_{i_2} \bar{I}^{(n i_1)}_{i_2-2, i_3+2} Y^{n-2}_{i_1, i_2-2, i_3+2} - y_{i_2} \bar{I}^{(n i_1)}_{i_2-2, i_3+2} \overset{\text{def}}{\equiv} \bar{\mathcal{T}}^{n}_{i_1 i_2 i_3} Y^{n}_{i_1 i_2 i_3} \end{aligned} \tag{2.35}$$

and

$$\begin{aligned} \beta^{n}_{i_1 i_2 i_3} \geqslant{} & \bar{\mathcal{T}}^{n-2}_{i_1-2, i_2, i_3} \bar{Y}^{n-2}_{i_1-2, i_2, i_3} \bar{y}_{i_1} I^{n}_{i_1-2, i_2+2} + (1 - \bar{y}_{i_1} I^{n}_{i_1-2, i_3+2}) + \bar{\mathcal{T}}^{(n-2, i_1)}_{i_2-2, i_3+2} \\ & \times \bar{y}_{i_2} I^{(n i_1)}_{i_2-2, i_3+2} \bar{Y}^{n-2}_{i_1, i_2-2, i_3+2} - \bar{y}_{i_1} I^{(n i_1)}_{i_2-2, i_3+2} \overset{\text{def}}{\equiv} \bar{\mathcal{T}}^{n}_{i_1 i_2 i_3} \bar{Y}^{n}_{i_1 i_2 i_3} \\ & [\text{for } I^{n}_{i_1-2, i_2+2},\ I^{(n i_1)}_{i_2-2, i_2+2}, \text{ etc., of (2.28a)}]. \end{aligned} \tag{2.36}$$

The rhs identifications of (2.35) [resp. (2.36)] yield the exact recurrent formulas (2.27) of $Y^{n}_{i_1 i_2 i_3}$ (resp. the corresponding formula for $\bar{Y}^{n}_{i_1 i_2 i_3}$). Q.E.D.

The bounds (2.35) and (2.36) yield the proof of the first properties of the recurrence hypothesis H.β.2 for the partition $w_n(I) = (i_1 i_2 i_3)$. We only notice that we have supposed $i_3 < i_{3\,\max}(i_1)$ [$\zeta(i_2 i_3) = 1$] and $i_3 = i_{3\,\min}(i_1 - 2)$ [$\zeta(i_1 i_2) = 1$] so (2.24a) is, in fact, obtained. If $i_3 > i_{3\,\min}(i_1 - 2)$ [$\zeta(i_1 i_2) = 0$] then (2.24b) and (2.28) are obtained because the corresponding (2.35) and (2.36) do not contain the first terms.

Let us now complete the proof of the secondary recurrence hypothesis. We consider the proved expression $Y^{n}_{i_1 i_2 i_3}$ [see (2.35)] (resp. $\bar{Y}^{n}_{i_1 i_2 i_3}$ [cf. (2.36)]).

By application of the recurrence hypothesis H.β.2 for $Y^{n-2}_{i_1-2, i_2, i_3}$ (resp. $\bar{Y}^{n-2}_{i_1-2, i_2, i_3}$) and $Y^{n-2}_{i_1, i_2-2, i_3}$ (resp. $\bar{Y}^{n-2}_{i_1, i_2-2, i_3}$) together with (2.18a) and (2.18b) for $y_{i_1}, y_{i_2}, \bar{y}_{i_1}, \bar{y}_{i_2}$, and the fact that

$$\bar{\mathcal{T}}^{n}_{i_1-2, i_2+2, i_3} + \bar{\mathcal{T}}^{(n i_1)}_{i_2-2, i_3+2} + 1 = \bar{\mathcal{T}}^{n}_{i_1 i_2 i_3}, \tag{2.37}$$

one directly verifies (2.29a) and (2.29b)

$$0 < \bar{Y}^{n}_{i_1 i_2 i_3}(\Lambda) \leqslant Y^{n}_{i_1 i_2 i_3}(\Lambda) \leqslant 1, \tag{2.38}$$

$$\lim_{\Lambda \to 0} Y^{n}_{i_1 i_2 i_3}(\Lambda) = \lim_{\Lambda \to 0} \bar{Y}^{n}_{i_1 i_2 i_3}(\Lambda) = Y^{n(0)}_{i_1 i_2 i_3} > 0, \tag{2.39}$$

with

$$Y^{n(0)}_{i_1i_2i_3} \equiv \frac{\mathscr{T}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^{n}_{i_1i_2i_3}} y^{(0)}_{i_1} I^{n(0)}_{i_1-2,i_2+2} Y^{n-2(0)}_{i_1-2,i_2,i_3} + \frac{(1-y^{(0)}_{i_1} I^{n(0)}_{i_1-2,i_2+2})}{\overline{\mathscr{T}}^{n}_{i_1i_2i_3}} + \frac{\mathscr{T}^{(n-2,i_1)}_{i_2-2,i_3+2}}{\overline{\mathscr{T}}^{n}_{i_1i_2i_3}} y^{(0)}_{i_2} I^{(ni_1)(0)}_{i_2-2,i_3+2} Y^{n-2(0)}_{i_1,i_2-2,i_3} - \frac{y^{(0)}_{i_2} I^{(ni_1)(0)}_{i_2-2,i_3+2}}{\overline{\mathscr{T}}^{n}_{i_1i_2i_3}}. \tag{2.40}$$

For the proof of (2.30) we take into account the analogs of formula (2.27) corresponding to $Y^n_{i_1i_2i_3}$ and $Y^{n-2}_{i_1-2,i_2,i_3}$ (resp. to $\overline{Y}^n_{i_1i_2i_3}$ and $\overline{Y}^{n-2}_{i_1-2,i_2,i_3}$) and subtract them. Using the recurrence hypothesis of (2.30), for $Y^{n-4}_{i_1-4,i_2,i_3}/\overline{Y}^{n-2}_{i_1-2,i_2,i_3}$ and $Y^{n-4}_{i_1,i_2-4,i_3}/\overline{Y}^{n-2}_{i_1,i_2-2,i_3}$ (resp. $\overline{Y}^{n-4}_{i_1-4,i_2,i_3}/Y^{n-2}_{i_1-2,i_2,i_3}$ and $\overline{Y}^{n-4}_{i_1,i_2-4,i_3}/Y^{n-2}_{i_1,i_2-2,i_3}$), we finally obtain

$$Y^{n-2}_{i_1-2,i_2,i_3}(\Lambda) - \overline{Y}^n_{i_1i_2i_3}(\Lambda) \leqslant \overline{Y}^n_{i_1i_2i_3}(\Lambda) \begin{cases} \dfrac{\mathscr{T}^n_{i_1i_2i_3} - \mathscr{T}^{n-2}_{i_1-2,i_2,i_3}}{\mathscr{T}^{n-2}_{i_1-2,i_2,i_3}} - \Psi^n_{i_1i_2}(\Lambda)\dfrac{\mathscr{T}^n_{i_1i_2i_3}}{\mathscr{T}^{n-2}_{i_1-2,i_2,i_3}}, & \text{if } i_1 \leqslant N_\Lambda \\ \left(\text{resp}.\dfrac{\mathscr{T}^{(ni_1)}_{i_2i_3} - \mathscr{T}^{(n-2,i_1-2)}_{i_2i_3}}{\mathscr{T}^{(n-2,i_1-2)}_{i_2i_3}} - \Psi^{(ni_1)}_{i_2i_3}(\Lambda)\dfrac{\mathscr{T}^{(ni_1)}_{i_2i_3}}{\mathscr{T}^{(n-2,i_1-2)}_{i_2i_3}}, \quad \text{if } i_2 \leqslant N_\Lambda\right), \\ \text{or} \\ \xi^n_{i_1i_2}/\mathscr{T}^n_{i_1i_2i_3}, \quad \text{if } i_1 > N_\Lambda \\ (\text{resp. } \xi^{(ni_1)}_{i_2i_3}/\mathscr{T}^{(ni_1)}_{i_2i_3}, \quad \text{if } i_2 > N_\Lambda) \end{cases} \tag{2.41a}$$

and, respectively,

$$\overline{Y}^{n-2}_{i_1-2,i_2,i_3}(\Lambda) - Y^n_{i_1i_2i_3}(\Lambda) \geqslant Y^n_{i_1i_2i_3}(\Lambda) \begin{cases} \dfrac{\overline{\mathscr{T}}^n_{i_1i_2i_3} - \overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}} - \overline{\Psi}^n_{i_1i_2}\dfrac{\overline{\mathscr{T}}^n_{i_1i_2i_3}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}, & \text{if } i_1 \leqslant N_\Lambda \\ \left(\text{resp. }\dfrac{\overline{\mathscr{T}}^{(ni_1)}_{i_2i_3} - \overline{\mathscr{T}}^{(n-2,i_1-2)}_{i_2i_3}}{\overline{\mathscr{T}}^{(n-2,i_1-2)}_{i_2i_3}} - \overline{\Psi}^{(ni_1)}_{i_2i_3}\dfrac{\overline{\mathscr{T}}^{(ni_1)}_{i_2i_3}}{\overline{\mathscr{T}}^{(n-2,i_1-2)}_{i_2i_3}}, \quad \text{if } i_2 \leqslant N_\Lambda\right) \\ \text{or} \\ \overline{\xi}^n_{i_1i_2}/\overline{\mathscr{T}}^n_{i_1i_2i_3}, \quad \text{if } i_1 > N_\Lambda \\ (\text{resp. } \overline{\xi}^{(ni_1)}_{i_2i_3}/\overline{\mathscr{T}}^{(ni_1)}_{i_2i_3}, \quad \text{if } i_2 > N_\Lambda). \end{cases} \tag{2.41b}$$

Here the quantities $\Psi^n_{i_1i_2}$, $\Psi^{(ni_1)}_{i_2i_3}$ (resp. $\overline{\Psi}^n_{i_1i_2}$, $\overline{\Psi}^{(ni_1)}_{i_2i_3}$) and $\xi^n_{i_1i_2}$, $\xi^{(ni_1)}_{i_2i_3}$ (resp. $\overline{\xi}^n_{i_1i_2}$, $\overline{\xi}^{(ni_1)}_{i_2i_3}$) are defined recurrently (in terms of Y^n's and $\Psi^n_{i_1-2,i_2+2}$, $\xi^n_{i_1-2,i_2+2}$, etc.) by the following (complicated, but easily obtained) formulas:

$$\Psi^n_{i_1i_2}(\Lambda) \overset{\text{def}}{\equiv} \frac{\mathscr{T}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}\overline{Y}^n_{i_1i_2i_3}}\left[\mathscr{D}^{(ni_1)}_{(Y)} + \frac{\overline{\mathscr{T}}^n_{i_1i_2i_3} - \overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}\overline{Y}^n_{i_1i_2i_3} - \mathscr{D}^{(ni_1)}_{(\mathscr{T})} - \mathscr{D}^{(ni_1)}_{(R)} - \mathscr{D}^{(ni_1)}_{(y)} - \mathscr{D}^{(ni_1)}_{(I)}\right], \tag{2.42}$$

$$\xi^n_{i_1i_2}(\Lambda) \overset{\text{def}}{\equiv} \frac{\overline{\mathscr{T}}^n_{i_1i_2i_3}}{\overline{Y}^n_{i_1i_2i_3}}\left[\mathscr{D}^{(ni_1)}_{(Y)} + \mathscr{D}^{(ni_1)}_{(\mathscr{T})} + \mathscr{D}^{(ni_1)}_{(R)} + \mathscr{D}^{(ni_1)}_{(y)} + \mathscr{D}^{(ni_1)}_{(I)}\right]. \tag{2.43}$$

Here the following abbreviated notations have been used:

$$\begin{aligned} \mathscr{D}^{(ni_1)}_{(Y)} \overset{\text{def}}{\equiv} {} & \overline{y}_{i_1}\frac{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}_{i_1i_2i_3}} I^n_{i_1-2,i_2+2}\overline{Y}^{n-2}_{i_1-2,i_2,i_3}\begin{cases} \Psi^{n-2}_{i_1-2,i_2}\dfrac{\mathscr{T}^{n-2}_{i_1-2,i_2,i_3}}{\mathscr{T}^{n-4}_{i_1-4,i_2,i_3}} - \dfrac{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3} - \overline{\mathscr{T}}^{n-4}_{i_1-4,i_2,i_3}}{\overline{\mathscr{T}}^{n-4}_{i_1-4,i_2,i_3}}, & \text{if } i_1-2 \leqslant N_\Lambda, \\ \dfrac{\xi^{n-2}_{i_1-2,i_2}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}, & \text{if } i_1-2 \geqslant N_\Lambda, \end{cases} \\ & + \overline{y}_{i_2}\frac{\mathscr{T}^{(n-2,i_1)}_{i_2-2,i_3}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}}\overline{Y}^{n-2}_{i_1,i_2-2,i_3} I^{(ni_1)}_{i_2-2,i_3+2}\begin{cases} \Psi^{(n-2,i_1)}_{i_2-2,i_3}\dfrac{\mathscr{T}^{(n-2,i_1)}_{i_2-2,i_3}}{\mathscr{T}^{(n-4,i_1)}_{i_2-4,i_3}} - \dfrac{\overline{\mathscr{T}}^{(n-2,i_1)}_{i_2-2,i_3} - \overline{\mathscr{T}}^{(n-4,i_1)}_{i_2-4,i_3}}{\overline{\mathscr{T}}^{(n-4,i_1)}_{i_2-4,i_3}}, & \text{if } i_2-2 \leqslant N_\Lambda, \\ \dfrac{\xi^{(n-2,i_1)}_{i_2-2,i_3}}{\mathscr{T}^{(n-2,i_1)}_{i_2-2,i_3}}, & \text{if } i_2-2 \geqslant N_\Lambda, \end{cases} \end{aligned} \tag{2.44a}$$

$$\mathscr{D}^{(ni_1)}_{(\mathscr{T})} \overset{\text{def}}{\equiv} \frac{1-y_{i_1-2}\overline{I}^{n-2}_{i_1-4,i_2+2}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}} - \frac{1-\overline{y}_{i_1}I^n_{i_1-2,i_2+2}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}} + \frac{\overline{y}_{i_2}I^{(ni_1)}_{i_2-2,i_3+2}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}} - \frac{y_{i_2-2}\overline{I}^{(n+2,i_1)}_{i_2-4,i_3+2}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}, \tag{2.44b}$$

$$\mathscr{D}^{(ni_1)}_{(y)} \overset{\text{def}}{\equiv} (y_{i_1-2} - \overline{y}_{i_1})\frac{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}} Y^{n-4}_{i_1+4,i_2,i_3}\overline{I}^{n-2}_{i_1-4,i_2+2} + (y_{i_2-2} - \overline{y}_{i_2})\frac{\overline{\mathscr{T}}^{(n-2,i_1)}_{i_2-2,i_3}}{\overline{\mathscr{T}}^n_{i_1i_2i_3}}\overline{I}^{(n-2,i_1)}_{i_2-4,i_3+2}Y^{n-4}_{i_1,i_2-4,i_3}, \tag{2.44c}$$

$$\mathscr{D}^{(ni_1)}_{(R)} \overset{\text{def}}{\equiv} y_{i_1-2}\left(\frac{\overline{\mathscr{T}}^{n-4}_{i_1-4,i_2,i_3}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}} - \frac{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^{n}_{i_1 i_2 i_3}}\right) Y^{n-4}_{i_1-4,i_2,i_3} \bar{I}^{n-2}_{i_1-4,i_2+2}$$

$$+ y_{i_2-2}\left\{\frac{\overline{\mathscr{T}}^{(n-4,i_1)}_{i_2-4,i_3}}{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}} - \frac{\overline{\mathscr{T}}^{(n-2,i_1)}_{i_2-2,i_3}}{\overline{\mathscr{T}}^{n}_{i_1 i_2 i_3}}\right\} Y^{n-4}_{i_1,i_2-4,i_3} \bar{I}^{(n-2,i_1)}_{i_2-4,i_3+2}, \tag{2.44d}$$

$$\mathscr{D}^{(ni_1)}_{(I)} \overset{\text{def}}{\equiv} (\bar{I}^{n-2}_{i_1-4,i_2+2} - I^{n}_{i_1-2,i_2+2}) \frac{\overline{\mathscr{T}}^{n-2}_{i_1-2,i_2,i_3}}{\overline{\mathscr{T}}^{n}_{i_1 i_2 i_3}} Y^{n-4}_{i_1-4,i_2,i_3} \bar{y}_{i_1}$$

$$+ (\bar{I}^{(n-2,i_1)}_{i_2-4,i_3+2} - I^{(ni_1)}_{i_2-2,i_3+2}) y_{i_2} \frac{\overline{\mathscr{T}}^{(n-2,i_1)}_{i_2-2,i_3}}{\overline{\mathscr{T}}^{n}_{i_1 i_2 i_3}} Y^{n-4}_{i_1,i_2-4,i_3}. \tag{2.44e}$$

Qualitatively, the contributions (2.44a)–(2.44e) correspond, for $\mathscr{D}^{(ni_1)}_{(\mathscr{T})}$ [(2.44b)], to the differences of the inverse numbers $[\overline{\mathscr{T}}^{(n)}]^{-1}$ of partitions for $n,n-2$ (the smallest contribution in the sum of the order $1/n^3$); for $\mathscr{D}^{(ni_1)}_{(Y)}$ [(2.44a)], to the differences in $Y^{n-4},\overline{Y}^{n-2}$ at the preceding step (recurrence hypothesis and the largest contribution); for $\mathscr{D}^{(ni_1)}_{(R)}$ [(2.44d)], to the differences of the ratios $\mathscr{T}^{n-4}_{i_1}/\mathscr{T}^{n-2}_{i_1}$ (order $\sim 1/i_1^2$); and for $\mathscr{D}^{(ni_1)}_{(y)} + \mathscr{D}^{(ni_1)}_{(I)}$, to the difference of the products $y_{i_1} I^{n}_{i_1-2,i_2+2}$ (small contribution of order $\sim 1/n^2$).

Analogous recurrent definitions hold for $\Psi^{n}_{i_1 i_2},\xi^{n}_{i_1 i_2}$ and $\Psi^{(ni_1)}_{i_2 i_3},\xi^{(ni_1)}_{i_2 i_3}$, etc. To obtain them one only has to consider the following transformations inside (2.42), (2.43), and (2.44a)–(2.44e):

$$\begin{aligned}
&\overline{Y}^{n}_{i_1 i_2 i_3} \to Y^{n}_{i_1 i_2 i_3}, \quad \bar{y}_{i_1} \to y_{i_1}, \quad \bar{y}_{i_2} \to y_{i_2},\\
&\overline{Y}^{n-2}_{i_1-2,i_2,i_3} \to Y^{n-2}_{i_1-2,i_2,i_3}, \quad \Psi^{n}_{i_1-2,i_2} \to \overline{\Psi}^{n}_{i_1-2,i_2}, \quad \xi^{n}_{i_1-2,i_2} \to \bar{\xi}^{n}_{i_1-2,i_2},\\
&\overline{Y}^{n}_{i_1,i_2-2,i_3} \to Y^{n}_{i_1,i_2-2,i_3}, \quad \Psi^{(n-2,i_1)}_{i_2-2,i_3} \to \overline{\Psi}^{(n-2,i_1)}_{i_2-2,i_3}, \quad \xi^{(n-2,i_1)}_{i_2-2,i_3} \to \bar{\xi}^{(n-2,i_1)}_{i_2-2,i_3}.
\end{aligned} \tag{2.45}$$

Despite the complicated expressions (2.42) and (2.44), we can, by using properties (2.29a), (2.29b), (2.30), and (2.30c) of the recurrence hypothesis, for $\Psi^{n-2}_{i_1-2,i_2}$, $(\overline{\Psi}^{n-2}_{i_1-2,i_2})$, $\Psi^{(n-2,i_1)}_{i_2-2,i_3}$, $(\overline{\Psi}^{(n-2,i_1)}_{i_2-2,i_3})$, $\xi^{n-2}_{i_1-2,i_2}$, $(\bar{\xi}^{n-2}_{i_1-2,i_2})$, $\xi^{(n-2,i_1)}_{i_2-2,i_3}$, $(\bar{\xi}^{(n-2,i_1)}_{i_2-2,i_3})$, together with the combinatorial property (2.37), prove (when $0<\Lambda \lesssim 0.1$)

$$\begin{aligned}
&0<\Psi^{n}_{i_1 i_2} \leqslant \overline{\Psi}^{n}_{i_1 i_2} \leqslant 4/i_1, \quad 0<\Psi^{(ni_1)}_{i_2 i_3} \leqslant \overline{\Psi}^{(ni_1)}_{i_2 i_3} \leqslant 4/i_2,\\
&0 \leqslant \bar{\xi}^{n}_{i_1 i_2} \leqslant \xi^{n}_{i_1 i_2} \leqslant \xi^{n}_{i_1-2,i_2+2} \leqslant 1,\\
&0 \leqslant \bar{\xi}^{(ni_1)}_{i_2 i_3} \leqslant \xi^{ni_1}_{i_2 i_3} \leqslant \xi^{(ni_1)}_{i_2-2,i_3+2} \leqslant 1,\\
&\lim_{\Lambda \to 0} \Psi^{n}_{i_1 i_2}(\Lambda) = \lim_{\Lambda \to 0} \overline{\Psi}^{n}_{i_1 i_2}(\Lambda) = \Psi^{n(0)}_{i_1 i_2},\\
&\lim_{\Lambda \to 0} \Psi^{(ni_1)}_{i_2 i_3}(\Lambda) = \lim_{\Lambda \to 0} \overline{\Psi}^{(ni_1)}_{i_2 i_3}(\Lambda) = \Psi^{(ni_1)(0)}_{i_2 i_3}.
\end{aligned} \tag{2.46}$$

From (2.41a), (2.41b), (2.42), (2.43), (2.44), and (2.46) we have the complete proof of the recurrence hypothesis H.β.2. Q.E.D.

Now, application of H.β.2 for $i_1 = n-2$ yields the proof of H.β.1 for $\bar{n} = n$, i.e.,

$$\overline{\mathscr{T}}_n \overline{Y}_n \leqslant \beta_n \leqslant \overline{\mathscr{T}}_n Y_n, \tag{2.47}$$

where we have identified

$$Y_n \equiv Y^{n}_{n-2,1,1} \tag{2.47a}$$

(resp.

$$\overline{Y}_n \equiv \overline{Y}^{n}_{n-2,1,1}) \tag{2.47b}$$

with the corresponding recurrent definitions (2.27) and (2.28) and properties (2.12), (2.12a), (2.12b), and (2.12d) satisfied [see properties (2.29) and (2.30) of the recurrent hypothesis]; notice also that we identify

$$\begin{aligned}
&\Psi^{n}_{n-2,1}(\Lambda) \equiv \Psi_n(\Lambda), \quad \overline{\Psi}^{n}_{n-2,1}(\Lambda) \equiv \overline{\Psi}_n(\Lambda),\\
&\xi^{n}_{n-2,1} \equiv \xi_n(\Lambda), \quad \bar{\xi}^{n}_{n-2,1} \equiv \bar{\xi}_n(\Lambda).
\end{aligned} \tag{2.47c}$$

To prove the last property of H.β.1, that is, the lower bound (2.12c) of the limit at infinity ($n \to \infty$) of $Y_n(\Lambda)$, we write, in view of the bounds (2.12d1) and Appendix B for $\overline{\mathscr{T}}_n$ when $n \geqslant 15$, i.e., $n \geqslant N_\Lambda$,

$$\lim_{n \to \infty} Y_n(\Lambda) \geqslant \operatorname*{Inf}_{\delta \in \overline{\mathscr{C}}_\Lambda}\left(\frac{\beta_n}{\overline{\mathscr{T}}_{N_\Lambda}}\right) \prod_{\bar{n}=N_\Lambda+2}^{\infty}\left[1 - \frac{48}{(\bar{n}-3)^2}\right]$$

$$\text{(for } \Lambda \text{ fixed inside: } (0;0.1]). \tag{2.48}$$

Again we put $N_\Lambda \equiv 4/3\Lambda + 1$ for the characteristic value of n corresponding to the change of the increasing mode of the δ_n's in $\overline{\mathscr{C}}_\Lambda$, and $\overline{\mathscr{T}}_{N_\Lambda}$ is explicitly given in Appendix B.

The limit when $n \to \infty$ of the second product in formula (2.48) exists and it is a positive constant. So that $\exists\, Y_\infty(\Lambda) > 0$, where

$$\lim_{n \to \infty} Y_n(\Lambda) \geqslant Y_\infty(\Lambda) \cong (1{,}3\pi)/4\overline{\mathscr{T}}_{N_\Lambda},$$

$$\text{for } \Lambda \text{ fixed in } (0;0.1]. \tag{2.49}$$

Q.E.D.

We conclude that H.β.1 is proved and, in other words, the validity of Proposition 2.3 is ensured.

Following analogous lines we can show the corresponding bounds for the sweeping factors ${}^{\Phi}\beta_n$ in the dimensional case. In view of the above detailed presentation, we consider that it is not necessary to give the proof. We state this result by the following proposition, which is the equivalent of statement (ii) in Lemma 2.6 of II. We notice that these bounds are uniform, i.e., they do not depend on the particular coherent sequence of ΦC's, $\{\Phi^{(\bar{n},n)}\}_n \in \mathscr{F}$. [Recall that the notation $\mathscr{F} \equiv \mathscr{F}(\{\Phi^{(\bar{n},n)}\})$ has been introduced in I and II for the infinite family of coherent sequences of ΦC's.]

Proposition 2.4: Let $H \in \Phi_\Lambda$ (or $\delta \in \overline{\mathscr{C}}_\Lambda$). Then the following bounds hold for every element of $\mathscr{F}$:

$$\overline{\mathscr{T}}_n \overline{\overline{Y}}_n(\Lambda)/\underline{c}_n \leqslant {}^{\Phi}\beta_n \leqslant \overline{\mathscr{T}}_n \widetilde{Y}_n(\Lambda)/\overline{c}_n. \tag{2.50}$$

Here $\widetilde{Y}_n(\Lambda)$ and $\overline{\overline{Y}}_n(\Lambda)$ satisfy the analogs of limit values and bounds (2.12a)–(2.12d) of $Y_n(\Lambda)$ and $\overline{Y}_n(\Lambda)$, respectively, presented by Proposition 2.3. The positive constants $\overline{c}_n$ and $\underline{c}_n$ are the maximal and minimal values, respectively, of the parameter ${}^{\Phi}C^n_{n-2,1,1}$ taken inside $\mathscr{F}$ (cf. Proposition 2.2 of II).

III. RECURSIVE PROCEDURE AND BOUNDS FOR α_n's

We start this section by proving the sweeping procedure through the sum B_0^{n+1} defined by

$$B_0^{n+1} = -3\Lambda \sum_{w(J)} \theta^n_{j_1 j_2} H_0^{j_2+2} H_0^{j_1+1}, \quad H_0 \in \mathscr{B}_0$$

[cf. Definition 3(b) of I] in terms of the corresponding sweeping factors $\alpha^n_{j_1 j_2}$, for every partition $w(J)$ of n and the ratios of β_n and δ_n sequences. In II this procedure was presented in terms of the Green's functions H_0^{n+1} [cf. Definition 2(c) of II] and then stated without proof by explicit recursive formulas in terms of the splitting sequences (when $H_0 \in \Phi_{0\Lambda}$), in Lemma 2.4. Besides this recursive procedure, Proposition 3.1 below also yields the corresponding bounds and limit values of these sweeping factors $\alpha^n_{j_1 j_2}$ (cf. the second part of Lemma 2.5 of II).

Proposition 3.1: Let $H_0 \in \Phi_{0\Lambda}$, $\forall n \geqslant 5$, and for every fixed partition $w(J) = (j_1 j_2)$ with $[(n+1)/2] \leqslant j_2 \leqslant n-1$.

(i) The sweeping factors $\alpha_n, \alpha^n_{j_1 j_2}$ [cf. Definition 2(c) of II] are given recurrently as follows:

$$\alpha_n \equiv \alpha^n_{1,n-1}, \tag{3.1a}$$

$$\alpha^n_{(n+1)/2,(n-1)/2} = 1, \quad \text{if } (n+1)/2 \text{ odd}, \tag{3.1b}$$

$$\alpha^n_{(n-1)/2,(n+1)/2} = 1 + (n-1)/(n+3), \quad \text{if } (n+1)/2 \text{ even}, \tag{3.1c}$$

and

$$\alpha^n_{j_1 j_2} = 1 + \frac{j_2(j_2-1)}{(j_1+2)(j_1+1)} \frac{\delta_{j_1+2}}{\delta_{j_2+1}} \times \frac{z(\beta_{j_1+2})}{z(\beta_{j_2+1})} \alpha^n_{j_1+2, j_2-2} + \frac{j_1 s(j_1-1)}{j_2+1},$$

$$(n+1)/2 < j_2 \leqslant n-1, \tag{3.2}$$

with $z(\beta_i)$ as in (2.1a) and

$$s(j_1-1) = \begin{cases} 1, & \text{if } j_1 > 1, \\ 0, & \text{if } j_1 = 1. \end{cases} \tag{3.2a}$$

(ii) The following bounds hold:

$$\mathscr{T}_n \overline{X}_n(\Lambda) \leqslant \alpha_n \leqslant \mathscr{T}_n X_n(\Lambda), \tag{3.3}$$

with $\mathscr{T}_n = (n-1)/2$. Moreover

if $H_{(1)}, H_{(2)} \in \Phi_{0\Lambda}$ with $\delta_{(1)} > \delta_{(2)} \in \mathscr{C}'_\Lambda$

and

$$0 < \Lambda \lesssim 0.01, \quad \alpha_{n(2)} - \alpha_{n(1)} \geqslant 0. \tag{3.3a1}$$

The quantities $X_n(\Lambda)$ and $\overline{X}_n(\Lambda)$ are given recurrently in the proof below and they satisfy the following properties for all Λ fixed inside: (0;0.1]. Let $N_\Lambda = 4/3\Lambda + 1$. Then

(a) $0 < \overline{X}_n(\Lambda) \leqslant X_n(\Lambda) \leqslant 1;$ (3.3a)

(b) $\exists X_n^{(0)} > 0\colon \ \lim_{\Lambda \to 0} X_n(\Lambda) = \lim_{\Lambda \to 0} \overline{X}_n(\Lambda) = X_n^{(0)};$ (3.3b)

(c) $\exists X_\infty(\Lambda) > 0\colon \ \lim_{n \to \infty} X_n(\Lambda) = \lim_{n \to \infty} \overline{X}_n(\Lambda) \geqslant X_\infty(\Lambda);$ (3.3c)

(d) $\forall n \geqslant 7$,

$$\frac{X_{n-2}}{\overline{X}_n} \leqslant \begin{cases} [(n-1)/(n-3)](1-\eta_n(\Lambda)), & \text{with } 0 < \eta_n(\Lambda) \leqslant 2/n, \quad \text{if } n \leqslant N_\Lambda, \\ (1 + k_n/(\mathscr{T}_n)^2), & \text{with } 0 < k_n(\Lambda) \leqslant 1, \quad n \geqslant N_\Lambda, \end{cases} \tag{3.3d1}$$

$$\frac{\overline{X}_{n-2}}{X_n} \geqslant \begin{cases} [(n-1)/(n-3)](1-\overline{\eta}_n(\Lambda)^2), & \text{with } 0 < \overline{\eta}_n(\Lambda) \leqslant 2/n, \quad \text{if } n \leqslant N_\Lambda, \\ (1 + \overline{k}_n/(\mathscr{T}_n)^2), & \text{with } 0 < \overline{k}_n(\Lambda) \leqslant 1, \quad \text{if } n \geqslant N_\Lambda. \end{cases} \tag{3.3d2}$$

(e) $\exists n_0 > N_\Lambda\colon \ \forall n \geqslant n_0, \ \overline{X}_n^2 \geqslant Y_{n+2}.$ (3.3e)

Proof of Proposition 3.1: The proofs of properties (i) and (ii) for the sweeping factors $\alpha^n_{j_1 j_2}$ and α_n proceed by recursion and in a way exactly analogous to that of Propositions 2.1 and 2.3 for the corresponding sweeping factors $\beta_n, \beta^n_{i_1 i_2 i_3}$ [in particular, property (3.3d1) can be shown using arguments analogous to (2.12a1) (cf. Appendix D)]. We shall therefore mention only the most important steps. We suppose that $H_0 \in \Phi_{0\Lambda}$ (i.e., $\delta \in \overline{\mathscr{C}}_\Lambda$). Then, as in the proof of Proposition 2.3, we establish two nested recursions. One principal called H.α.1, for $3 \leqslant \overline{n} \leqslant n-2$ and one secondary called H.α.2 valid at fixed n and for $[(n+1)/2] \leqslant \overline{j}_2 \leqslant j_2 - 2$.

Hypothesis H.α.1: Following this recurrence hypothesis we suppose that the Proposition 3.1 holds for every $3 \leqslant \overline{n} \leqslant n-2$ and prove it for $\overline{n} = n$. The trivial starting point of this procedure is when $n = 3$, because

$$\alpha^3_{1,2} = 1. \tag{3.4}$$

The first nontrivial sweeping factors α_n correspond to $n=5,7$. In Ref. 4 we have calculated them explicitly and found

$$\alpha_5 \equiv \alpha^5_{1,4} = 1 + \tfrac{1}{2}\delta_3/\delta_5, \tag{3.5a}$$

$$\alpha_7 \equiv \alpha^7_{1,6} = 1 + 5(1+\tfrac{3}{5})\delta_3/\delta_7\beta_7. \tag{3.5b}$$

Moreover, in view of the hypothesis $\delta \in \overline{\mathscr{C}}_\Lambda$ we obtain bounds of them in the form

$$2\underline{X}_5(\Lambda) \leqslant \alpha_5 \leqslant 2\overline{X}_5(\Lambda), \tag{3.6a}$$

$$3\underline{X}_7(\Lambda) \leqslant \alpha_7 \leqslant 3\overline{X}_7(\Lambda), \tag{3.6b}$$

where $\overline{X}_5$, X_5, X_7, and $\overline{X}_7$ satisfy the corresponding properties (ii) of the proposition. To show H.α.1 for $\bar{n} = n$ we need the secondary one H.α.2, which is defined as follows.

Hypothesis H.α.2: For every partition $w(\overline{J}) \equiv (\bar{j}_1 \bar{j}_2)$ with $[(n+1)/2] \leqslant \bar{j}_2 \leqslant j_2 - 2$.

(a) The definitions (3.1) and (3.2) are valid.

(b) The following properties are satisfied:

$$\overline{X}^n_{\bar{j}_1\bar{j}_2} \mathscr{T}^n_{\bar{j}_1\bar{j}_2} \leqslant \alpha^n_{\bar{j}_1\bar{j}_2} \leqslant \mathscr{T}^n_{\bar{j}_1\bar{j}_2}, \quad \text{with } \mathscr{T}^n_{\bar{j}_1\bar{j}_2} = (2\bar{j}_2 - n + 1)/2, \tag{3.7}$$

where $\mathscr{T}^n_{j_1j_2}$ represents the number of terms inside the partial sum of B_0^{n+1} when $2 \leqslant \bar{j}_2 \leqslant j_2$, in particular, $\mathscr{T}_n = (n-1/2 \equiv \mathscr{T}^n_{1,n-1}$ (cf. Appendix B). Here the quantities $X^n_{\bar{j}_1\bar{j}_2}$, $\overline{X}^n_{j_1j_2}$ are defined recurrently by

$$X^n_{j_1j_2} = \frac{\mathscr{T}^{n-2}_{j_1j_2-2}}{\mathscr{T}^n_{j_1j_2}} X^{n-2}_{j_1j_2-2} x_{j_2+1} \overline{J}^n_{j_1+2,j_2-2} + \frac{1 - x_{j_2+1}\overline{J}^n_{j_1+2,j_2-2}}{\mathscr{T}^n_{j_1j_2}} + \frac{j_1}{\mathscr{T}^n_{j_1j_2}}\left(\frac{1}{j_2+1}\right) - \frac{x_{j_2+1}}{j_2-1}\overline{J}^n_{j_1+2,j_2-2}, \tag{3.8}$$

and the analogous definition holds for $\overline{X}^n_{\bar{j}_1\bar{j}_2}$. The quantities x_{j_2+1} (resp. $\bar{x}_{j_2+1}$), $\overline{J}^n_{j_1+2,j_2-2}$ (resp. $J^n_{j_1+2,j_2-2}$) are the analogs of y_{i_1} (resp. $\bar{y}_{i_1}$) and $\overline{I}^n_{i_1-2,i_2+2}$ (resp. $I^n_{i_1-2,i_2+2}$) (cf. the proof of Proposition 2.3) for the sweeping factors $\beta^n_{i_1i_2i_3}$. Precisely

$$x_{j_2+1} \text{ (resp. } \bar{x}_{j_2+1}) = \sup_{\delta\in\overline{\mathscr{C}}_\Lambda}\left(\text{resp. } \inf_{\delta\in\overline{\mathscr{C}}_\Lambda}\right) \frac{j_2(j_2-1)}{(j_2-2)(j_2-3)} \frac{\beta_{j_2-1}}{\beta_{j_2+1}} \frac{\delta_{j_2-1}}{\delta_{j_2+1}}, \tag{3.9}$$

$$\overline{J}^n_{j_1+2,j_2-2} = \begin{cases} (1 - \bar{\eta}^n_{j_1+2,j_2-2})^{-1}, & \text{if } j_2 - 1 \leqslant N_\Lambda, \\ (\mathscr{T}^n_{j_1+2,j_2-2}/\mathscr{T}^{n-2}_{j_1+2,j_2-4})[1 + \bar{k}^n_{j_1+2,j_2-2}/\mathscr{T}^n_{j_1+2,j_2-2}]^{-1}, & \text{if } j_2 - 1 \geqslant N_\Lambda \end{cases} \tag{3.10}$$

[cf. (biii) below for $\bar{\eta}^n_{j_1+2,j_2-2}$, $\bar{k}^n_{j_1+2,j_2-2}$] and, respectively, an analogous definition for $J^n_{j_1+2,j_2-2}$. Moreover,

(bi) $0 < X^n_{\bar{j}_1\bar{j}_2} \leqslant \overline{X}^n_{\bar{j}_1\bar{j}_2} \leqslant 1$, (3.11a)

(bii) $\exists X^{n(0)}_{\bar{j}_1\bar{j}_2} > 0$: $\lim_{\Lambda\to 0} \overline{X}^n_{\bar{j}_1\bar{j}_2}(\Lambda) = \lim_{\Lambda\to 0} X^n_{j_1j_2}(\Lambda) = X^{n(0)}_{\bar{j}_1\bar{j}_2}$, (3.11b)

(biii) there exist positive constants $0 < \eta^n_{\bar{j}_1\bar{j}_2}(\Lambda) \leqslant \bar{\eta}^n_{j_1j_2} \leqslant 2/j_2$, $0 \leqslant k^n_{\bar{j}_1\bar{j}_2} \leqslant k^n_{\bar{j}_1\bar{j}_2} \leqslant 1$,

such that

$$\frac{X^{n-2}_{\bar{j}_1,\bar{j}_2-2}}{\overline{X}^n_{\bar{j}_1\bar{j}_2}} \leqslant \begin{cases} (\mathscr{T}^n_{\bar{j}_1\bar{j}_2}/\mathscr{T}^{n-2}_{\bar{j}_1\bar{j}_2-2})(1 - \eta^n_{\bar{j}_1\bar{j}_2}), & \text{if } \bar{j}_2 - 1 \leqslant N_\Lambda, \\ (1 + k^n_{j_1j_2}/(\mathscr{T}^n_{j_1j_2})^2), & \text{if } j_2 - 1 \geqslant N_\Lambda, \end{cases} \tag{3.11c}$$

and, respectively,

$$\frac{\overline{X}^{n-2}_{\bar{j}_1,\bar{j}_2-2}}{X^n_{j_1j_2}} \geqslant \begin{cases} (\mathscr{T}^n_{\bar{j}_1\bar{j}_2}/\mathscr{T}^{n-2}_{\bar{j}_1\bar{j}_2-2})(1 - \eta^n_{\bar{j}_1\bar{j}_2}), & \text{if } j_2 - 1 \leqslant N_\Lambda, \\ (1 + \bar{k}^n_{j_1j_2}/(\mathscr{T}^n_{j_1j_2})^2), & \text{if } j_2 - 1 \geqslant N_\Lambda, \end{cases} \tag{3.11d}$$

with

$$\lim_{\Lambda\to 0} \eta^n_{j_1j_2}(\Lambda) = \lim_{\Lambda\to 0} \bar{\eta}^n_{j_1j_2}(\Lambda) = \eta^{n(0)}_{j_1j_2} > 0, \tag{3.11d1}$$

(biv) $\exists n_0 > N_\Lambda$: $Y^{n+2}_{\bar{j}_2+1,\bar{j}_1,1} \leqslant (\overline{X}^n_{\bar{j}_1\bar{j}_2})^2$, $\forall n \geqslant n_0$. (3.11e)

Properties (a), (b), and (bi)–(biv) give the complete definition of H.α.2. The starting points of H.α.2 are given by the two initial possible values of $\bar{j}_2$, $\bar{j}_2 = (n-1)/2$ and $\bar{j}_2 = (n+1)/2$, for which (3.1b) and (3.1c) are trivially satisfied. The recurrence procedure of H.α.2 goes from $\bar{j}_2 = (n-1)/2$ [or $(n+1)/2$] to $\bar{j}_2 = n-1$. To show (3.2) the crucial technical tool is again [as in Eq. (2.5) for the proof of the corresponding recurrent formula (2.1) of $\beta^n_{i_1i_2i_3}$ in Proposition 2.1] a "shifting" equation in order to obtain from $H_0^{\bar{j}_2}$ the next step $H_0^{\bar{j}_2+2}$, namely,

$$H_0^{j_2} H_0^{j_1+3} = \frac{\delta_{j_1+2}}{\delta_{j_2+1}} \frac{z(\beta_{j_1+2})}{z(\beta_{j_2+1})} H_0^{j_2+2} H_0^{j_1+1}. \tag{3.12}$$

Apart from this shifting equation one must also take into account the contribution of the symmetric partition $w_n^{\text{sym}}(j_1j_2) = (j_2+1, j_1-1)$. This constitutes the difference between the two sweeping procedures through the sums C_0^{n+1} and B_0^{n+1},

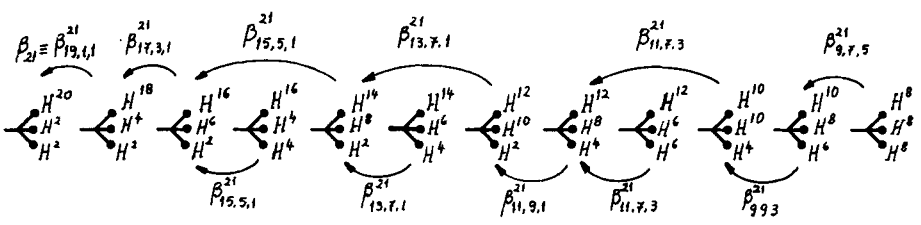


FIG. 1. The sweeping procedure β_n for n = 21.

because, in the latter case, the sweeping follows the direction of increasing j_2 from the middle to the end of B_0^{n+1} (cf. graphical representations of Figs. 1 and 2, respectively).

Using the signs together with the recurrence hypothesis H.α.2 and (3.12), we write [in view of Definition 2(c) of II]

$$\sum_{\substack{w(J)\\ [(n+1)/2]\leqslant \bar{j}_2 < \bar{j}_2}} \theta^n_{\bar{j}_1 \bar{j}_2} H_0^{\bar{j}_2+2} H_0^{\bar{j}_1+1} = \theta^n_{j_1 j_2}\left[1 + \frac{\theta^n_{j_1+2, j_2-2}}{\theta^n_{j_1 j_2}} \frac{\delta_{j_1+2}}{\delta_{j_2+1}} \frac{z(\beta_{j_1+2})}{z(\beta_{j_2+1})} \alpha^n_{j_1+2, j_2-2} + \frac{\theta^n_{j_2+1, j_1-1}}{\theta^n_{j_1 j_2}}\right] H_0^{j_2+2} H_0^{j_1+1}. \tag{3.13}$$

Then, by defining $\alpha^n_{j_1 j_2}$ as the factor multiplying $\theta^n_{j_1 j_2} H_0^{j_2+2} H_0^{j_1+2}$ in (3.13), we obtain Proposition 3.2. Q.E.D.

For the proof of properties (b) [(3.7)–(3.11d)] one only has to repeat exactly the analogous arguments we presented in detail for the proof of the corresponding properties of the β_n's and $Y_n(\Lambda)$'s in Proposition 2.3.

We nevertheless present a hint for the proof of (3.11e) that relates the "decreasing" factors of the β_n's and α_n's. The simplest way to demonstrate it is to show that there exists $\rho^n_{j_2}(\Lambda) \geqslant 0$ such that

$$(X^n_{j_1 j_2})^2 - Y^{n+2}_{j_2+1, j_1, 1} \geqslant \rho^n_{j_2}(\Lambda), \quad \forall n \geqslant n_0 > N_\Lambda, \quad j_2 > (n+1)/2. \tag{3.14}$$

Using the fact that this bound is true by H.α.2 we subtract the corresponding expressions $(\bar{X}^n_{j_1 j_2})^2$ [cf. the analog of (3.8) above] and $Y^{n+2}_{j_2+1, j_1, 1}(\Lambda)$ [cf. the analogs of (2.27) and (2.28) in the proof of Proposition 2.3] and obtain the following expression for $\rho^n_{j_2}(\Lambda)$:

$$\begin{aligned}
\rho^n_{j_2}(\Lambda) = {} & \left(\frac{\mathscr{T}^{n-2}_{j_1, j_2-2}}{\mathscr{T}^n_{j_1, j_2}}\right)^2 y_{j_2+1} \bar{I}^{n+2}_{j_2-1, j_1+2} \rho^n_{j_2-2} \\
& + \left(\frac{\mathscr{T}^{n-2}_{j_1 j_2-2}}{\mathscr{T}^n_{j_1 j_2}}\right)^2 (\bar{X}^{n-2}_{j_1 j_2-2})^2 \{(\bar{\alpha}_{j_2+1})^2 [(J^n_{j_1+2, j_2-2})^2 - \bar{I}^{n+1}_{j_2-1, j_1+2}] + \bar{I}^{n+2}_{j_2-1, j_1+2} [(\bar{\alpha}_{j_2+1})^2 - y_{j_2+1}]\} \\
& + Y^n_{j_2-1, j_1+2, 1}\, y_{j_2+1} \bar{I}^{n+2}_{j_2-1, j_1+2} \left\{\left(\frac{\mathscr{T}^{n-2}_{j_1, j_2-2}}{\mathscr{T}^n_{j_1 j_2}}\right)^2 - \frac{\bar{\mathscr{T}}^n_{j_2-1, j_1, 1}}{\bar{\mathscr{T}}^{n+2}_{j_2+1, j_1, 1}}\right\} \\
& + \left\{2 \frac{\mathscr{T}^{n-2}_{j_1, j_2-2}}{\mathscr{T}^n_{j_1 j_2}} \bar{X}^{n-2}_{j_1, j_2-2} \bar{x}_{j_2+1} J^n_{j_1+2, j_2-2} \left[\frac{1 - \bar{x}_{j_2+1} \bar{J}^n_{j_1+2, j_2-2}}{\mathscr{T}^n_{j_1 j_2}} + \frac{j_1}{\mathscr{T}^n_{j_1 j_2}} \left(\frac{1}{j_2+1} - \frac{\bar{\alpha}_{j_2+1}}{j_2-1} J^n_{j_1+2, j_2-2}\right)\right]\right. \\
& + \left[\frac{1 - \bar{x}_{j_2+1} \bar{J}^n_{j_1+2, j_2+2}}{\mathscr{T}^n_{j_1 j_2}} + \frac{j_1}{\mathscr{T}^n_{j_1 j_2}} \left(\frac{1}{j_2+1} - \frac{\bar{x}_{j_2+1}}{j_2-1} J^n_{j_1+2, j_2-2}\right)\right]^2 - \frac{(i - y_{j_2+1} \bar{I}^{n+2}_{j_2-1, j_1+2})}{\bar{\mathscr{T}}^{n+2}_{j_2+1, j_1, 1}}.
\end{aligned} \tag{3.15}$$

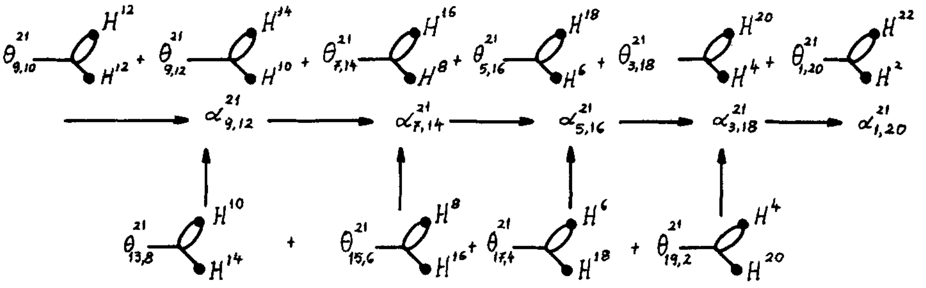


FIG. 2. The sweeping procedure α_n for $n = 21$.

One then verifies that $\rho^n_{j_2}(\Lambda)\geqslant 0$ because the sum of the differences in the brackets above yields a non-negative contribution in view of the values of $\overline{\mathscr{T}}^{n+2}_{j_2+1,j_1,1}$, $\mathscr{T}^n_{j_1j_2}$, etc. (cf. Appendix B) and the recurrence hypothesis H.α.2. From this result we obtain that $Y^{n+2}_{j_2+1,j_1,1}<(\overline{X}^n_{j_1j_2})^2$ and this completes the proof of H.α.2.

Application of the secondary recursion H.α.2 for $j_2=n-1$ yields the proof of all properties of the principal recurrence hypothesis H.α.1 except the $n\to\infty$ limit property (3.3c) for the lower bound of $\overline{X}_n(\Lambda)$. This can be obtained by application of the property (3.3e) $\overline{X}^2_n\geqslant Y_{n+2}(\Lambda)$ and the corresponding positive lower bound at infinity of $Y_{n+2}(\Lambda)$ [(2.12c)] proved in Proposition 2.3. Precisely we have

$$\lim_{n\to\infty}\overline{X}_n(\Lambda)=\lim_{n\to\infty}(\overline{X}^2_n)^{1/2}\geqslant\lim_{n\to\infty}(Y_{n+2})^{1/2}=Y^{1/2}_\infty(\Lambda), \tag{3.16}$$

which yields that

$$\exists X_\infty(\Lambda)\equiv Y^{1/2}_\infty>0\quad(\forall\Lambda\text{ fixed with }0<\Lambda\leqslant 0.1).$$

Q.E.D.

We now proceed to the proof of limit values and absolute and relative bounds of the functionals $\Delta_n(\delta(\Lambda))$, $n\geqslant 5$ [cf. definition (3.1d1) of the mapping (3.1) in II], i.e.,

$$\Delta_5\overset{\text{def}}{\equiv}\alpha_5(\delta)-\delta_7(\Lambda)\beta_7(\delta)/15 \tag{3.17a}$$

and,

$$\forall n\geqslant 7,$$

$$\Delta_n\overset{\text{def}}{\equiv}\alpha_n/(n-1)-\delta_{n+2}\beta_{n+2}/3n(n-1). \tag{3.17b}$$

As will be clear below in the proof of Proposition 3.2 these properties of Δ_n's result from the corresponding properties of the sweeping factors β_n's and α_n's proved by Propositions 2.1, 2,2, and 3.1, respectively. They have been crucially used in Sec. III of II for the proof of the stability of $\overline{\mathscr{C}}_\Lambda$ (cf. Lemma 3.2, and the proof of Theorem 3.2 of II).

Proposition 3.2: Let $\delta\in\overline{\mathscr{C}}_\Lambda$. Then for all $n\geqslant 5$, the functionals $\Delta_n(\delta(\Lambda))$ of definitions (2.17a) and (2.17b) satisfy the following properties (Λ fixed and $0<\Lambda\leqslant 0.1$):

(i) (a) $\lim_{\Lambda\to 0}\Delta_n(\Lambda)\equiv\Delta^{(0)}_n=X^{(0)}_n/2,$ (3.18)

(b) $\forall n\geqslant 7,\quad \Delta_n(\Lambda)\leqslant\frac{1}{2},$ (3.19)

(c) $\Delta_5(\Lambda)\geqslant\frac{1}{2}$ and $\forall n\geqslant 7,$ (3.20)

$(\Delta_n(\Lambda)\geqslant 1/\delta^\Lambda_0$ iff $2/X_\infty\leqslant\delta^\Lambda_\infty\leqslant 70/X_\infty$; (3.20a)

(ii) $2\leqslant\Delta_5/\Delta_7\leqslant 6$ (3.21a)

and for all $n\geqslant 9$ there exists

$$0<\nu_n(\Lambda)\leqslant\bar{\nu}_n(\Lambda)\leqslant 2/(n-2),\quad 0\leqslant\bar{\omega}_n(\Lambda)\leqslant\omega_n(\Lambda)<\infty$$

such that

$$\frac{\Delta_{n-2}}{\Delta_n}\leqslant\begin{cases}(1-\nu_n(\Lambda))(n-1)/(n-3), & \text{if } n\leqslant N_\Lambda,\\ 1+\omega_n(\Lambda)/(n-2)(n-3), & \text{if } n\geqslant N_\Lambda,\end{cases} \tag{3.21b}$$

and

$$\frac{\Delta_{n-2}}{\Delta_n}\geqslant\begin{cases}(1-\bar{\nu}_n(\Lambda))(n-1)/(n-3), & \text{if } n\leqslant N_\Lambda,\\ 1+\bar{\omega}_n/(n-2)(n-3), & \text{if } n\geqslant N_\Lambda,\end{cases} \tag{3.21c}$$

where $X^{(0)}_n$ and $X_\infty(\Lambda)$ are defined in Proposition 3.1, and $N_\Lambda\equiv(4/3\Lambda)+1$;

(iii) $\Delta_{n(2)}-\Delta_{n(1)}\geqslant 0,$

$\Delta_{n-2(2)}-\Delta_{n-2(1)}\geqslant\Delta_{n(2)}-\Delta_{n(1)},$

$\forall 0<\Lambda\lesssim 0.01.$ (3.21d)

Proof of Proposition 3.1: (i) Taking into account Propositions 2.1 and 3.1 for β_7 and α_5, we have

$$\beta_7=1+\tfrac{5}{6}\delta_3/\delta_5,\quad \alpha_5=1+\tfrac{1}{2}\delta_3/\delta_5. \tag{3.22}$$

Thus (3.17a) yields

$$\Delta_5=1+\frac{1}{2}\frac{\delta_3}{\delta_5}-\frac{\delta_7}{15}\left(1+\frac{5}{6}\frac{\delta_3}{\delta_5}\right). \tag{3.23}$$

In view of the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$, we obtain, on one hand, the limit value

$$\lim_{\Lambda\to 0}\Delta_5(\Lambda)=\tfrac{6}{5} \tag{3.23a}$$

and, on the other hand, the lower bound of Δ_5:

$$\Delta_5(\Lambda)\geqslant 1-\frac{42\Lambda\bar{S}_7}{5}+\frac{\bar{I}_3}{\bar{S}_5}\left(1-\frac{42\Lambda\bar{S}_7}{5}\right). \tag{3.23b}$$

Insertion of the definitions (2.4a)–(2.4c) of II for $\bar{I}_3,\bar{S}_5,\bar{S}_7$ yields, after some estimations, that $\Delta_5(\Lambda)\geqslant\frac{1}{2}$ for $\Lambda\lesssim 0.1$. Q.E.D.

From (3.17b) one easily obtains (in view of the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$ and Propositions 3.1 and 2.3 for α_n and β_{n+2}, respectively), on one hand, the limit value

$$\lim_{\Lambda\to 0}\Delta_n(\Lambda)=X^{(0)}_n/2 \tag{3.24}$$

[where $X^{(0)}_n$ is given by (3.3b)] and, on the other hand, the upper bound $\Delta_n\leqslant\frac{1}{2}$. Q.E.D.

In a similar way for the lower bound of Δ_n's we apply the definitions (2.4a)–(2.4c) of $\overline{\mathscr{C}}_\Lambda$ in II and the lower (resp. upper) bounds of α_n's (resp. β_{n+2}) given by Proposition 3.1 (resp. 2.3):

$$\Delta_n\geqslant\overline{X}_n/2-[\overline{\mathscr{T}}_{n+2}Y_{n+2}/3n(n-1)]\delta^\Lambda_\infty. \tag{3.25}$$

In order to prove (3.20) we require that the rhs of (3.25) be bounded below by $1/\delta^\Lambda_\infty$. Using the properties $Y_{n+2}(\Lambda)\leqslant\overline{X}^2_n$ (in view of Proposition 3.1) and the bound $\overline{\mathscr{T}}_{n+2}\leqslant(n-1)^2/48$ (cf. Appendix B), this condition can be written in a stronger form:

$$X_\infty(\Lambda)\left(\frac{1}{2}-\frac{\delta^\Lambda_\infty X_\infty(\Lambda)}{144}\right)\geqslant\frac{1}{\delta^\Lambda_\infty}. \tag{3.26}$$

The latter is verified iff δ^Λ_∞ takes values in the interval

$$2/X_\infty(\Lambda)\leqslant\delta^\Lambda_\infty\leqslant 70/X_\infty(\Lambda). \tag{3.27}$$

Finally, taking into account the bound (2.49) for $X_\infty(\Lambda)\leqslant X^2_\infty$, we can write (3.27) more explicitly as

$$4(\overline{\mathscr{T}}_{N_\Lambda})^{1/2}/[1.3\pi]^{1/2}\leqslant\delta^\Lambda_\infty\leqslant 140(\overline{\mathscr{T}}_{N_\Lambda})^{1/2}/[1.3\pi]^{1/2}, \tag{3.27'}$$

where $N_\Lambda\equiv 4/3\Lambda+1$ and $\overline{\mathscr{T}}_{N_\Lambda}$ means the number of different partitions $w(i_1i_2i_3)$ of N_Λ and it is evaluated exactly in Appendix B.

The last condition on the upper bound of the sequences $\{\delta_n\}_n$ in $\overline{\mathscr{C}}_\Lambda$ ensures the existence of a finite positive lower bound of Δ_n [property (3.20)] which is the most crucial point for the stability of $\overline{\mathscr{C}}_\Lambda$ (cf. Sec. III of II).

(ii) To show properties (3.21a)–(3.21c) we use the corresponding definitions (3.17b) of Δ_n and Δ_{n-2} and write down the difference $\Delta_{n-2}-\Delta_n$ in terms of Δ_{n-2}. Then by Propositions 2.3 and 3.1 and the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$ we evaluate the corresponding upper and lower bounds. We do not give the details for all the cases but we present the argument explicitly only for one of them, for example, the upper bound (3.21b) where $7\leqslant n\leqslant N_\Lambda$. Thus we write

$$\Delta_{n-2}-\Delta_n=\frac{\alpha_n}{n-1}\left[\frac{\alpha_{n-2}}{\alpha_n}\frac{n-1}{n-3}-1\right]-\frac{\delta_{n+2}\beta_{n+2}}{3n(n-1)}\left[\frac{\delta_n\beta_n}{\delta_{n+2}\beta_{n+2}}\times\frac{n(n-1)}{(n-2)(n-3)}-1\right]. \quad (3.28)$$

In view of the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$ [cf. Definitions 2(a) and 2(a1) of II], we apply the relative upper bounds of $X_{n-2}/\overline{X}_n$ given by Proposition 3.1, in the first term of the rhs of (3.28) and obtain

$$\frac{\alpha_{n-2}}{\alpha_n}\frac{n-1}{n-3}-1\leqslant\frac{2}{n-3}-\eta_n-\frac{2\eta_n}{n-3}\equiv\sigma_1. \quad (3.29)$$

In an analogous way (by the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$) we apply the relative lower bounds of δ_n/δ_{n+2} and β_n/β_{n+2} (the latter given by Proposition 2.3) in the second term of (3.28) and using the decomposition

$$\frac{n^2(n-1)^2}{(n^2-4)(n-3)(n+1)}=1+\frac{8}{n^2-2n-6}-\frac{8}{(n^2-4)(n-3)}+\frac{12}{(n^2-4)(n-3)(n+1)},$$

we finally have

$$\frac{\delta_n}{\delta_{n+2}}\frac{\beta_n}{\beta_{n+2}}\frac{n(n-1)}{(n-2)(n-3)}-1\geqslant(\overline{\gamma}_{n+2}(\Lambda)-\overline{\Psi}_{n+2})-\gamma_{n+2}\overline{\Psi}_{n+2}+\frac{(1-\overline{\Psi}_{n+2})}{1-\gamma_{n+2}}\times\left[\frac{8}{n^2-2n-6}-\frac{8}{(n^2-4)(n-3)}\right]\equiv\sigma_2. \quad (3.30)$$

By insertion of the two bounds (3.29) and (3.30) in (3.28) and taking into account the absolute lower bounds, $\delta_{n\pm2}\geqslant3\Lambda(n+1)(n+2)\overline{I}_{n+2}$ ($\delta\in\overline{\mathscr{C}}_\Lambda$) and $\beta_{n+2}\geqslant\overline{\mathscr{T}}_{n+2}Y_{n+2}$ (Proposition 2.3), and the upper bound $\Delta_n<\frac{1}{2}$ (proved before), we obtain, after some manipulations,

$$\Delta_{n-2}-\Delta_n\leqslant\Delta_n\left\{\sigma_1-\frac{(n-1)^2(\sigma_2-\sigma_1)}{12[2+3\Lambda(n+1)(n+2)]}\right\}. \quad (3.31)$$

Now we identify

$$2-\nu_n(\Lambda)(n-1)\equiv(n-3)\left\{\sigma_1-\frac{(n-1)^2(\sigma_2-\sigma_1)}{12[2+3\Lambda(n+1)(n+2)]}\right\}, \quad (3.32)$$

so that by (3.29) we obtain the definition of $\nu_n(\Lambda)$,

$$\nu_n(\Lambda)=\eta_n+\frac{(n-1)(n-3)(\sigma_2-\sigma_1)}{12[2+3\Lambda(n+1)(n+2)]}, \quad (3.33)$$

which means that $\nu_n(\Lambda)$ is proportional to $\eta_n(\Lambda)$ (cf. Proposition 3.1). From (3.31) and using (3.32), we can write

$$\Delta_{n-2}/\Delta_n\leqslant[(n-1)/(n-3)]\{1-\nu_n(\Lambda)\}, \quad (3.34)$$

where $\nu_n(\Lambda)$ is defined by (3.33). Finally one easily verifies the bounds $0<\nu_n(\Lambda)\leqslant2/(n-2)$ by (3.33), in view of definitions (3.29) and (3.30) of σ_1,σ_2, respectively, and the corresponding bounds of $\eta_n,\overline{\Psi}_{n+2},\gamma_{n+2}$ (by the hypothesis $\delta\in\overline{\mathscr{C}}_\Lambda$ and by application of Propositions 2.3 and 3.1). An analogous demonstration holds for the case $n\geqslant N_\Lambda$ with the corresponding expression of ω_n. [Properties (iii) are easily obtained using the recursive inequalities (2.12a1) and (3.3a1) for the differences of β_n's and α_n's, respectively.] With these last considerations we complete the proof of Proposition 3.2.

Exactly analogous properties to that presented by Propositions 3.1 and 3.2, respectively, hold for the sweeping factors ${}^\Phi\alpha_n$ of ${}^\Phi B_n^{n+1}$ [cf. definition (4.12d) in II]. We present them by Proposition 3.3 below without their proofs, because they can be easily obtained by using arguments analogous to that explained above.

Proposition 3.3: (i) Let $H\in\Phi_\Lambda$. For every coherent sequence of ΦC's$\in\mathscr{F}$, the sweeping factors ${}^\Phi\alpha_n$, defined by definition (2.34) of II, are recurrently defined by the formulas presented in Lemma 2.7 of II, in terms of the ratios of the corresponding splitting constants ${}^\Phi\delta_n$. Moreover, the corresponding limit values and absolute and relative bounds—analogous to those of Proposition 3.1—also hold (cf. Lemma 2.6 of II).

(ii) For every coherent sequence of ΦC's, $\Phi\in\mathscr{F}$, the functionals ${}^\Phi\Delta_n$, defined by

$${}^\Phi\Delta_n\equiv\frac{{}^\Phi\alpha_n\,{}^\Phi b_n}{n-1}-\frac{{}^\Phi\delta_{n+2}\,{}^\Phi\beta_{n+2}\,{}^\Phi a_n}{3n(n-1)}$$

[definition (4.12d) of II], (3.35)

satisfy, for all $n\geqslant7$,

$$\lim_{\Lambda\to0}{}^\Phi\Delta_n(\Lambda)={}^\Phi\Delta_n^{(0)}>0, \quad (3.36)$$

$${}^\Phi\Delta_n\leqslant\frac{1}{2},\quad {}^\Phi\Delta_n\geqslant\frac{1}{\tilde{\delta}_\infty^\Lambda}\ \text{iff}\ \frac{2}{\overline{X}_\infty}<\tilde{\delta}_\infty^\Lambda<\frac{13}{\overline{X}_\infty} \quad (3.37)$$

(here again $\tilde{\delta}_\infty^\Lambda$ is the corresponding upper bound of every sequence ${}^\Phi\delta\in\overline{\mathscr{C}}_\Lambda$). Moreover, relative bounds (decreasing properties) analogous to those of Proposition 3.2 (for Δ_n) are satisfied.

ACKNOLWEDGMENTS

The confirmation of the present method, through the identification of the two "zero-dimensional" solutions, has

been made possible thanks to the private communications of Professor Wightman and A. Voros to whom the author wishes to express her sincere gratitude.

The numerical test concerning the zero-dimensional problem originated from a suggestion of Professor Ph. Blanchard and Professor K. Osterwalder. The author is deeply indebted to both.

She wishes also to acknowledge the constant interest in her work of Professor L. Michel, Professor A. Wightman, Professor J. Iliopoulos, and Professor Ph. Blanchard.

Interesting discussions with Professor J. Ginibre and stimulating criticisms of A. Voros and B. Grammaticos are also gratefully acknowledged.

Finally she is indebted to the Institut des Hautes Etudes for the material presentation of the article.

APPENDIX A: NUMERICAL RESULTS

In this appendix we present some numerical verifications of the method, provided by the numerical evaluations of Voros.[2] These results are based on the numerical comparison (for fixed values of the coupling constant Λ in the interval $0 < \Lambda \leqslant 0.1$) of our equations for the Schwinger functions [system (A6) below] with the analogous equations one obtains by the functional integral formalism. More precisely, starting from the generating functional[3]

$$f(h) = \int e^{h\Phi}\, d\mu_{m,\lambda}(\Phi) = \sum_{N=0}^{\infty} \frac{S_N(m,\lambda)}{N!} h^N, \tag{A1}$$

where

$$S_N(m,\lambda) = \int \Phi^N\, d\mu_{m,\lambda}(\Phi) \tag{A2}$$

and

$$d\mu_{m,\lambda}(\Phi) = \frac{\exp[-(m\Phi^2 + \lambda\Phi^4)]d\Phi}{\int \exp[-(m\Phi^2 + \lambda\Phi^4)]d\Phi}, \tag{A3}$$

one can obtain the equations of motion for the truncated functions g_N defined by

$$g(h) = \operatorname{Inf}(h) = \sum_{N=1}^{\infty} \frac{g_N(m,\lambda)}{N!} h^N. \tag{A4}$$

These equations have the form

$$g_2 = \frac{1}{2m} - \frac{6\lambda}{m}(g_2)^2 - \frac{2\lambda}{m} g_4 \tag{A5a}$$

and, in general,

$$\begin{aligned}\frac{g_{2N+2}}{(2N+1)!} &= \frac{-2\lambda}{m} \frac{g_{2N+4}}{(2N+1)!} - \frac{6\lambda}{m} \\ &\quad\times \sum_{k=0}^{N} \frac{g_{2k+2}}{(2k+1)!} \frac{g_{2(N-k)+2}}{[2(N-k)]!} \\ &\quad - \frac{2\lambda}{m} \sum_{k_1,k_2,k_3 \geqslant 0} \frac{g_{2k_1+2}\, g_{2k_2+2}\, g_{2k_3+2}}{(2k_1+1)!(2k_2+1)!(2k_3+1)!}, \\ &\quad k_1 + k_2 + k_3 = N-1.\end{aligned} \tag{A5b}$$

As a result of this direct derivative of the system of equations (A5a) and (A5b), the truncated functions $g_N(\beta,\alpha)$ are automatically solutions of it. So a good test of our method[5] should be the comparison of the system (A5a) and (A5b) with the zero-dimensional system (cf. Sec. I of II and Sec. III of I)

$$H_0^2 = -\Lambda H_0^4 + 1, \tag{A6a}$$

$$\begin{aligned}H_0^{n+1} &= -\Lambda H_0^{n+3} - 3\Lambda \sum_{w(J)} \frac{n!}{j_1!\, j_2!} H_0^{j_2+2} H_0^{j_1+1} \\ &\quad - 6\Lambda \sum_{w(I)} \frac{n!}{i_1! i_2! i_3! \sigma_{\text{sym}}(i_1 i_2 i_3)} \prod_{l=1}^{3} H_0^{i_l+1}.\end{aligned} \tag{A6b}$$

As we explained in I and II, this zero-dimensional system (A6) results from the corresponding two-dimensional equations of motion where divergent (even in two dimensions) terms of self-energy type graphically represented, for example, by  vanish. This annihilation comes from the renormalization operator, intrinsically present in the definition of the renormalized normal product,[6] i.e., the corresponding rhs's (A6a) and (A6b). This fact explains the absence of the terms $(6\lambda/m)(g_2)^2$ in the first equation of (A6) or $(6\lambda/m) g_{2N+2} g_2$ in the second one. So, finally, if we put

$$\Lambda = 4\lambda, \tag{A7}$$

$$2m + 12\lambda g_2 = 1, \tag{A8}$$

we obtain the complete equivalence between the systems [(A5a) and (A5b)] and (A6).

Taking into account the renormalization conditions (A7) and (A8), Voros[2] has obtained numerically the results of Tables I and II in terms of the corresponding values of the splitting sequences $\{\delta_n^w\}$ (cf. Ref. 1). From these numerical values we concluded precisely the following.

(i) The two systems [(A5a) and (A5b)] and (A6) are completely equivalent modulo the renormalization conditions (A7) and (A8).

(ii) The solution $\{g_{2N+2}\}_N$ of (A5), when the "splitting" and the "sweeping" procedures defined by (1.4) and (1.5) of II[1] and Proposition 2.1 are considered, can be expressed in terms of the corresponding splitting sweeping sequences $\{\delta_n^{(W)}\}_{n=2N+1}$ and $\{\beta_n^{(W)}(\delta_n^{(W)})\}$ so that we can define the analogous system (3.1) in II. This system (or

TABLE I. The values of $\delta_n^{(W)}$ for $1 \leqslant n \leqslant 29$ when $\Lambda = 0.099\,179\,835$.

n	$\delta_n^{(W)}$
1	0.404 423 77
3	0.373 832 93
5	0.660 979 21
7	4.242 520 82
9	5.805 739 75
11	7.289 414 16
13	8.677 898 91
15	9.968 741 47
17	11.165 585 49
19	12.274 673 13
21	13.303 102 39
23	14.257 977 45
25	15.146 017 43
27	15.973 398 81
29	16.745 714 55

TABLE II. The values of $\delta_n^{(W)}$, maximal and minimal values $\delta_{n\,\max}$ and $\delta_{n\,\min}$, for $1<n\leqslant 55$, when $\Lambda=0.099\,179\,835$.

n	$\delta_{n\,\min}$	$\delta_n^{(W)}$	$\delta_{n\,\max}$
1	0.305 90	0.404 42	0.595 08
3	0.305 90	0.373 83	0.483 87
5	0.479 23	0.660 98	1.529 98
7	1.640 60	4.242 52	10.587 93
9	1.735 54	5.805 74	15.882 33
11	1.785 51	7.289 41	20.948 45
13	1.814 70	8.677 90	25.491 39
15	1.833 13	9.968 74	29.413 72
17	1.845 47	11.165 59	32.730 00
19	1.854 13	12.274 67	35.505 19
21	1.860 42	13.303 10	37.819 25
23	1.865 14	14.257 98	39.749 88
25	1.868 77	15.146 02	41.365 54
27	1.871 61	15.973 40	42.723 77
29	1.873 89	16.745 71	43.871 67
31	1.875 74	17.467 99	44.847 37
33	1.877 26	18.144 72	45.681 53
35	1.878 52	18.779 92	46.398 82
37	1.879 58	19.377 19	47.019 07
39	1.880 48	19.939 73	47.558 32
41	1.881 26	20.470 41	48.029 57
43	1.881 92	20.971 82	48.443 42
45	1.882 50	21.446 26	48 808 55
47	1.883 01	21.895 82	49.132 12
49	1.883 46	22.322 38	49.420 05
51	1.883 85	22.727 64	49.677 26
53	1.884 20	23.113 13	49.907 90
55	1.884 51	23.480 26	50.115 42

mapping $\mathscr{M}_\delta$ in the space of the splitting sequences) has as its solution the corresponding sequence $\{\delta_n^{(W)}\}$. In other words, the solution of our system (31) in II (by contraction of $\mathscr{M}_\delta$) coincides with the corresponding solution $\{\delta_n^{(W)}\}$ coming from the functional integral method.

Both results (i) and (ii) appear in Table I. Voros has used two values of Λ: 0.099 17 . . . and 0.0644. In Table I we present the results (up to $n=29$) corresponding to the value 0.099 179 and analogous numerical values for the splitting constants $\delta_n^W(\Lambda)$ exist corresponding to $\Lambda=0.064\,48\ldots$ (up to $n=69$). For a convincingly large number of iterations the values of Table I (and the ones corresponding to $\Lambda=0.0644$) remain invariant under the mapping $\mathscr{M}_\delta$ [or (3.1) of II].

(iii) This solution $\{\delta_n^W\}$ satisfies all absolute and relative bounds that describe the subset $\overline{\mathscr{C}}_\Lambda$ (i.e., the fine structure of $\Phi_{0\Lambda}$) in Sec. II of II.[1] Moreover, using these values one verifies numerically all bounds for the sweeping factors β_n's,α_n's that constitute the key for our combinatorial technique. From all these results the consistency of the scheme has been established.

The numerical verification of the absolute bounds are presented in the Table II for $\Lambda=0.099\,17$. There $\delta_n^{(W)}(\Lambda)$ means the "true" values of the δ_n's coming from the application of the splitting and sweeping formulas on the sequence $\{g_{n+1}^W\}$, as in Table I [cf. (3.55), Corollary 3.0 of Sec. III of II and Proposition 2.1 of this paper] but here for $=1,3,\ldots,69$ (in the table the values up to $n=55$ appear). Here $\delta_{n\,\max}$ means the absolute upper bounds of δ_n^W's following the formulas

$$\delta_{1\,\max}=6\Lambda,\quad \delta_{3\,\max}=\frac{6\Lambda(1+6\Lambda^2)}{1+9\Lambda-60\Lambda^2},$$
$$\delta_{5\,\max}=\frac{2(1+6\Lambda^2)15\Lambda}{2+15\Lambda(1+6\Lambda^2)},\tag{A9}$$
$$\forall n\geqslant 7,\quad \delta_{n\,\max}=\frac{3\Lambda n(n-1)\delta_\infty^\Lambda(1+6\Lambda^2)}{\delta_\infty^\Lambda+3\Lambda n(n-1)}$$

[cf. Definition 2(a1) of II]. Here δ_∞^Λ is fixed $\delta_\infty^{\Lambda\simeq 0.1}=50$ following the bounds (3.20a) in Proposition 3.3.

Respectively, $\delta_{n\,\min}$ means the absolute lower bounds of $\overline{\mathscr{C}}_\Lambda$ (because $\{\delta_n^W\}\in\overline{\mathscr{C}}_\Lambda$) following the formulas

$$\delta_{1\,\min}=\delta_{3\,\min}=6\Lambda[1+9\Lambda(1+6\Lambda^2)]^{-1},$$
$$\delta_{5\,\min}=\frac{15\Lambda(1-3\Lambda S_3)}{1+15\Lambda(1+6\Lambda^2)},$$
$$\delta_{n\,\min}=6\Lambda n(n-1)[2+3\Lambda n(n-1)(1+6\Lambda^2)]^{-1},\quad \forall n\geqslant 7\tag{A10}$$

[cf. definition 2(a1) for $\overline{\mathscr{C}}_\Lambda$ of II].

Analogous results have been obtained by Voros for Λ fixed at $\Lambda\simeq 0.064\,48\ldots$ and up to $=69$.

The relative bounds of the sequence $\{\delta_n^W\}\in\overline{\mathscr{C}}_\Lambda$ are also verified numerically using the values presented in Tables I and II.

APPENDIX B: THE NUMBERS $\mathscr{T}_n,\overline{\mathscr{T}}_n$

Let $\mathscr{T}_n$ (resp. $\overline{\mathscr{T}}_n$) be the number of different partitions (doublets) $w_n(j)$ such that $j_1+j_2=n$ ($j_1=$ odd) [resp. the number of different partitions (triplets) $w_n(I)=(i_1i_2i_3)$ such that $\Sigma_{1<l<3}\, i_l=n$]; then we prove the following combinatorial properties.

Lemma B.1:

$$\text{(i)}\ \mathscr{T}_n=(n-1)/2;\tag{B1}$$

$$\text{(ii)}\ \overline{\mathscr{T}}_n=\begin{cases}\dfrac{(\bar n-3)^2}{48}+\dfrac{\bar n-3}{4}+1, & \text{for } n=\bar n=12l+3,\quad l=0,1,2,\ldots, \qquad\text{(B2a)}\\[2mm] \dfrac{(\bar n-3)^2}{48}+\dfrac{\bar n-3}{4}+1+\dfrac{\bar n-3}{12}, & \text{for } n=\bar n+2, \qquad\text{(B2b)}\\[2mm] \dfrac{(\bar n-3)^2}{48}+\dfrac{\bar n-3}{4}+1+\dfrac{\bar n-3}{12}+\dfrac{\bar n+9}{12}g, & \text{for } n=\bar n+2(g+1),\quad 1\leqslant g\leqslant 5. \qquad\text{(B2c)}\end{cases}$$

Proof: (i) We note that the number $\mathscr{T}_n$ is exactly the number of different values taken by the even number j_2 inside the interval $2 \leqslant j_2 \leqslant n-1$, i.e.,

$$\mathscr{T}_n = \tfrac{1}{2}(n-1-2)+1 = (n-1)/2. \qquad \text{Q.E.D.}$$

(ii) (a) We shall first prove the formula (B2a): Let $n = 12l + 3$, $l = 0,1,2$. We also suppose that $i_1 \geqslant i_2 \geqslant i_3$; then in view of the definition $\Sigma_{1<j<3}\, i_j = n$, one finds

$$i_{1\,\max} = n-2, \tag{B3a}$$

$$i_{1\,\min} = n/3. \tag{B3b}$$

In view of the hypothesis (B2a) both of them are odd numbers.

We first keep i_1 fixed. Then, i_2 varies in the interval $i_{2\,\min} \leqslant i_2 \leqslant i_{2\,\max}$ and if we denote by $i_{10} \equiv (n-1)/2$ (here i_{10} = odd int. in view of hyp.), then

$$i_{2\,\max} = \begin{cases} n - i_1 - 1, & \text{if } i_1 > i_{10}, & \text{(B4a)} \\ i_1, & \text{if } i_1 \leqslant i_{10}, & \text{(B4b)} \end{cases}$$

and

$$i_{2\,\min} = \begin{cases} (n-i_1)/2, & \text{if } (n-i_1)/2 = 2k+1, \quad k = 0,1,..., \\ (n-i_1)/2+1, & \text{if } (n-i_1)/2 = 2k. \end{cases} \tag{B5}$$

We now define a number $\overline{\mathscr{T}}^{(ni_1)}$ representing the number of different partitions $w_n(I)$ when i_1 is fixed and i_2,i_3 vary. We note that $\overline{\mathscr{T}}^{(ni_1)}$ equals exactly the number of different values i_2 can take inside the interval defined by the extremal values (B4) and (B5):

$$\overline{\mathscr{T}}^{(ni_1)} = (i_{2\,\max} - i_{2\,\min})/2 + 1. \tag{B6}$$

From the above considerations, it follows that the total number $\overline{\mathscr{T}}_n$ equals

$$\overline{\mathscr{T}}_n = \sum_{i_1 = n/3}^{n-2} \overline{\mathscr{T}}^{(ni_1)}. \tag{B7}$$

For $i_1 \leqslant i_{10}$ we calculate [following (B4b) and the two cases of (B5)]

$$\overline{\mathscr{T}}^{(ni_1)} = \begin{cases} \dfrac{3i_1 - n}{4} + 1, & \text{if } i_1 = \dfrac{n}{3}, \dfrac{n}{3}+4, ..., \dfrac{n-1}{2}, & \text{(B8a)} \\[2ex] \dfrac{3i_1 - n - 2}{4}, & \text{if } i_1 = \dfrac{n}{3}+2, \dfrac{n}{3}+6, ..., \dfrac{n-1}{2} - 2, & \text{(B8b)} \end{cases}$$

so when l of (B2a) is even we have

$$\sum_{n/3 < i_1 < (n-1)/2} \overline{\mathscr{T}}^{(ni_1)} = \Sigma^{\pi_1} + \Sigma^{\pi_2} \tag{B9}$$

$$\text{with } \begin{cases} \Sigma^{\pi_1} = 1 + 4 + 7 + \cdots + (n-3)/8 + 1, \\ \Sigma^{\pi_2} = 2 + 5 + 8 + \cdots + (n-3)/8 - 1, \end{cases} \tag{B9a}$$

so finally

$$\sum_{n/3 \leqslant i_1 \leqslant (n-1)/2} \overline{\mathscr{T}}^{(ni_1)} = \frac{(n-3)^2}{192} + \frac{n-3}{8} + 1. \tag{B10}$$

In an analogous way for $i_1 > i_{10}$ one obtains, by using (B4a) and (B5) [l = even in (B2a)] $\overline{\mathscr{T}}^{(ni_1)} = (n - i_1 + 2)/4$ and

$$\sum_{(n-1)/2+2 \leqslant i_1 \leqslant n-2} \overline{\mathscr{T}}^{(ni_1)} = 2\Sigma^{\pi_3}$$

with

$$\Sigma^{\pi_3} = 1 + 2 + 3 + \cdots + \frac{n-3}{8} = \frac{n-3}{16} + \frac{(n-3)^2}{128}. \tag{B11}$$

Notice that, for $n = 3$, $\Sigma^{\pi_1} = 1$, $\Sigma^{\pi_2} = \Sigma^{\pi_3} = 0$, so $\mathscr{C}_{(3)} = 1$. For l = odd integer, analogous formulas to (B9)–(B11) hold. Insertion of (B10) and (B11) into (B7) yields

$$\overline{\mathscr{T}}_n = (n-3)^2/48 + (n-3)/4 + 1, \tag{B12}$$

and the same result is obtained for l = odd.

(b) For the cases $n = 12l + 3 + 2g$ with $1 \leqslant g \leqslant 5$, we use a recurrence procedure. We suppose that $\overline{\mathscr{T}}_{n-2}$ is calculated for $g = g' - 1$ and given by (B2) and find it for $g = g'$.

Let us call i_1^0, i_2^0, i_3^0 the corresponding odd numbers when $g = g' - 1$ and satisfying

$$i_1^0 \geqslant i_2^0 \geqslant i_3^0, \quad \sum_{l=1}^{3} i_l^0 = \bar{n} + 2(g'-1).$$

The new triplets (i_1, i_2, i_3), $\Sigma_{j=1}^3\, i_j = n$, also satisfy $i_1 \geqslant i_2 \geqslant i_3$ and

$$i_1 = i_1^0 + 2. \tag{B13}$$

Now let us define $\hat{i}_{10}$ by

$$\hat{i}_{10}=\begin{cases}(n+2g'-1)/2, & \text{if } g'=\text{even},\\ (n+2g'-1)/2, & \text{if } g'=\text{odd}.\end{cases} \tag{B14}$$

Then we can verify that if $i_1>\hat{i}_{10}$, in view of (B13),

$$\sum_{i_{10}+2<i_1<i_{1\,\max}} \mathscr{T}^{(ni_1)}=\sum_{i_{10}<i_1^0<i_{1\,\max}-2} \mathscr{T}^{(n-2,i_1^0)}, \tag{B15}$$

because $\mathscr{T}_{n}^{(ni_1)}=\mathscr{T}_{n-2}^{(n-2,i_1^0)}$ in the corresponding ranges of i_1 and i_1^0.

On the other hand, for $i_1\leqslant\hat{i}_{10}$ we find

$$\mathscr{T}^{(ni_1)}=\mathscr{T}^{(n-2,i_1^0)}+1, \quad \text{when } i_{1\,\min}<i_1<\hat{i}_{10} \quad (\text{resp. } i^0_{1\,\min}<i_1^0<\hat{i}_{10}-2), \tag{B16}$$

$$\text{with } i_{1\,\min}=\begin{cases}(\bar{n}+6)/3, & \text{if } g'<3\\ (\bar{n}+12)/3, & \text{if } g'>3\end{cases} \left(\text{resp. } i^0_{1\,\min}=\begin{cases}(\bar{n}+6)/3, & \text{if } g'-1<3\\ (\bar{n}+12)/3, & \text{if } g'-1>3\end{cases}\right). \tag{B16a}$$

We conclude that an additional set of partitions $w_n(I)$ appears relative to the set of partitions $w_{n-2}(I^0)$, only when $i_1<i_{10}$. The number of these supplementary partitions equals exactly

$$\sum_{i_{1\,\min}<i_1<\hat{i}_{10}} \mathscr{T}^{(ni_1)}-\sum_{i^0_{1\,\min}<i_1^0<\hat{i}_{10}-2} \mathscr{T}^{(n-2,i_1^0)}$$

$$=\frac{\hat{i}_{10}-i_{1\,\min}}{2}+1. \tag{B17}$$

Finally from (B15) and (B17) we obtain

$$\mathscr{T}_n=\mathscr{T}_{n-2}+\frac{\hat{i}_{10}-i_{1\,\min}}{2}+1. \tag{B18}$$

Insertion of the values (B15) and (B16a) into (B18) yields the numbers $\mathscr{T}_n$ of Lemma B.1 for $n=\bar{n}+2$ and $n=\bar{n}+2(g+1)$, $1\leqslant g\leqslant 4$. Q.E.D.

Remark: We notice that a more generalized formula can be derived, which concerns the number of terms of every ordered *partial* sum of C_0^{n+1} when i_1 varies from the minimal value $n/3$ till a fixed $\bar{i}_1\leqslant n-2$: $n/3\leqslant i_1\leqslant\bar{i}_1$. This formula will be essentially useful for the proof of the absolute bounds of every $\beta^n_{i_1i_2i_3}$.

From Eq. (B7) we write (using the notation $\mathscr{T}^{(n)}_{\bar{i}_1}$ for the above number)

$$\mathscr{T}^{(n)}_{\bar{i}_1}=\sum_{i_1=n/3}^{\bar{i}_1} \mathscr{T}^{(ni_1)}.$$

(a) Let $\bar{i}_1\leqslant(n-1)/2$. Then by (B4b), (B8a), and (B8b) we find two arithmetic progressions

$$\mathscr{T}^{(n)}_{\bar{i}_1}=\Sigma^{\pi_1(\bar{i}_1)}+\Sigma^{\pi_2(\bar{i}_1)},$$

with

$$\Sigma^{\pi_1(\bar{i}_1)}=\tfrac{1}{24}(3\bar{i}_1-n+12)(1+(3\bar{i}_1-n)/4+1)$$
$$=\tfrac{1}{96}(3\bar{i}_1-n+12)(3\bar{i}_1-n+8)$$

and

$$\Sigma^{\pi_2(\bar{i}_1)}=\tfrac{1}{24}(3\bar{i}_1-n+8)(2+(3\bar{i}_1-n-2)/4+1)$$
$$=\tfrac{1}{96}(3\bar{i}_1-n+8)(3\bar{i}_1-n+12),$$

so that

$$\mathscr{T}^{(n)}_{\bar{i}_1}=\tfrac{1}{48}(3\bar{i}_1-n+8)(3\bar{i}_1-n+12), \quad \bar{i}_1\leqslant(n-1)/2. \tag{B19}$$

(b) Let $i_1\geqslant(n-1)/2$. Then, using (B10) and (B4a),

$$\mathscr{T}^{(n)}_{\bar{i}_1}=\frac{(n-3)^2}{192}+\frac{n-3}{8}+1+\Sigma^{\pi_3(\bar{i}_1)}$$
$$=\frac{(n-3)^2}{192}+\frac{n-3}{8}+1$$
$$+\frac{(2\bar{i}_1-n-5)(3n-2\bar{i}_1-5)}{32}. \tag{B20}$$

APPENDIX C: RESULTS OF THE Φ ITERATION UP TO $\nu=3$

In this appendix we show that at the first three orders of the Φ iteration (see Sec. II) and $\forall i\leqslant n\leqslant 9$, $0<\Lambda\leqslant 0.01$ there exist well defined positive constants (continuous functions of the coupling constant Λ): $\delta_{\nu,n}(\Lambda)$, $\hat{\delta}_{\nu,n}(\Lambda)$, $\hat{\epsilon}_{\nu,n}(\Lambda)$, $\epsilon_{\nu,n}(\Lambda)$, $\varphi_{\nu,n}(\Lambda)$, $d_{\nu,n}(\Lambda)$, $\tilde{\varphi}_{\nu,n}(\Lambda)$ such that the following properties are verified: Let

$$\Phi_\nu=\prod_{l=1}^{m} H_\nu^{i_l+1} \quad (i_l\leqslant 9).$$

Then

$$(-1)^{(n-1)/2}s_\Phi H_\nu^{n+1}(q_{\bar{I}}=0,\Phi_\nu,\Lambda)\geqslant 0,$$
$$H_\nu^2(q,\Lambda)\Delta_F(q)>1,$$

with

$$s_\Phi=\prod_{l=1}^{m}(-1)^{(i_l-1)/2}\int H_\nu^4(q(q_{\bar{I}},k_{\bar{J}}))\prod_i \Delta_F(\tilde{l}_i)dk_{\bar{J}}$$
$$<0 \quad (\text{signs}); \tag{C1}$$

$\forall n\geqslant 3$, $\exists$ constant $1>\delta_0(\Lambda)$, $\gamma_0(\Lambda)>0$,

$$|N_2H_\nu^{n+1}|_{q_{\bar{I}}=0}\leqslant\delta_{\nu,n}(\Lambda)|\Delta_\nu^{n+1}|_{q_{\bar{I}}=0},$$
$$\gamma_0(\Lambda)\leqslant\delta_{\nu,n}(\Lambda)\leqslant\delta_0(\Lambda),$$
$$\lim_{\Lambda\to 0}(\delta_{\nu,n}(\Lambda)/\Lambda)=\text{const}>0 \quad (\text{simple splitting}); \tag{C2}$$

$\forall n\geqslant 3$, $\exists$ constant $1>\hat{\delta}_0(\Lambda)>\hat{\gamma}_0(\Lambda)>0$,

$$\Lambda|N_3H_\nu^{n+1}|_{q_{\bar{I}}=0}\leqslant\hat{\delta}_{\nu,n}(\Lambda)|C_\nu^{n-1}|_{q_{\bar{I}}=0},$$
$$\hat{\gamma}_0(\Lambda)\leqslant\hat{\delta}_{\nu,n}(\Lambda)\leqslant\hat{\delta}_0(\Lambda),$$
$$\lim_{\Lambda^2\to 0}(\hat{\delta}_{\nu,n}(\Lambda)/\Lambda^2)=\text{const}>0 \quad (\text{double splitting}); \tag{C3}$$

$$\forall n \geqslant 1,\ \nu \geqslant \overline{\nu_0}(n),\quad \exists 0 \leqslant \hat{\epsilon}_{\nu,n}(\Lambda),\quad \epsilon_{\nu,n}(\Lambda) < 1,$$

$$|H_\nu^{n+1}\tilde{|}_{q_{\bar{I}}=0} \leqslant (1 + \hat{\epsilon}_{\nu,n}(\Lambda))|\tilde{H}_{\nu-1}^{n+1}\tilde{|}_{q_{\bar{I}}=0},$$

$$|H_\nu^{n+1}|_{q_{\bar{I}}=0} \geqslant (1 - \epsilon_{\nu,n}(\Lambda))|H_{\nu-1}^{n+1}|_{q_{\bar{I}}=0}$$

$$\text{(convergence properties);} \qquad \text{(C4)}$$

$$\forall n \geqslant 1,\quad \nu \geqslant \overline{\nu_0}(n),\quad \exists \varphi_{\nu,n}(\Lambda) \geqslant 0,$$

$$|H_\nu^{n+1}\tilde{|} \leqslant (1 + \varphi_{\nu,n}(\Lambda))|H^{n+1}\tilde{|}_{q_{\bar{I}}=0}$$

$$\text{(zero momentum dominance);} \qquad \text{(C5)}$$

$$\forall n \geqslant 3,\quad \exists 0 < d_n(\Lambda) < 1,$$

$$|B_\nu^{n+1}\tilde{|}_{q_{\bar{I}}=0} \leqslant d_{\nu,n}(\Lambda)|\tilde{C}_\nu^{n+1}|_{q_{\bar{I}}=0} \quad \text{(tree dominance);} \qquad \text{(C6)}$$

$$\forall n \geqslant 1,\ \nu \geqslant \nu_0(n),\quad \exists \tilde{\varphi}_{\nu,n}(\Lambda) \geqslant 0,$$

$$|H_\nu^{n+1}|_{q=0} \leqslant (1 + \tilde{\varphi}_{\nu,n}(\Lambda))c_{n+1}(n+1)!\Lambda^{(n-1)/2}$$

$$\times [c_{n+1} \text{ as in (3.2b)}] \quad \text{(norms).} \qquad \text{(C7)}$$

When the calculations are trivial we give only the results:

$$\nu = 0,\quad H_0^2 = q^2 + 1,\quad H_0^{n+1} = 0,\quad \forall n \geqslant 3;$$

$$\nu = 1,\quad H_1^2(q)\Delta_F(q) = 1,\quad H_1^4(q,\Lambda) = -6\Lambda,$$

$$H_1^{n+1}(q,\Lambda) = 0,\quad \forall n \geqslant 5. \qquad \text{(C8)}$$

By (C8), properties (C1), (C5), (C6), and (C7) are trivially satisfied. Moreover

$$|N_2 H_1^4(q_{\bar{I}},\Phi_0,\Lambda)\tilde{|}_{q_{\bar{I}}=0}$$

$$\leqslant 6\Lambda\pi \left| \int \prod_{l=1}^{2} H_1^2(q_l)\Delta_F(q_l(q_{\bar{I}}k_{\bar{J}})) \times \Phi_0(H) \prod \Delta_F(\tilde{l}_i)dk_{\bar{J}} \right|_{q_{\bar{I}}=0}, \qquad \text{(C9)}$$

$$\Lambda|N_3 H_1^4(q_{\bar{I}},\Phi_0,\Lambda)\tilde{|}_{q_{\bar{I}}=0}$$

$$\leqslant 6\Lambda^2 c_0 \left| \int \Phi_0(H) \prod \Delta_F(\tilde{l}_i)dk_{\bar{J}} \right|_{q_{\bar{I}}=0}. \qquad \text{(C10)}$$

The two last inequalities imply that (C2) and (C3) are also satisfied with

$$\delta_{1,3}(\Lambda) = 6\Lambda\pi, \qquad \text{(C9a)}$$

$$\hat{\delta}_{1,3}(\Lambda) = 6\Lambda^2 c_0. \qquad \text{(C10a)}$$

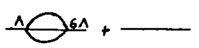
Here

$$c_0 = \int \prod \Delta_F(l_i)dk_1\, dk_2 \Big|_{q=0} \approx 47.3.$$

1. ν=2

(2.a) $n = 1$:

$$H_2^2(q,\Lambda) = 6\Lambda^2 \int \prod \Delta_F(l_i)dk_1\, dk_2 + q^2 + 1, \qquad \text{(C11)}$$



We obtain

$$H_2^2(q,\Lambda)\Delta_F(q) > 1, \qquad \text{(C12)}$$

$$\Phi H_2^2(q_{\bar{I}},\Phi_1,\Lambda)\tilde{|}_{q_{\bar{I}}=0} > |H_1^2(q_{\bar{I}},\Phi_1,\Lambda)\tilde{|}_{q_{\bar{I}}=0}, \qquad \text{(C13)}$$

$$|H_2^2(q_{\bar{I}},\Phi_1,\Lambda)|_{q_{\bar{I}}=0} \leqslant (1 + 6\Lambda^2 c_0)|\tilde{H}_1^2(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0}, \qquad \text{(C14)}$$

$$|H_2^2(q_{\bar{I}},\Phi_1,\Lambda)\tilde{|} \leqslant |\tilde{H}_2^2(q_{\bar{I}},\Phi_1,\Lambda)|_{q_{\bar{I}}=0}, \qquad \text{(C15)}$$

$$|H_2^2(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant 1 + 6\Lambda^2 c_0. \qquad \text{(C16)}$$

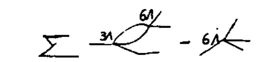
From the above inequalities properties (C1), (C4), (C5), and (C7) are satisfied with

$$\hat{\epsilon}_{2,1}(\Lambda) = 6\Lambda^2 c_0,\quad \epsilon_{2,1}(\Lambda) = 0,\quad \varphi_{2,1}(\Lambda) = 0,$$

$$\tilde{\varphi}_{2,1}(\Lambda) = 6\Lambda^2 c_0,\quad \Lambda \lesssim 0.5. \qquad \text{(C17)}$$

(2.b) $n = 3$: Using definitions (2.14) and (2.15), we have

$$H_2^4(q,\Lambda) = 18\Lambda^2 \sum_{w(J)} \prod_i \Delta_F(l_i(q_{j_i},k))dk - 6\Lambda, \qquad \text{(C18)}$$



We obtain

$$\int H_2^4(q,\Lambda) \prod_i \Delta_F(\tilde{l}_i)dk_{\bar{J}} < 0,\quad \text{if } \Lambda < \frac{1}{9\pi}, \qquad \text{(C19)}$$

$$|N_2 H_2^4(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0}$$

$$\leqslant \delta_{2,3}(\Lambda) \left| \int \prod_{l=1}^{2} H_2^2(q_{i_l}(q_{\bar{I}},k_{\bar{J}}))\Delta_F(Q_{I_l}) \times \Phi_1(H) \prod \Delta_F(\tilde{l}_i)dk_{\bar{J}} \right|_{q_{\bar{I}}=0},$$

with $\delta_{2,3}(\Lambda) = 6\pi\Lambda = \delta_{1,3}(\Lambda)$, (C20)

$$\Lambda|N_3 H_2^4(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0}$$

$$\leqslant 6\Lambda^2 c_0 \left| \int \Phi_1(H) \prod_i \Delta_F(\tilde{l}_i)dk_{\bar{J}} \right|_{q_{\bar{I}}=0}, \qquad \text{(C21)}$$

so $\hat{\delta}_{2,3}(\Lambda) = \hat{\delta}_{1,3}(\Lambda) = 6\Lambda^2 c_0$, with $\Lambda < 1/9\pi$.

The above inequalities prove (C1)–(C3) for $H_2^4(q,\Lambda)$. Moreover,

$$|H_2^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant |H_1^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0}, \qquad \text{(C22)}$$

$$|H_2^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \geqslant (1 - 9\pi\Lambda)|H_1^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0}, \qquad \text{(C23)}$$

$$|H_2^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|} \leqslant \left(1 + \frac{18\pi\Lambda}{1 - 9\pi\Lambda}\right)|H_2^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0}, \qquad \text{(C24)}$$

$$|B_1(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant 9\pi\Lambda|C_1^4(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0}. \qquad \text{(C25)}$$

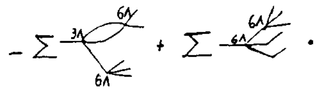
The last inequalities ensure properties (C4)–(C7) with

$$\hat{\epsilon}_{2,3}(\Lambda) = 0,\quad \epsilon_{2,3}(\Lambda) = d_{2,3}(\Lambda) = 9\pi\Lambda,$$

$$\varphi_{2,3}(\Lambda) = 18\pi\Lambda/(1 - 9\pi\Lambda),\quad \tilde{\varphi}_{2,3}(\Lambda) = 0.$$

(2.c) $n=5$: By definitions (2.14) and (2.15), we have

$$H_2^6(q,\Lambda) = -3\Lambda \sum_{w(J)} (6\Lambda)^2 \int \prod_{i=1}^{2} \Delta_F(l_i) dk \Delta_F(Q_{J_1}) + (6\Lambda)^2 \sum_{w(I)} \prod_{l=1}^{3} \Delta_F(Q_{I_l}), \quad \text{(C26)}$$



We obtain

$$s_F \widetilde{H}_2^6(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} > 0, \quad \text{if } \Lambda \leqslant 3\pi \quad \text{(C27)}$$

[notice that under the same condition in this particular order ($\nu=2$),

$H_2^6(q,\Lambda) > 0, \quad \forall q \in \epsilon_{(q)}^{10}$],

$$|N_2 H_2^6(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} \leqslant \delta_{2,5}(\Lambda) \sum_{w(J)} \left| \int H_2^4 H_2^2 \Phi_1(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0},$$

$$\text{with } \delta_{2,5}(\Lambda) \equiv \tfrac{3}{2}\delta_{1,4}(\Lambda)/[1-d_{2,3}(\Lambda)], \quad \text{(C28)}$$

$$\Lambda |N_3 H_2^6(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant \hat{\delta}_{2,5}(\Lambda) 6\Lambda \left| \int \prod_{l=1}^{3} H_2^2(q_{i_l},(q_{\bar{I}}k_{\bar{J}})) \times \Delta_F(q_{i_l}) \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0},$$

$$\text{with } \hat{\delta}_{2,5}(\Lambda) = \tfrac{2}{3}\delta_{2,3}(\Lambda)\delta_{2,5}(\Lambda). \quad \text{(C29)}$$

Inequalities (C27)–(C29) prove the sign properties simple and double "splitting" [(C1)–(C3)]. Moreover,

$$|H_2^6(q_{\bar{I}},\Phi,\Lambda)\tilde{|} \leqslant (1+\varphi_{2,5}(\Lambda)) |H_2^6(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0},$$
$$\text{with } \varphi_{2,5}(\Lambda) = 6\pi\Lambda/(1-6\pi\Lambda). \quad \text{(C30)}$$

So the zero momentum dominance is satisfied and

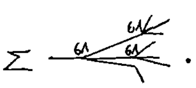
$$|B_1^6\tilde{|}_{q_{\bar{I}}=0} \leqslant 3\Lambda\pi |C_1^6\tilde{|}_{q_{\bar{I}}=0}$$
$$[\text{tree dominance with } d_{2,5}(\Lambda) = 3\Lambda\pi], \quad \text{(C31)}$$

$$|H_2^6(q,\Lambda)|_{q=0} \leqslant 6!\Lambda^2/2. \quad \text{(C32)}$$

The last inequality proves the norm property. Q.E.D.

(2.d) $n=7$: From definitions (2.14) and (2.15), we obtain

$$H_2^8(q,\Lambda) = -(6\Lambda)^3 \sum_{w(I)} \Delta_F(Q_{I_1}) \Delta_F(Q_{I_2}), \quad \text{(C33)}$$



From (C33) we obtain successively

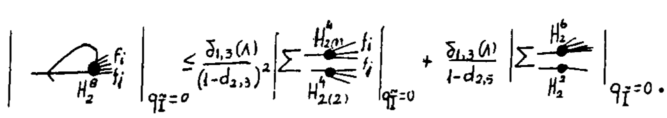
$$s_F \widetilde{H}_2^8(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} < 0 \quad \text{(signs)}, \quad \text{(C34)}$$

$$|N_2 H_2^8(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant \frac{\delta_{1,3}(\Lambda)}{(1-d_{2,3}(\Lambda))^2} \sum_{w(I)} \left| \int \prod_{l=1}^{2} H_{2(l)}^4 \Phi(H) \times \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0} + \frac{\delta_{1,3}(\Lambda)}{1-d_{2,5}(\Lambda)} \times \sum_{w(I)} \left| \int H_2^6 H_2^2 \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0}, \quad \text{(C35)}$$

which is the simple splitting property with

$$\delta_{2,7}(\Lambda) = \frac{\delta_{1,3}(\Lambda)}{(1-d_{2,3}(\Lambda))^2} \quad \left(\text{notice } \frac{1}{1-d_{2,3}} \geqslant \frac{1}{1-d_{2,5}}\right). \quad \text{(C35a)}$$

In the diagram below we show graphically the simple splitting property for H_2^8 and denote by f_i, f_j "internal" lines of the F-convolution loops:



In an analogous way the double splitting property for H_2^8 holds:

$$\Lambda |N_3 H_2^8(q_{\bar{I}},\Phi,\Lambda)\tilde{|}_{q_{\bar{I}}=0} \leqslant \hat{\delta}_{2,7}(\Lambda) \left| 6\Lambda \sum_{w(I)} \int H_2^4 H_2^2 H_2^2 \Phi(H) \times \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0},$$

$$\text{with } \hat{\delta}_{2,7}(\Lambda) = 7\Lambda\delta_{1,3}(\Lambda)/[1-d_{2,3}(\Lambda)]. \quad \text{(C36)}$$

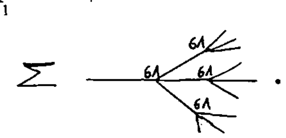
The remaining properties (zero momentum and tree dominance) hold trivially with $\varphi_{2,7} = d_{2,7} = 0$. The norm property finally is valid also trivially by (C33) precisely:

$$|H_2^8(q,\Lambda)| \leqslant \Lambda^3 8! \tfrac{3}{8} \quad [\text{i.e., } \tilde{\varphi}_{2,7}(\Lambda) = 0]. \quad \text{(C37)}$$

Q.E.D.

(2.e) $n=9$: Definitions (2.14) and (2.15) yield

$$H_2^{10} = (6\Lambda)^4 \sum_{w(I)} \prod_{l=1}^{3} \Delta_F(Q_{I_l}), \quad \text{(C38)}$$



From (C38) we obtain the signs

$$s_F \widetilde{H}_2^{10}(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} > 0; \quad \text{(C39)}$$

the simple splitting property,

$$|N_2 H_2^{10}(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} \leqslant \frac{\delta_{1,3}(\Lambda)}{4(1-d_{2,5}(\Lambda))(1-\epsilon_{2,3}(\Lambda))} \times \left\{ \sum_{w(I_1)} \left| \int H_2^6 H_2^4 \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0} + \sum_{w(I_2)} \left| \int H_2^8 H_2^2 \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0} \right\}, \quad \text{(C40)}$$

with splitting constant

$$\delta_{2,9}(\Lambda) \equiv \delta_{1,3}(\Lambda)/4(1 - d_{2,5}(\Lambda))(1 - \epsilon_{2,3}(\Lambda)); \tag{C40a}$$

the double splitting property

$$\Lambda|N_3 H_2^{10}(q_{\bar{I}},\Phi,\Lambda)\bar{|}_{q_{\bar{I}}=0} \leqslant \frac{4\hat{\delta}_3(\Lambda)}{(1 - d_{2,3}(\Lambda))^2} 6\Lambda \left\{ \sum_{w(\bar{I})} \left| \int H^4_{2(1)} H^4_{2(2)} H^2_2 \times \Phi(H) \prod_i \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0} + \sum \left| \int H^6 H^2_{2(1)} \times H^2_{2(2)} \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0} \right\}, \tag{C41}$$

with constant corresponding constant

$$\hat{\delta}_{2,9}(\Lambda) \equiv 4\hat{\delta}_3(\Lambda)/(1 - d_{2,3}(\Lambda))^2; \tag{C41a}$$

the zero momentum and tree dominance (trivially satisfied)

$$\varphi_{2,9} = d_{2,9}(\Lambda) = 0;$$

and the norms

$$|H_2^{10}(q,\Lambda)|_{q=0} \leqslant \Lambda^4 10!/10, \tag{C42}$$

with $\tilde{\varphi}_{2,9}(\Lambda) = 0$. Q.E.D.

(2.f) $\forall n \geqslant 11, H_2^{n+1}(q,\Lambda) = 0$.

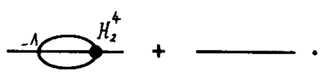
2. $\nu=3$

For this order we have proved all the above properties and calculated the analogous constants exactly for every $1 \leqslant n \leqslant 27$ [notice that $H_3^{n+1}(q,\Lambda) = 0, \forall n \geqslant 29$].

We only present here the results concerning $H_3^2(q,\Lambda)$, $H_3^4(q,\Lambda)$, in order to make evident the "stationarity" of the corresponding constants during the Φ iteration.

(3.a) $n = 1$: From definitions (2.14) and (2.15), we have

$$H_3^2 = -\Lambda \int H_2^4(q,k) \prod_{l=1}^{3} \Delta_F(l_i) dk_1\, dk_2 + q^2 + 1, \tag{C43}$$



We first note that, in view of (C19),

$$H_3^2(q,\Lambda)\Delta_F(q) > 1. \tag{C44}$$

Using property (C23) inside (C43) we can write

$$|H_3^2(q_I,\Phi_2,\Lambda)\bar{|}_{q_{\bar{I}}=0} \geqslant (1 - \epsilon_{2,3}(\Lambda))|\tilde{H}_2^2(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} + \epsilon_{2,3}(\Lambda) \left| \int (q^2 + 1)(q_{\bar{I}},k_{\bar{J}}) \times \Phi(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0}. \tag{C45}$$

Taking into account the definition $H_1^2 = q^2 + 1$ and the property (C14) in the second term on the rhs of (C44) yields

$$|H_3^2(q_I,\Phi_2,\Lambda)|_{q_{\bar{I}}=0} \geqslant (1 - \epsilon_{3,1}(\Lambda^3))|\tilde{H}_2^2(q_{\bar{I}},\Phi,\Lambda)|_{q_I} = 0. \tag{C46}$$

Here we have defined

$$\epsilon_{3,1}(\Lambda) \equiv \epsilon_{2,3}(\Lambda)\hat{\epsilon}_{2,1}(\Lambda^2) \approx 5400\Lambda^3. \tag{C46a}$$

On the other hand, by (C22) we also obtain from (C43)

$$|\tilde{H}_3^2(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0} \leqslant |\tilde{H}_2^2(q_{\bar{I}},\Phi_2,\Lambda)|_{q_{\bar{I}}=0}, \tag{C47}$$

which implies

$$\hat{\epsilon}_{3,1}(\Lambda) = 0. \tag{C47a}$$

It is worthwhile to compare $\hat{\epsilon}_{3,1}(\Lambda)$ with $\hat{\epsilon}_{2,1}(\Lambda)$: there is an indication of strong convergence of the Φ iteration.

Moreover, using (C24) and (C14) inside (C43), we obtain, for the zero momentum dominance property of H_3^2,

$$|H_3^2(q_{\bar{I}},\Phi_2,\Lambda)\bar{|} \leqslant (1 + \varphi_{2,3}(\Lambda)\hat{\epsilon}_{2,1}(\Lambda))|H_3^2(q_{\bar{I}},\Phi,\Lambda)|_{q_{\bar{I}}=0}, \tag{C48}$$

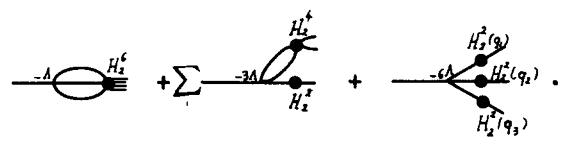
which implies

$$\varphi_{3,1}(\Lambda) \equiv [18\pi\Lambda/(1 - 9\pi\Lambda)]6\Lambda^2 c_0. \tag{C48a}$$

Q.E.D.

(3.b) $n = 3$: By (2.14) and (2.15) and using (C19), we now obtain

$$H_3^4 = -\Lambda \int H_2^6 \prod \Delta_F(l_i) dk_1\, dk_2 + 3\Lambda \sum_{w(J)} \int |H_2^4(q_{j_2}k)| \prod \Delta_F(l_i) dk \times H^2(g_{j_1}\Lambda)\Delta_F(g_{j_1}) - 6\Lambda \prod_{l=1}^{3} H_2^2(q_l)\Delta_F(q_l), \tag{C49}$$



Taking into account the sign properties (C27) [more precisely the positivity of $H_2^6(q,\Lambda)$—see the corresponding footnote—for all $q \in \epsilon_{(q)}^{10}$] and (C12) for H_2^6] and H_2^2, respectively, together with the bounds (C20), (C22), and (C24) for the second term on the rhs

$$\int H_3^4(q(q_{\bar{I}},k_{\bar{J}}),\Lambda) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} < 0, \quad \text{if } \Lambda < 1/9\pi. \tag{C50}$$

In other words, the negative sign of H_3^4 and every simple convolution integral (integration only with free propagators) at the third order of the Φ iteration is ensured under the same condition imposed on the coupling constant as we have found at the second order [see (C19)].

Moreover, application of signs and simple double splitting properties (at zero external momenta) of H_2^4, H_2^6, respectively, yields

$$|N_2 H_3^4\bar{|}_{q_{\bar{I}}=0} \leqslant \delta_{3,3}(\Lambda) \left| \int \prod_{l=1}^{2} H_3^2(q_l(q_{\bar{I}}k_{\bar{J}}))\Delta_F(q_l) \times \Phi_2(H) \prod \Delta_F(\tilde{l}_i) dk_{\bar{J}} \right|_{q_{\bar{I}}=0},$$

with $\delta_{3,3}(\Lambda) \equiv 6\Lambda\pi(1 - 9\pi)/(1 - \epsilon_{3,1}(\Lambda))^2$. (C51)

One verifies that $\delta_{3,3}(\Lambda) < \delta_{2,3}(\Lambda)$ [using (C20), (C46), and (C46a)], and

$$\Lambda |N_3 H_3^4 \bar{]}_{q_{\bar{l}}=0} \leqslant \hat{\delta}_{3,3}(\Lambda) \left| \int \Phi_2(H) \prod \Delta_F(\tilde{l}_i) dk_J \right|_{q_{\bar{l}}=0},$$

$$\text{with } \hat{\delta}_{3,3}(\Lambda) \equiv 6\Lambda^2 c_0 (1 - 9\Lambda\pi). \quad \text{(C52)}$$

One can also verify [using (C21)] that $\hat{\delta}_{3,3}(\Lambda) < \hat{\delta}_{2,3}(\Lambda)$. Q.E.D.

To prove properties (C4) we proceed in an analogous way [using the second-order double and simple splitting properties together with the signs of H_2^6, H_2^4, and the relative bounds (C13), (C14), and (C22), (C23)].

$$|H_3^4(q_{\bar{l}}, \Phi_2, \Lambda)\bar{]}_{q_{\bar{l}}=0} \leqslant (1 + \hat{\epsilon}_{3,3}(\Lambda^2)) |H_2^4(q_{\bar{l}}, \Phi_2, \Lambda)|_{q_{\bar{l}}=0}, \quad \text{(C53)}$$

where

$$\hat{\epsilon}_{3,3}(\Lambda^2) \equiv (1 + 6\Lambda^2 c_0)^3 / (1 - 9\pi\Lambda) - 1 \quad \text{(C53b)}$$

and

$$|H_3^4(q_{\bar{l}}, \Phi_2, \Lambda)\bar{]}_{q_{\bar{l}}=0} \geqslant |H_2^4(q_{\bar{l}}, \Phi_2, \Lambda)\bar{]}_{q_{\bar{l}}=0}, \quad \text{(C53b)}$$

i.e., $\epsilon_{3,3}(\Lambda) = 0$.

Finally insertion of the sign properties and (C22), (C23), (C25), (C20), (C29), and (C30) into (C49) yields the zero momentum dominance, explicitly

$$|H_3^4(q_{\bar{l}}, \Phi_2, \Lambda)\bar{]} \leqslant (1 + \varphi_{3,3}(\Lambda)) |H_3^4(q_{\bar{l}}, \Phi, \Lambda)|_{q_{\bar{l}}=0}, \quad \text{(C54)}$$

with

$$\varphi_{3,3}(\Lambda) = [\varphi_{2,5} + \hat{\delta}_{2,5}(\Lambda)] / (1 - 9\pi\Lambda) < \varphi_{2,3}(\Lambda); \quad \text{(C54a)}$$

the tree dominance property

$$|B_2^4(q_{\bar{l}}, F_2 \Lambda)|_{q_{\bar{l}}=0} \leqslant 9\pi\Lambda |C_2^4(q_{\bar{l}} F_2, \Lambda)|_{q_{\bar{l}}=0}, \quad \text{(C55)}$$

$$\text{i.e., } d_{3,3}(\Lambda) = d_{2,3}(\Lambda);$$

the norm property

$$|H_3^4(q, \Lambda)| \leqslant 6\Lambda (1 + \tilde{\varphi}_{3,3}(\Lambda)), \quad \text{(C56)}$$

$$\text{with } \tilde{\varphi}_{3,3}(\Lambda) = \hat{\delta}_{2,5}(\Lambda) + \varphi_{2,5} + 18\Lambda^2 c_0. \quad \text{(C56a)}$$

Q.E.D.

APPENDIX D: PROOF OF PROPERTIES (2.12d1) OF PROPOSITION 2.3

The properties are

$$\beta_{n(2)} - \beta_{n(1)} \geqslant 0, \quad \text{(p1)}$$

$$\delta_{n(1)} \beta_{n(1)} - \delta_{n(2)} \beta_{n(2)} \geqslant 0, \quad \text{(p2)}$$

$$\delta_{n(1)} \beta_{n(1)} - \delta_{n(2)} \beta_{n(2)} \geqslant \delta_{n-2(1)} \beta_{n-2(1)} - \delta_{n-2(2)} \beta_{n-2(2)}, \quad \text{(p3)}$$

$$\beta_{n-2(2)} / \beta_{n-2(1)} \geqslant \beta^n_{n-4,3,1(2)} / \beta^n_{n-4,3,1(1)}. \quad \text{(p4)}$$

The hypothesis is $\delta_{(1)} > \delta_{(2)} \in \mathscr{C}^{\mathscr{O}}_{\Lambda}$ and $0 < \Lambda \lesssim 0.01$.

All properties can be obtained recursively. We only present the argument for property (p4) (which is the most complicated), i.e., for a general partition (i_1, i_2, i_3),

$$\beta^{n-2}_{i_1, i_2 - 2, i_3(2)} / \beta^{n-2}_{i_1, i_2 - 2, i_3(1)} \geqslant \beta^n_{i_1, i_2, i_3(2)} / \beta^n_{i_1, i_2, i_3(1)}, \quad \text{(D1)}$$

using the simplest case of definitions (2.1), namely,

$$\beta^n_{i_1 i_2 i_3} = 1 + \frac{i_1(i_1 1)}{(i_2+1)(i_2+2)} \frac{\delta_{i_2+2}}{\delta_{i_1}} \frac{\beta_{i_2+2}}{\beta_{i_1}} \beta^n_{i_1 - 2, i_2 + 2, i_3}, \quad \text{(D2a)}$$

$$\beta^{n-2}_{i_1, i_2 - 2, i_3} = 1 + \frac{i_1(i_1 - 1)}{i_2(i_2 - 1)} \frac{\delta_{i_2}}{\delta_{i_1}} \frac{\beta_{i_2}}{\beta_{i_1}} \beta^{n-2}_{i_1 - 2, i_2, i_3}. \quad \text{(D2b)}$$

The remaining properties can be shown by analogy. Nevertheless we remark that it is essential for all proofs that the recurrence hypothesis contains all of them, i.e., a multiple recursion in which the system (2.12d1) of relations occurs simultaneously must be taken into account.

The first step of this recursion is verified by $n = 7$ and $n = 9$, i.e., $\forall 0 < \Lambda \lesssim 0.01$,

$$\beta_{7(2)} - \beta_{7(1)} \geqslant 0, \quad \beta_{9(2)} - \beta_{9(1)} \geqslant 0,$$
$$\delta_{9(1)} \beta_{9(1)} - \delta_{9(2)} \beta_{9(2)} \geqslant \delta_{7(1)} \beta_{7(1)} - \delta_{7(2)} \beta_{7(2)} \geqslant 0,$$
$$\beta_{7(2)} / \beta_{7(1)} \geqslant \beta^9_{531(2)} / \beta^9_{531(1)}, \quad \text{(D3)}$$

where

$$\beta^4_{531} = 1 + \frac{5}{36} \frac{\delta_3}{\delta_5}, \quad \beta_7 = 1 + \frac{5}{6} \frac{\delta_3}{\delta_5},$$
$$\beta_7 = 1 + 14 \frac{\delta_3}{\delta_7 \beta_7} \beta^9_{531}. \quad \text{(D3a)}$$

We then suppose that all properties (p1)–(p4) hold for every $7 \leqslant \bar{n} \leqslant n - 2$ and for $\bar{n} = n$, $\forall \bar{i}_1$ in $n/3 \leqslant \bar{i}_1 \leqslant i_1 - 2$. We then show (D1) for $\bar{n} = n, \bar{i}_1 = i_1$. (Recall $i_1 \geqslant i_2 \geqslant i_3$ always.)

Using the definitions (D2a) and (D2b), we write the inequality (D1) as follows:

$$R_n = \frac{i_1(i_1 - 1)}{i_2(i_2 - 1)} a - \frac{i_1(i_1 - 1)}{i_2(i_2 - 1)(i_2 + 1)(i_2 + 2)} c_0 c \geqslant 0, \quad \text{(D4)}$$

where

$$a \equiv \frac{\delta_{i_2(2)}}{\delta_{i_1(2)}} \frac{\beta_{i_2(2)}}{\beta_{i_1(2)}} \beta^{n-2}_{i_1 - 2, i_2, i_3(2)} - \frac{\delta_{i_2(1)}}{\delta_{i_1(1)}} \frac{\beta_{i_2(1)}}{\beta_{i_1(1)}} \beta^{n-2}_{i_1 - 2, i_2, i_3(1)}, \quad \text{(D5)}$$

$$b \equiv \frac{\delta_{i_2+2(2)}}{\delta_{i_1(2)}} \frac{\beta_{i_2+2(2)}}{\beta_{i_1(2)}} \beta^n_{i_1 - 2, i_2 + 2, i_3(2)} - \frac{\delta_{i_2+2(2)}}{\delta_{i_1(1)}} \frac{\beta_{i_2+2(2)}}{\beta_{i_1(1)}} \beta^n_{i_1 - 2, i_2 + 2, i_3(1)}, \quad \text{(D6)}$$

$$c_0 \equiv [\delta_{i_1(1)} \beta_{i_1(1)} \delta_{i_1(2)} \beta_{i_1(2)}]^{-1}, \quad \text{(D7)}$$

$$
\begin{aligned}
c \equiv{} & \overbrace{(\delta_{i_2+2(1)}\beta_{i_2+2(1)} - \delta_{i_2+2(2)}\beta_{i_2+2(2)})}^{A_1}\overbrace{\delta_{i_2(2)}\beta_{i_2(2)}}^{c_i}\overbrace{\beta^{n-2}_{i_1-2,i_2,i_3(2)}\beta^{n}_{i_1-2,i_2+2,i_3(1)}}^{B_2} \\
& - (\delta_{i_2(1)}\beta_{i_2(1)} - \delta_{i_2(2)}\beta_{i_2(2)})\beta^{n-2}_{i_1-2,i_3,i_3(1)}\beta^{n}_{i_1-2,i_2+2,i_3(2)}\delta_{i_2+2(2)}\beta_{i_2+2(2)} \\
& + (\beta^{n-2}_{i_1-2,i_2,i_3(2)}\beta^{n}_{i_1-2,i_2+2,i_3(1)} - \beta^{n-2}_{i_1-2,i_2,i_3(1)}\beta^{n}_{i_1-2,i_2,i_3(2)})\delta_{i_2+2(2)}\beta_{i_1+2(2)}\delta_{i_2(2)}\beta_{i_2(2)}.
\end{aligned}
\tag{D8}
$$

Let us first show that $c \geqslant 0$.

Using the recurrence hypothesis, we factorize out in the first two terms of (D7) the non-negative differences and, by the identity

$$A_1C_iB_2 - A_2B_1C_{i+2} \equiv (A_1 - A_2)C_iB_2 + (B_2 - B_{12})C_iA_2 - (C_{i+2} - C_i)A_2B_1,$$

we finally write (D7) as follows:

$$
\begin{aligned}
c \equiv{} & \{[(\delta_{i_2+2(1)}\beta_{i_2+2(1)} - \delta_{i_2+2(2)}\beta_{i_2+2(2)}) - (\delta_{i_2(1)}\beta_{i_2(1)} - \delta_{i_2(2)}\beta_{i_2(2)})]\delta_{i_2(2)}\beta_{i_2(2)}\beta^{n-2}_{i_1-2,i_2,i_3(2)}\beta^{n}_{i_1-2,i_2+2,i_3(1)} \\
& + (\beta^{n-2}_{i_1-2,i_2,i_3(2)}\beta^{n}_{i_1-2,i_2+2,i_3(1)} - \beta^{n-2}_{i_1-2,i_2,i_3(1)}\beta^{n}_{i_1-2,i_2+2,i_3(2)})(\delta_{i_2(1)}\beta_{i_2(1)} - \delta_{i_2(2)}\beta_{i_2(2)})\delta_{i_1(2)}\beta_{i_2(2)} \\
& - (\delta_{i_2+2(2)}\beta_{i_2+2(2)} - \delta_{i_2(2)} - \delta_{i_2(2)}\beta_{i_2(2)})(\delta_{i_2(1)}\beta_{i_2(1)} - \delta_{i_2(2)}\beta_{i_2(2)})\beta^{n-2}_{i_1-2,i_2,i_3(1)}\beta^{n}_{i_1-2,i_2+2,i_3(2)}\} \\
& + (\beta^{n-2}_{i_1-2,i_2,i_3(2)}\beta^{n}_{i_1-2,i_2+2,i_3(1)} - \beta^{n-2}_{i_1-2,i_2,i_3(1)}\beta^{n}_{i_1-2,i_2,i_3(2)})\delta_{i_2+2(2)}\beta_{i_2+2(2)}\delta_{i_2(2)}\beta_{i_2(2)}.
\end{aligned}
$$

We notice that the three terms inside the brackets are second-order differences. The first two of them have non-negative sign and only the third is negative (in view of the recurrence hypothesis). These considerations allow us to state that the sum of the bracket, even if eventually negative for small n, is dominated by the first-order difference outside the bracket, which is a non-negative quantity in view of the recurrence hypothesis. By this last remark we conclude that

$$c \geqslant 0. \tag{D9}$$

Q.E.D.

Insertion of this result inside (D4) yields that R_n is a non-negative quantity in view of the fact that the dominant term $\sim c_0 c$ [proportional to $i_1^2(i_1 - 1)^2/i_2(i_2 - 1)(i_2 + 1)(i_2 + 2)$] is a non-negative quantity. Q.E.D.

[1] M. Manolessou, "The Φ^4 equations of motion. I. A new constructive method: The Φ iteration," J. Math. Phys. **29**, 2092 (1988); "The Φ^4 equations of motion. II. The zero-, one-, and two-dimensional solutions," J. Math. Phys. **30**, 175 (1988).

[2] A. Voros (private communication).

[3] A. Wightman (private communication).

[4] M. Manolessou, Bielefeld University Research Center preprint BIBOS 52 a/ b/ c/, June, 1985.

[5] Ph. Blanchard and K. Osterwalder (private communication).

[6] M. Manolessou, Ann. Phys. (NY) **152**, 327 (1984).

Coupled nonlinear Schrödinger equations arising in the study of monomode step-index optical fibers

G. K. Newboult, D. F. Parker, and T. R. Faulkner
Department of Theoretical Mechanics, Nottingham University, Nottingham, NG7 2RD, England

(Received 27 July 1988; accepted for publication 30 November 1988)

In this paper it is shown that nonlinear propagation in a step-index, 'monomode' optical fiber is not generally governed by the nonlinear Schrödinger equation, even when the fiber is axially symmetric. It is governed by a coupled pair of nonlinear partial differential equations that includes the nonlinear Schrödinger equations only as a special case. Three simple types of solution to the coupled system are analyzed and the corresponding field patterns are interpreted. One case shows that, for uniform wavetrains, nonlinearity not only alters the phase speed but also causes the field pattern in an 'elliptically polarized' mode to rotate gradually about the fiber axis. The other two cases each allow the system to be reduced to a single nonlinear Schrödinger equation, so showing two distinct situations in which solutions have soliton properties.

I. INTRODUCTION

When signals are transmitted along a fiber as a series of light pulses, one of the limiting factors to the rate at which the data can be transmitted over long distances is the pulse distortion due to the dispersive effects of the medium. Hasegawa and Tappert[1] were the first to suggest that this effect could be balanced against the sharpening phenomenon due to the nonlinearity of the material, hence creating a soliton. However it was not until 1980 that Mollenauer, Stolen, and Gordon[2] made the first experimental observations of solitons in optical fibers and demonstrated conclusively many of the soliton properties. Much work has been undertaken to show mathematically how the two phenomena affect the propagation of waves in a fiber. Many authors[3–5] have incorporated a cubically nonlinear polarization into a one-dimensional model, and obtained the nonlinear Schrödinger (NSL) equation, which possesses single-soliton and multi-soliton solutions. Hasegawa and Kodama[6] and Potasek, Agrawal, and Pinault[5] have extended this to higher-order approximations and have obtained a third-order equation.

Several authors have worked on modelling a cubically nonlinear fiber in three space dimensions.[7,8] In these papers it was assumed that the electric field took the form

$$\mathbf{E}(\mathbf{r},t) = \mathbf{A}(\mathbf{r},t)e^{i(kz-\omega t)}$$

and that the dispersion relation was amplitude dependent. This allowed the expansion of the dispersion relation around the carrier frequency and some suitable wave number, to give the Fourier space equivalent of the operator equation that when operating on the amplitude $\mathbf{A}(r,t)$ was written in the form

$$\nabla_1^2\mathbf{A} + \frac{\partial^2\mathbf{A}}{\partial z^2} - q^2\mathbf{A} + 2iq\frac{\partial\mathbf{A}}{\partial z} + f(r)k_0^2\mathbf{A} + 2ia(r)\frac{\partial\mathbf{A}}{\partial t} + bf(r)\frac{\partial^2\mathbf{A}}{\partial t^2} + D|\mathbf{A}(r,t)|^2\mathbf{A} = 0, \quad (1.1)$$

where

$$k(\omega) = (\omega/c)n_1(\omega), \quad n_1(\mathbf{r},\omega) = n_1(\omega)f(\mathbf{r}),$$
$$b = k_0k_0'' + k_0'^2, \quad a(\mathbf{r}) = k_0k_0'f(\mathbf{r}), \quad D = 2(n_2/n_0)k_0^2.$$

Analysis usually assumes that $\mathbf{A}$ has fixed direction, with $\mathbf{A} = \mathbf{e}A(\mathbf{r},t)$, and either neglects transverse variations so that $A = A(z,t)$ or assumes that A is separable as $A = \Phi(x,y)\theta(z,t)$. This second assumption is, of course, incompatible with (1.1), but the equation is then replaced by its average taken over each cross section. Neither of the above assumptions satisfies both the field equations and the boundary conditions. Moreover, neither takes due account of the fact that a "single-mode" fiber really has two equivalent modes, since modes with orthogonal polarizations are degenerate, having the same dispersion relation. This follows from the widely known result that the axisymmetric TE_{01} and TM_{01} modes are not the modes to propagate at the lowest frequencies.

For birefringent fibers (which lack axial symmetry) Blow *et al.*[9] and Menyuk[10] develop approximate treatments that show that the interaction between two orthogonal linearly polarized modes having slightly differing phase speeds is governed by two coupled equations. However, the analysis we present in this paper shows that, even in cylindrically symmetric fibers, two coupled equations are required to describe propagation of a general signal.

In this paper we shall be concerned with a cladded, cylindrically symmetric fiber. For simplicity, we consider the specific example in which both the core and the cladding are taken to be homogeneous, isotropic, and nonlinear dielectrics with a discontinuity in a permittivity at the boundary $r = a$, between the two. The core occupies the region

$$0 \leqslant r \leqslant a, \quad 0 \leqslant \theta < 2\pi, \quad -\infty < z < \infty,$$

whilst, as is usual in studies of fibers for which the outer radius of the cladding is very much greater than a, the cladding is treated as extending to infinity. The nonlinearity will be assumed to be cubic and present in both the core and the cladding.

In Sec. II the basic equations and boundary conditions governing the fields are given and the form of the nonlinear-

 0022-2488/89/040930-07$02.50

ity used throughout the paper is stated. Since a multiple scales perturbation expansion of the fields will be used, appropriate stretched coordinates are defined. Sec. III contains the derivation of equations governing the nonlinear modulation of amplitude of the appropriate modes resulting from linear optical waveguide theory. Two independent complex amplitudes are required to describe these fields and as a consequence the modulation equations are found to be a pair of nonlinear, coupled, second-order, partial differential equations. These equations are, in fact, a special case of Eqs. (3) of Blow *et al.*[9] They show that, even in a cylindrically symmetric, isotropic fiber, the signals with different polarizations interact nonlinearly. It is shown that this system of partial differential equations contains the usual NLS equation as a special case, but clearly there are many situations in which solutions are not described by a single NLS equation. In Sec. IV, three special types of solution to the coupled system are analyzed and the corresponding fields are interpreted. These show that nonlinearity can induce double circular refraction, since generally an elliptically polarized wave field suffers gradual rotation as it travels along the fiber—the ellipse rotation phenomenon first predicted for plane waves in nonlinear isotropic media by Maker *et al.*[11] Also, solitons may exist for both circularly polarized and linearly polarized wave fields.

II. PROCEDURE

The fields in the core and the cladding are described by Maxwell's equations

$$\nabla\times\mathbf{E}+\mu_0\frac{\partial\mathbf{H}}{\partial t}=\mathbf{0},\quad \nabla\cdot\mathbf{D}=0,$$
$$\nabla\cdot\mathbf{B}=0,\quad \nabla\times\mathbf{H}=\frac{\partial\mathbf{D}}{\partial t}, \tag{2.1}$$

in the absence of free charge, free current, and magnetic susceptibility. At the interface $r=a$, the corresponding boundary conditions are that: (a) the tangential components of $\mathbf{E}$ and $\mathbf{H}$ are continuous; (b) the normal components of $\mathbf{B}(=\mu_0\mathbf{H})$ and $\mathbf{D}$ are continuous. Additionally we require that the fields are finite along the core axis, $r=0$, and that they decay to zero as r tends to infinity.

As in previous works in the field[4,12] the electric displacement $\mathbf{D}$ is assumed to be cubic in $\mathbf{E}$ and given by

$$\mathbf{D}=\epsilon_0 n_j(n_j+2n_q|\mathbf{E}|^2)\mathbf{E}\quad (j=1 \text{ or } 2),$$

where n_j is the usual refractive index, which takes the values n_1 in the core and n_2 in the cladding. Here, n_j is related to the permittivity by the equation

$$n_j=\sqrt{\epsilon_j/\epsilon_0}\quad (j=1 \text{ or } 2).$$

For simplicity of presentation, the nonlinear coefficient n_q of the refractive index is taken to be the same in both the core and the cladding and explicit dependence of n_1, n_2, and n_q on frequency is omitted.

The modulation of quasimonochromatic signals will be analyzed using modal analysis for the basic approximation and incorporating the modulation by employing the multiple scale technique in an asymptotic perturbation method or more specifically the method of derivative expansion.[13] The method requires the expansion of the fields in terms of a small amplitude parameter ν, for example,

$$\mathbf{F}=\nu\mathbf{F}^{(1)}+\nu^2\mathbf{F}^{(2)}+\nu^3\mathbf{F}^{(3)}+\cdots\quad (\mathbf{F}=\mathbf{E} \text{ or } \mathbf{H})$$

and also introduces the stretched coordinates

$$Z_n=\nu^n z\quad (n=1,2,3...),$$
$$T_n=\nu^n t\quad (n=1,2,3...),$$

which imply the derivative expansions

$$\frac{\partial}{\partial z}\to\frac{\partial}{\partial z}+\nu\frac{\partial}{\partial Z_1}+\nu^2\frac{\partial}{\partial Z_2}+\cdots$$

and

$$\frac{\partial}{\partial t}\to\frac{\partial}{\partial t}+\nu\frac{\partial}{\partial T_1}+\nu^2\frac{\partial}{\partial T_2}+\cdots.$$

By equating successive powers of ν, sets of equations and boundary conditions governing successive approximations to the fields $\mathbf{E}$ and $\mathbf{H}$ can then be determined and solved.

III. GENERAL FORMULATION

When the multiple scale expansions are substituted into (2.1) the equations obtained from equating the terms of first order in ν generate the standard linear boundary value problem. Separable solutions exist in the form of circularly polarized modes

$$\mathbf{F}^{(1)}(r,\theta,z,t)=\mathbf{F}^{(1)}(r)e^{i(l\theta+kz-\omega t)},$$

for each azimuthal mode number $l=0,\pm1,\pm2,...$, provided that k and ω satisfy the linear dispersion relation, which is obtained by applying the matching conditions at $r=a$ and the conditions as $r\to0$ and $r\to\infty$. It is readily shown[13] that the dispersion relation for a mode with l negative is identical to the relation with positive mode number $|l|$. Thus the corresponding modes may be written in the form

$$\mathbf{F}^{(1)}(r,\theta,z,t)=\mathbf{F}^{\pm(1)}(r)e^{i(\pm l\theta+kz-\omega t)},$$
$$\omega=\omega_l(k),\quad l=0,1,2,...,$$

where the $\pm$ superscript indicating the two possible solutions for each positive integer l. Moreover it is found that the vector functions $\mathbf{E}^{\pm(1)}(r)$ and $\mathbf{H}^{\pm(1)}(r)$ may be written as

$$\mathbf{E}^{\pm(1)}(r)=i\widetilde{E}_1(r)\mathbf{e}_r\pm\widetilde{E}_2(r)\mathbf{e}_\theta+\widetilde{E}_3(r)\mathbf{e}_z,$$
$$\mathbf{H}^{\pm(1)}(r)=\pm\widetilde{H}_1(r)\mathbf{e}_r+i\widetilde{H}_2(r)\mathbf{e}_\theta\pm i\widetilde{H}_3(r)\mathbf{e}_z, \tag{3.1}$$

where $\widetilde{E}_i(r)$ and $\widetilde{H}_i(r)$ $(i=1,2,3)$ are real and $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_z$ are unit vectors associated with the coordinate directions. Marcuse[14] also shows that only the HE_{11} mode (i.e., having $l=1$) has zero cutoff frequency. This means, that at the lower frequencies, both the $+$ and the $-$ mode for $l=1$ propagate and hence we need to use two independent amplitudes in representing the corresponding fields. Since both modes propagate with the same phase factor $\psi\equiv kz-\omega t$, even though they have independent complex amplitudes $A^{\pm}$, the general solution with frequency ω, wavenumber k, and azimuthal number 1 has the form

$$\mathbf{E}^{(1)} \equiv \mathbf{E}^{(1)}(r,\theta,\psi) = [A^{+}e^{i\theta}\mathbf{E}^{+(1)}(r) + A^{-}e^{-i\theta}\mathbf{E}^{-(1)}(r)]e^{i\psi} + \text{c.c.},$$

$$\mathbf{H}^{(1)} \equiv \mathbf{H}^{(1)}(r,\theta,\psi) = [A^{+}e^{i\theta}\mathbf{H}^{+(1)}(r) + A^{-}e^{-i\theta}\mathbf{H}^{-(1)}(r)]e^{i\psi} + \text{c.c.}, \tag{3.2}$$

$$\psi = kz - \omega t,$$

where c.c. denotes the complex conjugate. The complex amplitudes A^{+} and A^{-} are independent of z and t but are allowed to depend on the slow variables Z_n and T_n ($n = 1,2,3,\ldots$). Since $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ depend on z and t only through the phase variable ψ, and they are 2π-periodic in ψ, we insist likewise that all higher-order terms $\mathbf{E}^{(m)}$ and $\mathbf{H}^{(m)}$ depend only on r,θ,ψ and Z_n, T_n ($n = 1,2,\ldots$), and are 2π-periodic in both θ and ψ.

Substituting expressions (3.2) for $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ into the perturbation equations obtained from (2.1), and then comparing terms of second order in ν, we obtain the following system of linear inhomogeneous equations:

$$\nabla' \times \mathbf{E}^{(2)} - \mu_0 \frac{\omega}{k}\frac{\partial \mathbf{H}^{(2)}}{\partial \zeta} = \mathbf{g}, \quad \zeta = \frac{\psi}{k} = z - \frac{\omega t}{k}, \tag{3.3}$$

$$\nabla' \cdot \mathbf{E}^{(2)} = -\frac{\partial E_3^{(1)}}{\partial \zeta}, \tag{3.4}$$

$$\nabla' \times \mathbf{H}^{(2)} + \frac{\epsilon_0 n_j^2 \omega}{k}\frac{\partial \mathbf{E}^{(2)}}{\partial \zeta} = \mathbf{f}, \tag{3.5}$$

and

$$\nabla' \cdot \mathbf{H}^{(2)} = -\frac{\partial H_3^{(1)}}{\partial Z_1}. \tag{3.6}$$

Here $\mathbf{f} = \mathbf{f}(r,\theta,\zeta;Z_1,T_1,\ldots)$ and $\mathbf{g} = \mathbf{g}(r,\theta,\zeta;Z_1,T_1,\ldots)$ are expressions known explicitly as linear combinations of the first derivatives of $H_i^{(1)}$ and $E_i^{(1)}$, with respect to Z_1 and T_1, and ∇' is the usual del operator with the z derivative replaced by $\partial/\partial\zeta$ ($= k\,\partial/\partial\psi$) and where

$$F_1^{(1)}\mathbf{e}_r + F_2^{(1)}\mathbf{e}_\theta + F_3^{(1)}\mathbf{e}_z = \mathbf{F}^{(1)}.$$

Since it can be shown that the two scalar equations are in fact consequences of the two vector equations and the periodicity requirements, the fields $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$ are completely determined by (3.3), (3.5) and associated boundary conditions. In this inhomogeneous system which governs the dependence of $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$ on r,θ, and ψ, the self-adjoint differential operators on the left-hand sides are the same as those in the homogeneous system governing $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$. The condition that the inhomogeneous system has solutions which are 2π-periodic in both θ and ψ can be shown to be an orthogonality condition of the form

$$\int_0^\infty \int_0^{2\pi} \int_0^{2\pi} (\mathbf{u}\cdot\mathbf{f} - \mathbf{v}\cdot\mathbf{g})\, r\, d\theta\, d\psi\, dr = 0, \tag{3.7}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the most general periodic solution of the corresponding homogeneous boundary value problem. Explicitly, they have the form

$$\mathbf{u} = [\alpha^{+}e^{i\theta}\mathbf{E}^{+(1)}(r) + \alpha^{-}e^{-i\theta}\mathbf{E}^{-(1)}(r)]e^{i\psi} + \text{c.c.},$$

$$\mathbf{v} = [\alpha^{+}e^{i\theta}\mathbf{H}^{+(1)}(r) + \alpha^{-}e^{-i\theta}\mathbf{H}^{-(1)}(r)]e^{i\psi} + \text{c.c.}, \tag{3.8}$$

where α^{+} and α^{-} are two independent complex constants.

Evaluating the integrals with respect to ψ and θ and observing that Eq. (3.7) must be true for all constants α^{+} and α^{-} we obtain

$$\frac{\partial A^{\pm}}{\partial Z_1}\left[\int_0^\infty (\mathbf{E}^{\pm(1)} \times \mathbf{H}^{\pm(1)*} - \mathbf{H}^{\pm(1)} \times \mathbf{E}^{\pm(1)*})\cdot \mathbf{e}_z\, r\, dr\right] + \frac{\partial A^{\pm}}{\partial T_1}\left[\int_0^\infty (\epsilon_0 n_j^2 \mathbf{E}^{\pm(1)*}\cdot\mathbf{E}^{\pm(1)} + \mu_0 \mathbf{H}^{\pm(1)*}\cdot\mathbf{H}^{\pm(1)})\, r\, dr\right] = 0. \tag{3.9}$$

Taking into account the expressions (3.1) for $\mathbf{E}^{\pm(1)}(r)$ and $\mathbf{H}^{\pm(1)}(r)$, it is readily shown that in each case the ratio of the two bracketed terms reduces to

$$S = \frac{\int_0^\infty 2(\tilde{E}_1\tilde{H}_2 - \tilde{E}_2\tilde{H}_1)\, r\, dr}{\int_0^\infty (\epsilon_0 n_j^2(\tilde{E}_1^2 + \tilde{E}_2^2 + \tilde{E}_3^2) + \mu_0(\tilde{H}_1^2 + \tilde{H}_2^2 + \tilde{H}_3^2))\, r\, dr}, \tag{3.10}$$

which is in fact the group speed $S = d\omega/dk$ of each of the circularly polarized HE_{11} modes. Consequently, Eqs. (3.9) reduce to

$$\frac{\partial A^{+}}{\partial T_1} + S\frac{\partial A^{+}}{\partial Z_1} = 0, \quad \frac{\partial A^{-}}{\partial T_1} + S\frac{\partial A^{-}}{\partial Z_1} = 0, \tag{3.11}$$

thus confirming that, to first order, each complex amplitude travels at the group speed. Since A^{+} and A^{-} depend on Z_1 and T_1 only through the combination

$$\chi_1 = Z_1 - ST_1,$$

it is convenient, at each order $n = 1,2,3,\ldots$, to change from the variables Z_n, T_n to $\chi_n = Z_n - ST_n$ and Z_n. Hence $A^{\pm}$ may be regarded as of the form $A^{\pm}(\chi_1,\chi_2,Z_2,\ldots)$.

After expressions (3.2) are substituted into Eqs. (3.3) and (3.5) and the boundary conditions appropriate to $O(\nu^2)$, an inhomogeneous, linear boundary value problem for $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$ is obtained. Its complementary solution has the same form as the $O(\nu)$ solution (3.2), while a particular solution may be sought as a linear combination of terms having factors $\exp i(\psi \pm \theta)$. It is then found that the general solution may be written as

$$\mathbf{E}^{(2)} = \left\{ \left(b^{+}\mathbf{E}^{+(1)}(r) - i\frac{\partial A^{+}}{\partial \chi_1}\frac{d\mathbf{E}^{+(1)}}{dk} \right) e^{i\theta} + \left(b^{-}\mathbf{E}^{-(1)}(r) - i\frac{\partial A^{-}}{\partial \chi_1}\frac{d\mathbf{E}^{-(1)}}{dk} \right) e^{-i\theta} \right\} e^{i\psi} + \text{c.c.}, \tag{3.12}$$

with similar expressions for $\mathbf{H}^{(2)}$. The complementary solution involves two new arbitrary amplitudes b^{+} and b^{-} multiplying the fields of (3.1) while the particular solution involves the total derivative

$$\frac{d}{dk} = \frac{\partial}{\partial k} + \omega'(k)\frac{\partial}{\partial \omega}$$

of these same fields. For the step-index fiber the components $E_i(r)$ and $H_i(r)$ may be deduced from formulas on pages 293–295 of Marcuse.[14]

The expressions $\mathbf{E} \simeq \nu \mathbf{E}^{(1)} + \nu^2 \mathbf{E}^{(2)}$, $\mathbf{H} \simeq \nu \mathbf{H}^{(1)} + \nu^2 \mathbf{H}^{(2)}$, describe quasimonochromatic signals in which both of the circularly polarized modes modulate according to standard linear theory. Nonlinear effects enter only at $O(\nu^3)$ and are found to alter Eqs. (3.11) by introducing both nonlinearity and interaction. The equations obtained from the terms of $O(\nu^3)$ in Eqs. (2.1) are the following system of nonlinear, inhomogeneous, partial differential equations:

$$\nabla' \times \mathbf{E}^{(3)} - \mu_0 \frac{\omega}{k}\frac{\partial \mathbf{H}^{(3)}}{\partial \zeta} = \mathbf{G}, \tag{3.13}$$

$$\epsilon_0 n_j^2 \nabla' \cdot \mathbf{E}^{(3)} = -\epsilon_0 n_j^2 \left(\frac{\partial E_3^{(2)}}{\partial Z_1} + \frac{\partial E_3^{(1)}}{\partial Z_2} \right) - \epsilon_0 n_j^2 \left(\frac{\partial E_3^{(2)}}{\partial \chi_1} + \frac{\partial E_3^{(1)}}{\partial \chi_2} \right) - 2\epsilon_0 n_j n_q \nabla' \cdot (|\mathbf{E}^{(1)}|^2 \mathbf{E}^{(1)}), \tag{3.14}$$

$$\nabla' \times \mathbf{H}^{(3)} + \frac{\epsilon_0 n_j^2 \omega}{k}\frac{\partial \mathbf{E}^{(3)}}{\partial \zeta} = \mathbf{F}, \tag{3.15}$$

and

$$\nabla' \cdot \mathbf{H}^{(3)} = -\frac{\partial H_3^{(2)}}{\partial Z_1} - \frac{\partial H_3^{(1)}}{\partial Z_2} - \frac{\partial H_3^{(2)}}{\partial \chi_1} - \frac{\partial H_3^{(1)}}{\partial \chi_2}. \tag{3.16}$$

Here $\mathbf{F}$ and $\mathbf{G}$ denote linear combinations of the derivatives of $E_i^{(2)}$, $H_i^{(2)}$, $E_i^{(1)}$, and $H_i^{(1)}$ similar to those in (3.14) and (3.16) and, in the case of $\mathbf{F}$, also a term in $\partial(|\mathbf{E}^{(1)}|^2\mathbf{E}^{(1)})/\partial\psi$. Explicit forms are given by Newboult.[15] At $r = a$, all components of $\mathbf{E}^{(3)}$ are continuous, except for $E_1^{(3)}$, which satisfies

$$[n_j^2 E_1^{(3)}] = -2n_q [n_j |\mathbf{E}^{(1)}|^2 E_1^{(1)}],$$

where [ ] denotes the jump across the interface at $r = a$, while n_j equals n_1 in the core and n_2 in the cladding. As before, conditions at $r = 0$ and as $r \to \infty$ are also imposed.

As for the system (3.3)–(3.6), it can be shown that the system (3.13)–(3.16) is not overdetermined and that the two scalar equations (3.14) and (3.16) may be omitted. Since Eqs. (3.13) and (3.15) and associated boundary conditions have exactly the same form as the linear, inhomogeneous system governing $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$, there is a similar solvability condition

$$\int_0^\infty \int_0^{2\pi} \int_0^{2\pi} (\mathbf{u} \cdot \mathbf{F} - \mathbf{v} \cdot \mathbf{G}) r \, d\theta \, d\psi \, dr = 0, \tag{3.17}$$

where $\mathbf{u}$ and $\mathbf{v}$ are as given in (3.8). When the integrals with respect to ψ and θ are evaluated, again observing that we require Eq. (3.17) to be true for all constants $\alpha^{\pm}$, two scalar equations involving derivatives with respect to the slow variables χ_1, Z_1, and Z_2 are obtained, namely:

$$\left(\frac{\partial b^{\pm}}{\partial Z_1} + \frac{\partial A^{\pm}}{\partial Z_2} \right) f_1^{\pm} + \frac{\partial^2 A^{\pm}}{\partial \chi_1^2} g^{\pm} + A^{\pm *} A^{\pm 2} i f_2^{\pm} + A^{\mp *} A^{\mp} A^{\pm} i f_3^{\pm} = 0, \tag{3.18}$$

where the coefficients involve integrals over the fiber cross section of the fields $\mathbf{E}^{\pm(1)}$, $\mathbf{H}^{\pm(1)}$, $d\mathbf{E}^{\pm(1)}/dk$, and $d\mathbf{H}^{\pm(1)}/dk$ associated with the linearized modes.

After using expressions (3.1), it is found that the coefficients of the two equations are identical and may be written as

$$f_1^{\pm} = 2\int_0^\infty (\tilde{H}_1 \tilde{E}_2 - \tilde{H}_2 \tilde{E}_1) r \, dr \equiv f_1,$$

$$g^{\pm} = i\frac{d}{dk}\int_0^\infty (\tilde{E}_1 \tilde{H}_2 - \tilde{E}_2 \tilde{H}_1) r \, dr - \frac{1}{2} iS \frac{d}{dk}\int_0^\infty (\epsilon_0 n_j^2 (\tilde{E}_1^2 + \tilde{E}_2^2 + \tilde{E}_3^2) + \mu_0 (\tilde{H}_1^2 + \tilde{H}_2^2 + \tilde{H}_3^2)) r \, dr \equiv ig,$$

$$f_2^{\pm} = 2\omega\epsilon_0 n_q \int_0^\infty n_j [(\tilde{E}_1^2 - \tilde{E}_2^2 - \tilde{E}_3^2)^2 + 2(\tilde{E}_1^2 + \tilde{E}_2^2 + \tilde{E}_3^2)^2] r \, dr \equiv f_2,$$

and

$$f_3^{\pm} = 4\omega\epsilon_0 n_q \int_0^\infty n_j [(\tilde{E}_1^2 + \tilde{E}_2^2 - \tilde{E}_3^2)^2 + (\tilde{E}_1^2 + \tilde{E}_2^2 + \tilde{E}_3^2)^2 + (\tilde{E}_1^2 - \tilde{E}_2^2 + \tilde{E}_3^2)^2] r \, dr \equiv f_3.$$

It may be noted that f_3 and f_2 are related by

$$f_3 = 2f_2 - 32\omega\epsilon_0 n_q \int_0^\infty n_j \tilde{E}_2^2 \tilde{E}_3^2 r \, dr,$$

while use of standard results from linear theory yields the relation

$$g\frac{d\omega}{dk} = -\frac{1}{2} f \frac{d^2\omega}{dk^2}.$$

Since in (3.18), $b^{\pm}$ are the only quantities which may depend on Z_1, they are bounded only if $\partial b^{\pm}/\partial Z_1 = 0$. Consequently, $b^{\pm} = b^{\pm}(\chi_1, Z_2, \chi_2, Z_3, \ldots)$ is merely an $O(\nu^2)$ correction of $\nu A^{\pm}$ and can be omitted without loss of generality. Thus (3.18) may be written as a pair of cubically nonlinear Schrödinger equations

$$i\frac{\partial A^{+}}{\partial Z_2} = \frac{g}{f_1}\frac{\partial^2 A^{+}}{\partial \chi_1^2} + \left(\frac{f_2}{f_1}|A^{+}|^2 + \frac{f_3}{f_1}|A^{-}|^2 \right) A^{+},$$
$$i\frac{\partial A^{-}}{\partial Z_2} = \frac{g}{f_1}\frac{\partial^2 A^{-}}{\partial \chi_1^2} + \left(\frac{f_2}{f_1}|A^{-}|^2 + \frac{f_3}{f_1}|A^{+}|^2 \right) A^{-}, \tag{3.19}$$

which couples together the evolution of the complex amplitudes A^{+} and A^{-} of the two circularly polarized modes contained in the $O(\nu)$ fields (3.2). When expression (3.2) for $\mathbf{E}^{(1)}$ is rearranged as

$$\mathbf{E}^{(1)} = 2a^+\{ -\widetilde{E}_1\mathbf{e}_r \sin(\theta + kz - \omega t + \varphi^+) + (\widetilde{E}_2\mathbf{e}_\theta + \widetilde{E}_3\mathbf{e}_z)\cos(\theta + kz - \omega t + \varphi^+)\} + 2a^-\{ -\widetilde{E}_1\mathbf{e}_r \sin(-\theta + kz - \omega t + \varphi^-) + (-\widetilde{E}_2\mathbf{e}_\theta + \widetilde{E}_3\mathbf{e}_z)\cos(-\theta + kz - \omega t + \varphi^-)\}, \tag{3.20}$$

where $A^+ = a^+e^{i\varphi^+}$, $A^- = a^-e^{i\varphi^-}$, it is seen that A^+ describes a left-handed, circularly polarized (corkscrew) mode of amplitude $a^+ = |A^+|$ with radial field including the factor $|\widetilde{E}_1(r)|$ vanishing on $\theta = \omega t + kz - \varphi^+ + n\pi$ and having maxima in directions $\theta = (1/2)\pi + \omega t + kz - \varphi^+ + n\pi$. Similarly, A^- describes right-handed modes.

Although Eqs. (3.19) (having only three independent real coefficients $g/f_1, f_2/f_1, f_3/f_1$) are a special case of the coupled NLS equations occurring in the mathematical literature,[16,17] their relevance to the propagation of signals in perfect "monomode" fibers has not previously been noted, even in treatments[9,10] of birefringent fibers. They are equivalent to a pair of equations for the complex amplitudes of linearly polarized modes. Indeed, by writing

$$B_1 = i(A^+ + A^-) = b_1 e^{i\varphi_1}, \quad B_2 = A^- - A^+ = b_2 e^{i\varphi_2}$$

is found that

$$\mathbf{E}^{(1)} = 2b_1\{(\widetilde{E}_1\mathbf{e}_r \cos\theta + \widetilde{E}_2\mathbf{e}_\theta \sin\theta)\cos(kz - \omega t + \varphi_1) + \widetilde{E}_3\mathbf{e}_z \cos\theta \sin(kz - \omega t + \varphi_1)\} + 2b_2\{(\widetilde{E}_1\mathbf{e}_r \sin\theta - \widetilde{E}_2\mathbf{e}_\theta \cos\theta)\cos(kz - \omega t + \varphi_2) - \widetilde{E}_3\mathbf{e}_z \sin\theta \sin(kz - \omega t + \varphi_2)\}, \tag{3.21}$$

so that $b_1 = |B_1|$ and $b_2 = |B_2|$ are the amplitudes, and $\varphi_1 = \arg B_1$ and $\varphi_2 = -\arg B_2$ are the phases of modes polarized linearly in the $\theta = 0$ and $\theta = \pi/2$ directions, respectively.

The corresponding propagation equations are

$$\begin{aligned} i\frac{\partial B_1}{\partial Z_2} &= \frac{g}{f_1}\frac{\partial^2 B_1}{\partial \chi_1^2} + \frac{g_2}{f_1}(|B_1|^2 + |B_2|^2)B_1 + \frac{g_3}{f_1}(B_2^2B_1^* - |B_2|^2B_1), \\ i\frac{\partial B_2}{\partial Z_2} &= \frac{g}{f_1}\frac{\partial^2 B_2}{\partial \chi_1^2} + \frac{g_2}{f_1}(|B_1|^2 + |B_2|^2)B_2 + \frac{g_3}{f_1}(B_1^2B_2^* - |B_1|^2B_2), \end{aligned} \tag{3.22}$$

where $g_2 = \frac{1}{4}(f_2 + f_3)$, $g_3 = \frac{1}{4}(f_3 - f_2)$.

It may be noted that the systems (3.19) and (3.22) are equivalent to the systems (2) and (1), respectively, of Blow *et al.*[9] in the absence of birefringence. Moreover it can be readily seen that Eqs. (3.19) are satisfied by taking $A^- \equiv 0$, in which case the system reduces to the NLS equation

$$i\frac{\partial A^+}{\partial Z_2} = \frac{g}{f_1}\frac{\partial^2 A^+}{\partial \chi_1^2} + \frac{f_2}{f_1}|A^+|^2A^+, \tag{3.23}$$

which, as discussed in Sec. I, is usually taken to govern the competing nonlinear and dispersive effects in optical fibers without birefringence. Similarly the cases $A^+ = 0$, $B_2 = 0$ and $B_1 = 0$ each reduce to a NLS equation. Thus the solutions to Eq. (3.23) are included amongst the solutions to the general system (3.18), but this system clearly has many other solutions.

IV. SOME SPECIAL SOLUTIONS

Solutions to (3.19) are readily found in a few specific cases and this section will deal with the construction and interpretation of these. In particular, the coefficients g/f_1, $f_2/f_1, f_3/f_1$, g_2/f_1 and g_3/f_1, which in Sec. III are defined in terms of the modes capable of propagating in the axisymmetric waveguide, are readily identified as the associated "group delay dispersion" and as nonlinear susceptibilities.[9]

For ease of manipulation, it helps to rescale the variables Z_2 and χ_1 in (3.18), by introducing $\tau = (f_2/f_1)Z_2$, $x = (f_2/g)^{1/2}\chi_1$ and $h = f_3/f_2$, so giving

$$i\frac{\partial A^\pm}{\partial \tau} = \frac{\partial^2 A^\pm}{\partial x^2} + A^\pm(|A^\pm|^2 + h|A^\mp|^2). \tag{4.1}$$

[The condition that Eq. (3.23) possesses "bright" soliton solutions is $f_2/g > 0$, and this has been assumed here.] The corresponding leading order approximation to the fields in the "monomode" fiber ($l = 1$) have the form given in (3.20).

A. Case 1: Uniform wavetrains

It is easily confirmed that Eqs. (4.1) possess solutions of the form

$$A^+ = a \exp i[cx + (c^2 - a^2 - hb^2)\tau],$$
$$A^- = b \exp i[dx + (d^2 - b^2 - ha^2)\tau],$$

where a, b, c, d are independent real constants. Clearly these represent two wavetrains with wave speeds dependent on the (constant) amplitudes $a^+ = a$ and $a^- = b$ of the two circularly polarized (corkscrew) modes.

Substituting into (3.20) gives

$$\widetilde{\mathbf{E}}^{(1)} = 2a[-\widetilde{E}_1\mathbf{e}_r \sin(\theta + k_1z - \omega_1t) + (\widetilde{E}_2\mathbf{e}_\theta + \widetilde{E}_3\mathbf{e}_z)\cos(\theta + k_1z - \omega_1t)] + 2b[-\widetilde{E}_1\mathbf{e}_r \sin(-\theta + k_2z - \omega_2t) + (-\widetilde{E}_2\mathbf{e}_\theta + \widetilde{E}_3\mathbf{e}_z)\cos(-\theta + k_2z - \omega_2t)], \tag{4.2}$$

where

$$\begin{aligned} k_1 &= k + vc\sqrt{f_2/g} + v^2(f_2/f_1)(c^2 - a^2 - hb^2), \\ \omega_1 &= \omega + vc\sqrt{f_2/g}S, \\ k_2 &= k + vd\sqrt{f_2/g} + v^2(f_2/f_1)(d^2 - b^2 - ha^2), \\ \omega_2 &= \omega + vd\sqrt{f_2/g}S. \end{aligned}$$

This shows how nonlinearity induces both double circular refraction and interaction between the left- and right-handed circularly polarized modes that have wave speeds ω_1/k_1 and ω_2/k_2, respectively. Moreover, by writing $k_i - k = v\hat{k}_i$ ($i = 1,2$), it is seen that

$$\begin{aligned} \omega_1 &= \omega + Sv\hat{k}_1 - (g/f_1)Sv^2\hat{k}_1^2 + S(f_2/f_1)[v^2a^2 + hv^2b^2] + O(v^3), \\ \omega_2 &= \omega + Sv\hat{k}_2 - (g/f_1)Sv^2\hat{k}_2^2 + S(f_2/f_1)[v^2b^2 + hv^2a^2] + O(v^3), \end{aligned}$$

where

$$\mathbf{E}\cdot\mathbf{e}_r \simeq \nu E_1^{(1)} = -2\widetilde{E}_1(r) \times [\nu a \sin(\theta + k_1 z - \omega_1 t) + \nu b \sin(-\theta + k_2 z - \omega_2 t)].$$

This not only confirms the results of Sec. III that S is the group speed $\omega'(k)$ and that g/f_1 is the group delay dispersion $-\frac{1}{2}\omega''(k)/\omega'(k)$, but also shows that f_2/f_1 is the nonlinear susceptibility associated with either of the pure circularly polarized (corkscrew) modes.

The parameter h ($=f_3/f_2$) may be regarded as a coupling constant describing the interaction between these two modes. Alternatively, it may be defined in terms of plane-polarized modes such as the case $b=a$, $c=d$ [giving $k_2=k_1$, $\omega_2=\omega_1$ and $B_2=0$, in (3.20) and (3.21)]. Thus the wave field with

$$\mathbf{E}\cdot\mathbf{e}_r \simeq \nu E_1^{(1)} = 4a\widetilde{E}_1(r)\cos\theta\cos(k_1 z - \omega_1 t + \varphi_1),$$

$$k_1 = k + \nu c\sqrt{\frac{f_2}{g}} - (1+h)\frac{f_2}{f_1}4\nu^2 a^2 + \frac{f_2}{f_1}\nu^2 c^2,$$

$$\omega_1 = \omega + \nu c\sqrt{f_2/g}S,$$

has nonlinear susceptibility $(1+h)f_2/f_1 = 4g_2/f_1$. Consequently, f_3/f_1 ($=hf_2/f_1$) measures the difference between nonlinear effects for plane- and circularly polarized modes.

Generally, for monochromatic waves $\omega_2=\omega_1$ (with $c=d=0$) the difference in wave speeds causes gradual rotation of the field pattern, since expression (4.2) then becomes

$$\mathbf{E}^{(1)} = 2(a+b)\{(-\widetilde{E}_1\mathbf{e}_r\cos\Psi - \widetilde{E}_2\mathbf{e}_\theta\sin\Psi)\times\sin(\bar{\bar{k}}z-\omega t) + \widetilde{E}_3\mathbf{e}_z\cos\Psi\cos(\bar{\bar{k}}z-\omega t)\} + 2(a-b)\{(-\widetilde{E}_1\mathbf{e}_r\sin\Psi + \widetilde{E}_2\mathbf{e}_\theta\cos\Psi)\times\cos(\bar{\bar{k}}z-\omega t) - \widetilde{E}_3\mathbf{e}_z\sin\Psi\sin(\bar{\bar{k}}-\omega t)\},$$

where

$$\Psi \equiv \theta + \tfrac{1}{2}(k_1-k_2)z = \theta + (\nu^2 f_2/2f_1)(h-1)(a^2-b^2)z = \Psi(\theta, Z_2),$$

$$\bar{\bar{k}} \equiv \tfrac{1}{2}(k_1+k_2) = k - (\nu^2 f_2/2f_1)(h+1)(a^2+b^2).$$

Except in the special case $h=1$, the general "elliptically polarized" waveform ($b \neq \pm a$) exhibits rotation on the scale of $Z_2 = \nu^2 z$, analogous to the ellipse rotation phenomenon predicted for plane waves by Maker *et al.*[11] This is a nonlinearly induced double circular refraction.

B. Case 2: Circularly polarized modes

As suggested in Sec. III, we let $A^- = 0$ so giving the NLS equation (3.23) which yields

$$i\frac{\partial A^+}{\partial\tau} = \frac{\partial^2 A^+}{\partial x^2} + |A^+|^2 A^+, \tag{4.3}$$

with a and τ defined as in (4.1). This governs the evolution of a general left-handed mode.

There is an extensive theory concerning this equation, which is completely integrable so that many types of exact solutions are known.[18,19] One of the simplest, yet most important types is the single soliton

$$A^+ = a\sqrt{2}e^{-i\varphi}\operatorname{sech} a(x-b\tau), \tag{4.4}$$

where

$$\varphi = \tfrac{1}{2}bx - (\tfrac{1}{4}b^2 - a^2)\tau$$

and a and b are arbitrary constants.

Inserting the corresponding expressions $a^+ = a\sqrt{2}\operatorname{sech} a(x-b\tau)$, $\varphi^+ = -\varphi$, $a^- = 0$ into (3.20) shows that the $O(\nu)$ radial component of $\mathbf{E}^{(1)}$ is

$$E_1^{(1)} = -2\sqrt{2}a\operatorname{sech} a(\tilde{k}z - \tilde{\omega}t)\widetilde{E}_1(r)\sin(\theta + \bar{k}z - \bar{\omega}t),$$

where

$$\tilde{k} = \nu\sqrt{\frac{f_2}{g}} - \nu^2 b\frac{f_2}{f_1}, \quad \tilde{\omega} = \nu S\sqrt{\frac{f_2}{g}},$$

$$\bar{k} = k - \frac{\nu b}{2}\sqrt{\frac{f_2}{g}} + \nu^2\frac{f_2}{f_1}\left(\frac{1}{4}b^2 - a^2\right),$$

$$\bar{\omega} = \omega - \frac{\nu b}{2}S\sqrt{\frac{f_2}{g}}.$$

This solution describes a gradual amplitude modulation of a circular mode ($A^- \equiv 0$). The function $\widetilde{E}_1(r)$ describes (as in linear theory) the radial dependence of $E_1^{(1)}$. The signal amplitude is given by

$$2\sqrt{2}a\operatorname{sech} a(\tilde{k}z - \tilde{\omega}t),$$

which is a nondistorting waveform traveling at a speed close to the group speed $S=\omega'(k)$. The wavenumber $\bar{k}$ and angular frequency $\bar{\omega}$ are each constant. The radial field $\nu E_1^{(1)}$ has maximum modulus on lines $\theta = \pm\frac{1}{2}\pi - \bar{k}z + \bar{\omega}t$ which (for $\bar{k}>0$) describe left-handed corkscrews of pitch equal to the wavelength $2\pi/\bar{k}$.

Similarly the other components of $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ describe a field pattern which spirals around the fiber axis, and with amplitude having the same "sech envelope." Thus the solution (4.4) describes a soliton for a left-handed circularly polarized mode.

Solutions having $A^+ = 0$ and with A^- of the form (4.4) describe right-handed solitons, which are closely analogous to the above solution.

C. Case 3: Linearly polarized modes

When $A^- = A^+ e^{2i\alpha}$, where α is any real constant, then $|A^-| = |A^+|$ and the system (4.1) reduces to consideration of the single equation

$$i\frac{\partial A^+}{\partial\tau} = \frac{\partial^2 A^+}{\partial x^2} + (1+h)|A^+|^2 A^+, \tag{4.5}$$

which is merely a rescaled version of the NLS equation. Consequently it possesses solutions analogous to all the solutions of (4.3). In particular, soliton solutions are

$$A^+ = a\sqrt{2/(1+h)}e^{-i\varphi}\operatorname{sech} a(x-b\tau), \tag{4.6}$$

where $\varphi = \frac{1}{2}bx - (\frac{1}{4}b^2 - a^2)\tau$.

As in case 2, the field pattern within this mode having the sech envelope may be characterized by investigating the component $E_1^{(1)}(r,\theta,z,t)$. From (3.20) it is found that

$$E_1^{(1)} = -4a\sqrt{2/(1+h)}\widetilde{E}_1(r)\cos(\theta-\alpha)\times\operatorname{sech}[a(\tilde{k}z-\tilde{\omega}t)]\sin(\bar{k}z-\bar{\omega}t+\alpha),$$

with $\tilde{k}, \tilde{\omega}, \bar{k}, \bar{\omega}$ defined as in case 2. Consequently the solution (4.6) describes the familiar single soliton with arbitrary fixed polarization angle α. Special cases are the soliton solutions of the system (3.22) with either $B_2 = 0$ or $B_1 = 0$.

Although it is well known that (4.5) possesses multisoliton solutions, so that many solitons having the same polarization angle α may collide and yet emerge with unchanged form, the consequences of a collision between linearly polarized solitons of differing polarization angle are only just being investigated numerically. First results suggest that such solitons emerge as waves of permanent form, similar to the "vector solitons" recently reported by Christodoulides and Joseph.[20] These calculations are reported in Parker and Newboult.[21]

ACKNOWLEDGMENT

During preparation of this work, one of the authors (GKN) was supported by a studentship from the Science and Engineering Research Council.

[1] A. Hasegawa and F. Tappert, Appl. Phys. Lett. **23**, 142 (1973).
[2] L. F. Mollenauer, R. H. Stolen, and J. P. Gordon, Phys. Rev. Lett. **45**, 1095 (1980).
[3] D. Anderson and M. Lisak, Phys. Rev. A **27**, 1393 (1983).
[4] V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[5] M. J. Potasek, G. P. Agrawal, and S. C. Pinault, J. Opt. Soc. Am. B **3**, 205 (1986).
[6] A. Hasegawa and Y. Kodama, Proc. IEEE **69**, 1145 (1981).
[7] B. Bendow and P. D. Gianino, J. Opt. Soc. Am. **70**, 539 (1980).
[8] M. Jain and N. Tzoar, J. App. Phys. **49**, 4649 (1978).
[9] K. J. Blow, N. J. Doran, and D. Wood, Opt. Lett. **12**, 202 (1987).
[10] C. J. Menyuk, IEEE J. Quantum Electron. **QE-23**, 174 (1987).
[11] P. D. Maker, R. W. Terhune, and C. M. Savage, Phys. Rev. Lett. **12**, 507 (1964).
[12] N. N. Akhmedieva, V. M. Eleonskii, and N. E. Kulagin, Sov. Phys. JETP **62**, 894 (1985).
[13] A. Jeffrey and T. Kawahara, *Asymptotic Methods in Nonlinear Wave Theory* (Pitman, London, 1982).
[14] D. Marcuse, *Light Transmission Optics* (Van Nostrand, New York, 1972).
[15] G. K. Newboult, Ph.D. thesis, University of Nottingham, 1989.
[16] V. E. Zakharov and E. I. Schulman, Physica **4D**, 270 (1982).
[17] R. Sahadevan, K. M. Tamizhmani, and M. Lakshmanan, J. Phys. A: Math. **19**, 1783 (1986).
[18] R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, London, 1982).
[19] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[20] D. N. Christodoulides and R. I. Joseph, Opt. Lett. **13**, 53 (1988).
[21] D. F. Parker and G. K. Newboult, to appear in J. Phys. Colloq.

Vertex operator construction of nonassociative algebras and their affinizations

M. Günaydin
Department of Physics, Penn State University, University Park, Pennsylvania 16802

(Received 27 May 1988; accepted for publication 14 December 1988)

A general method for the vertex operator construction of nonassociative algebras and triple systems and their super analogs is given. Using this method the natural affine extensions of these algebraic structures can be defined and constructed. Essentially all the algebraic structures of physical theories based on "point particles" can be constructed this way. Conversely, these methods can be used to define the "stringy" analogs of these algebraic structures.

I. INTRODUCTION

One of the most important algebraic tools of the string theories is the Fubini–Veneziano vertex operator.[1] It plays a crucial role in various formulations of string theories and in calculating their scattering amplitudes. The vertex operators form the backbone of the representation theory of affine Kac–Moody algebras as formulated by Frenkel and Kac.[2,3] The level one representations of untwisted affine Kac–Moody algebras $\hat{g}$ defined by simply laced Lie algebras g were constructed by Frenkel and Kac[2] and Segal.[4] More recently, the results of Frenkel and Kac have been extended to the vertex operator construction of the nonsimply laced affine Kac–Moody algebras.[5,6]

Our aim in this paper is to give a method for a vertex operator construction of a very general class of nonassociative algebras and triple systems. They include practically all the algebraic structures that have appeared in theoretical physics, such as Jordan algebras, division algebras, and their tensor products, etc. Using this method one can define and construct natural "affinized" extensions of these algebraic structures. Therefore we expect this method to have two general applications to string theories. First, it can be used to identify and analyze the various underlying algebraic structures of string theories, especially those of the "compactified" string theories, in a point particle physics language. Second, the method can be used to define stringy analogs of the algebraic structures of point particle physics such as density matrices. Our method extends in full generality to the construction of the super analogs of these algebraic structures such as affine Jordan superalgebras. Therefore, it is equally applicable to bosonic, as well as superstring theories. We expect it to have applications in other areas of theoretical physics, such as the study of operator product expansion in two-dimensional conformal field theories.

II. THE CONSTRUCTION OF LIE ALGEBRAS FROM TERNARY ALGEBRAS

Every simple Lie algebra g has a five-dimensional graded decomposition (Kantor structure) with respect to a subalgebra $g^{(0)}$ of maximal rank[7,8]

$$g = g^{(-2)} \oplus g^{(-1)} \oplus g^{(0)} \oplus g^{(+1)} \oplus g^{(+2)},$$

where $\oplus$ denotes the vector space direct sum and the commutation relations of the elements of various graded spaces $g^{(m)}$ are such that the following formal relations hold:

$$[g^{(m)},g^{(n)}] = g^{(m+n)}, \quad m,n = 0, \mp 1, \mp 2 \qquad (2.1)$$

and $g^{(m+n)} = 0$ if $|m+n| > 2$. [For the Lie algebra of SU(2), the $g^{(2)}$ and $g^{(-2)}$ spaces vanish.] One can define a conjugation $\tilde{}$ in g such that

$$\begin{aligned} \tilde{g}^{(+1)} &= g^{(-1)}, \\ \tilde{g}^{(+2)} &= g^{(-2)}, \\ \tilde{g}^{(0)} &= g^{(0)}. \end{aligned} \qquad (2.2)$$

The elements of g belonging to the $g^{(+1)}$ space can be labeled by the elements of a ternary algebra V,[7,8]

$$U_a \in g^{(+1)} \Leftrightarrow a \in V.$$

A ternary algebra V is a vector space on which is defined a ternary product (, ,) such that

$$(a,b,c) \in V, \quad \text{for all } a,b,c \in V. \qquad (2.3)$$

Using the conjugation $\tilde{}$ one can label the elements of L belonging to the $g^{(-1)}$ space by the elements of V as well,

$$\widetilde{U}_a \in g^{(-1)} \Leftrightarrow U_a \in g^{(+1)}.$$

We define the commutators of the elements belonging to $g^{(+1)}$ and $g^{(-1)}$ spaces as follows:

$$\begin{aligned} [U_a,\widetilde{U}_b] &= S_{ab} \in g^{(0)}, \\ [U_a,U_b] &= K_{ab} \in g^{(+2)}, \\ [\widetilde{U}_a,\widetilde{U}_b] &= \widetilde{K}_{ab} \in g^{(-2)}, \\ [S_{ab},U_c] &= U_{(abc)} \in g^{(+1)}. \end{aligned} \qquad (2.4)$$

Then the remaining commutators of g can all be expressed in terms of the ternary product[7,8]

$$\begin{aligned} [S_{ab},\widetilde{U}_c] &= -\widetilde{U}_{(bac)}, \\ [K_{ab},\widetilde{U}_c] &= -U_{(bca)} + U_{(acb)}, \\ [\widetilde{K}_{ab},U_c] &= \widetilde{U}_{(bca)} - \widetilde{U}_{(acb)}, \\ [S_{ab},S_{cd}] &= S_{(abc)d} - S_{c(bad)} = -S_{(cda)b} + S_{a(dcb)}, \\ [S_{ab},K_{cd}] &= K_{(abc)d} + K_{c(abd)}, \\ [S_{ab},\widetilde{K}_{cd}] &= -\widetilde{K}_{(bac)d} - \widetilde{K}_{c(bad)}, \\ [K_{ab},\widetilde{K}_{cd}] &= S_{(acb)d} - S_{(bca)d} - S_{(adb)c} + S_{(bda)c}. \end{aligned} \qquad (2.5)$$

The Jacobi identities in g are all satisfied if the ternary algebra satisfies the identities[7,8]

 0022-2488/89/040937-06$02.50

$$(ab(cdx)) - (cd(abx)) - (a(dcb)x) + ((cda)bx) = 0 \tag{2.6a}$$

and

$$\{(ax(cbd)) - ((cbd)xa) + (ab(cxd)) + (c(bax)d)\} - \{c \leftrightarrow d\} = 0. \tag{2.6b}$$

This construction of all simple Lie algebras from ternary algebras satisfying the identities (2.6a) and (2.6b) was generalized to a unified construction of simple Lie algebras and Lie superalgebras in Refs. 8 and 9. The ternary superalgebras underlying the Lie superalgebras satisfy the graded forms of the identities (2.6a) and (2.6b).[8,9]

A large class of ternary (super) algebras V can be defined in terms of some underlying binary (super) algebras A with a binary product. An important subclass of such algebras have the ternary product[7,8]

$$(abc) = a\cdot(\bar{b}\cdot c) + c\cdot(\bar{b}\cdot a) - b\cdot(\bar{a}\cdot c), \tag{2.7}$$

where $\cdot$ and $\bar{\ }$ denote the product and the conjugation operation in the underlying binary algebra. Every associative algebra with the ternary product (2.7) defines a ternary algebra satisfying the identities (2.6a) and (2.6b).[7,8]

If the ternary product satisfies the symmetry condition

$$(abc) = (cba), \tag{2.8}$$

then the left-hand side of the identity (2.6b) vanishes identically and the condition (2.8) together with (2.6a) define a Jordan triple system.[10] In this case, the $g^{(+2)}$ and $g^{(-2)}$ spaces of g vanish and one has a three-graded (Jordan) structure:

$$g = g^{(-1)} \oplus g^{(0)} \oplus g^{(+1)}. \tag{2.9}$$

For Jordan triple systems the above construction reduces to the well-known Tits–Koecher–Kantor construction of Lie algebras from Jordan triple systems.[11-13] With the exception of G_2, F_4, and E_8, all simple Lie algebras can be constructed from some underlying Jordan triple system. The Jordan algebras with the symmetric Jordan product $\circ$ define a Jordan triple system under the triple product

$$\{abc\} \equiv a\circ(b\circ c) + c\circ(b\circ a) - (a\circ c)\circ b. \tag{2.10}$$

A Jordan triple system corresponds to a Jordan pair with involution. A Jordan pair is defined as a pair of spaces (V_+, V_-) that act on each other like a Jordan triple system. There is a one-to-one correspondence between three-graded Lie algebras $g = g^{(-1)} \oplus g^{(0)} \oplus g^{(+1)}$ and Jordan pairs (V_+, V_-). This correspondence can be seen from the fact that both the $g^{(+)}$ and $g^{(-1)}$ spaces can be mapped onto some Jordan triple systems. For a detailed study of the Jordan pairs and their connection to Jordan triple systems and Jordan algebras, we refer to the book by Loos[14] and the review article by McCrimmon.[15] Jordan triple systems and Jordan pairs have also appeared in the physics literature.[16]

The construction of Lie algebras from Jordan triple systems was generalized to construction of Lie superalgebras from super Jordan triple systems in Refs. 8 and 9. For a complete list of Lie superalgebras that can be constructed from simple Jordan superalgebras we refer to Ref. 9.

III. VERTEX OPERATOR CONSTRUCTION OF TERNARY ALGEBRAS AND THEIR AFFINIZATIONS

In the previous section we summarized briefly the general construction of Lie algebras and Lie superalgebras from ternary algebras. In this section we shall give a vertex operator construction of these ternary algebras. Furthermore, we shall define their natural "affine extensions" and construct them using vertex operators. Our starting point is the vertex operator construction of affine Kac–Moody algebras[2,5,6] and of the affine super Kac–Moody algebras.[17] The method is of full generality in that it can be applied to simply as well as nonsimply laced algebras and is independent of the level of the representation. To keep the discussion simple and concrete we shall, however, first summarize the Frenkel–Kac construction of simply laced algebras following Ref. 6.

Consider a simply laced Lie algebra of rank l, i.e., A_l, D_l, or E_l with a root lattice Λ and a weight lattice W. Normalize the roots α_i $(i = 1,...,l)$ such that they have length $\sqrt{2}$

$$\alpha\cdot\alpha \equiv \alpha_i \alpha_j a^{ij} = 2, \tag{3.1}$$

where a_{ij} is the Cartan matrix of the algebra. Construct a Fock space using a set of l Fubini–Veneziano fields $X^i(z)$ $(i = 1,...,l)$ over each of the normalized vacuum states $|p\rangle$ labeled by the elements p of the weight lattice W. The Fubini–Veneziano field $X^i(z)$ is defined as

$$X^i(z) = q^i - ip^i \ln z - i \sum_{n\neq 0} X^i_{-n} \frac{z^n}{n}, \tag{3.2}$$

and the l operators $h^i(z) = i(\partial/\partial z)X^i(z)$ represent the Heisenberg (affine Cartan) subalgebra. The Vertex operators $U(\alpha,z)$ are defined by the fields $X^i(z)$,

$$U(\alpha,z) = :e^{i\alpha\cdot X(z)}:, \tag{3.3}$$

and realize the current algebra modulo the cocycle factors. In terms of the moments of $U(\alpha,z)$ and $h^i(z)$ defined by

$$\begin{aligned} U_n(\alpha) &= \oint \frac{dz}{2\pi i} z^n U(\alpha,z), \quad n\in\mathbb{Z}, \\ h_n{}^i &= \oint \frac{dz}{2\pi i} z^n h^i(z), \end{aligned} \tag{3.4}$$

we have

$$U_m(\alpha)U_n(\beta) - (-1)^{\alpha\cdot\beta}U_n(\beta)U_m(\alpha) = \begin{cases} 0, & \text{for } \alpha\cdot\beta \geqslant 0, \\ U_{m+n}(\alpha+\beta), & \text{for } \alpha\cdot\beta = -1, \\ \alpha\cdot h_{m+n} + m\delta_{m+n,0}, & \text{for } \alpha = -\beta. \end{cases} \tag{3.5}$$

To turn the above commutation and anticommutation relations into commutators only one multiplies the vertex operator with the appropriate cocycle factor $\hat{\epsilon}_\alpha$

$$V(\alpha,z) = U(\alpha,z)\hat{\epsilon}_\alpha. \tag{3.6}$$

Then the moments $V_n(\alpha)$ and $h_n{}^i$ close under commutation and give a level 1 realization of the simply laced affine Kac–Moody algebra over the Hilbert space

$$\mathscr{H} = \text{Fock}\,(X^i) \otimes W, \tag{3.7}$$

where Fock (X^i) represents the Fock space generated by the l Fubini–Veneziano boson fields $X^i(z)$ over the vacua labeled by $p\in W$. The products of cocycle factors $\hat{\epsilon}_\alpha$ are such that

$$\hat{\epsilon}_\alpha \hat{\epsilon}_\beta = \epsilon(\alpha,\beta)\,\hat{\epsilon}_{\alpha+\beta}, \tag{3.8}$$

where $\epsilon(\alpha,\beta)$ are the cocycles of the corresponding Lie algebra.[2]

Consider now a five-graded (Kantor) decomposition of a simple affine Kac–Moody algebra $\hat{g}$ induced by the Kantor decomposition of the finite-dimensional Lie algebra g

$$\hat{g} = \hat{g}^{(-2)} \oplus \hat{g}^{(-1)} \oplus \hat{g}^{(0)} \oplus \hat{g}^{(+1)} \oplus \hat{g}^{(+2)}, \tag{3.9}$$

where each of the subspaces $\hat{g}^{(+1)}$, etc., is now infinite dimensional due to the standard infinite gradation of $\hat{g}$. From here on we shall denote the root vectors associated with the space $\hat{g}^{(+1)}$ as α,β, etc.,

$$V_n^{(+1)}(\alpha)\epsilon\hat{g}^{(+1)},$$

where

$$V_n^{(+1)}(\alpha) = \oint \frac{dz}{2\pi i} z^n V(\alpha,z), \quad n\in\mathbb{Z}. \tag{3.10}$$

Now by commutation of the elements $V_n^{(+1)}(\alpha)\in\hat{g}^{(+1)}$ we generate the elements of the spaces $\hat{g}^{(+2)}$

$$[V_m^{(+1)}(\alpha),V_n^{(+1)}(\beta)] = \begin{cases} \epsilon(\alpha,\beta)\,V_{m+n}^{(+2)}(\alpha+\beta), & \text{for } \alpha\cdot\beta = -1,\\ 0, & \text{for } \alpha\cdot\beta \geqslant 0. \end{cases} \tag{3.11}$$

The elements of the space $\hat{g}^{(-1)}$ are obtained from those of the $\hat{g}^{(+1)}$ space by replacing the root vectors $\alpha,\beta,\ldots$, by $-\alpha, -\beta,\ldots$, respectively,

$$V_n^{(-1)}(-\alpha) = \oint \frac{dz}{2\pi i} z^n V(-\alpha,z)\in\hat{g}^{(-1)}. \tag{3.12}$$

By commutation of $V_n^{(-1)}(-\alpha)$ we generate the elements of the space $\hat{g}^{(-2)}$

$$[V_n^{(-1)}(-\alpha),V_m^{(-1)}(-\beta)] = \begin{cases} \epsilon(-\alpha,-\beta)\,V_{m+n}^{(-2)}(-\alpha-\beta), & \text{for } \alpha\cdot\beta = -1,\\ 0, & \text{for } \alpha\cdot\beta \geqslant 0. \end{cases} \tag{3.13}$$

To obtain the elements of the space $\hat{g}^{(0)}$ we need the commutators of the elements of the spaces $\hat{g}^{(+1)}$ and $\hat{g}^{(-1)}$

$$[V_m^{(+1)}(\alpha),V_n^{(-1)}(-\beta)] = \begin{cases} 0, & \text{for } \alpha\cdot\beta = -1,\\ \epsilon(\alpha,-\beta)\,V_{m+n}^{(0)}(\alpha-\beta), & \text{for } \alpha\cdot\beta = +1,\\ \alpha\cdot h_{m+n} + m\delta_{m+n,0}, & \text{for } \alpha=\beta. \end{cases} \tag{3.14}$$

The central charge and Cartan subalgebra generators belong to the space $\hat{g}^{(0)}$. Note that the root vectors $(\alpha+\beta)$ that label the space $\hat{g}^{(+2)}$ all have length squared 2 since for every $V_m^{(+2)}(\alpha+\beta)$ we have $(\alpha+\beta)\cdot(\alpha+\beta) = \alpha\cdot\alpha + \beta\cdot\beta + 2\alpha\cdot\beta = 2$, using the fact that $\alpha\cdot\beta = -1$. Therefore we have

$$[V_m^{(+2)}(\alpha+\beta),V_n^{(-2)}(-\gamma-\delta)] = \begin{cases} 0, & \text{for } (\alpha+\beta)\cdot(\gamma+\delta) = -1,\\ \epsilon(\alpha+\beta,-\gamma-\delta)\,V_{m+n}^{(0)}(\alpha+\beta-\gamma-\delta), & \text{for } (\alpha+\beta)\cdot(\gamma+\delta) = +1,\\ (\alpha+\beta)\cdot h_{m+n} + m\delta_{m+n,0}, & \text{for } \alpha+\beta=\gamma+\delta. \end{cases} \tag{3.15}$$

Denoting the elements $V_m^{(+1)}(\alpha)$ as $J_m(\alpha)$ and the elements $V_m^{(-1)}(-\alpha)$ as $J_m(-\alpha)$, we define a ternary product $(J_m(\alpha),\tilde{J}_n(\beta),J_p(\gamma))$ over the space $\hat{g}^{+1}$ as

$$(J_m(\alpha),\tilde{J}_n(\beta),J_p(\gamma)) \equiv [[J_m(\alpha),J_n(-\beta)],J_p(\gamma)]. \tag{3.16}$$

Under this product the subspace $\hat{g}^{(+1)}$ of $\hat{g}$ becomes an "affine" ternary algebra $\hat{\mathscr{T}}$ satisfying the graded extension of the identities (2.6a) and (2.6b):

$$\begin{aligned} &(J_m(\alpha),\tilde{J}_n(\beta)(J_p(\gamma),\tilde{J}_q(\delta),J_r(\epsilon)))\\ &\quad - (J_p(\gamma),\tilde{J}_q(\delta),(J_m(\alpha),\tilde{J}_n(\beta),J_r(\epsilon)))\\ &\quad = (J_m(\alpha),\widetilde{(J_q(\delta),\tilde{J}_p(\gamma),J_n(\beta))},J_r(\epsilon))\\ &\quad - ((J_p(\gamma),\tilde{J}_q(\delta),J_m(\alpha)),\tilde{J}_n(\beta),J_r(\epsilon)), \end{aligned} \tag{3.17a}$$

$$\begin{aligned} &\{(J_m(\alpha),\tilde{J}_r(\epsilon),(J_p(\gamma),\tilde{J}_n(\beta),J_q(\delta)))\\ &\quad - ((J_p(\gamma),\tilde{J}_n(\beta),J_q(\delta)),\tilde{J}_r(\epsilon),J_m(\alpha))\\ &\quad + (J_m(\alpha),\tilde{J}_n(\beta),(J_p(\gamma),\tilde{J}_r(\epsilon),J_q(\delta)))\\ &\quad + (J_p(\gamma),\widetilde{(J_n(\beta),\tilde{J}_m(\alpha),J_r(\epsilon))},J_q(\delta))\}\\ &\quad - \{J_p(\gamma)\leftrightarrow J_q(\delta)\} = 0. \end{aligned} \tag{3.17b}$$

These two identities follow directly from the Jacobi identities of $\hat{g}$. The important fact to stress here is that the cocycle factors of $\hat{g}$ induce the right cocycle factors in $\hat{\mathscr{T}}$ so as to satisfy the above identities and lead to the definition of a natural affine extension of the ternary algebra that underlies the finite-dimensional Lie algebra g. A given Lie algebra g may have inequivalent Kantor (five-graded) decompositions corresponding to inequivalent underlying ternary algebras. Thus through the above construction one may obtain different affine ternary algebras starting from the same affine Kac–Moody algebra.

So far our discussion was centered around the simply laced affine Kac–Moody algebras and their realization using bosonic vertex operators. However, it must be clear from the above analysis that we can start from a general vertex operator (bosonic and/or fermionic) realization of any affine Kac–Moody algebra $\hat{g}$ (not necessarily simply laced) and construct the underlying affine ternary algebra using our methods. One may also choose different level representations of $\hat{g}$ to construct $\hat{\mathscr{T}}$. Different realizations of $\hat{g}$ lead to different realizations of $\hat{\mathscr{T}}$, in general.

This construction extends to superalgebras in a straightforward manner. Starting from a vertex operator realization of an affine super Kac–Moody algebra with a Kantor structure, one can construct the underlying affine superternary algebra just as explained above. The construction of finite-dimensional Lie superalgebras from superternary algebras was given in Refs. 8 and 9.

If the Lie algebra g has a Jordan structure (three-grading) with respect to its subalgebra $g^{(0)}$, then the ternary algebra $\hat{\mathscr{T}}$ constructed by our method is an affine Jordan triple system (JTS),

$$\hat{g} = \hat{g}^{(-1)} \oplus \hat{g}^{(0)} \oplus \hat{g}^{(+1)}. \tag{3.18}$$

The triple product defined by (3.16) then becomes the Jordan triple product,

$$(J_m(\alpha),\tilde{J}_n(\beta),J_p(\gamma)) = \{J_m(\alpha),\tilde{J}_n(\beta),J_p(\gamma)\},$$

and satisfies symmetry property

$$\{J_m(\alpha),\tilde{J}_n(\beta),J_p(\gamma)\} = \{J_p(\gamma),\tilde{J}_n(\beta),J_m(\alpha)\}$$

and the identity (3.17a). Similarly, if $\hat{g}$ is an affine super Kac–Moody algebra, the corresponding ternary algebra $\hat{\mathscr{T}}$ will be a super Jordan triple system satisfying (super) graded versions of the identities (3.17a) and (3.18).[8,9]

IV. VERTEX OPERATOR CONSTRUCTION OF NONASSOCIATIVE ALGEBRAS AND THEIR AFFINIZATIONS

In many cases the ternary algebra $\mathscr{T}$ associated with a Lie algebra g can be defined in terms of an underlying (binary) algebra $\mathscr{A}$. In such cases, the method explained in the previous section can be used to define and construct the affine extension $\hat{\mathscr{A}}$ of the underlying algebra $\mathscr{A}$, which is, in general, nonassociative. (Note the distinction between the terms not associative and nonassociative algebras. The latter includes both the not associative and the associative algebras.) To achieve this one simply chooses a fixed element $J_0(e)$ of the subspace $g^{(+1)}$ and defines the binary product $*$ of any two elements $J_m(\alpha)$ and $J_n(\beta)$ as

$$J_m(\alpha)*J_n(\beta) = (J_m(\alpha),\tilde{J}_0(e),J_n(\beta)). \tag{4.1}$$

Then the elements of $\hat{g}^{(+1)}$ under the $*$ product generate the affine extension of the underlying binary algebra $\mathscr{A}$. Different choices of the fixed element $J_0(e)$ lead to different "isotopes" of the algebra $\hat{\mathscr{A}}$. For example, a large class of Jordan triple systems are defined by Jordan algebras J under the Jordan triple product $\{\ ,\ ,\ \}$ defined in terms of the Jordan product $\circ$ as

$$\{a,b,c\} = a\circ(b\circ c) + c\circ(a\circ b) - b\circ(a\circ c). \tag{4.2}$$

If one chooses b to be the identity element $\mathbb{1}$ of J, then the triple product $\{a,\mathbb{1},c\}$ is simply equal to the Jordan product $a\circ c$. By defining the symmetric product between any two elements a and c as $\{apc\}$, where p is a fixed element, one generates an isotope of the Jordan algebra J.[18] To give a vertex operator construction of the affine Jordan algebra $\hat{J}$, we start from the vertex operator construction of the corresponding affine Lie algebra $\hat{g}$ and define the Jordan product of the elements of the $\hat{g}^{(+1)}$ subspace of $\hat{g}$ as

$$J_m(\alpha)\circ J_n(\beta) \equiv \{J_m(\alpha),\tilde{J}_0(e),J_n(\beta)\}, \tag{4.3}$$

where $J_0(e)$ is again a fixed element of $\hat{g}^{(+1)}$. Different choices of this fixed element lead to different isotopes of the affine Jordan algebra $\hat{J}$. In Table I we give a complete list of simple affine Jordan algebras $\hat{J}$ and the corresponding affine Kac–Moody algebras $\hat{g}$. This list follows directly from the list of simple Jordan algebras that can be obtained via the Tits–Koecher–Kantor construction.[19]

Similarly one can construct all the affine simple Jordan superalgebras by our method. In Table II we give the complete list of affine simple Jordan superalgebras $\hat{J}$ and the corresponding affine super Kac–Moody algebras. This list follows frim the super generalization of the Tits–Koecher construction[8,9] and the classification of finite-dimensional simple Jordan superalgebras.[20] The complete construction of all finite-dimensional Lie superalgebras from Jordan superalgebras was given in Ref. 9.

If the affine Kac–Moody algebra $\hat{g}$ has a Kantor structure (five-grading) with respect to its maximal subalgebra $\hat{g}^{(0)}$, then the product $*$ of the underlying binary algebra is, in general, not symmetric. Consider, for example, the affine extensions of the Lie algebras g of the magic square.[21] With the exception of SO(3) they all have a Kantor decomposition with respect to some subalgebra $g^{(0)}$. For a definite choice of $g^{(0)}$ the underlying binary algebras turn out to be the tensor products of the four division algebras R (reals), C (complex numbers), H (quaternions), and O (octonions) with each other. Thus starting from the affine Kac–Moody algebra $\hat{g}$ we obtain a vertex operator construction of the affine extensions of the tensor products of the division algebras.

The ternary product defined by (3.16) in the Lie algebra g is

$$(abc) = a(\bar{b}c) + c(\bar{b}a) - b(\bar{a}c), \tag{4.4}$$

where $a,b,c\in\mathscr{T}$. The binary product $a*c$ is defined as

$$a*c = (apc). \tag{4.5}$$

If we choose p to be the identity element then

$$a*c = ac + ca - \bar{a}c, \tag{4.6}$$

where ac refers to the usual product of a and c in the division algebras or their tensor products. Different choices for p lead to different isotopes.

In Table III we list the affine Lie algebras $\hat{g}$ corresponding to the magic square and the corresponding affine algebras.

TABLE I. The complete list of affine simple Jordan algebras $\hat{J}$ and the affine Lie algebras $\hat{g}$ from which they can be constructed: $\hat{g}$ has a Jordan structure with respect to $\hat{g}^{(0)}$; $J_n(A)$ denotes the Jordan algebra generated by $n\times n$ Hermitian matrices over the division algebra A; $\Gamma(d)$ denotes the Jordan algebra of Dirac γ matrices in d dimensions. Note that we use the same symbol for a (super) group and its Lie (super) algebra throughout the paper. The real and complex numbers are denoted as R and C while the quaternions and octonions as H and O, respectively.

$\hat{g}$	$\hat{g}^{(0)}$	$\hat{J}$
$\widehat{\mathrm{Sp}}(2n)$	$\hat{\mathrm{U}}(n)$	$\hat{J}_n(R)$
$\widehat{\mathrm{SU}}(2n)$	$\widehat{\mathrm{SU}}(n)\times\widehat{\mathrm{SU}}(n)\times\hat{\mathrm{U}}(1)$	$\hat{J}_n(C)$
$\widehat{\mathrm{SO}}(4n)$	$\hat{\mathrm{U}}(2n)$	$\hat{J}_n(\mathrm{H})$
$\widehat{\mathrm{SO}}(d+3)$	$\widehat{\mathrm{SO}}(d+1)\times\widehat{\mathrm{SO}}(2)$	$\hat{\Gamma}(d)$
$\hat{\mathrm{E}}_7$	$\hat{\mathrm{E}}_6\times\hat{\mathrm{U}}(1)$	$\hat{J}_3(O)$

TABLE II. The complete list of affine simple Jordan superalgebras $\hat{J}$ and the affine Lie superalgebras $\hat{g}$ from which they can be constructed. Again $\hat{g}$ has a Jordan structure with respect to its subalgebra $\hat{g}^{(0)}$. Our labeling of the Jordan superalgebras follows that of Kac (Ref. 20).

$\hat{g}$	$\hat{g}^{(0)}$	$\hat{J}$
$\widehat{\mathrm{SU}}(2m/2n)$	$\widehat{\mathrm{SU}}(m/n)\times\widehat{\mathrm{SU}}(m/n)\times\hat{\mathrm{U}}(1)$	$\hat{A}^s(m,n)$
$\widehat{\mathrm{OSp}}(4n/2m)$	$\widehat{\mathrm{SU}}(m/2n)\times\hat{\mathrm{U}}(1)$	$\widehat{BC}^s(m,2n)$
$\widehat{\mathrm{OSp}}(m+2/2n)$	$\widehat{\mathrm{OSp}}(m/2n)\times\hat{\mathrm{U}}(1)$	$\hat{D}^s(m,2n)$
$\hat{P}(2n-1)$	$\widehat{\mathrm{SU}}(n/n)$	$\hat{P}^s(n,n)$
$\hat{Q}(2n-1)$	$\hat{Q}(n-1)\times\hat{Q}(n-1)\times\hat{Q}(1)$	$\hat{Q}^s(n,n)$
$\hat{D}(2/1;\alpha)$	$\widehat{\mathrm{SU}}(2/1)\times\hat{\mathrm{U}}(1)$	$\hat{D}^s_\alpha$
$\hat{F}(4)$	$\widehat{\mathrm{OSp}}(2/4)\times\hat{\mathrm{U}}(1)$	$\hat{F}^s$
$\widehat{\mathrm{SU}}(2/2)$	$\widehat{\mathrm{SU}}(2/1)$	$\hat{K}$

The Kantor construction of the Lie algebras of the magic square extends to the construction of other classical Lie algebras from ternary algebras defined over the higher tensor product spaces of *associative* division algebras.[8] Thus starting from the vertex operator construction of the corresponding affine Kac–Moody algebras one can realize the affine extensions of the higher tensor products of associative division algebras. For example, the affine Lie algebra $\widehat{\mathrm{SO}}(3\times 2^n)$ has a Kantor decomposition with respect to its subalgebra $\widehat{\mathrm{SO}}(2^n)\times\widehat{\mathrm{SU}}(2^n)\times\hat{\mathrm{U}}(1)$ for even n and through our construction gives the affine extension $\hat{\mathscr{T}}=\hat{H}^n$ of the tensor product of n copies of the quaternion algebra $H(H^n\equiv H\otimes\cdots\otimes H)$. Similarly for odd n, the affine Lie algebra $\widehat{\mathrm{Sp}}(3\times 2^n)$ has a Kantor decomposition with respect to its subalgebra $\widehat{\mathrm{Sp}}(2^n)\times\widehat{\mathrm{SU}}(2^n)\times\hat{\mathrm{U}}(1)$ and leads to a vertex operator construction of $\hat{\mathscr{T}}=\hat{H}^n$ for odd n.

The unified construction of Lie algebras and Lie superalgebras given in Ref. 8 associates with a given Lie algebra, a corresponding Lie superalgebra in general. The only exceptions occur for the exceptional Lie algebras E_6, E_7, and E_8, which have no superanalogs. The Lie superalgebras of the "super magic square" were constructed from ternary algebras defined over the tensor product space of division algebras with Grassmann coefficients.[8] Thus the vertex operator construction of the affine super Kac–Moody algebras of the super magic square can be used to realize the affine extension of the corresponding tensor products of division algebras with Grassmann coefficients.

TABLE III. The affine Lie algebras $\hat{g}$ of the magic square and the subalgebras $\hat{g}^{(0)}$ with respect to which they have a Kantor structure. The last column lists the affine algebras that they give through our construction.

$\hat{g}$	$\hat{g}^{(0)}$	$\hat{\mathscr{T}}$
$\widehat{\mathrm{SO}}(3)$	$\hat{\mathrm{U}}(1)$	$\hat{R}$
$\widehat{\mathrm{SU}}(3)$	$\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)$	$\hat{C}$
$\widehat{\mathrm{Sp}}(6)$	$\widehat{\mathrm{SO}}(4)\times\hat{\mathrm{U}}(1)$	$\hat{H}$
$\hat{F}_4$	$\widehat{\mathrm{SO}}(7)\times\hat{\mathrm{U}}(1)$	$\hat{O}$
$\widehat{\mathrm{SU}}(3)\times\widehat{\mathrm{SU}}(3)$	$\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)$	$\widehat{C\otimes C}$
$\widehat{\mathrm{SU}}(6)$	$\widehat{\mathrm{SO}}(4)\times\widehat{\mathrm{SU}}(2)\times\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)$	$\widehat{C\otimes H}$
$\hat{E}_6$	$\widehat{\mathrm{SO}}(8)\times\hat{\mathrm{U}}(1)\times\hat{\mathrm{U}}(1)$	$\widehat{C\otimes O}$
$\widehat{\mathrm{SO}}(12)$	$\widehat{\mathrm{SO}}(6)\times\widehat{\mathrm{SO}}(4)\times\hat{\mathrm{U}}(1)$	$\widehat{H\otimes H}$
$\hat{E}_7$	$\widehat{\mathrm{SO}}(10)\times\widehat{\mathrm{SU}}(2)\times\hat{\mathrm{U}}(1)$	$\widehat{H\otimes O}$
$\hat{E}_8$	$\widehat{\mathrm{SO}}(14)\times\hat{\mathrm{U}}(1)$	$\widehat{O\otimes O}$

So far we have concentrated on the vertex operator construction of the affine extensions of binary algebras that underlie ternary algebras. An important class of ternary algebras that has no underlying binary algebras are the Freudenthal triple systems[22] which through the construction outlined in Sec. II lead to the exceptional Lie algebras F_4, E_6, E_7, and E_8. The elements of the Freudenthal triple system can be represented by 2×2 formal "matrices"

$$\begin{pmatrix} \alpha_1 & J_1 \\ J_2 & \alpha_2 \end{pmatrix}, \tag{4.7}$$

where $\alpha_1,\alpha_2\in R$ and J_1 and J_2 are elements of a simple Jordan algebra of degree 3, i.e., $J_3(R)$, $J_3(C)$, $J_3(H)$, and $J_3(O)$. Under a suitable triple product these matrices close and form a ternary algebra satisfying conditions (2.6a) and (2.6b). Thus starting from a vertex operator representation of $\hat{F}_4$, $\hat{E}_6$, $\hat{E}_7$, and $\hat{E}_8$ we can construct the affine extensions of the corresponding Freudenthal triple systems.

V. GENERALIZATIONS AND DISCUSSION

Above we have given the constructions of affine ternary algebras and the corresponding binary algebras whenever they exist. Our starting point was the vertex operator realization of an affine Kac–Moody algebra $\hat{g}$ (or superalgebra) which had either a Jordan or a Kantor decomposition with respect to the subalgebra $\hat{g}^{(0)}$. From the known classifications of Lie algebras g with a Jordan[10] or a Kantor structure[7,8] we can immediately give a complete list of all the affine ternary algebras or binary algebras that can be constructed by our method. In Sec. IV, we listed some important classes of these algebras. Using the methods and results of Ref. 8, one can similarly give a complete list of affine superternary algebras or affine super binary algebras that can be constructed from affine super Kac–Moody algebras with a Kantor or a Jordan decomposition.

One can easily generalize our construction to those cases when $\hat{g}$ has a higher grading with respect to its subalgebra $\hat{g}^{(0)}$ of maximal rank:

$$\hat{g}=\hat{g}^{(-k)}\oplus\hat{g}^{(-k+1)}\oplus\cdots\oplus\hat{g}^{(0)}\oplus\cdots\oplus\hat{g}^{(k-1)}\oplus\hat{g}^{(k)},$$

where k is some positive integer determined by $\hat{g}$ and $\hat{g}^0$. To our knowledge, a complete classification of such higher gradings of semisimple Lie algebras has not appeared in literature. The most extensive study of such higher gradings was given by Kantor.[7] These higher-dimensional gradings will give rise to the vertex operator construction of different affine ternary or binary algebras. The work of Kantor implies that a rank l Lie algebra has l inequivalent graded decompositions. Therefore a given Lie algebra or superalgebra g of rank l can be constructed from l different ternary algebras. Conversely, through the generalization of the construction given in this paper, one can construct l inequivalent affine ternary (super) algebras starting from a vertex operator realization of an affine (super) Kac–Moody algebra defined by a Lie (super) algebra of rank l.

Recently Goddard *et al.*[23] have established a connection between certain fermionic vertex operators and the division algebras. We believe that a precise relation between their results and our construction of the division algebras and their tensor products given in Sec. III can be established by starting from the fermionic construction of the affine Kac–Moody algebras of the magic square. For example, for the affine $\hat{E}_8$ this corresponds to the fermionic realization that was given for the heterotic string.[24] The bosonic vertex operator whose moments give the affine $\widehat{O \otimes O}$ can be written as a product of fermionic vertex operators corresponding to a single octonion field $\hat{O}$. For $\hat{F}_4$ the corresponding construction gives directly the affine $\hat{O}$. However, we should stress that our framework is very general, in that an algebraic structure underlying a Lie algebra g or its affine extension $\hat{g}$ can be constructed starting from any realization of $\hat{g}$. Furthermore, if g is of rank l then our construction leads to l inequivalent underlying algebraic structures or their affine extensions. A detailed construction of some of the algebraic structures relevant to string theory will be the subject of separate studies.

[1]S. Fubini and G. Veneziano, Nuovo Cimento A **64**, 811 (1969).

[2]I. Frenkel and V. G. Kac, Invent. Math. **62**, 23 (1980).

[3]The first constructions of Kac–Moody algebras using vertex operators appeared in K. Bardakci and M. Halpern, Phys. Rev. D **3**, 2493 (1971). The construction of the affine algebra $A_1^{(1)}$ was given in J. Lepowski and R. L. Wilson, Commun. Math. Phys. **62**, 43 (1978).

[4]G. Segal, Commun. Math. Phys. **80**, 301 (1981).

[5]P. Goddard, W. Nahm, D. Olive, and A. Schwimmer, Commun. Math. Phys. **107**, 179 (1986).

[6]D. Bernard and J. Thierry-Mieg, Commun. Math. Phys. **111**, 181 (1987).

[7]I. L. Kantor, Sov. Math. Dokl. **14**, 254 (1973); Trudy Sem. Vektor. Tenzor Anal. **16**, 407 (1972).

[8]I. Bars and M. Günaydin, J. Math. Phys. **20**, 1977 (1979).

[9]M. Günaydin, Ann. Israel Phys. Soc. **3**, 279 (1980).

[10]For a complete list of references on Jordan triple systems, see N. Jacobson, *Structure and Representations of Jordan Algebras* (Am. Math. Soc., Providence, RI, 1968), Publ. XXXIX, and the articles by E. Neher, Manuscripta Math. **31**, 197 (1980) and J. Reine, Angew. Math. **322**, 145 (1981).

[11]J. Tits, Nederl. Akad. Wetensch. Proc. Ser. A **65**, 530 (1962).

[12]M. Koecher, Am. J. Math. **89**, 787 (1967).

[13]I. L. Kantor, Dokl. Akad. Nauk. SSSR **158**, 1271 (1964).

[14]O. Loos, *Jordan Pairs, Lecture Notes in Mathematics*, Vol. 460 (Springer, New York, 1975).

[15]K. McCrimmon, Bull. Am. Math. Soc. **84**, 612 (1978).

[16]F. Gürsey, invited paper at the Conference on Non-associative Algebras, Univ. of Virginia, Charlottesville, VA 1977 (unpublished); M. Günaydin, invited paper at the second Johns Hopkins Workshop, Baltimore, 1978, edited by G. Domokos and S. Kövesi-Domokos; M. Günaydin, Ref. 9; L. C. Biedenharn and L. P. Horwitz, in *Differential Geometric Methods in Physics*, Springer Lectures, Vol. 139, edited by D. H. Doebner (Springer, New York, 1981); M. Günaydin, G. Sierra, and P. K. Townsend, Phys. Lett. B **133**, 72 (1983); Nucl. Phys. B **242**, 244 (1984); P. Truini, G. Olivieri, and L. C. Biedenharn, Lett. Math. Phys. **9**, 255 (1985).

[17]P. Goddard, D. Olive, and G. Waterson, Commun. Math. Phys. **112**, 591 (1987).

[18]See, for example, the first citation in Ref. 10.

[19]See, for example, K. Meyberg, Math. Z. **115**, 58 (1970).

[20]V. G. Kac, Commun. Algebra **5**, 1375 (1977).

[21]H. Freudenthal, Proc. Koninkl. Ned. Akad. Wetenschap A **62**, 447 (1959); B. A. Rozenfeld, Dokl. Adak. Nauk SSSR **106**, 600 (1956); J. Tits, Mem. Acad. R. Belg. Sci. **29**, 3 (1955).

[22]H. Freudenthal, Proc. Koninkl. Ned. Akad. Wetenschap A **57**, 363 (1954); K. Meyberg, *ibid.* **71**, 162 (1968); J. Faulkner, Mem. Am. Math. Soc. **10**, 185 (1977).

[23]P. Goddard, W. Nahm, D. I. Olive, H. Ruegg, and A. Schwimmer, Commun. Math. Phys. **112**, 385 (1987).

[24]D. Gross, J. Harvey, E. Martinec, and R. Rohm, Nucl. Phys. B **256**, 253 (1985).

On the symmetric representations of SU(5). Matrix elements of the generators in the SO(3) basis

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

(Received 22 June 1988; accepted for publication 14 December 1988)

By making use of the results of a previous paper [J. Math. Phys. **29**, 1958 (1988)], in which the matrix elements of the SU(5) generators were found in closed form for symmetric representations and in an SO(4) labeled state basis, it is shown how the matrix elements of the generators can also be derived in the physical SO(3) state basis. The main step consists in studying the action of the SU(5) generators upon a particular subset of so-called intrinsic SU(2)×SU(2) states out of which the physical SO(3) states are projected by means of the Hill–Wheeler technique.

I. INTRODUCTION

The original version of the interacting boson model[1,2] for describing even–even nuclei exhibits three dynamical symmetries, corresponding to three limits of collective motion, i.e., the limits SU(3), SO(6), and SU(5). Recently, there has been some interest in the construction of cubic interaction terms in the Hamiltonian and also of terms of higher degree in the SU(6) group generators,[3,4] which preserve one of the possible dynamical symmetries. As far as the SU(5) limit is concerned, it is therefore necessary to have at ones disposal closed form expressions of the matrix elements of the SU(5) generators in the symmetric representations of SU(5). Furthermore, from a physical point of view, it is required to label the representation states by the angular momentum labels l and m, hence to establish a physical basis consisting of SO(3) states, where SO(3) is the principal subgroup of SU(5).

In the previous paper[5] (hereafter to be referred to as I), we have obtained the matrix elements of the SU(5) generators expressed in a basis of SO(4)≈SU(2)×SU(2) states. In fact, according to the reduction chain SU(5) ⊃SO(5)⊃SO(4) the symmetric irreducible representations (irreps) of SU(5) decompose without degeneracy into SO(4) irreps, whereby one extra label is provided by the intermediate SO(5) group. In I we have also established the explicit relationships between the generators of SU(5) either realized as SU(2)×SU(2) bitensors or as SO(3) tensors. The main problem, however, remains in relating an SO(3) state basis with the SU(2)×SU(2) state basis, whereby it should be noticed that the principal SO(3) subgroup of SU(5) is indeed a subgroup of SO(5) but not of SU(2)×SU(2). The same problem has been encountered in our recent study of the SO(6) limit of the interacting boson model[4] and the way of solution may be drawn back to the work of Kemmer, Pursey, and Williams[6,7] concerning the irreducible representations of the rotation group SO(5). Hence we shall apply again the so-called Hill–Wheeler technique that consists in defining a restricted subset of SU(2)×SU(2) states, called intrinsic states, out of which is projected the complete set of physical SO(3) states by integrations over the physical SO(3) group manifold. With this technique there remains the problem of expanding every state that one obtains by letting an SU(5) generator act upon an intrinsic SU(2)×SU(2) state, in terms of intrinsic states and of states obtained by letting operator products of SO(3) generators [more precisely elements of the SO(3) enveloping algebra] upon intrinsic states. This task is carried out in Sec. III. In Sec. IV we present all the formulas that permit us to establish the matrix elements of the SU(5) generators in the SO(3) basis.

II. SU(5) IN THE SO(3) BASIS

In I it has been shown that the symmetric SU(5) irreducible representations (irreps) $[n,0,0,0]$ decompose into symmetric irreps of SO(5) according to the rule

$$\mathrm{SU}(5)\to\mathrm{SO}(5):[n,0,0,0]\to\sum_{\tau}[\tau,0],\quad \text{with } \tau=n,n-2,\dots,1 \text{ or } 0. \tag{2.1}$$

Furthermore, we have taken advantage of the fact that these SO(5) irreps themselves decompose without degeneracies into SU(2)×SU(2) irreps, i.e.,

$$\mathrm{SO}(5)\to\mathrm{SU}(2)\times\mathrm{SU}(2):\ [\tau,0]\to\sum_{s=t}(s,t)\ \text{ with } s=t=\tau/2,\tau/2-\tfrac{1}{2},\tau/2-1,\dots,0, \tag{2.2}$$

in order to calculate all the matrix elements of the generators of SU(5) in the orthonormal SU(2)×SU(2) state basis

$$|n,\tau,s,m_s,s,m_t),(m_s,m_t=-s,-s+1,\dots,s). \tag{2.3}$$

In this paper we are concerned with the reduction of the SO(5) irreps $[\tau,0]$ into irreps of the physical SO(3) algebra, which is the principal SO(3) subalgebra of SO(5). This reduction is not without degeneracies, and more precisely one extra label is

[a] Research Director at the National Fund for Scientific Research (NFWO), Belguim.

required to distinguish between the SO(3) states. As an extra label, ν say, we choose the one that is usually considered in the interacting boson model[1,2] and which is obtained in a way closely analogous to that of Elliott's SU(3) model,[8] i.e.,

$$\begin{aligned}
&\mathrm{SO}(5)\to\mathrm{SO}(3):\ [\tau,0]\to\sum_{l,m,\nu}(l,m,\nu)\\
&\qquad\text{with } \nu=0,1,2,\ldots,[\tau/3]\\
&\qquad l=2(\tau-3\nu),2(\tau-3\nu)-2,2(\tau-3\nu)-3,\ldots,\tau-3\nu+1,\tau-3\nu,\\
&\qquad m=-l,-l+1,\ldots,+l.
\end{aligned}\tag{2.4}$$

In fact, the label ν can be related to the angular momentum projection of SU(2)×SU(2) intrinsic states, introduced further on. As a consequence of the branching rule (2.4), in any SU(5) symmetric irrep $[n,0,0,0]$ the SO(3) states that constitute a physical basis are labeled as

$$|n,\tau,\nu,l,m\rangle,\tag{2.5}$$

where one has to take into account the restrictions imposed by (2.1) and (2.4). Clearly, the basis spanned by the states (2.5) is orthogonal except for the label ν. Explicitly,

$$\langle n\tau'\nu'l'm'|n\tau\nu lm\rangle=\delta_{\tau'\tau}\delta_{l'l}\delta_{m'm}A_l^\tau(\nu',\nu).\tag{2.6}$$

Several equivalent formulas are available for the calculation of the overlap integrals $A_l^\tau(\nu',\nu)$. For the sake of self-containedness we mention the following expression taken from Ref. 7:

$$\begin{aligned}
A_l^\tau(\nu',\nu)&=2^{\nu'-\nu}[(2l+1)(3\nu'-3\nu)!]^{-1}[(\tau-\nu)!(\tau-\nu')!\nu!\nu'!(l+3\nu'-\tau)!(l+\tau-3\nu)!]^{1/2}\\
&\times[(l+\tau-3\nu')!(l+3\nu-\tau)!]^{-1/2}\sum_{\alpha,\beta}(-4)^{\nu+\alpha-\beta}(3\nu'-3\beta+\alpha)!(2\tau-2\nu-2\nu'+2\beta)!\\
&\times[(\tau-\nu-\nu'+\beta-\alpha)!(\nu'-\beta)!(\nu-\beta)!\alpha!\beta!(2\tau+\nu'-2\nu+\alpha-\beta+1)!]^{-1}\\
&\times{}_3F_2(\tau-3\nu-l,\tau+l,-3\nu+1,3\nu'-3\beta+\alpha+1;3\nu'-3\nu+1,2\tau+\nu'-2\nu-\beta+\alpha+2;1).
\end{aligned}\tag{2.7}$$

Herein ${}_3F_2$ is a generalized hypergeometric function that in the present case necessarily reduces to a polynomial.

It is our aim to calculate the matrix elements of the SU(5) generators between the physical states (2.5). Let us recall that with respect to SO(3) the adjoint irrep [1001] of SU(5) decomposes into the SO(3) irreps $(1)+(2)+(3)+(4)$, proving that the SU(5) algebra can be realized in terms of SO(3) tensors G_μ^k ($k=1,2,3,4$, $\mu=-k,-k+1,\ldots,+k$). These tensors satisfy the commutation relations (I 2.1). The standard commutation relations amongst the SO(3) subalgebra generators $l_0,l_{\pm1}$ are retrieved by making the identification $l_\mu=\sqrt{10}G_\mu^1$. In I we have introduced a realization of the SU(5) algebra in terms of SU(2)×SU(2) bitensors $s_\mu^{(1,0)}$, $t_\nu^{(1,0)}$, $Z_{00}^{(0,0)}$, $T_{\alpha\beta}^{(1/2,1/2)}$, $U_{\alpha\beta}^{(1/2,1/2)}$, and $V_{\mu\nu}^{(1,1)}$ ($\mu,\nu=-1,0,+1$, $\alpha,\beta=-\frac12,+\frac12$), and all the matrix elements of these generators have been established in the SU(2)×SU(2) state basis (2.3). Henceforth, through the relationships (I 2.6), which for the present purposes are reformulated as

$$\begin{aligned}
&l_0=3s_0+t_0,\\
&l_{\pm1}=2t_{\pm1}\mp\sqrt3T_{\pm1/2\mp1/2},\\
&G_0^3=(1/\sqrt{10})(-s_0+3t_0),\\
&G_{\pm1}^3=(1/\sqrt5)(\sqrt3t_{\pm1}\pm T_{\pm1/2\mp1/2}),\\
&G_{\pm2}^3=\pm(1/\sqrt2)T_{\pm1/2\pm1/2},\\
&G_{\pm3}^3=s_{\pm1},\\
&G_0^2=(1/2\sqrt{14})(5Z_{00}+6V_{00}),\\
&G_{\pm1}^2=(1/\sqrt{14})(-U_{\pm1/2\mp1/2}+2\sqrt3V_{0\pm1}),\\
&G_{\pm2}^2=(1/\sqrt7)(\sqrt2U_{\pm1/2\pm1/2}-\sqrt3V_{\pm1\mp1}),\\
&G_0^4=(\sqrt5/2\sqrt{14})(-3Z_{00}+2V_{00}),\\
&G_{\pm1}^4=(1/\sqrt7)(\sqrt3U_{\pm1/2\mp1/2}+V_{0\pm1}),\\
&G_{\pm2}^4=(1/\sqrt{14})(\sqrt3U_{\pm1/2\pm1/2}+2\sqrt2V_{\pm1\mp1}),\\
&G_{\pm3}^4=V_{\pm10},\\
&G_{\pm4}^4=V_{\pm1\pm1},
\end{aligned}\tag{2.8}$$

we also know the matrix elements of the tensor components G_μ^k in that same basis (2.3). Let us recall that $\{l_\mu,G_\mu^3\}$ constitutes the SO(5) subalgebra. The matrix elements of the SO(5) generators in the SO(3) basis have been completely determined by Kemmer *et al.*[6,7] Hence we may concentrate on the calculation of the matrix elements of the remaining tensors G^2 and G^4.

It is well known from the work of Williams and Pursey[7] that the physical states can be projected out of the set of SU(2)×SU(2) intrinsic states $\chi(n,\tau,\nu)$ defined by

$$\begin{aligned}
\chi(n,\tau,\nu)&=|n,\tau,s=\tau/2,m_s\\
&=\tau/2-\nu,t=\tau/2,m_t=-\tau/2)
\end{aligned}$$

with

$$\tau=n,n-2,\ldots,1\text{ or }0\text{ and }\nu=0,1,2,\ldots,[\tau/3].\tag{2.9}$$

It follows from the first equality in (2.8) that $l_0\chi(n,\tau,\nu)=(\tau-3\nu)\chi(n,\tau,\nu)$, showing that $\tau-3\nu$ is the m value of the intrinsic state $\chi(n,\tau,\nu)$. The restrictions imposed by the Hill–Wheeler projection technique[9] imply that, in order to calculate closed form expressions for the matrix elements of the SU(5) generators in the SO(3) basis, we need to express the action of these generators upon intrinsic states (2.9), either in terms of pure intrinsic states again, or in terms of intrinsic states upon which one or more SO(3) generators l_μ operate. In the next section this task will be carried out in some detail.

III. THE ACTION OF SU(5) GENERATORS UPON INTRINSIC STATES

The action of the SO(5) generators $s_\mu, t_\mu, T_{\alpha\beta}$ upon intrinsic states $\chi(n,\tau,\nu)$ has been expressed previously in terms on intrinsic states. Let us, for further use, recall certain results

$$t_{+1}\chi(n,\tau,\nu) = \tfrac{1}{2}l_{+1}\chi(n,\tau,\nu), \tag{3.1}$$

$$T_{-1/2\;1/2}\chi(n,\tau,\nu) = (1/\sqrt{3})l_{-1}\chi(n,\tau,\nu), \tag{3.2}$$

$$T_{1/2\;1/2}\chi(n,\tau,\nu) = -[3(\tau-\nu+1)/\nu]^{-1/2}l_{-1}\chi(n,\tau,\nu-1). \tag{3.3}$$

Notice, that we have dropped the bitensor superscripts.

As an example, we shall study in some detail the action of Z_{00} upon $\chi(n,\tau,\nu)$. Since Z_{00} behaves as an SU(2)$\times$SU(2) biscalar and on the other hand it is an SU(5) generator not in SO(5), it can change the τ value to $\tau'=\tau+2$, $\tau'=\tau-2$ or $\tau'=\tau$, and one can immediately propose that

$$\begin{aligned} Z_{00}\chi(n,\tau,\nu) &= Z_{00}|n,\tau,\tau/2,\tau/2-\nu,\tau/2,-\tau/2) \\ &= a|n,\tau,\tau/2,\tau/2,-\nu,\tau/2,-\tau/2) \\ &\quad + b\,|n,\tau+2,\tau/2,\tau/2-\nu,\tau/2,-\tau/2) \end{aligned} \tag{3.4}$$

whereby the coefficients a and b can be calculated by the application of the well-known relationship

$$\begin{aligned} &(n,\tau',s',m_s',t',m_t'|X^{uv}_{\mu\nu}|n,\tau,s,m_s t,m_t) \\ &= (-1)^{s'+t'-m_s'-m_t'}\begin{pmatrix} s' & u & s \\ -m_s' & \mu & m_s\end{pmatrix} \\ &\quad\times\begin{pmatrix} t' & v & t \\ -m_t' & \nu & m_t\end{pmatrix} \\ &\quad\times(n,\tau',s',t'\|X^{uv}\|m,\tau,s,t). \end{aligned} \tag{3.5}$$

This formula is valid for any SU(2)$\times$SU(2) bitensor component $X^{uv}_{\mu\nu}$ whereas the last part on the right-hand side represents an SU(2)$\times$SU(2) reduced matrix element. It follows that

$$\begin{aligned} a &= (-1)^{\tau+\nu}\begin{pmatrix} \tau/2 & 0 & \tau/2 \\ -\tau/2+\nu & 0 & \tau/2-\nu\end{pmatrix} \\ &\quad\times\begin{pmatrix} \tau/2 & 0 & \tau/2 \\ \tau/2 & 0 & -\tau/2\end{pmatrix} \\ &\quad\times(n,\tau,\tau/2,\tau/2\|Z\|n,\tau,\tau/2,\tau/2) \\ &= (\tau+1)^{-1}(n,\tau,\tau/2,\tau/2\|Z\|n,\tau,\tau/2,\tau/2). \end{aligned}$$

Here we have made use of well-known closed form expressions for the occurring $3j$ symbols. The reduced matrix element of Z has been found in (I 5.1). The calculation of the coefficient b is similar and we arrive at the intermediate result

$$\begin{aligned} Z_{00}\chi(n,\tau,\nu) &= [\tau(2n+5)/5(2\tau+5)]\chi(n,\tau,\nu) \\ &\quad+[2/(2\tau+5)][(n-\tau)(n+\tau+5) \\ &\quad\times(\tau+2)/(2\tau+7)]^{1/2}|n,\tau+2,\tau/2,\tau/2 \\ &\quad-\nu,\tau/2,-\tau/2). \end{aligned} \tag{3.6}$$

Next, we have to express the state on the right-hand side of (3.6) in terms of intrinsic states. To that aim, we first consider

$$\begin{aligned} &T_{-1/2\;1/2}\chi(n,\tau+2,\nu) \\ &= (1/\sqrt{3})l_{-1}\chi(n,\tau+2,\nu) \\ &= (\tau+2-\nu)^{1/2}|n,\tau+2,\tau/2+\tfrac{1}{2},\tau/2 \\ &\quad+\tfrac{1}{2}-\nu,\tau/2+\tfrac{1}{2},-\tau/2-\tfrac{1}{2}), \end{aligned} \tag{3.7}$$

where the first equality follows from (3.2), and the second equality is obtained in an analogous way as for the proof of (3.6), with the use, however, of the reduced T-matrix elements as given in (I 2.14). Letting $T_{-1/2\;1/2}$ act once more upon each member of (3.7), we obtain in particular that

$$\begin{aligned} &T_{-1/2\;1/2}|n,\tau+2,\tau/2+\tfrac{1}{2},\tau/2+\tfrac{1}{2}-\nu,\tau/2+\tfrac{1}{2},-\tau/2-\tfrac{1}{2}) \\ &= [(\nu+1)/(\tau+2)]^{1/2}|n,\tau+2,\tau/2 \\ &\quad+1,\tau/2-\nu,\tau/2+1,-\tau/2) \\ &\quad+[(\tau+1-\nu)(2\tau+5)/(\tau+2)]^{1,2} \\ &\quad\times|n,\tau+2,\tau/2,\tau/2-\nu,\tau/2,-\tau/2), \end{aligned} \tag{3.8}$$

where again use has been made of formula (3.5) and of the reduced matrix elements (I 2.14).

Since $[T_{-1/2\;1/2},l_{-1}] = -\sqrt{2}T_{-1/2-1/2}$, $T_{-1/2-1/2}$ clearly vanishes when acting upon any intrinsic state, and it also follows that

$$\begin{aligned} T_{-1/2\,1/2}l_{-1}\chi(n,\tau+2,\nu) &= l_{-1}T_{-1/2\,1/2}\chi(n,\tau+2,\nu) \\ &= (1/\sqrt{3})(l_{-1})^2\chi(n,\tau+2,\nu). \end{aligned} \tag{3.9}$$

Combining (3.7)–(3.9) we arrive at the result

$$\begin{aligned} &|n,\tau+2,\tau/2-\nu,\tau/2,-\tau/2) \\ &= \tfrac{1}{3}[(\tau+2)/(\tau+1-\nu)(\tau+2-\nu)(2\tau+5)]^{1/2} \\ &\quad\times(l_{-1})^2\chi(n,\tau+2,\nu) \\ &\quad-[(\nu+1)/(\tau+1-\nu)(2\tau+5)]^{1/2} \\ &\quad\times|n,\tau+2,\tau/2+1,\tau/2-\nu,\tau/2+1,-\tau/2). \end{aligned} \tag{3.10}$$

Finally, it is easily shown that

$$\begin{aligned} t_{+1}\chi(n,\tau+2,\nu+1) &= \tfrac{1}{2}l_{+1}\chi(n,\tau+2,\nu+1) \\ &= -[(\tau+2)/2)]^{1/2}|n,\tau+2,\tau/2 \\ &\quad+1,\tau/2-\nu,\tau/2+1,-\tau/2). \end{aligned} \tag{3.11}$$

Hence putting (3.6), (3.10), and (3.11) together, we obtain the expression that brings $Z_{00}\chi(n,\tau,\nu)$ into the appropriate form, namely

$$Z_{00}\chi(n,\tau,\nu) = [\tau(2n+5)/5(2\tau+5)]\chi(n,\tau,\nu) + [2(\tau+2)/3(2\tau+5)][(n-\tau)(n+\tau+5)/(\tau+1-\nu)$$
$$\times(\tau+2-\nu)(2\tau+5)(2\tau+7)]^{1/2}(l_{-1})^2\chi(n,\tau+2,\nu) + [1/(2\tau+5)]$$
$$\times[2(n-\tau)(n+\tau+5)(\nu+1)/(\tau+1-\nu)$$
$$\times(2\tau+5)(2\tau+7)]^{1/2}l_{+1}\chi(n,\tau+2,\nu+1).$$

The calculation of the action of each of the tensor components $U_{\alpha\beta}$ and $V_{\mu\nu}$ upon $\chi(n,\tau,\nu)$ proceeds in an analogous way, although sometimes many more intermediate steps are required. After lengthy calculations one arrives at the following results:

$$U_{1/2\ -1/2}\chi(n,\tau,\nu) = -2[\Delta(\tau+1-\nu)/3(\tau+2-\nu)]^{1/2}l_{-1}\chi(n,\tau+2,\nu), \tag{3.13}$$

$$U_{-1/2\ -1/2}\chi(n,\tau,\nu) = -2[\Delta(\nu+1)/3(\tau+1-\nu)]^{1/2}l_{-1}\chi(n,\tau+2,\nu+1), \tag{3.14}$$

$$\begin{aligned}U_{1/2\ 1/2}\chi(n,\tau,\nu) &= [2/9(2\tau+5)][3\nu\Delta/(\tau+1-\nu)(\tau+2-\nu)(\tau+3-\nu)]^{1/2}(l_{-1})^3\chi(n,\tau+2,\nu-1)\\ &\quad+[(2n+5)/(2\tau+5)][\nu/3(\tau+1-\nu)]^{1/2}l_{-1}\chi(n,\tau,\nu-1)\\ &\quad+[(2\tau-2\nu+5)/(2\tau+5)][2\Delta/3(\tau+1-\nu)(\tau+2-\nu)]^{1/2}l_{-1}l_{+1}(n,\tau+2,\nu)\\ &\quad+2[2\Delta(\tau+2-\nu)/3(\tau+1-\nu)]^{1/2}\chi(n,\tau+2,\nu),\end{aligned} \tag{3.15}$$

$$\begin{aligned}&U_{-1/2\ 1/2}\chi(n,\tau,\nu)\\ &\quad=-[2/9(2\tau+5)][3\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}(l_{-1})^3\chi(n,\tau+2,\nu)-[(2n+5)/(2\tau+5)][\tfrac{1}{3}]^{1/2}l_{-1}\chi(n,\tau,\nu)\\ &\quad\quad+[2/(2\tau+5)][2\Delta(\nu+1)/3(\tau+1-\nu)]^{1/2}l_{-1}l_{+1}\chi(n,\tau+2,\nu+1)\\ &\quad\quad+2[2\Delta(\nu+1)/3(\tau+1-\nu)]^{1/2}\chi(n,\tau+2,\nu+1),\end{aligned} \tag{3.16}$$

$$V_{1\ -1}\chi(n,\tau,\nu) = -[\Delta(\tau+1-\nu)(\tau+2-\nu)]^{1/2}\chi(n,\tau+2,\nu), \tag{3.17}$$

$$V_{0\ -1}\chi(n,\tau,\nu) = -[2\Delta(\nu+1)(\tau+1-\nu)]^{1/2}\chi(n,\tau+2,\nu+1), \tag{3.18}$$

$$V_{-1\ -1}\chi(n,\tau,\nu) = -[\Delta(\nu+1)(\nu+2)]^{1/2}\chi(n,\tau+2,\nu+2), \tag{3.19}$$

$$\begin{aligned}V_{1\ 0}\chi(n,\tau,\nu) &= [(2n+5)/(2\tau+5)][\nu(\tau+1-\nu)/2]^{1/2}\chi(n,\tau,\nu-1)\\ &\quad+[1/3(2\tau+5)][2\Delta\nu(\tau+1-\nu)/(\tau+2-\nu)(\tau+3-\nu)]^{1/2}(l_{-1})^2\chi(n,\tau+2,\nu-1)\\ &\quad+[(2\tau+5-2\nu)/(2\tau+5)][\Delta\nu(\tau+1-\nu)/(\tau+2-\nu)]^{1/2}l_{+1}\chi(n,\tau+2,\nu),\end{aligned} \tag{3.20}$$

$$\begin{aligned}V_{0\ 0}\chi(n,\tau,\nu) &= -[(2n+5)(\tau-2\nu)/2(2\tau+5)]\chi(n,\tau,\nu)\\ &\quad-[(\tau-2\nu)/3(2\tau+5)][\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}(l_{-1})^2\chi(n,\tau+2,\nu)\\ &\quad+[(4\tau+5-4\nu)/(2\tau+5)][\Delta(\nu+1)/2(\tau+1-\nu)]^{1/2}(l_{+1})\chi(n,\tau+2,\nu+1),\end{aligned} \tag{3.21}$$

$$\begin{aligned}V_{-1\ 0}\chi(n,\tau,\nu) &= -[(2n+5)/(2\tau+5)][(\nu+1)(\tau-\nu)/2]^{1/2}\chi(n,\tau,\nu+1)\\ &\quad-[1/3(2\tau+5)][2\Delta(\nu+1)/(\tau+1-\nu)]^{1/2}(l_{-1})^2\chi(n,\tau+2,\nu+1)\\ &\quad+[2/(2\tau+5)][\Delta(\nu+1)(\nu+2)]^{1/2}l_{+1}\chi(n,\tau+2,\nu+2),\end{aligned} \tag{3.22}$$

$$\begin{aligned}V_{1\ 1}\chi(n,\tau,\nu) &= -[1/9(2\tau+3)(2\tau+5)]\\ &\quad\times[\Delta\nu(\nu-1)/(\tau+1-\nu)(\tau+2-\nu)(\tau+3-\nu)(\tau+4-\nu)]^{1/2}(l_{-1})^4\chi(n,\tau+2,\nu-2)\\ &\quad-[(2n+5)/3(2\tau+1)(2\tau+5)][\nu(\nu-1)/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}(l_{-1})^2\chi(n,\tau,\nu-2)\\ &\quad-[(n-\tau+2)(n+\tau+3)\nu(\nu-1)/(2\tau+1)(2\tau+3)]^{1/2}\chi(n,\tau-2,\nu-2)\\ &\quad-[(2n+5)(2\tau+3-2\nu)/(2\tau+1)(2\tau+5)][\nu/2(\tau+1-\nu)]^{1/2}l_{+1}\chi(n,\tau,\nu-1)\\ &\quad-[(2\tau+3-2\nu)(2\tau+5-2\nu)/2(2\tau+3)(2\tau+5)]\\ &\quad\times[\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}(l_{+1})^2\chi(n,\tau+2,\nu)\\ &\quad-[(2\tau+5-2\nu)/3(2\tau+3)(2\tau+5)]\\ &\quad\times[2\nu\Delta/(\tau+1-\nu)(\tau+2-\nu)(\tau+3-\nu)]^{1/2}l_{+1}(l_{-1})^2\chi(n,\tau+2,\nu-1)\\ &\quad-[(12\tau^2+72\tau-24\nu\tau+12\nu^2-80\nu+113)/6(2\tau+3)(2\tau+5)]\\ &\quad\times[2\nu\Delta/(\tau+1-\nu)(\tau+2-\nu)(\tau+3-\nu)]^{1/2}l_{-1}\chi(n,\tau+2,\nu-1),\end{aligned} \tag{3.23}$$

$$
\begin{aligned}
V_{0\,1}\chi(n,\tau,\nu) = {} & [1/9(2\tau+3)(2\tau+5)][2\Delta\nu/(\tau+1-\nu)(\tau+2-\nu)(\tau+3-\nu)]^{1/2}(l_{-1})^4\chi(n,\tau+2,\nu-1) \\
& + [(2n+5)/3(2\tau+1)(2\tau+5)][2\nu/(\tau+1-\nu)]^{1/2}(l_{-1})^2\chi(n,\tau,\nu-1) \\
& + [2\nu(\tau-\nu)(n-\tau+2)(n+\tau+3)/(2\tau+1)(2\tau+3)]^{1/2}\chi(n,\tau-2,\nu-1) \\
& + [(2n+5)(2\tau+1-4\nu)/2(2\tau+1)(2\tau+5)][l_{+1}\chi(n,\tau,\nu)] \\
& - [2(2\tau+3-2\nu)/(2\tau+3)(2\tau+5)][\Delta(\nu+1)/2(\tau+1-\nu)]^{1/2}(l_{+1})^2\chi(n,\tau+2,\nu+1) \\
& + [(2\tau+3-4\nu)/3(2\tau+3)(2\tau+5)][\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}l_{+1}(l_{-1})^2\chi(n,\tau+2,\nu) \\
& + [(12\tau^2-24\nu\tau+12\nu^2+48\tau-56\nu+45)/3(2\tau+3)(2\tau+5)] \\
& \times[\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}l_{-1}\chi(n,\tau+2,\nu), \qquad (3.24) \\
V_{-1\,1}\chi(n,\tau,\nu) = {} & [-1/9(2\tau+3)(2\tau+5)][\Delta/(\tau+1-\nu)(\tau+2-\nu)]^{1/2}(l_{-1})^4\chi(n,\tau+2,\nu) \\
& - [(2n+5)/3(2\tau+1)(2\tau+5)](l_{-1})^2\chi(n,\tau,\nu) \\
& - [(n-\tau+2)(n+\tau+3)(\tau-1-\nu)(\tau-\nu)/(2\tau+1)(2\tau+3)]^{1/2}\chi(n,\tau-2,\nu), \\
& + [(2n+5)/(2\tau+1)(2\tau+5)][2(\nu+1)(\tau-\nu)]^{1/2}l_{+1}\chi(n,\tau,\nu+1) \\
& - [2/(2\tau+3)(2\tau+5)][\Delta(\nu+1)(\nu+2)]^{1/2}(l_{+1})^2\chi(n,\tau+2,\nu+2) \\
& + [2/3(2\tau+3)(2\tau+5)][2\Delta/(\nu+1)/(\tau+1-\nu)]^{1/2}l_{+1}(l_{-1})^2\chi(n,\tau+2,\nu+1) \\
& + [2(6\tau+11-3\nu)/3(2\tau+3)(2\tau+5)][2\Delta(\nu+1)/(\tau+1-\nu)]^{1/2}l_{-1}\chi(n,\tau+2,\nu+1), \qquad (3.25)
\end{aligned}
$$

where we have introduced the shorthand notation

$$\Delta = (n-\tau)(n+\tau+5)/(2\tau+5)(2\tau+7). \qquad (3.26)$$

IV. MATRIX ELEMENTS IN THE SO(3) BASIS

We now have at our disposal all the intermediate results required for deriving matrix elements of the SU(5) generators in the SO(3) state basis (2.5). The method used to perform the remaining calculational steps is based upon the Hill–Wheeler projection technique[9] and has been applied for similar purposes by Williams *et al.*[7] and by the present authors.[4] Therefore, we only present here some essential formulas that enable us to carry out easily all the necessary calculations. The physical states (2.5) are projected out of the intrinsic states (2.9) by means of an invariant integral over the group manifold of SO(3), i.e.,

$$|n,\tau,\nu,l,m\rangle = \int D^{l*}_{m,K}(\Omega)\chi_\Omega(n,\tau,\nu)\,d\Omega \quad (K=\tau-3\nu), \qquad (4.1)$$

whereby $D^{l}_{m,K}(\Omega)$ is a rotation matrix and $\chi_\Omega(n,\tau,\nu)$ the intrinsic state on which the rotation Ω is applied. It follows that the effect of any of the SO(3) tensor zero components G^k_0 upon a physical state can be brought into the form

$$
\begin{aligned}
& G^k_0|n,\tau,\nu,l,m\rangle \\
& = \sum_{\mu,l'}\langle l\,m\,k\,o|l'\,m\rangle\langle l\,K\,k\,\mu|l'\,K+\mu\rangle \\
& \quad\times\int D^{l'*}_{m,K+\mu}(\Omega)G^k_\mu(\Omega)\chi_\Omega(n,\tau,\nu)\,d\Omega \\
& \quad (K=\tau-3\nu). \qquad (4.2)
\end{aligned}
$$

Since we are actually interested in the cases $k=2$ and $k=4$ for both situations separately, we substitute under the integral symbol the expressions for G^k_μ obtained in (2.8). Then, by means of the formulas (3.12)–(3.26), the action of the SU(2)$\times$SU(2) bitensors upon intrinsic states in the rotated frame is replaced by the action of SO(3) tensor components upon similar states. By doing this carefully, there remains the problem of the calculation of integrals of the type

$$\mathscr{F}_{a,b} = \int D^{l'*}_{m,K+\mu}(\Omega)(l_{-1})^a(l_{+1})^b\chi_\Omega(n,\tau,\nu)\,d\Omega, \qquad (4.3)$$

whereby a and b can only take non-negative integer values. It can be shown that

$$
\begin{aligned}
\mathscr{F}_{a,b} = {} & [l'(l'+1)]^{(a+b)/2} \\
& \times\prod_{\alpha=0}^{a-1}\langle l'\,K+\mu-b+\alpha+1\,1 \\
& -1|l'\,K+\mu-b+\alpha\rangle \\
& \times\prod_{\beta=0}^{b-1}\langle l'\,K+\mu-\beta-1\,1\,1|l'\,K \\
& +\mu-\beta\rangle|n,\tau,\nu,l',m\rangle \qquad (4.4)
\end{aligned}
$$

with the convention that a nonexistent product ($a=0$ or $b=0$) reduces to the unity. With the help of (4.4) all the integrations on the right-hand side of (4.2) can be done explicitly and we arrive at an expression of the form

$$
\begin{aligned}
& G^k_0|n,\tau,\nu,l,m\rangle \\
& = \sum_{l''\nu''\tau''}\langle l\,m\,k\,0|l''\,m\rangle\Phi(k,n;\tau,\nu,l;\tau'',\nu'',l'') \\
& \quad\times|n,\tau'',\nu'',l'',m\rangle \quad (k\in\{2,4\}), \qquad (4.5)
\end{aligned}
$$

where the coefficients Φ depend upon all the parameters indicated and are composed amongst others by a number of Clebsch–Gordan coefficients. The reduced matrix elements of the tensors G^2 and G^4 then immediately are found to be given by

$$\langle n,\tau',v',l'\|G^k\|n,\tau,v,l\rangle$$
$$= (2l'+1)^{1/2}\sum_{v''}\Phi(k,n;\tau,v,l;\tau',v'',l')A_{l'}^{\tau'}(v',v'') \tag{4.6}$$

the overlap-integral $A_{l'}^{\tau'}(v',v'')$ being defined in (2.6) and (2.7).

Clearly, it is of no use to give here the closed form expressions of the Φ coefficients since in practice their numerical values corresponding to given parameter values can be very easily established by means of a computer program.

Finally, it should be mentioned that other techniques exist to calculate reduced matrix elements of the type (4.6). Indeed, as it has been mentioned in I, this problem has been treated already by Chacon and Moshinsky.[10,11] It is, however, not possible to compare algebraically their final results with ours. The present derivation has the advantage that it is self-contained, easy to turn into a program, and without reference to additional tables.

[1]A. Arima and F. Iachello, Phys. Rev. Lett. **35**, 1069 (1975); Ann. Phys. (NY) **99**, 253 (1976); **111**, 201 (1978).

[2]A. Arima and F. Iachello, Phys. Rev. Lett. **40**, 385 (1978); Ann. Phys. (NY) **123**, 468 (1979).

[3]H. De Meyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **26**, 3109 (1985).

[4]J. Vanthournout, J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **28**, 2529 (1987).

[5]J. Vanthournout, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **29**, 1958 (1988).

[6]N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1124 (1968).

[7]S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).

[8]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128 (1958).

[9]D. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).

[10]E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976).

[11]E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).

Variational principle and conservation laws for nonbarotropic flows

George Mathew and M. J. Vedan
Department of Mathematics and Statistics, Cochin University of Science and Technology, Cochin 682022, India

(Received 11 August 1987; accepted for publication 19 October 1988)

A variational formulation for nonbarotropic perfect fluid flow is presented. By introducing suitable field variables, Euler's equations of motion from the variational principles are obtained. The conservation laws of energy, impulse, and angular momenta follow from application of Noether's theorem.

I. INTRODUCTION

Variational descriptions of the dynamics of perfect incompressible fluids seem to have been first introduced by Bateman,[1] followed by Lichtenstein,[2] and Lamb.[3] Since then, several versions of the variational principle for perfect fluid flows based on Hamilton's principle have appeared in the literature—Herivel,[4] Serrin,[5] Eckart,[6] Lin,[7] and Seliger and Whitham,[8] to mention a few. However, few applications have been attempted to specific problems due to either the inherent limitations of a Lagrangian approach or the indeterminacies and redundancies in the definition of potentials in an Eulerian approach. It has been pointed out by Mobbs[9] that the most general version is the one attributed to Serrin.[5]

The present paper is an attempt to generalize the variational principle for barotropic flow of Drobot and Rybarski[10] to the case of nonbarotropic flow.

II. HYDROMECHANICAL VARIATIONS

We consider the Euclidean four-dimensional space X. A point x in X has coordinates x^α, $\alpha = 0,1,2,3$, where x^0 is time t and x^α, $\alpha = 1,2,3$, are spacelike coordinates. The terms $\bar{p}(x)$ and $\bar{s}(x)$ are four-dimensional vector fields with components p^α and s^α, $\alpha = 0,1,2,3$. Here, p^0 is the density and p^α, $\alpha = 1,2,3$, are the impulses $p^0 v^\alpha$ $(\alpha = 1,2,3)$, v^α being the components of the velocity $\bar{v}$. Again $\bar{s}(x)$ has the components s^α, $\alpha = 0,1,2,3$, where $(s^0, s^1, s^2, s^3) = (p^0 S, p^0 v^1 S, p^0 v^2 S, p^0 v^3 S)$, where S is specific entropy.

Let H denote the three-dimensional hypersurface in X and dH_α denote the oriented element on H,

$$dH_\alpha = \varepsilon_{\alpha\beta\gamma\delta}\, dl^\beta\, dl^\gamma\, dl^\delta, \tag{1}$$

where $\varepsilon_{\alpha\beta\gamma\delta}$ is "Levi-Civita tensor" and dl^α, dl^β, dl^γ are three linearly independent vectors lying on H so that dH_α is normal to the hypersurface H. The mass contained on H will be represented by the integral $\int_H dH_\alpha\, p^\alpha$, called the complete matter flow. In particular, when the hypersurface H is the spacelike three-dimensional volume V, we have $dH_0 = dV$, $dH_1 = dH_2 = dH_3 = 0$ and

$$\int_H dH_\alpha\, p^\alpha = \int_V p^0\, dV,$$

reduces to the usual mass. If H is closed and consists of V_{t_0}, V_{t_1} and the moving two-dimensional boundary S_t of V_t for $t_0 < t < t_1$, then

$$\oint_H dH_\alpha\, p^\alpha = \int_{V_{t_1}} p^0\, dV_t - \int_{V_{t_0}} p^0\, dV_t + \int_{t_0}^{t_1} dt \int_{S_t} p^0 \cdot v \cdot dS_t. \tag{2}$$

This is called the matter balance for the moving region. In general, the hypersurface may be open. In this case, the complete matter flow represents a generalization of the notion of the mass contained in H. If V is any four-dimensional region contained in X and ∂V is its boundary, by Gauss' theorem applied to $\oint_V dH_\alpha\, p^\alpha$, we obtain

$$\partial_\alpha p^\alpha = \frac{\partial p^0}{\partial t} + \operatorname{div}(p^0 \bar{v}), \tag{3}$$

representing the density of the source of matter existing in V. We have similar results for $\bar{s}(x)$ as well. The action W will be assumed to have the form,

$$W = \int_V dV\, L(x, p(x), s(x)), \tag{4}$$

in which the Lagrangian L is any given function depending on x, $\bar{p}(x)$, and $\bar{s}(x)$ only.

The usual Lagrangian is given by

$$L = (1/2p^0)((p^1)^2 + (p^2)^2 + (p^3)^2) - E(p^0, s^0) - V(x). \tag{5}$$

Let F be a function space of vector valued functions $p^\alpha(x)$ and $s^\alpha(x)$, $\alpha = 0,1,2,3$, supposed to be regular in X. We consider the following infinitesimal transformations of X and F into themselves:

$$\begin{aligned} \bar{x}^\alpha &= x^\alpha + \delta x^\alpha(x), \\ \bar{p}^\alpha &= p^\alpha + \delta p^\alpha(x), \\ \bar{s}^\alpha &= s^\alpha + \delta s^\alpha(x), \end{aligned} \tag{6}$$

where

$$\begin{aligned} \delta x^\alpha(x) &= e\xi^\alpha(x) + o(e), \\ \delta p^\alpha(x) &= e\pi^\alpha(x) + o(e), \\ \delta s^\alpha(x) &= e\theta^\alpha(x) + o(e). \end{aligned} \tag{7}$$

In (7), ξ, π, θ are arbitrary functions belonging to the space F and e is a scalar parameter. The functions $\delta p^\alpha(x)$ and $\delta s^\alpha(x)$ are called the local variations of the fields $\bar{p}$ and $\bar{s}$, respectively.

We define

 0022-2488/89/040949-04$02.50

$$\Delta \int_H dH_\alpha\, p^\alpha = e\left[\frac{d}{de}\int_{\tilde H} dH_\alpha\, p^\alpha(x)\right]_{e=0}, \tag{8}$$

where $\tilde H$ denotes the hypersurface obtained from H by transformations (6), the total variation of the flow. Now, we have the identity

$$\Delta \int_H dH_\alpha\, p^\alpha = \int_H dH_\alpha\, [\delta p^\alpha - \partial_\beta(p^\beta \delta x^\alpha - p^\alpha \delta x^\beta) + \partial_\beta p^\beta \cdot \delta x^\alpha]. \tag{9}$$

Complete entropy flow is defined in the same way, by replacing p^α by s^α in (8), and we get an expression similar to (9). The functional

$$\delta W = e\left[\frac{d}{de}\int_{\tilde V} dV\, L(\tilde x,\tilde p(x),\tilde s(x))\right]_{e=0}, \tag{10}$$

is called the local variation of the action W given by (4), and

$$\Delta W = e\left[\frac{d}{de}\int_{\tilde V} dV\, L(\tilde x,\tilde p(\tilde x),\tilde s(\tilde x))\right]_{e=0}, \tag{11}$$

where $\tilde V$ is obtained from V by the transformations (6) and is called the total variation of the action. Then we have,

$$\Delta W = \int_V dV\left[\frac{\partial L}{\partial p^\alpha}\delta p^\alpha + \frac{\partial L}{\partial s^\alpha}\delta s^\alpha + \partial_\alpha(L\delta x^\alpha)\right]. \tag{12}$$

Now we impose certain conditions on the variations of $\bar p$ and $\bar s$. (i) For every hypersurface H for which $dH_\alpha\, \delta x^\alpha = 0$,

$$\Delta \int_H dH_\alpha\, p^\alpha = 0, \quad \Delta \int_H dH_\alpha\, s^\alpha = 0.$$

(ii) The variations δp^α, δs^α shall satisfy the equations

$$\partial_\alpha \delta p^\alpha = 0, \quad \partial_\alpha \delta s^\alpha = 0.$$

The variations $\delta_0 p^\alpha$, $\delta_0 s^\alpha$ satisfying these conditions will be called local hydromechanical variations of the fields $\bar p(x)$ and $\bar s(x)$, respectively. The total hydromechanical variation of $\bar p$ and $\bar s$ are defined by

$$\Delta_0 p^\alpha = \delta_0 p^\alpha + \partial_\beta p^\alpha \cdot \delta x^\beta$$

and

$$\Delta_0 s^\alpha = \delta_0 s^\alpha + \partial_\beta s^\alpha \cdot \delta x^\beta, \tag{13}$$

respectively.

It can be shown that all local hydromechanical variations are of the form

$$\delta_0 p^\alpha = \partial_\beta(p^\beta \delta x^\alpha - p^\alpha \delta x^\beta)$$

and

$$\delta_0 s^\alpha = \partial_\beta(s^\beta \delta x^\alpha - s^\alpha \delta x^\alpha), \tag{14}$$

where δx^α are arbitrary infinitesimal functions belonging to the space F. The proof is similar to that given by Drobot and Rybarski.[10]

Thus Eqs. (14) define an infinitesimal group of transformations of the vector fields $\bar p(x)$ and $\bar s(x)$, depending on arbitrary functions δx^α.

We can see that the conditions defining the hydromechanical variations are appropriate counterparts of the necessary conditions for the motion.

III. GENERALIZED VARIATIONAL PRINCIPLE

We now state the variational principle from which the equations of motion follow.

For all δx^α vanishing on the boundary of the region V,

$$\Delta W = \Delta \int_V dV\, L\,[x,p(x),s(x)] = \int_V dv\left[\left(\frac{\partial L}{\partial p^\alpha}\right)\delta p^\alpha + \left(\frac{\partial L}{\partial s^\alpha}\right)\delta s^\alpha + \partial_\alpha(L\delta x^\alpha)\right] = 0, \tag{15}$$

provided that δp^α, δs^α are hydromechanical variations. Then

$$\Delta W = \int_V dV\left[\frac{\partial L}{\partial p^\alpha}\partial_\beta(p^\beta\delta x^\alpha - p^\alpha \delta x^\beta) + \frac{\partial L}{\partial s^\alpha}\partial_\beta(s^\beta \delta x^\alpha - s^\alpha \delta x^\beta) + \partial_\alpha(L\delta x^\alpha)\right] = \int_V dV\, \partial_\beta(T^\beta_\alpha \delta x^\alpha) - \int_V dV\, \psi_\alpha \delta x^\alpha, \tag{16}$$

where

$$T^\beta_\alpha = p^\beta \frac{\partial L}{\partial p^\alpha} + s^\beta \frac{\partial L}{\partial s^\alpha} + \delta^\beta_\alpha\left(L - p^\gamma \frac{\partial L}{\partial p^\gamma} - s^\gamma \frac{\partial L}{\partial s^\gamma}\right) \tag{17}$$

and

$$\psi_\alpha = p^\beta\left[\partial_\beta\left(\frac{\partial L}{\partial p^\alpha}\right) - \partial_\alpha\left(\frac{\partial L}{\partial p^\beta}\right)\right] + s^\beta\left[\partial_\beta\left(\frac{\partial L}{\partial s^\alpha}\right) - \partial_\alpha\left(\frac{\partial L}{\partial s^\beta}\right)\right]. \tag{18}$$

The expressions ψ_α are called hydromechanical Euler expressions.

Since δx^α vanish on the boundary of V, the first integral on the right-hand side of (16) vanishes. Thus we obtain from (15)

$$\int_V dV\, \psi_\alpha \delta x^\alpha = 0. \tag{19}$$

Since δx^α are arbitrary in the interior of V, we get the equations of motion,

$$\psi_\alpha = 0. \tag{20}$$

Since $s^\alpha\, \psi_\alpha = 0$, the four equations of motion are linearly dependent and we get only three linearly independent solutions. In the case the Lagrangian takes the usual form we get the following equations:

$\alpha = 0$:

$$\frac{\partial}{\partial t}\left(\frac{1}{2}|\bar v|^2\right) + \bar v\cdot\nabla\left(\frac{1}{2}|\bar v|^2\right) + \bar v\cdot\left(\frac{1}{p^0}\nabla P + \nabla U\right) = 0, \tag{21}$$

$\alpha = 1,2,3$:

$$\frac{\partial \bar v}{\partial t} + \bar v\cdot\nabla\bar v = -\frac{1}{p^0}\nabla P - \nabla U, \tag{22}$$

where

$$P = p^0\frac{\partial E}{\partial p^0} + s^0\frac{\partial E}{\partial s^0} - E, \tag{23}$$

is the pressure.

All these equations of motion are deduced under the assumption that the functions p^α, s^α are continuous in the whole region. However, if there exists some hypersurfaces on which p^α or s^α ceases to be continuous, jump conditions for them can be deduced from the variational principle.

IV. NOETHER'S THEOREMS AND CONSERVATION LAWS

We consider a class of transformations, depending on scalar parameters. Let

$$W = \int_V dV L(x,p(x),s(x)), \tag{24}$$

be a functional defined on the space F, and V be any given region contained in X. We consider the infinitesimal transformations defined by

$$\tilde{x}^\alpha = x^\alpha + \Delta x^\alpha, \tag{25}$$

where Δx^α are functions of x^β, $p^\beta(x)$, $s^\beta(x)$ and their derivatives.

Then the functional W is transformed to

$$\begin{aligned}\widetilde{W} &= W + \Delta_0 W \\ &= \int_{\widetilde{V}} dV L(\tilde{x},\tilde{p}(\tilde{x}),\tilde{s}(\tilde{x})),\end{aligned} \tag{26}$$

where $\widetilde{V}$ is the transformed region of V.

Now we have

$$\Delta_0 W = \int_V dV \left[\frac{\partial L}{\partial p^\alpha}\delta_0 p^\alpha + \frac{\partial L}{\partial s^\alpha}\delta_0 s^\alpha + \partial_\alpha (L\delta x^\alpha)\right], \tag{27}$$

where

$$\begin{aligned}\delta_0 p^\alpha &= \partial_\beta (p^\beta \Delta x^\alpha - p^\alpha \Delta x^\beta), \\ \delta_0 s^\alpha &= \partial_\beta (s^\beta \Delta x^\alpha - s^\alpha \Delta x^\beta).\end{aligned} \tag{28}$$

The functional W is said to be hydromechanically invariant up to a divergence or div-invariant with respect to the transformations (25) if there exists a vector C^α such that

$$\Delta_0 W = \partial_\alpha C^\alpha \quad \text{identically in} \quad V.$$

If, in particular, $C^\alpha \equiv 0$, so that $\Delta_0 W = 0$, then W is said to be absolutely invariant with respect to the transformations (25).

Noether's theorems describe a relationship between the invariance of an action integral with respect to given infinitesimal transformations and some identities satisfied by its Euler expressions. It has been shown by Drobot and Rybarski[10] that the conservation laws for barotropic fluid flow follow from the application of Noether's theorems if the variations are restricted to hydromechanical variations.

Noether's first theorem can be adapted to our variational principle as follows.

Theorem: If the function W is div-invariant with respect to the transformations (25) depending essentially on k arbitrary parameters, then exactly k linearly independent linear forms of the Euler expressions ψ_α are divergences, the variations of field variables being restricted to hydromechanical variations.

The proof is similar to that of Theorem 2 by Drobot and Rybarski.[10]

Thus if the transformations are of the form,

$$\Delta x^\alpha = g^\alpha_m \varepsilon^m, \quad m = 1,2,3,...k, \tag{29}$$

and

$$C^\alpha = C^\alpha_m \varepsilon^m,$$

where $\varepsilon^1,...,\varepsilon^m$ are infinitesimal scalar parameters and g^α_m, C^α_m are given functions. Then

$$\psi_\alpha g^\alpha_m = \partial_\beta (T^\beta_\alpha g^\alpha_m - C^\beta_m). \tag{30}$$

Since during the motion, $\psi_\alpha = 0$, it follows that

$$\partial_\beta (T^\beta_\alpha g^\alpha_m - C^\beta_m) = 0. \tag{31}$$

We now apply the formula (30) to the case when (25) is the group of Galilean transformations.

For this we suppose that the action W is absolutely invariant with respect to the Galilean transformation,

$$\tilde{x}^\alpha = a^\alpha + a^\alpha_\beta x^\beta, \quad \alpha,\beta = 0,1,2,3, \tag{32}$$

in which the infinitesimal parameters a^α and a^α_β satisfy the conditions,

$$\begin{aligned} a^0_\beta &= a^\alpha_0 = 0, \quad \text{for } \alpha,\beta = 0,1,2,3, \\ a^\alpha_\beta &= a^\beta_\alpha = 0, \quad \text{for } \alpha,\beta = 1,2,3.\end{aligned} \tag{33}$$

In this case $C^\alpha = 0$ and by (31) we have the conservation laws,

$$\partial_\alpha T^\alpha_\beta = 0, \quad \alpha,\beta = 0,1,2,3, \tag{34}$$

and

$$\partial_\alpha M^\alpha_{\beta\gamma} = 0, \quad \alpha = 0,1,2,3, \quad \beta,\gamma = 1,2,3, \tag{35}$$

where

$$M^\alpha_{\beta\gamma} = T^\alpha_\beta x^\gamma - T^\alpha_\gamma x^\beta. \tag{36}$$

Substituting for T^α_β the expression given by (17) we get the conservation laws of energy, impulse, and angular momenta, respectively, for inviscid fluid flows. If L takes the form given by (5), these laws become the usual ones.

V. CONCLUSION

We have extended the variational principle of Drobot and Rybarski to nonbarotropic adiabatic inviscid fluid flows and obtained Euler's equations of motion as Euler–Lagrange equations of the variation. As a consequence of Noether's first theorem we obtained usual conservation laws of energy, impulse, and angular momenta from the invariance of action by the Galilean group of transformations of the independent space variables. This method avoids some of the difficulties encountered in other variational formulations.[9,11,12]

ACKNOWLEDGMENT

The authors wish to thank the referee, whose careful reading of an earlier manuscript helped to improve this paper significantly.

[1]H. Bateman, Proc. R. Soc. London, Ser. A **125**, 598 (1929).
[2]L. Lichtenstein, Grundlagen der Hydromechanik (Berlin, 1929).
[3]H. Lamb, *Hydrodynamics* (Cambridge U.P., Cambridge, 1932), Third Ed.
[4]J. W. Herivel, Proc. Cambridge Philos. Soc. **51**, 344 (1959).

[5]J. Serrin, *Handbuch der Physik* (Springer, Berlin, 1959), Vol. 8.
[6]C. Eckart, Phys. Fluids **3**, 421 (1960).
[7]C. C. Lin, Proceedings of the International School of Physics "Enrico Fermi" (New York), Course XXI (1963).
[8]R. L. Seliger and G. B. Whitham, Proc. R. Soc. London, Ser. A **305**, 1 (1968).
[9]S. D. Mobbs, Proc. R. Soc. London, Ser. A **381**, 457 (1982).
[10]S. Drobot and A. Rybarski, Arch. Rational Mech. Anal. **2**, 393 (1959).
[11]F. P. Bretherton, J. Fluid Mech. **44**, 19 (1970).
[12]B. F. Schutz and F. Sorkin, Ann. Phys. **107**, 1 (1976).

On the log-normal diffusion process

Siegfried H. Lehnigk
Research, Development, and Engineering Center, U.S. Army Missile Command, Redstone Arsenal, Alabama 35898-5248

(Received 26 July 1988; accepted for publication 14 December 1988)

The log-normal conditional density function is the delta function initial condition solution of a four-parameter Fokker–Planck equation. It defines a diffusion process over the open first quadrant of the (x,t) plane. This process reaches a nonzero steady state as t increases indefinitely if the drift parameter is positive. The process may be monotonic or by expansion and contraction (breathing). If the drift parameter is negative the process goes to zero by expansion and contraction towards $x = 0$ as t increases indefinitely.

I. INTRODUCTION

Investigations of the size distribution of metal particles formed by nucleation[1] have led to the conclusion that it is of log-normal type. This result may possibly be best understood, from an evolution point of view, by interpreting the associated density as the terminal steady state of a process governed by the log-normal Fokker–Planck equation (1.1), which is initiated at time zero from a completely concentrated initial state at initial particle size $y > 0$. It is the objective of this paper to show that the underlying evolution process may be monotonic or by expansion and contraction (breathing).

The parabolic equation

$$[(a_2(x)z)_x - a_1(x)z]_x - z_t = 0,$$
$$z = z(x,t), \quad x > 0, \quad t > 0, \tag{1.1a}$$

with diffusion and drift coefficients

$$a_2(x) = \alpha x^2, \; a_1(x) = (4\alpha - \beta_2)x - \beta_1 x \log \rho x, \tag{1.1b}$$

$\alpha > 0, \rho > 0, \beta_1 \neq 0, \beta_2 \in \Re$, has been established[2] as one of the one-dimensional, autonomous, Fokker–Planck type of equations that admit linear similarity solutions. It shall be designated the log-normal equation. Another one, the generalized Feller equation, has been discussed elsewhere.[3]

The change in variables

$$\log \rho x = -\lambda\xi + \log \rho b, \quad z(x,t) = (\rho b)^{-1} e^{\lambda\xi} v(\xi,t), \tag{1.2}$$

with

$$\lambda = i\sqrt{2}(\alpha^{-1}\beta_1)^{-1/2}, \quad \rho b = \exp \beta_1^{-1}(3\alpha - \beta_2), \tag{1.3}$$

transforms (1.1) into the equation of Gaussian type

$$v_{\xi\xi} - 2\xi v_\xi - 2v + 2\beta_1^{-1} v_t = 0.$$

From its well-known fundamental solution (transition density)[4] $v(\xi,t;\xi_0,s)$, which is of Gaussian type with mean $\xi_0 \exp - \beta_1(t-s)$ and variance $-1 + \exp - 2\beta_1(t-s)$, we obtain, by means of (1.2), the fundamental solution of (1.1) in the form

$$z(x,t;y,s) = (1/\sqrt{\pi})x^{-1}c^{-1/2}(t-s)\exp - [\log b^{-1}x - e^{-\beta_1(t-s)} \log b^{-1}y]^2 c^{-1}(t-s), \tag{1.4a}$$

with b defined by (1.3) and

$$c(t) = 2\alpha\beta_1^{-1}(1 - e^{-2\beta_1 t}). \tag{1.4b}$$

As a function of x and t, $z(x,t;y,s)$ is a solution of (1.1) and, as a function of y and s, it is a solution of the adjoint of (1.1).

The conditional density function $f(x,t;y) = z(x,t;y,0)$ of (1.1) plays an essential role in diffusion applications and, as an integral transform kernel, in the construction of initial and boundary condition solutions of (1.1). By construction, $f(x,t;y)$ is the delta function initial condition solution of the log-normal equation with the delta function applied at $x = y > 0$ and $t = 0$. Furthermore, for fixed $t > 0$ and $y > 0$, the function $f(x,t;y)$ becomes the log-normal probability density well known in statistics and statistical physics. To reduce it to its standard form[5] we set, for fixed $t = t_0 > 0$, $c(t_0) = 2\sigma^2$ and $\log b + (\exp - \beta_1 t_0)\log b^{-1}y = \mu$ and obtain

$$f(x) = (2\pi\sigma^2)^{-1/2}x^{-1} \exp - (\log x - \mu)^2/2\sigma^2.$$

This establishes the diffusion character of the log-normal probability distribution.

Functional properties of $f(x,t;y)$ can be derived along the lines of theorems established for the conditional density of the generalized Feller equation.[6] They will not be discussed here. They are relevant only to the differential equation theoretical approach to the structure of the solutions of (1.1) and have no bearing on the present subject matter, which concerns the evolutionary behavior of the log-normal diffusion process.

II. THE LOG-NORMAL DIFFUSION PROCESS

The log-normal conditional density function

$$f(x,t;y) = (1/\sqrt{\pi})x^{-1}c^{-1/2}(t)\exp - [\log b^{-1}x - e^{-\beta_1 t} \log b^{-1}y]^2 c^{-1}(t), \tag{2.1}$$

with b defined by (1.3), $c(t)$ by (1.4b), is unimodal and takes its maximum at

$$x_m(t) = b \exp(e^{-\beta_1 t} \log b^{-1}y - c(t)/2). \tag{2.2}$$

We also note the following facts:

$f(x,t;y) \downarrow 0$, as $t \downarrow 0$, $x \neq y$,

$f(x,t;x) \uparrow \infty$, as $t \downarrow 0$,

$$f(x,t;y) \to \begin{cases} (2\pi\alpha\beta_1^{-1})^{-1/2} x^{-1} \exp - [(2\alpha\beta_1^{-1})^{-1} \log^2 b^{-1}x], & \text{as } t \uparrow \infty, \quad \beta_1 > 0, \\ 0, & \text{as } t \uparrow \infty, \quad \beta_1 < 0. \end{cases} \tag{2.3}$$

The first limit relation in (2.3) shows that the conservative log-normal diffusion process, which is initiated from a completely concentrated state at $t = 0$, $x = y > 0$, approaches a nonzero steady state as $t \uparrow \infty$, if the drfit parameter β_1 is positive.

Let us investigate the time behavior of the maximum $\varphi(t) = f(x_m(t), t;y)$ of the log-normal diffusion process. With f given in (2.1) and $x_m(t)$ in (2.2), the function $\varphi(t)$ takes the form

$$\begin{aligned} \varphi(t) &= b^{-1}[2\pi\alpha\beta_1^{-1}(1 - e^{-2\beta_1 t})^{-1/2}] \\ &\quad \times \exp - \tfrac{1}{2}(\alpha\beta_1^{-1} e^{-2\beta_1 t} \\ &\quad + 2e^{-\beta_1 t} \log b^{-1}y - \alpha\beta_1^{-1}), \quad 0 < t < \infty, \end{aligned} \tag{2.4}$$

of (2.1).

First of all, we note that, according to (2.2) and (1.4b),

$$x_m(t) \to y > 0, \quad \text{as } t \downarrow 0, \quad \beta_1 \neq 0,$$

and that, according to (2.4),

$$\varphi(t) \uparrow \infty, \text{ as } t \downarrow 0, \beta_1 \neq 0.$$

Furthermore, according to (2.2) and (1.4b),

$$x_m(t) \to \begin{cases} b e^{-\alpha\beta_1^{-1}} < b, & \text{as } t \uparrow \infty, \quad \beta_1 > 0, \\ 0, & \text{as } t \uparrow \infty, \quad \beta_1 < 0, \end{cases} \tag{2.5}$$

and

$$\varphi(t) \to \begin{cases} b^{-1}(2\pi\alpha\beta_1^{-1})^{-1/2} \exp \tfrac{1}{2}\alpha\beta_1^{-1} > 0, & \text{as } t \uparrow \infty, \quad \beta_1 > 0, \\ +\infty, & \text{as } t \uparrow \infty, \quad \beta_1 < 0. \end{cases} \tag{2.6}$$

It is instructive to compare the limit relations (2.5) and (2.6) for $\beta_1 > 0$ with the corresponding limits in (2.3).

We want to investigate next whether $\varphi(t)$ takes extreme values in compact subintervals of $(0, \infty)$. The equation $d\varphi / dt = 0$ is equivalent to

$$\begin{aligned} &(\log b^{-1}y) e^{3\beta_1 t} \\ &\quad - (1 - \alpha\beta_1^{-1}) e^{2\beta_1 t} - (\log b^{-1}y) e^{\beta_1 t} - \alpha\beta_1^{-1} = 0, \end{aligned}$$

which, if we set $\log b^{-1}y = \gamma$, $\exp \beta_1 t = \sigma$, $1 < \sigma < \infty$, if $\beta_1 > 0$, $0 < \sigma < 1$, if $\beta_1 < 0$, can be written in the form

$$h(\sigma) = \gamma\sigma^3 - (1 - \alpha\beta_1^{-1})\sigma^2 - \gamma\sigma - \alpha\beta_1^{-1} = 0. \tag{2.7}$$

(i) The singular case: $\log b^{-1}y = \gamma = 0$. If $\alpha = \beta_1 > 0$ then, as (2.7) shows, $h(\sigma)$ reduces to the negative constant -1. In other words, in this situation $\varphi(t)$ is a strictly monotonically decreasing function of $t > 0$. An example parameter set for this case is $\alpha = \beta_1 = \frac{1}{2}$, $y = b = 1$.

If $\alpha \neq \beta_1$, we have from (2.7)

$$\sigma^2 = \alpha\beta_1^{-1}(\alpha\beta_1^{-1} - 1)^{-1}.$$

Therefore, $h(\sigma)$ has a positive zero if $\alpha\beta_1^{-1}(\alpha\beta_1^{-1} - 1)^{-1} > 0$.

If $\beta_1 > 0$ this requires $0 < \beta_1 < \alpha$. In this situation, $\varphi(t)$ takes a minimum at

$$t_1 = \beta_1^{-1} \log[\alpha\beta_1^{-1}(\alpha\beta_1^{-1} - 1)^{-1}]^{1/2} > 0.$$

An example set of parameters is $\alpha = 1, \beta_1 = \frac{1}{2}, y = b = 1$. If $0 < \alpha < \beta_1$, $\varphi(t)$ is strictly monotonically decreasing ($\alpha = \frac{1}{4}$, $\beta_1 = \frac{1}{2}, y = b = 1$).

If $\beta_1 < 0$, we have $0 < \alpha\beta_1^{-1}(\alpha\beta_1^{-1} - 1)^{-1} < 1$, and $\varphi(t)$ takes a minimum at

$$t_1 = \beta_1^{-1} \log[\alpha\beta_1^{-1}(\alpha\beta_1^{-1} - 1)^{-1}]^{1/2}$$

$(\alpha = \frac{1}{2}, \beta_1 = -\frac{1}{2}, y = b = 1)$.

(ii) The regular case: $\log b^{-1}y = \gamma \neq 0$.

(a) $\gamma > 0$, i.e., $0 < b < y$. We are now dealing with the polynomial

$$h(\sigma) = \gamma\sigma^3 - (1 - \alpha\beta_1^{-1})\sigma^2 - \gamma\sigma - \alpha\beta_1^{-1}.$$

If $\beta_1 > 0$, the coefficient sequence of $h(\sigma)$ contains one variation in sign. Descartes' rule[7] implies that $h(\sigma)$ has exactly one positive zero σ_1. Since $h(0) = -\alpha\beta_1^{-1} < 0$, $h(1) = -1$, this zero is greater than unity. The function $\varphi(t)$ takes a minimum at $t_1 = \beta_1^{-1} \log \sigma_1 > 0$ [$\alpha = \frac{1}{4}, \beta_1 = \frac{1}{2}$. $b = 1, y = e$ ($\gamma = 1$)].

If $\beta_1 < 0$, the coefficient sequence of $h(\sigma)$ has two variations in sign. Since $h(0) = -\alpha\beta_1^{-1} > 0$, $h(1) = -1$, $h(\sigma)$ has exactly one zero σ_1 in $(0,1)$. Thus $\varphi(t)$ takes a minimum at $t_1 = \beta_1^{-1} \log \sigma_1$ [$\alpha = \frac{1}{2}, \beta_1 = -\frac{1}{2}, b = 1, y = e$ ($\gamma = 1$)].

(b) $\gamma < 0$, i.e., $0 < y < b$. We set $\gamma = -\kappa, \kappa > 0$, and investigate the polynomial

$$g(\sigma) = -h(\sigma) = \kappa\sigma^3 + (1 - \alpha\beta_1^{-1})\sigma^2 - \kappa\sigma + \alpha\beta_1^{-1}. \tag{2.8}$$

If $\beta_1 > 0$, the coefficient sequence of $g(\sigma)$ contains two variations in sign. To ascertain whether there are two or no positive zeros we apply Sturm's chain.[7] Disregarding positive constant factors and setting $\alpha\beta_1^{-1} = \rho > 0$, we obtain Sturm's chain of the polynomial (2.8),

$$\begin{aligned} g(\sigma) &= r_1(\sigma) = \kappa\sigma^3 + (1 - \rho)\sigma^2 - \kappa\sigma + \rho, \qquad (2.9) \\ g'(\sigma) &= r_2(\sigma) = 3\kappa\sigma^2 + 2(1 - \rho)\sigma - \kappa, \\ r_3(\sigma) &= 2[3\kappa^2 + (1 - \rho)^2]\sigma - (1 + 8\rho)\kappa, \\ r_4 &= 4[3\kappa^2 + (1 - \rho)^2]^2 \\ &\quad - 4(1 - \rho)(1 + 8\rho)[3\kappa^2 + (1 - \rho)^2] \\ &\quad - 3\kappa^2(1 + 8\rho)^2. \end{aligned}$$

The sign sequences at 0, 1, and $+\infty$ are shown in Table I. Inspection of Table I shows that there are three possibilities. If $r_4 < 0$, $r_1(\sigma)$ has no positive zeros, which means that $\varphi(t)$ is strictly montonically decreasing [$\alpha = \frac{1}{2}, \beta_1 = \frac{1}{2}$ ($\rho = 1$), $b = 1$, $y = e^{-1}$ ($\kappa = 1$)]. If $r_4 > 0$, $r_1(\sigma)$ has two positive zeros $\sigma_1 < \sigma_2$. If $r_2(1) > 0$, $r_3(1) > 0$, we have $0 < \sigma_1 < \sigma_2 < 1$

TABLE I. The sign sequences at 0, 1, and $+\infty$.

σ	r_1	r_2	r_3	r_4
0	+	−	−	r_4
1	+	$r_2(1)$	$r_3(1)$	r_4
$+\infty$	+	+	+	r_4

so that, again, $\varphi(t)$ is strictly montonically decreasing $[\alpha=1,\ \beta_1=2\ \ (\rho=\frac{1}{2}),\ b=1,\ y=e^{-1}\ \ (\kappa=1)]$. If $r_2(1)<0$ or $r_3(1)<0$ or both, then $1<\sigma_1<\sigma_2$. This implies that $\varphi(t)$ takes a minimum at $t_1=\beta_1^{-1}\log\sigma_1$ and a maximum at $t_2=\beta_1^{-1}\log\sigma_2$ $[\alpha=10, \beta_1=1\ (\rho=10), b=1, y=e^{-1}\ (\kappa=1)]$. If $r_4=0$, $r_1(\sigma)$ has a positive zero s of multiplicity 2, i.e.,

$$\begin{aligned} r_1(\sigma) &= \kappa(\sigma-s)^2(\sigma+\sigma_3) \\ &= \kappa[\sigma^3+(\sigma_3-2s)\sigma^2+s(s-2\sigma_3)\sigma+s^2\sigma_3], \\ \sigma_3 &> 0. \end{aligned} \tag{2.10}$$

If $0<s<1$, $\varphi(t)$ is strictly monotonically decreasing. However, if $s>1$, $\varphi(t)$ has an inflection point at $t=\beta_1^{-1}\log s$. For example, for $\kappa=1$, comparison of (2.9) and (2.10) shows that s must be that positive zero (of the two positive ones that exist) of the polynomial s^4-2s^2-2s+1, which is greater than unity. The numerical value is approximately $s=1.683\,772$. The corresponding value for $\rho=\alpha\beta_1^{-1}$ is approximately 3.228 706.

Finally, if $\beta_1<0$, the coefficient sequence of $g(\sigma)$ given in (2.8) contains one variation in sign which means that $g(\sigma)$ has exactly one positive zero. Since $g(0)=\alpha\beta_1^{-1}<0$, $g(1)=1$, this zero σ_1 is located in (0,1). Consequently, $\varphi(t)$ takes a minimum at $t_1=\beta_1^{-1}\log\sigma_1[\alpha=\frac{1}{2},\beta_1=-\frac{1}{2}, b=1, y=e^{-1}\ (\kappa=1)]$.

In summary, we can say this. If the drift parameter β_1 is positive, the diffusion process $f(x,t;y)$, initiating at $t=0$ from the point $x=y>0$, ultimately approaches a nonzero steady state as $t\uparrow\infty$. The process may expand in the direction of x, its maximum decreasing strictly monotonically toward the limit value given in the first limit relation in (2.6). It may also first expand until its maximum reaches a minimum value and then contract toward the steady state, its maximum increasing toward its final limit value. The process may also first expand until its maximum reaches a minimum, then contract until the maximum reaches a maximum and then expand again toward the steady state, its maximum decreasing toward the final limit value. If the drift parameter β_1 is negative, the diffusion process always expands at first until its maximum reaches a minimum. Thereafter it contracts, its maximum approaching $+\infty$ at $x=0$ as $t\uparrow\infty$.

[1]J. A. A. J. Perenboom and P. Wyder, "Electronic properties of small metallic particles," Phys. Rep. **78**, 173 (1981), see especially pp. 269–272.

[2]S. H. Lehnigk, "Conservative similarity solutions of the one-dimensional autonomous parabolic equation," J. Appl. Math. Phys. **27**, 385 (1976).

[3]S. H. Lehnigk, "A class of conservative diffusion processes with delta function initial conditions," J. Math Phys. **17**, 973 (1976).

[4]T. T. Soong, *Random Differential Equations in Science and Engineering* (Academic, New York, 1973).

[5]A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis* (McGraw-Hill, New York, 1982), see Sec. 5.2.2.

[6]S. H. Lehnigk, "Initial condition solutions of the generalized Feller equation," J. Appl. Math. Phys. **29**, 273 (1978).

[7]S. H. Lehnigk, *Stability Theorems for Linear Motions* (Prentice-Hall, Englewood Cliffs, NJ, 1966), see Secs. IV 5,6.