The relativistic coma aberration. I. Geometrical optics

Natig M. Atakishiyev,[a] Wolfgang Lassner,[b] and Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas/Cuernavaca, Universidad Nacional Autónoma de México, Apdo. Postal 20-726, 01000 México D.F., Mexico

(Received 18 November 1988; accepted for publication 24 May 1989)

It is shown, in the framework of Hamilton–Lie geometrical optics, that the image on a moving screen undergoes comatic aberration as the conjugate sphere of ray directions distorts under Lorentz boosts.

I. INTRODUCTION

Stellar aberration is a phenomenon known for centuries in positional astronomy: As a result of the Earth's motion in orbit, the directions to stars on the celestial sphere suffer distortion toward the direction of motion. In relativity we know that, corresponding to a ray with angle θ measured from the motion vector, and a velocity $v = c \tanh \alpha$, the distortion is given by the transformation of the *circle*[1]

$$\tan \tfrac{1}{2}\theta \to \tan \tfrac{1}{2}\theta' = e^{-\alpha} \tan \tfrac{1}{2}\theta. \tag{1.1}$$

Hamilton–Lie geometrical optics[2,3] works with phase-space observables on plane screens. It is usually natural to distinguish an *optical axis* when working with optical image-forming systems or optical fibers. In that case it is convenient to perform the aberration expansions of classical geometrical optics.[4] In this paper we treat the aberration phenomenon *globally*, i.e., through exact (closed) expressions valid on the whole optical phase-space manifold: optical momentum is directly related to points on the direction sphere, and this is a *compact* manifold (*unlike* the phase space of point particles).

Distortion of the sphere of directions entails a corresponding *comatic* aberration of ray position at the screen, if the relativistic transformation is to be canonical on optical phase space.

In Sec. II we assemble the basic facts of the Hamilton–Lie account of local and global properties of the phase space of geometric optics. In Sec. III we use this formalism in the framework of Euclidean and special relativity: screens may be translated to new origins, rotated to new optical axes, or boosted to motion. This last transformation is performed by group *deformation*[1,5,6] of the Euclidean to the Lorentz algebra and group, in Sec. IV. In Sec. V the specific aberration due to screen motion along the optical axis is studied as are some of its basic geometric properties for all velocities. Caustic phenomena are highly visible and could be observable in appropriate experimental situations. In the concluding section (VI) some considerations of a mathematical nature are added.

II. OPTICAL PHASE SPACE

Optical phase space in the three-dimensional world of geometric optics is referred to a two-dimentional screen of positions and a sphere of ray directions. It is a four-dimensional manifold where it is convenient to introduce Cartesian coordinates[7] and write its points as $w = (\mathbf{p};\mathbf{q})$, $\mathbf{p} = (p_x,p_y)$, $\mathbf{q} = (q_x,q_y)$, with $\mathbf{q} \in \mathscr{R}^2$ (the real plane) the *position* vector of the ray's intersection, and $\mathbf{p}$ the *momentum* coordinate. The latter is the projection on the plane of the screen of a three-vector $\vec{n} = (p_x,p_y,h)$ along the ray whose length is n, the refractive index of the medium (constant in this paper, corresponding to a homogeneous optical medium). The two coordinate sets are canonically conjugate, i.e., the Poisson bracket[8] relations hold:

$$\{q_i,p_j\} = \delta_{ij}, \quad \{q_i,q_j\} = 0, \quad \{p_i,p_j\} = 0. \tag{2.1}$$

The origin of phase space is $\mathbf{q} = \mathbf{0}$ (the optical center), and $\mathbf{p} = \mathbf{0}$ (the optical axis).

We note that the *range* of the momentum coordinates is limited by $p^2 \leqslant n^2$, and is the projection of the sphere S^2 of ray directions on the screen plane. It consists of the disk $p^2 < n^2$ counted once for $h > 0$ ("forward" rays), and once for $h < 0$ ("backward" rays); the two disks are at the boundary $p^2 = n^2$ when $h = 0$. We may assume the sign of the z component of $\vec{n}$, i.e., h, is always available to distinguish between the two disks, and we may freely revert to the direction sphere coordinates. (The range of $\mathbf{p}$ in two-dimensional mechanics, in contrast, is the full $\mathscr{R}^2$ plane.)

The z component of the direction vector $\vec{n}$ is

$$h = (n^2 - p^2)^{1/2} \tag{2.2}$$

and serves as (minus) the optical Hamiltonian.[3] (The series expansion $h = n - p^2/2n \cdots$ suggests giving n the analog role of a potential, notwithstanding that it also appears in the denominator, where mass ought to be in mechanics.)

Lie optics uses the symplectic structure (2.1) to build *Lie operators* $\hat{f} = \{f,\cdot\}$ associated to differentiable functions $f(\mathbf{p},\mathbf{q})$. Their action on phase space is

$$\hat{f}\mathbf{p} = \{f,\mathbf{p}\} = \frac{\partial f}{\partial \mathbf{q}}, \quad \hat{f}\mathbf{q} = \{f,\mathbf{q}\} = -\frac{\partial f}{\partial \mathbf{p}}. \tag{2.3}$$

Various properties follow,[9] such as

$$\hat{f}g(\mathbf{p},\mathbf{q}) = \{f,g\}(\mathbf{p},\mathbf{q}) = g(\hat{f}\mathbf{p},\hat{f}\mathbf{q}), \tag{2.4a}$$

$$(\{f,g\}\hat{}) = [\hat{f},\hat{g}], \tag{2.4b}$$

[a] Permanent address: Institute of Physics, Azerbaijan Academy of Sciences, Baku 370143, USSR.

[b] Permanent address: Department of Mathematics, Karl-Marx University, 701 Leipzig, German Democratic Republic.

 0022-2488/89/112457-06$02.50

where $[\cdot,\cdot]$ is the *commutator* of operators. They allow us to work with the enveloping algebra of (2.1), and exponentiate to the corresponding *Lie transformation* generated by f:

$$F_\alpha = \exp(\alpha\hat{f}) = 1 + \alpha\hat{f} + 1/2!\alpha^2\hat{f}^2 + \cdots, \tag{2.5}$$

so that acting on suitably smooth functions $g(\mathbf{p},\mathbf{q})$ (such as $\mathbf{p}$ and $\mathbf{q}$ themselves),

$$F_\alpha g = g + \alpha\{f,g\} + 1/2!\alpha^2\{f,\{f,g\}\} + \cdots. \tag{2.6}$$

Also, Lie transformations are *canonical*,[9] i.e., for arbitrary $g_1(\mathbf{p},\mathbf{q})$ and $g_2(\mathbf{p},\mathbf{q})$,

$$\{F_\alpha g_1, F_\alpha g_2\} = \{g_1,g_2\}. \tag{2.7}$$

As a first (counter-) example, consider f to be a function *quadratic* in the components of $\mathbf{p}$ and $\mathbf{q}$. Then[2,9] F_α will map the components of $\mathbf{p}$ and $\mathbf{q}$ linearly among themselves, thus generating Sp $(4,R)$, the group of linear symplectomorphisms of phase space. In this example, however, the natural range of optical momentum $p^2 \leqslant n^2$ is *not* preserved. [In spite of not globally respecting optics, Sp $(4,R)$ has been extremely useful in treating aberration expansions by order around an optical center and axis.[10]]

The position coordinates (q_x,q_y), in particular, are also *not* good functions to generate Lie transformations for global optics, since they translate the $\mathbf{p}$ plane, as in mechanics, and do not respect the natural range $p^2 \leqslant n^2$ of optical momentum.

In fact, it seems rather difficult to write Lie transformations that *do* not preserve the optical momentum range, except for one very obvious class: point-to-point mappings of the sphere, i.e., *rotations* and *distortions* $S^2 \to S^2$ so that $\mathbf{p} \to \mathbf{p}' = \mathbf{p}'(\mathbf{p}, \operatorname{sgn} h)$. These are distortions in the sense that $\mathbf{p}'$ is *not* a function of $\mathbf{q}$. [In the optical distortion aberration,[4] $\mathbf{q}'(\mathbf{q})$ is independent of $\mathbf{p}$; the latter is the Fourier conjugate variable except for ranges.]

To avoid uncomfortable formulas at the joining of the two momentum disks, let us use explicit spherical coordinates for the three-vector of ray directions:

$$p_x = n \sin\theta \sin\phi, \tag{2.8a}$$

$$p_y = n \sin\theta \cos\phi, \tag{2.8b}$$

$$h = n \cos\theta. \tag{2.8c}$$

We now define the Lorentz group action of special relativity on the phase space of geometrical optics through binding (2.8) to be the three-vector parallel to the three-vector part of a lightlike four-vector (k_x,k_y,k_z,k_0) undergoing such tranformations. Thus setting $\mathbf{k} = (k_x,k_y)$, $k = (k_x^2 + k_y^2)^{1/2}$, $\mathbf{p} = n\mathbf{k}/k_0$, and $h = nk_z/k_0$, we obtain the following relations:

$$p/n = \sin\theta = k/k_0, \tag{2.9a}$$

$$h/n = \cos\theta = k_z/k_0, \tag{2.9b}$$

$$p/(n+h) = \tan\tfrac{1}{2}\theta = k/(k_z + k_0). \tag{2.9c}$$

Hence when the lightlike vector $(\mathbf{k},k_z,k_0)$ undergoes a boost in the z direction, the transformation of (2.9c) yields (1.1).

III. EUCLIDEAN TRANSFORMATIONS

Three functions that generate Lie transformations that map optical phase space onto itself properly are the components of $\vec{n} = (p_x,p_y,h)$. The first two generate translations in $\mathbf{q} \in \mathscr{R}^2$,

$$\exp(\mathbf{a}\cdot\hat{\mathbf{p}})g(\mathbf{p},\mathbf{q}) = g(\mathbf{p},\mathbf{q} - \mathbf{a}); \tag{3.1}$$

while the last one generates translations along the optical axis normal to the screen,

$$\exp(z\hat{h})g(\mathbf{p},\mathbf{q}) = g(\mathbf{p},\mathbf{q} + z\,\mathbf{p}/h). \tag{3.2}$$

The transformation of $\mathbf{q}$ in the last argument reads $q + z\tan\theta$ in the direction of $\mathbf{p}$, as is clear from simple geometry. The three generating functions commute under the Lie bracket: $\{p_i,p_j\} = 0$, $\{p_i,h\} = 0$.

Another set of S^2-preserving Lie transformations is the group of rotations of the screen in three-space.[11] To simplify arguments, consider the *two*-dimensional optics case depicted in Fig. 1, where a ray is seen in two different frames rotated by γ, as θ and as $\theta' = \theta + \gamma$, or, in two-dimensional phase space,

$$p \to p' = p\cos\gamma + h\sin\gamma, \tag{3.3a}$$

$$h \to h' = -p\sin\gamma + h\cos\gamma, \tag{3.3b}$$

$$q \to q' = q/(\cos\gamma - p/h\sin\gamma). \tag{3.3c}$$

The last relation is obtained from the law of sines in the triangle of the figure. From here, the generator of two-dimensional rotations $\exp(\gamma\hat{m})$ may be found through

$$\frac{\partial m}{\partial q} = \hat{m}p = \frac{\partial p'}{\partial\gamma}\bigg|_{\gamma=0}, \quad \frac{\partial m}{\partial p} = -\hat{m}q = \frac{\partial q'}{\partial\gamma}\bigg|_{\gamma=0},$$

and is[11] $m = qh$. In three-dimensional optics, if Fig. 1 is the x-z plane, the generator will be that of rotation r_y around the y axis, and if it is the y-z plane the generator will be $-r_x$. Hence the generators of rotations of the direction sphere are [cf. Ref. 11, Eq. (2.9)]

$$r_x = q_y h, \tag{3.4a}$$

$$r_y = -q_x h, \tag{3.4b}$$

$$r_z = q_x p_y - q_y p_x = \mathbf{q}\times\mathbf{p}, \tag{3.4c}$$

and are easily checked to close into an so (3) algebra under the Lie bracket of geometrical optics:

$$\{r_x,r_y\} = r_z, \quad \{r_y,r_z\} = r_x, \quad \{r_z,r_x\} = r_y. \tag{3.5}$$

The first of Eqs. (3.5) may be used to define r_z in (3.4c); this quantity generates rotations in the plane of the screen around the optical center, together with rotations of the direction sphere around the optical axis. Its square is the Petzval invariant of optics.[10]

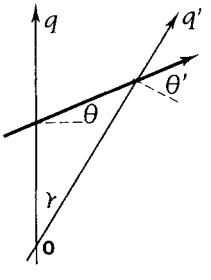


FIG. 1. The transformation of optical phase space due to rotation of the screen about the origin by γ. The ray $(p = n\sin\theta, q)$ transforms to $(p' = n\sin\theta', q')$.

The six functions $(p_x,p_y,h;\ r_x,r_y,r_z)$ close under Poisson brackets into the Lie algebra iso (3) of the Euclidean group of motions: the three translations leaving the direction sphere invariant and the three joint rotations intertwine through

$$\{r_i,p_j\} = p_k, \tag{3.6}$$

with i,j,k cyclic permutations of x,y,z, and $p_z = h$. The Euclidean group ISO (3), containing the Hamiltonian among its generators, is the dynamical group of[11] geometric optics in a homogeneous medium. The two Euclidean invariants are $p^2 + h^2 = n^2$ and $\vec{n}\cdot\vec{p} = 0$.

IV. THE DEFORMATION ISO (3)→SO (3,1)

We recall the classic deformation process[1,5] that builds the Lorentz algebra so (3,1) out of the generators of the Euclidean iso (3), realized on a sphere, and generalizations thereof.[6] Basically, one builds bilinear functions of the generators of iso (3) with the right transformation properties under so (3). These will close into so (3,1) on the sphere. One may also add linearly the generators of the translation subalgebra, thus arriving at all representations of the nonexceptional continuous series. In geometric optics we may propose the three-vector

$$\vec{b} = \vec{r}\times\vec{p} + \sigma\vec{p}. \tag{4.1}$$

As vector functions in the phase-space coordinates, the components are

$$\mathbf{b} = n\mathbf{q} - \mathbf{p}\cdot\mathbf{q}\mathbf{p}/n + \sigma\mathbf{p}, \tag{4.2a}$$

$$b_z = -\mathbf{p}\cdot\mathbf{q}h/n + \sigma h. \tag{4.2b}$$

These three functions transform under the so (3) subalgebra (3.4) as the components of a proper three-vector,

$$\{r_x,b_y\} = b_z \quad \text{(and cyclically)}. \tag{4.3}$$

Finally, they close under the Lie (Poisson) bracket of the algebra, into the Lorentz algebra so (3,1):

$$\{b_x,b_y\} = -r_z \quad \text{(and cyclically)}. \tag{4.4}$$

The constant σ in the boost generators (4.2) is also in the Lorentz invariant $\vec{b}^2 - \vec{r}^2 = n^2\sigma^2$ while $\vec{r}\cdot\vec{b} = 0$.

It is noteworthy that we may express the ray position coordinate $\mathbf{q}$ in terms of the functions generating Lorentz boosts and Euclidean translations:

$$\mathbf{q} = \mathbf{b}/n - b_z\mathbf{p}/nh. \tag{4.5}$$

We shall examine in detail the boosts along the optical axis; these are generated by b_z in (4.2b) as the Lie transformation $\exp(\alpha\hat{b}_z)$ acting on the reference (stationary) screen phase space $(\mathbf{p},\mathbf{q})$, to produce the phase space $(\mathbf{p}'(\mathbf{p},\mathbf{q},\alpha),\mathbf{q}'(\mathbf{p},\mathbf{q},\alpha))$ of a screen in motion with velocity $v = c\tanh\alpha$. On the momentum coordinates, we find the integrated group action to be

$$\mathbf{p}'(\mathbf{p},\alpha) = \exp(\alpha\hat{b}_z)\mathbf{p} = \mathbf{p}/(\cosh\alpha + h/n\sinh\alpha). \tag{4.6}$$

This, we duly note, is a mapping $\mathbf{p}'(\mathbf{p})$ independent of position $\mathbf{q}$ and the "Lorentz constant" σ. The momentum distortion of S^2 is precisely—of course—the stellar aberration (1.1), as may be verified through (2.9).

We may also find the action of this boost on the position coordinate $\mathbf{q}$ with the help of (4.5) and (2.4). The transformed position coordinate is

$$\mathbf{q}'(p,q,\alpha) = \exp(\alpha\hat{b}_z)\mathbf{q} = (\cosh\alpha + (h/n)\sinh\alpha) \times\left(\mathbf{q} + \frac{\sinh\alpha}{n\sinh\alpha + h\cosh\alpha} \times\left[-\frac{\mathbf{p}\cdot\mathbf{q}}{n} + \sigma\right]\mathbf{p}\right). \tag{4.7}$$

The magnification and aberrations present in (4.7) will be studied in Sec. V. We only point out here that the meaning of the arbitrary "Lorentz" constant σ may be elucidated in the Inönü–Wigner contraction of SO (3,1) to ISO (3), when $\alpha\to 0$, $\sigma\to\infty$, with finite $z = \alpha\sigma$. Then $\mathbf{q}'\to\mathbf{q} + z\mathbf{p}/h$, showing an (arbitrary) amount of z translation (3.2), which will not affect ray direction. We will disregard this (purely spherical[2,4]) aberration and set $\sigma = 0$ henceforth. The transformation (4.6) and (4.7) of phase space may be verified to be canonical.

V. THE RELATIVISTIC COMA

Transformations of a four-dimensional manifold are difficult to visualize. The canonicity of the transformation only assures us that the manifold of rays will move as specks of dust in an incompressible fluid (Louville's theorem). A *section* of much use in optics is to choose a single "object" point $\mathbf{q}_0$, and plot $\mathbf{q}'(\mathbf{p},\mathbf{q}_0)$ as a function of $\mathbf{p}$ on part (or the whole) of its range. This corresponds to a bundle (or all) rays passing through the chosen $\mathbf{q}_0$ (as a point light source) imaged after the transformation. In the figures of this section we let $\mathbf{p}$ draw a polar coordinate grid around the optical axis, and plot the image $\mathbf{q}'\in\mathscr{R}^2$; this is the *spot* diagram of the optical transformation for $\mathbf{q}_0$.

When we take a square lattice of such object points a distance d apart, at $\mathbf{q}_0 + n_x(d,0) + n_y(0,d)$; $n_x n_y$ integers, we obtain the *spots* diagram (as in our figures), usually also called "spot." It depicts what is seen on the screen of an array of luminous points after the transformation to $(\mathbf{p}'(\mathbf{p},\mathbf{q}),\mathbf{q}'(\mathbf{p},\mathbf{q}))$. [The spot diagram *before* the transformation, i.e., $(\mathbf{p},\mathbf{q})$ is simply a square array of points, a perfectly focused 1:1 unit transformation of the object.]

We start the analysis of relativistic coma in the context of *aberration–expansion* optics, and will later consider its *global* characteristics. We must assume p/n to be less than unity so that the expansion of (4.6) and (4.7) may be performed by powers of p^2. This may mean $p^2 < n^2/10$ ($\theta < 18°26'\ldots$) or $p^2 < n^2/2$ ($\theta < 45°$), according to how high the order of aberration we are willing to calculate. To fifth aberration order we have the following fifth-degree approximation of relativistic coma:

$$\mathbf{p}' = e^{-\alpha}\mathbf{p} + \tfrac{1}{2}n^{-2}\sinh\alpha\, e^{-2\alpha}p^2\mathbf{p} + \tfrac{1}{4}n^{-4}\sinh\alpha\, e^{-2\alpha}(1-\tfrac{1}{2}e^{-2\alpha})p^2\mathbf{p} + \cdots, \quad (5.1a)$$

$$\mathbf{q}' = e^{\alpha}\mathbf{q} - n^{-1}\sinh\alpha\,\mathbf{p}\cdot\mathbf{q}\mathbf{p} - \tfrac{1}{2}n^{-2}\sinh\alpha\, p^2\mathbf{q} - \tfrac{1}{2}n^{-3}\sinh\alpha\, e^{-2\alpha}p^2\mathbf{p}\cdot\mathbf{q}\mathbf{p} - \tfrac{1}{8}n^{-4}\sinh\alpha(p^2)^2\mathbf{q} - \cdots, \quad (5.1b)$$

with increasingly complicated coefficients for higher $(p^2)^m\mathbf{p}$ in (5.1a), and $(p^2)^{m-1}\mathbf{p}\cdot\mathbf{q}\mathbf{p}$ and $(p^2)^m\mathbf{q}$ in (5.1b).

The first term on the right-hand sides of (5.1) is the *linear* part of the mapping. This falls within Gaussian (paraxial, linear) optics: $\mathbf{p}' = e^{-\alpha}\mathbf{p}$ is a contraction of ray momentum that necessitates (for canonicity) the expansion $\mathbf{q}' = e^{\alpha}\mathbf{q}$ of ray positions.

The rest of the series (5.1) is nonlinear and contains the *aberration* due to boost. It should be noted carefully that the only smallness parameter is p^2. Indeed, in α, the magnification part $e^{-\alpha}\mathbf{p} \simeq (1-\alpha)\mathbf{p}$ and the aberration part $\simeq \alpha(p^2/2n^2 + (p^2)^2/8n^4 + \cdots)\mathbf{p}$ in relativistic coma are of the same order; similarly for $\mathbf{q}'$.

In the expansion (5.1b) [*and* in the exact form (4.7)], it is useful to note that the particular function form $C(\mathbf{p},\mathbf{q}) = A\mathbf{p}\cdot\mathbf{q}\mathbf{p} + Bp^2\mathbf{q}$ maps a cone of rays around the optical axis ($\mathbf{q}$ and $\mathrm{p} = |\mathbf{p}|$ fixed, *twice* (for $\pm\mathbf{p}$) onto a *circle* in the spot diagram, with center at $(A/2 + B)p^2\mathbf{q}$, of radius $Ap^2q/2$, and extending between $(A+B)p^2\mathbf{q}$ (the image of the two *meridional* rays, i.e., in a plane with the optical axis, $\mathbf{p}\cdot\mathbf{q} = \pm pq$) and $Bp^2\mathbf{q}$ (the image of the two *saggital* rays across, $\mathbf{p}\cdot\mathbf{q} = 0$).

In Lie optics[2] the generator of *circular coma* aberration of order $2m+1$ is $f^c = (p^2)^m\mathbf{p}\cdot\mathbf{q}$. (This is M_{m10} in the monomial basis[10] and ${}^{m+1}\chi_m^{m+1}$ in the symplectic basis.[11]) The action of $\exp\kappa\hat{f}^c$ on phase space, to the aberration order, is

$$\mathbf{p}\to\mathbf{p} + \kappa(p^2)^m\mathbf{p}$$

and

$$\mathbf{q}\to\mathbf{q} - \kappa[2m(p^2)^{m-1}\mathbf{p}\cdot\mathbf{q}\mathbf{p} - (p^2)^m\mathbf{q}].$$

On this basis we recognize the relativistic aberration as *circular coma*. The third-order ($m=1$) comatic parameter is thus

$$\kappa^{(3)} = \tfrac{1}{2}n^{-2}e^{-\alpha}\sinh\alpha. \quad (5.2a)$$

In the factorization order[2]

$$e^{\alpha\hat{b}_z} = \cdots e^{\kappa^{(7)}\hat{f}_7^c}e^{\kappa^{(5)}\hat{f}_5^c}e^{\kappa^{(3)}\hat{f}_3^c}e^{-\alpha\widehat{\mathbf{p}\cdot\mathbf{q}}},$$

the fifth- and seventh-order coma parameters are found to be[9]

$$\kappa^{(5)} = \tfrac{1}{16}n^{-4}e^{-2\alpha}\sinh 2\alpha, \quad (5.2b)$$

$$\kappa^{(7)} = \tfrac{1}{32}n^{-6}e^{-3\alpha}\sinh 3\alpha. \quad (5.2c)$$

In Figs. 2 and 3 we show the spot(s) diagram of relativistic coma at values $\alpha = 0.3$ and $\alpha = -0.3$, respectively, for a 4×4 array of sources. The tips of the "comets" (wherefrom the name for *coma* aberration) exhibit the familiar 60° opening angle characteristic of third-order Seidel coma.[4] The angles τ which the circles subtend from the tip are not constant, however, but

$$\sin\tfrac{1}{2}\tau(p^2) = \tfrac{1}{2} + p^2(2e^{-2\alpha} - 1)/16n^2 + \cdots.$$

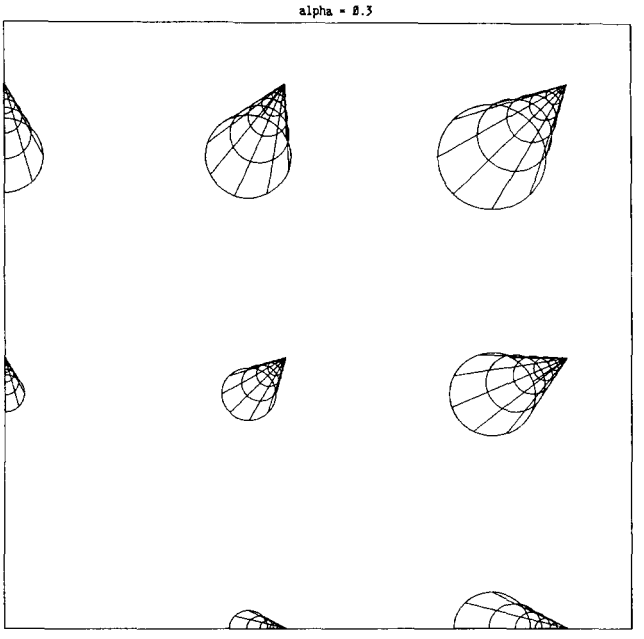


FIG. 2. Spots diagram of the relativistic coma transformation with positive $\alpha = 0.3$. A 4×4 array of object sources (the last row and column of which fall entirely outside the figure) is shown for ray angles of up to 45° (the values of momentum p are spaced by 0.101, up to 0.7071, corresponding to seven circles. The optical center is at the lower left corner.

For $\alpha = 0.3$, the p^2 coefficient is positive and so the comet opens; for values beyond $\alpha = \tfrac{1}{2}\ln 2 \simeq 0.3466$, from 60°, τ closes somewhat before opening again for p^2 in the far-metaxial region, to be discussed below.

The figures were drawn for p^2 up to $n^2/2$, i.e., for rays with angles θ of up to 45° from the direction of motion. This is more than what most instruments are designed for, but it allows us to discuss relativistic coma as a *global* aberration

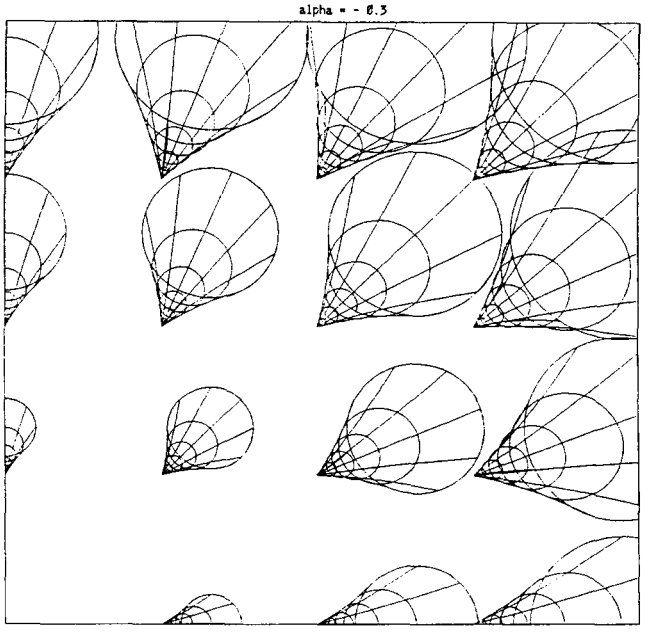


FIG. 3. Spots diagram of the relativistic coma transformation with negative $\alpha = -0.3$. The array of sources and angles are the same as in Fig. 2.

phenomenon. The figures were plotted using the *exact* expression (4.7) rather than any truncated aberration expansion (5.1b).

Consider what happens for negative α: as the screen moves in the $-z$ direction, (1.1) shows that some "critical" rays with angle θ_c to the optical axis will map onto rays with angle $\theta_c' = \pi/2$. This happens for $\tan \frac{1}{2}\theta_c = e^{\alpha}$ or

$$p_c = n \sin \theta_c = n \operatorname{sech} \alpha, \quad h_c = n \cos \theta_c = -n \tanh \alpha.$$

At this value, the denominator in (4.7) vanishes, and that cone of rays will map to infinity at the moving screen. The Poisson bracket $\{\mathbf{q}',\mathbf{p}'\}$ remains constant: the blowup of $\mathbf{q}'(\mathbf{q},\mathbf{p}_c)$ at θ_c is compensated by $p'(p_c)$ reaching its maximum at $\theta' = \pi/2$ and having zero derivative there. There is, of course, no physical singularity, as there is none for rotations in (3.3c) vis-à-vis Fig. 1, when $\theta + \gamma \to \pi/2$. Up to θ_c, the circles subtend angles up to 180°, while the distance from the circle to the comet tip slowly increases up to and beyond θ_c.

The *global* picture of the relativistic coma aberration is the mapping of the whole direction *sphere*. We note that forward rays ($h>0$) under backward motion ($\alpha<0$) are the same as backward rays ($h<0$) under forward motion ($\alpha>0$); indeed, Eq. (4.7) is invariant under the exchange $(h,\alpha)\leftrightarrow(-h,-\alpha)$. Thus while Fig. 2 is the image of rays around the *forward* pole of the direction sphere, $\alpha>0$, Fig. 3 is the image of rays around the *backward* pole, also for $|\alpha|$. To see the spot diagram of the whole direction sphere we may superpose both figures: the spot will extend from $\mathbf{q}_F' = e^{\alpha}\mathbf{q}$ (the image of the forward ray, along the optical axis) to $\mathbf{q}_B' = e^{-\alpha}\mathbf{q}$ (the image of the backward ray, counter to that axis). The full comatic *caustic* acquires a diamond shape, with two 60° angles at the two finite tips, $\mathbf{q}_F$ and $\mathbf{q}_B$, and two "vertices" at infinity in the perpendicular direction. The location of the asymptotic caustic line may be found in (4.7) as the position of the saggital rays ($\mathbf{p}\cdot\mathbf{q}=0$) at the critical angle θ_c: it is $\mathbf{q}_c = \operatorname{sech}\alpha\ \mathbf{q}$. The three points: $\mathbf{q}_B$, $\mathbf{q}_c$, and $\mathbf{q}_F$ lie in a line in that order. The region of the direction sphere accessible to optical focusing is in practice very limited, of course.

Thus far we referred to the boost aberration as coma, because of its striking appearance when the screen movement is in the z direction. The effect of boosts *in* the screen plane, specifically, $\hat{b}_x$, will be described now more succinctly.

The boost function b_x in (4.2a) for $\sigma=0$ will generate the Lie transformation on the ray position plane,

$$e^{\alpha \hat{b}_x} q_x = q_x' = (\cosh\alpha + p_x n^{-1}\sinh\alpha) \times (q_x \cosh\alpha + \mathbf{p}\cdot\mathbf{q} n^{-1}\sinh\alpha), \tag{5.3a}$$

$$e^{\alpha \hat{b}_x} q_y = q_y' = (\cosh\alpha + p_x n^{-1}\sinh\alpha) q_y, \tag{5.3b}$$

while on the ray direction sphere it yields the familiar stellar aberration in the x direction.

To first order in α, the aberration of phase space produces spot diagrams with elliptical spots. If $q_y/q_x = \tan\tau$, τ is the angle between the object point and the direction of motion (the x axis here), the ellipses are tilted by $\kappa = \tau/2$, have "major" axis $a(\tau) = \alpha pq(1+\cos\tau)/n$, and "minor" axis $b(\tau) = \alpha pq(1-\cos\tau)/n$. We note that $a(\tau) = b(\pi-\tau)$, $b(\tau) = a(\pi-\tau)$, so that the longer axes of the ellipses are closer to the x axis. For object points *on* the x axis the ellipse degenerates into a line segment of length $2a(0) = 4pq\alpha/n$, mimicking third-order Seidel astigmatism[2] on that axis. Object points on the y axis have their spots circular, as in third-order Seidel curvature of field.

This first-order description holds up to $p=n$, i.e., for the whole forward ($0\leqslant\theta\leqslant\pi/2$) hemisphere of rays. The *same* spots are obtained from the backward ($\pi/2\leqslant\theta\leqslant\pi$) hemisphere: note that (5.3) do not depend on the sign of h. The global mapping of the direction sphere on the image plane generated by b_x is thus *also* a 2:1 mapping. The global coma of boosts has a variety of faces according to the orientation of the observer screen with respect to the boost direction, resembling Seidel coma in the z direction, and an asymmetric kind of Seidel astigmatism/curvature of field aberration for directions of boost in the x-y plane. Intermediate orientations should interpolate between these faces.

Regarding the observability of relativistic coma, stellar aberration is the ray direction aspect of the phenomenon. To observe it in ray *position* space (regardless of the imaging apparatus one may contrive) we may present the following estimate of aberration *size*: at satellite speeds of $V\sim 10$ km/sec, $\alpha\sim 3\times10^{-5}$, a $\theta = 0.1r\sim 5.7°$ cone of ray directions around the optical axis will yield a factor of $\alpha p/n\sim 3\times10^{-6}$. Under z boosts this will spread into a circle of radius $1.5\times10^{-7}q$ in a coma whose *caustic* has a relative size of the order of $10^{-7}q$. For boosts in the screen plane, we may have from circles of radius $3\times10^{-6}q$ to caustic segments of length $6\times10^{-6}q$. The linear factor q gives the relative scale of the aberration to object size.

VI. CONCLUDING REMARKS

The spirit of our prediction of a relativistic comatic aberration due to screen motion has been Lie theoretical. In that vein we should add the following glossary and comments beyond geometric optics.

The three-dimensional Euclidean group is the dynamical group of optics in a homogeneous medium.[12] The deformation[5,6] of this group on the (ray direction) sphere leads to the Lorentz group of special relativity acting on the same sphere. When the projection of this sphere is called momentum space for a system, the canonically conjugate position space will undergo spherical aberration and circular coma when the screen is boosted (in *first* aberration order these are free flight and pure magnification) perpendicular to itself.

The Casimir invariant of the Lorentz group is related to the freedom in translating position space without affecting ray direction. This role of the invariant seems to be new and needs to be further exploited.

Finally, the relativistic transformation is *global* over the phase space of geometric optics, singularities notwithstanding.

The group theoretical objects mentioned above were seen here in the geometric optics realization. They possess other realizations, however, that are better known,[13] and

that will be explored to clarify further the "wavization"[14,15] process. It seems this should be parallel to quantization,[7] but based on the Euclidean, rather than Heisenberg–Weyl, algebra.

ACKNOWLEDGMENTS

We would like to thank Mat. Guillermo Correa for the SPOT_D program for screen and printer output of spot diagrams. It works with exact formulas, not only aberration expansions, as we discovered. The two visitors (NMA, WL) would like to acknowledge the hospitality of the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, in México D.F. and especially in Cuernavaca.

[1]V. Bargmann, Ann. Math. **48**, 568 (1947), Sect. 6; S. Ström, Ark. Fys. **30**, 455 (1965); A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967); K. B. Wolf, *ibid.* **12**, 197 (1971), Eq. (4.13).

[2]A. J. Dragt, E. Forest, and K. B. Wolf, "Foundations of a Lie algebraic theory of geometrical optics," in *Lie Methods in Optics, Lecture Notes in Physics*, Vol. 250, edited by J. Sánchez-Mondragón and K. B. Wolf (Springer, Heidelberg, 1986), Chap. 4.

[3]T. Sekiguchi and K. B. Wolf, Am. J. Phys. **55**, 830 (1987).

[4]H. Buchdahl, *An Introduction to Hamiltonian Optics* (Cambridge U. P., Cambridge, 1970).

[5]R. Gilmore, *Lie Groups, Lie Algebras, and Some of their Applications* (Wiley, New York, 1974), Chap. 10.

[6]C. P. Boyer and K. B. Wolf, J. Math. Phys. **14**, 1853 (1973); K. B. Wolf and C. P. Boyer, *ibid.* **15**, 2096 (1974).

[7]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U.P., Cambridge, 1984).

[8]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed.

[9]S. Steinberg, "Lie series, Lie transformations, and their applications," in Ref. 2, pp. 45–104.

[10]A. J. Dragt, *Lectures in Nonlinear Orbit Dynamics, American Institute of Physics, Conference Proceedings*, Vol. 87 (AIP, New York, 1982).

[11]K. B. Wolf, Ann. Phys. (NY) **172**, 1 (1986).

[12]K. B. Wolf, "Dynamical groups in Lie optics," in *XV International Colloquium on Group Theoretical Methods in Physics*, edited by R. Gilmore (World Scientific, Singapore, 1987).

[13]N. Mukunda, R. Simon, and E. C. G. Sudarshan, Phys. Rev. A **28**, 2921, 2933 (1983), J. Opt. Soc. Am. A **2**, 416 (1985).

[14]K. B. Wolf, Kinam **6**, 141 (1984).

[15]T. Pradhan, Phys.Lett. A **122**, 397 (1987).

The relativistic coma aberration. II. Helmholtz wave optics

Natig M. Atakishiyev,[a] Wolfgang Lassner,[b] and Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas/Cuernavaca, Universidad Nacional Autónoma de México, Apdo. Postal 20-726, 01000 México D.F., Mexico

(Received 18 November 1988; accepted for publication 24 May 1989)

The relativistic coma aberration of geometric optics was examined in the first paper [J. Math. Phys. **30**, 2457 (1989)]. Here is a study of a unitary realization of the Euclidean group, the dynamic group for global optics, on the space of solutions of the Helmholtz equation. Deformation to the Lorentz group of relativity yields the self-adjoint generators of boost transformations on that space. Graphic results for the action of a boost normal to the screen on an off-axis Gaussian beam, that may be compared with classical wave optics results on diffraction in aberration, are presented.

I. INTRODUCTION

In the first part of this work[1] we posed the well-known phenomenon of stellar aberration in the context of geometrical Lie optics in phase space.[2] Given a relativistic distortion of ray directions on the sphere, we showed that the canonically conjugate ray positions undergo magnification and a circular comatic aberration. This we called relativistic coma. The phenomenon was analyzed globally, i.e., on the full, proper phase space manifold of rays, and given in closed, explicit formulas, and also as an expansion in aberration up to seventh order. We pointed out that the basic dynamical group of optics is the Euclidean group,[3] rather than the Heisenberg–Weyl,[4] diamond,[5] or Weyl-symplectic[6] groups used in nonrelativistic quantum mechanics. The latter group appears as a *contraction* of the first in the paraxial approximation.[7]

The Euclidean ISO(3) group of rigid motions of three-space was *deformed*[8,9] to relativistic SO(3,1) transformations. This provided stellar aberration for boosts on the ray directions of geometric optics. The spot diagrams were obtained for the full direction sphere, with the standard comatic appearance for boosts normal to the screen plane. For boosts in the screen plane, the aberration had the characteristics of an astigmatism and curvature of field, along and normal to the boost.[1] The group realization we used was that of geometrical optics phase space.

In this paper we apply the same construction to the same groups, but in the realization on the space of solutions of the Helmholtz equation studied in Ref. 10. Geometric optics has no time variable; the Helmholtz equation does not contain it either. The space of solutions f of

$$(\partial_x^2 + \partial_y^2 + \partial_z^2 + k^2) f(x,y,z) = 0 \tag{1.1}$$

that are square integrable over any plane *screen* and whose spectrum there is bounded by k, is invariant under the Euclidean group ISO(3) of screen translations and rotations, and we call it $\mathscr{H}_k$. For $\mathbf{f} \in \mathscr{H}_k$, written as a two-function column $\mathbf{f}(\mathbf{q}) = (f(\mathbf{q}), f_z(\mathbf{q}))$ on the plane $\mathbf{q} = (x,y)$, we can present the Helmholtz equation in the form

$$\mathbf{H}\mathbf{f}(\mathbf{q},z) = \begin{pmatrix} 0 & 1 \\ -\Delta_k & 0 \end{pmatrix} \begin{pmatrix} f(\mathbf{q},z) \\ f_z(\mathbf{q},z) \end{pmatrix} = \frac{\partial}{\partial z} \begin{pmatrix} f(\mathbf{q},z) \\ f_z(\mathbf{q},z) \end{pmatrix}, \tag{1.2}$$

where $\Delta_k = \partial_x^2 + \partial_y^2 + k^2$. The formal solution may be written as an evolution of initial conditions:

$$\mathbf{f}(\mathbf{q},z) = \exp(z\mathbf{H})\mathbf{f}(\mathbf{q}), \tag{1.3}$$

for $f(\mathbf{q})$ and its normal derivative $f_z(\mathbf{q})$ at the reference screen $z = 0$ to all of $\mathfrak{R}^3$. In $\mathscr{H}_k$, the initial value problem for the system is well posed.[10] As in geometrical optics, we work with observables and wavefunctions in the plane of a screen.

We present the realization of the Euclidean group ISO(3) on $\mathscr{H}_k$ in Sec. II, completed with respect to an inner product that is conserved under Euclidean transformations of the screen. We regard the ensuing unitary representation of the Euclidean group as defining the Helmholtz wavization of geometrical optics. In Sec. III we proceed to deform ISO(3) to the Lorentz group SO(3,1) on $\mathscr{H}_k$. In Sec. IV we present the results for the z boost studied in Ref. 1 on an off-center forward Gaussian beam. The explicit computation is done to fifth order in the relativistic boost parameter. The "isophotes" of $|f(\mathbf{q})|^2$ are comparable to those seen and calculated[11] for diffraction in third-order pure Seidel coma. The closing section presents some conclusions and open comments on nonlocality, observability, and the role of the normal derivative $f_z(\mathbf{q})$ of the field at the screen.

II. THE EUCLIDEAN GROUP OF HELMHOLTZ WAVE OPTICS

A well known realization of the generators of the Euclidean algebra of translations and rotations on smooth functions of $\mathfrak{R}^3$, is given by

$$P_x' = -i\lambda\partial_x, \quad P_y' = -i\lambda\partial_y, \quad P_z' = -i\lambda\partial_z, \tag{2.1a}$$

$$R_x' = i(y\partial_z - z\partial_y), \quad R_y' = i(z\partial_x - x\partial_z), \quad R_z' = i(x\partial_y - y\partial_x), \tag{2.1b}$$

where λ is a consant with units of length, to render the operators dimensionless. The action of the corresponding Lie exponential group ISO(3) is that of ordinary, rigid transformations of $\mathfrak{R}^3$. These operators are self-adjoint in $\mathscr{L}^2(\mathfrak{R}^3)$

[a] Permanent address: Institute of Physics, Azerbaijan Academy of Sciences, Baku 370143, Union of Soviet Socialist Republics.
[b] Permanent address: Naturwissenschaftlich-theoretisches Zentrum, Karl-Marx University, Leipzig 7010, German Democratic Republic.

 0022-2488/89/112463-06$02.50

and their (i) exponential, i.e., $\exp(i\alpha_j P_j')$, $\exp(i\beta_j R_j')$, $j = x,y,z$, $\alpha \in \Re$, is unitary.

For functions in the solution space $\mathscr{H}_k$ of the Helmholtz equation (1.1), we may replace[10] ∂_z by the matrix operator in (1.2), acting on two-functions $\mathbf{f}(\mathbf{q}) = (\mathbf{f}(\mathbf{q}), f_z(\mathbf{q}))$ at the reference $z = 0$ plane, and identify $\lambda\!\!\!^{-}$ with the reduced wavelength of (1.1), namely, $\lambda\!\!\!^{-} = \lambda/2\pi = 1/k$. This takes the place of $\hbar$ in quantum mechanics. In the Helmholtz realization, the generators of translations become

$$P_x = -i\lambda\!\!\!^{-}\begin{pmatrix} \partial_x & 0 \\ 0 & \partial_x \end{pmatrix}, \quad P_y = -i\lambda\!\!\!^{-}\begin{pmatrix} \partial_y & 0 \\ 0 & \partial_y \end{pmatrix}, \quad P_z = -i\lambda\!\!\!^{-}\begin{pmatrix} 0 & 1 \\ -\Delta_k & 0 \end{pmatrix}. \tag{2.2a}$$

As in geometrical optics, P_z takes the role of the Hamiltonian operator, generating *z evolution* in the system. The generators of rotations become

$$R_x = -i\begin{pmatrix} 0 & y \\ -y\Delta_k - \partial_y & 0 \end{pmatrix},$$
$$R_y = -i\begin{pmatrix} 0 & -x \\ x\Delta_k + \partial_x & 0 \end{pmatrix}, \tag{2.2b}$$
$$R_z = -i\begin{pmatrix} x\partial_y - y\partial_x & 0 \\ 0 & x\partial_y - y\partial_x \end{pmatrix}.$$

Note that in the 2-1 elements of R_x and R_y there could be an ordering ambiguity between x, y, and Δ_k. These elements are given by the anticommutator, $\frac{1}{2}(x\Delta_k + \Delta_k x)$; that is the only possibility when we demand closure under commutation

$$[R_x,R_y] = iR_z, \quad [R_x,P_y] = iP_z, \quad [P_x,P_y] = 0, \quad \text{and cyclically.} \tag{2.3}$$

The i's above fit generators that are self-adjoint under an inner product. It is easy to see that an $\mathscr{L}^2(\Re^2)$ product allows for the self-adjointness of the diagonal matrix operators P_x, P_y, and R_z, generators of the ISO(2) symmetry group of screen motions in its plane, but not for the z evolution P_z and the out-of-screen rotations R_x and R_y.

In Ref. 10, Steinberg and Wolf found the (unique) Euclidean-invariant inner product of solutions of the (well-posed) Helmholtz equations in two dimensions, through proposing a nonlocal sesquilinear form. In the case of *three* dimensions, we have

$$(\mathbf{f},\mathbf{g})_{\mathscr{H}_k} = \int_{\Re^2} d^2\mathbf{q} \int_{\Re^2} d^2\mathbf{q}'\mathbf{f}(\mathbf{q})^\dagger \mathbf{M}^k(\mathbf{q},\mathbf{q}')\mathbf{g}(\mathbf{q}'). \tag{2.4a}$$

The 2×2 matrix $\mathbf{M}^k(\mathbf{q},\mathbf{q}') = \|M^k_{j,j'}(x,y,x',y')\|$ is determined by the requirement that the algebra operators P_j, R_j, satisfy $(\mathbf{f},P_j\mathbf{g})_{\mathscr{H}_k} = (P_j\mathbf{f},\mathbf{g})_{\mathscr{H}_k}$, etc. We find here

$$\mathbf{M}^k(\mathbf{q},\mathbf{q}') = \begin{pmatrix} kj_1(k|\mathbf{q}-\mathbf{q}'|)/|\mathbf{q}-\mathbf{q}'| & 0 \\ 0 & j_0(k|\mathbf{q}-\mathbf{q}'|) \end{pmatrix}, \tag{2.4b}$$

where j_0 and j_1 are the spherical Bessel functions. The form is positive definite on $\mathscr{H}_k$. Completion with respect to this inner product turns $\mathscr{H}_k$ into a Hilbert space where the Euclidean transformations are unitary. This realization we call the Helmholtz representation of ISO(3). The Casimir invariants are $P^2 = P_x^2 + P_y^2 + P_z^2 = 1$, and $\mathbf{P}\cdot\mathbf{R} = 0$. This means that we have a sphere for homogeneous space under the Euclidean algebra, its enveloping algebra, and the group; physically, it is a homogeneous space of unit refractive index. The SO(3) subgroup Casimir operator is diagonal (but not simply a multiple of the unit operator):

$$\mathbf{R}^2 = R_x^2 + R_y^2 + R_z^2 = \begin{pmatrix} D(D+1) + q^2k^2 & 0 \\ 0 & (D+1)(D+2) + q^2k^2 \end{pmatrix}, \tag{2.5}$$

where

$$D = x\partial_x + y\partial_y. \tag{2.6}$$

The Helmholtz representation space $\mathscr{H}_k$ has an inner product (2.4) that is nonlocal in the wave function $f(\mathbf{q})$ to the extent of $j_1(|\mathbf{q}|/\lambda\!\!\!^{-})/(|\mathbf{q}|/\lambda\!\!\!^{-})$, and in the normal derivative function $f_z(\mathbf{q})$ to the extent of $j_0(|\mathbf{q}|/\lambda\!\!\!^{-})$, of the order of $\lambda\!\!\!^{-}$. Both f and f_z contribute to the energy of an elastic medium[12] so we may identify $(\mathbf{f},\mathbf{f})_{\mathscr{H}_k}$ with total field energy on the screen. As we shall show below, this inner product may be brought to local form in an appropriate transform space. Finally, it seems we should identify $|f(\mathbf{q})|^2$ with the visible image illumination.

The Euclidean algebra and its covering have been used by Vilenkin[13] and Miller[14] to find all separable solutions of the Helmholtz equation (1.1). The algebra itself yields the three subalgebra bases of Cartesian, cylindrical, and spherical coordinates; the corresponding separated functions are plane waves, the nondiffracting J_m beams of Durnin *et al.*,[15] and multipole solutions. The covering algebra provides the rest of the 11 coordinate system where the equation separates. Let us draw attention here to the plane-wave generalized eigenbasis of the translation subalgebra P_x, P_y, and a sign (of P_z). Up to an arbitrary normalization constant κ, with units of (illumination/area)$^{1/2}$,

$$\Phi_{k_x,k_y,\sigma} = \frac{\kappa}{2\pi}\begin{pmatrix} 1 \\ ik_z \end{pmatrix} e^{i(xk_x + yk_y)} \tag{2.7}$$

is a plane wave on the screen, labeled by the respective dimensionless eigenvalues, $\lambda\!\!\!^{-}k_x$, $\lambda\!\!\!^{-}k_y$, and $\sigma \in \{-1,0,+1\}$ ($k_z = \sigma\sqrt{k^2 - k_x^2 - k_y^2}$). The manifold of plane waves (k_x,k_y,σ) is that of two disks δ_k^σ, of radius k, sown at the edge of δ_k^0. This is the sphere of ordinary plane-wave direction three-vectors, projected on its equatorial screen plane. Solutions in $\mathscr{H}_k$ may be written as a generalized linear combination of these basis functions,

$$\mathbf{f}(\mathbf{q}) = \lambda^2 \sum_\sigma \int_{\delta_k^\sigma} d^2\mathbf{k}\, \tilde{f}_\sigma(\mathbf{k})\Phi_{\mathbf{k},\sigma}(\mathbf{q}). \tag{2.8}$$

We refer to the ordinary two-dimensional Fourier transform to write explicitly

$$f(\mathbf{q}) = \kappa\lambda\!\!\!^{-2} \int_{\delta_k^+} d^2\mathbf{k}\, [\tilde{f}_+(\mathbf{k}) + \tilde{f}_-(\mathbf{k})]e^{i\mathbf{k}\cdot\mathbf{q}}, \tag{2.9a}$$

$$f_z(\mathbf{q}) = i\kappa\lambda\!\!\!^{-2} \int_{\delta_k^+} d^2\mathbf{k}\, k_z[\tilde{f}_+(\mathbf{k}) - \tilde{f}_-(\mathbf{k})]e^{i\mathbf{k}\cdot\mathbf{q}}. \tag{2.9b}$$

Now, replacing (2.9) in the convoluted inner product (2.4), we exchange integrals and note that the **q**-Fourier transform of $kj_1(k|\mathbf{q}|)/|\mathbf{q}|$ is k_z, and that of $j_0(k|\mathbf{q}|)$ is $1/k_z$ on $\delta_k^{\pm}$, and zero outside. We may thus write the Parseval relation between (2.4) and the local form on one disk δ_k,

$$(\mathbf{f},\mathbf{g})_{\mathscr{H}_k} = 2\pi|\kappa\hbar|^2 \int_{\delta_k} d^2\mathbf{k}\,\frac{k}{k_z}\,[\tilde{f}_+(\mathbf{k})^*\tilde{g}_+(\mathbf{k}) + \tilde{f}_-(\mathbf{k})^*\tilde{g}_-(\mathbf{k})]. \tag{2.10}$$

This is a local integration over the wave-vector sphere projected on the screen plane, including both forward (+) and backward (−) waves, with the obliquity factor $k/k_z = \sec\theta$, where θ is the angle between the wave three-vector and the normal to the screen. The $\mathscr{L}^2(\Re^2)$ norm majorizes the $\mathscr{H}_k$ norm.

The Euclidean group has thus a geometrical optics model and a Helmholtz optics model. The generators of the Abelian translation ideal of the abstract Lie algebra, P_x, P_y, and P_z, are in geometrical optics the optical momenta and the Hamiltonian p_x, p_y, and $p_z = h = \sqrt{1-|\mathbf{p}|^2}$ (for unit refractive index); the Lie bracket is the Poisson bracket. In Helmholtz wave optics, the homomorphic realization of this subalgebra is given by (2.3a) acting on $\mathscr{H}_k$ described above, and the Lie bracket is the commutator.

III. THE DEFORMATION OF THE EUCLIDEAN TO THE LORENTZ GROUP

Out of the Helmholtz representation of the Euclidean algebra (2.2)–(2.4) we may construct a representation of the Lorentz algebra through deformation. The deformation extends to the corresponding groups. The generators of the SO(3,1) Lorentz group are the following: the SO(3) generators are those of the Euclidean group R_x, R_y, and R_z in (2.2b); the boosts are built as $B_j = [R^2,P_j] + (\tau + i)P_j$, $j = x,y,z$. This formula is the heart of the deformation process.[16] For real τ, the boost generators are self-adjoint in $\mathscr{H}_k$, and belong to the nonexceptional continuous representation series. As in Ref. 1, we set $\tau = 0$ on the grounds that this parameter only reflects an admixture of translations to the basic boost transformation.

We build the following explicit boost matrix operators:

$$B_x = R_zP_y - R_yP_z + iP_x = \hbar\begin{pmatrix}(D+1)\partial_x + k^2x & 0\\ 0 & (D+2)\partial_x + k^2x\end{pmatrix}, \tag{3.1a}$$

$$B_y = R_xP_z - R_zP_x + iP_y = \hbar\begin{pmatrix}(D+1)\partial_y + k^2y & 0\\ 0 & (D+2)\partial_y + k^2y\end{pmatrix}, \tag{3.1b}$$

$$B_z = R_yP_x - R_xP_y + iP_z = \hbar\begin{pmatrix}0 & D+1\\ -(D+2)\Delta_k + k^2 & 0\end{pmatrix}. \tag{3.1c}$$

We may verify that these indeed close into the Lie algebra of the SO(3,1) group:

$$[R_x,R_y] = iR_z, \quad [R_x,B_y] = iB_z, \quad [B_x,B_y] = -iR_z, \quad \text{and cyclically.} \tag{3.2}$$

The Lorentz Casimir invariants are $B^2 - R^2 = \mathbf{1}$ [for arbitrary τ it would be $(\tau+1)^2\mathbf{1}$], and $\mathbf{B}\cdot\mathbf{R} = 0$.

In most applications of Lie theory—including geometric optics[1]—the algebra generators are functions of some variable(s) ξ, times a first-order differential operator in ξ, that we write $\partial^{(1)}$ for short. These lead to point-to-point transformations $\exp(i\alpha\,\partial^{(1)})\,f(\xi) = f(\xi'(\xi;\alpha))$. Bargmann[8] deformed the Euclidean algebra with $\partial^{(1)}$ plus $\partial^{(0)}$—functions of ξ; this leads to multiplier representations of the group as $\exp(i\alpha[\partial^{(1)} + \partial^{(0)}])\,f(\xi) = \mu(\xi;\alpha)\,f(\xi'(\xi;\alpha))$. The multiplier function μ just offsets the change in the measure $\partial\xi/\partial\xi'$ so that the generators be Hermitian; when technical conditions on the domain hold, they are self-adjoint.[8] The exponentiation of second-[17,18] and higher-order[19] differential operators leads in general to integral transforms: $\exp(i\alpha\partial^{(>2)})\,f(\xi) = \int d\xi'\,K(\xi,\xi';\alpha)\,f(\xi')$. The particular case of paraxial scalar wave optics[20] is well studied and requires of $\partial^{(0)}$, $\partial^{(1)}$, and $\partial^{(2)}$; it leads to the canonical integral transforms.[6,18]

Of the boost generators (3.1), B_x and B_y are matrices with elements $\partial^{(0)} + \partial^{(1)} + \partial^{(2)}$ on the diagonal, while B_z has $\partial^{(0)} + \partial^{(1)}$ and $\partial^{(0)} + \partial^{(1)} + \partial^{(2)} + \partial^{(3)}$ on the antidiagonal. They seem to be more difficult to exponentiate in closed form than they were in Ref. 1, and here we do not attempt to do so. We can state, however, that because of the homomorphism of transformations of ray momenta in geometric optics and plane-wave vectors under SO(3,1), the latter will behave in the same way as the former [viz., Ref. 1, Eq. (4.6)], conforming to the distortion of directions characteristics of stellar aberration, namely,

$$\tan\tfrac{1}{2}\theta \mapsto \tan\tfrac{1}{2}\theta' = e^{-\alpha}\tan\tfrac{1}{2}\theta, \tag{3.3}$$

where $\theta,\theta' \in [0,\pi]$ are the angles between the rays or wave vectors and the direction of the boost to velocity $v = c\tanh\alpha$. We now examine the effect of these unitary transformations on the screen images for the case of boosts in the z direction.

IV. THE COMA ABERRATION IN z BOOSTS

We consider boosts in the direction normal to the screen. The antidiagonal matrix elements of B_z in (3.1c) may be written in terms of formal Schrödinger operators [$\mathsf{Q}_j f(\mathbf{q}) = q_j f(\mathbf{q})$, $\mathsf{P}_j f(\mathbf{q}) = -i\hbar\,\partial_{q_j} f(\mathbf{q})$, $j = x,y$, distinguished in sans serif font] in the following way:

$$\mathscr{D} = \hbar(D+1) = i\mathsf{Q}\cdot\mathsf{P} + \hbar = i(\mathbf{p}\cdot\mathbf{q})_Q = 2iX_0^1, \tag{4.1a}$$

$$\mathscr{E} = \hbar[k^2 - (D+2)\Delta_k] = ik^2[\mathsf{Q}\cdot\mathsf{P}\mathsf{P}^2 - \mathsf{Q}\cdot\mathsf{P} - 2i\hbar\mathsf{P}^2 + i\hbar] = ik^2[(p^2\mathbf{p}\cdot\mathbf{q})_Q - (\mathbf{p}\cdot\mathbf{q})_Q] = ik^2(X_1^2 - 2X_0^1). \tag{4.1b}$$

In these expressions, $\mathscr{D}$ and $\mathscr{E}$ are written in terms of standard ordered operators (derivatives to the right), and in terms of the [$\mathscr{L}^2(\Re^2)$-Hermitian] operators that quantize, à la Schrödinger (Q), the classical observables $\mathbf{p}\cdot\mathbf{q}$ and $p^2\mathbf{p}\cdot\mathbf{q}$, namely, $X_0^1 = \frac{1}{2}(\mathbf{p}\cdot\mathbf{q})_Q$ and $X_1^2 = (p^2\mathbf{p}\cdot\mathbf{q})_Q$. We recall that the quantization of functions of the general form $A(p)q + B(p)$ to $\mathscr{L}^2$-Hermitian operators is (modulo weak tech-

nical assumptions) independent of the scheme (Weyl, Born–Jordan, symmetrization, etc.).[4] The operator X_0^1 generates $\mathscr{L}^2(\Re^2)$-unitary *dilatations* on functions of position and, since $p^2\mathbf{p}\cdot\mathbf{q}$ generates the geometrical Seidel *coma* aberration, X_1^2 will be the coma operator. The indices reflect the placement of the X's within the symplectic aberration multiplets classified in Ref. 3, and the results on their quantization in Ref. 21. Note that the off-diagonal operator $\mathscr{E}$ is a combination of magnification and coma with the same order in α. [The exponentiation of the boost generators in the plane of the screen, B_x and B_y entails exponentiating the diagonal elements $\mathscr{D}\nabla + k\mathbf{q} = k\mathsf{Q} - (\mathbf{p}\cdot\mathbf{q})_Q\mathsf{P}$ and $\nabla\mathscr{D} + k\mathbf{q} = k\mathsf{Q} - \mathsf{P}(\mathbf{p}\cdot\mathbf{q})_Q$.]

The effect of a finite boost in the z direction on wave functions on the screen and their normal derivatives may be written formally in terms of the operators $\mathscr{F}^2 = \mathscr{D}\mathscr{E}$ and $\mathscr{G}^2 = \mathscr{E}\mathscr{D}$ [see (3.5)], as

$$\begin{pmatrix} f(\alpha) \\ f_z(\alpha) \end{pmatrix} = \exp i\alpha \begin{pmatrix} 0 & \mathscr{D} \\ \mathscr{E} & 0 \end{pmatrix} \begin{pmatrix} f(0) \\ f_z(0) \end{pmatrix} = \begin{pmatrix} \cos\alpha\mathscr{F} & i\alpha\mathscr{D}\ \operatorname{sinc}\alpha\mathscr{G} \\ i\alpha\mathscr{E}\ \operatorname{sinc}\alpha\mathscr{F} & \cos\alpha\mathscr{G} \end{pmatrix} \begin{pmatrix} f(0) \\ f_z(0) \end{pmatrix} \tag{4.2a}$$

where only even powers of $\mathscr{F}$ and $\mathscr{G}$ appear in the cosine and sinc functions [$\operatorname{sinc} x = x^{-1}\sin x = 1 - (1/3!)x^2 + (1/5!)x^4 - \cdots$]. The expansion of the matrix to fifth order in α is

$$\begin{pmatrix} 1 - (1/2!)\alpha^2\mathscr{D}\mathscr{E} + (1/4!)\alpha^4\mathscr{D}\mathscr{E}\mathscr{D}\mathscr{E} + \cdots \\ i\alpha\mathscr{E} - i(1/3!)\alpha^3\mathscr{E}\mathscr{D}\mathscr{E} + i(1/5!)\alpha^5\mathscr{E}\mathscr{D}\mathscr{E}\mathscr{D}\mathscr{E} + \cdots \\ i\alpha\mathscr{D} - i(1/3!)\alpha^3\mathscr{D}\mathscr{E}\mathscr{D} + i(1/5!)\alpha^5\mathscr{D}\mathscr{E}\mathscr{D}\mathscr{E}\mathscr{D} + \cdots \\ 1 - (1/2!)\alpha^2\mathscr{E}\mathscr{D} + (1/4!)\alpha^4\mathscr{E}\mathscr{D}\mathscr{E}\mathscr{D} + \cdots \end{pmatrix}. \tag{4.2b}$$

In order to present concrete results comparable with other developments in Fourier aberration optics,[22] we apply this expansion to a forward Gaussian beam with waist at the screen plane and centered at $\mathbf{q} = \mathbf{a}$. This we write as

$$\mathbf{G}_w^a(\mathbf{q}) = \begin{pmatrix} 1 \\ ik \end{pmatrix} E_w(\mathbf{q} - \mathbf{a}), \quad E_w(\mathbf{q}) = \exp(-|\mathbf{q}|^2/2w). \tag{4.3}$$

Of course, a Gaussian is not strictly in $\mathscr{H}_k$, since its Fourier transform is a Gaussian (in $\mathbf{k}$) of width $1/w$, that has generally small but nonzero support outside the disk δ_k. We assume that the spread of directions off the $+z$ axis is small, so that the approximation holds good and that we may replace the obliquity factor k_z in the normal derivative by the constant k in (4.3). Gaussians beams in the $-z$ direction reverse the sign of the second component, i.e., complex conjugate (4.3). (A null second component would indicate it is a solution even in z, with maximum amplitude at the screen.)

Equation (4.2b) gives the effect of a z boost as a series of derivative operators that is straightforward to apply to the Gaussian (4.3) algorithmically through symbolic computation, albeit the approximation errors of the assumption that are not easy to estimate except by examining the stability of the main features of the graphic outcome. The series for the amplitude f is

$$f_{w,a}^{\alpha}(\mathbf{q}) = (1 - \alpha k\mathscr{D} - (1/2!)\alpha^2\mathscr{D}\mathscr{E} + (1/3!)\alpha^3 k\mathscr{D}\mathscr{E}\mathscr{D} + (1/4!)\alpha^4\mathscr{D}\mathscr{E}\mathscr{D}\mathscr{E} - (1/5!)\alpha^5 k\mathscr{D}\mathscr{E}\mathscr{D}\mathscr{E}\mathscr{D} + \cdots)E_w(\mathbf{q} - \mathbf{a}). \tag{4.4}$$

We recall that $\mathscr{D}$ and $\mathscr{E}$ are real derivative operators; when acting on Gaussian functions they yield polynomial factors of degrees 2 and 4, respectively. The summands in (4.4) thus yield polynomials of degrees 0, 2, 6, 8, 12, and 14 in x and y, followed by degrees 18, 20, 24,..., that we disregard on account of the power of α of the approximation.

To first order in α, (4.2b) and (4.4) represent magnification $f \to f + i\alpha f_z + \cdots$ by the normal derivative. The factor polynomial of the Gaussian is $1 - \alpha[1 - \mathbf{q}\cdot(\mathbf{q} - \mathbf{a})]$. For $\alpha > 1/\sqrt{1 + |\mathbf{a}|^2/4w}$ this polynominal vanishes on a circle with center at $\mathbf{q} = \frac{1}{2}\mathbf{a}$, and radius $\sqrt{|\mathbf{a}|^2/4 + w(1 - 1/\alpha)}$; the radius grows with α bounded by $\sqrt{|\mathbf{a}|^2/4 + w}$. To second order in α, the operator $\mathscr{E}$, containing coma, appears in product with a $\mathscr{D}$, acting on the wave function. Two $\mathscr{D}$'s with an $\mathscr{E}$ appear for third order in α acting on the normal derivative, and so forth.

In Fig. 1 we show the squared first component of a forward Gaussian beam. In units of λ we have placed the center of the Gaussian at the point $\mathbf{a} = (10,0)$. We have set the width of the Gaussian to be $w = 4$, so the squared amplitude drops to $e^{-1} = 0.3679...$ of its maximum value at $|\mathbf{q} - \mathbf{a}| = 2$. The conjugate wavenumber Gaussian has width $\frac{1}{4}$ and is thus comfortably concentrated within the $|\mathbf{k}| = 1$ disk.

Figures 2 and 3 show the square of the resulting aberrated function on the screen, $|f_{w,a}^{\alpha}(\mathbf{q})|^2$ for $\alpha = 0.3$ and $\alpha = -0.3$. We have chosen these values so that the figures will be comparable with those of Ref. 1. The geometric coma caustic angle (60°) is superposed on the figures, with the

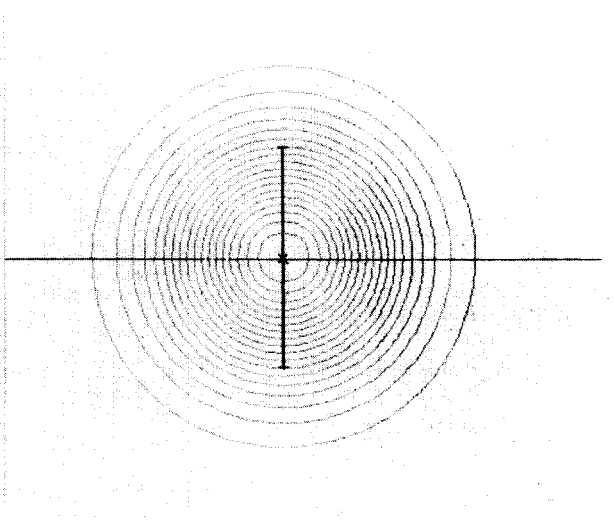


FIG. 1. Contours of the square of the amplitude of a reference Gaussian placed at $\times$. The width is $w = 4\lambda^2$ (we mark the $|\mathbf{q} - \mathbf{a}| = 2$ distance at which the function drops to $e^{-1} = 0.3679...$of its maximum). The optical center is 10λ to the left of $\times$; the vertical line stands at $x = 5\lambda$. We have used 20 "isophote" contours spaced by 0.05, from 0.0 (hence not shown in the figure) to 1.0 (marked by $\times$).

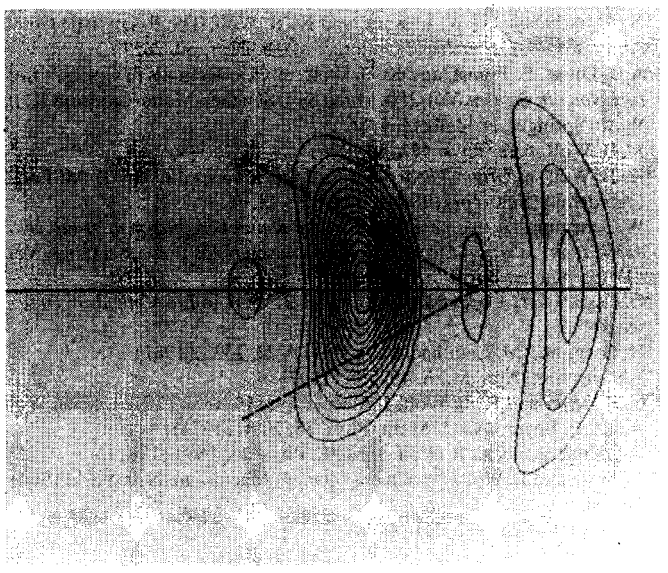


FIG. 2. The relativistic coma-aberrated forward Gaussian of Fig. 1, for $\alpha = +0.3$ ($v = c \tanh \alpha$). We indicate the apex and the opening 60°-caustic angle of the geometric Seidel coma image.

apex at $e^{\alpha}\mathbf{a}$ (13.5 and 7.408 units from the optical center for $\alpha = \pm 0.3$). The figures were drawn after evaluating the polynomial factor in the series (4.4) to fifth degree in α for the above parameter values. They show that the single Gaussian peak unfolds into several local maxima, separated by crescent-shaped "dark fringes," whose number was seen to increase with the truncation degree of α in the series (4.4). The location of the global maxima (0.804 and 0.818 of the reference Gaussian maximum) changed only slowly from first degree on in the direction of magnification. New, smaller local maxima are added with increasing degree.

We should compare these features to those calculated for diffraction in aberration under pure Seidel coma [Ref. 11, Figs. 9.6(a) and 9.6(b)]. Our Fig. 3 seems to conform better than Fig. 2 to the familiar pattern of fringes of pure coma, where crescents bend in the direction of the apex; however, for the parameters of the second figure, $\alpha = +0.3$, the first-degree term is magnification. As we saw above, this will introduce a circular dark fringe of radius nearly 4 with the center at (5,0). This fringe seems to be the dominant feature that keeps the higher-degree crescents bending toward the optical center over the coma bending of the same. We note that the crescents of Fig. 2 are slightly "stiffer" than those of Fig. 3; this may indicate that in the former, the purely comatic bending weakly counteracts the basic magnification bending.

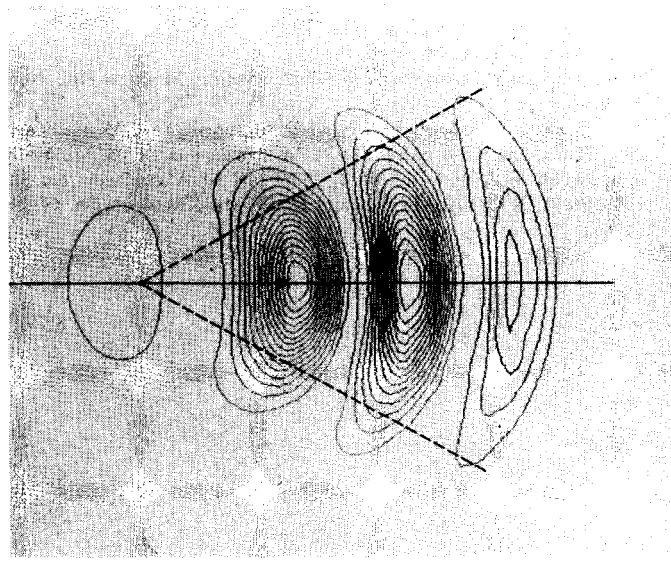


FIG. 3. The relativistic-coma-aberrated Gaussian of Fig. 1, for $\alpha = -0.3$. This is equivalent to a backward-directed Gaussian beam with positive α parameter. We indicate the geometric opening caustic angle for Seidel coma.

V. RECAPITULATION AND CONCLUDING REMARKS

The quantization of a system on the level of its dynamical group has been proposed in Refs. 23, among many others, providing self-adjoint representations on a space with a physical interpretation, such as the $\mathscr{L}^2(\mathfrak{R}^3)$ Hilbert space of quantum wave functions. The dynamical group of optics in homogeneous media is the Euclidean group. The representations we have explored are that of directed lines through a screen in geometrical optics, and that of a two-functions on a reference plane in Helmholtz optics. Both remain homogeneous spaces under the deformation of the Euclidean to the Lorentz group.

In this way, boosted screens are described on par with rotated or translated screens, and special relativity is brought in contact with geometric and Helmholtz optics, that *prima facie* had little to do with motion because they contain no time variable. Solutions of the two-dimensional Helmholtz equation have been subjected to the Lorentz group before,[24] but we did not realize then that the Euclidean group has a transparent optical interpretation. The group action is correct as far as the prediction of the familiar stellar aberration for ray directions and plane waves. The relativistic coma phenomenon is the "Fourier conjugate" of that distortion of the sphere. The quotation marks are to withold a precise definition that encompasses canonical conjugation in geometric optics, and integral transformation into the basis of the plane waves (2.7) by (2.8) in Helmholtz optics.

Even more pressing than the question of a time variable, is the absence of a space variable, $\mathbf{q}$, within the dynamical group. In geometric optics, $\mathbf{q}$ is the canonical conjugate to ray momentum, $\mathbf{p}$, within the Heisenberg–Weyl algebra under Poisson brackets. In Helmholtz wave optics, a position operator $\mathsf{Q}f(\mathbf{q}) = \mathbf{q}f(\mathbf{q})$ is not self-adjoint in $\mathscr{H}_k$, and hence does not lead to a standard observable within the framework of wave mechanics. We see this as a welcome feature of our theory, since Dirac δ's on the screen cannot be strictly produced. Sinc-type or J_0 beams[15] may be the best approximations. Here, we have an inner product (2.4) that has a Bessel-function nonlocality. Mathematically, this is Parseval equivalent to the presence of the obliquity factor k_z/k in the plane-wave basis inner product (2.10). The obliquity factor must be there for geometric reasons. Both f and f_z should be present if the squared norm is to mean total field energy of the system.

In fact, in two-dimensional optics,[25,6] where ray directions θ range over a circle S_1, the Fourier conjugate observable will generate rotations of that circle, indicating that $i\hbar d/d\theta$ may be an appropriate position operator (cf. Ref. 4, Sec. V, for the conjugate problem of quantum mechanics on S_1). The spectrum of such an operator in $\mathcal{L}^2(S_1)$ is discrete and equally spaced by $\hbar$, consistent with the sampling theorem of Whittaker and Shannon.[26] In the three-dimensional case we have a direction sphere S_2 (*not* a torus), so the identification of the traditional position operators with our rotation generators $\hbar R_x, \hbar R_y$, could be appropriate in view of the paraxial contraction limit, where they commute. Alternatively, the position observable in geometrical optics was written in Ref. 1, Eq. (4.5), as an algebraic function of the translation and boost operators, but the wave version of this relation is not obvious. The plane q of the figures, however intuitive as the screen where images form, still requires further understanding. This also applies to the role of the normal derivative f_z that does not seem to be directly observable on the screen, but could be inferred from the values of the field amplitude at two different locations near the screen.

The endowment of a physical system with a Lie algebraic structure allows the compact statement of a cornucopia of properties, such as polarization, separation of variables, classification of solutions, and transformations, Clebsch–Gordon coupling, and aberration expansions—to mention the most obvious ones for the Euclidean groups. These will be developed elsewhere for optics and relativistic oscillator mechanics,[27] going beyond the present description of scalar fields in homogeneous, empty space.

ACKNOWLEDGMENTS

We would like to thank José Fernando Barral, Instituto de Astronomía, UNAM/DF for the indispensable help with the figures in this article and the spur of interest in applications to digital image processing. A forum and a fine evening of discussions on Fourier/Helmholtz optics is acknowledged to Jorge Ojeda-Castañeda, Raúl Berriel (Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla), and José Luis Morales (ESFM, Instituto Politécnico Nacional). Two of us (N.M.A. and W.L.) would like to acknowledge the hospitality of the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), UNAM, in Cuernavaca.

[1] N. M. Atakishiyev, W. Lassner, and K. B. Wolf, The Relativistic Coma Aberration. I. Geometrical Optics, J. Math. Phys. **30**, 2457 (1989).
[2] A. J. Dragt, E. Forest, and K. B. Wolf, in "Lie Methods in Optics," *Lecture Notes in Physics*, Vol. 250, edited by J. Sánchez-Mondragón and K. B. Wolf (Springer, Heidelberg, 1986).
[3] K. B. Wolf, Ann. Phys. **172**, 1 (1986).
[4] K. B. Wolf, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. 3.
[5] W. Schempp, *Harmonic analysis on the Heisenberg nilpotent group, with applications to signal theory*, Pitman research notes in mathematics, Vol. 147 (Longman, Burnt Mill, Essex, 1986).
[6] O. Castaños, E. López-Moreno, and K. B. Wolf, in *Lie Methods in Optics*, see Ref. 2.
[7] H. Bacry and M. Cadilhac, Phys. Rev. A **23**, 2533 (1981).
[8] V Bargmann, Ann. Math. **48**, 568 (1947).
[9] C. P. Boyer and K. B. Wolf, J. Math. Phys. **14**, 1853 (1973); K. B. Wolf and C. P. Boyer, *ibid.* J. Math. Phys. **15**, 2096 (1974).
[10] S. Steinberg and K. B. Wolf, J. Math. Phys. **22**, 1660 (1981).
[11] M. Born and E. Wolf, *Principles of Optics* (Macmillan, New York, 1964), 2nd ed.
[12] K. B. Wolf, Kinam **2**, 223 (1980).
[13] N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (AMS, Providence, RI, 1968) (Russian edition: Nauka, Moscow, 1965), Chaps. 4 and 9.
[14] W. Miller, Jr., *Symmetry and Separation of Variables*, Encyclopedia of Mathematics and its Applications (Addison–Wesley, Reading, MA, 1977), Vol. 4, Chap. 3.
[15] J. Durnin, J. Opt. Soc. Am. A **4**, 651 (1987); J. Durnin, J. J. Miceli, and J. H. Eberly, Phys. Rev. Lett. **58**, 1499 (1987); cf. K. B. Wolf, Phys. Rev. Lett. **60**, 757 (1988).
[16] R. Gilmore, *Lie groups, Lie algebras, and some of their Applications* (Wiley, New York, 1974), Chap. 10; cf. Refs. 8 and 9.
[17] A. Weil, Acta Math. **11**, 143 (1963); M. Hávliček and W. Lassner, Int. J. Theor. Phys. **15**, 867 (1976).
[18] K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979).
[19] S. Steinberg and K. B. Wolf, Nuovo Cimento A **53**, 149 (1979), Sec. 5.
[20] M. Nazarathy, A. Hardy, and J. Shamir, J. Opt. Soc. Am. A **3**, 1360 (1986).
[21] M.García-Bullé, W. Lassner, and K. B. Wolf, J. Math. Phys. **27**, 29 (1986), Sec. V.
[22] J. Ojeda-Castañeda and A. Boivin, Can. J. Phys. **63**, 250 (1985).
[23] J. M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970); B. Kostant, "Quantization and Unitary Representations," *Lecture Notes in Mathematics*, Vol. 170 (Springer, New York, 1970); J. Śniatycki, *Geometric Quantization and Quantum Mechanics* (Springer, New York, 1980); V. Aldaya and J. A. de Azcárraga, J. Math. Phys.**23**, 1297 (1982).
[24] F. Soto, "La Ecuación de Helmholtz y el Grupo Tridimensional de Lorentz," thesis, Fac. Ciencias, UNAM, 1977.
[25] K. B. Wolf, Kinam **6**, 141 (1985).
[26] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
[27] A. D. Donkov, V. G. Kadyshevsky, M. D. Matveev, and R. M. Mir-Kasimov, Teor. Mat. Fiz. **8**, 61 (1971); N. M. Atakishiyev, R. M. Mir-Kasimov, and Sh. M. Nagiev, *ibid.* **44**, 47 (1980); N. M. Atakishiyev, *ibid.* **58**, 254 (1984); N. M. Atakishiyev and R. M. Mir-Kasimov, *ibid.* **67**, 68 (1986).

Branching rules for the hyperoctahedral group

J. P. Doeraene
Aspirant F.N.R.S., Institut de Mathématiques, Louvain La Neuve, Belgium
G. Iommi Amunategui
Instituto de Física, Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

(Received 19 May 1988; accepted for publication 31 May 1989)

Reduction theorems for the decomposition of induced and irreducible characters of $W(B_n)$ in terms of induced and irreducible characters of $W(B_{n-1})$, respectively, are given.

I. INTRODUCTION

To each classical Lie group corresponds a finite group generated by the reflections of its root system, called the Weyl group. For $B_n = \mathrm{SO}(2n+1)$ and $C_n = \mathrm{Sp}(2n)$, the Weyl groups $W(B_n)$ and $W(C_n)$ are isomorphic. $W(B_n) = Z_2^n \otimes S_n$ is the semidirect product of the Abelian group Z_2^n generated by the n sign changes $(+i, -i)$, $1 \leqslant i \leqslant n$, and the symmetric group S_n. The order of $W(B_n)$ is $2^n n!$ (Refs. 1, 2). Moreover, let K_n be defined as the convex hull of points $\pm e_i$, $1 \leqslant i \leqslant n$, where $e_1, \ldots, e_n$ are the unit coordinate vectors in $\mathbb{R}^n$. It is the n-dimensional generalization of the octahedron K_3. The group of symmetries of K_n called the hyperoctahedral group is $W(B_n)$.

The structure and representations of this group have been studied (Refs. 3–6), and applications to physical problems have been considered, especially on lattices (for instance: discrete σ models, lattice gauge theories, and chiral models; see Refs. 7, 8). The purpose of this paper is to solve the reduction problem, i.e., to decompose the character of a representation of $W(B_n)$ into characters of $W(B_{n-1})$ representations. We shall consider characters of induced or permutation representations as well as of irreducible representations.

Our starting point will be an algorithm giving the induced characters of $W(B_n)$. In order to make the article reasonably self-contained, this algorithm and some other results, previously published elsewhere, will be exposed anew (Refs. 9, 10).

This work is organized as follows. In Sec. II, the group $W(B_n) = Z_2^n \otimes S_n$ and the algorithm for the characters of $W(B_n)$ induced in it by the one-dimensional representations of the so-called canonical subgroups are treated. In Sec. III A we establish Theorem 1 dealing with the induced characters; in Sec. III B, Theorem 2 corresponding to the reduction of irreducible characters is given, and Sec. III C is concerned with a discussion of the dimensions of the induced and irreducible representations of $W(B_n)$.

II. THE INDUCED CHARACTERS OF $W(B_n)$

The set of all $g = (\sigma; f)$, where $\sigma \in S_n$ and f is a mapping of $[1,n] \subset \mathbb{N}$ into Z_2, together with the composition defined by

$$(\sigma'; f')(\sigma; f) = (\sigma'\sigma; f'(f\sigma'^{-1}))$$

form the group $W(B_n) = Z_2^n \otimes S_n$.

The cycles of the permutation are called "cycles of g." A cycle $(a_1, \ldots, a_\beta)$ of g is positive or negative if $f(a_1)\cdots f(a_\beta) = +1$ or -1.

Let $\beta = (\beta_1, \ldots, \beta_k)$ be the β system of cycles of σ, and suppose the cycles are arranged in such a way that a negative cycle necessarily precedes a positive cycle of equal length. Then (β, b) is called the β system of cycles of g, where $b := (b_1, \ldots, b_k)$ with $b_i := 1$ or 0 if the ith cycle is positive or negative (remark: if $\beta_i = \beta_{i+1}$, then $b_i \leqslant b_{i+1}$). Moreover if α_i^+ and α_i^- denote the number of positive and negative cycles, respectively, of length i of g, then

$$\alpha = (\alpha_1^+, \alpha_1^-, \alpha_2^+, \alpha_2^-, \ldots, \alpha_l^+, \alpha_l^-)$$

is called the α system of cycles of g (remark: if $\alpha_i := \alpha_i^+ + \alpha_i^-$, then $\sum_i i\alpha_i = n$).

The elements of $W(B_n)$ are conjugated iff they have the same α system of cycles and iff they have the same β system of cycles. The class of elements with α system $\alpha = (\alpha_1^+, \ldots, \alpha_l^-)$ is denoted $C(\alpha)$.

Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be a partition of n $(\lambda_1 \geq \lambda_2 \cdots \geq \lambda_k)$ and $b = (b_1, \ldots, b_k)$ be such that $b_i = 1$ or 0 (remark: if $\lambda_i = \lambda_{i+1}$, then $b_i \leqslant b_{i+1}$).

The subgroup $(Z_2^{(\lambda_1 - b_1)} \otimes S_{\lambda_1}) \times (Z_2^{(\lambda_2 - b_2)} \otimes S_{\lambda_2}) \ldots$, denoted by $S(\lambda, b)$, is a canonical subgroup of $W(B_n)$. Then, for the class $C(\alpha)$ and the canonical subgroup $S(\lambda, b)$ the algorithm giving the character $\phi_{S(\lambda,b)}(C(\alpha))$ of the representation of $W(B_n)$ induced by the identity representation of $S(\lambda, b)$ is

$$\phi_{S(\lambda,b)}(C(\alpha)) = 2^{(\Sigma_i bi)} \left(\sum \frac{\Pi_{i=1}^{l} (\alpha_i^+)!(\alpha_i^-)!}{\Pi_{i=1}^{l} \Pi_{j=1}^{k} (\alpha_{ij}^+)!(\alpha_{ij}^-)!} \right).$$

The sum concerns the matrices $(\alpha_{ij}^{+/-})$ of dim $l \times k \times 2$ where

$$\forall i_0, \sum_{j=1}^{k} \alpha_{i_0 j}^+ = \alpha_{i_0}^+$$

and

$$\sum_{j=1}^{k} \alpha_{i_0 j}^- = \alpha_{i_0}^-,$$

$$\forall j_0, \sum_{i=1}^{l} i(\alpha_{i j_0}^+ + \alpha_{i j_0}^-) = \lambda_{j_0}.$$

 0022-2488/89/112469-07$02.50

Besides, $\forall j_0$, if $b_{j_0}=1$, then $\sum_i \alpha_{ij_0}^-$ is an even number.

The order of the class $C(\alpha)$ is

$$|C(\alpha)| = n! \prod_{i=1}^{l} \left(\frac{2^{\alpha_i(i-1)}}{i^{\alpha_i}(\alpha_i^+)!(\alpha_i^-)!} \right)$$

(for details, see Appendix A). With this algorithm, the induced character table $\phi\{W(B_n)\}$ is obtained. Each row of the table is given by the corresponding $\phi_{S(\lambda,b)}(C(\alpha))$.

For instance, the table of the induced characters of $W(B_2)$, the Weyl group of SO(5), is

$\phi\{W(B_2)\} =$

	$\pm\square$	$\pm\square$	$\square\!-\!\square$	$\square\!-\!\square$	classes
	1	2	1	2	2	order
$-\square$	1	1	1	1	1	
$+\square$	2	0	2	2	0	
$\square\!-\!\square$	2	2	2	0	0	
$\pm\square$	4	2	0	0	0	
$\pm\square$	8	0	0	0	0	

.

III. THE REDUCTION $W(B_n) \to W(B_{n-1})$

A. The induced characters

To state the reduction theorem, we denote by $(i^+)^{a_i}, a_i$ subpartitions of n, of length i, and of sign $+$, by $(i^-)^{a'_i}, a'_i$ subpartitions of n of length i, and sign $-$, $\phi\{(1^+)^{a_1}(1^-)^{a'_1}(2^+)^{a_2}(2^-)^{a'_2}\cdots\}$ represents an induced character of $W(B_n)$.

Theorem 1: The induced character of $W(B_n)$, $\phi\{(1^+)^{a_1}(1^-)^{a'_1}(2^+)^{a_2}(2^-)^{a'_2}\cdots\}$ reduces into $W(B_{n-1})$ induced characters as follows:

$$\begin{aligned}
&\phi\{(1^+)^{a_1}(1^-)^{a'_1}\cdots\} \\
&\quad = 2a_1\phi\{(1^+)^{a_1-1}(1^-)^{a'_1}\cdots\} \\
&\qquad + a'_1\phi\{(1^+)^{a_1}(1^-)^{a'_1-1}\cdots\} \\
&\qquad + a_2\phi\{(1^+)^{a_1+1}(1^-)^{a'_1}(2^+)^{a_2-1}\cdots\} \\
&\qquad + a'_2\phi\{(1^+)^{a_1}(1^-)^{a'_1+1}(2^+)^{a_2}(2^-)^{a'_2-1}\cdots\} \\
&\qquad + \cdots .
\end{aligned}$$

Proof: (i) Let us take an element of $W(B_{n-1})$ belonging to the class $(\alpha_1^+ - 1, \alpha_1^-, \alpha_2^+, \alpha_2^-, \ldots)$ (hereafter only classes with $\alpha_1^+ \neq 0$ will be considered).

From $\forall i_0$,

$$\sum_{j=1}^{k} \alpha_{i_0 j}^+ = \alpha_{i_0}^+,$$

i.e.,

$$\alpha_1^+ = \alpha_{11}^+ + \alpha_{12}^+ + \cdots \alpha_{1k}^+,$$

it is clear that there are k ways to subtract a unity from α_1^+: $(\alpha_{11}^+ - 1) + \alpha_{12}^+ + \cdots + \alpha_{1k}^+$,

$$\begin{aligned}
&\alpha_{11}^+ + (\alpha_{12}^+ - 1) + \cdots + \alpha_{1k}^+, \\
&\quad \vdots \\
&\alpha_{11}^+ + \alpha_{12}^+ + \cdots + (\alpha_{1k}^+ - 1),
\end{aligned}$$

and from

$$\lambda_{j_0} = \sum_{i=1}^{l} i(\alpha_{ij_0}^+ + \alpha_{ij_0}^-)$$

we get

$$\begin{aligned}
\lambda_1 - 1 &= (\alpha_{11}^+ - 1 + \alpha_{11}^-) + \cdots \\
\lambda_2 - 1 &= (\alpha_{12}^+ - 1 + \alpha_{12}^-) + \cdots \\
&\vdots \\
\lambda_k - 1 &= (\alpha_{1k}^+ - 1 + \alpha_{1k}^-) + \cdots .
\end{aligned}$$

Hence, the induced character of $W(B_n)$, $\phi(\lambda_1 \cdots \lambda_k)$ reduces into the sum of induced characters of $W(B_{n-1})$:

$$\begin{aligned}
&\phi(\lambda_1 - 1,\ldots,\lambda_k) + \phi(\lambda_1,\lambda_2 - 1,\ldots,\lambda_k) \\
&\quad + \cdots + \phi(\lambda_1,\lambda_2,\ldots,\lambda_k - 1).
\end{aligned}$$

(ii) Besides, if a subpartition λ_l appears p times, (i) yields

$$\begin{aligned}
\lambda_{l_1} - 1 &= (\alpha_{1l_1}^+ - 1 + \alpha_{1l_1}^-) + \cdots, \\
\lambda_{l_2} - 1 &= (\alpha_{1l_2}^+ - 1 + \alpha_{1l_2}^-) + \cdots, \\
&\vdots \\
\lambda_{l_p} - 1 &= (\alpha_{1l_p}^+ - 1 + \alpha_{1l_p}^-) + \cdots,
\end{aligned}$$

and the corresponding induced character appears p times in the decomposition, i.e.,

$$p\phi(\ldots,\lambda_{l-1} + 1,\lambda_l - 1,\lambda_{l+1},\ldots).$$

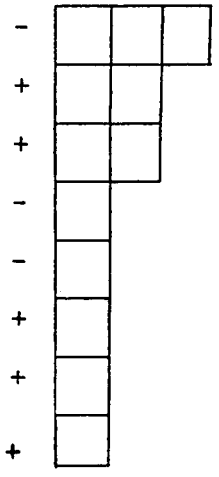
(iii) Through the reduction process, the factors $2^{\Sigma_i b_i}$ do not change, that is,

$$\begin{aligned}
&(2^{\Sigma b_i} \text{ for the induced character of } W(B_n)) \\
&\quad = \begin{pmatrix} 2^{\Sigma b_i} \text{ for each of the induced} \\ \text{characters of } W(B_{n-1}) \\ \text{resulting from the reduction} \end{pmatrix}.
\end{aligned}$$

It is to be noted that the reduction of the subpartitions $(1^+)^{a_1}$ implies the diminution of one term $b_i = 1$ in the exponent. This is not the case for the subpartitions > 1 with $b_i = 1$: (2^+), (3^+),... since subpartitions (1^+), (2^+), ..., respectively, occur and the number of $b_i = 1$ terms do not diminish. Consequently, to keep the $2^{\Sigma b_i}$ constant, the coefficient of the induced character issued from the reduction of the subpartitions $(1^+)^{a_1}$ must be multiplied by 2.

Note that (i), (ii), and (iii) imply the statement.

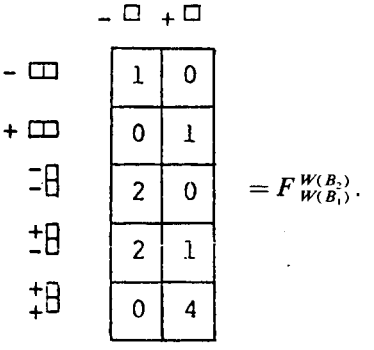
Example: Consider the representation of $W(B_{12})$ given by



Then, the branching rule for $W(B_{12}) \to W(B_{11})$ gives

$$\begin{aligned}\phi\{(1^+)^3(1^-)^2(2^+)^2(3^-)\} &= \phi\{(1^+)^3(1^-)^2(2^+)^2(2^-)\} \\ &\quad + 2\phi\{(1^+)^4(1^-)^2(2^+)(3^-)\} \\ &\quad + 2\phi\{(1^+)^3(1^-)(2^+)^2(3^-)\} \\ &\quad + 6\phi\{(1^+)^2(1^-)^2(2^+)^2(3^-)\}.\end{aligned}$$

This algorithm allows us to construct a reduction matrix $F^{W(B_n)}_{W(B_{n-1})}$; each row of this matrix is given by the corresponding reduction (in Ref. 10, we introduce the matrices F corresponding to the reduction of the induced characters of the symmetric group S_n). To illustrate the construction we apply the algorithm to $W(B_2) \to W(B_1)$:



$$= F^{W(B_2)}_{W(B_1)}.$$

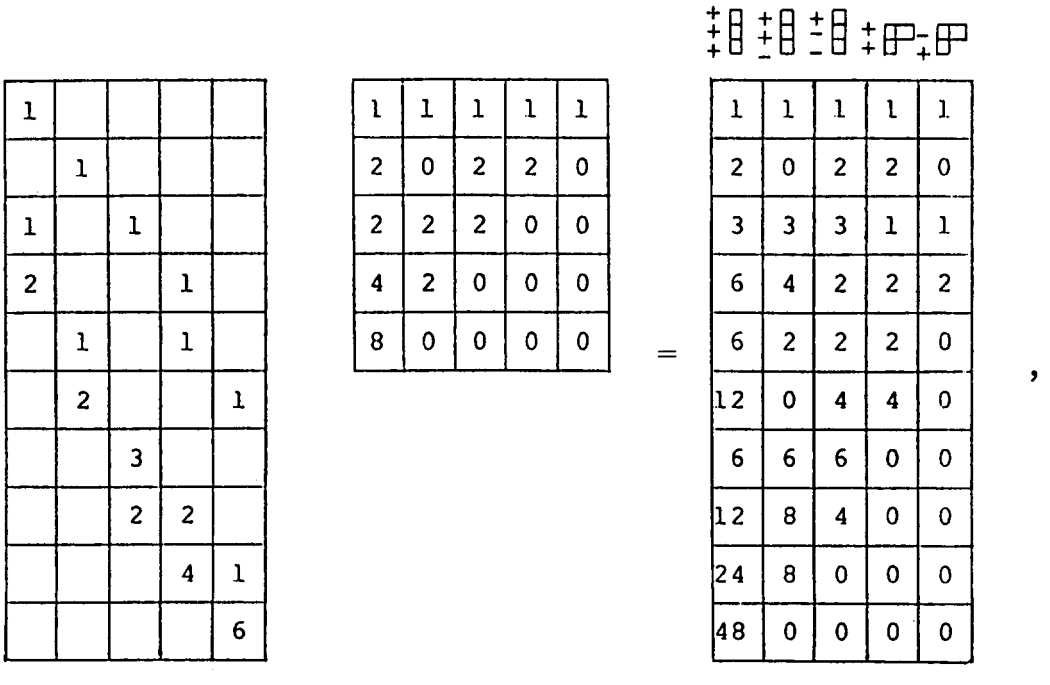
Hence, in general, the reduction of the induced characters may be written as

$$\phi\{W(B_n)\} = F^{W(B_n)}_{W(B_{n-1})}\phi\{W(B_{n-1})\}.$$

It must be noted that this result concerns only classes with $\alpha_1^+ \neq 0$; consequently the induced character table $\phi\{W(B_n)\}$ does not contain the other classes. To clarify this point, let us consider

$$\phi\{W(B_3)\} = F^{W(B_3)}_{W(B_2)}\,\phi\{W(B_2)\}.$$

We have



$$F^{W(B_3)}_{W(B_2)}\phi\{W(B_2)\} = \phi\{W(B_3)\}.$$

In $\phi\{W(B_3)\}$ the characters corresponding to the classes $\begin{smallmatrix}-\\-\end{smallmatrix}\square$, $\begin{smallmatrix}+\\-\end{smallmatrix}\square$, $\begin{smallmatrix}-\\-\end{smallmatrix}\square$, etc. (i.e., classes with $\alpha_1^+ = 0$) do not appear.

B. The irreducible (or simple) characters

We recall here some results (see Appendices B, C, D) which allow us to establish the reduction of the irreducible characters using the reduction of the induced characters.

First of all, it must be remembered that the table of irreducible characters $X\{W(B_n)\}$ can be obtained from $\phi\{W(B_n)\}$. To carry out this calculation, each row ϕ_i of $\phi\{W(B_n)\}$ must be considered as a vector; it suffices to orthonormalize them via the Gram–Schmidt procedure to get the rows X_i of the table $X\{W(B_n)\}$. In general, it is shown that

$$X_i = \phi_i - \sum_{k=1}^{i-1} (\phi_i K X_k) X_k \,(\text{for } i = 1, X_1 = \phi_1), \quad (3.1)$$

where X_i and ϕ_i are the ith rows of $X\{W(B_n)\}$ and $\phi\{W(B_n)\}$, respectively, and K is a diagonal matrix whose elements are

$$(K_{\alpha\beta}) = \delta_{\alpha\beta}(|C(\alpha)|/2^n n!), \quad |C(\alpha)|$$

is the order of class $C(\alpha)$ of $W(B_n)$.

As a by-product of this method, we obtain the triangular matrix Δ which plays an important role in the sequel. We are going to illustrate these results—up to this point—through the $W(B_2)$ case.

Expression (3.1) yields

$$\begin{aligned} X_1 &= \phi_1, \\ X_2 &= \phi_2 - X_1, \\ X_3 &= \phi_3 - X_1 - 0\cdot X_2, \\ X_4 &= \phi_4 - X_1 - 0\cdot X_2 - X_3, \\ X_5 &= \phi_5 - X_1 - X_2 - X_3 - 2X_4 \end{aligned}$$

or

$$\begin{aligned} \phi_1 &= X_1, \\ \phi_2 &= X_1 + X_2, \\ \phi_3 &= X_1 + 0 + X_3, \\ \phi_4 &= X_1 + 0 + X_3 + X_4, \\ \phi_5 &= X_1 + X_2 + X_3 + 2X_4 + X_5. \end{aligned} \quad (3.2)$$

Considering the coefficients of the X_i, we may write

$$\Delta\{W(B_2)\} = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 0 & 1 & & \\ 1 & 0 & 1 & 1 & \\ 1 & 1 & 1 & 2 & 1 \end{pmatrix}.$$

We remark that Δ is nonsingular (in fact $\forall n$, $\det \Delta = 1$; see Appendices B, C).

With the aid of Δ (3.2) becomes

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & 2 & 2 & \\ 2 & 2 & 2 & & \\ 4 & 2 & & & \\ 8 & & & & \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 0 & 1 & & \\ 1 & 0 & 1 & 1 & \\ 1 & 1 & 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 \\ 2 & 0 & -2 & 0 & 0 \\ 1 & -1 & 1 & -1 & 1 \end{pmatrix},$$

i.e., $\phi\{W(B_2)\} = \Delta X\{W(B_2)\}$. In general we have

$$\phi\{W(B_n)\} = \Delta\{W(B_n)\} X\{W(B_n)\}.$$

We may now state the reduction of the irreducible characters.

Theorem 2: The irreducible characters X_n of $W(B_n)$ reduces into irreducible characters X_{n-1} of $W(B_{n-1})$ according to the equation

$$X_n = W_{n-1}^n X_{n-1},$$

where

$$W_{n-1}^n = \Delta_n^{-1} F_{n-1}^n \Delta_{n-1}.$$

Proof: For $W(B_n)$, $\phi_n = \Delta_n X_n$ (i) and for $W(B_{n-1})$, $\phi_{n-1} = \Delta_{n-1} X_{n-1}$ (ii). Besides, from Theorem 1, $\phi_n = F_{n-1}^n \phi_{n-1}$ (iii). Putting (i) and (ii) in (iii),

$$\Delta_n X_n = F_{n-1}^n \Delta_{n-1} X_{n-1},$$

$$\therefore X_n = \Delta_n^{-1} F_{n-1}^n \Delta_{n-1} X_{n-1}.$$

Example: $W(B_2) \to W(B_1)$:

1				
-1	1			
-1	0	1		
0	0	-1	1	
1	-1	1	-2	1

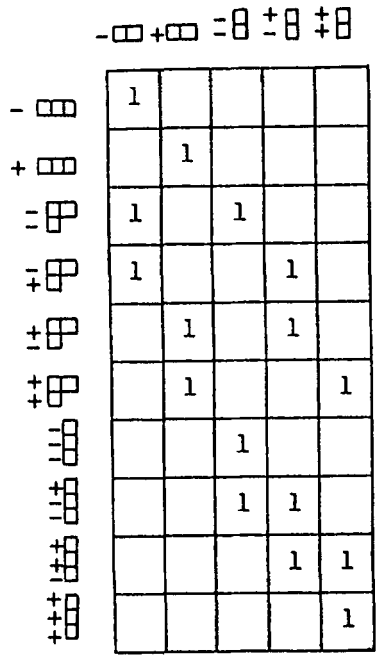
1	0
0	1
2	0
2	1
0	4

1	0
1	1

$= W_1^2 =$

	$-\square$	$+\square$
$-\square\square$	1	0
$+\square\square$	0	1
$\begin{smallmatrix}-\\-\end{smallmatrix}\square$	1	0
$\begin{smallmatrix}+\\-\end{smallmatrix}\square$	1	1
$\begin{smallmatrix}+\\+\end{smallmatrix}\square$	0	1

Remark: This reduction is partially equivalent to the Weyl branching rule for S_n, that is, the allowed diagrams are obtained from the original one by removing a square in turn from the end of each row which is longer than the following one. The peculiarity of the $W(B_n)$ case stems from the signs $+$ and $-$ assigned to the patterns. As an example, we give here the reduction matrix for $W(B_3) \to W(B_2)$:



For all n, the matrices F^n_{n-1} and W^n_{n-1} have the same entries $\neq 0$. For F^n_{n-1} the multiplicities of these entries are ≥ 1 and for W^n_{n-1}, equal to 1.

C. On the dimensions of the representations of $W(B_n)$

The dimension of a representation, irreducible as well as induced, is nothing but the number of times that a one-dimensional representation is contained in it. As a matter of fact, the reduction matrices W and F provide a method to evaluate the dimensions of the irreducible and induced representations: for a group $W(B_n)$, these dimensions are given via the products

$$W^{W(B_n)}_{W(B_{n-1})}\, W^{W(B_{n-1})}_{W(B_{n-2})} \cdots W^{S_2}_{S_1}$$

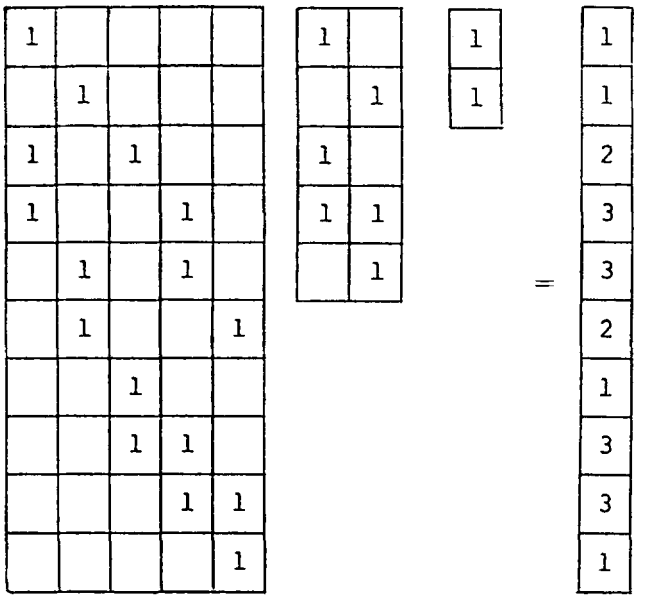
and

$$F^{W(B_n)}_{W(B_{n-1})}\, F^{W(B_{n-1})}_{W(B_{n-2})} \cdots F^{S_2}_{S_1},$$

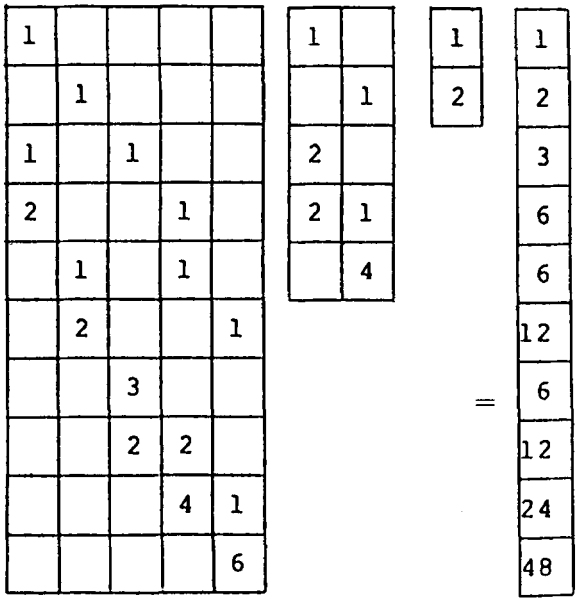
respectively.

The last term of these sequences concerns the reduction of the symmetric groups $S_2 \to S_1$. This is due to the structure of $W(B_1) = Z_2 \otimes S_1$; $W(B_1)$ is isomorphic to S_2, and to obtain the dimensions, the reduction must be continued until S_1 (dimension 1). (For details concerning the symmetric group reduction see Ref. 10.)

Example: (a) Dimensions of the irreducible representations of $W(B_3) = W^{W(B_3)}_{W(B_2)}\, W^{W(B_2)}_{W(B_1)}\, W^{S_2}_{S_1}$:



(b) Dimensions of the induced representations of $W(B_3) = F^{W(B_3)}_{W(B_2)}\, F^{W(B_2)}_{W(B_1)}\, F^{S_2}_{S_1}$:



APPENDIX A: THE ALGORITHM FOR $\phi_{S(\lambda,b)}(C(\alpha))$

We give here the main steps required to obtain the algorithm for the induced characters $\phi_{S(\lambda,b)}(C(\alpha))$.

From the general theory of characters, we have that if H is a subgroup of a finite group G, g an element of G, $C(g)$ the conjugacy class of g in G, the induced character $\phi^G_H(g)$ is given by

$$\phi^G_H(g) = \frac{|G|}{|H|} \frac{|C(g) \cap H|}{|C(g)|}. \tag{A1}$$

In the present case,

$$G = Z_2^n \wr S_n, \quad |G| = 2^n n! \tag{A2}$$

$$H = S(\lambda,b) = (Z_2^{(\lambda_1 - b_1)} \wr S_{\lambda_1}) \times \cdots \times (Z_2^{(\lambda_k - b_k)} \wr S_{\lambda_k})$$

and

$$|H| = 2^{\lambda_1 - b_1}\lambda_1! \cdots 2^{\lambda_k - b_k}\lambda_k!,$$

i.e.,

$$|H| = (2^{\lambda_1 + \cdots + \lambda_k}/2^{b_1 + \cdots + b_k})\lambda_1! \cdots \lambda_k!,$$

but

$$\lambda_1 + \cdots + \lambda_k = n,$$

hence

$$|H| = \frac{2^n}{2^{\Sigma_i b_i}} \lambda_1! \cdots \lambda_k!. \tag{A3}$$

The order of the class $C(\alpha)$ is

$$|C(\alpha)| = n! \prod_{i=1}^{l} \frac{2^{\alpha_i(i-1)}}{i^{\alpha_i}(\alpha_i^+)!(\alpha_i^-)!}. \tag{A4}$$

To evaluate $|S(\lambda,b) \cap C(\alpha)|$ it suffices to replace in (A4) $n!$ by $\lambda_1! \cdots \lambda_k!$ and $\Pi_{i=1}^{l} (\alpha_i^+)!(\alpha_i^-)!$ by $\Pi_{i=1}^{l} \Pi_{j=1}^{k} (\alpha_{ij}^+)!(\alpha^-{}_{ij})!$.
Putting (A2), (A3), (A4), and $|S(\ ,b) \cap C(c)|$ in (A1), the expression given in Sec. II for $\phi_{S(\lambda,b)}(C(\alpha))$ is obtained.

APPENDIX B: SOME PROPERTIES OF THE CHARACTERS OF A FINITE GROUP

In this Appendix we state some basic results concerning the characters of a finite group which are implicit in some subsequent derivations.

(1) The orthogonality relations,

$$XKX^T = E, \quad X^T X = K^{-1},$$

where X is the character table of a finite group G, K is the matrix $(K_{uv}) = \delta_{uv}(|C_u|/|G|)$, $|C_u|$ is the order of the conjugacy class of u, $|G|$ is the order of G, and E is the identity matrix.

(2) Every representation D is the direct sum of irreducible representation D_i, over $\mathbb{C}$:

$$D = \oplus n_i D_i, \quad n_i = \text{non-negative integers.}$$

(3) Every character (ϕ) can be expressed via the $\mathbb{Z}$-basis of the irreducible characters over $\mathbb{C}$:

$$(\phi) = \sum n_i \chi_i, \quad n_i \in \mathbb{N}.$$

Besides $n_i = \langle (\phi), \chi_i \rangle$, where, by definition,

$$\langle x,y \rangle = xKy^T.$$

In Ref. 2 (pages 38,39) it is shown that:

(4) For the symmetric group S_n, the multiplicities n_i can be written in matrix form and this matrix, Δ, is lower triangular with ones along the main diagonal, i.e., unitriangular (we remark that James and Kerber use a different notation and a different label for the matrix entries, so it turns out that for them Δ is upper triangular). Hence $\det \Delta = 1$ and Δ^{-1} is a matrix over $\mathbb{Z}$.

(5) The matrix of induced characters ϕ may be written as $\phi = \Delta X$ [see (3) above], and the matrix of irreducible characters is given by $X = \Delta^{-1}\phi$ (see Ref. 2, 2.2.9, p. 39).

APPENDIX C: THE MATRIX Δ FOR $Z_2^n \wr S_n$

This Appendix contains a characterization of the Δ matrix for $Z_2^n \wr S_n$. To establish the properties of the Δ matrix for S_n, the basic tool is the theorem of Ruch and Schönhofer (Ref. 2, p. 27); moreover, two orders between partitions λ and λ' of n are required:

(1) The lexicographic order $\leqslant$:

$$\lambda \leqslant \lambda' :\leftrightarrow \exists i(\lambda_1 = \lambda'_1, \ldots, \lambda_{i-1} = \lambda'_{i-1}, \lambda_i < \lambda'_i),$$

$\leqslant$ is a total order.

(2) The dominance order $\trianglelefteq$: This order is defined in terms of partial sums $\Sigma_1^i \gamma_\nu$ of the parts of the partition in question:

$$\lambda \trianglelefteq \lambda' :\leftrightarrow \forall i \left(\sum_1^i \lambda_\nu \leqslant \sum_1^i \lambda'_\nu \right),$$

$\trianglelefteq$ is a partial order (note that $\lambda \trianglelefteq \lambda' \rightarrow \lambda \leqslant \lambda'$). These two orders differ only if $n \geqslant 6$. For $Z_2^n \wr S_n$ we must introduce a new condition into the definition of the dominance order, which takes into account $+1$ and -1 (denoted by ϵ_ν). Hence, for $Z_2^n \wr S_n$,

$$\lambda \trianglelefteq \lambda' :\leftrightarrow \begin{cases} \forall i, \sum_1^i \lambda_i \leqslant \sum_1^i \lambda'_i \\ (\forall i, \lambda_i = \lambda'_i) \rightarrow \left(\forall i, \sum_1^i \epsilon_\nu^\lambda \geqslant \sum_1^i \epsilon_\nu^{\lambda'} \right) \end{cases}.$$

Moreover, the entries of Δ are labelled:

(i) By partitions arranged in lexicographic order (λ for the rows; λ' for the columns).

(ii) By the combination of the Z_2^n elements $(+1, -1)$ and the Young diagrams of the partitions. The resulting arrangements must be ordered from the sign $-$ to the sign $+$ (top to bottom, left to right).

For example, $\mp$  precedes $\pm$ (the number of nodes — of the first diagram is greater than the number of nodes — of the second one).

With these modifications, the line of argument displayed in Ref. 2 for the Δ matrix of S_n can be followed; if λ is the row label and λ' the column label we have: the entires of Δ are different from 0 if and only if $\lambda \trianglelefteq \lambda'$. When $\lambda \geqslant \lambda'$, the entries are 0.

The diagonal terms correspond to the same partition and to the same arrangement, hence these diagonal elements are ones (consequence of the theorem of Ruch and Schönhofer).

APPENDIX D: A PROOF OF (3.1)

In this Appendix we outline the proof of result (3.1):

"The characters of the irreducible representations of $Z_2^n \otimes S_n$, X_i, are given by

$$X_i = \phi_i - \sum_{k=1}^{i-1} (\phi_i K X_k) X_k$$

(for $i = 1$, $X_1 = \phi_1$)."

Proof: (i) Let us consider the columns of the character table of $Z_2^n \otimes S_n$ as orthogonal vectors

(i.e., $X^T X = K^{-1}$).

(ii) The matrix Δ is lower unitriangular as well as Δ^{-1}. The equation $X = \Delta^{-1}\phi$ may be written

$$\begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \begin{pmatrix} (\phi_1) \\ (\phi_2) \\ \vdots \end{pmatrix} = X$$

(note that the nondiagonal entries of Δ^{-1} are not considered).

By recurrence: for $i = 1$ $(\phi_1) = (X_1)$. At the ith place of the ith row of Δ^{-1} there is a coefficient equal to 1 (on the diagonal) if the other coefficients are $a_1, a_2, \ldots, a_{i-1}$. We have

$$X_i = a_1(\phi_1) + a_2(\phi_2) + \cdots + a_{i-1}(\phi_{i-1}) + (\phi_i).$$

Reciprocally:

(iii) $X_i = b_1(X_1) + b_2(X_2) + \cdots + b_{i-1}(X_{i-1}) + (\phi_i)$ but the (X_i) are orthogonal. For (iii), we have

$$\forall k,\ (X_i)K(X_k) = b_1(X_1)K(X_k) + \cdots + b_{i-1}(X_{i-1})K(X_k) + (\phi_i)K(X_k),$$

$$\forall k < i,\ 0 = b_k + (\phi_i)K(X_k),$$

then

$$b_k = -(\phi_i)K(X_k).$$

Replacing this result in (iii), we obtain

$$X_i = \phi_i - \sum_{k=1}^{i-1} (\phi_i K X_k) X_k.$$

ACKNOWLEDGMENTS

The authors are grateful to D. Speiser for several helpful discussions. This work was supported in part by FONDECYT (Contract No. 650-88).

[1] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972).
[2] G. James and A. Kerber, *The Representation Theory of the Symmetric Group* (Addison-Wesley, Reading, MA, 1981).
[3] R. P. Stanley, J. Combin. Theory A **32**, 132 (1982).
[4] A. Kerber, "Representations of permutations groups II," *Lecture Notes in Mathematics*, Vol. 495 (Springer, Berlin, 1975).
[5] L. Geissinger and D. Kinch, J. Algebra **53**, 1 (1978).
[6] J. P. Doeraene, "Caracteres de certains groupes finis," Rapport 80 Institut de Math. U. C. L., Cabay, Louvain La Neuve, Belgium, 1986.
[7] J. E. Mandula, G. Zweig, and J. Govaerts, Nucl. Phys. B **228**, 109 (1983).
[8] M. Baake, J. Math, Phys. **25**, 3171 (1984).
[9] G. Iommi Amunategui and D. Speiser, Ann. Soc. Sci. Bruxelles **98**, 161 (1984).
[10] G. Iommi Amunategui, Ann. Soc. Sci. Bruxelles **100**, 115 (1986).

The complete classification of generalized homogeneous symplectic manifolds

D. R. Grigore and O. T. Popp
Department of Theoretical Physics, Central Institute of Physics, P.O.B. MG-6, Bucharest-Magurele, Romania

(Received 25 May 1988; accepted for publication 31 May 1989)

The classification scheme for homogeneous symplectic manifolds is completed and generalized including the case when the symmetries are symplectic only up to a factor and the group is nonconnected. This improved classification also describes the possible coverings in terms of classes of some discrete subgroups. As an application, the possibility of mixed symplectic and antisymplectic symmetries is studied for some physical groups: rotations in three dimensions, the Poincaré group and the Galilei group (including inversions). Some new possibilities (with compactified energy), not appearing previously in the literature, have been found even for the connected Galilei group.

I. INTRODUCTION

The classification of transitive G-symplectic actions for connected Lie groups, useful for a classical counterpart of an elementary system, has been done by Kostant,[1] Souriau,[2] and Kirillov[3] in the hypothesis of existence of the momentum map. It was observed[4,5] that these actions can be classified without this hypothesis, using a different technique based on the reduction principle.[6] This shows that the homogeneous G-symplectic spaces are classified up to covering by some "regular" orbits of the natural action of the group in $Z^2(\text{Lie G}, \mathbb{R})$ (the $\mathbb{R}$-valued two-cocycles of the Lie algebra of G).

In this paper we prove, by refining this analysis, that the transitive smyplectic actions of a Lie group (not necessarily connected) can be described in a complete way, not only up to covering. Moreover we also treat the case when the symmetries are symplectic up to a factor. For connected Lie groups this was studied in Ref. 7 using the reduction principle. Also, we give a classification of Hamiltonian G spaces, not only up to covering, but including symmetries up to a "discrete" factor. This refines a well known analysis.[1-3] As an interesting application we consider the case of mixed symplectic or antisymplectic[2] symmetries in analogy with the similar problem encountered in quantum mechanics (unitary or antiunitary symmetries[8-10]). These are done in Sec. II.

As applications we consider some important nonconnected Lie groups, namely, the orthogonal group in three dimensions, the Poincaré group, and the Galilei group (including inversions). The first two cases can be treated using a simplified form of the main theorem and this is done in Sec. III. The analysis for the Galilei group in more involved; it is the subject of Sec. IV. Some new transitive actions appear. These G-symplectic manifolds are covered by already known coadjoint orbits (corresponding to zero mass); they do not admit a momentum map and have a "compactified energy." It would be interesting to see if they have a reasonable physical interpretation.

II. THE CLASSIFICATION THEOREM

Let (M_i,Ω_i) $(i = 1,2)$ by symplectic manifolds. A diffeomorphism $\phi: M_1 \to M_2$ is called *symplectic up to a factor* if, for a $\chi \in \mathbb{R}^* = \mathbb{R} - \{0\}$,

$$\phi^*\Omega_2 = \chi\Omega_1, \tag{2.1}$$

If $\chi = 1$, ϕ is called *symplectic*, and, if $\chi = -1$, *antisymplectic*. Two symplectic manifolds are called *symplectomorphic* if there exists a symplectic map between them. They are of the same *symplectic type* if the map is symplectic up to a factor.

A symplectic manifold (M,Ω) is called a *G-symplectic manifold up to a factor* if the Lie group G acts on M by maps $G \ni g \to \phi_g \in \text{Diff}(M)$, which are symplectic up to a factor:

$$\phi_g^*\Omega = \chi(g)\Omega \tag{2.2}$$

(note that $\chi: G \to \mathbb{R}^*$ is a group homomorphism).

Suppose for two G-symplectic manifolds, there exists a G-equivariant map between them which is also symplectic; then the manifolds are called *G-symplectomorphic*. The aim of this section is to classify up to a G-symplectomorphism the homogeneous G-symplectic manifolds.

Up to a point the analysis goes on the lines in Ref. 6. Suppose (Ω,M) is a G-symplectic manifold up to a χ factor.

(1) Define for any $x \in M$ the map $\Psi_x: G \to M$ by $\Psi_x(g) = \phi_g(x)$. Then it follows easily that

$$\text{Ker } \Psi_{x^*,e} = \text{Lie } G_x = \mathfrak{G}_x, \tag{2.3}$$

where G_x is the stability subgroup for $x \in M$. If the action is transitive, then

$$\text{Ran } \Psi_{x^*,e} = T_x(M). \tag{2.4}$$

(2) Now define the map ψ: $M \to \Lambda^2\mathfrak{G}^*$ (with Lie $G = \mathfrak{G}$) by

$$\psi(x) = (\Psi_x^*\Omega)_e. \tag{2.5}$$

Then the image of ψ is, in fact, in $Z^2_\chi(\mathfrak{G},\mathbb{R})$. Here $\chi = \chi_{*,e}$ and $Z^2_\chi(\mathfrak{G},\mathbb{R})$ is a linear subspace in $\Lambda^2\mathfrak{G}^*$ defined by elements σ satisfying

 0022-2488/89/112476-08$02.50

$$(\partial\sigma)(\xi_1,\xi_2,\xi_3)$$
$$=\dot{\chi}(\xi_1)\sigma(\xi_2,\xi_3)-\sigma([\xi_1,\xi_2],\xi_3)+\text{cyclic perm}$$
$$=0. \tag{2.6}$$

Here G acts naturally in $\Lambda^2\mathfrak{G}^*$ and restricts to $Z^2_\chi(\mathfrak{G},\mathbb{R})$ by

$$g\to \mathrm{Ad}^\chi(g)\equiv\chi(g)\mathrm{Ad}(g^{-1})^*. \tag{2.7}$$

Moreover, ψ is G-equivariant; it follows that it maps a homogeneous G space onto an Ad^χ orbit.

(3) From now on, M is supposed G-homogeneous. An important object for the classification theorem is the following set, defined for every $\sigma\in Z^2_\chi(\mathfrak{G},\mathbb{R})$:

$$h_\sigma=\mathrm{Ker}\,\sigma=\{\xi\in\mathfrak{G}\,|\,\sigma(\xi,\eta)=0,\ \forall\eta\in\mathfrak{G}\}. \tag{2.8}$$

From the cocycle identity (2.6) it follows that h_σ is a Lie subalgebra of $\mathfrak{G}$. The main property of this subalgebra is

$$h_{\psi(x)}=\mathfrak{G}_x, \tag{2.9}$$

for every $x\in M$. Denote by $H_\sigma\subset G$ the connected Lie group immersed in G associated to the Lie subalgebra h_σ. From (2.9) it follows that $H_{\psi(x)}$ is closed in G and

$$H_{\psi(x)}=(G_x)^0\subset G_x, \tag{2.10}$$

where $(G_x)^0$ is the connected open subgroup of G_x.

An Ad^χ orbit $\mathcal{O}$ is called *regular* if for a $\sigma\in\mathcal{O}$ (then for any $\sigma\in\mathcal{O}$), H_σ is closed.

Using the preceding facts and the reduction principle the classification of transitive actions symplectic up to a factor can be done up to covering as in Refs. 4–6.

(4) Using a somehow modified method we shall do this classification, not only up to covering. The first simple observation following from the equivariance of ψ is

$$G_x\subset G_{\psi(x)} \tag{2.11}$$

(G_σ is the stability subgroup for σ with respect to the Ad^χ action of G in $\Lambda^2 g$). Second, H_σ is a normal subgroup of G_σ so that a discrete subgroup G_x/H_σ in G_σ/H_σ corresponds to G_x.

(5) If $N\subset K$ is a normal subgroup then two subgroups $Q,Q'\subset K/N$ are called *conjugated by elements of* K if there is $k_0\in K$ such that

$$Q'=\{k_0kk_0^{-1}N\,|\,kN\in Q\}.$$

Theorem 1: Take one representative σ from every regular Ad^χ orbit in $Z^2_\chi(\mathfrak{G},\mathbb{R})$. Let $\mathscr{H}_\sigma$ be the set of discrete subgroups of G_σ/H_σ and $\mathscr{C}_\sigma$ be the set of conjugacy classes in $\mathscr{H}_\sigma$ by elements of G_σ. If $\mathscr{C}$ is the union of $\mathscr{C}_\sigma$ over all regular orbits, then it is into a one to one correspondence with the set of G-symplectomorphic classes of homogeneous G-symplectic manifolds up to a χ factor (2.2).

An explicit construction of this map is contained in the proof.

Proof: (i) First we define the one to one map j. Let M be a homogeneous G-symplectic manifold up to a χ factor. We can find $x\in M$ such that $\sigma=\psi(x)$ is the representative chosen in the statement of the theorem for the image orbit. Then it follows that G_x satisfies (2.10) and (2.11). If M' is G-symplectomorphic with M, the image orbits in $Z^2_\chi(\mathfrak{G},\mathbb{R})$ coincide, so if $x'\in M'$ is chosen as above, then $G_{x'}$ is conjugated with G_x by an element in G_σ. By definition, j maps the equivalence class of M into the equivalence class of G_x/H_σ. From above, this is consistent.

(ii) Now we prove that j is injective. Let M and M' be homogeneous G-symplectic manifolds, such that the corresponding elements in $\mathscr{C}$ are identical. We prove that M and M' are G-symplectomorphic. From the definition of j and the above assumption we can find $x\in M$, $x'\in M'$ such that

$$\psi(x)=\psi'(x')=\sigma, \tag{2.12}$$

and also G_x is conjugated to $G_{x'}$ by an element of G_σ. This allows us to choose x' such that $G_x=G_{x'}$, and (2.12) remains valid. Then we can take $M=M'=G/G_x$. Also, from (2.12),

$$(\Psi_x^*\Omega)_e=(\Psi_x^*\Omega')_e=\sigma;$$

since Ψ_x is a submersion, Ψ_x^* is injective, and thus $\Omega_x=\Omega'_x$. Using now the invariance of Ω and Ω' [(2.1)] we have $\Omega=\Omega'$ all over M. The injectivity of j follows.

(iii) Here we show that j is surjective. Let $\overline{H}\in\mathscr{H}_\sigma$ be a representative for a given element in $\mathscr{C}_\sigma$ and

$$H=\{h\in G_\sigma\,|\,hH_\sigma\in\overline{H}\}.$$

Because $\overline{H}$ is discrete it follows that H is closed in G_σ (and thus in G) and satisfies

$$H_\sigma=H^0\subset H\subset G_\sigma. \tag{2.13}$$

Lemma: Let $H\subset G$ be a closed subgroup such that (2.13) is true. Then there exists a unique symplectic form Ω on G/H, G-invariant up to a χ factor such that

$$\sigma=(\pi^*\Omega)_e \tag{2.14}$$

(here π: $G\to G/H$ is the canonical submersion).

Proof of the lemma: The action of G on G/H is denoted by $\phi_g(g'H)=gg'H$. Then $\Psi_H=\pi$; thus, using (2.3), we get

$$\mathrm{Ker}\,\pi_{*,e}=\mathrm{Lie}\,H=\mathrm{Lie}\,H_\sigma=\mathrm{Ker}\,\sigma. \tag{2.15}$$

From (2.14), we must have

$$\Omega_H(\pi_{*,e}\xi,\pi_{*,e}\eta)=\sigma(\xi,\eta). \tag{2.16}$$

This Ω_H is well defined by (2.16) because of (2.15) and the surjectivity of π_*. Also, from (2.15), it follows that Ω_H is nondegenerated. But Ω is G-invariant as in (2.2) *iff*

$$\Omega_{gH}=\chi(g)\phi_{g^{-1}}^*\Omega_H. \tag{2.17}$$

This can be taken as a consistent definition of Ω *iff*

$$\phi_h^*\Omega_H=\chi(h)\Omega_H,\quad \forall h\in H,$$

which is an easy consequence of $H\subset G_\sigma$. From (2.17) and the nondegeneracy of Ω_H it follows that Ω as a two-form is nondegenerated. From (2.6), Ω is closed. The proof of the lemma is finished.

This lemma proves that the symplectic manifold G/H generates the equivalence class, which is the preimage we are looking for. The proof of the theorem is now finished.

Remarks: (a) Loosely speaking, the classes of homogeneous G-symplectic manifolds are classified by regular orbits above, up to covering. The possible coverings are classified by classes of discrete subgroups in an appropriate group.

(b) To every regular orbit we can associate a maximal homogeneous G-symplectic manifold G/H_σ covering those classified by the set $\mathscr{C}_\sigma$.

(c) The classification already done can be made only up

to the symplectic type if we consider, instead of regular orbits in $Z^2_\chi(\mathfrak{G},\mathbb{R})$, regular orbits in the corresponding projective space. From this classification we can recover the preceding one by rescaling.

(6) The simple criterion for regular orbits provided by Chu[4] can be adapted easily to this more general situation.

Proposition 1: Let G be a connected and simply connected Lie group. Then, every orbit in $Z^2_\chi(\mathfrak{G},\mathbb{R})$ is regular.

Proof: Let $\sigma\in Z^2_\chi(\mathfrak{G},\mathbb{R})$; then $\mathrm{ad}(\xi)$ leaves $\mathrm{Ker}\,\dot\chi$ invariant in $\mathfrak{G}$. We define $\mathrm{ad}^{\#}(\xi)\in\mathrm{End}((\mathrm{Ker}\,\dot\chi)^*)$ as the dual of $\mathrm{ad}(\xi)$ restricted to $\mathrm{Ker}\,\dot\chi$. Remark that $\mathrm{ad}^{\#}(\xi)$ is a linear representation of the Lie algebra $\mathfrak{G}$. We now extend this representation to $(\mathrm{Ker}\,\dot\chi)^* + \mathbb{R}$ by the following formula:

$$\rho(\xi) = \begin{pmatrix} \mathrm{ad}^{\#}(\xi) + \dot\chi(\xi)id & i_\xi\sigma \\ 0 & 0 \end{pmatrix}.$$

It is a consequence of $\partial\sigma = 0$ that ρ is also a Lie algebra representation. Since G is connected and simply connected, there exists a group representation T of G in $(\mathrm{Ker}\,\dot\chi)^* + \mathbb{R}$ such that $\dot T = \rho$. Denote by $G_0 \subset G$ the stability subgroup of $(0,1)\in(\mathrm{Ker}\,\dot\chi)^* + \mathbb{R}$. Then $\mathrm{Lie}\,G_0 = h_\sigma$, such that H_σ is the connected component of the identity of G_0; but G_0 is closed and so is H_σ.

Remarks: (a) The hypothesis of connectedness for G is not necessary.

(b) This proposition is also valid for Lie groups with the first homotopy group finite, because the covering map is proper, and then closed.

(7) Now we discuss some important simplifications valid in the case $\dot\chi = 0$; in particular, this happens when $\chi \equiv 1$ or $\chi(g) = \pm 1$ (i.e., G acts by symplectic or antisymplectic maps). Suppose $\sigma = \partial\beta$. Then it is easy to prove

$$H_{\partial\beta} = (G_\beta)^0, \tag{2.18}$$

so that the corresponding orbits are regular. Also

$$G_\beta \subset G_{\partial\beta}, \tag{2.19}$$

and hence we have the following corollary.

Corollary: Every Ad^χ orbit in $\mathfrak{G}^*$ is a homogeneous G-symplectic manifold up to a χ factor ($\dot\chi = 0$!) with the symplectic form

$$\Omega_\beta(\xi_{\mathfrak{G}^*},\eta_{\mathfrak{G}^*}) = -\beta([\xi,\eta]). \tag{2.20}$$

Remark: For $\chi\equiv 1$ this is well known.[1-6] If $H^1_\chi(\mathfrak{G},\mathbb{R}) = 0$ then (2.19) (valid for any χ) can be easily strengthened to

$$G_\beta = G_{\partial\beta}. \tag{2.21}$$

(8) The following proposition is useful in some applications:

Proposition 2: Suppose $H^1(\mathfrak{G},\mathbb{R}) = 0 = H^2(\mathfrak{G},\mathbb{R})$. Take one representative β from every Ad^χ orbit in $\mathfrak{G}^*$. Denote the set of all subgroups of $G_\beta/(G_\beta)^0$ by $\mathscr{H}_\beta$. Finally, denote by $\mathscr{C}_\beta$ the conjugacy classes in $\mathscr{H}_\beta$ by elements of G_β, and their union over all orbits in $\mathfrak{G}^*$ by $\mathscr{C}$. There exists a one to one map between equivalence classes of G-symplectomorphic homogeneous G-symplectic manifolds up to a χ factor (2.2) and $\mathscr{C}$.

Proof: It is contained in the following remarks (combined with Theorem 1).

(a) $H^1(\mathfrak{G},\mathbb{R}) = 0$ implies $\dot\chi = 0$.

(b) $\partial: \mathfrak{G}^* \to Z^2(\mathfrak{G},\mathbb{R})$ is a one to one Ad^χ morphism.

(c) (2.18) and (2.21) are valid under the conditions of the above proposition.

(9) In this context we can define a momentum map as follows. For every $\xi\in\mathrm{Ker}\,\dot\chi$, $i_{\xi_M}\Omega$ is closed. If it is exact, let $\psi_\xi: M\to\mathbb{R}$ be defined by

$$i_{\xi_M}\Omega = d\psi_\xi; \tag{2.22a}$$

ψ can be chosen linear in ξ. Then the momentum map J: $M\to(\mathrm{Ker}\,\dot\chi)^*$ is defined by

$$J(x)(\xi) = \psi_\xi(x). \tag{2.22b}$$

The homogeneous cocycle measuring the obstruction to equivariance of the momentum map is

$$c(g_0,g_1) = \mathrm{Ad}^\chi(g_0)\phi^*_{g_0^{-1}}J - \mathrm{Ad}^\chi(g_1)\phi^*_{g_1^{-1}}J;$$

c takes values in the set of maps from $\pi_0(G)$ to $\mathfrak{G}^*$. When c is trivial, J can be taken to be equivariant.

In this case (M,Ω) is called a *χ-Hamiltonian G space.*

The equivariance of J and (2.22) determines it up to an additive element from F:

$$F = \{\lambda\in(\mathrm{Ker}\,\dot\chi)^* | \mathrm{Ad}^\chi(g)\lambda = \lambda, \quad \forall g\in G\}. \tag{2.23}$$

It follows that for χ-Hamiltonian G spaces we have

$$G_x \subset G_{J(x)}. \tag{2.24}$$

So (2.18), (2.24), and

$$\psi = -\partial J \tag{2.25}$$

imply that J is a covering map for the image orbit.

With the remark that the Ad^χ action in $(\mathrm{Ker}\,\dot\chi)^*$ factorizes to an action in $(\mathrm{Ker}\,\dot\chi)^*/F$, a classifying theorem, as in (5) can be proved for χ-Hamiltonian G spaces with $\dot\chi = 0$.

Threorem 2: For every G orbit in $\mathfrak{G}^*/F$ let us choose a representative point $\beta\in\mathfrak{G}^*$. For such β, denote by $\mathscr{H}'_\beta$ the set of all subgroups of $G_\beta/(G_\beta)^0$. Finally, denote by $\mathscr{C}'_\beta$ the conjugacy classes in $\mathscr{H}'_\beta$ by elements in G_β and by $\mathscr{C}'$ their union over orbits in $\mathfrak{G}^*/F$. There exists a one to one map between $\mathscr{C}'$ and equivalence classes of χ-Hamiltonian homogeneous G spaces (for $\dot\chi = 0$).

Proof: We follow the lines of the proof of Theorem 1.

(i) We define the one to one map j'. Let M be a χ-Hamiltonian homogeneous G space. We choose $x\in M$ such that $\beta = J(x)$ is a representative point as in the statement of the theorem (by adding to J an element in F if necessary). Then from (2.18), (2.25), (2.10), and (2.24),

$$(G_\beta)^0 = H_{\partial\beta} = H_{\psi(x)} = (G_x)^0 \subseteq G_x \subset G_\beta,$$

so that G verifies

$$(G_\beta)^0 \subseteq G_x \subset G_\beta \tag{2.26}$$

and G_x projects into a subgroup of $G_\beta/(G_\beta)^0$. The definition of j' and its injectivity follows as in parts (i) and (ii) of the proof of Theorem 1, respectively.

(ii) For surjectivity we proceed as in part (iii) of the proof of Theorem 1 with $\sigma = \partial\beta$.

It remains to show that a G/G_x thus constructed is also Hamiltonian. From (2.24) we have a unique equivariant map from $M = G/G_x$ onto $G/G_\beta \subset \mathfrak{G}^*$, sending x on β. The proof that this map verifies (2.22) and hence is a momentum map follows from (2.5):

$$(i_{\xi_M}\Omega)(\eta_M) = \Omega(\xi_M,\eta_M)_x = \psi(x)(\xi,\eta) = J(x)([\xi,\eta]).$$

Using the equivariance of J we have

$$(i_{\xi_M}\Omega)(\eta_M) = J([\xi,\eta]),$$

and now, from the infinitesimal version of equivariance,

$$\eta_M(\psi_\xi) = \psi_{[\xi,\eta]},$$

(2.22) comes. The proof of the theorem is finished.

Remarks of the same type as in (5) are also true here. We note that under the conditions of Proposition 2, every homogeneous G-symplectic manifold covers a coadjoint orbit, and the covering map is an Ad^χ-equivariant momentum map. So in this case, all these manifolds are also χ-Hamiltonian G spaces.

(10) The classification of all χ factors with $\dot\chi = 0$ reduces to finding all group homomorphisms $\tilde\chi$: $G/G^0 \to \mathbb{R}^*$. Here $\tilde\chi$ is uniquely defined by $\chi = \tilde\chi \circ \pi$ (π: $G \to G/G^0$ is the canonical surjection). For our applications we have three cases:

(i) $G/G^0 \simeq \{1\}$, for SO(3), $\mathscr{P}^\uparrow_+$, $\mathscr{G}^\uparrow_+$;

(ii) $G/G^0 \simeq \mathbb{Z}_2$, for O(3), $\mathscr{P}^\uparrow$, $\mathscr{P}_+$, $\mathscr{P}' \equiv \mathscr{P}^\uparrow_+ \cup \mathscr{P}^\downarrow_-$, $\mathscr{G}^\uparrow$, $\mathscr{G}_+$, $\mathscr{G}' \equiv \mathscr{G}^\uparrow_+ \cup \mathscr{G}^\downarrow_-$;

(iii) $G/G^0 \simeq \mathbb{Z}_2 \times \mathbb{Z}_2$, for $\mathscr{P}$, $\mathscr{G}$.

In the first case, $\tilde\chi \equiv 1$. In the second, $\tilde\chi \equiv 1$ or $\tilde\chi(\pm 1) = \pm 1$. In the third one, $\tilde\chi = 1$, $\tilde\chi = \chi_s$, χ_t, χ_{st}. Here χ_s is trivial on $\mathscr{P}_+$ or $\mathscr{G}_+$, χ_t is trivial on $\mathscr{P}^\uparrow$ or $\mathscr{G}^\uparrow$, and $\chi_{st} = \chi_s \cdot \chi_t$. Of course, χ_s, χ_t, and χ_{st} are not trivial. (For these conventions see Ref. 2.)

III. APPLICATIONS. I

(1) For SO(3) and $\mathscr{P}^\uparrow_+$, Proposition 2 in Sec. II (7) gives as symplectic homogeneous manifolds only the coadjoint orbits, because the stability subgroups G_β for these orbits are connected. This is a result of Souriau[2] obtained with the aid of the momentum map.

(2) For the study of nonconnected groups [O(3), $\mathscr{P}^\uparrow$, and $\mathscr{P}$], we shall use the following general fact. Let $G = G_1 \times_\tau G_2$ (here the homomorphism τ: $G_2 \to \mathrm{Aut}\, G_1$ gives a semidirect product structure) and $H = H_1 \times_\tau H_2$, where H_1 and H_2 are subgroups of G_1 and G_2, respectively, such that $\tau(G_2)H_1 \subset H_1$. Then the factor manifold G/H is canonically diffeomorphic to $(G_1/H_1) \times (G_2/H_2)$ with the action of G:

$$(g_1^0,g_2^0)(g_1H_1,g_2H_2) = (g_1^0\tau(g_2^0)g_1H_1, g_2^0g_2H_2). \quad (3.1)$$

In our applications, $G_1 = G^0$, $G_2 \simeq G/G^0$, and $H_1 = G_\beta \cap G^0$. So G_1/H_1 is a coadjoint orbit of G^0, and (3.1) becomes

$$(g_1^0,g_2^0)(\beta',g_2H_2) = (\mathrm{Ad}^\chi(g_1^0,g_2^0)\beta', g_2^0g_2H_2), \quad (3.2)$$

where Ad^χ is the action of G in $\mathfrak{G}^*$.

We list now, without providing explicit details of computation (much of the technique is essentially contained in Ref. 6), the homogeneous G-symplectic manifolds for the indicated groups and χ factors. Zero-dimensional manifolds are omitted.

(3) We discuss the rotation group (and inversion).

(A) For $G = \mathrm{SO}(3)$, $\chi \equiv 1$, we have $M = S^2$ (the unit sphere in $\mathbb{R}^3$); $\Omega = s \times$ surface two-form on S^2, $s \in \mathbb{R}_+$; and the action of SO(3) is the natural one. There is only one symplectic type.

Remark: All the homogeneous SU(2)-symplectic manifolds are obtained from the ones above by naturally lifting the action.

(B) For $G = \mathrm{O}(3)$, $\chi \equiv 1$, we have two outcomes.

(1) $M = S^2$; Ω as in A; and SO(3) acts as in A and I_s (spatial inversion) acts trivially.

(2) $M = S^2 \times \{1, -1\}$; Ω as in A on each sphere; and SO(3) acts as in A on each sphere and $I_s(x,\epsilon) = (x, -\epsilon)$.

There are two symplectic types

(C) For $G = \mathrm{O}(3)$, $\chi(R) = \det R$, for $R \in \mathrm{O}(3)$, we have two outcomes.

(1) $M = S^2$; Ω as in A; and O(3) acts naturally on S^2.

(2) $M = S^2 \times \{1, -1\}$; Ω as in A on each sphere; and SO(3) acts as in A on each sphere and $I_s(x,\epsilon) = (-x, -\epsilon)$.

There are two symplectic types.

(4) For the Poincaré group, all the homogeneous symplectic manifolds can be constructed from some coadjoint $\mathscr{P}^\uparrow_+$ orbits in $(\mathrm{Lie}\,\mathscr{P}^\uparrow_+)^*$ naturally identified with $(\Lambda^2\mathbb{R}^4) + \mathbb{R}^4$ with respect to the coadjoint action:[6]

$$(\mathrm{Ad}^\#(L,a))(\Gamma,P) = (L\Gamma + a \wedge LP, LP). \quad (3.3)$$

Here $a \in \mathbb{R}^4$, $L \in \mathscr{L}^\uparrow_+$, $\Gamma \in \Lambda^2\mathbb{R}^4$, $P \in \mathbb{R}^4$. The Minkowski bilinear form in $\mathbb{R}^4$ is

$$(a,b) = a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3;$$

this induces a bilinear form in $\Lambda^2\mathbb{R}^4$ so we have a natural bilinear symetric form in $(\mathrm{Lie}\,\mathscr{P}^\uparrow_+)^*$.

The symplectic form on each orbit is given by (2.20).

(A) For $G = \mathscr{P}^\uparrow_+$, $\chi \equiv 1$,

(1) $M^\eta_{m,0} = \{(\Gamma,P) \mid \|P\|^2 = m^2, \mathrm{sgn}\, P_0 = \eta, \|\Gamma \wedge P\| = 0\}$, $m \in \mathbb{R}_+$, $\eta = \pm 1$.

(2) $M^\eta_{m,s} = \{(\Gamma,P) \mid \|P\|^2 = m^2, \mathrm{sgn}\, P_0 = \eta, \|\Gamma \wedge P\|^2 = m^2s^2\}$, $s,m \in \mathbb{R}_+$, $\eta = \pm 1$.

(3) $M^\eta_s = \{(\Gamma,P) \mid \|P\|^2 = 0, \mathrm{sgn}\, P_0 = \eta, *(\Gamma \wedge P) = sP\}$, $s \in \mathbb{R}$, $\eta = \pm 1$

(here $*$ is the Hodge operator);

(4) $M'^\eta_\rho = \{(\Gamma,P) \mid \|P\|^2 = 0, \mathrm{sgn}\, P_0 = \eta, \|\Gamma \wedge P\|^2 = \rho^2\}$, $\rho \in \mathbb{R}_+$, $\eta = \pm 1$;

(5) $M'_{m,\rho} = \{(\Gamma,P) \mid \|P\|^2 = -m^2, \|\Gamma \wedge P\|^2 = m^2\rho^2\}$, $m,\rho \in \mathbb{R}_+$;

(6) $M'^\eta_{m,s} = \{(\Gamma,P) \mid \|P\|^2 = -m^2, \|\Gamma \wedge P\|^2 = -m^2s^2, \mathrm{sgn}\,(*(\Gamma \wedge P))_0 = \eta\}$, $m \in \mathbb{R}_+$, $s \in \mathbb{R}$, $\eta = \pm 1$;

(7) $M_{\lambda_1,\lambda_2} = \{(\Gamma,O) \mid \|\Gamma\|^2 = \lambda_1, \frac{1}{2}\Gamma \wedge \Gamma = \lambda_2 \times \text{vol. elem. in } \mathbb{R}^4\}$.

In (1) and (4), we have a single symplectic type. In (2), the symplectic type is given by s/m; in (3), we have three

symplectic types indexed by $\mathrm{sgn}(s\eta)$; in (5), the symplectic type is given by ρ/m; in (6), by s/m. In (7), for $\lambda_1 \neq 0$, the symplectic type is indexed by λ_2/λ_1^2; for $\lambda_1 = 0$, there are three more types ($\lambda_2 > 0$, $\lambda_2 = 0$, $\lambda_2 < 0$).

Remark: It is usually assumed[2] that only the first three cases have a direct physical meaning. On the other hand, if M is a homogeneous G-symplectic manifold, up to a χ factor, it is easy to see that every connected component of M is a homogeneous G^0-symplectic manifold up to a χ/G^0 factor. So in the following we list for the nonconnected groups only those manifolds that have at least one physical component. Moreover, we shall see that in the cases above all the components are physical.

(B) For $G = \mathscr{P}^\uparrow$, $\chi \equiv 1$,

(1) $M^\eta_{m,0}$ with the action of the inversion given also by (3.3);

(2) $M^\eta_{m,s}$ with the same action as above, $s \in \mathbb{R}_+$;

(3) $M^\eta_{m,0} \times \{1, -1\}$ with the action of $\mathscr{P}^\uparrow_+$ as in (3.3) on each component and $\varphi_{I_s}(x,\epsilon) = (\mathrm{Ad}^\#_{I_s}(x), -\epsilon)$;

(4) $M^\eta_{m,s} \times \{1, -1\}$ as above;

(5) $M^\eta_s \cup M^\eta_{-s}$, $s \in \mathbb{R}_+$, $\eta = \pm 1$ with the coadjoint action of $\mathscr{P}^\uparrow$;

(6) M^η_0 with the coadjoint action of $\mathscr{P}^\uparrow$;

(7) $M^\eta_0 \times \{1, -1\}$ as in (3) above.

Remarks: (a) In the cases (B)(1), (B)(2), and (B)(6) the manifolds are connected. A similar analysis for $\mathscr{P}_+$ and $\mathscr{P}'$ shows that there are no connected manifolds.

(b) The symplectic types can be found as in (A).

(C) For $G = \mathscr{P}$, $\chi \equiv 1$,

(1) $M^+_{m,0} \cup M^-_{m,0}$, $\mathscr{P}$ acting as in (3.3);

(2) $M^+_{m,s} \cup M^-_{m,s}$ acting as in (3.3);

(3) $(M^+_{m,0} \cup M^-_{m,0}) \times \{1, -1\}$, I_s acting as in (A) (3);

(4) $(M^+_{m,s} \cup M^-_{m,s}) \times \{1, -1\}$ as above;

(5) $M^+_s \cup M^-_s \cup M^+_{-s} \cup M^-_{-s}$, $s \in \mathbb{R}_+$, $\mathscr{P}$ acting as in (3.3);

(6) $(M^+_0 \cup M^-_0)$, $\mathscr{P}$ acting as in (3.3);

(7) $(M^+_0 \cup M^-_0) \times \{1, -1\}$, as in (3) above.

Remarks: (a) All the manifolds are disconnected.

(b) These results were obtained in Ref. 2 using the existence of the momentum map for $\mathscr{P}$ [as a consequence of $H^1(\mathrm{Lie}\,\mathscr{P}, \mathbb{R}) = 0$].

(c) The symplectic types can be analyzed as in (A).

(D) For $G = \mathscr{P}$, $\chi = \chi_t$,

(1) $M^\eta_{m,0}$;

(2) $M^\eta_{m,0} \times \{1, -1\}$ with the action $\phi_{(L,a)}(x,\epsilon) = ((\mathrm{Ad}^{\chi_t}(L,a))(x), \chi_s(L)\epsilon)$;

(3) $M^\eta_{m,0} \times \{1, -1\}$ with the action $\phi_{(L,a)}(x,\epsilon) = ((\mathrm{Ad}^{\chi_t}(L,a))(x), \chi_t(L)\epsilon)$;

(4) $M^\eta_{m,0} \times \{1, -1\}$ with the action $\phi_{(L,a)}(x,\epsilon) = ((\mathrm{Ad}^{\chi_t}(L,a))(x), \chi_{st}(L)\epsilon)$;

(5) $M^\eta_{m,0} \times \{1, -1\} \times \{1, -1\}$ with the action $\phi_{(L,a)}(x,\epsilon,\epsilon') = ((\mathrm{Ad}^{\chi_t}(L,a))(x), \chi_s(L)\epsilon, \chi_t(L)\epsilon')$;

(6)–(10) the same possibilities as in (1)–(5) with $M^\eta_{m,s}$ ($s \in \mathbb{R}_+$) instead of $M^\eta_{m,0}$;

(11) $M^\eta_s \cup M^\eta_{-s}$ with the Ad^{χ_t} action of $\mathscr{P}$, $s \in \mathbb{R}_+$;

(12) $(M^\eta_s \cup M^\eta_{-s}) \times \{1, -1\}$ with the action $\phi_{(L,a)}(x,\epsilon) = ((\mathrm{Ad}^{\chi_t}(L,a))(x), \chi_t(L)\epsilon)$;

(13)–(17) the same as in (1)–(5) above but with M^η_0 instead of $M^\eta_{m,0}$.

Remarks: (a) In the cases $G = \mathscr{P}$ and $\chi = \chi_s$ or χ_{st} the analysis is similar and shows that connected manifolds do not appear.

(b) The possibilities (1) and (6) above also appear in Ref. 2, without a proof that these are the only connected manifolds for $m \in \mathbb{R}_+$.

IV. APPLICATIONS. II. THE GALILEI GROUP

As in Ref. 6 we identify $\mathscr{G}^\uparrow$ with the set of 5×5 matrices of the form

$$(R,v,\tau,a) \equiv \begin{pmatrix} R & v & a \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{pmatrix}, \quad R \in O(3), \quad v,a \in \mathbb{R}^3, \quad \tau \in \mathbb{R}.$$

Now $\mathscr{G}^\uparrow_+$ is defined by $R \in SO(3)$ and the temporal inversion is

$$I_t \equiv \begin{pmatrix} \mathbb{1} & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We study only the case $\dot{\chi} = 0$. For this group there are no essential simplifications for the use of Theorem 1. Using remark (b) in Sec. II (6), it follows that all orbits in the two cocycle space are regular.

(A) For $G = \mathscr{G}^\uparrow_+$, $\chi \equiv 1$, $Z^2(\mathrm{Lie}\,\mathscr{G}^\uparrow_+, \mathbb{R})$ is isomorphic to $\mathbb{R} \times \Lambda^2\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$ with the following action of $\mathscr{G}^\uparrow_+$:

$$\mathrm{Ad}^{\#}_{(R,v,\tau,a)}[m,l,G,P] = [m,Rl + a\wedge(RP + mv) + RG\wedge v, R(G - \tau P) + m(a - \tau v), RP + mv]. \tag{4.1}$$

The corresponding $\mathscr{G}^{\uparrow}_{+}$ orbits are of five types:

(a) $\mathscr{O}_{m,s} \equiv \{[m,l,G,P] \mid \|ml - G\wedge P\| = ms\}$, $m\in\mathbb{R}^*$, $s\in\mathbb{R}_+\cup\{0\}$;

(b) $\tilde{\mathscr{O}}_{k,s} \equiv \{[0,l,G,P] \mid \|P\| = k,\ G\wedge P = 0,\ l\wedge P = ks\times\text{vol. elem. in } \mathbb{R}^3\}$, $k\in\mathbb{R}_+$, $s\in\mathbb{R}$;

(c) $\tilde{\mathscr{O}}'_{k,k'} \equiv \{[0,l,G,P] \mid \|P\| = k,\ \|G\wedge P\| = kk'\}$, $k,k'\in\mathbb{R}_+$;

(d) $\mathscr{O}'_{k,r} \equiv \{[0,l,G,0] \mid \|G\| = k,\ l\wedge G = kr\times\text{vol. elem. in } \mathbb{R}^3\}$, $k\in\mathbb{R}_+$, $r\in\mathbb{R}$;

(e) $\mathscr{O}_s \equiv \{[0,l,0,0] \mid \|l\| = s\}$, $s\in\mathbb{R}_+\cup\{0\}$.

We remark that for orbits of the same type [(a)–(e)] G_σ/H_σ are the same (up to isomorphism). We also note that in cases (b)–(e), the homogeneous $\mathscr{G}^{\uparrow}_{+}$-symplectic manifolds are coadjoint orbits with the symplectic form given by (2.20) or manifolds covered by them. (Lie $\mathscr{G}^{\uparrow}_{+}$)* is isomorphic to $\Lambda^2\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3$ with the coadjoint action:

$$\mathrm{Ad}^{\#}_{(R,v,\tau,a)}(l,G,E,P) = (Rl + a\wedge RP + RG\wedge v, R(G - \tau P), E + (RP,v), RP). \tag{4.2}$$

We now list the corresponding discrete subgroups and homogeneous symplectic manifolds.

(a) $G_\sigma/H_\sigma \cong \{1\}$.

(1) For $s = 0$; $M_m = \mathbb{R}^3\times\mathbb{R}^3$; the symplectic form is

$$\Omega^{M_m} = dq_i\wedge dp^i,$$

and the action of $\mathscr{G}^{\uparrow}_{+}$ is

$$\phi^{(m)}_{(R,v,\tau,a)}(q,p) = (Rq + a - \tau(1/m)(Rp + mv), Rp + mv). \tag{4.3}$$

(2) For $s\in\mathbb{R}_+$, $M_{m,s} = M_m\times S^2$; $\Omega^{M_{m,s}} = \Omega^{M_m} + s\times$surface two-form on S^2; and the action of $\mathscr{G}^{\uparrow}_{+}$ is given by (4.3) on the first entry and by rotation on the second one.

(b) $G_\sigma/H_\sigma \cong \mathbb{R}$. The discrete subgroups of $\mathbb{R}$ are indexed by $\gamma\in\mathbb{R}_+\cup\{0\}$ and are of the form $\gamma\mathbb{Z}$.

(1) For $\gamma = 0$ (a coadjoint orbit),

$$\tilde{M}_{k,s} = \{(l,G,E,P) \mid \|P\| = k,\ G\wedge P = 0,\ l\wedge P = ks\times\text{vol. elem. in } \mathbb{R}^3\}.$$

(2) For $\gamma\in\mathbb{R}_+$, let $\mathbb{Z}$ act on $\tilde{M}_{k,s}$ by

$$n,(l,G,E,P)\mapsto(l,G,E + nk\gamma,P). \tag{4.4}$$

Then the manifold is $\tilde{M}_{k,s,\gamma} = \tilde{M}_{k,s}/\mathbb{Z}$. The action (4.2) of $\mathscr{G}^{\uparrow}_{+}$ and the symplectic form (2.20) factorize to $\tilde{M}_{k,s,\gamma}$.

(c) $G_\sigma/H_\sigma \simeq \mathbb{R}$. The discrete subgroups are those of (b) above

(1) For $\gamma = 0$ (a coadjoint orbit),

$$\tilde{M}'_{k,k'} \equiv \{(l,G,E,P) \mid \|P\| = k',\quad \|G\wedge P\| = kk'\},$$

(2) For $\gamma\in\mathbb{R}_+$, the manifold is also $\tilde{M}'_{k,k',\gamma} \equiv \tilde{M}'_{k,k'}/\mathbb{Z}$, where $\mathbb{Z}$ acts as in (4.4).

(d) $G_\sigma/H_\sigma \simeq \{1\}$ (a coadjoint orbit). We have

$$M'_{k,r} \equiv \{(l,G,0,0) \mid \|G\| = k,\quad l\wedge G = kr\times\text{vol. elem. in } \mathbb{R}^3\}.$$

(e) $G_\sigma/H_\sigma \simeq \{1\}$ (a coadjoint orbit). We have

$$M_s \equiv \{(l,0,0,0) \mid \|l\| = s\}.$$

Remark: As for the Poincaré group, only the first three cases are considered with physical meaning.[2] We discuss the inversions in the spirit of the remark in Sec. II (4) (A).

(B) $G = \mathscr{G}^{\uparrow}, \chi\equiv 1$. The $\mathscr{G}^{\uparrow}$ orbits in Z^2 (Lie $\mathscr{G}^{\uparrow}, \mathbb{R}$) are $\mathscr{O}_{m,s}$, $\tilde{\mathscr{O}}_{k,0}$, $\tilde{\mathscr{O}}'_{k,k'}$, $\mathscr{O}'_{k,0}$, $\mathscr{O}_s$, and $\tilde{\mathscr{O}}_{k,s}\cup\tilde{\mathscr{O}}_{k,-s}$, $\mathscr{O}'_{k,s}\cup\mathscr{O}'_{k,-s}$ (for $s\in\mathbb{R}_+$).

(a) For $\mathscr{O}_{m,s}$, $G_\sigma/H_\sigma \simeq \mathbb{Z}_2$,

(1) M_m from (A) with (4.3) valid for all $\mathscr{G}^{\uparrow}$;

(2) $M_{m,s}$ with the same action as in (A) (2) above, $s\in\mathbb{R}_+$;

(3) $M_m\times\{1,-1\}$ constructed as in Sec. III (3) (B) (2);

(4) $M_{m,s}\times\{1,-1\}$ as above, $s\in\mathbb{R}_+$.

(b) For $\tilde{\mathscr{O}}_{k,s}\cup\tilde{\mathscr{O}}_{k,-s}$ ($s\in\mathbb{R}_+$), $G_\sigma/H_\sigma \simeq \mathbb{R}$, where the discrete subgroups are $\gamma\mathbb{Z}$:

(1) $\gamma = 0$, $\tilde{M}_{k,s}\cup\tilde{M}_{k,-s}$ with the coadjoint action;

(2) $\gamma\in\mathbb{R}_+$ $\tilde{M}_{k,s,\gamma}\cup\tilde{M}_{k,-s,\gamma}$ with the factorized action.

(c) For $\tilde{\mathscr{O}}_{k,0}$, $G_\sigma/H_\sigma \simeq \mathbb{Z}_2\times\mathbb{R}$, the discrete subgroups of $\mathbb{Z}_2\times\mathbb{R}$ are $\{1\}\times\gamma\mathbb{Z}$, $\mathbb{Z}_2\times\gamma\mathbb{Z}$,

$(1,\gamma\mathbb{Z})\cup(-1,\gamma/2 + \gamma\mathbb{Z})$, for $\gamma\in\mathbb{R}_+\cup\{0\}$.

(1) For $\gamma = 0$, (i) $\tilde{M}_{k,0}$ with the coadjoint action; and (ii) $\tilde{M}_{k,0}\times\{1,-1\}$ constructed as in Sec. III (4)(B)(3).

(2) For $\gamma\in\mathbb{R}_+$, (i) $\tilde{M}_{k,0,\gamma}$ with the factorized action; (ii) $\tilde{M}_{k,0,\gamma}\times\{1,-1\}$ with the factorized action; and (iii) $\tilde{M}_{k,0,\gamma}$ with the action

$$\varphi_{(R,v,\tau,a)}(l,G,E(\mathrm{mod}\,k\gamma),P) = (Rl + a\wedge RP + RG\wedge v, R(G - \tau P),\quad E + (RP,v) + (k\gamma/2)(\mathrm{mod}\,k\gamma), RP). \tag{4.5}$$

Remark: In the cases (A) and (B) above the equivalence classes in $\mathscr{H}_\sigma$ are formed by a single element.

(C) $G=\mathscr{G}, \chi\equiv 1$. The orbits in Z^2 (Lie $\mathscr{G}, \mathbb{R}$) are $\mathcal{O}_{m,s}\cup\mathcal{O}_{-m,s}$ (for $m\in\mathbb{R}_+$), $\tilde{\mathcal{O}}_{k,0}$, $\tilde{\mathcal{O}}'_{k,k'}$, $\mathcal{O}'_{k,0}$, $\mathcal{O}_s$, and $\tilde{\mathcal{O}}_{k,s}\cup\tilde{\mathcal{O}}_{k,-s}$, $\mathcal{O}'_{k,s}\cup\mathcal{O}'_{k,-s}$ (for $s\in\mathbb{R}_+$).

(a) For $\mathcal{O}_{m,s}\cup\mathcal{O}_{-m,s}$, $G_\sigma/H_\sigma\simeq\mathbb{Z}_2$, (1) $M_m\times\{1,-1\}$, $m\in\mathbb{R}_+$, $\mathscr{G}^\uparrow$ acts as in (B) (1) on each component and $\varphi_{I_t}(x,\epsilon)=(x,-\epsilon)$; (2) $M_m\times\{1,-1\}\times\{1,-1\}$, $m\in\mathbb{R}_+$, $\mathscr{G}^\uparrow_+$ acts as in (A) (1) on each component and $\varphi_{I_s}(x,\epsilon,\epsilon')=(x,-\epsilon,\epsilon')$, $\varphi_{I_t}(x,\epsilon,\epsilon')=(x,\epsilon,-\epsilon')$, and (3), (4) the same as above with $M_{m,s}$ instead of M_m, for $s\in\mathbb{R}_+$.

(b) For $\tilde{\mathcal{O}}_{k,s}\cup\tilde{\mathcal{O}}_{k,-s}$, $\mathscr{G}_\sigma/\mathscr{H}_\sigma\simeq\mathbb{R}\times_\tau\mathbb{Z}_2$, where the homomorphism $\tau\colon\mathbb{Z}_2\to\mathrm{Aut}\,\mathbb{R}$, which gives the semidirect product structure, is $\tau(\epsilon)=\epsilon\,\mathrm{Id}$. The discrete subgroups (one from each equivalence class in $\mathscr{H}_\sigma$) are $\gamma\mathbb{Z}\times_\tau\{1\}$, $\gamma\mathbb{Z}\times_\tau\mathbb{Z}_2$, for $\gamma\in\mathbb{R}_+\cup\{0\}$.

(1) For $\gamma=0$,

(i) $\tilde{M}_{k,s}\cup\tilde{M}_{k,-s}$ with the coadjoint action;

(ii) $(\tilde{M}_{k,s}\cup\tilde{M}_{k,-s})\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\mathrm{Ad}_g^{\#}(x),\ \chi_t(g)\epsilon)$.

(2) For $\gamma\in\mathbb{R}_+$,

(i) $\tilde{M}_{k,s,\gamma}\cup\tilde{M}_{k,-s,\gamma}$ with the factorized action;

(ii) $(\tilde{M}_{k,s,\gamma}\cup\tilde{M}_{k,-s,\gamma})\times\{1,-1\}$ with the factorized action.

(c) For $\tilde{\mathcal{O}}_{k,0}$, $G_\sigma/H_\sigma\simeq(\mathbb{R}\times_\tau\mathbb{Z}_2)\times\mathbb{Z}_2$ with τ as in (b); the discrete subgroups (one from each equivalence class) are

$(\gamma\mathbb{Z},1,1)$, $(\gamma\mathbb{Z},1,1)\cup(\gamma\mathbb{Z},1,-1)$, $(\gamma\mathbb{Z},1,1)\cup(\gamma\mathbb{Z},-1,1)$, $(\gamma\mathbb{Z},1,1)\cup(\gamma\mathbb{Z},-1,-1)$,

$(\gamma\mathbb{Z},1,1)\cup(\gamma\mathbb{Z},1,-1)\cup(\gamma\mathbb{Z},-1,1)\cup(\gamma\mathbb{Z},-1,-1)$

(for $\gamma\in\mathbb{R}_+\cup\{0\}$),

$(\gamma\mathbb{Z},1,1)\cup(\gamma/2+\gamma\mathbb{Z},1,-1)$, $(\gamma\mathbb{Z},1,1)\cup(\gamma/2+\gamma\mathbb{Z},1,-1)\cup(\gamma\mathbb{Z},-1,1)\cup(\gamma/2+\gamma\mathbb{Z},-1,-1)$

(for $\gamma\in\mathbb{R}_+$).

(1) For $\gamma=0$,

(i) $\tilde{M}_{k,0}$ with the coadjoint action;

(ii) $\tilde{M}_{k,0}\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\mathrm{Ad}_g^{\#}(x),\ \chi_s(g)\epsilon)$;

(iii) $\tilde{M}_{k,0}\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\mathrm{Ad}_g^{\#}(x),\ \chi_t(g)\epsilon)$;

(iv) $\tilde{M}_{k,0}\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\mathrm{Ad}_g^{\#}(x),\ \chi_{st}(g)\epsilon)$;

(v) $\tilde{M}_{k,0}\times\{1,-1\}\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon,\epsilon')=(\mathrm{Ad}_g^{\#}(x),\ \chi_s(g)\epsilon,\ \chi_t(g)\epsilon')$.

(2) For $\gamma\in\mathbb{R}_+$,

(i)–(v) the same as above with $\tilde{M}_{k,0,\gamma}$ instead of $\tilde{M}_{k,0}$ and the corresponding factorized actions;

(vi) $\tilde{M}_{k,0,\gamma}$ with the action of $\mathscr{G}^\uparrow$ given by (4.5) and $\varphi_{I_t}(l,G,E(\mathrm{mod}\,k\gamma),P)=(l,-G,-E(\mathrm{mod}\,k\gamma),P)$;

(vii) $\tilde{M}_{k,0,\gamma}\times\{1,-1\}$ constructed from (vi) as in (1) (ii).

(D) $G=\mathscr{G}, \chi=\chi_t$. The orbits in Z^2 (Lie $\mathscr{G}$,$\mathbb{R}$) are $\mathcal{O}_{m,s}$, $\tilde{\mathcal{O}}_{k,0}$, $\tilde{\mathcal{O}}'_{k,k'}$, $\mathcal{O}'_{k,0}$, $\mathcal{O}_s$, and $\tilde{\mathcal{O}}_{k,s}\cup\tilde{\mathcal{O}}_{k,-s}$, $\mathcal{O}'_{k,s}\cup\mathcal{O}'_{k,-s}$ (for $s\in\mathbb{R}_+$).

(a) For $\mathcal{O}_{m,s}$, $G_\sigma/H_\sigma\simeq\mathbb{Z}_2\times\mathbb{Z}_2$,

(1) M_m with the action of $\mathscr{G}^\uparrow$ given by (4.3) and $\varphi^{(m)}_{I_t}(q,p)=(q,-p)$;

(2) $M_m\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\phi^{(m)}_g(x),\ \chi_s(g)\epsilon)$;

(3) $M_m\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\phi^{(m)}_g(x),\ \chi_t(g)\epsilon)$;

(4) $M_m\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon)=(\phi^{(m)}_g(x),\ \chi_{st}(g)\epsilon)$;

(5) $M_m\times\{1,-1\}$ with the action $\varphi_g(x,\epsilon,\epsilon')=(\phi^{(m)}_g(x),\ \chi_s(g)\epsilon,\ \chi_t(g)\epsilon')$;

(6)–(10) the same possibilities as in (1)–(5) with $M_{m,s}$ instead of M_m, $s\in\mathbb{R}_+$.

(b) For $\tilde{\mathcal{O}}_{k,s}\cup\tilde{\mathcal{O}}_{k,-s}$, $G_\sigma/H_\sigma\simeq\mathbb{Z}_2\times\mathbb{R}$, $s\in\mathbb{R}_+$, the discrete subgroups are those of (B) (c).

(1) For $\gamma=0$,

(i) $\tilde{M}_{k,s}\cup\tilde{M}_{k,-s}$ with the Ad^{χ_t} action of $\mathscr{G}$;

(ii) $(\tilde{M}_{k,s}\cup\tilde{M}_{k,-s})\times\{1,-1\}$ constructed as in (3) above.

(2) For $\gamma\in\mathbb{R}_+$,

(i) $\tilde{M}_{k,s,\gamma}\cup\tilde{M}_{k,-s,\gamma}$;

(ii) $(\tilde{M}_{k,s,\gamma}\cup\tilde{M}_{k,-s,\gamma})\times\{1,-1\}$;

with the factorized actions.

(c) For $\tilde{\mathcal{O}}_{k,s}$, $G_\sigma/H_\sigma\simeq(\mathbb{R}\times_\tau\mathbb{Z}_2)\times\mathbb{Z}_2$ with τ as before, the discrete subgroups are those of (C) (c).

(1) For $\gamma=0$, (i)–(iv) the same possibilities as (1)–(5) above with $\tilde{M}_{k,0}$ and Ad^{χ_t} instead of M_m and $\phi^{(m)}$, respectively.

(2) For $\gamma\in\mathbb{R}_+$, (i)–(iv) the same possibilities as above, only with the factorized actions; (vi) $\tilde{M}_{k,0,\gamma}$ with the action of $\mathscr{G}^\uparrow$ given by (4.5) and

$$\varphi_{I_t}(l,G,E(\mathrm{mod}\,k\gamma),P)=(-l,G,E(\mathrm{mod}\,k\gamma),-P);$$

(vii) $\tilde{M}_{k,0,\gamma}\times\{1,-1\}$ constructed from above as in (ii).

V. CONCLUSION

In this paper, using the general classification theorem (Theorem 1 in Sec. II), we have succeeded in completely classifying the homogeneous G-symplectic manifolds for an arbitrary χ factor and/or nonconnected symmetry groups. This includes the important case of mixed symplectic and antisymplectic symmetries used to describe inversions. Also, we do not need the hypothesis of the existence of the momentum map which, for instance, permits a clearcut analysis of the Galilei group. This analysis shows that a complete list of homogeneous Galilean symplectic manifolds includes, besides those already known in the literature,[5,6] some new ones, corresponding to zero mass, covered by coadjoint orbits, and not having a momentum map.

For the physical interpretation it is desirable to investigate the existence of an evolution space with an associated extended configuration space, as in Refs. 2 and 6. This will be done elsewhere.

It will be profitable to find groups for which $\chi \neq 0$ can occur and that are related to interesting physical systems.

Comparing the lists in Sec. IV (3) with the well known results in quantum mechanics,[8–10] we remark that "doubling" and "quadrupling" phenomena appear that do not have a quantum counterpart, e.g., B(2), (F)(2) in Sec. III. The same can be said for the new possibilities for the Galilei group. It is worthwhile to investigate in detail these problems using the ideas of geometric quantization.

[1] B. Kostant, *Quantization and Unitary Representations, Lecture Notes in Mathematics*, Vol. 70 (Springer, New York, 1970).
[2] J. M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970).
[3] A, Kirillov, *Elèments des la theorie des représentations* (Mir, Moscow, 1974).
[4] B. Y. Chu, "Symplectic homogeneous spaces," Trans. Am. Math. Soc. **197**, 145 (1975).
[5] S. Sternberg, "Symplectic homogeneous spaces," Trans. Am. Math. Soc. **212**, 113 (1975).
[6] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U. P., Cambridge, 1984).
[7] D. R. Grigore, "A possible generalization of the concept of symmetry in analytical mechanics," preprint FT-316-1987; J. Math. Phys. **30**, 2638 (1989).
[8] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, New York, 1970), Vol. II.
[9] L. Fonda and G. C. Ghirardi, *Symmetry Principles in Quantum Physics* (Marcel Dekker, New York, 1970).
[10] A. S. Wightman (Les Houches, Hermann, Paris, 1960).

Relativistic and nonrelativistic plasma dispersion functions

P. A. Robinson
Department of Astrophysical, Planetary and Atmospheric Sciences, Campus Box 391, University of Colorado at Boulder, Boulder, Colorado 80309-0391

(Received 9 September 1988; accepted for publication 21 June 1989)

A number of new results involving the functions that determine the emission, dispersion, and absorption of waves in nonrelativistic and weakly relativistic plasmas are presented. These results comprise series, integrals, recursion relations, symmetry properties, interrelations, approximations, and connections with standard transcendental functions.

I. INTRODUCTION

The dielectric properties of a plasma determine the amplification, dispersion, absorption, and mode conversion of the waves it supports. These properties can be expressed analytically in terms of plasma disperison functions (PDF's), which arise from integrals over the momentum distribution function of the plasma, and also relate to the emission of waves. The most commonly considered PDF's are those of thermal plasmas, which include the nonrelativistic Z function[1] and the weakly relativistic PDF's discussed in Refs. 2 and 3. These and closely related functions also occur in the quantum electrodynamics of strongly magnetized plasmas, and in problems of acoustics[1] and heat conduction.[4] (For more extensive references of these applications, see Ref. 2.)

The purpose of this work is to present a number of new properties of the PDF's of thermal plasmas, and other related PDF's, thereby expanding the range of mathematical tools available when investigating the dispersion, absorption, and interaction of waves in plasmas. Section II is concerned with the nonrelativistic PDF's that occur in expressions for the dielectric properties of unmagnetized plasmas. The results presented are relevant to approximation of the dielectric properties of Maxwellian plasmas and to their calculation for plasmas in which superthermal particles are present.[5] Moreover, these dispersion functions occur in orthogonal-polynomial decompositions of the dielectric properties of arbitrary unmagnetized plasmas, as described in a related paper in which these results are applied.[6] In Sec. III, we discuss new properties and approximations of the PDF's of weakly relativistic magnetized plasmas.[2,3] These functions are used in determining the properties of waves in hot magnetized plasmas. In many cases, a barrier to analytic, semiquantitative, and numerical work on such wave problems is the general paucity of known properties of the dispersion functions. In this context, the results obtained here provide new properties and approximations of use in analytical and numerical studies, as well as tests of the accuracy of numerical calculations. Several new integral relationships involving standard transcendental function are also obtained, which have wider applicability, in general.

II. NONRELATIVISTIC PDF'S

The Z function and its first derivative have long been used in describing dispersion in thermal plasmas,[1] while higher derivatives are employed in a companion paper in which the dielectric properties of more general nonrelativistic plasmas are expanded via Hermite polynomials.[6] These functions also arise in the approximation of weakly relativistic PDF's[2,3,7] and in problems of acoustics.[1] Here, we first consider a class of PDF's consisting of the standard PDF Z for a nonrelativistic thermal plasma and its derivatives $Z^{(n)}$.

An orthogonal–polynomial expansion of the dielectric properties of arbitrary plasmas is developed in a companion paper.[6] One realization of this method, via Hermite polynomials, motivates consideration of the derivative dispersion functions $Z^{(n)}$ here, in order to provide the necessary mathematical basis for this work in Ref. 6. The function $Z^{(n)}(u)$ may be defined by

$$Z^{(n)}(u) = i\int_0^\infty dt(it)^n \exp\left(iut - \frac{1}{4}t^2\right), \tag{1}$$

for all complex u. The Z function is also related to the complementary error function by[1]

$$Z(u) = i\pi^{1/2}e^{-u^2}\operatorname{erfc}(-iu).$$

Use of this expression, together with Eq. 7.2.9 of Ref. 8, implies that $Z^{(n)}(u)$ is related to the nth integral of the complementary error function $I^n\operatorname{erfc}(-iu)$ by the following formula:

$$Z^{(n)}(u) = i\pi^{1/2}(2i)^n n! e^{-u^2} I^n \operatorname{erfc}(-iu). \tag{2}$$

Repeated application of the basic formula[1]

$$Z^{(1)}(u) = -2 - 2uZ(u), \tag{3}$$

leads to the recursion relation

$$Z^{(n+1)}(u) = -2uZ^{(n)}(u) - 2nZ^{(n-1)}(u), \tag{4}$$

for $n \geqslant 1$. Substitution of the recursion relation $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$ for Hermite polynomials into the integral

$$\int_{-\infty}^{\infty} dx \frac{(-1)^n H_n(x)e^{-x^2}}{x-u} \tag{5a}$$

(which occurs in the orthogonal–polynomial decomposition of Ref. 6) shows that integrals of this form also satisfy the recursion relation (4) with respect to n. In the case $n = 0$, Eq. (5a) defines $Z(u)$ for $\operatorname{Im} u > 0$ (see Ref. 1) and, hence, the following alternative definition of $Z^{(n)}$ can be stated:

$$Z^{(n)}(u) = \frac{1}{\pi^{1/2}}\int_{-\infty}^{\infty} dx \frac{(-1)^n H_n(x)e^{-x^2}}{x-u}. \tag{5b}$$

Equation (5b) generalizes the corresponding result for Z and is valid for $\operatorname{Im} u > 0$; whereas analytic continuation of (5b) is used for $\operatorname{Im} u \leqslant 0$. Using the above expressions, one

 0022-2488/89/112484-04$02.50

can show that the functions $Z^{(n)}$ satisfy the symmetry relations

$$Z^{(n)}(u) = (-1)^{n+1}[Z^{(n)}(-u^*)]^*, \tag{6}$$

$$Z^{(n)}(u^*) = Z^{(n)}(u)^* + 2i\pi^{1/2}(-1)^n[e^{-u^2}H_n(u)]^*, \tag{7}$$

which generalize the analogous results for $Z(u)$, and can be most easily proved directly by induction.

Asymptotic series and power series for Z have been given in Refs. 1 and 2, from which it is straightforward to obtain the corresponding series for $Z^{(n)}$. Hence we omit the results here, but note that the definition (1) enables one to evaluate the following infinite series involving $Z^{(n)}(u)$:

$$\sum_{n=0}^{\infty} \frac{(-u)^n}{n!} Z^{(n)}(u) = i\pi^{1/2}, \tag{8}$$

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} Z^{(2n)}(u) = -\frac{1}{u}. \tag{9}$$

In applications to analysis of dispersion in arbitrary plasmas, sequences of $Z^{(n)}(u)$ are needed (Ref. 6). Equations (8) and (9) may be used to normalize these sequences when downward recursion via the functions I^n erfc is used [upward recursion using Eq. (6) is unstable for large u, because of the effects of cancellation].[4,6,9] This removes the need to calculate Z spearately. Physically, the cases of interest involve values of u near the real axis (corresponding to weakly damped waves). In such cases (8) is not an alternating series for large u, where alternation would lead to numerical cancellation problems because $Z^{(n)}$ alternates in sign with n. Likewise, Eq. (9) ceases to oscillate as the imaginary axis is approached.

Recent work has shown that the dispersion of generalized-Lorentzian plasmas can be used to approximate that of Maxwellian plasmas in analytic work, in order to speed numerical calculations, and to explore the sensitivity of the physical wave properties to the high-velocity tail of the distribution.[5] Lorentzian plasmas are also of importance in their own right, because they can be used to model the power-law tails produced by many plasma heating and acceleration mechanisms. For example, this problem is of current interest in understanding the dispersion of magnetohydrodynamic waves in the laboratory and space, where superthermal tails are commonplace (Refs. 10–12).

Approximation of the PDF Z by the PDF's L_n of a plasma with a generalized-Lorentzian velocity distribution may be accomplished if we make the definition

$$L_n(u,a) = \int_{-\infty}^{\infty} \frac{dx}{(x^2+a^2)^n(x-u)} \times \left[\int_{-\infty}^{\infty} \frac{dx}{(x^2+a^2)^n}\right]^{-1}, \tag{10}$$

and note the identity $\exp(-x^2) = \lim_{n\to\infty}(1+x^2/n)^{-n}$. We then have

$$Z(u) = \lim_{n\to\infty} L_n(u,n^{1/2}), \tag{11}$$

which leads to a set of approximations of the form $Z(u) \approx L_n(u,n^{1/2})$ for finite n. More generally, the parameter a is the characteristic dimensionless velocity above which the near-Maxwellian core of the distribution passes over into the power-law tail.

In Ref. 5 it was shown that the functions L_n satisfy the equation

$$L_n(u,a) = [\pi^{1/2}a^{2n-1}\Gamma(n)/\Gamma(n-\tfrac{1}{2})]I_n(u,a), \tag{12}$$

with

$$I_1(u,a) = \frac{-1}{a(u+ia)}, \tag{13a}$$

$$I_{n+1}(u,a) = \left(\frac{-1}{2an}\frac{\partial}{\partial a}\right) I_n(u,a). \tag{13b}$$

Alternatively, $L_n(u,a)$ can be written in the form[5]

$$L_n(u,a) = \frac{\pi^{1/2}\Gamma(n)}{\Gamma(n-\frac{1}{2})(u+ia)^n} \sum_{j=0}^{n-1} d_j^{(n)}(ia)^j u^{n-1-j}, \tag{14}$$

where the coefficients $d_j^{(n)}$ satisfy

$$\sum_{k=0}^{n} (-1)^k d_{2p+1-k}^{(n)} \binom{n}{k} = 0, \quad 0 \leqslant p < n-1; \tag{15a}$$

$$= (-1)^p, \quad p = n-1. \tag{15b}$$

III. WEAKLY RELATIVISTIC PDF'S

The PDF's appropriate to weakly relativistic magnetized thermal plasmas are the Shkarofsky functions, which are defined[2]

$$\mathcal{F}_q(z,a) = -i\int_0^{\infty} \frac{dt}{(1-it)^q} \exp\left[izt - \frac{at^2}{1-it}\right]. \tag{16}$$

The closely related Dnestrovskii Functions $F_q(z)$ satisfy $F_q(z) = \mathcal{F}_q(z,0)$ and have the alternative definition[2]

$$F_q(z) = \frac{1}{\Gamma(q)} \int_0^{\infty} du \frac{u^{q-1}e^{-u}}{u+z}. \tag{17}$$

These functions are the standard dispersion functions for thermal magnetized plasmas, and comprise the appropriate weakly relativistic generalizations of the Z function. A major barrier to the use of these functions in analytic work, and sometimes in computations, has been the general scarcity of analytic results bearing on their properties. The existing known properties of the Dnestrovskii and Shkarofsky functions have been discussed in Refs. 2 and 3. Here, we obtain several new results of interest in providing alternative expansion methods, tests of numerical calculations, analytically tractable approximations that preserve physical wave properties of interest, and new relationships with standard transcendental functions.

The Dnestrovskii functions of most interest in plasma physics are those of half-integer and integer order q. These functions can be written in terms of the PDF's $Z^{(n)}$ if we set $u = x^2$ in (17) and expand the numerator of the integrand in Hermite polynomials. These steps yield the expansion [cf. Eq. (47) of Ref. 2]

$$F_q(z) = \frac{1}{\Gamma(q)} \sum_{n=0}^{2q-1} c_n \int_{-\infty}^{\infty} dx \frac{H_n(x)e^{-x^2}}{x^2+z}, \tag{18a}$$

$$= \frac{\pi^{1/2}}{iz^{1/2}\Gamma(q)} \sum_{n=0}^{2q-1} (-1)^n c_n Z^{(n)}(iz^{1/2}), \tag{18b}$$

with

$$c_n = \frac{1}{2^n n! \pi^{1/2}} \int_{-\infty}^{\infty} dx\, x^{2q-1} H_n(x) e^{-x^2};$$

$$= \frac{(-1)^r}{(2r)!\pi} \Gamma(q) \Gamma\left(r+\frac{1}{2}\right) F\left(-r,q;\frac{1}{2};1\right),$$

$$n = 2r,\ 2q-1 \text{ even};$$

$$= \frac{(-1)^r 2}{(2r+1)!\pi} \Gamma\left(q+\frac{1}{2}\right) \Gamma\left(r+\frac{3}{2}\right)$$

$$\times F\left(-r,q+\frac{1}{2};\frac{3}{2};1\right),$$

$$n = 2r+1, 2q-1 \text{ odd};$$

$$= 0, \quad \text{otherwise};$$

where r is an integer and $F(a,b;c;1)$ is a hypergeometric function of unit argument (Ref. 8, Eqs. 7.376.2 and 7.376.3).

When calculating the optical depth for cyclotron absorption in plasmas, the integral

$$\int_{-\infty}^{\infty} dz\, \mathrm{Im}\, F_q(z) = -\pi, \tag{19a}$$

is encountered.[13] The form (18a) is particularly convenient to demonstrate the related result

$$\int_{-\infty}^{\infty} dz\, \mathrm{Re}\, F_q(z) = 0, \tag{19b}$$

which arises in mode-coupling problems in inhomogeneous plasmas, where z corresponds to the distance along the trajectory of the coupled waves. The identity[2]

$$\mathcal{F}_q(z,a) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} F_{q+k}(z-a), \tag{20}$$

implies that these results are also valid for $\mathcal{F}_q(z,a)$, provided a is real. Futhermore, we note that (19a) and (19b) hold for the dispersion functions of Sec. II, with (19a) being related to the normalization of the velcoity distribution in the unmagnetized-plasma case.[5]

In general, plasma dispersion properties involve singular integrals over the distribution function, with the forms (1), (10), and (17) providing specific examples of this principle. For general distributions, it would be computationally useful to obtain forms in which the singularity was either absent, or of a weaker form, but it is unclear whether this can be done. Here, we explore whether such forms can be obtained even in principle, using the specific example of the Shkarofsky functions, which arise from singular integrals over a weakly relativistic Maxwellian distribution.

Equations (80) and (86) of Ref. 2 yield the following relation:

$$\mathcal{F}_{j+3/2}(z,a) = \frac{2}{\pi^{1/2}\Gamma(j+1)} \int_0^{\infty} dy\, y^{2j+1} e^{-y^2} \times \int_{-\infty}^{\infty} dx \frac{e^{-x^2}}{x^2+y^2-2a^{1/2}x+z}. \tag{21}$$

If we treat x and y in (21) as the components of a three-dimensional vector parallel and perpendicular to some axis, respectively, and transform to spherical coordinates, we find

$$\mathcal{F}_{j+3/2}(z,a) = \frac{2}{\pi^{1/2}\Gamma(j+1)} \int_0^{\infty} du\, u^{2j+2} e^{-u^2} \times \int_{-1}^{1} du \frac{(1-\mu^2)^j}{z+u^2-2a^{1/2}u\mu}. \tag{22}$$

In the special case $j=0$, (22) implies

$$\mathcal{F}_{3/2}(z,a) = \frac{1}{\pi^{1/2}a^{1/2}} \int_0^{\infty} du\, u e^{-u^2} \times \ln\left[\frac{z+u^2+2a^{1/2}u}{z+u^2-2a^{1/2}u}\right], \tag{23}$$

where this integral involving the logarithm function does not appear to have been previously written down. More generally, a Legendre-polynomial expansion of $(1-\mu^2)^j$ in the μ integrand yields

$$\mathcal{F}_{j+3/2}(z,a) = \frac{2}{a^{1/2}\pi^{1/2}\Gamma(j+1)} \sum_{r=0}^{j} c_{2r}^{(j)} \times \int_0^{\infty} du\, u^{2j+1} e^{-u^2} Q_{2r}\left(\frac{z+u^2}{2a^{1/2}u}\right), \tag{24}$$

where Q_n is a Legendre function of the second kind and (Ref. 8, Eq. 8.14.16)

$$c_{2r}^{(j)} = \left(2r+\frac{1}{2}\right) \int_{-1}^{1} du (1-\mu^2)^j P_{2r}(\mu) = \frac{\pi(j!)^2}{[r!\Gamma(j+r+\frac{3}{2})\Gamma(j-r+1)\Gamma(\frac{1}{2}-r)]}. \tag{25}$$

Equation (24) provides a form of $\mathcal{F}_q$ in which the singularity has been weakened from a pole to a logarithmic form, leading to the conclusion that a similar approach may be viable in more general cases. Additionally, (24) can be used to express the new integrals of Q_{2r} that appear on the right, in terms of the functions $\mathcal{F}_q$, for whose evaluation algorithms exist.

The Shkarofsky functions can be related to the modified Bessel functions by considering the identity (Ref. 14, Eq. 3.471.9)

$$\int_0^{\infty} dx\, x^{\nu-1} \exp\left(-\gamma x - \frac{\beta}{x}\right) = 2\left(\frac{\beta}{\gamma}\right)^{\nu/2} K_{\nu}[2(\beta\gamma)^{1/2}], \tag{26}$$

which is valid for $\mathrm{Re}\,\beta > 0$ and $\mathrm{Re}\,\gamma > 0$. Comparison of this result with the expression [Ref. 2, Eq. (26)]

$$\mathcal{F}_q(z,a) = e^{z-2a} \int_1^{\infty} dt\, t^{-q} \exp\left((a-z)t + \frac{a}{t}\right),$$

yields

$$2(\beta/\gamma)^{\nu/2} K_{\nu}[2(\beta\gamma)^{1/2}] e^{\beta+\gamma} = \mathcal{F}_{1+\nu}(\beta-\gamma,-\gamma) + \mathcal{F}_{1-\nu}(\gamma-\beta,-\beta), \tag{27}$$

which can be extended to other values of β and γ by analytic

continuation. Special cases of (27) include the following:

$$(\pi/\beta)^{1/2} = \mathcal{F}_{3/2}(0,-\beta) + \mathcal{F}_{1/2}(0,-\beta), \tag{28a}$$

$$K_0(2\beta)\, e^{2\beta} = \mathcal{F}_1(0,-\beta), \tag{28b}$$

$$\Gamma(q-1)e^{-a}(-a)^{1-q} = \mathcal{F}_q(a,a) - F_{2-q}(-a), \tag{28c}$$

where Eq. (28c) reproduces Eq. (45) of Ref. 2 after correction of a typographical error in the earlier result. Substitution of (20) into Eqs. (27) and (28a)–(28c) then enables a variety of infinite series to be evaluated in terms of the modified Bessel functions K_ν. In addition to their analytic applications, results such as (27) and (28a)–(28c) are useful in checking the accuracy of numerical calculations of the dispersion functions.

In both analytic and numerical work it is often of interest to approximate the Shkarofsky functions in order to obtain semiquantitative results for the dispersion of specific types of waves (e.g., Bernstein waves). One such approximation, in terms of derivatives of the Z function, was obtained by Maroli and Petrillo.[7] This approximation is obtained by writing the factor $\exp[-at^2/(1-it)](1-it)^{-q}$ in (16) as $\exp[iqt - qt^2/2 - at^2 + S(t)]$, where $S(t)$ is cubic in t, and then expanding $\exp[S(t)]$ in powers of t. Use of Eq. (3) then yields

$$\begin{aligned}\mathcal{F}_q(z,a) = & -hZ(\psi) - (3a+q)\frac{h^4}{3}Z^{(3)}(\psi) \\ & - (4a+q)\frac{h^5}{4}Z^{(4)}(\psi) \\ & - (5a+q)\frac{h^6}{5}Z^{(5)}(\psi) \\ & - [3(6a+q) + 2(3a+q)^2]\frac{h^7}{18}Z^{(6)}(\psi) \\ & + O(q^{-2},a^{-2}),\end{aligned} \tag{29}$$

with $h = (4a+2q)^{-1/2}$ and $\psi = (z+q)h$; Eq. (29) includes corrections to the result of Ref. 7. Comparison of the approximation (29) for $\mathcal{F}_q(z,a)$ with that for $F_q(z)$ shows that $\mathcal{F}_q(z,a) \approx F_{q+2a}(z)$.

One disadvantage of the approximation (29) is that it does not preserve the known property that $\operatorname{Im} \mathcal{F}_q(z,a) = 0$ for $z > a$. This corresponds to the existence of weak damping of waves that should be strictly undamped, on physical grounds. Furthermore, the analytic properties of the Riemann surface of $\mathcal{F}_q(z,a)$ are not retained intact ($\mathcal{F}_q$ has a multisheeted Riemann surface,[2] whereas that of $Z^{(n)}$ is single sheeted). Preservation of the form of the Riemann surface is essental in studies of the properties of wave modes that pass from one sheet to another[15–17] and, hence, it is of interest to find approximate forms of the dispersion functions that preserve this property. An alternative approximation, which does preserve the above properties, can be obtained from (16) by writing

$$(1-it)^{-q} = (1-it)^{n-q}\exp[int - nt^2/(1-it) + T(t)], \tag{30}$$

where n is an integer less than q. Retention of terms up to order t^2 in $T(t)$ yields

$$\mathcal{F}_q(z,a) \approx \mathcal{F}_{q-n}(z+n,a+n), \tag{31}$$

which enables $\mathcal{F}_q$ to be approximated in terms of a lower-order Shkarofsky function, while preserving the relevant analytic properties of the original function. Refinements of the approximation (31) may be obtained by expanding $\exp[T(t)]$ in (30) to higher order in t; this approximation becomes more accurate but less advantageous as n is reduced and the order $q-n$ of the approximating function increases. If n is nonintegral, the property $\operatorname{Im} \mathcal{F}_q(z,a) = 0$ for $z > a$ is preserved, but the form of the Riemann surface is altered.[2]

IV. SUMMARY

We have presented a number of new results for the properties of nonrelativistic and weakly relativistic PDF's. These results include series, integrals, recursion relations, symmetry properties, approximations, interrelation, and connections with standard transcendental functions. Moreover, new integrals involving standard transcendental functions were also obtained. These results add significantly to the set of known properties of these PDF's, and are expected to be useful in both analytic and numerical work involving the dispersion and absorption of waves in plasmas. In particular, the results given here are relevant to systematic calculation of the dielectric properties of Maxwellian, generalized-Lorentzian, and more general plasmas, evaluation of dispersion functions, cyclotron absorption and mode-coupling problems, analytical approximation of dispersion functions in ways that preserve physical wave properties of interest, and the provision of tests of numerical calculations.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation's Solar-Terrestrial Division under Grants No. ATM-8511906 and ATM-8813255, and by NASA's Solar Terrestrial Theory and Solar Physics Programs under Grants No. NAGW-91 and NSG-7287, respectively, to the University of Colorado at Boulder.

[1] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
[2] P. A. Robinson, J. Math. Phys. **27**, 1206 (1986).
[3] P. A. Robinson, J. Math. Phys. **28**, 1203 (1987).
[4] W. Gautschi, SIAM Rev. **9**, 24 (1967).
[5] P. A. Robinson and D. L. Newman, J. Plasma Phys. **40**, 553 (1988).
[6] P. A. Robinson, J. Comput. Phys. (in press).
[7] C. Maroli and V. Petrillo, Phys. Scr. **24**, 955 (1981).
[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathametical Functions* (Dover, New York, 1970).
[9] W. Gautschi, Math. Comput. **15**, 227 (1961).
[10] M. F. Thomsen, H. C. Barr, S. P. Gary, W. C. Feldman, and T. E. Cole, J. Geophys. Res. **88**, 3035 (1983).
[11] I. H. Cairns, Bull. Am. Phys. Soc. **33**, 2076 (1988).
[12] I. H. Cairns, private communication.
[13] M. Bornatici, R. Cano, O. De Barbieri, and F. Engelmann, Nucl. Fusion **23**, 1153 (1983).
[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).
[15] P. L. Pritchett, Geophys. Res. Lett. **11**, 143 (1984).
[16] R. J. Strangeway, J. Geophys. Res. **91**, 3152 (1986).
[17] P. A. Robinson, J. Plasma Phys. **37**, 149 (1987).

Topology of parametrized systems

Peter Hajicek
Institute for Theoretical Physics, University of Berne, Sidlerstrasse 5, CH-3012 Bern, Switzerland

(Received 19 July 1988; accepted for publication 28 June 1989)

The discussion is limited to finite-dimensional parametrized systems with one constraint—the so-called super-Hamiltonian. The topology of the corresponding Hamiltonian vector field on the constraint hypersurface, in particular the question of the existence of cross sections, is studied. This has some bearing on the applicability of the reduction method, as well as on unitarity within the Dirac method, of canonical quantization [see the previous paper, Phys. Rev. D **34**, 1040 (1986)]. The main theorem of the present paper states that a cross section will exist if and only if the following conditions are both satisfied: (i) no dynamical trajectory is closed or almost closed; and (ii) the quotient set topology of the set of all dynamical trajectories is Hausdorff. Examples of systems that separately violate each of these conditions are given.

I. INTRODUCTION AND SUMMARY

Finite-dimensional parametrized systems, which will be studied in this paper, have some features in common with gravity theories. Quantization of these systems can, therefore, give hints of how some typical problems of quantum gravity could be solved.

With this motivation, canonical quantization of the parametrized systems has been investigated in Refs. 1 and 2. We have found that some special topological conditions on the system have to be satisfied if the reduction method is to work at all. It has also turned out that similar conditions are sufficient for the Dirac method to lead to unitary evolution.

The topology in consideration is that of the set of *all* classical dynamical trajectories of the system: there must be a function T on the phase space of the system such that T increases along each trajectory. This T has been called "global phase time." We have also found that such a function will exist iff (i) there is a hypersurface that intersects each trajectory in exactly one point, and (ii) there are no pointlike or closed trajectories.

In the present paper, we study these properties by rigorous mathematical methods. The tools we will use are borrowed from the theory of global structure of space-times (for a representative review, see Ref. 3), from the theory of differentiable manifolds as it is given, e.g., in Ref. 4, and from that of fiber bundles.[5]

The program of the paper is as follows. First, we give a rigorous definition of finite-dimensional parametrized systems with one constraint; then we describe some of their properties and summarize the most important of our results. In Sec. II, we give all proofs, lemmas, and auxiliary definitions step by step. Finally, in Sec. III the relevant properties of parametrized systems are illustrated by means of examples.

Let $(\mathscr{P},\Omega)$ be a symplectic $2N$-manifold endowed with the symplectic form Ω (see, e.g., Ref. 6). The action of a finite-dimensional parametrized system is given by

$$S=\int_{\gamma}(\Theta-\alpha\mathscr{H}\,dt).$$

Here, $\tilde{\gamma}$ is a curve in the space $\mathscr{P}\times\mathbb{R}^1$,

$$\tilde{\gamma}\colon\ (a,b)\to\mathscr{P}\times\mathbb{R}^1,$$

with components

$$\gamma\colon\ (a,b)\to\mathscr{P}$$

and

$$\mathrm{id}\colon\ (a,b)\to\mathbb{R}^1;$$

Θ satisfies, locally, the equation

$$\Omega=d\Theta;$$

α is a function on γ,

$$\alpha\colon\ (a,b)\to\mathbb{R}^1;$$

and $\mathscr{H}$ is a function on $\mathscr{P}$

$$\mathscr{H}\colon\ \mathscr{P}\to\mathbb{R}^1;$$

where α is called the *Lagrange muliplier* and $\mathscr{H}$ the *super-Hamiltonian*. The dynamics of the system is invariant with respect to the change of the super-Hamiltonian:

$$\mathscr{H}\to\Lambda\mathscr{H}, \tag{1}$$

where Λ is a positive differentiable function, $\Lambda\colon\mathscr{P}\to(0,\infty)$, with open domain containing all zero points of $\mathscr{H}$. Let us denote the corresponding class of super-Hamiltonians by $\{\mathscr{H}\}$.

The Euler–Lagrange equations that follow from the action S are, in a coordinate charge $(\mathscr{U},z)$,

$$\dot{z}^{\mu}=\alpha X^{\mu},\quad \mu=1,\ldots,2N, \tag{2}$$

$$\mathscr{H}=0, \tag{3}$$

where X^{μ} is the so-called *super-Hamiltonian vector field* defined by

$$X^{\mu}=\Omega^{\mu\nu}\partial_{\nu}\mathscr{H}.$$

The transformation (1) leads to the rescaling

$$X^{\mu}\to\Lambda X^{\mu}$$

of the super-Hamiltonian vector field.

We assume that (a) there is only one constraint $\mathscr{H}$, and that (b) $\mathscr{H}$ is differentiable everywhere on $\mathscr{P}$. Assumption (a) is a simplification justified solely as a first contact with the problem; multiplicity of constraints could possibly lead

 0022-2488/89/112488-10$02.50

to qualitatively new features. Assumption (b), however, does not seem to imply any important restriction of generality. Even if there exist physical systems such as cosmological models, whose super-Hamiltonian is not differentiable everywhere (for example, at the points corresponding to the "big bang" or "big crunch"), we can always cut away these points from $\mathscr{P}$.

We will call a point $p \in \mathscr{P}$ *critical* if Eq. (3) is satisfied at p together with

$$d\mathscr{H} = 0. \tag{4}$$

Critical points are often present, see, e.g., examples 1 and 2 in Sec. III. The set of all *noncritical* points of $\mathscr{P}$ that satisfy Eq. (3) will, in this paper, be called the *constraint hypersurface* and will be denoted by Γ [usually, the full solution to (3) is called the constraint hypersurface]. Indeed Γ is a hypersurface, at least locally, because Eq. (4) is violated at all points of Γ. However, if there are critical points, Γ need not be connected and/or closed. We will use the abbreviation n for the dimension of Γ ($n = 2N - 1$).

The antisymmetry of Ω implies that the super-Hamiltonian vector field X is tangent to Γ at any point of Γ; X also designates the degenerate direction for the pullback Ω_Γ of Ω to Γ:

$$X^\mu \Omega_{\Gamma\mu\nu} = 0.$$

A (finite-dimensional) *parametrized system* is the triple $(\mathscr{P},\Omega,\{\mathscr{H}\})$. Working with a parametrized system, we will always choose a particular representative $\mathscr{H}$ from the class $\{\mathscr{H}\}$. However, only those properties of $\mathscr{H}$ that are invariant with respect to the transformation (1) are interesting.

A maximal integral curve of X through a point $p \in \Gamma$ will be denoted by γ_p. It is a map

$$\gamma_p\colon J(p) \to \Gamma,$$

where $J(p)$ is an open interval containing 0 and $\gamma_p(0) = p$. For a definition of a maximal integral curve of X, see Ref. 4, p. 135, from where we have taken most of our notation. Clearly $J(p)$ depends on the choice of $\mathscr{H}$. The set $\gamma_p(J(p))$ is called an *orbit* through p. Orbits are independent of $\mathscr{H}$.

There is a formal relation between the parametrized and the Hamiltonian systems of the symplectic mechanics (see Ref. 6). A Hamiltonian system is a triple $(\mathscr{P},\Omega,\mathscr{H})$, where $\mathscr{H}$ is the energy. Dynamical trajectories of the Hamiltonian system lie on the surfaces $\mathscr{H} = \text{const}$, and possible values of the energy form some set, usually $(0,\infty)$. Thus a parametrized system can *formally* be considered as a *class* of Hamiltonian systems for each of which only the value zero of energy is allowed. The relation is formal, because the phase space of parametrized systems also contains the "time dimension."

The topological problems of Hamiltonian systems (perturbations around critical elements of phase portraits, stability) are rather different from those of parametrized systems: as the latter are a sort of gauge theory, the existence of *cross sections* through "gauge trajectories" is important. Thus our main problem will be, "under which conditions does a system of orbits admit a cross section."

Definition 1: A cross section is a hypersurface Σ (without boundary) in Γ satisfying the following condition: for each $p \in \Gamma$, there is a unique $t(p) \in J(p)$ such that

$$\gamma_p(t(p)) \in \Sigma.$$

This definition of the cross section is different from that used in Ref. 2. Here, we require that there be exactly one value of the *parameter* along any maximal integral curve at which the curve intersects the cross section Σ. In Ref. 2, we have required that the orbits intersect Σ in exactly one *point* of Γ. Thus the two notions are equivalent only if each γ_p is an injection, and this is true only if there are no pointlike or closed maximal integral curves (critical points or cycles of X).

We are going to derive the existence of such cross sections from properties that either have more direct physical meaning for the parametrized systems or are more immediately recognizable. In Sec. II, we will find two such properties: namely, that the parametrized system is *strongly nonrepetitive and Hausdorff*. Let us define these properties

Definition 2: A point $p \in \Gamma$ is called strongly nonrepetitive if each neighborhood $\mathscr{U}_1$ of p in Γ has a subneighborhood $\mathscr{U}_2$ with the following property: for each $q \in \mathscr{U}_2$, $\gamma_q^{-1}(\gamma_q(J(q)) \cap \mathscr{U}_2)$ is a connected subset of $J(q)$. A point p is called *repetitive* if the orbit $\gamma_p(J(p))$ is compact. The system $(\mathscr{P},\Omega,\{\mathscr{H}\})$ itself is called (strongly) nonrepetitive, if each point of Γ is (strongly) nonrepetitive.

Thus a strongly nonrepetitive point in nonrepetitive; the property implies absence of closed (compact) orbits, but it is stronger than that: "almost closed"orbits are also excluded. The property is very similar to that of "strong causality" for space-times (cf. Ref. 3). It has, however, *nothing* to do with causal properties of *space-times*. If the dynamical trajectories of our system are to have the meaning of a classical space-time geometry, then all these space-times must be globally hyperbolic,[2] so they will necessarily not only be causal, but stably causal. The notion of a weakly repetitive (the converse of "strongly nonrepetitive") point is also related to that of a "nonwandering" or "recurrent" point of the topological dynamics (see Ref. 6). We prefer another name, however, because we want to emphasize that the "return" includes a return *of the same time*, in the case of parametrized systems.

Examples of repetitive systems are given in Ref. 1. Among these, there are some cosmological models. The meaning of closed orbits in phase space is also illustrated in Ref. 1. Example 1 of Sec. III gives a parametrized system that is nonrepetitive but weakly repetitive and not Hausdorff: all orbits are "almost closed."

Definition 3: An equivalence relation "$\equiv$" on Γ is defined as follows: let $p \equiv q$ if p and q lie at the same orbit. A system $(\mathscr{P},\Omega,\{\mathscr{H}\})$ is called Hausdorff if the quotient set topology with respect to the relation "$\equiv$" is Hausdorff.

We will denote the quotient space with its quotient set topology by $\mathscr{Q}$ and the natural projection by π [$\pi(p) = \gamma_p(J(p))$, for any $p \in \Gamma$]. Example 2 of Sec. III describes a system that is strongly nonrepetitive but not Hausdorff. In Ref. 1, some examples with closed orbits are given; these systems are, therefore, strongly nonrepetitive, but Hausdorff. Constructing an example of a nonrepetitive Hausdorff system that is weakly repetitive or proving that

each nonrepetitive Hausdorff system must be strongly nonrepetitive remains an open problem.

The main theorems of Sec. II imply that these two properties are necessary and sufficient for the existence of cross sections.

We will also show the following. If there is a cross section Σ, then Γ is diffeomorphic to the set $\Sigma' \times \mathbb{R}^1$, where Σ' is a hypersurface in Γ, and the corresponding diffeo Ψ,

$$\Psi: \Gamma \to \Sigma' \times \mathbb{R}^1,$$

maps each maximal integral curve on a set of the form $(A \times \mathbb{R}^1)$ for some $A \in \Sigma'$. Hence the function T,

$$T: \Gamma \to \mathbb{R}^1,$$

defined by

$$T = \pi_2 \circ \Psi,$$

where π_2 is the natural projection,

$$\pi_2: \Sigma \times \mathbb{R}^1 \to \mathbb{R}^1,$$

is a global phase time. The definition of the cross section as given in the present paper is so strong that the existence of a cross section implies that of a global time function.

Finally, a simple criterion for nonexistence of a cross section will be given: Each cross section must be diffeomorphic to the quotient manifold $\mathcal{Q}$. Thus if $\mathcal{Q}$ has a bad topology, there will be no cross section. This criterion will be used in example 2.

It is important to notice that all our proofs hold for any differentiable vector field X on any differentiable manifold Γ. However, the assumption that the pair (Γ, X) originates from a parametrized system cannot replace any of the premises of the lemmas and theorems of Sec. II. This is, in fact, shown by the examples of Sec. III.

Moreover, these examples seem to lead to some interesting conjectures about the general case. Thus the pathology in example 1 is entirely caused by identifications in the "time direction" y, say, of the phase space. Taking a covering space, i.e., introducing suitable angular coordinates, we can make the system strongly nonrepetitive. Such a method might be quite generally applicable. The rigorous procedure could be analogous to that given in Ref. 2 for removal of closed orbits.

So far, we can say something about topology only after having solved the dynamical equations and investigated the totality of the integral curves. Example 2 seems to suggest that this may be superfluous in some cases and that a closer study of small neighborhoods of critical points or cycles may yield useful information about the topology of the system. Here, one could use the well-known results on Hamiltonian systems.[6]

Nevertheless, I have to admit that the examples of Sec. III are somewhat artificial from the physical point of view. Their super-Hamiltonians are simpler than those of various cosmological models, and they are constructed just with the aim of producing a complicated topology. Thus one is tempted to ask whether general relativity and the models derived from it can have a similar structure at all. However, in order to answer this question and to understand cosmological models, we have to develop methods that would not require a general solution to their dynamical equations, for example, such as those mentioned above.

II. LEMMAS AND PROOFS

In this section, we will assume that Γ is connected. If it is not, then all claims of the present section hold only for components of Γ. This Γ is a paracompact manifold (that is, Γ is Hausdorff and has a countable atlas, see Ref. 4, p. 51). This follows from critical points forming a closed subset of $\mathcal{P}$, and from Proposition 2.5.1 of Ref. 4.

One of the important questions is that of the completeness of the vector field X on Γ, see Ref. 4, p. 138. We will call the super-Hamiltonian $\mathcal{H}$ *complete* if the super-Hamiltonian vector field X is complete on Γ. It turns out that one can often rescale the super-Hamiltonian so that it becomes complete. This is the content of the following lemma.

Lemma 1: Let the system $(\mathcal{P}, \Omega, \{\mathcal{H}\})$ be strongly nonrepetitive. Then there is a super-Hamiltonian $\mathcal{H} \in \{\mathcal{H}\}$ that is complete.

Proof: Lemma 1 is analogous to the theorem of Clarke[7] for null geodesic completeness on strongly causal space-times, and the corresponding proof can easily be adapted to our case. Clarke's proof is based on two lemmas that hold for space-times. We first show analogous lemmas for Γ, and then leave the (trivial) modification of Clarke's proof to the reader.

Lemma 1.1: Let $\mathcal{K}$ be a compact subset of Γ. If γ is a maximal integral curve of X for some $\mathcal{H} \in \{\mathcal{H}\}$, then there is a value t_0 of its parameter t such that $\gamma(t) \notin \mathcal{K}$ for all $t > t_0$.

Proof of Lemma 1.1: This is a simple modification of the proof of Proposition 6.4.7, p. 195 in Ref. 3.

Lemma 1.2: Let X' be the super-Hamiltonian vector field on Γ corresponding to the super-Hamiltonian $\mathcal{H}' = \Lambda \mathcal{H}$. Let γ_p and γ'_p be integral curves of X and X', respectively, that both start at the same point $p \in \Gamma$. Then

$$\gamma'(t) = \gamma(f(t)),$$

where

$$f(t) = \int_0^t d\tau \, \Lambda(\gamma(\tau)).$$

Proof of Lemma 1.2: This lemma is evident.

An important tool in many proofs of the present section will be the *flow* of the vector field X on the Hausdorff manifold Γ. A definition and some properties of the flow are described in Ref. 4, Sec. 8.3. We briefly reproduce the main points here in order to keep the paper self-contained.

Let γ_p be the maximal integral curve through p and $J(p)$ be its domain. Let $\mathcal{D}$ denote the set of points (t,p) of $\mathbb{R}^1 \times \Gamma$ such that $t \in J(p)$. The function

$$\Phi: \mathbb{R}^1 \times \Gamma \to \Gamma$$

with an open domain $\mathcal{D}$ defined by

$$(t,p) \to \gamma_p(t)$$

ia called the flow of the vector field X (see Ref. 4, p. 139).

The flow Φ of X is a differentiable function, because X is a differentiable vector field and Γ is Hausdorff (see Ref. 4, Proposition 8.3.1, p. 140).

For each point $p \in \Gamma$, there is a chart $(\mathcal{U}, x)$ around p

such that

$$X = \frac{\partial}{\partial x^n}, \tag{5}$$

on $\mathcal{U}$, where $x^1,...,x^n$ are the coordinate functions corresponding to the chart (see Ref. 4, Proposition 8.3.2, p. 140).

Definition 4: The chart $(\mathcal{U},x)$ is called an "*X-adapted chart around p.*"

If $\mathcal{H}$ is complete, then the super-Hamiltonian vector field X defines a one-dimensional diffeomorphism transformation group $\{\varphi_t\}$ on Γ by

$$\varphi_t(p) = \Phi(t,p), \tag{6}$$

for any t and p (see Ref. 4, example 13.1.1, p. 241).

Another crucial notion will be that of a *partial cross section* Σ. These (partial) cross sections will play a role similar to that of (partial) Cauchy hypersurfaces in space-times (see Ref. 3).

Definition 5: Let Σ be a hypersurface in Γ and γ be a maximal integral curve of X with domain J. We say that γ intersects Σ *once*, if there is a unique $t\in J$ such that $\gamma(t)\in\Sigma$. Then Σ is called a *partial cross section* if each maximal integral curve of X intersects Σ at most once. A partial cross section Σ is called *regular* if the vector field X is nowhere tangent to Σ.

Thus the cross section of Definition 1 is a partial cross section intersected by *all* maximal integral curves of X. Clearly, if Σ is a (regular) cross section with respect to one super-Hamiltonian, then it is so with respect to any.

Lemma 2: There is a partial cross section Σ through a point $p\in\Gamma$ if and only if p is strongly nonrepetitive.

Proof: Let $(\mathcal{U},x)$ be an X-adapted chart around p and let Σ_p be defined by

$$\Sigma_p = \{q\in\mathcal{U} \mid x^n(q) = x^n(p)\}. \tag{7}$$

Then Σ_p is a hypersurface in $\mathcal{U}$ containing p.

Suppose that p is strongly nonrepetitive, and let $\mathcal{V}$ be a neighborhood of p intersected by each maximal integral curve of X at most once. Then $\Sigma_p\cap\mathcal{V}$ is intersected by each such curve at most once, and so $\Sigma_p\cap\mathcal{V}$ is a partial cross section.

Suppose that Σ is a partial cross section through p. As $\Sigma\cap\mathcal{U}$ is also a partial cross section, each maximal integral curve of X intersects $\Sigma\cap\mathcal{U}$ at most once. Define $\mathcal{V}$ as follows: $\mathcal{V}$ consists of all points q of $\mathcal{U}$ such that $\gamma_q\cap\Sigma\cap\mathcal{U}\neq\emptyset$ and q lies in the component of $\gamma_q\cap\mathcal{U}$, which contains $\gamma_q\cap\Sigma$. Then $\mathcal{V}$ is an open subset of $\mathcal{U}$ containing p. Each maximal integral curve of X intersects $\mathcal{V}$ at most once; thus p is strongly nonrepetitive. Q.E.D.

Two partial cross sections define the following very useful mappings. Let Σ_1 and Σ_2 be two partial cross sections. We define the "(Σ_1,Σ_2) projection Π,"

$$\Pi\colon \Sigma_1\to\Sigma_2,$$

by

$$\Pi(p) = \gamma_p(J(p))\cap\Sigma_2,$$

for $p\in\Sigma_1$. For each $p\in\Sigma_1$, there is at most one $\Pi(p)$, and the domain $\mathcal{U}\in\Sigma_1$ of Π is open, as $\mathcal{D}$ is open ($\mathcal{D}$ is the domain of the flow of X). Clearly Π has an inverse, defined on $\Pi(\mathcal{U})$, that is the (Σ_2,Σ_1) projection. Furthermore, we define the "(Σ_1,Σ_2) shift f,"

$$f\colon \Sigma_1\to\mathbb{R}^1,$$

by

$$f(p) = t(\Pi(p)),$$

where $t(q)$ is the value of the parameter t along γ_p at the point q (q must lie on γ_p). The domain of f is $\mathcal{U}$.

We have the following transitivity properties of projections and shifts. Let Σ_1, Σ_2, and Σ_3 be three partial cross sections; Π_{12}, Π_{23}, and Π_{13} be the corresponding projections; and f_{12}, f_{23}, and f_{13} be the corresponding shifts. Then

$$\Pi_{13} = \Pi_{23}\circ\Pi_{12}, \quad f_{13} = f_{12} + f_{23}\circ\Pi_{12}.$$

Thus the function f_{21}, defined by

$$f_{21} = -f_{12}\circ\Pi_{12}^{-1},$$

is nothing but the (Σ_2,Σ_1) shift. The projections and shifts between two partial cross sections have important differentiability properties described by Lemmas 3 and 5 to follow.

Lemma 3: Let $(\mathcal{U},x)$ be an X-adapted chart around $p\in\Gamma$ and let Σ_p be the hypersurface in $\mathcal{U}$ defined by (7). Let Σ be a partial cross section through p. Then there is a neighborhood V of p in Σ_p such that (i) V is a partial cross section, (ii) the (V,Σ) shift f is continuous on V, and (iii) the (V,Σ) projection Π is a continuous bijection and Π^{-1} is differentiable on $\Pi(V)$. Moreover, the following three statements are equivalent to each other: (a) $\Pi(V)$ is a regular partial cross section, (b) f is differentiable on V, and (c) Π is a diffeo on V.

Proof: We decompose the proof into several steps.

(i) The existence of a partial cross-section $V\subset\Sigma_p$ follows from Lemma 2 and its proof. If V is a partial cross section, then f and Π are well defined and Π must be a bijection.

(ii) We show that Π and f are continuous. For any $q\in\Sigma_p$, we have

$$f(q) = x^n(\Pi(q)) - x^n(p);$$

this follows from relation (5). Thus f is continuous if Π is, because the function x^n is differentiable on $\mathcal{U}$ and so on Σ.

Let $q_1\in V$, $q_2 = \Pi(q_1)$, and let V_2 be any neighborhood of q_2 in $\Sigma\cap\mathcal{U}$. Then there is a neighborhood $\mathcal{U}_2$ of q_2 in $\mathcal{U}$ such that $\mathcal{U}_2\cap\Sigma = V_2$. This follows from Σ being a hypersurface: the topology on Σ is the subset topology.

We define $q_3\in\mathbb{R}^{n-1}$ and $q_4\in\mathbb{R}^1$ as follows:

$$q_3 = (x^1(q_2),...,x^{n-1}(q_2)), \quad q_4 = x^n(q_2).$$

Let $D^k(a,b)$ be an open disk in $\mathbb{R}^k$ with center a and radius b. Then there is $\delta_1 > 0$ such that

$$x^{-1}(D^{n-1}(q_3,\delta_1)\times D^1(q_4,\delta_1))\subset\mathcal{U}.$$

We define the map $y\colon \Sigma_p\to\mathbb{R}^{n-1}$ by

$$y(q) = (x^1(q),...,x^{n-1}(q)),$$

for any $q\in\Sigma_p$. Then y is a diffeo with domain Σ_p. The range of y contains q_3 and so it contains the whole disk $D^{n-1}(q_3,\delta_2)$ around q_3 for some positive δ_2. Let $\delta = \min(\delta_1,\delta_2)$, and let a neighborhood V_1 of q_2 in V be defined as follows:

$$V_1 = y^{-1}(D^{n-1}(q_3,\delta)).$$

We have

$$\Pi(V_1)\subset V_2.$$

Since V_2 was arbitrary, we obtain that for any neighborhood V_2 of q_2 in $\Sigma\cap\mathcal{U}$, there is a neighborhood V_1 of q_1 in V such that $\Pi(V_1)\subset V_2$, and Π is continuous.

(iii) Π^{-1} is even differentiable. Indeed, the maps i: $\Sigma\to\Gamma$ and x: $\mathcal{U}\to\mathbb{R}^n$ are differentiable. We define a differentiable map π_{n-1}: $\mathbb{R}^n\to\mathbb{R}^{n-1}$ by

$$\pi(x^1,...,x^n)=(x^1,...,x^{n-1}).$$

We have

$$\Pi^{-1}=y^{-1}\circ\pi_{n-1}\circ x\circ i;$$

thus Π^{-1} is differentiable.

(iv) Suppose (a) is true. Then $\Sigma\cap\Pi(V)$ is a submanifold. Thus each point $q_2\in\Sigma\cap\Pi(V)$ has a neighborhood $\mathcal{U}_2$ in Γ such that $\Sigma\cap\mathcal{U}_2$ is defined by the equation

$$u(x^1,...,x^n)=0, \tag{8}$$

where u: $\mathbb{R}^n\to\mathbb{R}^1$ is a differentiable function. The regularity of Σ implies

$$X^\mu\frac{\partial u}{\partial x^\mu}=\frac{\partial u}{\partial x^n}\neq 0.$$

Thus there is a neighborhood V_2 of q_2 in Σ such that Eq. (8) can be solved for x^n in V^2:

$$x^n=F(x^1,...,x^{n-1}),$$

and F is differentiable in V_2 (cf. Ref. 4, p. 14). We have however,

$$f=F\circ y.$$

Hence f is differentiable at $\Pi^{-1}(q_2)$, and, as q_2 was an arbitrary point of $\Sigma\cap\Pi(V)$, statement (b) holds.

(v) Suppose (b) is true. We define the map Π': $\mathbb{R}^{n-1}\to\Sigma$ by

$$\Pi'(x',...,x^{n-1})=x^{-1}\circ(x^1,...,x^{n-1},F(x^1,...,x^{n-1})).$$

Since F is differentiable, Π' is differentiable and Π'_* maps the $(n-1)$ independent vectors δ^μ_α, $\alpha=1,...,n-1$, at a point $(x^1,...,x^{n-1})\in\mathbb{R}^{n-1}$ into the vectors

$$x^{-1}_*\circ(\delta^\mu_\alpha,F_*(\delta^\mu_\alpha))$$

at $x^{-1}\circ(x^1,...,x^{n-1},F(x^1,...,x^{n-1}))$, which are also linearly independent. Thus Π' is a diffeo. We have however,

$$\Pi=\Pi'\circ y;$$

thus Π is a diffeo, and statement (c) holds.

(vi) Suppose (c) is true. Then Π^{-1} is a diffeo with domain $\Pi(V)$. Let $q\in\Pi(V)$ and suppose that $X(q)$ is tangential to Σ at q. We have

$$\Pi^{-1}=y^{-1}\circ\pi_{n-1}\circ x\circ i$$

and

$$(\pi_{n-1}\circ x)_*X=0.$$

Thus

$$\Pi^{-1}_*(X(q))=0.$$

Then Π^{-1} cannot be a diffeo, and this is a contradiction. Q.E.D.

Lemma 4: Let $p\in\Gamma$ and $a\in J(p)$ [$J(p)$ is the domain of the maximal integral of X through p]. Then the map

$$\varphi_a\colon\Gamma\to\Gamma,$$

given by Eq. (6), is well-defined on a neighborhood $\mathcal{U}$ of p; moreover, φ_a is a diffeo on any open set on which it is well-defined.

Proof: The domain $\mathcal{D}$ of Φ: $\Gamma\times\mathbb{R}^1\to\Gamma$ is open. As the point (p,a) belongs to $\mathcal{D}$, there is a whole neighborhood $\mathcal{V}$ of (p,a) in $\mathcal{D}$ on which Φ is well defined. Let

$$\mathcal{U}=\{q\in\Gamma\mid(q,a)\in\mathcal{V}\}.$$

Then $\mathcal{U}$ is open and contains p. As Φ is well defined on $\mathcal{U}\times a$, φ_a is well defined on $\mathcal{U}$.

Let φ_a be well defined on an open set $\mathcal{U}_1$. Then φ_a is a composition of two differential maps on $\mathcal{U}_1$, namely,

$$\varphi_a(p)=\Phi(p,f(p)),$$

where f: $\mathcal{U}\to\mathbb{R}^{-1}$ is given by

$$f(p)=a,\quad\forall p\in\mathcal{U}.$$

Hence φ_a is differentiable on $\mathcal{U}_1$. Moreover, the map φ_{-a}, defined by Eq. (6), is clearly well defined on $\varphi_a(\mathcal{U}_1)$ and

$$\varphi_a\circ\varphi_{-a}=\mathrm{id}$$

on $\mathcal{U}_1$. Thus φ_a has an inverse on $\mathcal{U}_1$. This inverse is differentiable for the same reason as φ_a is. Then φ_a is a diffeo. Q.E.D.

Lemma 5: Let Σ_1 be a partial cross section, Σ_2 a hypersurface, and f: $\Sigma_1\to\mathbb{R}^1$ a function such that the map Π: $\Sigma_1\to\Sigma_2$, defined by

$$\Pi(p)=\gamma_p(f(p)),$$

has a domain Σ_1 and a range Σ_2.

Then Σ_2 is a partial cross section, Π is an open and continuous bijection, and f is continuous.

Moreover, let Σ_1 be regular. Then Π^{-1} is differentiable, and the following three statements are equivalent to each other: (a) Σ_2 is a regular partial cross section, (b) f is a differentiable, and (c) Π is a diffeo.

Proof: We decompose the proof into several steps.

(i) We show that Π is a bijection. Suppose that $p\in\Sigma_1$, and $q\in\Sigma_1$ such that $\Pi(p)=\Pi(q)$. Then $\gamma_p\cap\gamma_q\neq\emptyset$, so γ_p intersects Σ_1 in p and q. However, Σ_1 is a partial cross section, so $p=q$, and Π is a bijection.

(ii) We show that Σ_2 is a partial cross section. Suppose that there are $p\in\Sigma_2$ and $a>0$ such that $q=\gamma_p(a)$ lies on Σ_2. Then the points r and s at Σ_1, defined by

$$r=\Pi^{-1}(p),\quad s=\Pi^{-1}(q),$$

satisfy

$$p=\gamma_r(f(r)),\quad q=\gamma_s(f(s)).$$

Hence r, s, p, and q all lie at the same maximal integral curve of X (say, γ_r) and

$$r=\gamma_r(0),\quad p=\gamma_r(f(r)),$$
$$q=\gamma_r(f(r)+a),\quad s=\gamma_r(f(r)+a-f(s)).$$

Let $p=q$. Then $r=s$, $f(r)=f(s)$, and $s=\gamma(a)$; thus Σ_1 is not a partial cross section. Let $p\neq q$. Then γ_r intersects Σ_1 in two different points, r and s, and Σ_1 is not a partial cross section. This is a contradiction; thus Σ_2 must be a partial cross section.

(iii) Choose a point $p\in\Sigma_1$ and denote by q the point of Σ_2 defined by

$$q=\varphi_{f(p)}(p).$$

According to Lemma 4, $\varphi_{f(p)}$ is well defined in a neighborhood $\mathcal{U}_1$ of p and it is a diffeo in $\mathcal{U}_1$. Let us define Σ' by

$$\Sigma'=\varphi_{f(p)}(\Sigma_1\cap\mathcal{U}_1).$$

Then Σ' is a partial cross section by the same argument as Σ_2 is. The map Π': $\Sigma_1\to\Sigma'$, defined by

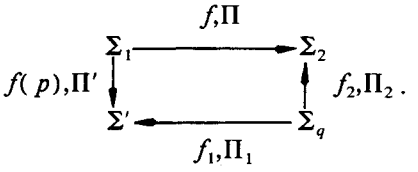
$$\Pi'=\varphi_{f(p)}|_{\Sigma_1\cap\mathcal{U}_1},$$

is, therefore, a diffeo with domain $\Sigma_1\cap\mathcal{U}_1$ and range Σ'.

Moreover, if Σ_1 is a regular partial cross section, then Σ' is. Indeed, $\varphi_{f(p)}$ is a diffeo that preserves X:

$$\varphi_{f(p)}\cdot X=X.$$

The various maps and their relations are illustrated in the following diagram:



(IV) Let $(\mathcal{U}_2,x)$ be an X-adapted coordinate neighborhood of q and Σ_q be the hypersurface in $\mathcal{U}_2$ defined by an equation analogous to (7). Apply Lemma 3 to the pair Σ_q and Σ'; let $\overline{V}$ be the corresponding neighborhood of q in Σ_q, f_1 the shift, and Π_1 the projection. Applying Lemma 3 to the pair Σ_q and Σ_2, we have the neighborhood $\tilde{V}$, the shift f_2, and the projection Π_2. Let $V=\overline{V}\cap\tilde{V}$, $V'=\Pi_1(V)$, $V_2=\Pi_2(V)$, and $V_1=\Pi'^{-1}(\Pi_1(V))$. We have

$$V_1\subset\mathcal{U}_1\cap\Sigma_1,$$

$$\Pi|_{V_1}=(\Pi_2\circ\Pi_1^{-1}\circ\Pi')|_{V_1} \tag{9}$$

where Π' is a diffeo, Π_1^{-1} is differentiable, and Π_2 is continuous. Thus $\Pi|_{V_1}$ is continuous. Similarly, Π^{-1} is continuous on V_1 because

$$\Pi^{-1}|_{\Pi(V_1)}=(\Pi'^{-1}\circ\Pi_1\circ\Pi_2^{-1})|_{\Pi(V_1)}, \tag{10}$$

and f is continuous on V_1, because

$$f=x^n\circ\Pi-x^n\circ\Pi'+f(p). \tag{11}$$

(V) Suppose that Σ_1 is regular. Then Σ' is regular. Applying Lemma 3 to the pair Σ_q and Σ', we obtain that f_1 is differentiable and Π_1 is a diffeo. Then Π^{-1} is differentiable on $\Pi(V_1)$ because of Eq. (10). Finally, applying Lemma 3 to the pair Σ_q and Σ_2, we obtain the last claim of Lemma 5, because of Eqs. (9) and (11). Q.E.D.

The shift and projection apparatus is now ready to help us in the proof the following two important theorems.

Theorem 1: Let Σ be a cross section. Then there is a regular cross section Σ'.

Proof: Each point p of Σ has an X-adapted coordinate neighborhood $(\mathcal{U}_p,x_p)$. Let Σ_p be the hypersurface in $\mathcal{U}_p$ defined by Eq. (7); Σ_p is a regular partial cross section for each p. Let $V_p\subset(\mathcal{U}_p\cap\Sigma)$ be defined as follows: $q\in\mathcal{U}_p\cap\Sigma$ lies in V_p, if γ_q intersects Σ_p. Then V_p is an open set containing p, and the collection $\{V_p\}$ of all $\{V_p$'s$\}$ cover Σ.

Σ is a regular submanifold of Γ with its induced differentiable structure. We can define another differentiable structure on Σ as follows. Consider a set V_p from $\{V_p\}$ and the corresponding X-adapted chart $(\mathcal{U}_p,x_p)$. The functions $x_p^1|V_p,\ldots,x_p^{n-1}|V_p$ are well defined on V_p and they define a map

$$h_p:V_p\to\mathbb{R}^{n-1}$$

by

$$h_p(q)=(x_p^1(q),\ldots,x_p^{n-1}(q)),$$

for any $q\in V_p$. This map is an injection: two different points, q_1 and q_2 from V_p, are mapped to two different $(n-1)$-tuples, because the first $(n-1)$ coordinates of $(\mathcal{U}_p,x_p)$ are constant along maximal integral curves of X, and each maximal integral curve of X intersects V_p at most once. $h_p(V_p)$ is open in $\mathbb{R}^{n-1}$, because Σ_p is. Thus (V_p,h_p) is a chart. The collection $\{(V_p,h_p)\}$ of these charts forms an atlas for Σ, if $h_p\circ h_q^{-1}$ is a diffeo for any two charts (V_p,h_p) and (V_q,h_q) that intersect each other. Consider the corresponding charts, $(\mathcal{U}_p,x_p)$ and $(\mathcal{U}_q,x_q)$. Let us define the diffeo σ_p: $\Sigma_p\to\mathbb{R}^{n-1}$ by

$$\sigma_p(r)=(x_p^1(r),\ldots,x_p^{n-1}(r)), \tag{12}$$

for each $r\in\Sigma_p$, and similarly for σ_q. Then, we clearly have

$$h_p=\sigma_p\circ\Pi_p,$$

where Π_p: $V_p\to\Sigma_p$ is the (V_p,Σ_p) projection. Let us calculate $h_p\circ h_q^{-1}$:

$$h_p\circ h_q^{-1}=\sigma_p\circ\Pi_p\circ\Pi_q^{-1}\circ\sigma_q^{-1}=\sigma_p\circ\Pi_{qp}\circ\sigma_q^{-1},$$

where Π_{qp} is the (Σ_q,Σ_p) projection. According to Lemma 5, Π_{qp} is a diffeo, so $h_p\circ h_q^{-1}$ is a diffeo, and the collection $\{(V_p,h_p)\}$ is an atlas.

The set Σ with the atlas $\{(V_p,h_p)\}$ is a differentiable manifold, which we denote by $\overline{\Sigma}$. The point set identity i: $\Sigma\to\overline{\Sigma}$ induces a bijection i between the manifolds Σ and $\overline{\Sigma}$ that is a differentiable map; but its inverse i^{-1} need not be differentiable. Indeed, the map $h_p=\sigma_p\circ\Pi_p$, considered as a map of the manifold Σ into $\mathbb{R}^{n-1}$, is differentiable, because Σ_p is; but its inverse need not be such, because Π_p^{-1} need not be differentiable.

The induced topological structure on $\overline{\Sigma}$ is, however, identical to that of the manifold Σ, because Π_p is an open and continuous bijection (Lemma 5). Thus $\overline{\Sigma}$ is a paracompact manifold.

On $\overline{\Sigma}$, there is a partition of unity $\{\varphi_\alpha\}$ subordinate to the collection $\{V_p\}$ (see Ref. 4, p. 51). Let the index set be $\mathcal{I}$ and let the support of φ_α be C_α. For each $\alpha\in\mathcal{I}$, we can choose V_α from $\{V_p\}$ such that $C_\alpha\subset V_\alpha$. The collection $\{V_\alpha\}$ still covers Σ. Let the corresponding $\mathcal{U}_p$, x_p, Σ_p, and Π_p be called $\mathcal{U}_\alpha$, x_α, Σ_α, and Π_α. The collection of functions $\{\varphi_\alpha\}$ also defines a partition of unity on the manifold Σ, because the maps $\varphi_\alpha\circ i$ are still differentiable. Not each partition of unity $\{\varphi'_\alpha\}$ on Σ defines, however, a partition of unity on $\overline{\Sigma}$. The collection $\{\varphi_\alpha\}$ as a partition of unity on Σ clearly has the following additional property: each function

$$\varphi_\alpha\circ\Pi_\alpha^{-1}:\quad\Sigma_\alpha\to\mathbb{R}^1$$

is differentiable.

For each $\alpha\in\mathcal{I}$, we have the (V_α,Σ_α) shift f_α. Define a

function

$$f:\ \Sigma\to\mathbb{R}^1$$

by

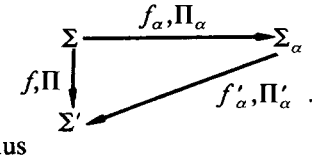
$$f=\Sigma_\alpha\varphi_\alpha f_\alpha.$$

If all f_α's are differentiable, then Σ is regular according to Lemma 5. Thus f need not be differentiable. However, f is a continuous function on Σ, and it defines a set Σ' in Γ by

$$\Sigma'=\{q\in\Gamma\,|\,\exists p\in\Sigma,\quad \gamma_p(f(p))=q\}$$

[f is the (Σ,Σ') shift].

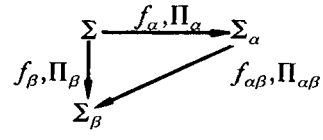
We show that Σ' is a regular cross section. Consider the (Σ_α,Σ') shift f_α. We have the diagram



Thus

$$\begin{aligned}f'_\alpha&=f\circ\Pi_\alpha^{-1}-f_\alpha\circ\Pi_\alpha^{-1}=(f-f_\alpha)\circ\Pi_\alpha^{-1}\\&=\left(\sum_\beta f_\beta\varphi_\beta-\sum_\beta f_\alpha\varphi_\beta\right)\circ\Pi_\alpha^{-1}\\&=\left(\sum_\beta (f_\beta-f_\alpha)\cdot\varphi_\beta\right)\circ\Pi_\alpha^{-1}\\&=\sum_\beta((f_\beta-f_\alpha)\circ\Pi_\alpha^{-1})\cdot(\varphi_\beta\circ\Pi_\alpha^{-1}).\end{aligned}$$

We have the diagram



so

$$f_{\alpha\beta}=-f_\alpha\circ\Pi_\alpha^{-1}+f_\beta\circ\Pi_\alpha^{-1}=(f_\beta-f_\alpha)\circ\Pi_\alpha^{-1}$$

and

$$\Pi_{\alpha\beta}=\Pi_\beta\circ\Pi_\alpha^{-1}.$$

According to Lemma 5, $f_{\alpha\beta}$ is a differentiable function of Σ_α and $\Pi_{\alpha\beta}$ is a diffeo. Thus we obtain, carefully applying each map within its domain,

$$f'_\alpha=\sum_\beta f_{\alpha\beta}\cdot((\varphi_\beta\circ\Pi_\beta^{-1})\circ\Pi_{\alpha\beta}^{-1}).$$

It follows that f'_α must be differentiable. Using Lemma 5 once more, we find that V_α is a regular partial cross section for each $\alpha\in\mathcal{I}$, so Σ' is nowhere tangent to X (and, in particular, Σ is a hypersurface). Q.E.D.

Theorem 2: Let there be a cross section Σ. Then the system is strongly nonrepetitive.

Proof: Let $q\in\Gamma$ be arbitrary. As Σ is a cross section, there is $a\in\mathbb{R}^1$ such that the point p defined by $p=\gamma_q(-a)$ exists and satisfies

$$p\in\Sigma.$$

According to Lemma 4, φ_a is well defined in a neighborhood $\mathcal{U}$ of p and it is a diffeo in $\mathcal{U}$. Thus the set Σ', defined by

$$\Sigma'=\varphi_a(\Sigma\cap\mathcal{U}),$$

is a hypersurface satisfying the conditions of Lemma 5 with $\Sigma_1=\Sigma\cap\mathcal{U}$, $\Sigma_2=\Sigma'$, and $f=a$. Then Σ' is a partial cross section. According to Lemma 2, q is strongly nonrepetitive. Q.E.D.

Theorem 3: Let there be a regular cross section Σ. Then there is a diffeo

$$\Psi:\ \Sigma\times\mathbb{R}^1\to\Gamma$$

such that $\Psi(p\times\mathbb{R}^1)$ is an orbit for any fixed $p\in\Sigma$ and $\Psi(\Sigma\times 0)=\Sigma$.

Proof: According to Theorem 2, the system is strongly nonrepetitive. Using Lemma 1, we construct a super-Hamiltonian $\mathcal{H}'$ that is complete. As X' is complete, its flow Φ defines a diffeomorphism transformation group $\{\varphi_t\}$ on Γ. Define the map

$$\Psi:\ \Sigma\times\mathbb{R}^1\to\Gamma$$

by

$$\Psi(p,t)=\Phi(t,p),$$

for $p\in\Sigma$ and $t\in\mathbb{R}^1$.

The domain of Ψ is $\Sigma\times\mathbb{R}^1$, because X' is complete. Let $p\in\Gamma$ be arbitrary. Then there is a unique $t\in\mathbb{R}^1$ such that

$$\varphi_{-t}(p)\in\Sigma,$$

because Σ is a cross section. Let $q=\varphi_{-t}(p)$. Then $\Psi(q,t)=p$. Thus Ψ has an inverse whose domain is Γ, and Ψ is a bijection. Furthermore, Ψ is differentiable, because it is a composition of differentiable maps.

At any point (p,t) of $\Sigma\times\mathbb{R}^1$, we can construct a frame of independent tangent vectors as follows. Let $e_1,\ldots,e_{n-1}$ be a set of independent tangent vectors to Σ at a point p of Σ. Then there is a unique $(n-1)$-tuple $\bar{e}_1,\ldots,\bar{e}_{n-1}$ of linearly independent tangent vectors at (p,t) satisfying

$$\pi_{1*}(\bar{e}_i)=e_i,\quad \pi_{2*}(\bar{e}_i)=0,$$

where $\pi_1:\ \Sigma\times\mathbb{R}^1\to\Sigma$ and $\pi_2:\ \Sigma\times\mathbb{R}^1\to\mathbb{R}^1$ are the natural projections. This means that $\bar{e}_1,\ldots,\bar{e}_{n-1}$ are tangent to the hypersurface consisting of all points $p\in\Sigma\times\mathbb{R}^1$ that satisfy $\pi_2(p)=t$. Let us denote by $\bar{e}_n$ the vector in $T(\Sigma\times\mathbb{R}^1)$ tangent to the curve (p,λ) in $\Sigma\times\mathbb{R}^1$, $\lambda\in\mathbb{R}^1$, where p is a fixed point of Σ. The vectors $\bar{e}_1,\ldots,\bar{e}_n$ form a linear frame at the point (p,t). The map Ψ_* maps $\bar{e}_n$ to X' and $\bar{e}_1,\ldots,\bar{e}_n$ to $\varphi_{t*}(\bar{e}_1),\ldots,\varphi_{t*}(e_{n-1})$. Then Ψ will be a diffeomorphism, if the vectors $\varphi_{t*}(e_1),\ldots,\varphi_{t*}(e_{n-1}),X'$ form a linear frame at the point $\varphi_t(p)$. However this must be the case if Σ is a regular cross section. Q.E.D.

Now, we are prepared to study the properties of the quotient space $\mathcal{Q}$. Let us assume that the system is strongly nonrepetitive. Then, the following notation will be useful. For any point $p\in\Gamma$, choose a fixed X-adapted chart $(\overline{\mathcal{U}}_p,x_p)$. As the system is strongly nonrepetitive, p has a neighborhood $\mathcal{U}_p\subset\overline{\mathcal{U}}_p$ that is intersected by each maximal integral curve of X at most once. Then $(\mathcal{U}_p,x_p)$ is another X-adapted chart around p, and the hypersurface Σ_p defined by Eq. (7) is intersected by each maximal integral curve of X at most once. The map

$$\rho_p:\Sigma_p\to\mathcal{Q},$$

defined by

$$\rho_p=\pi|_{\Sigma_p},$$

must, therefore, be an injection with domain Σ_p; let us denote its range by W_p. The collection $\{W_p\}$ forms an open covering of $\mathcal{Q}$, which will be used in various constructions.

First, we construct a differentiable structure on $\mathcal{Q}$. For this, we need the map $\sigma_p: \Sigma_p \to \mathbb{R}^{n-1}$, defined by Eq. (12) for any $q \in \Sigma_p$. Clearly σ_p is a diffeo with domain Σ_p and $\sigma_p(\Sigma_p)$ is an open set in $\mathbb{R}^{n-1}$.

Lemma 6: Let the system be strongly nonrepetitive. Then $\{W_p, \sigma_p \circ \rho_p^{-1}\}$ is an atlas on $\mathcal{Q}$ that defines the differentiable structure of the quotient manifold on $\mathcal{Q}$.

Proof: Let us introduce the abbreviation

$$h_p = \sigma_p \circ \rho_p^{-1}.$$

Clearly h_p is an injection with domain W_p and a range $\sigma_p(\Sigma_p)$ that is an open set. Thus (W_p, h_p) is an $(n-1)$-dimensional chart around $\pi(p)$.

The map $\pi|_{\mathcal{U}_p}$ with respect to the coordinates x_p and h_p is given by

$$h_p \circ \pi \circ x_p(x_p^1, \ldots, x_p^n) = (x_p^1, \ldots, x_p^{n-1}).$$

Thus $h_p \circ \pi \circ x_p|_{x_p}(\mathcal{U}_p)$ is a differentiable map of rank $n-1$. If $\{(W_p, h_p)\}$ form a chart on $\mathcal{Q}$, then π will be a submersion and the differentiable structure will be that of a quotient manifold (see Ref. 4, p. 93).

We must show that the transformation $h_p \circ h_q^{-1}$ between any two charts (W_p, h_p) and (W_q, h_q) is a diffeo on $h_q(W_p \cap W_q)$. We have

$$W_q = \rho_q(\Sigma_q), \quad W_p = \rho_p(\Sigma_p),$$

and

$$h_q = \sigma_q \circ \rho_q^{-1}, \quad h_p = \sigma_p \circ \rho_p^{-1}.$$

Let $W = W_q \cap W_p$ and $W \neq \emptyset$. Then, on $h_q(W)$, we obtain

$$h_p \circ h_q^{-1} = \sigma_p \circ (\rho_q^{-1} \circ \rho_q) \circ \sigma_q^{-1}.$$

As σ_p and σ_q are diffeos, $h_p \circ h_q^{-1}$ will also be one iff the map $\rho_p^{-1} \circ \rho_q$ is a diffeo. However, $\rho_p^{-1} \circ \rho_q$ is nothing but the (Σ_q, Σ_p) projection Π_{qp}. As both partial cross sections are regular, Lemma 5 implies that Π_{qp} is a diffeo. Q.E.D.

Propositions 6.1.5, 6.3.1, and 6.3.2 of Ref. 4 imply that the topology induced on $\mathcal{Q}$ by the differentiable structure constructed by Lemma 6 coincides with the quotient set topology, that the projection π is open, and that the manifold $\mathcal{Q}$ has a countable basis for its topology.

Theorems 2 and 3 imply that the quotient space $\mathcal{Q}$ has the following properties.

Corollary 1: Let the system possess a regular cross section Σ. Then $\mathcal{Q}$ can be given the structure of a quotient manifold, and the map

$$\rho: \Sigma \to \mathcal{Q},$$

defined by

$$\rho = \pi|_\Sigma,$$

is a diffeo.

Proof: According to Theorem 2, the system is strongly nonrepetitive, so Lemma 6 implies that $\mathcal{Q}$ can be given the structure of quotient manifold. Theorem 3 guarantees the existence of a diffeo $\Psi: \Sigma \times \mathbb{R}^1 \to \Gamma$ such that $\Psi(p \times \mathbb{R}^1)$ is the maximal integral curve of X for each p. Hence the map $\Pi_\Sigma: \Gamma \to \Sigma$, defined by

$$\Pi_\Sigma = \pi_1 \circ \Psi^{-1},$$

where $\pi_1: \Sigma \times \mathbb{R}^1 \to \Sigma$ is the natural projection, is invariant and differentiable [invariant means constant along $\pi^{-1}(p)$ for any $p \in \mathcal{Q}$ (see Ref. 41, p. 94)]. Then, its projection $\pi_\Sigma: \mathcal{Q} \to \Sigma$ is differentiable by Proposition 6.3.3 of Ref. 4. However, π_Σ is a bijection, whose inverse satisfies

$$\pi_\Sigma^{-1} = \pi|_\Sigma,$$

which is also differentiable. Thus $\pi|_\Sigma$ is a diffeo. Q.E.D.

Corollary 2: Let there be a regular cross section. Then the system is Hausdorff.

According to Corollary 2 and Theorem 2, existence of a cross section implies that the system must be strongly nonrepetitive and Hausdorff. We can also prove the inverse, and for this, we will need the following notation. Consider again the collection $\{\mathcal{U}_p, \Sigma_p, W_p, \rho_p\}$. Here $\{W_p\}$ is an open covering of $\mathcal{Q}$; if $W_q \cap W_p = W \neq \emptyset$ for some p and q from Γ, there will be a well-defined (Σ_q, Σ_p) shift with the domain $\rho_q^{-1}(W)$, which will be denoted by f_{qp}.

Lemma 7: If the system is strongly nonrepetitive and Hausdorff, then there exists a unique differentiable bundle Γ with base space $\mathcal{Q}$, projection π, fiber $\mathbb{R}^1$, group $\{\varphi_t\}$, coordinate neighborhoods $\{W_p\}$, and a system of coordinate transformations $f_{qp} \circ \rho_q^{-1}$.

Proof: If the system is strongly nonrepetitive and Hausdorff, $\mathcal{Q}$ is a paracompact manifold. The transitivity of the coordinate transformations,

$$f_{qp} \circ \rho_q^{-1} = f_{qr} \circ \rho_q^{-1} + f_{rp} \circ \rho_r^{-1},$$

as well as their differentiability follow directly from the transitivity property and the differentiability of the shifts. The existence and uniqueness of the bundle follows, then, from the existence theorem (Theorem 3.2 in Ref. 5).

Theorem 4: Let the system be strongly nonrepetitive and Hausdorff. Then there is a cross section.

Proof: According to Lemma 7, Γ is a bundle with base space $\mathcal{Q}$ and fiber $\mathbb{R}^1$. As $\mathcal{Q}$ is a paracompact manifold, it is metrizable and, therefore, normal. Then the conditions of theorem 12.2 of Ref. 5 are satisfied and the bundle Γ has a continuous cross section. Theorem 6.7 of Ref. 5 implies that there is also a differentiable cross section Σ that is a cross section in our sense. Q.E.D.

III. EXAMPLES

A. Example 1: A weakly repetitive system

Consider a space-time $(\mathcal{M}, g_{\mu\nu})$, and the action of the following form:

$$S = \int dt (p_\mu \dot{x}^\mu - \alpha \mathcal{H}),$$

where x^μ are coordinates on $\mathcal{M}$ and

$$\mathcal{H} = \tfrac{1}{2}(g^{\mu\nu} p_\mu p_\nu - m^2).$$

Such an action leads to an equation of motion that coincides with the geodesic equation for geodesics of the signature $\operatorname{sgn}(m^2)$ and with the affine parameter αt. The configuration space of the system is $\mathcal{M}$ and the phase space is the cotangent bundle $T^*(\mathcal{M})$. In this way, we can use some

well-known examples of weakly acausal space-times to construct weakly repetitive dynamical systems. The procedure is, however, not completely straightforward (almost closed orbits must exist in the phase space!). Here, we will freely use the example given on p. 195 of Ref. 3.

Consider the two-dimensional space-time $(\mathscr{M},g_{\mu\nu})$ with the metric

$$ds^2 = dx^2 - dy^2$$

and with toroidal manifold $\mathscr{M}$: in $\mathbb{R}^2$ with the coordinates x, y, we identify the points differing by the vectors (1,0) and $(0,\sqrt{2})$. Thus the space-time itself is acausal—there are closed timelike curves. However, the null geodesics are not closed: each of them winds infinitely many times around $\mathscr{M}$ and forms a dense set in $\mathscr{M}$. For example, the two null geodesics that start at the point p_0 with coordinates $(x_0,0)$ cut the spacelike "hypersurface" $y=0$ in a dense countable set of points p_n and q_n, whose coordinates are given by

$$p_n \cdots (u_n,0), \quad q_n \cdots (v_n,0),$$
$$u_n = x_0 + n\sqrt{2} - [x_0 + n\sqrt{2}],$$
$$v_n = x_0 - n\sqrt{2} - [x_0 - n\sqrt{2}],$$

where $[a]$ is the greatest integer smaller than a. Hence there is an uncountable number of different null geodesics.

This suggests setting $m=0$ in the action S. The phase space $\mathscr{P}$ is spanned by the coordinates (x,y,p_x,p_y) and Γ is given by the equation

$$p_x^2 - p_y^2 = 0.$$

The super-Hamiltonian vector field X has the components

$$X = (p_x, -p_y, 0, 0),$$

so there is a set of critical points given by $p_x = p_y = 0$. Thus Γ consists of four disconnected sheets of topology $S^1 \times S^1 \times \mathbb{R}^1$ given by the equations and inequalities

$$p_x = p_y, \quad p_y > 0 \text{ or } p_y < 0$$

and

$$p_x = -p_y, \quad p_y > 0 \text{ or } p_y < 0.$$

Along any maximal integral curve of X, both p_x and p_y remain constant. This guarantees that the orbits whose projections are almost closed in the configuration space will be almost closed in the phase space. Thus the system is nonrepetitive but weakly repetitive.

The structure of the quotient space $\mathscr{Q}$ is very interesting in this example. Indeed, as any orbit comes arbitrarily close to any point of any other orbit, there are only two open sets: $\mathscr{Q}$ itself and $\varnothing$. Thus the quotient set topology does not satisfy the separation axiom T_1, and the system is, of course, not Hausdorff. It seems that the separation axiom T_1 will not be satisfied for quite a large class of weakly repetitive systems, namely, for those in which some orbit spirals in approaching another orbit.

The nontrivial topology in the present example originates in the nontrivial topology of the phase space. In the next example, the topology of $\mathscr{P}$ will be trivial, so all of the nontrivial topology will originate from the super-Hamiltonian.

B. Example 2: Strongly nonrepetitive and non-Hausdorff system

The phase space $\mathscr{P}$ of this system is $\mathbb{R}^4$, spanned by canonical coordinates x^1, x^2, p_1, and p_2. The super-Hamiltonian is

$$\mathscr{H} = \tfrac{1}{2}((p_1)^2 - (p_2)^2 - (x^1)^2 + (x^2)^2).$$

It has a lot of symmetry: "Boosts" in the (x^1,x^2), (x^1,p_1), and (x^2,p_2) planes generated by

$$A = x^1p_2 + x^2p_1, \quad B = (p_1)^2 - (x^1)^2, \quad C = (p_2)^2 - (x^2)^2$$

(B and C are also the two "separation" integrals), satisfying the algebra

$$\{A,B\} = 2D, \quad \{A,C\} = 2D, \quad \{B,C\} = 0,$$
$$\{A,D\} = B + C, \quad \{B,D\} = -2A, \quad \{C,D\} = -2A,$$

where

$$D = x^1x^2 + p_1p_2$$

is the "double rotation" in the (p_1,x^2) and (p_2,x^1) planes, and one weak integral

$$I = (x^1x^2 - p_1p_2)/(x^1p_1 + x^2p_2),$$

which satisfies

$$\{\mathscr{H},I\} = -2D\mathscr{H}.$$

The super-Hamiltonian vector field has the components

$$X^\mu = (p_1, -p_2, x^1, -x^2),$$

so there is just one critical point, namely the origin $O = (0,0,0,0)$. If we delete O, the constraint hypersurface has two components with the topology $T^2 \times \mathbb{R}^1$, and it is advantageous to introduce the three coordinates x, ξ, and η on it according to the relations

$$p_1 = x \sin\tfrac{1}{2}(\xi + \eta), \quad p_2 = x \sin\tfrac{1}{2}(\eta - \xi),$$
$$x^1 = x \cos\tfrac{1}{2}(\eta - \xi), \quad x^2 = x \cos\tfrac{1}{2}(\xi + \eta),$$

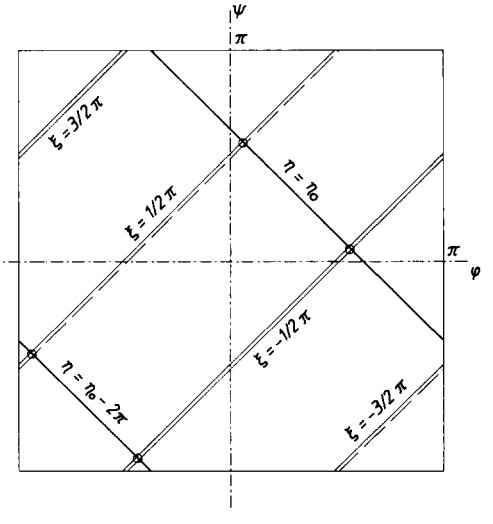


FIG. 1. The torus $x = \text{const}$. The opposite sides of the square are identified. The angular coordinates ψ and φ have periods 2π. For each η_0, the curves $\eta = \eta_0$ adn $\eta = \eta_0 - 2\pi$ together form a smooth closed curve that winds twice around the torus. The separatrices are represented by small circles.

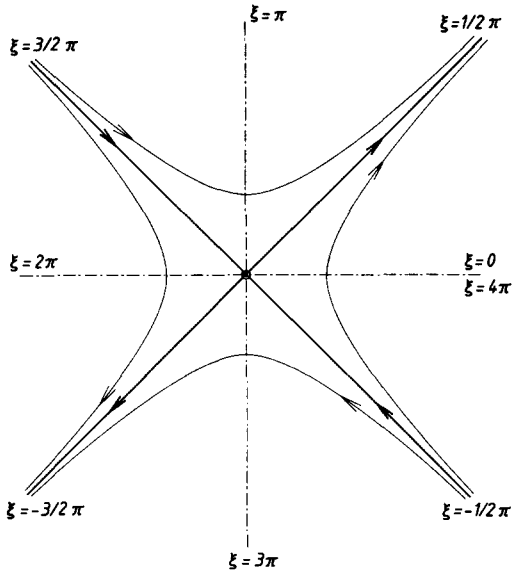


FIG. 2. The surface $\eta = \eta_0$, $\eta = \eta_0 - 2\pi$ in the polar coordinates $x, \frac{1}{2}\xi$ with some typical integral curves. The separatrices are represented by thick lines. The parameter orientation is given by arrows. The origin is cut away.

because of the following relations:

$$I = \cot \eta, \quad D = x^2 \cos \xi. \tag{13}$$

The period of the angles $\varphi = \frac{1}{2}(\eta - \xi)$ and $\psi = \frac{1}{2}(\xi + \eta)$ must be 2π and we take their ranges to be $(-\pi, \pi)$: the φ and ψ directions are those of the minimal periodicity of the torus. Then the period of the coordinates ξ and η is 4π and their ranges are

$$-2\pi < \xi < 2\pi, \quad -2\pi < \eta < 2\pi$$

(see Fig. 1).

The equations of motion in the coordinates x, ξ, and η, and with the parameter given by $d\lambda = \alpha\, dt$, read

$$\dot{x} = x \sin \xi, \quad \dot{\xi} = 2 \cos \xi, \quad \dot{\eta} = 0. \tag{14}$$

If we project all integral curves onto the torus, then we obtain the following picture. All solutions lie within the curves $\eta = \text{const}$. The curve $\eta = \eta_0$, $\eta_0 > 0$, on the torus is a smooth continuation of the curve $\eta = \eta_0 - 2\pi$ and together they form a smooth circle winding twice around the torus. On each of these circles there are four special points with $\cos \xi = 0$, where ξ remains constant. These points divide the circle into four segments within which all other solutions must lie. The solutions corresponding to the four points, the so-called *separatrices*, are given by

$$\xi = +\tfrac{1}{2}\pi x = Ke^{\lambda}, \quad \xi = -\tfrac{1}{2}\pi x = Ke^{-\lambda}, \tag{15}$$

$$\xi = +\tfrac{3}{2}\pi x = Ke^{-\lambda}, \quad \xi = -\tfrac{3}{2}\pi x = Ke^{\lambda}. \tag{16}$$

Consider the submanifold of Γ given by the equations

$$\eta = \eta_0, \quad \eta = \eta_0 - 2\pi,$$

where $\eta_0 > 0$. It is topologically a plane with the origin removed; the plane carries the "polar" coordinates x and $\frac{1}{2}\xi$ (see Fig. 2). There are, then, four purely radial integral curves—the separatrices given are the relations (15) and (16). They are, together with their orientations, illustrated in Fig. 2.

Other integral curves are obtained by integrating Eq. (14) and by setting the result in Eq. (13). This yields, within the different segments,

for $-\frac{1}{2}\pi < \xi < +\frac{1}{2}\pi$, $+\frac{3}{2}\pi < \xi < 2\pi$, $-2\pi < \xi < -\frac{3}{2}\pi$,

$$\sin \xi = \tanh 2(\lambda - \lambda_0),$$

$$\cos \xi = 1/\cosh 2(\lambda - \lambda_0),$$

$$x = \sqrt{|D| \cosh 2(\lambda - \lambda_0)};$$

for $+\frac{1}{2}\pi < \xi < +\frac{3}{2}\pi$, $-\frac{3}{2}\pi < \xi < -\frac{1}{2}\pi$,

$$\sin \xi = \tanh 2(\lambda - \lambda_0),$$

$$\cos \xi = 1/\cosh 2(\lambda - \lambda_0),$$

$$x = \sqrt{|D| \cosh 2(\lambda - \lambda_0)}.$$

The integral curves within the wedges between pairs of the four separatrices $\xi = +\frac{3}{2}\pi$, $+\frac{1}{2}\pi$, $-\frac{1}{2}\pi$, and $-\frac{3}{2}\pi$ form four one-parameter families, the parameter being $|D|$ (λ_0 determines the origin of the parameter λ along the solutions). They are illustrated by Fig. 2.

Within each of these families, we can choose a sequence of orbits, for example, $D = 1, 2^{-1}, \ldots, 2^{-n}, \ldots$, which converges to two different separatrices, namely, those forming the boundary of the corresponding wedge. Hence $\mathcal{Q}$ cannot be Hausdorff in this case. This property seems to have something to do with the local behavior of the super-Hamiltonian vector field in a neighborhood of the critical point O. We are going to study the relations of the topology of $\mathcal{Q}$ and the behavior around critical points in a separate paper.

The constraint hypersurface Γ of this example is multiply connected again. However, this time the pathology would not be removed by taking a covering space Γ' of Γ. The corresponding quotient space $\mathcal{Q}'$ would be non-Hausdorff in even a worse way than $\mathcal{Q}$: the number of pairs of points whose neighborhoods always intersect each other would grow from four to infinity in each ($\eta = \text{const}$)-surface!

ACKNOWLEDGMENTS

The author is thankful to Frank Tipler for stimulating discussions and useful remarks, to Andreas Meister for checking the equations, and to Alen Held for reading the manuscript.

This work was supported by Schweizerischer Nationalfonds.

[1] P. Hajicek, Phys. Rev. D **34**, 1040 (1986).
[2] P. Hajicek, "Dynamical cones of finite dimensional parametrized systems," BUTP-87/23, 1987.
[3] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).
[4] F. Brickell and R. S. Clark, *Differential Manifolds* (Van Nostrand, London, 1970).
[5] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1951).
[6] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, Reading, MA, 1967), 2nd ed.
[7] C. J. S. Clarke, Proc. Cambridge Philos. Soc. **69**, 319 (1970).

Linear kinetic theory in stochastic media

Richard Sanchez
Service d'Etudes de Réacteurs et de Mathématiques Appliquées, Centre d' Etudes Nucléaires de Saclay, 91191 Gif sur Yvette Cedex, France

(Received 1 February 1989; accepted for publication 12 July 1989)

Linear kinetic particle transport in stochastic heterogeneous media is discussed. The analysis includes scattering in a three-dimensional setting and deals with arbitrary time-dependent statistics. Ensemble-average operators are used to derive two independent complete descriptions for the ensemble-averaged angular flux. The first description consists of an infinite system of integral, renewal-like equations for averaged flux values over spatially dependent, increasingly smaller sets. The second approach results in an infinite system of kinetic, balancelike equations for locally averaged flux values. Both types of equations include averagings over transitional sets of states that change locally of physical properties. Previous results are recovered from the limit form of these equations for no-memory statistics and purely absorbing media. Also, the Levermore–Pomraning–Wong proposed models are shown to correspond to truncated forms of these equations.

I. INTRODUCTION

Linear, deterministic transport within a domain D is entirely defined by the angular flux $\psi(\mathbf{r},t,\mathbf{v})$ solution of the kinetic equation[1]

$$\begin{aligned}(L+v\Sigma)\psi &= q + H\psi, \quad \text{in X},\\ \psi &= \Phi, \qquad\quad \text{on } \partial \mathrm{X},\end{aligned} \tag{1}$$

where $L\equiv\partial_t + \mathbf{v}\cdot\nabla$ is the convective derivative, H is the collision operator,

$$(H\psi)(\mathbf{r},t,\mathbf{v}) = v\int d\mathbf{v}'\, \Sigma_s(\mathbf{r},t,\mathbf{v}'\to\mathbf{v})\psi(\mathbf{r},t,\mathbf{v}'),$$

$\Sigma(\mathbf{r},t,\mathbf{v})$ is the total cross section, and $q(\mathbf{r},t,\mathbf{v})$ is the source. We will use the notation $\mathbf{x} = (\mathbf{r},t,\mathbf{v})$ to denote a point in the phase space $\mathrm{X} = \{\mathbf{x}|\mathbf{r}\in D, t\geqslant t_0,\ v\leqslant v_{\max}\}$. In Eq. (1) initial and boundary conditions have been lumped together in the term $\Phi(\mathbf{x})$, which is defined on the set $\partial\mathrm{X} = \{\mathbf{x}\in\mathrm{X}|t = t_0 \text{ or } \mathbf{r}\in\partial D \text{ and } \mathbf{v}\cdot\mathbf{n}\leqslant 0\}$, where $\mathbf{n}$ is the outward normal on ∂D. Particle transport within D is determined by the physical properties of the host medium and by the limit condition Φ: once the quartet (Σ, H, q, Φ) is known, then there is a unique particle distribution solution of Eq. (1).

The problem we want to address here is that of stochastic transport, that is, the description of linear particle transport when the physical properties of the host medium are only known in a statistical manner. In order to set up a statistical formulation we assume that all that is known about the system is that it can be in one of a set of possible states. More precisely, we assume that we are given a set of states $\Omega = \{\omega\}$ together with a time-dependent probability density $p_t(\omega)$:

$$p_t:\Omega\to\mathscr{R}_+ \Rightarrow \int_\Omega d\omega\, p_t(\omega) = 1.$$

Each state ω represents a physical realization of the system, that is, a quartet $(\Sigma_\omega, H_\omega, q_\omega, \Phi_\omega)$. Although Ω is a fixed set, only the states with $p_t(\omega)\geqslant 0$ can be present at time t and therefore our description allows for a variable set of states. Furthermore, since $p_t(\omega)$ is the probability of observing state ω at time t, and therefore its corresponding angular flux ψ_ω, we conclude that the ensemble-averaged angular flux is given by

$$\psi = A\psi = \int_\Omega d\omega\, p_t(\omega)\psi_\omega,$$

where A stands for the time-dependent ensemble-averaging operator. More generally, to each measurable subset $X\subseteq\Omega$ we will associate the time-dependent operators

$$\begin{aligned}M_X\psi &= \int_X d\omega\, p_t(\omega)\psi_\omega,\\ A_X\psi &= M_X\psi / M_X 1,\end{aligned} \tag{2}$$

where $p_X = M_X 1$ is the probability of observing a state in X at time t, and $\psi_X = A_X\psi$ is the ensemble-averaged angular flux over the states that belong to X.

From now on and in order to lighten the notation we will omit the set index in set-related quantities and will write, for instance, M and A instead of M_X and A_X. Except if otherwise indicated our formulas will apply to any measurable subset $X\subseteq\Omega$.

Direct ensemble averaging of Eq. (1) results in the equation

$$\begin{aligned}L\psi + v(\Sigma\psi)_\Omega &= \int_\Omega d\omega\, p_t'(\omega)\psi_\omega + q + (H\psi)_\Omega,\\ \psi &= \Phi.\end{aligned} \tag{3}$$

Here, q and Φ are the Ω-averaged values of the source and of the limit condition, respectively, and $(\,)_\Omega$ represents also an ensemble average. Even in the simple case of stationary statistics, $p_t(\omega) = p(\omega)$, the values of $(\Sigma\psi)_\Omega$ and $(H\psi)_\Omega$ will be in general different from the respective products of average values, $\Sigma\psi$ and $H\psi$, and therefore Eq. (3) shows that the ensemble-averaged flux does not obey a classical kinetic equation. Thus the first objective of a stochastic theory of particle transport should be to derive an equation or set of equations for the ensemble-averaged angular flux ψ. In this paper we will obtain such a result for the case when the statistics can be related to a set of materials. Specifically, we assume that there is a finite set of materials $\{\alpha\}$, where each

 0022-2488/89/112498-14$02.50

material is characterized by a quartet $(\Sigma_\alpha, H_\alpha, q_\alpha, \Phi_\alpha)$, and that each state ω can be viewed as a mapping

$$\omega:\{(\mathbf{r},t)\,|\,\mathbf{r}\in D, t\geqslant t_0\}\rightarrow\{\alpha\}$$

that associates a material $\omega(\mathbf{P})$ to each $\mathbf{P}=(\mathbf{r},t)$. Consequently, the physical properties of a state $(\Sigma_\omega, H_\omega, q_\omega, \Phi_\omega)$ at point $\mathbf{r}$ and time t are those of the material $\alpha=\omega(\mathbf{P})$.

Such material-related statistics can represent, for example, the modeling of particle transport in a multiphase flow or in turbulent plasma. In such a case the suitability of adopting a stationary statistical description will depend on the existence of a well-defined time scaling, i.e., the characteristic time of motion of chunks of flow must be much slower than particle motion. We note also that a further simplification arises if the limit condition Φ is nonstochastic on a part (wall) or on the entire limit domain ∂X.

A basic property of these statistics is the existence of a natural, space-and time-dependent partition of the set of states Ω such that for every $X\subseteq\Omega$:

$$X=(\cup_\alpha X_\alpha(\mathbf{P}))\cup X_t(0), \tag{4}$$

where $\mathbf{P}=(\mathbf{r},t)$, $X_t(0)$ is the subset of states of X that are absent at time t, and

$$X_\alpha(\mathbf{P})=\{\omega\in X\,|\,\omega(\mathbf{P})=\alpha\}$$

is the subset of states of X that "have" material α at $\mathbf{r}$ at time t. Since formula (4) is a partition of X we have the result

$$M\equiv\sum_\alpha M_\alpha(\mathbf{P}),\quad A\equiv\sum_\alpha p_\alpha(\mathbf{P})A_\alpha(\mathbf{P}). \tag{5}$$

Here $M_\alpha(\mathbf{P})$ and $A_\alpha(\mathbf{P})$ are the operators (2) associated with $X_\alpha(\mathbf{P})$, and $p_\alpha(\mathbf{P})=M_\alpha(\mathbf{P})\,1/M\,1$ is the conditional probability of observing material α at $\mathbf{r}$ at time t knowing that the physical state is in X. From the second formula in (5) we obtain the local decomposition

$$\psi(\mathbf{x})=\sum_\alpha p_\alpha(\mathbf{P})\psi_\alpha(\mathbf{x}), \tag{6}$$

where $\mathbf{x}=(\mathbf{P},\mathbf{v})$. This equation gives the ensemble-average flux in terms of the local ensemble averages over the subsets of X which associate the same material to point $\mathbf{r}$ at time t. The above decomposition is local in the sense that the states participating to the averging of ψ_α vary with $\mathbf{r}$ and t.

Equation (6) reduces the task of finding a description for ψ to that of obtaining equations for each of the ψ_α's. As we will show the ψ_α's depend in turn on higher statistical moments of the angular flux and so do these new moments, so that this approach snowballs and results in an infinite set of coupled equations for increasingly singular averagings of the angular flux. Lately, the problem of stochastic transport in binary mixtures of fluids has been the subject of repeated attention.[2–5] However, a rigorous equation for the ensemble-averaged flux has only been obtained for collisionless, stationary transport with time-independent Markovian statistics.[2] Master equations[3] and renewal theory[4,5] have been invoked to extend this work to more general statistics and to include scattering, but results to date are limited and necessarily heuristic. An excellent summary of previous work is given in Ref. 5. To my knowledge, the only results that include scattering in a correct form have been obtained recently by Pomraning[6] for a particular case of stationary transport in systems with planar symmetry and by Vanderhaegen *et al.*[7] for the so-called rod problem. In Ref. 6 the author considers a planar system comprising a single material with constant macroscopic properties (σ,h,q) but variable density $N(x)$: each of the states ω in the stationary statistical ensemble Ω can then be viewed as a spatial density distribution $N_\omega(x)$ such that $\Sigma_\omega(\mathbf{r},v)=N_\omega(x)\sigma(v)$, $H_\omega(\mathbf{r},\mathbf{v})=N_\omega(x)q(\mu,v)$, and $q_\omega(\mathbf{r},\mathbf{v})=N_\omega(x)q(\mu,v)$. In these conditions, the author shows that, for nonstochastic boundary conditions, the ensemble-averaged angular flux $\psi(x,\mu,v)$ can be obtained from a joint probability density function that is directly related to basic properties of the statistical set Ω. A different approach, based on the application of the invariant imbedding method, has been used in Ref. 7 to incorporate scattering effects in a monodirectional description of radiative transfer for a binary Markovian mixture.

The main object of the present work is to derive a complete description of the ensemble-averaged angular flux for stochastic, time-dependent transport including scattering. Our only restriction is that the statistics are material-related but we believe that the method which we use can be easily generalized to other types of statistics.

In Sec. II an ensemble averaging of the integral form of the transport equation is used to derive an infinite set of coupled, renewal-like equations. Each one of these equations gives the ensemble-averaged flux on a set $X\subseteq\Omega$ in terms of the ensemble-averaged fluxes over two families of subsets of X. This result is then confronted with previous findings.

A different approach, based on the direct ensemble averaging of kinetic equation (1), is utilized in Sec. III to obtain this time a system of integrodifferential equations for local moments of ψ, that is, for ensemble-averaged values over subsets $X\subseteq\Omega$ which depend only on local coordinates. In contrast, the ensemble averagings that appear in the renewal equations of Sec. II are defined over subsets that depend on one or more $(\mathbf{r},t)$ values. We conclude in Sec. IV by specializing our equations to no-memory statistics, such as Markovian statistics, and by recovering previous results.

Since ψ does not obey a classical kinetic equation it is pertinent to obtain an approximate transport equation for ψ or, equivalently, to define a quartet $(\Sigma_{\text{eff}}, H_{\text{eff}}, q_{\text{eff}}, \Phi_{\text{eff}})$. The appropriate answer to this problem depends on the class of situations considered.[3–5] In a future paper we will construct an approximate transport equation for the double-heterogeneity problem.[8] A different approach consists of using a closure argument to obtain an approximate finite system of equations (not necessarily classical transport equations)[2] for ψ. Out conclusions are given in Sec. V where we also propose a truncation assumption to reduce the infinite system of equations to a finite system comprising a balance differential equation and a renewal integral equation.

Appendix A contains a discussion of some mathematical details regarding operators used throughout this work. In Appendix B we give a summary of results for the case of stationary statistics. Appendix C is dedicated to an analysis of the application of continuous Markovian process to the description of three-dimensional multicomponent fluid mixtures.

II. RENEWAL EQUATIONS

The integral form of the transport equation gives the angular flux at $\mathbf{x} = (\mathbf{P},\mathbf{v}) = (\mathbf{r},t,\mathbf{v})$ in terms of the local emissivity at velocity $\mathbf{v}$ along the past trajectory, i.e., along the set

$$\mathrm{Tr}(\theta_l,\mathbf{x}) = \{\mathbf{P}_\theta = (\mathbf{r} - \theta\mathbf{v}, t - \theta) | \theta \in [0, \theta_l)\},$$

where

$$\theta_l(\mathbf{x}) = \min[\theta_{\mathrm{bd}}(\mathbf{r},\mathbf{v}), \theta_{\mathrm{in}}(t)].$$

Here θ_{bd} is the time required for a particle of velocity $\mathbf{v}$ to travel from the boundary ∂D to location $\mathbf{r}$, and $\theta_{\mathrm{in}} = t - t_0$. Depending on whether $\theta_l = \theta_{\mathrm{bd}}$ or $\theta_l = \theta_{\mathrm{in}}$, the angular flux at $\mathbf{x}$ will be related to a boundary or to an initial condition. In order to ensemble-average the integral transport equation it is necessary to study the behavior of the states along the past trajectory. Assume that ω assigns material α to $\mathbf{P} = (\mathbf{r},t)$, then the local behavior of ω along the past trajectory can be characterized by the maximum length of time during which a particle of velocity $\mathbf{v}$ can travel so that it reaches $\mathbf{P}$ seeing continuously material α. More precisely, for each $\mathbf{x}$ we define the function $\theta(\mathbf{x})$: $\Omega \to \mathscr{R}_+$ such that

$$\theta_\omega(\mathbf{x}) = \max\{\theta \in \mathscr{R}_+ | \omega(\mathbf{P}') = \omega(\mathbf{P}), \forall \mathbf{P}' \in \mathrm{Tr}(\theta,\mathbf{x})\}.$$

Thus, along the past trajectory ω continuously takes on the value α over $\mathrm{Tr}(\theta_\omega,\mathbf{x})$ and changes of value at $\mathbf{P}_\omega = (\mathbf{r} - \theta_\omega\mathbf{v}, t - \theta_\omega)$: $\omega(\mathbf{P}_\omega) \neq \omega(\mathbf{P})$.

Then, for a state $\omega \in \Omega_\alpha(\mathbf{P})$ we write the integral transport equation at $\mathbf{x} = (\mathbf{P},\mathbf{v})$ as

$$\begin{aligned}\psi_\omega(\mathbf{x}) = {} & H(\theta_\omega - \theta_l) e^{-\tau_\alpha(\theta_l,\mathbf{x})} \Phi_\omega(\mathbf{x}_l) \\ & + H(\theta_l - \theta_\omega) e^{-\tau_\alpha(\theta_\omega,\mathbf{x})} \psi_\omega(\mathbf{x}_\omega) \\ & + \int_0^{\min(\theta_\omega,\theta_l)} d\theta\, e^{-\tau_\alpha(\theta,\mathbf{x})} F_\omega(\mathbf{x}_\theta). \end{aligned} \tag{7}$$

In this equation H represents Heaviside's function,

$$\tau_\alpha(\theta,\mathbf{x}) = v\int_0^\theta d\theta'\, \Sigma_\alpha(\mathbf{r} - \theta'\mathbf{v}, t - \theta', v)$$

is the optical distance along the past trajectory $\mathrm{Tr}(\theta,\mathbf{x})$, Φ_ω can be a boundary or initial value depending on whether θ_l equals θ_{bd} or θ_{in}, and F_ω is the local emissivity

$$F_\omega(\mathbf{x}) = q_\alpha(\mathbf{x}) + (H_\alpha\psi_\omega)(\mathbf{x}).$$

The notation adopted for the coordinates along the past trajectory is $\mathbf{x}_\theta = (\mathbf{P}_\theta,\mathbf{v})$ with $\mathbf{P}_\theta = (\mathbf{r} - \theta\mathbf{v}, t - \theta)$ and similarly for $\mathbf{x}_l(\theta = \theta_l)$ and $\mathbf{x}_\omega(\theta = \theta_\omega)$.

Our purpose now is to obtain an equation for the ensemble-averaged flux $\psi_\alpha = A_\alpha(\mathbf{P})\psi$ by ensemble-averaging Eq. (7) over the subset $X_\alpha(\mathbf{P})$. To carry out this operation we will use an integral representation of the operator $M_\alpha(\mathbf{P})$. As shown in Appendix A, for any $\theta \geqslant 0$:

$$M_\alpha(\mathbf{P}) \equiv M_{\alpha\alpha}(\theta,\mathbf{x}) + \int_0^\theta d\theta'\, M_{\cdot\alpha}(\theta',\mathbf{x}). \tag{8}$$

In this formula $M_{\alpha\alpha}(\theta,\mathbf{x})$ is the integral operator

$$M_{\alpha\alpha}(\theta,\mathbf{x})f = \int_{X_{\alpha\alpha}(\theta,\mathbf{x})} d\omega\, p_t(\omega) f_\omega$$

with support

$$X_{\alpha\alpha}(\theta,\mathbf{x}) = \{\omega \in X | \omega(\mathbf{P}') = \omega(\mathbf{P}), \forall \mathbf{P}' \in \mathrm{Tr}(\theta,\mathbf{x})\}.$$

For any point $\mathbf{P}_\theta$ in the past trajectory of $\mathbf{P}$ we define the transit time θ as the time required for a particle of velocity $\mathbf{v}$ to go from $\mathbf{P}_\theta$ to $\mathbf{P}$; then $X_{\alpha\alpha}(\theta,\mathbf{x})$ is the subset of states of X that assign material α to all points on the past trajectory with transit time greater than θ. Also the operator $M_{\cdot\alpha}(\theta,\mathbf{x})$ is the directional derivative of $M_{\alpha\alpha}(\theta,\mathbf{x})$ at $\mathbf{P}_\theta$, and its support is essentially the subset of states in $X_{\alpha\alpha}(\theta,\mathbf{x})$ that change from material α to a different material at $\mathbf{P}_\theta$.

For more details the reader should look over Appendix A. Here we give a somewhat intuitive derivation of formula (8) based on a change-of-variable argument. The integration over $X_\alpha(\mathbf{P})$

$$M_\alpha(\mathbf{P})f = \int_{X_\alpha(\mathbf{P})} d\omega\, p_t(\omega) f_\omega$$

can be carried out by selecting a past trajectory at $\mathbf{P}$ and by collecting the states that have the same value of $\theta_\omega(\mathbf{x})$. This is formally equivalent to the change of variables $\omega = (\theta_\omega,\omega_\perp)$, where $\omega_\perp$ stands for the remaining coordinates. We can therefore write

$$M_\alpha(\mathbf{P})f = \int_0^\infty d\theta_\omega \int d\omega_\perp\, P_t(\omega) f_\omega,$$

where the integral over $\omega_\perp$, which is symbolized by the operator $M_{\cdot\alpha}(\theta_\omega,\mathbf{x})$ in formula (8), is done over the subset of states of X that are continuously α along the past trajectory and that change to a different material within a $d\theta_\omega$ at $\mathbf{P}_\omega$. From this argument it also follows that

$$M_{\alpha\alpha}(\theta,\mathbf{x}) = \int_\theta^\infty d\theta'\, M_{\cdot\alpha}(\theta',\mathbf{x}). \tag{9}$$

The operators $M_{\cdot\alpha}(\theta,\mathbf{x})$ are directly related to the chord distribution of material α along the past trajectory at $\mathbf{x}$. Indeed, this is easily seen from the definitions

$$\begin{aligned} R_\alpha(\theta,\mathbf{x}) = {} & \text{probability for } \omega \in X_\alpha(\mathbf{P}) \text{ of having } \theta_\omega(\mathbf{x}) \geqslant \theta \\ = {} & \frac{M_{\alpha\alpha}(\theta,\mathbf{x})1}{M_\alpha(\mathbf{P})1} = \frac{\int_\theta^\infty d\theta'\, M_{\cdot\alpha}(\theta',\mathbf{x})1}{\int_0^\infty d\theta'\, M_{\cdot\alpha}(\theta',\mathbf{x})1}, \\ Q_\alpha(\theta,\mathbf{x}) = {} & -\partial_\theta R_\alpha(\theta,\mathbf{x}) = \frac{M_{\cdot\alpha}(\theta,\mathbf{x},)1}{M_\alpha(\mathbf{P})1}. \end{aligned} \tag{10}$$

Hence, $R_\alpha(\theta,\mathbf{x})$ is the conditional probability for the states in X that the entire interval $\mathrm{Tr}(\theta,\mathbf{x})$ is in material α given that $\mathbf{P}$ is in material α; and $Q_\alpha(\theta,\mathbf{x})$ is the conditional probability density of having the interval $\mathrm{Tr}(\theta,\mathbf{x})$ in material α and of changing to a different material at $\mathbf{P}_\theta$ given that $\mathbf{P}$ is in α. In particular, the averaged transit time through material α along the past trajectory is

$$t_\alpha(\mathbf{x}) = \int_0^\infty d\theta\, \theta Q_\alpha(\theta,\mathbf{x}) = \int_0^\infty d\theta\, R_\alpha(\theta,\mathbf{x}).$$

Finally, ensemble-averaging integral equation (7) and with the help of formula (8) for $\theta = \infty$ we obtain the result:

$$\begin{aligned}\psi_\alpha(\mathbf{x}) = {} & e^{-\tau_\alpha(\theta_l,\mathbf{x})} \mathrm{R}_\alpha(\theta_l,\mathbf{x}) \Phi_{\alpha\alpha}^{\theta_l\mathbf{x}}(\mathbf{x}_l) + \int_0^{\theta_l} d\theta\, e^{-\tau_\alpha(\theta,\mathbf{x})} \\ & \times [Q_\alpha(\theta,\mathbf{x}) \psi_{\cdot\alpha}^{\theta\mathbf{x}}(\mathbf{x}_\theta) + R_\alpha(\theta,\mathbf{x}) F_{\alpha\alpha}^{\theta\mathbf{x}}(\mathbf{x}_\theta)], \end{aligned} \tag{11}$$

where

$$F_{\alpha\alpha}^{\theta\mathbf{x}}(\mathbf{x}_\theta) = q_\alpha(\mathbf{x}_\theta) + (H_\alpha \psi_{\alpha\alpha}^{\theta\mathbf{x}})(\mathbf{x}_\theta),$$

and where we have defined the following ensemble-averaged values of the angular flux:

$$\psi_{.\alpha}^{\theta\mathbf{x}} = A_{.\alpha}(\theta,\mathbf{x})\psi = \frac{M_{.\alpha}(\theta,\mathbf{x})\psi}{M_{.\alpha}(\theta,\mathbf{x})1},$$
$$\psi_{\alpha\alpha}^{\theta\mathbf{x}} = A_{\alpha\alpha}(\theta,\mathbf{x})\psi = \frac{M_{\alpha\alpha}(\theta,\mathbf{x})\psi}{M_{\alpha\alpha}(\theta,\mathbf{x})1}. \tag{12}$$

Here $\psi_{\alpha\alpha}^{\theta\mathbf{x}}$ is the ensemble-averaged angular flux over the set of states $\omega \in X$ that have the entire interval $\mathrm{Tr}(\theta,\mathbf{x})$ in material α, and $\psi_{.\alpha}^{\theta\mathbf{x}}$ is essentially the ensemble-averaged value over the $\omega \in X$ that have $\mathrm{Tr}(\theta,\mathbf{x})$ in material α and that change to a different material at $\mathbf{P}_\theta$.

Furthermore, if we introduce the definition

$$\lambda_\alpha(\theta,\mathbf{x}) = \frac{Q_\alpha(\theta,\mathbf{x})}{R_\alpha(\theta,\mathbf{x})} = -\partial_\theta \ln R_\alpha(\theta,\mathbf{x}) \tag{13}$$

then Eq. (11) can be written in the more compact form:

$$\psi_\alpha(\mathbf{x}) = e^{-\tilde{\tau}_\alpha(\theta_l,\mathbf{x})}\Phi_{\alpha\alpha}^{\theta_l\mathbf{x}}(\mathbf{x}_l) + \int_0^{\theta_l} d\theta\, e^{-\tilde{\tau}_\alpha(\theta,\mathbf{x})} \times [\lambda_\alpha(\theta,\mathbf{x})\psi_{.\alpha}^{\theta\mathbf{x}}(\mathbf{x}_\theta) + F_{\alpha\alpha}^{\theta\mathbf{x}}(\mathbf{x}_\theta)], \tag{14}$$

where now

$$\tilde{\tau}_\alpha(\theta,\mathbf{x}) = \tau_\alpha(\theta,\mathbf{x}) + \int_0^\theta d\theta'\,\lambda_\alpha(\theta',\mathbf{x}).$$

The value of $R_\alpha(\theta + d\theta,\mathbf{x})$ can be written as the product of $R_\alpha(\theta,\mathbf{x})$ times the conditional probability $p_\alpha(d\theta,\theta,\mathbf{x})$ for states in $X_{\alpha\alpha}(\theta,\mathbf{x})$ to remain in α in the interval $(\theta,\theta + d\theta)$. At the limit when $d\theta$ goes to zero,

$$p_\alpha(d\theta,\theta,\mathbf{x}) = R_\alpha(\theta + d\theta,\mathbf{x})/R_\alpha(\theta,\mathbf{x}) \to 1 - d\theta\lambda_\alpha(\theta,\mathbf{x}).$$

This shows that $\lambda_\alpha(\theta,\mathbf{x})$ is the conditional probability density for states in $X_{\alpha\alpha}(\theta,\mathbf{x})$ to change to a different material per unit transit time.

Let us now come back to Eq. (11) and show how to use this result to generate a complete description of the ensemble-averaged angular flux ψ. First we recall that Eq. (11) has been obtained by averaging over the set $X_\alpha(\mathbf{P})$ and therefore it cannot be directly used to compute $\psi_{.\alpha}^{\theta\mathbf{x}}$. In order to be able to treat this latter ensemble-averaged value we will introduce the decomposition

$$\psi_{.\alpha}^{\theta\mathbf{x}} = \sum_{\beta\neq\alpha} p_{\beta\alpha}(\theta,\mathbf{x})\psi_{\beta\alpha}^{\theta\mathbf{x}}. \tag{15}$$

Here

$$\psi_{\beta\alpha}^{\theta\mathbf{x}} = A_{\beta\alpha}(\theta,\mathbf{x})\psi, \quad (\beta\neq\alpha),$$

is the ensemble-averaged angular flux over the states $\omega \in X$ that have the interval $\mathrm{Tr}(\theta,\mathbf{x})$ in material α and that change to β at $\mathbf{P}_\theta$. Also

$$p_{\beta\alpha}(\theta,\mathbf{x}) = \frac{M_{\beta\alpha}(\theta,\mathbf{x})1}{M_{.\alpha}(\theta,\mathbf{x})1} \quad (\beta\neq\alpha),$$

is the conditional transition probability density that the states $\omega \in X$ that have the interval $\mathrm{Tr}(\theta,\mathbf{x})$ in material α and that change to a different material at $\mathbf{P}_\theta$ will change to material β. Moreover, the continuity of the particle flux at the interface between two materials (we assume here that there are no surface sources) leads to the continuity of the ensemble-averaged values. Consequently, the mean value $\psi_{\beta\alpha}^{\theta\mathbf{x}}$ can be viewed as the ensemble-average angular flux leaving material β in direction $\mathbf{\Omega} = \mathbf{v}/v$ at $\mathbf{x}_\theta$ to travel through material α at least during a time θ.

We see now how the use of Eq. (11) for the determination of the ψ_α's (for $X = \Omega$) leads to an infinite number of renewal-like equations for an infinite set of ensemble-averaged fluxes over a set of subsets of Ω which starts with the $\Omega_\alpha(\mathbf{P})$'s. For $X = \Omega$ Eq. (10) gives ψ_α in terms of $\psi_{\alpha\alpha}^{\theta\mathbf{x}}$ and of the $\psi_{\beta\alpha}^{\theta\mathbf{x}}$'s for $\beta\neq\alpha$. Next, using again (10) to compute first $\psi_{\alpha\alpha}^{\theta\mathbf{x}}$ and then $\psi_{\beta\alpha}^{\theta\mathbf{x}}$ produces four new types of ensemble-average fluxes, and so on. Each new equation brings in two new ensemble-averaged fluxes over increasingly smaller subsets of $\Omega_\alpha(\mathbf{P})$. Table I gives a summary of this situation for the first three equations of the system. It is worth noting that the ensemble-averaged fluxes depend on more than one spatial coordinate; for instance, $\psi_{\gamma\beta\alpha}^{\theta'\theta\mathbf{x}}(\mathbf{P}',\mathbf{v})$ is the ensemble-averaged angular flux over the states that change from material γ to β at $\mathbf{P}'$, have the interval $\mathrm{Tr}(\theta',\mathbf{P}_\theta)$ in material β, change to material α at $\mathbf{P}_\theta$ and have the interval $\mathrm{Tr}(\theta,\mathbf{P})$ in material α.

The system of renewal-like equations remains of infinite order even at the collisionless limit: $H_\alpha \equiv 0$ for $\forall\alpha$. In this limit the ensemble-averaged fluxes with two identical adjacent indexes (such as $\psi_{\alpha\alpha}$ or $\psi_{\beta\beta\alpha}$) do not appear in the equations, and only ensemble-averaged fluxes with different adjacent indexes are necessary. Nevertheless, the remaining

TABLE I. Types of ensemble-averaged fluxes entering the first three renewal equations.

Equation	Main flux	Collision term	Interface term	Coordinates
1	$\psi_\alpha(\mathbf{P},\mathbf{v})$[a]	$\psi_{\alpha\alpha}^{\theta\mathbf{x}}(\mathbf{P}_\theta,\mathbf{v}')$	$\psi_{\beta\alpha}^{\theta\mathbf{x}}(\mathbf{P}_\theta,\mathbf{v})$	$\mathbf{x} = (\mathbf{P},\mathbf{v}) = (\mathbf{r},t,\mathbf{v})$ $\mathbf{P}_\theta = (\mathbf{r} - \theta\mathbf{v}, t - \theta)$
2	$\psi_{\alpha\alpha}^{\theta\mathbf{x}}(\mathbf{P}_\theta,\mathbf{v}')$	$\psi_{\alpha\alpha\alpha}^{\theta'\mathbf{x}'\mathbf{x}}(\mathbf{P}'',\mathbf{v}'')$	$\psi_{\beta\alpha\alpha}^{\theta'\mathbf{x}'\mathbf{x}}(\mathbf{P}'',\mathbf{v}')$	$\mathbf{x}' = (\mathbf{P}_\theta,\mathbf{v}')$ $\mathbf{P}'' = (\mathbf{r} - \theta\mathbf{v} - \theta'\mathbf{v}',$ $t - (\theta + \theta'))$
3	$\psi_{\beta\alpha}^{\theta\mathbf{x}}(\mathbf{P}_\theta,\mathbf{v})$	$\psi_{\beta\beta\alpha}^{\theta'\theta\mathbf{x}}(\mathbf{P}',\mathbf{v}')$	$\psi_{\gamma\beta\alpha}^{\theta'\theta\mathbf{x}}(\mathbf{P}',\mathbf{v})$	$\mathbf{P}' = (\mathbf{r} - (\theta + \theta')\mathbf{v},$ $t - (\theta + \theta'))$

[a] For consistency one should write $\psi_\alpha^{\mathbf{P}}$ to indicate that the averaging depends on the local point $\mathbf{P}$.

equations still form a system of infinite order. It is the opinion of this author that the reduction of the exact, infinite system of renewal equations to only the first two equations is only possible for a special type of statistics. This is particularly true for the no-memory statistics that we discuss at the end of Sec. III, and for renewal statistics.[4,5] In relation with the equations proposed in Ref. 5 we note further that for binary statistical mixtures the transition probability $p_{\beta\alpha}$ necessarily equals 1 and then $\psi_{.\alpha}$ reduces to $\psi_{\beta\alpha}$.

III. ENSEMBLE-AVERAGED BALANCE EQUATIONS

Because of the spatial coupling inherent to the integral transport equation, ensemble-averaging of this equation brings in nonlocal ensemble-averaged fluxes, i.e., averages over sets of states that depend on a continuous segment along the past trajectory. The velocity spreading produced by the collision operator introduces further complications in that the averages have to be done over sets of states depending on a zigzag-like subset of the past trajectory. In this section we will explore a different approach based on the local ensemble-averaging of the integrodifferential kinetic equation. The result is a balance-like equation that brings into play two new velocity-dependent, local ensemble-averaged fluxes.

Let X be a measurable, but otherwise arbitrary, subset of Ω and $X_\alpha(\mathbf{P})$ the locally defined subset of states in X that assign material α to $\mathbf{P}=(\mathbf{r},t)$. Then, ensemble-averaging of kinetic equation (1) gives the local balance relation:

$$\begin{aligned} A_\alpha(\mathbf{P})L\psi + v\Sigma_\alpha\psi_\alpha &= q_\alpha + H_\alpha\psi_\alpha, \quad \text{in } X, \\ \psi_\alpha &= \Phi_\alpha, \quad \text{on } \partial X, \end{aligned}$$

where $A_\alpha(\mathbf{P})$ is the ensemble-averaging operator associated with $X_\alpha(\mathbf{P})$ and $\psi_\alpha = A_\alpha(\mathbf{P})\psi$ is the corresponding averaged flux. In order to close this equation it remains to calculate the ensemble-averaging of the convective term $L\psi$ or, equivalently, to obtain the commutation rule for the operators $A_\alpha(\mathbf{P})$ and L. As shown in Appendix A the commutator of these two operators is

$$\begin{aligned} &A_\alpha(\mathbf{P})L - LA_\alpha(\mathbf{P}) \\ &\quad \equiv Q_\alpha^-(\mathbf{x})[A_{.\alpha}^-(\mathbf{x}) - A_\alpha(\mathbf{P})] \\ &\qquad - Q_\alpha^+(\mathbf{x})[A_{.\alpha}^+(\mathbf{x}) - A_\alpha(\mathbf{P})], \end{aligned} \tag{16}$$

where $\mathbf{x}=(\mathbf{P},\mathbf{v})$. The quantities $Q_\alpha^\pm(\mathbf{x})$ measure the ratio at which states enter (+) or leave (−) the set $X_\alpha(\mathbf{P})$ as one moves with velocity $\mathbf{v}$ along the trajectory at $\mathbf{P}$. In terms of the probabilities defined in Eq. (10):

$$Q_\alpha^\pm(\mathbf{x}) = \lim_{\theta\to 0_+} Q_\alpha(\pm\theta,\mathbf{x}). \tag{17}$$

Also, the operators $A_{.\alpha}^\pm(\mathbf{x})$ essentially indicate ensemble-averaging over the states that locally enter (+) or leave (−) the set $X_\alpha(\mathbf{P})$. If $A_{.\alpha}(\theta,\mathbf{x})$ denotes the averaging operator

$$A_{.\alpha}(\theta,\mathbf{x})f = \frac{M_{.\alpha}(\theta,\mathbf{x})f}{M_{.\alpha}(\theta,\mathbf{x})1},$$

then

$$A_{.\alpha}^\pm(\mathbf{x}) \equiv \lim_{\theta\to 0_+} A_{.\alpha}(\pm\theta,\mathbf{x}) \equiv \frac{M_{.\alpha}^\pm(\mathbf{x})}{M_{.\alpha}(\mathbf{x})1}.$$

In regard to this last formula we note that for $\theta \geqslant 0$ the operator $M_{.\alpha}(\theta,\mathbf{x})$ is the result of a derivative over the future trajectory.

With the help of commutation (16) the balance equation for ψ_α can be written in the more explicit form:

$$\begin{aligned} (L + v\Sigma_\alpha)\psi_\alpha &= q_\alpha + H_\alpha\psi_\alpha + \Delta_\alpha, \quad \text{in } X, \\ \psi_\alpha &= \Phi_\alpha, \quad \text{on } \partial X, \end{aligned} \tag{18a}$$

where the new term

$$\begin{aligned} \Delta_\alpha = &Q_\alpha^+(\mathbf{x})[\psi_{.\alpha}^+(\mathbf{x}) - \psi_\alpha(\mathbf{x})] \\ &- Q_\alpha^-(\mathbf{x})[\psi_{.\alpha}^-(\mathbf{x}) - \psi_\alpha(\mathbf{x})] \end{aligned} \tag{18b}$$

takes into account the convective derivative of the ensemble-average angular flux. This term, which vanishes at any $\mathbf{x}$ that has a finite neighborhood over which X takes on constant values, depends on the values of the local ensemble-averaged interface fluxes

$$\psi_{.\alpha}^\pm = A_{.\alpha}^\pm\psi = \lim_{\theta\to 0_+} \psi_{.\alpha}^{\pm\theta\mathbf{x}}. \tag{19}$$

By integrating Eq. (18) along the past trajectory one finds the equivalent integral equation

$$\begin{aligned} \psi_\alpha(\mathbf{x}) = &e^{-\hat{\tau}_\alpha(\theta_l,\mathbf{x})}\Phi_\alpha(\mathbf{x}_l) \\ &+ \int_0^{\theta_l} d\theta\, e^{-\hat{\tau}_\alpha(\theta,\mathbf{x})}[Q_\alpha^+(\mathbf{x}_\theta)\psi_{.\alpha}^+(\mathbf{x}_\theta) \\ &- Q_\alpha^-(\mathbf{x}_\theta)\psi_{.\alpha}^-(\mathbf{x}_\theta) + F_\alpha(\mathbf{x}_\theta)], \end{aligned} \tag{20}$$

where all the averages are local, $\mathbf{x}_\theta = (\mathbf{r}-\theta\mathbf{v}, t-\theta, \mathbf{v})$ and

$$\hat{\tau}_\alpha(\theta,\mathbf{x}) = \tau_\alpha(\theta,\mathbf{x}) + \int_0^\theta d\theta'[Q_\alpha^+(\mathbf{x}_{\theta'}) - Q_\alpha^-(\mathbf{x}_{\theta'})].$$

A comparison between (20) and renewal equation (14) shows that, in general, balance equation (18) is not equivalent to the renewal formulation. Similarly to the renewal equations, Eq. (18) can also be used to generate a system of infinite kineticlike equations for increasingly singular local flux averages. However, since the additional term Δ_α, brings in increasingly higher derivates of averaging operators, the implementation of such an approach is bound to be more difficult than that based on the renewal equations.

The new velocity-dependent local flux averages $\psi_{.\alpha}^\pm(\mathbf{x})$ are the ensemble averages of ψ over the set of states that per unit transit time enter (+) or leave (−) material α in the direction of $\mathbf{v}$ at $\mathbf{P}=(\mathbf{r},t)$. The decomposition in Eq. (15) of $\psi_{.\alpha}^{\theta\mathbf{x}}$ into the separate contributions from the material-to-material local transitions can be carried over to the present case to obtain:

$$\psi_{.\alpha}^\pm(\mathbf{x}) = \sum_{\beta\neq\alpha} p_{\beta\alpha}^\pm(\mathbf{x})\psi_{\beta\alpha}^\pm(\mathbf{x}), \tag{21}$$

where the

$$\psi_{\beta\alpha}^\pm = A_{\beta\alpha}^\pm(\mathbf{x})\psi = \lim_{\theta\to 0_+} A_{\beta\alpha}(\pm\theta,\mathbf{x})\psi,$$

are flux-averaged values over states changing from β into α (+) or from α into β(−) per unit transit time at $\mathbf{x}=(\mathbf{r},t,\mathbf{v})$. The coefficients

$$p^{\pm}_{\beta\alpha}(\mathbf{x}) = \lim_{\theta\to 0_+} \frac{M_{\beta\alpha}(\pm\theta,\mathbf{x})1}{M_{\cdot\alpha}(\pm\theta,\mathbf{x})1}$$

are local conditional probabilities for states that emerge from material β to enter material $\alpha(+)$ or for states that leave material α to enter material $\beta(-)$. Likewise the $p_{\beta\alpha}(\theta,\mathbf{x})$ these new probabilities are normalized to 1:

$$\sum_{\beta\neq\alpha} p^{\pm}_{\beta\alpha}\mathbf{x} = 1.$$

We close this section by establishing a direct connection between our balance equation (18) and a similar result recently obtained by Adams and co-workers in an independent manner.[9] By noticing that

$$Lp_\alpha = p_\alpha[Q^+_\alpha(\mathbf{x}) - Q^-_\alpha(\mathbf{x})],$$

one can rewrite Eq. (18) under the form

$$(L + v\Sigma_\alpha)(p_\alpha\psi_\alpha) = p_\alpha q_\alpha + H_\alpha p_\alpha\psi_\alpha + \theta_\alpha,$$

where

$$\theta_\alpha = p_\alpha[Q^+_\alpha(\mathbf{x})\psi^+_{\cdot\alpha}(\mathbf{x}) - Q^-_\alpha(\mathbf{x})\psi^-_{\cdot\alpha}(\mathbf{x})]$$
$$= [M^+_{\cdot\alpha}(\mathbf{x}) - M^-_{\cdot\alpha}(\mathbf{x})]\psi(\mathbf{x}).$$

This is the form of the balance equation that has been derived in Ref. 9 from the direct averaging of a binary statistical mixture in an arbitrarily small control volume.

IV. THE COLLISIONLESS CASE

In the collisionless case, $H = 0$, the angular flux depends on the values of the source q and the cross section Σ only along the past trajectory, and for all practical purposes the geometry of the problem simplifies to a half-line geometry. Then, the angular flux ψ_ω does not depend on the material composition over the future trajectory, and we can write $\psi_\omega = \psi_{\omega_-}$, where $\omega_\pm$ stands for the restriction of ω to the past $(-)$ or to the future $(+)$ trajectories. Moreover, for an arbitrary $X\subseteq\Omega$ we have with $\omega = (\omega_-,\omega_+)$:

$$(M_X\psi)(\mathbf{x}) = \int_X d\omega\, p_t(\omega)\psi_\omega(\mathbf{x})$$
$$= \int_{X_-} d\omega_-\, p_t(\omega_-,X,\mathbf{x})\psi_{\omega_-}(\mathbf{x}) \quad (H\equiv 0),$$

where X_- is the factor set X/R by the equivalence relation $\omega R\omega' \leftrightarrow \omega_- = \omega'_-$, and $p_t(\omega_-,X,\mathbf{x})$ is the density of probability for the states in X that have the same ω_- at $\mathbf{x}$. A further simplification occurs when the statistics are such that the chord distribution along the future trajectory is independent of the material composition over the past trajectory. More precisely, we will say that $\hat{\Omega}\equiv(\Omega,p_t)$ is a no-memory statistical set if at every $\mathbf{x}$ in phase space we can construct two statistical sets $\hat{\Omega}_\pm = (\Omega_\pm,p^\pm_t)$ such that $\Omega_\pm$ are the restrictions of Ω, respectively, to the past $(-)$ and to the future $(+)$ trajectories, and such that $\hat{\Omega}$ is the Cartesian product of $\hat{\Omega}_+$ and $\hat{\Omega}_-$:

$$\forall\omega\in\Omega:\omega = (\omega_-,\omega_+),$$
$$p_t(\omega)d\omega = p^+_t(\omega_+)p^-_t(\omega_-)d\omega_+\, d\omega_-. \tag{22}$$

An example of stationary no-memory statistics is that of a binary Markovian process over a line.[2] Further, all homogeneous no-memory statistics are necessarily Markovian. However, whether all no-memory processes are Markovian or not is still an open question for this author. In Appendix C we show that the only homogeneous Markovian process in three-dimensional geometry is that with slab symmetry.

No-memory statistics are important because they offer an example case for which the infinite system of equations for computing the ensemble-averaged flux can be reduced to a finite system. This is only true for collisionless transport where the relation $\psi_\omega = \psi_{\omega_-}$ together with the local factorization (22) inherent to no-memory statistics imply that

$$\psi^+_{\beta\alpha}(\mathbf{x}) = \psi_\beta(\mathbf{x}), \quad \psi^-_{\beta\alpha}(\mathbf{x}) = \psi_\alpha(\mathbf{x}), \forall\beta. \tag{23}$$

This identity simplifies the convective term in (18b) to

$$\Delta_\alpha = Q^+_\alpha(\mathbf{x})\left[\sum_{\beta\neq\alpha} p^+_{\beta\alpha}(\mathbf{x})\psi_\beta(\mathbf{x}) - \psi_\alpha(\mathbf{x})\right],$$

and therefore closes Eq. (18a). Since we are confronted with a line problem we may adopt as the coordinate the transit time from the position of the initial or boundary condition and write $\psi_\alpha(\mathbf{x}) = \psi_\alpha(\theta_l)$. Then the corresponding integral form of the balance equation reads

$$\psi_\alpha(\theta_l) = e^{-\hat{\tau}_\alpha(0,\theta_l)}\Phi_\alpha + \int_0^{\theta_l} d\theta\, e^{-\hat{\tau}_\alpha(\theta,\theta_l)}$$
$$\times\left[Q^+_\alpha(\theta)\sum_{\beta\neq\alpha} p^+_{\beta\alpha}(\theta)\psi_\beta(\theta) + q_\alpha(\theta)\right], \tag{24}$$

where now

$$\hat{\tau}_\alpha(\theta,\theta_l) = \tau_\alpha(\theta,\theta_l) + \int_\theta^{\theta_l} d\theta'\, Q^+_\alpha(\theta').$$

Next, we turn our attention to the analysis of the renewal equations for collisionless transport and no-memory statistics. Since condition (22) must be true at every point $\mathbf{x} = (\mathbf{P},\mathbf{v})$ of the trajectory, we conclude that it is also true for the subset $\Omega_{\alpha\alpha}(\theta,\mathbf{x})$ of states that are continuously α in portion $\mathrm{Tr}(\theta,\mathbf{x})$ of the trajectory. [If ω assigns continuously α between $\mathbf{P}_\theta = (\mathbf{r}-\theta\mathbf{v}, t-\theta)$ and $\mathbf{P} = (\mathbf{r},t)$, then $\omega = (\omega_-,\omega_+)$, where ω_- is the restriction of ω to the past trajectory of $\mathbf{P}_\theta$, and ω_+ its restriction to the future trajectory of $\mathbf{P}$.] Therefore the averages $\Phi^{\theta/\mathbf{x}}_{\alpha\alpha}(\mathbf{x}_l)$ and $\psi^{\theta\mathbf{x}}_{\cdot\alpha}(\mathbf{x}_\theta)$ over spatially dependent sets of states become local averages $\Phi_\alpha(\mathbf{x}_l)$ and $\psi_{\cdot\alpha}(\mathbf{x}_\theta)$ and renewal equation (14) becomes identical to Eq. (24). We note, in particular, that at the no-memory limit the conditional probability density to change to a new material, given that the state has kept a constant material during transit time θ, becomes independent of θ:

$$\lambda_\alpha(\theta,\mathbf{x}) \to Q^+_\alpha(\mathbf{x}_\theta).$$

We end this section with an analysis of the behavior of the ensemble-averaged angular flux in the collisionless, no-memory limit for homogeneous statistics. To define in a general way what we understand by homogeneous statistics we first introduce the concept of a group-invariant statistical set: we consider a group G of transformations acting on the set $\{\mathbf{P} = (\mathbf{r},t)\,|\,\mathbf{r}\in D, t\geqslant t_0\}$ and, in the usual way, lift G to act on Ω by defining $(g\omega)(\mathbf{P}) = \omega(g^{-1}\mathbf{P})$. But nothing ensures that $g\omega$ is a state in Ω and, if such is the case, no relation is given between $p_t(\omega)$ and $p_t(g\omega)$. This situation is regularized by defining a G-invariant statistical set Ω as a set that is

complete under transformations in G, $G\Omega \subseteq \Omega$, and whose measure $p_t(\omega)$ is G-invariant:

$$\int_X d\omega\, p_t(\omega) f_\omega = \int_{gX} d\omega\, p_{gt}(\omega) f_{g^{-1}}(\omega), \quad \forall g \in G,$$

where the notation for gt is $g(\mathbf{r},t) = (g\mathbf{r},gt)$. Furthermore, we will say that G acts transitively on Ω if Ω can be recovered from a single state: $\Omega = G\omega$. If G acts transitively on Ω and if Ω is G-invariant, then for any $\mathbf{P}_a$ and $\mathbf{P}$ such that $\mathbf{P}_a = g\mathbf{P}$ we have $X_\alpha(\mathbf{P}_a) \simeq X_\alpha(\mathbf{P})$ and

$$\int_{X_\alpha(\mathbf{P})} d\omega\, p_t(\omega) f_\omega = \int_{X_\alpha(\mathbf{P}_a)} d\omega\, p_{t_a}(\omega) f_{g^{-1}}(\omega).$$

Finally, we say that Ω is homogeneous if G acts transitively on the phase space and on Ω and if Ω is G-invariant. In this case the previous result is true for any two locations $\mathbf{P}$ and $\mathbf{P}_a$ since there is always a transformation g passing from $\mathbf{P}$ to $\mathbf{P}_a$. Therefore, for homogeneous statistics local averages are independent of position and spatially-dependent averages depend only on the G-invariant properties of the sets that are involved. For instance, the operator $M_\alpha(\mathbf{P})$ does not depend on $\mathbf{P}$ and the operator $M_{\alpha\alpha}(\theta,\mathbf{x})$ depends only on θ and $\mathbf{v}$. In particular, local probabilities and transition probabilities, p_α and $p^{\pm}_{\beta\alpha}$, are constant.

Coming back to our collisionless, no-memory equation (24) we now define the group of motions over the particle trajectory

$$G = \{g_\theta | g_\theta(\mathbf{r},t) = (\mathbf{r} - \theta\mathbf{v}, t - \theta), \theta \in \mathscr{R}\}$$

and assume that the statistics Ω are homogeneous with respect to G. Moreover, we will assume also that the materials have constant cross sections. In these conditions $\hat{\tau}_\alpha(\theta,\theta_l) = (\theta_l - \theta)\hat{\lambda}_\alpha$ with $\hat{\lambda}_\alpha = v\Sigma_\alpha + \lambda_\alpha$, and Eq. (24) becomes a convolution equation. The Laplace transform of this equation gives

$$(\hat{\lambda}_\alpha + s)\hat{\psi}_\alpha(s) = \Phi_\alpha + \lambda_\alpha \sum_{\beta \neq \alpha} p^{+}_{\beta\alpha} \hat{\psi}_\beta(s) + \hat{q}_\alpha(s), \quad \forall \alpha,$$

having defined

$$\hat{f}(s) = \int_0^\infty d\theta\, e^{-s\theta} f(\theta).$$

Solving this system of equations is only a matter of algebraic manipulation. For a binary mixture with constant sources and nonstochastic initial or boundary condition Φ we obtain

$$\psi_\alpha(\theta) = \frac{1}{s_+ - s_-}\left\{\left[(v\Sigma_\alpha - s_-)\Phi + \frac{(v\tilde{\Sigma} - s_+)q_\alpha - A_\alpha}{s_+}\right]e^{-s_+\theta} - \left[(v\Sigma_\alpha - s_+)\Phi + \frac{(v\tilde{\Sigma} - s_-)q_\alpha - A_\alpha}{s_-}\right] \times e^{-s_-\theta}\right\} + \frac{v\tilde{\Sigma}q_\alpha - A_\alpha}{s_+ s_-} \tag{25a}$$

and a similar result, interchanging α and β, for $\psi_\beta(\theta)$. In this expression:

$$v\hat{\tilde{\Sigma}} = \lambda + v(p_\alpha\Sigma_\beta + p_\beta\Sigma_\alpha),$$

$$A_\alpha = p_\beta[\lambda(q_\alpha - q_\beta) + q_\alpha v(\Sigma_\alpha - \Sigma_\beta)],$$

and

$$s_\pm = (v/2)\left[\Sigma_{\text{mix}} + \tilde{\Sigma} \pm \sqrt{(\Sigma_{\text{mix}} + \tilde{\Sigma})^2 - 4(\Sigma_\alpha\Sigma_\beta + \lambda\Sigma_{\text{mix}}/v)}\right],$$

where $\lambda = \lambda_\alpha + \lambda_\beta$ and $\Sigma_{\text{mix}} = p_\alpha\Sigma_\alpha + p_\beta\Sigma_\beta$ is the homogenized cross section value. Also, we have simplified the expression of A_α with the help of the relation $p_\alpha\lambda_\alpha = p_\beta\lambda_\beta$ which is a consequence of Debye's formula[10] for binary homogeneous statistics. The expression for the ensemble-averaged flux $\psi = p_\alpha\psi_\alpha + p_\beta\psi_\beta$ is then

$$\psi(\theta) = \frac{1}{s_+ - s_-}\left\{\left[(v\Sigma_{\text{mix}} - s_-)\Phi + \frac{(v\tilde{\Sigma} - s_+)q_{\text{mix}} - A}{s_+}\right]e^{-s_+\theta} - \left[(v\Sigma_{\text{mix}} - s_+)\Phi + \frac{(v\tilde{\Sigma} - s_-)q_{\text{mix}} - A}{s_-}\right] \times e^{-s_-\theta}\right\} + \frac{v\tilde{\Sigma}_{\text{mix}} - A}{s_+ s_-}, \tag{25b}$$

where $A = p_\alpha p_\beta(\Sigma_\alpha - \Sigma_\beta)(q_\alpha - q_\beta)$. This last result has already been obtained in Ref. 2 for a homogeneous, binary Markovian description for stationary, collisionless transport.

At any point $\mathbf{P} = (\mathbf{r},t)$ the value Σ_{mix} is only a local statistical average that does not directly affect neutron motion. Actually, neutrons moving around $\mathbf{P}$ "see" the cross sections of the materials defined by a given state ω. In general, these cross sections will differ from one state to another. However, if every state ω changes very rapidly from one material to another, then for any state the transport properties will be those of the averaged state modulated with a very fast oscillating behavior characteristic of the state. For instance, $\Sigma_\omega = \Sigma_{\text{mix}} + \delta\Sigma_\omega$, here $\delta\Sigma_\omega$ is a very fast oscillating cross section whose averaged value vanishes over very short transit times. In these conditions the angular flux will also be of the form $\psi_\omega = \psi_{\text{mix}} + \delta\psi_\omega$, where again $\delta\psi_\omega$ is a very fast oscillating function. We can expect that at the limit when the frequency of the oscillating goes to infinity: $\psi_\omega \to \psi_{\text{mix}}$. This so-called atomic mix limit occurs when the transition probability density $Q^{\pm}_{\alpha}(\mathbf{x})$ goes to infinity. For the case of homogeneous statistics described by Eqs. (25) the corresponding condition is $\lambda \to \infty$. This in turn implies $s_+ \to \lambda, s_- \to v\Sigma_{\text{mix}}$ and the averages

$$\psi_\alpha(\theta), \psi_\beta(\theta), \psi(\theta) \to \Phi e^{-v\Sigma_{\text{mix}}\theta} + \frac{q}{v\Sigma_{\text{mix}}}(1 - e^{-v\Sigma_{\text{mix}}\theta})$$

behave according to a classic transport equation. Notice that, although the bulk of the states will exhibit the same asymptotic behavior, one can have states of very small probability which behave quite differently.

V. CONCLUSIONS

A discussion of linear particle transport in random media necessarily has to account for the behavior of the ensemble-average angular flux ψ. In this work we have embodied the randomness in a statistical set Ω whose elements are all the admissible physical realizations of the transport host medium as a time-dependent, heterogeneous composite of a finite set of materials. Then we have used basic properties of the statistical set to derive two independent and complete kinetic descriptions for the average flux. Each one of these descriptions consists of an infinite number of equations but fully accounts for arbitrary statistics and includes scattering in the underlying kinetic equation in a three-dimensional setting.

The first description has been obtained by ensemble averagings of the integral form of the time-dependent transport equation. Each averaging is done over a set X_α of states which share locally the same physical properties. The result is an integral, renewal-like equation for the averaged flux in terms of scattering and interface contributions which depend in turn on new flux averages over spatially dependent subsets of X_α. The coefficients in this equation are directly related to conditional chord distributions along particle trajectories. This system of equations is a complete description for the average flux and fully generalizes all previous attempts at obtaining such a result in more restricted conditions.

Our second description, on the other hand, is based on the direct ensemble-averaging of the kinetic equation and provides a completely new set of kinetic, balancelike equations for local ensemble-averages fluxes. Each one of these equations incorporates an interface term that depends on flux averages over transitional states, that is, over states that change locally of physical properties. Accordingly, the statistics in these equations appear as local transition probabilities. Because the averages over transitional states depend on increasingly higher derivatives of averaging operators we believe, however, that this formalism will be more difficult to implement than that based on the renewal equations.

In any case, since both descriptions consist of systems of equations of infinite order they are of dubious interest for most practical applications. Their real interest resides in that they provide a sound theoretical basis for developing consistent approximate formalisms. In relation to this one has to mention that for no-memory statistics of collisionless transport both the renewal and the balance descriptions reduce to a single system of equations for the local flux averages, but such reduction is only possible for purely absorbing media. A first model inspired on the properties of binary Markovian statistics has been developed for collisionless transport using only two ensemble-average fluxes[5]: the global average ψ_α over all the states that have locally the same physical properties, and the average $\psi_{\beta\alpha}$ over the states that change locally of physical properties. With regard to our renewal description in Eq. (11) such a model can be viewed as a finite system of renewal equations obtained by forcing the truncation condition

$$\psi^{\theta x}_{\beta\alpha}(\mathbf{x}_\theta) \sim \psi_{\beta\alpha}(\mathbf{x}_\theta), \tag{26}$$

that is, by neglecting the dependence on the transit time of the fluxes averaged over transition states (see Appendix B). This is of course exact for no-memory statistics where, for the collisionless case, $\psi^{\theta x}_{\beta\alpha}(\mathbf{x}_\theta) = \psi_{\beta\alpha}(\mathbf{x}_\theta) = \psi_\beta(\mathbf{x}_\theta)$. Also, a phenomenological model for stationary transport including scattering has been developed via the master equation formalism[5] but proved to be unsatisfactory for the so-called rod problem.[6] From our point of view this last model can be derived from balance equation (18) by forcing closure relation (23). For a binary mixture the resulting equations read

$$\left[L + \begin{pmatrix} v\hat{\Sigma}_\alpha & -Q^+_\alpha \\ -Q^+_\beta & v\hat{\Sigma}_\beta \end{pmatrix} - \begin{pmatrix} H_\alpha & \\ & H_\beta \end{pmatrix}\right]\begin{pmatrix}\psi_\alpha \\ \psi_\beta\end{pmatrix} = \begin{pmatrix} q_\alpha \\ q_\beta \end{pmatrix}, \tag{27}$$

where $v\hat{\Sigma}_\alpha = v\Sigma_\alpha + Q^+_\alpha$. By introducing now the average flux ψ and the cross correlation function χ[5],

$$\begin{pmatrix}\psi \\ \chi\end{pmatrix} = A\begin{pmatrix}\psi_\alpha \\ \psi_\beta\end{pmatrix} = \begin{pmatrix} p_\alpha & p_\beta \\ \sqrt{p_\alpha p_\beta} & -\sqrt{p_\alpha p_\beta}\end{pmatrix}\begin{pmatrix}\psi_\alpha \\ \psi_\beta\end{pmatrix},$$

we obtain

$$\left[L + \begin{pmatrix} v\Sigma_{\text{mix}} & v\tilde{\Sigma} + \delta \\ v\tilde{\Sigma} & v\hat{\Sigma} + \lambda_c^{-1}\end{pmatrix} - \begin{pmatrix} H_{\text{mix}} & \hat{H} \\ \hat{H} & \tilde{H}\end{pmatrix} - B\right]\begin{pmatrix}\psi \\ \chi\end{pmatrix} = \begin{pmatrix} q_{\text{mix}} \\ \hat{q}\end{pmatrix}. \tag{28}$$

In this equation:

$$\Sigma_{\text{mix}} = p_\alpha\Sigma_\alpha + p_\beta\Sigma_\beta,$$
$$\hat{\Sigma} = p_\beta\Sigma_\alpha + p_\alpha\Sigma_\beta,$$
$$\tilde{\Sigma} = \sqrt{p_\alpha p_\beta}\,(\Sigma_\alpha - \Sigma_\beta),$$

and similar expressions for H_{mix}, q_{mix}, $\hat{H}$, $\hat{q}$, and $\tilde{H}$; we have also

$$v\lambda_c^{-1} = Q^+_\alpha + Q^+_\beta, \quad \delta = (p_\alpha Q^+_\alpha - p_\beta Q^+_\beta)/\sqrt{p_\alpha p_\beta}.$$

The matrix B in Eq. (28) comes from the commutation $AL = LA - B$. This matrix is of the form

$$B = \begin{pmatrix} b_\alpha & b_\beta \\ b & -b\end{pmatrix},$$

where $b_\alpha = Q^+_\alpha - Q^-_\alpha$ and $b = (p_\beta b_\alpha + p_\alpha b_\beta)/(2\sqrt{p_\alpha p_\beta})$.

For isotropic statistics $Q^+_\alpha = Q^-_\alpha$ and matrix B vanishes from Eq. (28). For Markovian statistics v/Q^+_α is the mean chord length in material α (noted λ_α in Ref. 5) and, according to Debye's formula for binary statistics,[10] we have $\delta = 0$ so that Eq. (28) reduces to the model proposed by Levermore and quoted as Eq. (4) in Ref. 5. It is important to note, however, that (23) is only true for no-memory statistics in purely absorbing media and that Debye's formula only applies to homogeneous statistics. Therefore the inclusion of the scattering term in Eq. (28) is incorrect even for inhomogeneous Markov statistics.

With regard to Eq. (27) we will mention finally that a more tractable, self-adjoint equation can be readily obtained by operating in this equation with the matrix

$$C = \begin{pmatrix} (Q^+_\alpha)^{-1} & \\ & (Q^+_\beta)^{-1}\end{pmatrix}.$$

The result of this operation is

$$\left[CL+\begin{pmatrix} d_\alpha & -1 \\ -1 & d_\beta \end{pmatrix}-\begin{pmatrix} h_\alpha & \\ & h_\beta \end{pmatrix}\right]\begin{pmatrix}\psi_\alpha \\ \psi_\beta\end{pmatrix}=C\begin{pmatrix}q_\alpha \\ q_\beta\end{pmatrix},$$

where $h_\alpha=(Q_\alpha^+)^{-1}H_\alpha$ and $d_\alpha=v(Q_\alpha^+)^{-1}\Sigma_\alpha+1$.

Truncation assumption (26) is exact only at the limit as $\theta\to 0$ and it is clear that problems would certainly arise if the statistics favors chord distributions with long optical paths. On the other hand, this assumption can be a good approximation when the statistics enforces chords with small optical length. An improvement may result if we replace (26) with the assumption

$$\psi_{\beta\alpha}^{\partial\mathbf{x}}(\mathbf{x}_\theta)\sim\epsilon_\alpha(\theta,\mathbf{x})\psi_\beta(\mathbf{x}_\theta)+[1-\epsilon_\alpha(\theta,\mathbf{x})]\psi_{\beta\alpha}(\mathbf{x}_\theta), \tag{29}$$

where ϵ_α is a positive, rapidly decreasing function of θ with the limiting values: $\epsilon_\alpha(0,\mathbf{x})=1$, and $\epsilon_\alpha(\infty,\mathbf{x})=0$. Possible candidates for ϵ_α could be a step function of $\tau_\alpha(\theta,\mathbf{x})$ or an exponential function of $\tau_\alpha(\theta,\mathbf{x})$.

To be sound a model for stochastic transport has to respect general properties of the exact equations, such as limit behaviors, and to be validated for a general class of statistical sets. In a future paper[11] we will examine a formalism that uses balance equation (18) for the local averages ψ_α, and a truncated version of renewal equation (11) for the local transition fluxes $\psi_{\beta\alpha}$. The truncation assumptions that we have selected are of the type of (29):

$$\psi_{\beta\beta\alpha}^{\partial\mathbf{x}}(\mathbf{x}_\theta)\sim\epsilon_\beta(\theta,\mathbf{x})\psi_{\beta\alpha}(\mathbf{x}_\theta)+[1-\epsilon_\beta(\theta,\mathbf{x})]\psi_\beta(\mathbf{x}_\theta)$$

for the collision term, and

$$\psi_{\gamma\beta\alpha}^{\partial\mathbf{x}}(\mathbf{x}_\theta)\sim\epsilon_\beta(\theta,\mathbf{x})\psi_{\gamma\alpha}(\mathbf{x}_\theta)+[1-\epsilon_\beta(\theta,\mathbf{x})]\psi_{\gamma\beta}(\mathbf{x}_\theta)$$

for the interface term.

As it stands the present work provides a complete description for the expected flux ψ of particle transport in random media, but opens more questions that it answers. It is clear that there remains a large area of research for: (a) developing approximate models for practical problems, (b) analyzing in depth the renewal and balance equations to establish general properties and limit behaviors of ψ, (c) obtaining general descriptions for different underlying equations or different statistics (such as the one in Ref. 5), (d) deriving a formalism to account for statistical fluctuations, and many others. Also, the validity of the descriptions presented here stands on assumptions of the existence of averaging operators and their Gateau's derivatives and it remains to provide the necessary rigorous mathematical description and, in particular, to build the statistical set into a measurable, topological space for which the assumptions in Appendix A will be true.

ACKNOWLEDGMENTS

I first learned about stochastic transport through Gerry Pomraning's stimulating talks. Without him this work would have been impossible. Many thanks are also due to Miguel Muñoz, Cristine Magnaud, Mireille Coste, and Jean-Jacques Lautard for listening to me when I was in need of a friendly "mathematical" ear. Normal J. McCormick did a very careful reading of the manuscript and suggested many improvements. Finally I wish to thank Anna K. for many helpful discussions.

APPENDIX A: BASIC AVERAGING OPERATORS FOR STOCHASTIC KINETICS

The object of this Appendix is twofold: to define in a more precise way the operators used in this work and to outline the derivation of some of the relations between these operators.

To any measurable subset of states $X\subseteq\Omega$ we associate the time-dependent operators:

$$\begin{aligned} M_X\psi&=\int_X d\omega\, p_t(\omega)\psi_\omega, \\ A_X\psi&=M_X\psi/M_X 1. \end{aligned} \tag{A1}$$

From now on we suppress the set index X and write simply M and A. These operators are positive in the sense that if $\psi_\omega\geqslant 0$ over X then $M\psi\geqslant 0$ and $A\psi\geqslant 0$.

Let now $\Xi=\{\mathbf{P}=(\mathbf{r},t)\}$ be an arbitrary subset of the space-time phase space $\Xi\subseteq X$. Then we define $X_\alpha(\Xi)$ as the subset of states in X that are constantly equal to α on Ξ,

$$X_\alpha(\Xi)=\{\omega\in X\,|\,\omega(\mathbf{P})=\alpha,\forall\mathbf{P}\in\Xi\}, \tag{A2}$$

and denote by $M_\alpha(\Xi)$ and $A_\alpha(\Xi)$ the corresponding operators. When Ξ reduces to a single point $\mathbf{P}$, then X is the disjoint union of the $X_\alpha(\mathbf{P})$'s plus a set of measure zero (absent states) and we have

$$M\equiv\sum_\alpha M_\alpha(\mathbf{P}),\quad A\equiv\sum_\alpha p_\alpha(\mathbf{P})A_\alpha(\mathbf{P}),$$

where $p_\alpha(\mathbf{P})=M_\alpha(\mathbf{P})1/M1$ is the conditional probability for states in X to affect material α at position $\mathbf{P}$. Subsets of the phase space of particular interest to us are the particle's past $(+)$ and future $(-)$ trajectories for finite transit times. The trajectories at $\mathbf{x}=(\mathbf{P},\mathbf{v})$ for a transit time θ are the sets:

$$\mathrm{Tr}(\pm\theta,\mathbf{x})=\{\mathbf{P}_{\theta'}=(\mathbf{r}-\theta'\mathbf{v},t-\theta')\,|\,\theta'\in[0,\theta)\},\quad \theta\geqslant 0.$$

For $\Xi=\mathrm{Tr}(\pm\theta,\mathbf{x})$ we will write $X_{\alpha\alpha}(\pm\theta,\mathbf{x})$ to denote the set (A2) of states in X that are constantly α over the past $(+)$ or future $(-)$ trajectories during transit time θ. We will adopt a similar notation for the associated operators and write $M_{\alpha\alpha}(\pm\theta,\mathbf{x})$ and $A_{\alpha\alpha}(\pm\theta,\mathbf{x})$. With the usual topology in $\mathscr{R}^4$ we note that

$$\lim_{\theta\to 0_+}\mathrm{Tr}(\pm\theta,\mathbf{x})=\mathbf{P},$$

and we assume that we can construct a topology in Ω such that

$$\lim_{\theta\to 0_+}X_{\alpha\alpha}(\pm\theta,\mathbf{x})=X_\alpha(\mathbf{P}).$$

and also

$$\lim_{\theta\to 0_+}M_{\alpha\alpha}(\pm\theta,\mathbf{x})\equiv M_\alpha(\mathbf{P}).$$

We also assume the existence of the Gateau's directional derivative

$$\begin{aligned} &M_{\cdot\alpha}(\pm\theta,\mathbf{x}) \\ &\quad\equiv\mp\partial_\theta M_{\cdot\alpha}(\pm\theta,\mathbf{x}) \\ &\quad\equiv\lim_{\epsilon\to 0}(1/\epsilon)\left[M_{\alpha\alpha}(\pm\theta,\mathbf{x})-M_{\alpha\alpha}[\pm(\theta+\epsilon)\mathbf{x}]\right], \end{aligned} \tag{A3}$$

which is a positive operator in the sense earlier indicated in this Appendix. This operator can also be written as

$$M_{.\alpha}(\pm\theta,\mathbf{x})\psi=\lim_{\epsilon\to 0}\frac{1}{|\epsilon|}\int_{X^{\epsilon}_{\alpha\alpha}(\pm\theta,\mathbf{x})}d\omega\, p_t(\omega)\psi_\omega, \quad \text{(A4)}$$

where for $\epsilon\geqslant 0$:

$$X^{\epsilon}_{\alpha\alpha}(\pm\theta,\mathbf{x})=\{\omega\in X_{\alpha\alpha}(\pm\theta,\mathbf{x})\,|\,\omega(\mathbf{P}_{\pm\theta'})\neq\alpha,\exists\theta'\in[\theta,\theta+\epsilon)\}$$

is the subset of states in X which affect material α to the past $(+)$ or future $(-)$ trajectory continuously through transit time θ, and that then change material at least once during the following ϵ flying time. A similar expression can be written for $\epsilon\leqslant 0$ but, for simplicity, we will consider only the case $\epsilon\geqslant 0$.

To compute the integral in (A4) we will use local coordinates on the "surface"

$$X_{\alpha\alpha}(\pm\theta,\mathbf{x})=\{\omega\in X_{\alpha\alpha}(\pm\theta,\mathbf{x})\,|\,\omega(\mathbf{P}_{\pm\theta})\neq\alpha\}$$

and write $\omega=(\bar{\omega},\omega_\perp)$, where $\bar{\omega}\in\partial X_{\alpha\alpha}(\pm\theta,\mathbf{x})$ and $\omega_\perp$ is an orthogonal coordinate such that $d\omega=d\bar{\omega}\,d\omega_\perp$. Then for all functions ψ_ω that are continuous in ω we can write:

$$M_{.\alpha}(\pm\theta,\mathbf{x})\equiv\int_{X_{.\alpha}(\pm\theta,\mathbf{x})}d\bar{\omega}\,\bar{p}_t(\bar{\omega}), \quad \text{(A5)}$$

where

$$\bar{p}_t(\bar{\omega})=\lim_{\epsilon\to 0_+}\int_0^{\omega_\perp(\bar{\omega},\epsilon)}d\omega_\perp\, p_t(\bar{\omega},\omega_\perp).$$

Therefore, $A_{.\alpha}(\pm\theta,\mathbf{x})\equiv M_{.\alpha}(\pm\theta,\mathbf{x})/M_{.\alpha}(\pm\theta,\mathbf{x})1$ is an averaging operator over the states that remain in material α along the trajectory from $\mathbf{P}$ to $\mathbf{P}_{\pm\theta}$ and change to a different material at $\mathbf{P}_{\pm\theta}$. Furthermore, this implies the decompositions

$$\begin{aligned}M_{.\alpha}(\pm\theta,\mathbf{x})&\equiv\sum_{\beta\neq\alpha}M_{\beta\alpha}(\pm\theta,\mathbf{x}),\\ A_{.\alpha}(\pm\theta,\mathbf{x})&\equiv\sum_{\beta\neq\alpha}p_{\beta\alpha}(\pm\theta,\mathbf{x})A_{\beta\alpha}(\pm\theta,\mathbf{x}),\end{aligned} \quad \text{(A6)}$$

where $M_{\beta\alpha}$ is an operator like $M_{.\alpha}$ in (A5) but with domain $X_{\beta\alpha}(\pm\theta,\mathbf{x})$, the subset of $X_{.\alpha}(\pm\theta,\mathbf{x})$ that affects material β at $\mathbf{P}_{\pm\theta}$. The operator $A_{\beta\alpha}$ is the corresponding averaging operator, and

$$p_{\beta\alpha}(\pm\theta,\mathbf{x})=\frac{M_{\beta\alpha}(\pm\theta,\mathbf{x})1}{M_{.\alpha}(\pm\theta,\mathbf{x})1} \quad \text{(A7)}$$

is the conditional probability for material β given that the state has kept material α during transit time θ and changes material at $\mathbf{P}_{\pm\theta}$.

Also, whenever $M_{\alpha\alpha}(\pm\theta,\mathbf{x})$ is continuously differentiable in θ we have the integral representation

$$M_{\alpha\alpha}(\pm a,\mathbf{x})\equiv M_{\alpha\alpha}(\pm\theta,\mathbf{x})+\int_a^{\theta}d\theta'\,M_{.\alpha}(\pm\theta',\mathbf{x}),\quad \forall a,\theta\geqslant 0,$$

and, in particular, for $a\to 0_+$:

$$M_\alpha(\mathbf{P})\equiv M_{\alpha\alpha}(\pm\theta,\mathbf{x})+\int_0^{\theta'}d\theta'\,M_{.\alpha}(\pm\theta',\mathbf{x}),\quad \forall\theta\geqslant 0. \quad \text{(A8)}$$

In this work we have also assumed that the set of states which keep a constant value for all transit times is of measure zero:

$$\lim_{\theta\to\infty}M_{\alpha\alpha}(\pm\theta,\mathbf{x})\equiv 0, \quad \text{(A9)}$$

which is equivalent to saying that the average transit time through any given material is finite.

Observe also that for $\epsilon\to 0$ we have, to first order in ϵ,

$$M_\alpha(\mathbf{P})\cong M_{\alpha\alpha}(\epsilon,\mathbf{x})+|\epsilon|M_{.\alpha}(\epsilon',\mathbf{x}),\quad \epsilon'\in[0,\epsilon]. \quad \text{(A10)}$$

We will use this last relation to compute the commutator of the convective derivative L with the averaging operator $M_\alpha(\mathbf{P})$. By definition

$$[LM_\alpha(\mathbf{P})]\psi(\mathbf{x})=\lim_{\epsilon\to 0}(1/\epsilon)[M_\alpha(\mathbf{P}_{-\epsilon})\psi(\mathbf{x}_{-\epsilon})-M_\alpha(\mathbf{P})\psi(\mathbf{x})], \quad \text{(A11)}$$

where $\mathbf{x}_{-\epsilon}=(\mathbf{P}_{-\epsilon},\mathbf{v})$ and $\mathbf{P}_{-\epsilon}=(\mathbf{r}+\epsilon\mathbf{v},t+\epsilon)$. But, to first order in ϵ, we can write $\psi(\mathbf{x}_{-\epsilon})\simeq\psi(\mathbf{x})+\epsilon(L\psi)(\mathbf{x})$, and using (A10) first with $-\epsilon$, and second replacing $\mathbf{P}$ with $\mathbf{P}_{-\epsilon}$:

$$M_\alpha(\mathbf{P})\cong M_{\alpha\alpha}(-\epsilon,\mathbf{x})+|\epsilon|M_{.\alpha}(-\epsilon',\mathbf{x}),$$

$$M_\alpha(\mathbf{P}_{-\epsilon})\cong M_{\alpha\alpha}(\epsilon,\mathbf{x}_{-\epsilon})+|\epsilon|M_{.\alpha}(\epsilon',\mathbf{x}_{-\epsilon}),$$

where $\epsilon'\in[0,\epsilon]$. Finally, by using these expressions in (A11) and by noticing that $M_{\alpha\alpha}(-\epsilon,\mathbf{x})\equiv M_{\alpha\alpha}(\epsilon,\mathbf{x}_{-\epsilon})$, we obtain

$$LM_\alpha(\mathbf{P})\equiv M_\alpha(\mathbf{P})L+M^{+}_{.\alpha}(\mathbf{x})-M^{-}_{.\alpha}(\mathbf{x}), \quad \text{(A12)}$$

where

$$M^{\pm}_{.\alpha}(\mathbf{x})\equiv\lim_{\theta\to 0_+}M_{.\alpha}(\pm\theta,\mathbf{x}).$$

Now, since $A_\alpha(\mathbf{P})\equiv M_\alpha(\mathbf{P})/M_\alpha(\mathbf{P})1$ it is a matter of simple algebraic manipulations to obtain the result

$$LA_\alpha(\mathbf{P})\equiv A_\alpha(\mathbf{P})L+Q^{+}_{\alpha}(\mathbf{x})[A^{+}_{.\alpha}(\mathbf{x})-A_\alpha(\mathbf{P})]-Q^{-}_{\alpha}(\mathbf{x})[A^{-}_{.\alpha}(\mathbf{x})-A_\alpha(\mathbf{P})], \quad \text{(A13)}$$

where

$$A^{\pm}_{.\alpha}(\mathbf{x})\equiv\frac{M^{\pm}_{.\alpha}(\mathbf{x})}{M^{\pm}_{.\alpha}(\mathbf{x})1}, \quad \text{(A14)}$$

and where

$$Q^{\pm}_{\alpha}(\mathbf{x})\equiv\frac{M^{\pm}_{\alpha}(\mathbf{x})1}{M_\alpha(\mathbf{P})1} \quad \text{(A15)}$$

is the density of probability for states in X to enter $(+)$ or to leave $(-)$ material α at $\mathbf{P}$ per unit transit time in direction $\mathbf{v}$. When the density of probability $p_t(\omega)$ is continuous in t the local averaging operators in (A14) are symmetric: $A^{\pm}_{.\alpha}(\mathbf{P},\mathbf{v})=A^{\pm}_{.\alpha}(\mathbf{P},-\mathbf{v})$.

The formulas in this Appendix are based on formal derivations from assumptions on the continuity and differentiability of the measure operators over Ω. For a deeper under-

standing it will be necessary to relate such properties to the basic topological and material properties of Ω and of the spaces of functions and operators that have been implicitly assumed. A natural step would be to lift topology and Borel sets from phase space into the space of functions $\psi_\omega(\mathbf{x})$ and into the statistical set and from there into the set of measure operators. In particular, the notion of closeness of states $\omega \sim \bar{\omega}$ should be related to that of the corresponding particle fluxes $\psi_\omega \sim \psi_{\bar{\omega}}$ and, from the properties of the kinetic operator, to the closeness of materials $\omega(\mathbf{P}) \sim \bar{\omega}(\mathbf{P})$. In this sense we can define a quasidistance in the statistical set $d(\omega,\bar{\omega})$ as the measure of the set $\{\mathbf{P}=\mathbf{r},t)|\omega(\mathbf{P}) \neq \bar{\omega}(\mathbf{P})\}$ on which the two states are different, and then built Ω into a metric space by identifying states which differ only over a set of measure zero.

Next we consider no-memory statistics for which the set $X_\alpha(\mathbf{P})$ can be viewed as the Cartesian product $X_\alpha^-(\mathbf{x}) \times X_\alpha^+(\mathbf{x})$, where $X_\alpha^\pm(\mathbf{x})$ are the sets of states defined over the past ($-$) or over the future ($+$) trajectory and whose limit value at $\mathbf{P}$ is α. Then, since $X_\alpha^-(\mathbf{x}) \times X_\alpha^+(\mathbf{x})$ preserves also the measure, we have for a purely absorbing medium

$$M_\alpha(\mathbf{P})\psi = \int_{X_\alpha(\mathbf{P})} d\omega\, p_t(\omega)\psi_{\omega_-} = W_\alpha^+(\mathbf{x}) M_\alpha^-(\mathbf{x})\psi,$$

where

$$M_\alpha^\pm(\mathbf{x})\psi = \int_{X_\alpha^\pm(\mathbf{x})} d\omega_\pm\, p_t^\pm(\omega_\pm)\psi_{\omega_\pm}$$

and $W_\alpha^\pm(\mathbf{x}) = M_\alpha^\pm(\mathbf{x})1$. Hence $\psi_\alpha(\mathbf{x}) = A_\alpha^-(\mathbf{x})\psi$ is the averaged flux over $X_\alpha^-(\mathbf{x})$.

Consider now the set $X_{\alpha\beta}(\theta,\mathbf{x})$ that for no-memory statistics can be written as $X_\alpha^-(\mathbf{x}_\theta) \times X_{\beta\beta}^+(\theta,\mathbf{x})$, where $X_{\beta\beta}^+(\theta,\mathbf{x})$ is the subset of $X_\beta^+(\mathbf{x}_\theta)$ whose states are constant on $\mathrm{Tr}(\theta,\mathbf{x})$. As before we have

$$M_{\alpha\beta}(\theta,\mathbf{x})\psi = \int_{X_{\alpha\beta}(\theta,\mathbf{x})} d\omega\, p_t(\omega)\psi_{\omega_-} = W_\beta^+(\theta,\mathbf{x}) M_\alpha^-(\mathbf{x}_\theta)\psi,$$

with

$$W_\beta^+(\theta,\mathbf{x})\psi = \int_{X_{\beta\beta}^+(\theta,\mathbf{x})} d\omega_+\, p_t^+(\omega_t).$$

Therefore $A_{\alpha\beta}(\theta,\mathbf{x})\psi = A_\alpha^-(\mathbf{x}_\theta)\psi$ which means that $\psi_{\alpha\beta}^{\theta\mathbf{x}}(\mathbf{x}_\theta) = \psi_\alpha(\mathbf{x}_\theta)$ for arbitrary θ. Furthermore, when $\theta \to 0_+$ we also get $\psi_{\alpha\beta} = \psi_\alpha$.

APPENDIX B: STATIONARY STATISTICS

A case of practical interest occurs when particle motion is much faster than the motion of the materials composing the host medium. Then one can assume that the statistical set is stationary, $p_t(\omega) = p(\omega)$, and view the states as mappings of the spatial domain into the set of materials $\omega: D \to \{\alpha\}$. The corresponding formulation can be obtained from the general, time-dependent formulation by formally posing $\mathbf{P} = \mathbf{r}$ in all $\mathbf{P}$-dependent subsets and operators. For easy reference we summarize in this Appendix the more important results.

We note first that all dependence of subsets of states and operators on the velocity variable $\mathbf{v}$ reduces presently to a dependence on only the direction $\mathbf{\Omega}$ of $\mathbf{v}$. Indeed, by using the flight distance $x = v\theta$ instead of the transit time θ, we introduce now the notion of back ($+$) and forth ($-$) trajectories

$$\mathrm{Tr}(\pm x,\mathbf{r},\mathbf{\Omega}) = \{\mathbf{r}_{x'} = \mathbf{r} - x'\mathbf{\Omega} | x' \in [0,x)\},$$

and note that $\mathrm{Tr}(\pm x,\mathbf{r},\mathbf{\Omega}) = \mathrm{Tr}(\mp x,\mathbf{r},-\mathbf{\Omega})$ so heretofore we will consider only the back ($+$) trajectory. We characterize a state ω over the back trajectory at $\mathbf{r}$ by the maximum distance $x_\omega(\mathbf{r},\mathbf{\Omega})$ over which the state keeps on a constant material. Then the equivalent of renewal equation (11) is now

$$\psi_\alpha(\mathbf{x}) = e^{-\tau_\alpha(\mathbf{r}_l,\mathbf{r})} R_\alpha(\mathbf{r}_l,\mathbf{r}) \Phi_{\alpha\alpha}^{\mathbf{r}_l\mathbf{r}}(\mathbf{x}_l) + \int_0^{x_l} dx\, e^{-\tau_\alpha(\mathbf{r}_x,\mathbf{r})} \left[Q_\alpha(\mathbf{r}_x,\mathbf{r}) \psi_{\cdot\alpha}^{\mathbf{r}_x\mathbf{r}}(\mathbf{x}_x) + v^{-1} R_\alpha(\mathbf{r}_x,\mathbf{r}) F_{\alpha\alpha}^{\mathbf{r}_x\mathbf{r}}(\mathbf{x}_x) \right], \tag{B1}$$

where $\mathbf{r}_x = \mathbf{r} - x\mathbf{\Omega}$, $\mathbf{x}_x = (\mathbf{r}_x, t - x/v, \mathbf{v})$, and $\tau(\mathbf{r}_x,\mathbf{r})$ is the optical distance between $\mathbf{r}_x$ and $\mathbf{r}$. The mean flux

$$\psi_{\alpha\alpha}^{\mathbf{r}'\mathbf{r}} = A_{\alpha\alpha}(\mathbf{r}',\mathbf{r})\psi = \frac{M_{\alpha\alpha}(\mathbf{r}',\mathbf{r})\psi}{M_{\alpha\alpha}(\mathbf{r}',\mathbf{r})1}$$

is the average over the set of states $X_{\alpha\alpha}(\mathbf{r}',\mathbf{r})$ which affect continuously material α to the segment $(\mathbf{r}',\mathbf{r}]$, and similarly for the other mean fluxes in (B1) with

$$M_{\cdot\alpha}(\mathbf{r}_x,\mathbf{r}) \equiv -\partial_x M_{\alpha\alpha}(\mathbf{r} - x\mathbf{\Omega},\mathbf{r}).$$

Also, in Eq. (B1) we have the definitions

$$R_\alpha(\mathbf{r}_x,\mathbf{r}) = \frac{M_{\alpha\alpha}(\mathbf{r}_x,\mathbf{r})1}{M_\alpha(\mathbf{r})1},$$

$$Q_\alpha(\mathbf{r}_x,\mathbf{r}) = -\partial_x R_\alpha(\mathbf{r}_x,\mathbf{r}).$$

Note that now Q_α is a density of probability per unit distance and not per unit time. By defining $\lambda_\alpha(\mathbf{r}_x,\mathbf{r}) = -\partial_x \ln R_\alpha(\mathbf{r}_x,\mathbf{r})$ we get the equivalent of Eq. (14):

$$\psi_\alpha(\mathbf{x}) = e^{-\tilde{\tau}_\alpha(\mathbf{r}_l,\mathbf{r})} \Phi_{\alpha\alpha}^{\mathbf{r}_l\mathbf{r}}(\mathbf{x}_l) + \int_0^{x_l} dx\, e^{-\tilde{\tau}_\alpha(\mathbf{r}_x,\mathbf{r})} \times \left[\lambda_\alpha(\mathbf{r}_x,\mathbf{r}) \psi_{\cdot\alpha}^{\mathbf{r}_x\mathbf{r}}(\mathbf{x}_x) + v^{-1} F_{\alpha\alpha}^{\mathbf{r}_x\mathbf{r}}(\mathbf{x}_x) \right], \tag{B2}$$

with

$$\tilde{\tau}_\alpha(\mathbf{r}_x,\mathbf{r}) = \tau_\alpha(\mathbf{r}_x,\mathbf{r}) + \int_0^x dx'\, \lambda_\alpha(\mathbf{r}_{x'},\mathbf{r}),$$

where presently λ_α is the conditional transitional probability density per unit length.

The stationary equivalent to the balancelike equation is like Eq. (18) with the only difference that Δ_α has to be replaced by $v\Delta_\alpha$:

$$(L + v\Sigma_\alpha)\psi_\alpha = q_\alpha + H_\alpha\psi_\alpha + v\Delta_\alpha. \tag{B3}$$

We end this Appendix by showing how to derive finite systems of renewal equations by forcing a truncation condition on the infinite system of equations (B1). We consider stationary transport in a purely absorbing binary mixture.

Then we can use as spatial coordinate the distance to the boundary and write (B1) under the form

$$\psi_\alpha(x) = e^{-\tau_\alpha(0,x)} R_\alpha(0,x)\Phi^{0x}_{\alpha\alpha} + \int_0^x dy\, e^{-\tau_\alpha(y,x)} \times\left[Q_\alpha(y,x)\psi^{yx}_{\beta\alpha}(y) + R_\alpha(y,x)S_\alpha(y)\right],$$

where $S_\alpha = vq_\alpha$. Next, by integrating by parts the source term and with the help of the relation $Q_\alpha(y,x) = \partial_y R_\alpha(y,x)$ we obtain

$$\psi_\alpha(x) = R_\alpha(0,x)\phi_\alpha[0,x;\Phi^{0x}_{\alpha\alpha}] + \int_0^x d_y R_\alpha(y,x) \times \phi_\alpha[y,x;\psi^{yx}_{\beta\alpha}(y)], \tag{B4a}$$

where

$$\phi_\alpha[y,x;b] = e^{-\tau_\alpha(y,x)} b + \int_y^x dx'\, e^{-\tau_\alpha(x',x)} S_\alpha(x'), \quad y \leqslant x,$$

is the nonstochastic flux at x with an angular flux b at y when material α occupies the entire segment $(y,x]$. In Eq. (B4a) the quantity $\psi^{yx}_{\beta\alpha}(y)$ is the flux at y averaged over the set of states that change from material β to α at y and then keep material α over the entire interval $(y,x]$.

The average flux $\psi^{xz}_{\alpha\beta}(x)$ can also be calculated from an equation similar to (B4a):

$$\psi^{xz}_{\alpha\beta}(x) = R^{xz}_{\alpha\beta}(0,x)\phi_\alpha[0,x;\Phi^{0xz}_{\alpha\alpha\beta}] + \int_0^x d_y\, R^{xz}_{\alpha\beta}(y,x)\phi_\alpha[y,x;\psi^{yxz}_{\beta\alpha\beta}(y)], \tag{B4b}$$

and so on for the new average $\psi^{yxz}_{\beta\alpha\beta}(y)$. Even in the present collisionless situation this process cascades into an infinite number of equations. We know, however, that for no-memory statistics

$$\psi^{yx}_{\beta\alpha}(y) = \psi_\beta(y), \quad \Phi^{0x}_{\alpha\alpha} = \Phi_\alpha, \tag{B5}$$

and (B4a) closes in a natural way; moreover, in this case (B4a) is identical to (B4b) and to all the other equations in the cascade. As we have seen such is not the case for general statistics and one is faced with the difficult task of solving an infinite system of integral equations. In practice one can choose instead to deal with an approximate system of equations obtained from an artificial truncation. The more rudimentary example of such a truncation will be to enforce condition (B5) an obtain a system of two renewal equations of the type of (B4) for ψ_α and ψ_β. A better approximation is to use the truncation condition

$$\psi^{yxz}_{\beta\alpha\beta}(y) = \psi^{yx}_{\beta\alpha}(y) = \psi_{\beta\alpha}(y), \quad \Phi^{0xz}_{\alpha\alpha\beta} = \Phi^{0x}_{\alpha\alpha} = \Phi_\alpha, \tag{B6}$$

which gives a model with two equations of the type of (B4a) for ψ_α and ψ_β, and two equations of the type of (B4b) for $\psi_{\beta\alpha}$ and $\psi_{\alpha\beta}$. With the notation $\Psi_\alpha = \psi_{\alpha\beta}$ and $\Psi_\beta = \psi_{\beta\alpha}$ these equations read:

$$\psi_\alpha(x) = R_\alpha(0,x)\phi_\alpha[0,x;\Phi_\alpha] + \int_0^x d_y\, R_\alpha(y,x)\phi_\alpha[y,x;\Psi_\beta(y)],$$

$$\Psi_\alpha(x) = \hat{R}_\alpha(0,x)\phi_\alpha[0,x;\Phi_\alpha] + \int_0^x d_y\, \hat{R}_\alpha(y,x)\phi_\alpha[y,x;\Psi_\beta(y)]. \tag{B7}$$

This model is exact, however, for alternate renewal processes[4,5] which fulfill rigorously relations (B6) in the collisionless case.

Notice that approximation (B5) implies that we do not make any difference between transitional and nontransitional states: $\psi_\beta(y)$ is an average over states that have material β at y regardless of whether y is an end point for material β; on the other hand $\psi^{yx}_{\beta\alpha}(y)$ is an average that is done exclusively on a subset of transitional states, i.e., over states that change at y from β to α. Approximation (B6) is less drastic because it only says that an average on a subset of transitional states equals the average over all transitional states $\psi_{\beta\alpha}(y)$. Accordingly $R^{xz}_{\alpha\beta}(y,x)$ in Eq. (B4b) has been replaced by $\hat{R}_\alpha(y,x)$ in the second of Eqs. (B7). The function $\hat{R}_\alpha(y,x)$ is the probability for the states that change from α to β at x to have material α in the entire interval (y,x). We recall that $R_\alpha(y,x)$ is a similar probability but for all states, transitional and nontransitional, that have material α at x. Finally, the original $R^{xz}_{\alpha\beta}(y,x)$ is the corresponding probability but for only the transitional states that have the interval $(x,z]$ in material β.

APPENDIX C: THREE-DIMENSIONAL STATIONARY MARKOVIAN PROCESSES

Stationary Markovian processes have been used to describe the statistics of one-dimensional turbulent binary mixtures and, although such processes can be extended to a plane,[12] questions have been raised with regard to the possibility of constructing similar statistics in the three-dimensional case.[2] The purpose of this Appendix is to investigate the use of continuous, three-dimensional Markovian processes for modeling random multicomponent fluid mixtures.

For simplicity we will consider stationary statistics. We assume that the fluid consists of a random mixture of a finite number of materials, and denote by $p_\alpha(\mathbf{r})$ the probability for material α at $\mathbf{r}$, and by $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$ the conditional transition probability for material β at $\mathbf{r}$ given that material is α at $\mathbf{s}$. Our aim is to analyze the properties of continuous Markovian processes with the properties:

$$\sum_\alpha p_\alpha(\mathbf{r}) = 1, \quad \forall \mathbf{r}, \qquad p_\alpha(\mathbf{r}) = \sum_\beta p_\beta(\mathbf{s})P_{\beta\alpha}(\mathbf{s}\to\mathbf{r}), \quad \forall \mathbf{s},\mathbf{r}. \tag{C1}$$

It is customary to work with the more basic quantities $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$). From Eq. (C1) it follows that the transition probabilities obey the relations

$$P_{\alpha\beta}(\mathbf{s}\to\mathbf{r}) = \sum_\gamma P_{\alpha\gamma}(\mathbf{s}\to\mathbf{t})P_{\gamma\beta}(\mathbf{t}\to\mathbf{r}), \quad \forall \mathbf{s},\mathbf{r},\mathbf{t}, \qquad \sum_\beta P_{\alpha\beta}(\mathbf{s}\to\mathbf{r}) = 1, \qquad P_{\alpha\beta}(\mathbf{r}\to\mathbf{r}) = \delta\alpha_\beta. \tag{C2}$$

To solve these equations we assume that the $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$ are differentiable with respect to $\mathbf{r}$ and write for small $d\mathbf{r}$:

$$P_{\alpha\beta}(\mathbf{s}\to\mathbf{r}+d\mathbf{r})\sim P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})+d\mathbf{r}\cdot\nabla P_{\alpha\beta}(\mathbf{s}\to\mathbf{r}). \quad \text{(C3)}$$

Finally, from this assumption and the first equation in (C2) we obtain the well-known Chapman–Kolmogorov forward equations:

$$\nabla P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=\sum_{\gamma}P_{\alpha\gamma}(\mathbf{s}\to\mathbf{r})\mathbf{f}_{\gamma\beta}(\mathbf{r}). \quad \text{(C4)}$$

Here the

$$\mathbf{f}_{\alpha\beta}(\mathbf{r})=\lim_{(\mathbf{s}\to\mathbf{r})}\nabla P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$$

have been defined from the limit form, $\mathbf{s}\to\mathbf{r}$, of Eq. (C3):

$$P_{\alpha\beta}(\mathbf{r}\to\mathbf{r}+d\mathbf{r})\sim\delta_{\alpha\beta}+d\mathbf{r}\cdot\mathbf{f}_{\alpha\beta}(\mathbf{r}).$$

Since we must always have $0\leqslant P_{\alpha\beta}\leqslant 1$, it is clear from the last relation that the value of $\mathbf{f}_{\alpha\beta}$ depends on the way in which $\mathbf{s}\to\mathbf{r}$ or, equivalently, that one should not naively accept (C3) without imposing a "flow" condition of the type $\mathbf{r}\geqslant\mathbf{s}$. We will return to this point later.

From (C2) and the solvability condition, $\nabla\times\nabla P_{\alpha\beta}=0$, of Eq. (C4) we obtain the following relations for the $\mathbf{f}_{\alpha\beta}$'s:

$$\begin{aligned}&\sum_{\beta}\mathbf{f}_{\alpha\beta}(\mathbf{r})=0,\\ &\nabla\times\mathbf{f}_{\alpha\beta}(\mathbf{r})=0,\\ &\sum_{\gamma}\mathbf{f}_{\alpha\gamma}(\mathbf{r})\times\mathbf{f}_{\gamma\beta}(\mathbf{r})=0,\end{aligned} \quad \text{(C5)}$$

which show, in particular, that these quantities derive from a set of potentials. Note also that (C4) implies an equivalent equation for the probabilities $p_{\alpha}(\mathbf{r})$'s:

$$\nabla p_{\alpha}(\mathbf{r})=\sum_{\beta}p_{\beta}(\mathbf{r})\mathbf{f}_{\beta\alpha}(\mathbf{r}). \quad \text{(C6)}$$

In order to solve the Chapman–Kolmogorov equations we will consider only the no-memory situation

$$\mathbf{f}_{\alpha\beta}(\mathbf{r})=\mathbf{f}_{\beta}(\mathbf{r}),\quad \forall\beta\neq\alpha, \quad \text{(C7)}$$

which is certainly the case for binary mixtures. Then, with

$$\mathbf{f}(\mathbf{r})=\sum_{\alpha}\mathbf{f}_{\alpha}(\mathbf{r}),$$

properties (C5) yield

$$\mathbf{f}_{\alpha}(\mathbf{r})=\rho_{\alpha}(\mathbf{r})\mathbf{f}(\mathbf{r}), \quad \text{(C8)}$$

with

$$\begin{aligned}&\sum_{\alpha}\rho_{\alpha}(\mathbf{r})=1,\\ &\mathbf{f}(\mathbf{r})\times\nabla\rho_{\alpha}(\mathbf{r})=0,\quad\forall\alpha,\\ &\nabla\times\mathbf{f}(\mathbf{r})=0.\end{aligned} \quad \text{(C9)}$$

Hence $\mathbf{f}$ comes from a potential $\mathbf{f}=\nabla\Phi$ and the ρ_{α}'s are constant along the trajectories parallel to $\mathbf{f}$. By adopting a local coordinate system which moves with the trajectories we can write $\mathbf{r}=r\mathbf{e}+\mathbf{r}_{\perp}$ where $\mathbf{e}=\mathbf{f}/f$ is the unit vector in the direction of $\mathbf{f}$ at $\mathbf{r}$, and $\mathbf{e}\cdot\mathbf{r}_{\perp}=0$. Then $\mathbf{f}=f(\mathbf{r})\mathbf{e}$ and $\rho_{\alpha}=\rho_{\alpha}(\mathbf{r}_{\perp})$.

The interest of assumptions (C7) is that now the Chapman–Kolmogorov equations uncouple

$$\nabla P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=[\rho_{\beta}(\mathbf{r})-P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})]\mathbf{f}(\mathbf{r}), \quad \text{(C10)}$$

and can then be integrated to yield:

$$P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=e^{-\tau(s,r)}P^{0}_{\alpha\beta}(\mathbf{s}_{\perp}\to\mathbf{r}_{\perp})+(1-e^{-\tau(s,r)})\rho_{\beta}(\mathbf{r}_{\perp}). \quad \text{(C11)}$$

In this equation $P^{0}_{\alpha\beta}$ is the value of $P_{\alpha\beta}$ on the equipotential surface $\Phi(\mathbf{r})=0$. These "boundary" values must satisfy conditions (C2). Also, the $\rho_{\beta}(\mathbf{r}_{\perp})$ represents the value of ρ_{β} on the surface $\Phi(\mathbf{r})=0$, and the function τ is defined as the integral of $\mathbf{f}\cdot\mathbf{e}$ along the trajectory:

$$\tau(s,r)=\int_{s}^{r}dt\,f(t), \quad \text{(C12)}$$

where $f\geqslant 0$ is the length of $\mathbf{f}$.

The $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$ of Eq. (C11) will be actual probabilities if and only if $\tau(s,r)\geqslant 0$ which, in view of definition (C12), is equivalent to the condition $r\geqslant s$. This last condition amounts to an ordering of the spatial locations and gives the answer to the riddle posed by assumption (C3) that ought to be considered valid only for $r\geqslant s$.

We can use (C1) together with (C11) to obtain an expression for the absolute probability of presence of a given material:

$$p_{\alpha}(\mathbf{r})=e^{-\tau(0,r)}p^{0}_{\alpha}(\mathbf{r}_{\perp})+(1-e^{-\tau(0,r)})\rho_{\alpha}(\mathbf{r}_{\perp}), \quad \text{(C13)}$$

where the p^{0}_{α} are "boundary" values on the equipotential surface $\Phi(\mathbf{r})=0$ which satisfy relations (C1) on this surface. Notice that a possible solution of Eq. (13) is $p_{\alpha}(\mathbf{r})=\rho_{\alpha}(\mathbf{r}_{\perp})$ with the ρ_{α}'s satisfying conditions similar to (C1). This is indeed the unique solution for homogeneous statistics for which $\mathbf{f}_{\alpha\beta}$ must be invariant under arbitrary translations and therefore the ρ_{α}'s and $\mathbf{f}=f\mathbf{e}$ are constants. In this case $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=P_{\alpha\beta}(\mathbf{r}-\mathbf{s})$ with

$$P_{\alpha\beta}(\mathbf{x})=e^{-f\chi\mu}P^{0}_{\alpha\beta}(\mathbf{x}_{\perp})+(1-e^{-f\chi\mu})p_{\beta},\quad \mu\geqslant 0,$$

where $\chi=|\mathbf{x}|$ and $\mu=\mathbf{e}\cdot\mathbf{x}/\chi$. When the $P^{0}_{\alpha\beta}$'s are constant this solution gives the slab statistics discussed elsewhere.[2]

We can also consider isotropic statistics for which $\mathbf{f}_{\alpha\beta}$ are invariant under arbitrary rotations with center the origin of coordinates. Then we have again $p_{\alpha}=\rho_{\alpha}$ and $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$ depends only on s,r and $\mu=\mathbf{s}\cdot\mathbf{r}/(sr)$:

$$P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=e^{-\tau(s,r)}P^{0}_{\alpha\beta}(\mu)+(1-e^{-\tau(s,r)})p_{\beta}.$$

We close this Appendix by noticing that it is also possible to write the transition probabilities in terms of operators defined over the associated statistical set. Indeed, if we define the symmetrical subsets of states $\hat{X}_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=X_{\alpha}(\mathbf{s})\cap X_{\beta}(\mathbf{r})$ then

$$P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=\frac{\hat{M}_{\alpha\beta}(\mathbf{s},\mathbf{r})1}{M_{\alpha}(\mathbf{s})1},$$

where $\hat{M}_{\alpha\beta}(\mathbf{s},\mathbf{r})$ is the operator associated to $\hat{X}_{\alpha\beta}(\mathbf{s},\mathbf{r})$. Moreover

$$p_{\alpha}(\mathbf{s})P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})=p_{\beta}(\mathbf{r})P_{\alpha\beta}(\mathbf{r}\to\mathbf{s}), \quad \text{(C14)}$$

in agreement with (C1) and (C2). Furthermore from the definition of the $\mathbf{f}_{\alpha\beta}$'s and that of Q_{α} in Appendix A we obtain

$$Q_\alpha(\mathbf{r},\mathbf{\Omega}) = |\mathbf{\Omega}\cdot \sum_{\beta\neq\alpha} \mathbf{f}_{\alpha\beta}(\mathbf{r})|,$$

which with assumption (C7) reads

$$Q_\alpha(\mathbf{r},\mathbf{\Omega}) = (1-\rho_\alpha)|\mathbf{\Omega}\cdot\mathbf{f}|.$$

Note finally that in a statistical set the notion of causality is irrelevant and that therefore (C14) gives the correct definition of $P_{\alpha\beta}(\mathbf{s}\to\mathbf{r})$ for $s\geqslant r$.

[1]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, MA, 1967).
[2]C. D. Levermore, G. C. Pomraning, D. L. Sanzo, and J. Wong, J. Math. Phys. **27**, 2526 (1986).
[3]D. Vanderhaegen, J. Quant. Spectrosc. Radiat. Transfer **36**, 557 (1986).
[4]D. Vanderhaegen, J. Quant. Spectrosc. Radiat. Transfer **39**, 333 (1988).
[5]C. D. Levermore, J. Wong, and G. C. Pomraning, J. Math. Phys. **29**, 995 (1988).
[6]G. C. Pomraning, "Radiative transfer in random media with scattering," J. Quant. Spectrosc. Radiat. Transfer (in press).
[7]D. Vanderhaegen and C. Deutsch, J. Stat. Phys. **54**, 331 (1989).
[8]R. Sanchez, J. Mondot, Z. Stankovski, A. Cossic, and I. Zmijarevic, Nucl. Sci. Eng. **100**, 352 (1988).
[9]M. L. Adams, E. W. Larsen, and G. C. Pomraning, "Benchmark results for particle transport in a binary Markov statistical medium," in preparation.
[10]P. Debye and A. M. Beuche, J. Appl. Phys. **20**, 518 (1949).
[11]R. Sanchez, "A finite model for transport in stochastic media" (in preparation).
[12]P. Switzer, Ann. Math. Statist. **36**, 1859 (1965).

Sequences of point transformations and linear canonical transformations in classical and quantum mechanics

F. Leyvraz and T. H. Seligman
Instituto de Física, Laboratorio de Cuernavaca, Universidad Nacional Autonoma de México, México D.F., Mexico

(Received 15 April 1988; accepted for publication 11 January 1989)

Sequences of point transformations and canonical linear transformations are considered in classical and quantum mechanics. It is shown that the unitary representations of such transformations can be obtained, in general, in the sense that in the limit, classical behavior is retrieved. In the particular case of one point transformation combined with two linear transformations, the results found in this way are exact. A new class of differential equations is thereby solved by quadratures.

I. INTRODUCTION

The representation in quantum mechanics of classical canonical transformations is a problem set by Dirac,[1] which has been treated extensively in the literature (see, e.g., Refs. 2–4). The classical transformations themselves exhibit fairly complicated behavior as soon as we leave the linear domain.[5] Clearly, the composition of two canonical transformations is again canonical, i.e., canonical transformations form a semigroup. If we restrict ourselves to bijective transformations, the group obtained is not understood well at all. Furthermore, it is very frequently necessary to deal with nonbijective transformations. The machinery developed for this purpose, however, does not restore the group property.[6]

A further restriction to linear transformations leads to ISp $(2N,\mathbb{R})$, the inhomogeneous group of real symplectic transformations in a $2N$-dimensional space, as well as its subgroups. This group and its unitary representations are well understood. Similarly, the group of bijective point transformations, while it has a complicated structure, carries a trivial unitary representation. In this paper, we shall analyze sequences of linear transformations and point transformations. A conjecture that arbitrary transformations can be approximated by such sequences is proposed. We shall give a number of relevant examples and then proceed to show that sequences containing at most one point transformation, are of particular interest because their unitary representations are exact solutions of the quantum mechanical problem. Therefore they yield a class of differential and hyperdifferential equations that can be solved by quadratures. We display some examples for this class.

We also show that the Feynman path integral representation of time evolution may be interpreted as the representation of a particular sequence of point and linear transformation.

II. CANONICAL SEQUENCES OF LINEAR AND POINT TRANSFORMATIONS

Linear canonical transformations in two-dimensional phase space are defined as maps L such that

$$L:\begin{pmatrix}\bar{q}\\ \bar{p}\end{pmatrix}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}q\\ p\end{pmatrix},\quad ad-bc=1. \tag{2.1}$$

Thus the transformations L form the group $\mathrm{Sp}(2,\mathbb{R})$ $\simeq \mathrm{SO}(2,1)\simeq \mathrm{SU}(1,1)$. The isomorphisms are useful when the representation properties of the group are needed, which are more readily available for unitary unimodular groups for arbitrary metric than for symplectic groups. Inhomogeneities could be added, but are irrelevant in the present context.

Point transformations P in the same space are defined by

$$\mathbf{P}:\quad \bar{q}=f(q),\quad \bar{p}=P/f'(q)+\chi(q), \tag{2.2}$$

where the derivative $f'(q)$ is assumed to exist, as, otherwise, the Poisson bracket is ill defined. These transformations form a group as long as the map $\mathbf{P}$ is bijective. We now consider more general canonical transformations that are alternating sequences of L's and P's. For example, the transformation to action and angle variables for the harmonic oscillator

$$\bar{q}=\tfrac{1}{2}(p^2+q^2),\quad \bar{p}=\arctan(p/q), \tag{2.3}$$

may be written as $P_1L_2P_3L_4$ with[7]

$$\begin{aligned}
&P_1:\quad |q_1|=\tfrac{1}{2}q^2,\quad (\operatorname{sgn} q_1)p_1=p/q,\\
&L_2:\quad q_2=p_1,\quad p_2=-q_1,\\
&P_3:\quad q_3=\arctan q_2,\quad p_3=p_2(1+q_2^2),\\
&L_4:\quad q_4=-p_3,\quad p_4=q_3.
\end{aligned} \tag{2.4}$$

Similarly, we have for a linearly increasing potential,

$$\begin{aligned}
&L_1:\quad q_1=p,\quad p_1=-q,\\
&P_2:\quad q_2=q_1|q_1|,\quad p_2=p_1/2|q_1|,\\
&L_3:\quad q_3=p_2,\quad p_3=-q_2,\\
&P_4:\quad q_4=\tfrac{4}{3}q_3^3,\quad p_4=p_3/4q_3^2,\\
&L_5:\quad q_5=-p_4,\quad p_5=q_4,\\
&P_6:\quad q_6=\tfrac{1}{4}\left[\left(-\operatorname{sgn} q_5/\sqrt{1+2|q_5|^{-1/2}}\right)+1\right],\\
&\qquad p_6=8p_5|q_5|^{3/2}(1+2|q_5|^{-1/2})^{3/2},\\
&L_7:\quad q_7=p_6,\quad p_7=-q_6,\\
&P_8:\quad q_8=|q_7|,\quad p_8=(\operatorname{sgn} q_7)p_7,\\
&L_9:\quad \bar{q}=q_9=p_8,\quad \bar{p}=p_9=-q_8.
\end{aligned} \tag{2.5}$$

Note that we used the absolute value of q and the sign, in order that q and p be defined throughout the plane. This is needed in order to lift ambiguities resulting from the nonbijective nature of this transformation.[5] For the present con-

 0022-2488/89/112512-04$02.50

siderations, however, this point is not relevant.

For unbound problems we can make similar transformations to energy and time variables. Thus, for the repulsive oscillator, we have the following transformation:

$$\bar{q} = \tfrac{1}{2}(p^2 - q^2), \quad \bar{p} = -\ln|p+q|, \tag{2.6}$$

which can be decomposed as $L_1 P_2 L_3$,

$$\begin{aligned} L_1&: \quad q_1 = q + p, \quad p_1 = p, \\ P_2&: \quad q_2 = \ln|q_1|, \quad p_2 = p_1 q_1 - \tfrac{1}{2}q_1^2, \\ L_3&: \quad q_3 = p_2, \quad p_3 = -q_2. \end{aligned} \tag{2.7}$$

As a further example, the free falling particle can be transformed to energy and time variables by the following sequence of transformations:

$$\begin{aligned} L_1&: \quad q_1 = -p, \quad p_1 = q, \\ P_2&: \quad q_2 = q_1, \quad p_2 = p_1 + \tfrac{1}{2}q_1^2, \\ L_3&: \quad \bar{q} = q_3 = p_2, \quad \bar{p} = p_3 = -q_2. \end{aligned} \tag{2.8}$$

Finally we may discretize the time evolution of a Hamiltonian of the form $H = \frac{1}{2}p^2 + V(q)$ with arbitrary $V(q)$ into canonical steps,

$$q_n = q_{n-1} + p_n \Delta t, \quad p_n = p_{n-1} - V^1(q_{n-1})\Delta t. \tag{2.9}$$

These can be composed as

$$\begin{aligned} P_n&: \quad q_n' = q_{n-1}, \quad p_n' = p_{n-1} - V^1(q_{n-1})\Delta t \\ L_n&: \quad q_n = q_n' + p_n' \Delta t, \quad P_n = p_n'. \end{aligned} \tag{2.10}$$

Thus the time evolution is given by the limit of a sequence $P_1 L_1 \cdots P_N L_N$.

Thus we find that sequences of linear and point transformations have many relevant applications. It is also obvious that very complex canonical transformations can be formed in this way. We therefore venture the conjecture: Every canonical transformations can be decomposed into a sequence of alternating linear and point transformations or can be approximated by such a sequence.

The basis for such a conjecture is as yet slim, but the time evolution is more than an isolated example; it is an important and general class of transformations. If the conjecture is wrong it will be important to determine the subclass of transformations that admit the decomposition.

III. THE QUANTUM REPRESENTATION

The quantum representation of linear transformation [Eq. (2.1)] is well known[2] and given by

$$\langle q|U|\bar{q}\rangle = (1/\sqrt{2\pi|b|})\exp[(i/b)(\tfrac{1}{2}aq^2 - q\bar{q} + \tfrac{1}{2}d\bar{q}^2)], \tag{3.1}$$

while the one of the point transformations is given by[8]

$$\langle q|U|\bar{q}\rangle = \delta(\bar{q} - f(q))\sqrt{|f'(q)|}\exp(-i\phi(q)), \tag{3.2}$$

where the phase is determined by the additive term $\chi(q)$ in Eq. (2.2) through the relation $\phi(q) = \int \chi(q) d\bar{q}$.

The representation of sequences $L_1 P_2 L_3 \cdots P_k$ is then given by the integral

$$\begin{aligned} &\langle q|L_1 P_2 \cdots P_k|\bar{q}\rangle = \int dq_1 \cdots dq_{k-1}, \\ &\langle q|L_1|q_1\rangle\langle q_1|P_2|q_2\rangle \cdots \langle q_{k-1}|P_k|\bar{q}\rangle, \end{aligned} \tag{3.3}$$

where half of these integrations will be trivial because of the presence of delta functions.

Note that each individual transformation will transform the operators for the coordinates and momenta exactly like classical variables, but that this is no longer true for the sequence. In general, the nonlinear nature of the point transformations will create a quantum mechanical ordering problem once coordinates and momenta are mixed. Thus the operator transforms will only coincide with the classical transformation to the loweset order in $\hbar$. Hence the above scheme does not yield a representation of the group of sequences. As is usual in this context, the representation property is retrieved only in the classical limit. Exact results are, in this sense "accidental."

Yet the occurrence of these additional terms of higher order in $\hbar$ is of crucial importance as far as practical applications are concerned. We shall therefore introduce the concept of a precise representation of a classical transformation. By definition, such a transformation transforms the quantum mechanical operators for the coordinates and the momenta exactly as their classical counterparts. In such a case, the unitary representation will lead to exact quantum mechanical results, whereas otherwise it will only yield a semiclassical limit.

In particular, we readily see that the subset of sequences of type $L_1 P_2 L_3$ will have precise representations as a result of the fact that the quantization of $f(q)$ and $p/f'(q)$ is always unique. On the other hand, a sequence involving more than one nonlinear term will not, in general, admit such a representation, as the resulting classical functions will in general contain nonuniquely quantizable terms.

Thus the transformation to action and angle variables of the harmonic oscillator and its decomposition given in Eq. (2.3) yields for the integral

$$\begin{aligned} \langle q|U|\bar{q}\rangle = \int dq_1\, dq_2\, dq_3\, \delta\left(|q_1| - \frac{q^2}{2}\right)\sqrt{|q|} \\ \times \frac{e^{iq_2 q_1}}{\sqrt{2\pi}}\, \delta(q_3 - \arctan q_2) \frac{1}{\sqrt{1+q_2^2}} \cdot \frac{e^{-i\bar{q}q_3}}{\sqrt{2\pi}}, \end{aligned} \tag{3.4}$$

which can be simplified to

$$\begin{aligned} \langle q|U|\bar{q}\rangle = \frac{2\sqrt{|q|}}{\pi} \int dx \cos\left(\frac{q^2}{2}x\right) \frac{1}{\sqrt{1+x^2}} \\ \times \left(\frac{1+ix}{1-ix}\right)^{\bar{q}/2} \cdot \sum_{m=-\infty}^{\infty} \delta(\bar{q} - 2m). \end{aligned} \tag{3.5}$$

The nonbijective nature of the first point transformation gives rise to the separation of odd and even states as well as to the appearance of negative actions. A careful discussion of such ambiguities allows us to resolve such problems in terms of Riemann-type sheets in phase space or ambiguity spin.[4,5,9]

More importantly, however, it can be seen that the result is wrong. Indeed, it has been shown[4] that $\langle q|U|\bar{q}\rangle$ is simply the expression for $\psi_{\bar{q}}(q)$. By inspection, we find that for any $\bar{q}$ the integral diverges logarithmically as $q \to 0$. Thus we have

$$\psi_{\bar{q}}(q) = O(q^{1/2}\ln q) \quad (q \to 0), \tag{3.6}$$

in manifest contradiction to the properties of the harmonic oscillator functions.

On the other hand, as we have seen, the repulsive oscillator can be reduced to energy and time variables using only one point transformation. This gives

$$\langle q|U|\bar{q}\rangle = \int dq_1\, dq_2 \frac{e^{-(i/2)(q-q_1)^2}}{\sqrt{2\pi}} \delta(q_2 - \ln|q_1|) \times e^{iq_1^2/4} \cdot \frac{1}{\sqrt{|q_1|}} \frac{e^{iq_2\bar{q}}}{\sqrt{2\pi}}, \tag{3.7}$$

which again simplifies to

$$\langle q|U|\bar{q}\rangle = \frac{e^{-iq^2/2}}{2\pi} \int dx\, e^{-i(x^2/4 - qx)} \cdot |x|^{i\bar{q}-1/2}. \tag{3.8}$$

It is then straightforward to relate this integral to the parabolic cylinder functions (see, e.g., Whittaker and Watson,[10] p. 349) or else to insert it in the Schrödinger equation and verify that it is indeed exactly satisfied.

The case of the free falling particle is handled similarly. One has

$$\begin{aligned}\langle q|U|\bar{q}\rangle &= \int dq_1\, dq_2 \frac{e^{-iqq_1}}{\sqrt{2\pi}} e^{-iq_2^2/b} \delta(q_1 - q_2) \frac{e^{i\bar{q}q_2}}{\sqrt{2\pi}} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dx\, e^{ix(\bar{q}-q)} e^{-ix^3/b} \\ &= 2^{-1/3}\mathrm{Ai}(2^{1/3}(q-\bar{q})),\end{aligned} \tag{3.9}$$

which is indeed the exact solution. Thus the above examples vindicate the claim that no ordering problems occur in combinations of the type $L_1P_2L_3$, while they do occur in the combination $L_1P_2L_3P_4$, which is next in order of complexity.

If we finally consider a single step of the time evolution transformation (2.9), it is readily represented by

$$\begin{aligned}\langle q_n|U|q_{n-1}\rangle &= \int dq_n' \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{i}{2\Delta t}(q_n - q_n')^2\right) \\ &\quad \times \delta(q_n' - q_{n-1}) e^{iV(q_{n-1})\Delta t} \\ &= \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-i\Delta t\left(\frac{1}{2}\left(\frac{q_n - q_{n-1}}{\Delta t}\right)^2 \right.\right. \\ &\quad \left.\left. - V(q_{n-1})\right)\right].\end{aligned} \tag{3.10}$$

Note that, in the limit $\Delta t \to 0$, this becomes $e^{-iL(q,\dot{q})\Delta t}$, where $L(q,\dot{q})$ is the Lagrangian of the system.

Thus the full time evolution operator is obtained as a limit of successive iterations of the transformations (2.10) and we recover the Feynman path integral representation. Furthermore, we know this representation to be precise. Note that this does not imply that the representation of a finite product $P_1L_1 \cdots P_NP_N$ is also precise.

IV. THE MOST GENERAL *LPL* TRANSFORMATION

The procedure outlined above can in fact be inverted: it is not hard to write down the most complicated form of p and q that can occur (as $\bar{p}$ or $\bar{q}$) as the result of a transformation of the type $L_1P_2L_3$. It is given by

$$F(p,q) = (cp + dq) f(ap + bq) + g(ap + bq), \tag{4.1}$$

where $f(x)$ and $g(x)$ are arbitrary functions of a, b, c, and d are real constants with $ad - bc \neq 0$ (or, without loss of generality, $ad - bc = 1$).

Thus a Hamiltonian of the form (4.1) can be diagonalized by quadratures. This means, in other words, that the entire class of differential and hyperdifferential equations given by Eq. (4.1) can be solved in closed form. While this is an interesting result of the theory of canonical transformations developed above, it also has a far simpler derivation: consider a basis where the operator $ap + bq$ is diagonal (i.e., is the multiplication operator). If $ad - bc = 1$, one has

$$[ap + bq,\ cp + dq] = i, \tag{4.2}$$

so that, calling this basis $|\lambda\rangle$, we have for $\psi = \int a(\lambda)|\lambda\rangle$,

$$\begin{aligned}(ap + bq)a(\lambda) &= \lambda a(\lambda), \\ (cp + dq)a(\lambda) &= \frac{1}{i}\frac{d}{d\lambda} a(\lambda),\end{aligned} \tag{4.3}$$

and hence the operator defined above becomes

$$\frac{1}{2}\left(\frac{1}{i}\frac{d}{d\lambda} f(\lambda) + f(\lambda)\frac{1}{i}\frac{d}{d\lambda}\right) + g(\lambda). \tag{4.4}$$

This is a linear differential operator of the first order, which is readily solved by quadratures, while the basis $|\lambda\rangle$ can readily be expressed in terms of elementary functions.

We have seen that the repulsive oscillator, the free falling particle, and the free particle belong to this class of Hamiltonians. Unfortunately, it can be shown that the only Hamiltonians of the form $H = T(p) + V(q)$, which can be cast into the form of Eq. (4.1) are the following: (i) $T(p)$ or $V(q)$ is linear and the other arbitrary; (ii) $T(p)$ and $V(q)$ are quadratic polynomials of opposite sign. Thus there are no physically relevant Hamiltonians significantly different from the two mentioned above.

To show this, note first that the first case is obvious. We therefore limit ourselves to the case where neither $T(p)$ nor $V(q)$ are linear. We can further assume that $a \neq 0$, since otherwise it is impossible to construct Hamiltonians that are not linear in p. But from Eq. (4.1) follows

$$F((1/a)(C - bq), q) = A(C)q + B(C), \tag{4.5}$$

for any value of C. Applying this to the above Hamiltonian one gets

$$T((1/a)(C - bq)) + V(q) = A(C)_q + B(C), \tag{4.6}$$

and taking the derivative with respect to C,

$$T'\left(\frac{1}{a}(C - bq)\right) = a\left(\frac{dA}{dC} q + \frac{dB}{dC}\right), \tag{4.7}$$

which is only possible if $T(p)$ is quadratic (since we have excluded the linear case). This, however, leads immediately to the fact that $V(q)$ is also quadratic in q, thus proving the claim. If a, b, c and d are real further it is easy to see that the leading coefficient in $V(q)$ must be negative, showing that the attractive oscillator cannot be incorporated in this formalism without using complex extensions. Such extensions have been shown to lead to the algebraic solution of the oscillator by introducing Bargmann–Hilbert spaces.[11] If we apply them in the spirit of this paper, we end up with the well-known expression of Hermite polynomials in terms of parabolic cylinder functions.

V. CONCLUSIONS

We show that sequences of linear and point transformations describe a number of relevant canonical transformations; a conjecture was put forward that such sequences may approximate an arbitrary canonical transformation, but a proof is not given. On the other hand, the subset of such transformations is readily represented by unitary transformations or, more generally, isometries. This is obtained by folding the representations for each member of the sequence. If the conjecture is true, this result is of great importance, because it solves the representation problem by quadratures. Should the conjecture prove false, we still have solved this problem for a large and relevant class of transformations.

We show that only a few examples correspond to Hamiltonians of the type $\frac{1}{2}p^2 + V(q)$ and these are discussed explicitly. On the other hand, interesting special cases that are not of this type may be found in this class of transformations.

The concept of precise representation has thus proven useful, and we may study whether other classes of transformations exist for which we can give precise representations.

[1]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U.P., London, 1947).
[2]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971); C. Itzykson, Commun. Math. Phys. **4**, 92 (1967).
[3]M. Moshinsky, in *Group Theory and its Applications in Physics*, Proceedings, 1980 (American Institute of Physics, New York, 1981), p. 312.
[4]M. Moshinsky and T. H. Seligman, Ann. Phys. (N.Y.) **114**, 243 (1978); I. Deenen, M. Moshinsky, and T. H. Seligman, *ibid.* **127**, 458 (1980).
[5]M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1338 (1981).
[6]A. Antillon *et al.*, J. Phys. A Gen. Phys. **15**, 2659 (1982).
[7]G. Espinoza, Ph.D. thesis, Universidad Nacional Autonoma de México (1980).
[8]K. B. Wolf, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. III.
[9]J. F. Plebański and T. H. Seligman, Rep. Math. Phys. **17**, 437 (1980).
[10]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1927).
[11]P. Kramer *et al.* in *Group Theory and its Applications* (Academic, New York, 1975).

The equations for isentropic motion of inviscid fluid in terms of wave function

Yu. A. Rylov
Institute for Problems in Mechanics, USSR Academy of Sciences, Prospect Vernadskogo 101, 117526 Moscow, Union of Soviet Socialist Republics

(Received 23 February 1989; accepted for publication 26 April 1989)

It is shown that by change of variables, the equations of motion of the inviscid fluid can transform into a differential equation for a wave function ψ, consisting of two complex components. A unit spin vector $\mathbf{s}$ can be formed by means of the wave function. The fluid flow vorticity is determined only by the spin vector. The dynamical equation for the ψ reduces to a linear equation with constant coefficients, if the flow is potential ($\mathbf{s}=\text{const}$) and the state equation has some special form. Use of the wave function ψ reduces the nonlinearity connected with the convective term $(\mathbf{v}\nabla)\mathbf{v}$. It means that the velocity expansion over powers of ϵ ($\epsilon \ll 1$) associates with the wave function expansion over powers of ϵ^2.

I. INTRODUCTION

It is known[1] that the wave function satisfying the Schrödinger equation describes a potential flow of some inviscid fluid. But a solution of the reciprocal problem of describing an arbitrary (nonpotential, generally speaking) flow in terms of the wave function meets some difficulties, because of a discrepancy in the number of variables (four variables ρ, $\mathbf{v}$ for the hydrodynamical description and two variables Re ψ, Im ψ for the quantum description). In the present paper this problem is solved by means of doubling the wave function components. It leads to appearance of the spin and spinors.

The equations of isentropic motion of the inviscid fluid have the form

$$\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\nabla)\mathbf{v}=\mathbf{F}\,,\qquad \frac{\partial \rho}{\partial t}+\nabla(\rho\mathbf{v})=0\,, \tag{1.1}$$

$$\mathbf{F}=\{F^1,F^2,F^3\},\quad F^\alpha=-(1/\rho)\partial_\beta P^{\alpha\beta},\quad \alpha=1,2,3\,,$$

$$P^{\alpha\beta}=\delta_{\alpha\beta}\left(\rho\frac{\partial U}{\partial \rho}-U+\frac{\partial U}{\partial \rho_\gamma}\rho_\gamma\right)-\rho\partial_\alpha\frac{\partial U}{\partial \rho_\beta}\,,\qquad \alpha,\beta=1,2,3\,, \tag{1.2}$$

$$\rho_\beta\equiv\partial_\beta\rho,\quad \partial_i\equiv\frac{\partial}{\partial x^i},\quad i=0,1,2,3\,,$$

where $x^0=t$ is the time, $\mathbf{x}=\{x^1,x^2,x^3\}$ is the position vector, and ρ, $\mathbf{v}=\{v^1,v^2,v^3\}$ are correspondently the density and the velocity of the fluid. $U=U(\rho,\nabla\rho,S)$ is the internal energy of the unit volume. The U depends on the density ρ, on the density gradient $\nabla\rho$, and on the entropy density S, which is supposed to be constant. Furthermore, the dependence of S will be omitted. In the special case when the U does not depend on $\nabla\rho$, the fluid is barotropic, and the stress tensor $P^{\alpha\beta}$ becomes isotropic. In relations (1.2), and later on, the summation is made over repeated latin sub- and superscripts from 0 to 3 and over Greek ones from 1 to 3.

Let us change variables,

$$j^0\equiv\rho=\psi^*\psi\,,\qquad \mathbf{j}\equiv\rho\mathbf{v}=-(ia/2)(\psi^*\nabla\psi-\nabla\psi^*\psi)\,, \tag{1.3}$$

where ψ and ψ^* are correspondently a column and a line of two complex variables ψ_1 and ψ_2,

$$\psi=\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix},\quad \psi^*=(\psi_1^*,\psi_2^*),\quad f^*g\equiv f_1^*g_1+f_2^*g_2\,. \tag{1.4}$$

An asterisk denotes the complex conjugate. The a is a constant, having dimensionality of the kinematic viscosity (cm^2/s).

Now let us consider the equation for ψ,

$$ia\frac{\partial\psi}{\partial t}=\hat{L}\psi,\quad \hat{L}=\hat{L}_0+\hat{L}_D+\hat{L}_c\,,$$

$$\hat{L}_0=-\frac{a^2}{2}\nabla^2,\quad \hat{L}_D=\hat{L}_D(\psi)=U'(\psi^*\psi)\,, \tag{1.5}$$

$$\hat{L}_c=\hat{L}_c(\psi)=\frac{a^2}{2}\frac{1}{\sqrt{\rho}}\nabla^2\sqrt{\rho}-\frac{a^2}{8}\nabla s_\alpha\nabla s_\alpha+\frac{a^2}{4}\left[\frac{\nabla\rho}{\rho}\nabla s_\alpha\sigma_\alpha-\nabla^2 s_\alpha(s_\alpha-\sigma_\alpha)\right],$$

where $\boldsymbol{\sigma}=\{\sigma_1,\sigma_2,\sigma_3\}$ are 2×2 complex matrices known as spin matrices or Pauli matrices and σ_0 is the unit matrix:

$$\sigma_0=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

$$\sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}, \tag{1.6}$$

$$\rho=\psi^*\psi,\quad s_\alpha=\psi^*\sigma_\alpha\psi/\psi^*\psi,\quad \alpha=1,2,3,\quad s_0=1\,,$$

$$U'=U'(\rho)\equiv\frac{\partial U}{\partial\rho}-\partial_\alpha\frac{\partial U}{\partial\rho_\alpha}\,. \tag{1.7}$$

The variable ψ will be called the wave function. The fact is that in the special case, when

$$U=\frac{a^2}{8}\frac{(\nabla\rho)^2}{\rho},\quad U'=-\frac{a^2}{2\sqrt{\rho}}\nabla^2\sqrt{\rho},\quad \psi_2=0\,, \tag{1.8}$$

one obtains $s_\alpha=\text{const}$ ($\alpha=1,2,3$). As a result, all terms depending on ψ disappear in the operator $\hat{L}$. Setting $a=\hbar/m$, $\psi_1=f/m$, where $\hbar$ is the Planck's constant, one obtains instead of (1.5) the Schrödinger equation

$$i\hbar\frac{\partial f}{\partial t}=-\frac{\hbar^2}{2m}\nabla^2 f \tag{1.9}$$

 0022-2488/89/112516-05$02.50

for wave function f, which describes the motion of a spinless particle with the mass m.

If ψ is a solution of the system of equations (1.3), where ρ, $\mathbf{v}$ is a solution of the system (1.1), then the ψ satisfies Eq. (1.5). Vice versa, if ψ is a solution of Eq. (1.5), then ρ, $\mathbf{v}$ determined by Eq. (1.3) satisfy the system (1.1). A direct verification of these statements is rather bulky and complicated. For instance, a direct substitution of Eqs. (1.3) into Eqs. (1.1) leads to four equations. Each term of these equations is an even combination of ψ, whereas each term of Eq. (1.5) is an odd combination of ψ. In other words, to obtain Eq. (1.5), some additional nontrivial transformation is needed.

Strictly speaking, one cannot say that the system (1.1) is equivalent to Eq. (1.5), because every solution of Eqs. (1.1) associates with a set of solutions of Eq. (1.5), whereas every solution of Eq. (1.5) associates with one and only one solution of Eqs. (1.1). It is associated with the fact that the relation (1.3) between ρ, $\mathbf{v}$, and ψ is differential in terms of ψ, but it is finite in terms of ρ, $\mathbf{v}$.

Let us constitute a system I of Eqs. (1.1)–(1.3), and a system II of Eqs. (1.3), (1.5)–(1.7). Then every solution ρ, $\mathbf{v}$, ψ of system I is a solution of system II. Vice versa, every solution ρ, $\mathbf{v}$, ψ of system II is a solution of system I. The following theorem takes place.

Theorem 1: The system of equations (1.1)–(1.3) for variables ρ, $\mathbf{v}$, ψ is equivalent to the system of equations (1.3), (1.5)–(1.7) for the same variables ρ, $\mathbf{v}$, ψ.

Let us show that both system I and system II are systems of Eulerian equations (extremals) for the same variational problem. Such a way of proving the theorem seems the most simple.

Proof of Theorem 1: Equations (1.1) and (1.3) can be obtained as extremals of the variational problem with the action functional

$$S[\rho,\mathbf{j},\varphi,\boldsymbol{\xi}] = \int \left\{ \frac{\mathbf{j}^2}{2\rho} - U(\rho) - a j^i (\partial_i \varphi + \epsilon_{\alpha\beta\gamma} b_\alpha \eta_\beta \partial_i \eta_\gamma) \right\} d^4x , \tag{1.10}$$

where

$$j^0 = \rho, \quad \mathbf{j} = \rho\mathbf{v}, \quad \boldsymbol{\xi} = \{\xi_1,\xi_2,\xi_3\}, \quad \boldsymbol{\eta} = \boldsymbol{\xi}/|\boldsymbol{\xi}| ,$$
$$\mathbf{b} = \text{const}, \quad \mathbf{b}^2 = 1, \quad \partial_i \equiv \frac{\partial}{\partial x^i}, \quad i = 0,1,2,3 . \tag{1.11}$$

Here, $\epsilon_{\alpha\beta\gamma}$ is the Levi-Civita pseudotensor ($\epsilon_{123} = 1$) and $\boldsymbol{\xi}$ are Lagrangian coordinates numbering the fluid particles. Among variables φ, $\boldsymbol{\eta}$ only three are independent, because $\boldsymbol{\eta}^2 = 1$. The variables φ, $\boldsymbol{\eta}$ are related to the wave function ψ by means of relations

$$\psi = \sqrt{\rho} e^{i\varphi} \eta_\alpha \sigma_\alpha \chi, \quad \psi^* = \chi^* \sigma_\alpha \eta_\alpha e^{-i\varphi} \sqrt{\rho} , \tag{1.12}$$

where

$$\chi = \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} = \text{const}, \quad \chi^* = (\chi_1^*, \chi_2^*), \quad \chi^* \chi = 1 , \tag{1.13}$$

$$b_\alpha = \chi^* \sigma_\alpha \chi, \quad \alpha = 1,2,3, \quad b_\alpha b_\alpha = 1 , \tag{1.14}$$

the last Eq. (1.14) being a corollary of Eq. (1.13) and of the identity

$$(\chi^* \sigma_\alpha \chi)(\chi^* \sigma_\alpha \chi) \equiv (\chi^* \chi)^2 . \tag{1.15}$$

Equations (1.12) determine one-to-one relation between ψ and ρ, φ, $\boldsymbol{\eta}$ ($\rho \geqslant 0$, $|\varphi| \leqslant \pi$).

Varying (1.10), one obtains

$$\begin{aligned} \frac{\delta S}{\delta \rho} &= -\frac{\mathbf{j}^2}{2\rho^2} - U'(\rho) - a(\partial_0 \varphi + \epsilon_{\alpha\beta\gamma} b_\alpha \eta_\beta \partial_0 \eta_\gamma) \\ &= 0 , \\ \frac{\delta S}{\delta j^\mu} &= \frac{j^\mu}{\rho} - a(\partial_\mu \varphi + \epsilon_{\alpha\beta\gamma} b_\alpha \eta_\beta \partial_\mu \eta_\gamma) \\ &= 0, \quad \mu = 1,2,3 , \\ \frac{\delta S}{\delta \varphi} &= a \partial_i j^i = 0 , \\ \frac{\delta S}{\delta \xi_\mu} &= -2a j^i \epsilon_{\alpha\beta\gamma} b_\alpha \partial_i \eta_\gamma \frac{\delta_\mu^\beta - \eta_\beta \eta_\mu}{|\boldsymbol{\xi}|} \\ &= 0, \quad \mu = 1,2,3 . \end{aligned} \tag{1.16}$$

Eliminating φ from the first two equations in (1.16), one obtains

$$\frac{\partial v^\mu}{\partial t} + \partial_\mu \left(\frac{\mathbf{v}^2}{2} \right) + \partial_\mu U'(\rho) = 2a \epsilon_{\alpha\beta\gamma} b_\alpha \partial_0 \eta_\beta \partial_\mu \eta_\gamma, \quad \mu = 1,2,3 , \tag{1.17}$$

$$\partial_\nu v^\mu - \partial_\mu v^\nu = 2a \epsilon_{\alpha\beta\gamma} b_\alpha \partial_\nu \eta_\beta \partial_\mu \eta_\gamma, \quad \mu,\nu = 1,2,3 . \tag{1.18}$$

Let us convolute the last equation in (1.16) with $\partial_\nu \eta_\mu$ and take into account that because of $\boldsymbol{\eta}^2 = 1$,

$$\eta_\mu \partial_i \eta_\mu = 0 . \tag{1.19}$$

Then one obtains

$$j^i \epsilon_{\alpha\beta\gamma} b_\alpha \partial_\nu \eta_\beta \partial_i \eta_\gamma = 0 . \tag{1.20}$$

Let us eliminate η_μ from Eqs. (1.17) and (1.18). For this purpose let us multiply Eq. (1.17) by $j^0 = \rho$, Eq. (1.18) by j^ν, and add them. Due to Eq. (1.20), the right-hand side of the equation vanishes. After transformations, the first equation in (1.1) arises. The third equation in (1.16) coincides with the second equation in (1.1). The relations (1.3) result from Eq. (1.12) and from the second equation in (1.16).

Thus system I of Eqs. (1.1), (1.3) describes extremals of the functional (1.10).

Let us add the conditions (1.3) to the variational problem (1.10) by means of the Lagrangian multipliers q_0 and q_α ($\alpha = 1,2,3$). Equations (1.3) are necessary conditions of the functional (1.10) extremum. Hence joining Eqs. (1.3) as additional conditions, one does not change the variational problem. As a result, one obtains instead of Eq. (1.10),

$$\begin{aligned} S_2[\rho,\mathbf{j},\psi,q] = \int d^4x \Big\{ & \frac{ia}{2}(\psi^* \partial_0 \psi - \partial_0 \psi^* \psi) \\ & - \frac{a^2}{8} \frac{[i(\psi^* \nabla \psi - \nabla \psi^* \psi)]^2}{\psi^* \psi} \\ & - U(\psi^* \psi) + q_0(\rho - \psi^* \psi) \\ & + q_\alpha \left[j^\alpha + \frac{ia}{2}(\psi^* \partial_\alpha \psi - \partial_\alpha \psi^* \psi) \right] \Big\} . \end{aligned} \tag{1.21}$$

By means of the identity

$$[i(\psi^*\nabla\psi - \nabla\psi^*\psi)]^2$$
$$\equiv 4(\psi^*\psi)(\nabla\psi^*\nabla\psi)$$
$$- [\nabla(\psi^*\psi)]^2 - (\psi^*\psi)\nabla s_\alpha \nabla s_\alpha , \quad (1.22)$$

where s_α is defined by Eq. (1.7), the functional (1.21) can be represented in the form

$$S_2[\rho,\mathbf{j},\psi,q] = \int d^4x \left\{ \frac{ia}{2}(\psi^*\partial_0\psi - \partial_0\psi^*\psi) - \frac{a^2}{2}\nabla\psi^*\nabla\psi \right.$$
$$+ \frac{a^2}{8}\left[\frac{[\nabla(\psi^*\psi)]^2}{\psi^*\psi} + \psi^*\psi\nabla s_\alpha \nabla s_\alpha\right]$$
$$- U(\psi^*\psi) + q_0(\rho - \psi^*\psi)$$
$$\left. + q_\alpha\left[j^\alpha + \frac{ia}{2}(\psi^*\partial_\alpha\psi - \partial_\alpha\psi^*\psi)\right]\right\} . \quad (1.23)$$

Variation over j^i leads to

$$q_i = 0, \quad i = 0,1,2,3 . \quad (1.24)$$

Variating (1.23) over q_i and ψ, and taking into account designations (1.7) and Eq. (1.24), one obtains correspondently Eq. (1.3) and Eq. (1.5).

Thus system II of Eqs. (1.3), (1.5) describes extremals of the action functional (1.10). It proves the theorem.

Let us refer to the unit vector $\mathbf{s} = \{s_1,s_2,s_3\}$ with components defined by Eqs. (1.7) as the spin vector, or merely spin. As a result of Eq. (1.15),

$$\mathbf{s}^2 = s_0^2 = 1 . \quad (1.25)$$

Theorem 2: The unit spin vector $\mathbf{s}$ is constant for any fluid particle of the inviscid fluid,

$$\frac{d\mathbf{s}}{dt} \equiv \frac{\partial \mathbf{s}}{\partial t} + (\mathbf{v}\nabla)\mathbf{s} = 0 . \quad (1.26)$$

Proof: The action functional (1.21) is invariant with respect to transformation

$$\psi \to \psi' = e^{i\omega^k\sigma_k}\psi, \quad \psi^* \to \psi^{*\prime} = \psi^* e^{-i\omega^k\sigma_k}, \quad (1.27)$$

where $\omega^k = \text{const}$ $(k = 0,1,2,3)$ are the transformation parameters. The action (1.23) is invariant with respect to transformation (1.27) also. As a result of Nëther's theorem[2] one obtains the conservation laws of the form

$$\frac{\partial}{\partial t}(\rho s_i) + \nabla(\rho \mathbf{v} s_i) = 0, \quad i = 0,1,2,3 . \quad (1.28)$$

For $i = 0$ Eq. (1.28) leads to the second equation in (1.1). For $i = 1,2,3$ Eqs. (1.28) coincide with Eq. (1.26) due to Eq. (1.28) for $i = 0$.

Theorem 3: The vorticity of the inviscid fluid is determined by the field of the unit spin vector $\mathbf{s}$,

$$\nabla\times\mathbf{v} = \frac{a}{12}\left\{\frac{1}{s_3}[\nabla s_1 \times \nabla s_2] + \frac{1}{s_2}[\nabla s_3 \times \nabla s_1]\right.$$
$$\left. + \frac{1}{s_1}[\nabla s_2 \times \nabla s_3]\right\} . \quad (1.29)$$

Proof: It follows from Eq. (1.3) and from the second equation (1.16) that

$$\nabla\times\mathbf{v} = a\{b_1[\nabla\eta_2\times\nabla\eta_3] + b_2[\nabla\eta_3\times\nabla\eta_1]$$
$$+ b_3[\nabla\eta_1\times\nabla\eta_2]\} , \quad (1.30)$$

i.e., curl $\mathbf{v}$ depends on $\boldsymbol{\eta}$, but not on ρ and φ. Using Eq. (1.1) and the properties of Pauli matrices,

$$\sigma_\alpha\sigma_\beta = \delta_{\alpha\beta}\sigma_0 + i\epsilon_{\alpha\beta\gamma}\sigma_\gamma, \quad \alpha,\beta = 1,2,3 , \quad (1.31)$$

one can obtain the relation between unit vectors $\mathbf{s}$, $\boldsymbol{\eta}$, $\mathbf{b}$. It has the form

$$\boldsymbol{\eta} = (\mathbf{b} + \mathbf{s})/\sqrt{2(1 + \mathbf{bs})} . \quad (1.32)$$

From a geometrical standpoint, it means that $\boldsymbol{\eta}$ is directed along the bisectrix of the angle between the constant vector $\mathbf{b}$ and the spin $\mathbf{s}$. Substituting Eq. (1.32) into Eq. (1.30), one obtains after some calculations the relation (1.29). As one should expect, the curl $\mathbf{v}$ does not depend on a choice of the constant vector $\mathbf{b}$, i.e., on the wave function representation. The vorticity is determined completely by the spin $\mathbf{s}$, but the reverse statement is not valid, generally speaking, because the spin $\mathbf{s}$ depends on the wave function representation.

It is worthwhile to note a more interesting circumstance. According to Eqs. (1.3) and (1.29), both $\mathbf{v}$ and curl $\mathbf{v}$ are expressed through first-order derivatives of the wave function. All second-order derivatives of ψ arising at action of the operation curl upon $\mathbf{v}$ vanish.

II. CONNECTION BETWEEN DIFFERENT WAYS OF DESCRIBING THE INVISCID FLUID

Usually such concepts as the spin and the wave function are associated with quantum mechanics. Their appearance in the continuous medium mechanics seems to be rather unexpected. To explain this question, let us try to establish a connection between the Lagrangian description (LD), the Eulerian description (ED), and the description in terms of wave function (DTWF). In this section only, the case will be considered when U does not depend on $\nabla\rho$, i.e.,

$$U'(\rho) = \frac{dU(\rho)}{d\rho} .$$

The system of equations of LD has the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(t,\mathbf{x}), \quad \frac{d\mathbf{v}}{dt} = \mathbf{F} . \quad (2.1)$$

It contains six equations for six dependent variables $\mathbf{x}$, $\mathbf{v}$, which depend on four independent variables t, $\boldsymbol{\xi}$. The $\boldsymbol{\xi} = \{\xi_1,\xi_2,\xi_3\}$ are the Lagrangian coordinates, which number the fluid particles.

The system of the ED equations has the form (1.1). It represents four equations for four dependent variables ρ, $\mathbf{v}$ which depend on four independent variables t, $\mathbf{x}$. The LD contains more equations than the ED, and solutions of the system (2.1) describe the trajectories and velocities of fluid particles, whereas a solution $\mathbf{v} = \mathbf{v}(t,\mathbf{x})$ of the system (1.1) describes only the fluid particle velocities. For describing the trajectories, it is more necessary to integrate the system of ordinary differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(t,\mathbf{x}) , \quad (2.2)$$

where $\mathbf{v}(t,\mathbf{x})$ is some solution of Eqs. (1.1).

The ED can be obtained from the LD by substitution of independent variables t, $\boldsymbol{\xi}$ by t, $\mathbf{x}$. At such a transformation, the system (2.1) takes the form

$$j^i = \frac{\partial J}{\partial \xi_{0,i}}$$

$$\equiv \frac{\partial(x^i,\xi_1,\xi_2,\xi_3)}{\partial(t,x^1,x^2,x^3)}, \quad i=0,1,2,3, \quad j^i=\{\rho,\rho\mathbf{v}\}, \quad (2.3)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\mathbf{v} = \mathbf{F}, \quad (2.4)$$

where

$$J = \det\|\xi_{i,k}\|$$

$$\equiv \frac{\partial(\xi_0,\xi_1,\xi_2,\xi_3)}{\partial(x^0,x^1,x^2,x^3)}, \quad \xi_{i,k} \equiv \partial_k \xi_i, \quad i,k=0,1,2,3 \quad (2.5)$$

is the Jacobian, which is a four-linear function of variables $\xi_{i,k} \equiv \partial_k \xi_i$ $(i,k=0,1,2,3)$, the ξ_0 being the time Lagrangian coordinate. The ξ_0 is fictitious, and it is not contained in Eqs. (2.3) and (2.4). The term ρ denotes the Jacobian,

$$\rho = \frac{\partial J}{\partial \xi_{0,0}} = \frac{\partial(\xi_1,\xi_2,\xi_3)}{\partial(x^1,x^2,x^3)}. \quad (2.6)$$

The system (1.1) is obtained from Eqs. (2.3) and (2.4), if one eliminates Lagrangian coordinates ξ, using identity

$$\partial_i \frac{\partial J}{\partial \xi_{0,i}} \equiv 0. \quad (2.7)$$

The substitution of Eq. (2.3) into Eq. (2.7) leads to the continuity equation [the second equation in (1.1)]. Equation (2.4) coincides with the first equation in (1.1).

Let us note that the Lagrangian coordinates ξ play the role of a vector stream function in the sense that the combination (2.3) of the first-order derivatives of ξ is a solution of the continuity equation (2.7) for any choice of the quantities ξ.

Elimination of the Lagrangian coordinates leads to reducing the order of the system, what is connected with some arbitrariness of their choice. Indeed, the quantities ρ, $\mathbf{v}$ are invariant with respect to the transformation

$$\xi \to \xi' = \xi'(\xi), \quad \frac{\partial(\xi'_1,\xi'_2,\xi'_3)}{\partial(\xi_1,\xi_2,\xi_3)} = 1. \quad (2.8)$$

It means that the ξ are determined to within two arbitrary functions ξ'_1, ξ'_2. But the ξ'_3 is determined by the relation (2.8). Thus the transition from the LD to the ED is a transformation of variables $(t,\xi) \to (t,\mathbf{x})$ which is accompanied by a substitution of three variables ξ by one variable ρ, the ρ being a function of spatial derivatives of ξ.

The transition from the LD to DTWF is also a transformation of variables $(t,\xi) \to (t,\mathbf{x})$, but the Lagrangian coordinates are not eliminated. The wave function ψ contains information about the fluid particles' trajectories in the finite form. For instance, if the ψ is a solution of Eq. (1.5), then three independent relations

$$\frac{\psi(t,\mathbf{x})}{\sqrt{\psi^*\psi}} = \begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix} = \text{const}, \quad \chi_1^*\chi_1 + \chi_2^*\chi_2 = 1, \quad (2.9)$$

describe the fluid particle motion in the finite form. In this respect ψ is distinguished from the solution ρ, $\mathbf{v}$ of Eq. (1.1), which describes the fluid particle motion only by means of the differential equations (2.2). The DTWF arises after partial integration of the system (2.3), (2.4).

The invariance of the system (2.3), (2.4) with respect to the transformation group (2.8) enables us to reduce the order of the system (2.3), (2.4). The last can be represented in the form of a system of four first-order equations for four variables $\xi = \{\xi_i\}$, $i=0,1,2,3$,

$$g^k \partial_0 \xi_k = -\frac{1}{2}\frac{\partial J}{\partial \xi_{0,\alpha}}\frac{\partial J}{\partial \xi_{0,\alpha}}\left(\frac{\partial J}{\partial \xi_{0,0}}\right)^{-2} - U'(\rho),$$

$$g^k \partial_\alpha \xi_k = \frac{\partial J}{\partial \xi_{0,\alpha}} \Big/ \frac{\partial J}{\partial \xi_{0,0}}, \quad \alpha = 1,2,3, \quad (2.10)$$

where g^k $(k=0,1,2,3)$ are functions of the Lagrangian coordinates ξ and satisfy the conditions

$$\frac{\partial g^0}{\partial \xi_\beta} = \frac{\partial g^\beta}{\partial \xi_0}, \quad \beta = 1,2,3. \quad (2.11)$$

In particular one can set $g^0 = 1$, $g^\alpha = g^\alpha(\xi)$. The relations (2.10) are a result of integration of the system (2.4). One can verify this by means of substituting the relations (2.10) into identities with respect to ξ,

$$\frac{\partial J}{\partial \xi_{0,i}}\partial_\beta(g^s \xi_{s,i}) - \frac{\partial J}{\partial \xi_{0,i}}\partial_i(g^s \xi_{s,\beta})$$

$$\equiv \xi_{\alpha,\beta}\partial_l\left(\frac{\partial^2 J}{\partial \xi_{0,i}\partial \xi_{\alpha,l}} g^s \partial_i \xi_s\right) = 0, \quad \beta = 1,2,3, \quad (2.12)$$

which are valid for g^s $(s=0,1,2,3)$ satisfying Eq. (2.11). Substituting relations (2.10) into the left-hand side of Eq. (2.12) and using designations (2.3), one obtains Eqs. (2.4) as a result. Thus the system of equations (2.10) is equivalent to the system of equations (2.3), (2.4). Its order is reduced because arbitrariness of the Lagrangian coordinates ξ is taken into account by a choice of functions g^s, ξ_0. The wave function (1.12) corresponds to the choice of the function g^s, ξ_0 in the form

$$\xi_0 = a\varphi, \quad g^0 = 1, \quad g^\mu = (a/|\xi|^2)\epsilon_{\alpha\beta\mu} b_\alpha \xi_\beta, \quad \mu = 1,2,3,$$

$$b_\alpha = \text{const}, \quad b_\alpha b_\alpha = 1, \quad a = \text{const}. \quad (2.13)$$

Essentially, Eq. (1.5) for ψ is the system (2.10) represented in terms of (1.12), (2.13), and (2.6). Appearance of the second-order spatial derivatives in Eq. (1.5) is connected with relation (2.7) which is used for the calculation of $\partial\rho/\partial t$.

III. APPLICATION OF THE DESCRIPTION IN TERMS OF WAVE FUNCTION

The DTWF reduces the nonlinearity connected with the convective term $(\mathbf{v}\nabla)\mathbf{v}$, but it does not concern nonlinearity connected with the dynamical term $\mathbf{F}$ in Eq. (1.1). Let us manifest this in the example of the "quantum fluid" with the internal energy U of the form (1.8). In this special case, the dynamical term $\hat{L}_D$ and the first term of $\hat{L}_c$ compensate for each other, and Eq. (1.5) takes the form

$$ia\frac{\partial \psi}{\partial t} = \hat{L}_0\psi + \hat{L}'_c(\psi)\psi,$$

$$\hat{L}'_c(\psi) = -\frac{a^2}{8}\nabla s_\alpha \nabla s_\alpha + \frac{a^2}{4}\left[\frac{\nabla\rho}{\rho}\nabla s_\alpha \sigma_\alpha + \nabla^2 s_\alpha (s_\alpha - \sigma_\alpha)\right]. \quad (3.1)$$

Let us consider the solutions of Eq. (3.1) for which $|\hat{L}'_c(\psi)\psi| \ll |\hat{L}_0\psi|$. Let

$$\psi = \psi_0 + \epsilon^2\psi_2 + \epsilon^4\psi_4 + \cdots \tag{3.2}$$

$$ia\frac{\partial\psi_0}{\partial t} - \hat{L}_0\psi_0 = 0\,,$$

$$ia\frac{\partial\psi_2}{\partial t} - \hat{L}_0\psi_2 = \frac{1}{\epsilon^2}\hat{L}'_c(\psi_0)\psi_0\,, \tag{3.3}$$

$$ia\frac{\partial\psi_4}{\partial t} - \hat{L}_0\psi_4 = \frac{1}{\epsilon^4}\{\hat{L}'_c(\psi_0+\epsilon^2\psi_2)(\psi_0+\epsilon^2\psi_2) - \hat{L}'_c(\psi_0)\psi_0\}\,.$$

Such an expansion arises, if the zeroth approximation ψ_0 is taken in the form

$$\psi_0(t,\mathbf{x}) = \int_\Omega d\,\mathbf{k} f(\mathbf{k})\exp\left\{-\frac{ia\mathbf{k}^2}{2}t + i\mathbf{k}\mathbf{x}\right\},$$

$$f(\mathbf{k}) = \begin{pmatrix} f_1(\mathbf{k}) \\ f_2(\mathbf{k}) \end{pmatrix}, \tag{3.4}$$

where $f(\mathbf{k})$ is a column of two complex functions of the wavenumber $\mathbf{k}$. The functions do not vanish only inside the volume Ω of the $\mathbf{k}$-space, the line size χ of the Ω being small as compared with the mean wavenumber $\mathbf{K}$ inside Ω, i.e.,

$$\mathbf{K} = \int_\Omega \mathbf{k}\,d\,\mathbf{k}, \quad \chi = \max\{|\mathbf{k}-\mathbf{k}'|\} \quad \text{for } \mathbf{k}\in\Omega, \quad \mathbf{k}'\in\Omega\,,$$

$$K = |\mathbf{K}|, \quad \chi = \epsilon K, \quad \epsilon \ll 1\,. \tag{3.5}$$

The following estimations can be obtained from the definitions (1.7) and relations (3.4), (3.5),

$$|\nabla s_\alpha| \simeq \epsilon K, \quad |\nabla\rho| \simeq \epsilon K|\rho|, \quad |\nabla^2 s_\alpha| \simeq \epsilon^2 K^2\,,$$

$$|\hat{L}'_c(\psi_0)\psi_0| \simeq \epsilon^2 a^2 K^2|\psi_0| \simeq \epsilon^2|\hat{L}_0\psi_0|\,. \tag{3.6}$$

Thus if the solutions of Eqs. (3.3), $\psi_0,\psi_2,\psi_4,\ldots$, have the spectrum unvanishing only inside Ω, then relation (3.2) is an approximate solution of Eq. (3.1).

On the other side, if one searches for the solution of Eq. (1.1) associating with Eq. (3.1), then the expansion has the form

$$\mathbf{v} = \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \epsilon^2\mathbf{v}_2 + \cdots\,,$$

$$\rho = \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots\,, \tag{3.7}$$

ρ, $\mathbf{v}$ being connected with $\psi_0,\psi_2,\ldots$, by means of relations of the form

$$\rho_0 = \text{const}, \quad \mathbf{v}_0 = \text{const}\,, \quad \rho_0 + \epsilon\rho_1 = \psi_0^*\psi_0\,,$$

$$\mathbf{v}_0 + \epsilon\mathbf{v}_1 = -\frac{ia}{2\psi_0^*\psi_0}(\psi_0^*\nabla\psi_0 - \nabla\psi_0^*\psi_0)\,,$$

$$\rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \epsilon^3\rho_3 = (\psi_0^* + \epsilon^2\psi_2^*)(\psi_0 + \epsilon^2\psi_2),\ldots\,. \tag{3.8}$$

Thus the transition from ED to DTWF enables us to substitute the expansion in series over powers of ϵ by the expansion in series over powers of ϵ^2.

ACKNOWLEDGMENTS

I would like to thank Dr. M. S. Ruderman and Dr. I. M. Yavorskaya, who have looked through the manuscript and made useful remarks.

[1]E. Madelung, "Quantentheorie in hydrodynamischer form," Z. Phys. **40**, 322 (1926).

[2]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Moscow, 1974), p. 76 (in Russian).

Nonpropagating solitons of the variable coefficient and nonisospectral Korteweg–de Vries equation

W. L. Chan and Li Kam-Shun
Department of Mathematics, Science Centre, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

(Received 2 February 1989; accepted for publication 14 June 1989)

The method of inverse scattering is extended to solve the initial value problem of a variable coefficient and nonisospectral Korteweg–de Vries equation with time varying boundary condition. One- and two-soliton solutions are examined in detail. By an appropriate decomposition, soliton interactions and asymptotic behaviors are investigated. Oscillating and asymptotically standing two-soliton solutions are discussed.

I. INTRODUCTION

The purpose of this paper is twofold. We first describe some extensions of the inverse scattering in solving a nonisospectral and variable coefficient Korteweg–de Vries (KdV) equation with time varying nonvanishing boundary condition. Second, we obtain some nonpropagating soliton solutions and demonstrate their behavior. The experimental discoveries of them have only been recently reported.[1,2] Our aim is, therefore, to provide a mathematical model for such solitons. At present, theoretical results on them are sketchy. However, here we show clearly that the factors that contributed to their presence are (1) the coefficients of the evolution equation are time varying, (2) the scattering problem is nonisospectral, and (3) the time varying boundary condition is nonvanishing. The first author[3] has initiated the study of a KdV equation with (1) and (2) and constructed one-soliton solutions by Bäcklund transformation. Au-Yeung *et al.* considered a special case of (3), i.e., with constant nonvanishing boundary condition.[4] They argued forcefully the relevance of such condition but no nonpropagating soliton appeared in their study since (1) and (2) were not treated. Our emphasis is on two-soliton solutions and we adopt the approach of Moloney and Hodnett[5] in decomposing them into individual solitons in order to examine their interactions. Specifically, we consider the following variable coefficient and nonisospectral KdV equation (h-t-KdV equation)

$$u_t + K_0(u_{xxx} + 6uu_x) + 4K_1u_x - h(2u + xu_x) = 0 \tag{1.1}$$

under the time varying boundary conditions

$$u(x,t) \to L(t), \quad u_x(x,t) \to 0 \quad \text{as } x \to \mp\infty, \tag{1.2}$$

and

$$\int_{-\infty}^{\infty} (1 + |x|)|u - L|dx < \infty. \tag{1.3}$$

where (A) K_0, K_1, h, and L are smooth functions of t; (B) the Lax pair of (1.1), from Ref. 3, is

$$\psi_{xx} + u\psi = \lambda\psi, \tag{1.4}$$

$$\psi_t = K_0u_x\psi + [hx - 2K_0(u + 2\lambda) - 4K_1]\psi_x; \tag{1.5}$$

$$\text{(C)} \quad \lambda_t = 2h\lambda, \tag{1.6}$$

where λ is a function of t.

This paper is organized as follows. In Sec. II, the method of inverse scattering is extended to solve the initial value problem of (1.1). In particular, we establish a relation connecting the nonisospectral condition and the nonvanishing boundary condition. Examples of one-soliton solutions are given in Sec. III while two-soliton solutions are examined in detail in Sec. IV. The method of Ref. 5 is extended to investigate soliton interactions and their asymptotic behavior. In Sec. V oscillating two-soliton solutions and asymptotically standing two-soliton solutions are discussed. Qualitatively they reproduce experimental observations in Ref. 1.

II. INVERSE SCATTERING

We consider the initial value problem of Eq. (1.1) with the initial value condition

$$u(x,0) = u_0(x), \tag{2.1}$$

and

$$\int_{-\infty}^{\infty} (1 + |x|)|u_0 - L(0)|dx < \infty \tag{2.2}$$

under the conditions (1.2), (1.3), (1.6).

Theorem 1: A necessary condition for the above problem to be solvable is

$$L_t = 2hL. \tag{2.3}$$

Proof: Let $u(x,t)$ be the desired solution. Taking transform

$$u = u' + L(t) \tag{2.4}$$

and substituting it into (1.2)–(1.5), so that

$$u' \to 0, \quad u'_x \to 0, \quad \text{as } x \to \mp\infty, \tag{2.5}$$

$$u'(x,0) = u_0(x) - L(0), \tag{2.6}$$

$$\int_{-\infty}^{\infty} (1 + |x|)|u'|dx < \infty, \quad \text{for all } t, \tag{2.7}$$

and (1.4) and (1.5) become

$$-\psi_{xx} + (-u')\psi = k^2\psi, \quad k^2 = L(t) - \lambda, \tag{2.8}$$

$$\psi_t = K_0u'_x\psi + [hx - 2K_0(u' - 2k^2 + 3L(t)) - 4k_1]\psi_x. \tag{2.9}$$

 0022-2488/89/112521-06$02.50

From Ref. 6, it is clear that the system (2.8) and (2.9) has a solution

$$f(x,k,t) = c_1(t)\psi_1(x,k,t) + c_2(t)\psi_2(x,k,t). \tag{2.10}$$

where

$$\psi_1(x,k,t) \to \exp(-ikx),$$

$$\psi_2(x,k,t) \to \exp(ikx) \quad \text{as } x \to \infty$$

and $c_1^2(t) + c_2^2(t) \neq 0$. Hence

$$f(x,k,t) \to c_1(t)\exp(-ikx) + c_2(t)\exp(ikx) \quad \text{as } x \to \infty. \tag{2.11}$$

Substituting it into (2.9), and as $x \to \infty$,

$$\begin{aligned} &c_1'(t)\exp(-ikx) - ik_t c_1(t)x\exp(-ikx) + c_2'(t)\exp(ikx) + ik_t c_2(t)x\exp(ikx) \\ &\quad = 4ik\left[-K_0k^2 + \tfrac{3}{2}K_0L + K_1\right]c_1(t)\exp(-ikx) - ihkc_1(t)x\exp(-ikx) \\ &\quad\quad - 4ik\left[-K_0k^2 + \tfrac{3}{2}k_0L + K_1\right]c_2(t)\exp(ikx) + ihkc_2(t)x\exp(ikx). \end{aligned} \tag{2.12}$$

Comparing the coefficients of the terms $x\exp(-ikx)$ on both sides of (2.12), we have

$$k_t = hk. \tag{2.13}$$

Hence $2kk_t = 2h\,k^2$, $(k^2)_t = 2hk^2$. By (2.8) and (1.6), we have (2.3). The proof is completed.

Remark: From this theorem, if $K_0 = 1$, $K_1 = 0$, $h = 0$, then $L(t)$ can only be a constant and Eq. (1.1) becomes the standard KdV equation. This is the case studied by Au-Yeung *et al.*

In the proof of Theorem 1, it is shown that the eigenvalues k^2 of the Schrödinger equation should satisfy

$$\begin{aligned} k^2(t) &= L(t) - \lambda(t) \\ &= [L(0) - \lambda(0)]\exp\left(-2\int_0^t h\,dt\right) \\ &= k^2(0)\exp\left(-2\int_0^t h\,dt\right). \end{aligned}$$

Theorem 2: Let $\{a(k,t), r(k,t), k_l, b_l(t)\}$ be the scattering data with respect to $-u' = -u(x,t) + L(t)$, where u is a solution of (1.1) under the conditions (1.2), (1.3), (1.6), (2.1), (2.2), and (2.3). Then the following equalities are valid.

The transmission coefficient

$$a(k,t) = A\left[k\exp\left(-\int_0^t h\,dt\right)\right], \tag{2.14}$$

the reflection coefficient

$$r(k,t) = r(0)\exp\left[-8i\int_0^t k(-K_0k^2 + \tfrac{3}{2}K_0L + K_1)dt\right]. \tag{2.15}$$

$$A\left[k_l\exp\left(-\int_0^t h\,dt\right)\right] = 0, \quad k_l = i\eta_l, \quad l = 1,2,3,\ldots,N, \tag{2.16}$$

and

$$b_l(t) = b_l\exp[f_l(t)], \tag{2.17}$$

where $k = \sqrt{L-\lambda}$, $r(0) = r[k(0),0]$; $A(s)$ is a function of s; $b_l = r(0)\,A\,[k_l(0)]$ and

$$\begin{aligned} f_l(t) &= -8i\int_0^t k_l\left[-K_0k_l^2 + \frac{3}{2}K_0L + K_1\right]dt \\ &= 8\int_0^t \eta_l\left[K_0\eta_l^2 + \frac{3}{2}K_0L + K_1\right]dt. \end{aligned} \tag{2.18}$$

Proof: First we recall that the scattering data with respect to $-u'$ is derived from the Schrödinger equation (2.8). Following the proof of Theorem 1 and using (2.13), (2.12) becomes

$$\begin{aligned} &c_1'(t)\exp(-ikx) + c_2'(t)\exp(ikx) \\ &\quad = 4ik\left[-K_0k^2 + \tfrac{3}{2}K_0L + K_1\right]c_1(t)\exp(-ikx) \\ &\quad\quad - 4ik\left[-K_0k^2 + \tfrac{3}{2}K_0L + K_1\right]c_2(t)\exp(ikx). \end{aligned} \tag{2.19}$$

Comparing the coefficients of terms $\exp(-ikx)$, $\exp(ikx)$, respectively, on both sides of (2.19), we obtain the system

$$c_1'(t) = 4ik\left[-K_0k^2 + \tfrac{3}{2}K_0L + K_1\right]c_1(t), \tag{2.20}$$

$$c_2'(t) = -4ik\left[-K_0k^2 + \tfrac{3}{2}K_0L + K_1\right]c_2(t). \tag{2.21}$$

Solving for $c_1(t)$ and $c_2(t)$, we deduce

$$c_1(t) = c_1(0)\exp\left\{4i\int_0^t k\left[-K_0k^2 + \frac{3}{2}K_0L + K_1\right]dt\right\}, \tag{2.22}$$

$$c_2(t) = c_2(0)\exp\left\{-4i\int_0^t k\left[-K_0k^2 + \frac{3}{2}K_0L + K_1\right]dt\right\}. \tag{2.23}$$

In particular, taking

$$f(x,k,t) = \phi(x,k,t)\exp\left\{4i\int_0^t k\left[-K_0k^2 + \frac{3}{2}K_0L + K_1\right]dt\right\}, \tag{2.24}$$

where $\phi(x,k,t) \to \exp(-ikx)$, as $x \to -\infty$. From Ref. 6, we have

$$f(x,k,t) = a(k,t)\psi_1(x,k,t)W(t) + b(k,t)\psi_2(x,k,t)W(t),$$

where

$$W(t) = \exp\left\{4i\int_0^t k\left[-K_0k^2 + \frac{3}{2}K_0L + K_1\right]dt\right\}.$$

In this case, $c_1(t)$ and $c_2(t)$ of (2.22) and (2.23) become

$$c_1(t) = a[k(t),t]W(t), \quad c_2(t) = b[k(t),t]W(t)$$

and

$$c_1(0) = a[k(0),0], \quad c_2(0) = b[k(0),0].$$

Hence we obtain

$$a(k,t) = a[k(0),0], \tag{2.25}$$

$$b(k,t) = b\,[k(0),0]\,[W(t)]^{-2}. \tag{2.26}$$

By (2.13),

$$a[k(0),0] = a\left[k\exp\left(-\int_0^t h\,dt\right),0\right],$$

and setting

$$A\left[k\exp\left(-\int_0^t h\,dt\right)\right] = a\left[k\exp\left(-\int_0^t h\,dt\right),0\right],$$

then (2.14) and (2.15) are thus true, since $r(k,t) = b(k,t)/a(k,t)$.

It is clear[6] that

$$A\left[k\exp\left(-\int_0^t h\,dt\right)\right] = 0$$

has N pure imaginary roots $k_l = i\eta_l (l = 1,2,...,N)$, i.e., the Schrödinger equation (2.8) has N negative eigenvalues $k_l^2 = -\eta_l^2$. This is (2.16). Hence

$$b_l(t) = b(k_l,t) = r(k_l,t)A\left[k_l\exp\left(-\int_0^t h\,dt\right)\right]$$
$$= r[k_l(0),0]A\,[k_l(0)]$$
$$\times\exp\left\{-8i\int_0^t k_l\left[-K_0k_l^2 + \frac{3}{2}K_0L + K_1\right]dt\right\}.$$

Setting $b_l = r[k_l(0),0]\,A\,[k_l(0)]$ and

$$f_l(t) = -8i\int_0^t k_l\left[-K_0k_l^2 + \frac{3}{2}K_0L + K_1\right]dt$$
$$= 8\int_0^t \eta_l\left[K_0\eta_l^2 + \frac{3}{2}K_0L + K_1\right]dt, \tag{2.27}$$

we have (2.17). The proof is completed.

If $K_0 = 1$, $K_1 = 0$ and $h = 0$, then this theorem reduces to the case of the KdV equation.

Now according to Ref. 6, we can recover $-u'$ from the scattering data $\{a(k,t), r(k,t), k_l, b_l(t)\}$ by solving the Gelfand–Levitan integral equation

$$K(x,y;t) + F(x,y;t) + \int_x^\infty F(z+y;t)K(x,z;t)dz = 0, \tag{2.28}$$

where

$$F(\zeta;t) = \left(\frac{1}{2\pi}\right)\int_{-\infty}^{\infty} r(k,t)\exp(ik\zeta)dk$$
$$+\sum_{l=1}^{N}\left\{b_l\exp\left[f_l(t) + \int_0^t h\,dt - \eta_l\zeta\right]\right\}$$
$$\times(\{iA'[i\eta(0)]\})^{-1} \tag{2.29}$$

to get $K(x,y;t)$. Then we compute

$$u'(x,t) = 2\frac{\partial}{\partial x}K(x,x;t) \tag{2.30}$$

and by (2.4) recover $u(x,t)$, which is a solution of (1.1),

$$u(x,t) = L(t) + 2\frac{\partial}{\partial x}K(x,x;t). \tag{2.31}$$

III. ONE-SOLITON SOLUTIONS

When $r(0) = 0$, i.e., $r(k,t) = 0$, the solution (2.31) is called the N-soliton solution to Eq. (1.1) under the conditions (1.2), (1.3), (1.6), (2.1), (2.2), and (2.3). For simplicity, we only consider the cases $N = 1,2$ below.

Let

$$K(x,y;t) = P(x,t)\exp\left(-\eta_1 y + \int_0^t h\,dt\right). \tag{3.1}$$

Substituting it into (2.28) and (2.29), and by direct computation, we obtain

$$K(x,y;t) = -\{2\eta_1 b_1\exp[f_1(t) - \eta_1(x+y)]\}$$
$$\times(\{2i\eta_1(0)A'[i\eta_1(0)]$$
$$+ b_1\exp[f_1(t) - 2\eta_1 x]\})^{-1}. \tag{3.2}$$

From (2.31), we have the one-soliton solution

$$u(x,t) = L(t) + 2\eta_1^2\,\mathrm{sech}^2\,\theta_1, \tag{3.2}$$

where

$$\theta_1 = \alpha_1 + f_1(t)/2 - \eta_1 x, \tag{3.3}$$
$$\alpha_1 = \tfrac{1}{2}\ln\{b_1/2i\eta_1(0)A'[i\eta_1(0)]\}, \tag{3.4}$$
$$L(t) = L(0)\exp\left(2\int_0^t h\,dt\right), \tag{3.5}$$
$$\eta_1 = \eta_1(0)\exp\left(\int_0^t h\,dt\right). \tag{3.6}$$

From (3.2), we can show some of the properties of the soliton $u(x,t)$. Thus, for a given t, the value of x where the peak of the soliton appearing satisfies the following condition:

$$\theta_1 = \alpha_1 + f_1(t)/2 - \eta_1 x = 0. \tag{3.7}$$

Solving for x, we obtain

$$x = f_1(t)/(2\eta_1) + \alpha_1/\eta_1 \tag{3.8}$$

and the traveling velocity of the soliton is

$$V_1 = \frac{dx}{dt} = \frac{[f_1'(t)\eta_1 - \eta_1' f_1(t)]}{(2\eta_1^2)} - \frac{(\alpha_1\eta_1')}{\eta_1^2} \tag{3.9}$$

where $f_1(t)$ is defined by (2.18).

Because K_0, K_1, and h are functions of t, the properties of the soliton are more complex than the one for the KdV equation. They are illustrated by the following examples.

Example 1: Suppose $K_0 = 1$, $K_1 = 0$, $L(0) = \frac{2}{3}$, $\alpha_1 = 0$, $\eta_1(0) = 1$, and $h = 2t/[3(1+t^2)(2+t^2)]$, we obtain

$$u(x,t) = \tfrac{2}{3}[(2+2t^2)/(2+t^2)]^{2/3}$$
$$+ 2[(2+2t^2)/(2+t^2)]^{2/3}\,\mathrm{sech}^2\,\theta_1,$$

where

$$\theta_1 = 8[2t - \sqrt{2}\arctan(t/\sqrt{2})]$$
$$- x[(2+2t^2)/(2+t^2)]^{1/3}.$$

For fixed t, the peak appears at

$$x = 8[(2+t^2)/(2+2t^2)]^{1/3}$$
$$\times[2t - \sqrt{2}\arctan(t/\sqrt{2})]$$

with the traveling velocity of the soliton

$$V = \tfrac{32}{3}[(2+t^2)/(2+2t^2)]^{1/3}$$
$$\times[4t^2 + 3t^4 + 3 + \sqrt{2}t\arctan(t/\sqrt{2})]$$
$$\times\{[(2+t^2)(2+2t^2)]\}^{-1}.$$

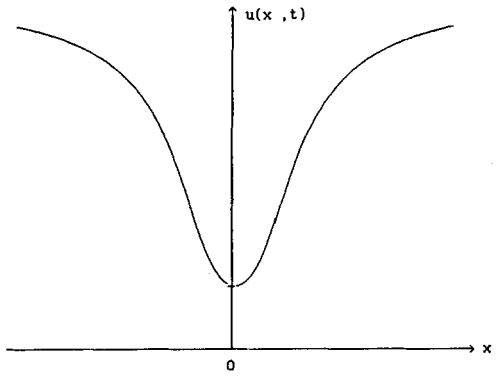


FIG. 1. $u(x,t)$ along the path $\theta_1 = 0$.

Figure 1 shows how the peak values of $u(x,t)$ depend on x.

Example 2: Suppose $K_0 = h = \cos t/[3(2+\sin t)]$ and the other conditions remain the same as Example 1. We obtain

$$u(x,t) = \tfrac{2}{3}[(2+\sin t)/2]^{(2/3)} + 2[(2+\sin t)/2]^{(2/3)}\operatorname{sech}^2\theta_1,$$

where

$$\theta_1 = \tfrac{4}{3}\sin t - x[(2+\sin t)/2]^{(1/3)}.$$

Here the soliton has the peak at

$$x = \tfrac{4}{3}[2/(2+\sin t)]^{1/3}\sin t,$$

for fixed t.

We can see that for each t, the value of x where the peak of the soliton appearing falls in the finite interval $[-2^{(7/3)}/3, 2(\frac{2}{3})^{(4/3)}]$ on the x axis. One of the reasons for this is that K_0 depends on t. Figure 2 shows how the peak values of $u(x,t)$ depend on x, along the path

$$x = \tfrac{4}{3}[2/(2+\sin t)]^{(1/3)}\sin t.$$

The soliton $u(x,t)$ only oscillates periodically in the above interval but never outside of it.

IV. TWO-SOLITON SOLUTIONS

Let $K(x,y;t)$ be

$$K(x,y;t) = \sum_{m=1}^{2} W_m(x,t)\exp\left(-\eta_m y + \int_0^t h\,dt\right).$$

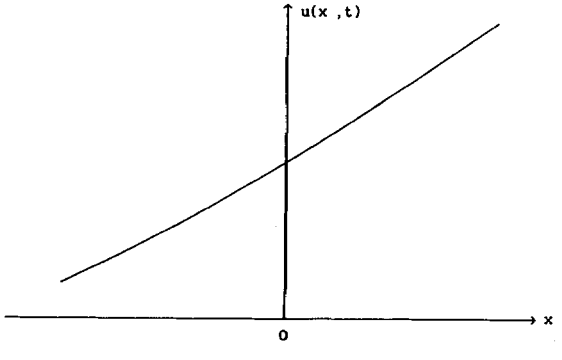


FIG. 2. $u(x,t)$ along the path $\theta_1 = 0$.

and substituting it into (2.28) and (2.29), then solving for W_m, we obtain

$$K(x,x;t) = -f_x/f, \tag{4.1}$$

where

$$f = 1 + \exp(2\theta_1) + \exp(2\theta_2) + A\exp(2\theta_1 + 2\theta_2), \tag{4.2}$$

$$\theta_m = \alpha_m + f_m(t)/2 - \eta_m(t)x, \tag{4.3}$$

$$\alpha_m = (1/2)\ln\{b_m/[iA'(i\eta_m(0))\eta_m(0)]\}, \quad m = 1,2, \tag{4.4}$$

$$A = \{[\eta_1(0) - \eta_2(0)]/[\eta_1(0) + \eta_2(0)]\}^2. \tag{4.5}$$

Hence,

$$\frac{\partial}{\partial x}K(x,x;t) = (f_{xx}f - f_x^2)/f^2$$

and

$$u(x,t) = L(t) + 2(f_{xx}f - f_x^2)/f^2. \tag{4.6}$$

In the spirit of Ref. 5, we obtain the decomposition of the solution $u(x,t)$, i.e.,

$$u(x,t) = L(t) + u_1 + u_2, \tag{4.7}$$

where

$$u_1 = 2\eta_1^2(t)H_1(\theta_2)\operatorname{sech}^2[\theta_1 + G(\theta_2)], \tag{4.8}$$

$$u_2 = 2\eta_2^2(t)H_2(\theta_1)\operatorname{sech}^2[\theta_2 + G(\theta_1)], \tag{4.9}$$

$$H_1(\theta_2) = [1 + A^{(1/2)}\exp(2\theta_2)]^2 \times(\{[1+\exp(2\theta_2)][1 + A\exp(2\theta_2)]\})^{-1}, \tag{4.10}$$

$$H_2(\theta_1) = [1 - A^{(1/2)}\exp(2\theta_1)]^2 \times(\{[1+\exp(2\theta_1)][1 + A\exp(2\theta_1)]\})^{-1}, \tag{4.11}$$

$$G(\theta_m) = \tfrac{1}{2}\ln\{[1 + A\exp(2\theta_m)]/[1+\exp(2\theta_m)]\}. \tag{4.12}$$

We call u_1 and u_2 soliton also.

Considering the asymptotic behavior of the second term of (4.7) for any fixed θ_1 (as $t \to \mp\infty$) and the asymptotic behavior of the third term of (4.7) for any fixed θ_2 (as $x \to \mp\infty$), we have the following theorem.

Theorem 3: Suppose that $h^- = \int_0^{-\infty} h\,dt$ and $h^+ = \int_0^{\infty} h\,dt$ are all finite but

$$\int_0^{-\infty}\left[K_0\exp\left(3\int_0^t h\,dt\right)\right]dt = p^-,$$

$$\int_0^{\infty}\left[K_0\exp\left(3\int_0^t h\,dt\right)\right]dt = p^+,$$

where p^- and p^+ are either finite or ∞ (or $-\infty$). Then, for any fixed θ_1, it holds that

$$u_1 \approx 2\eta_1^2(0)\exp(2h^-)H_1(\phi_2^-)\operatorname{sech}^2[\theta_1 + G(\phi_2^-)],$$

$$u_2 \approx 2\eta_2^2(0)\exp(2h^-)H_2(\theta_1)\operatorname{sech}^2[\phi_2^- + G(\theta_1)]$$

$$\text{as } t \to -\infty \tag{4.13}$$

and

$$u_1 \approx 2\eta_1^2(0)\exp(2h^+)H_1(\phi_2^+)\operatorname{sech}^2[\theta_1 + G(\phi_2^+)],$$

$$u_2 \approx 2\eta_2^2(0)\exp(2h^+)H_2(\theta_1)\operatorname{sech}^2[\phi_2^+ + G(\theta_1)]$$

$$\text{as } t \to \infty, \tag{4.14}$$

where

$$\phi_2^{\mp} = 4\eta_2(0)[\eta_2^2(0) - \eta_1^2(0)]p^{\mp} - [\eta_2(0)/\eta_1(0)](\alpha_1 - \theta_1) + \alpha_2. \quad (4.15)$$

Proof: First, in view of (2.13) and (2.16),

$$\eta_1(t) = \eta_1(0)\exp\left(\int_0^t h\,dt\right) \to \eta_1(0)\exp(h^-) \quad \text{as } t \to -\infty, \quad (4.16)$$

$$\eta_1(t) = \eta_1(0)\exp\left(\int_0^t h\,dt\right) \to \eta_1(0)\exp(h^+) \quad \text{as } t \to \infty. \quad (4.17)$$

Besides, from (4.3), for any fixed θ_1, we express θ_2 in terms of θ_1 and t to get

$$\theta_2 = \alpha_2 + f_2(t)/2 - [\eta_2(t)/\eta_1(t)] \times [\alpha_1 + f_1(t)/2 - \theta_1].$$

By (2.18), (2.16), and (2.13), it is rewritten as

$$\theta_2 = 4\eta_2(0)[\eta_2^2(0) - \eta_1^2(0)] \times \int_0^t \left[K_0 \exp\left(3\int_0^t h\,dt\right)\right]dt - [\eta_2(0)/\eta_1(0)](\alpha_1 - \theta_1) + \alpha_2. \quad (4.18)$$

Hence

$$\theta_2 \to 4\eta_2(0)[\eta_2^2(0) - \eta_1^2(0)]p^- - [\eta_2(0)/\eta_1(0)](\alpha_1 - \theta_1) + \alpha_2 = \phi_2^-, \quad \text{as } t \to -\infty,$$

$$\theta_2 \to 4\eta_2(0)[\eta_2^2(0) - \eta_1^2(0)]p^+ - [\eta_2(0)/\eta_1(0)](\alpha_1 - \theta_1) + \alpha_2 = \phi_2^+, \quad \text{as } t \to \infty.$$

Taking $t \to -\infty$ and $t \to \infty$ in Eq. (4.8) and (4.9), and using the results above, we have (4.13) and (4.14). Hence the proof is completed.

Under the conditions of Theorem 3, generally the effects of the interaction of u_1 and u_2 are, as t varies from $-\infty$ to ∞, (1) the amplitude of u_1 increases $\exp[2(h^+ - h^-)] \times H_1(\phi_2^-)/H_1(\phi_2^+)$ times; (2) the phase shift is $4\,\eta_2(0) \times [\eta_2^2(0) - \eta_1^2(0)](p^+ - p^-)$.

In particular, if $h^- = h^+, p^- = \infty, p^+ = -\infty$, then

$$u_1 \approx 2\eta_1^2(0)\exp(2h^-)\operatorname{sech}^2(\theta_1),$$
$$u_2 \approx 2\eta_2^2(0)\exp(2h^-)H_2(\theta_1), \quad \text{as } t \to -\infty,$$
$$u_1 \approx 2\eta_1^2(0)\exp(2h^-)\operatorname{sech}^2[\theta_1 + \tfrac{1}{2}\ln A\,],$$
$$u_2 \approx 0, \quad \text{as } t \to \infty.$$

This case is the same as the two-soliton solution of the KdV equation discussed in Ref. 5, i.e., the effect of interaction of u_1 and u_2 is that only u_1 has the phase shift $\frac{1}{2}\ln A$. Similar results are given for u_2.

Example 3: Suppose $K_0 = 2/(1+t^2)$, $K_1 = h = 0$, $L(0) = -\frac{1}{3}$, $\eta_1(0) = 2\,\eta_2(0) = 1$, $\alpha_1 = \alpha_2 = 0$. We obtain

$$u(x,t) = -(1/3) + u_1 + u_2, \quad (4.19)$$

$$u_1 = 2H_1(\theta_2)\operatorname{sech}^2[\theta_1 + G(\theta_2)], \quad (4.20)$$

$$u_2 = \tfrac{1}{2}H_2(\theta_1)\operatorname{sech}^2[\theta_2 + G(\theta_1)], \quad (4.21)$$

where

$$\theta_1 = 4\arctan t - x, \quad \theta_2 = -\arctan t - \tfrac{1}{2}x.$$

Here, $h^- = h^+ = 0$, $p^- = -\pi$, $p^+ = \pi$. Hence, for any fixed θ_1, we have

$$u_1 \approx 2H_1[\tfrac{1}{2}\theta_1 + \tfrac{3}{2}\pi]\operatorname{sech}^2\{\theta_1 + G[\tfrac{1}{2}\theta_1 + \tfrac{3}{2}\pi]\},$$
$$u_2 \approx \tfrac{1}{2}H_2(\theta_1)\operatorname{sech}^2[\tfrac{1}{2}\theta_1 + \tfrac{3}{2}\pi + G(\theta_1)], \quad \text{as } t \to -\infty,$$

and

$$u_1 \approx 2H_1[\tfrac{1}{2}\theta_1 - \tfrac{3}{2}\pi]\operatorname{sech}^2\{\theta_1 + G[\tfrac{1}{2}\theta_1 - \tfrac{3}{2}\pi]\},$$
$$u_2 \approx \tfrac{1}{2}H_2(\theta_1)\operatorname{sech}^2[\tfrac{1}{2}\theta_1 - \tfrac{3}{2}\pi + G(\theta_1)], \quad \text{as } t \to \infty.$$

Figure 3 shows the dependence of u_1 and u_2 on t, for $\theta_1 = 0$.

If the conditions of Theorem 3 are not satisfied, the status is more complex. Let us consider the following example.

Example 4: Suppose $K_0 = 2\cos t$ and the other conditions remain the same as in Example 3. Then we have (4.19), (4.20), and (4.21) but $\theta_1 = 4\sin t - x$ and $\theta_2 = -\sin t - \frac{1}{2}x$. Here, $h^- = h^+ = 0$, but $p^- = \lim_{t\to-\infty} 2\sin t$, $p^+ = \lim_{t\to\infty} 2\sin t$, i.e., p^- and p^+ do not exist. Thus the result of Theorem 3 is not valid. For any fixed θ_1, since θ_2 is a periodic function of t with period 2π, $H_1(\theta_2)$, $G(\theta_2)$, u_1, u_2, and u are so as well.

V. OSCILLATING AND ASYMPTOTIC STANDING SOLITONS

Following Ref. 5, we define the interaction point of the solutions u_1 and u_2 as the point of intersection of the paths along which u moves

$$\theta_1 = -\tfrac{1}{4}\ln A \quad (5.1)$$

and

$$\theta_2 = -\tfrac{1}{4}\ln A, \quad (5.2)$$

i.e.

$$\alpha_1 + \tfrac{1}{2}f_1(t) - \eta_1(t)x = -\tfrac{1}{4}\ln A, \quad (5.3)$$

$$\alpha_2 + \tfrac{1}{2}f_2(t) - \eta_2(t)x = -\tfrac{1}{4}\ln A. \quad (5.4)$$

Solving for x and t, we obtain the time t_d and the coordinate x_d corresponding to the interaction point satisfying

$$\int_0^{t_d}\left[K_0\exp\left(3\int_0^t h\,dt\right)\right]dt = \delta \quad (5.5)$$

and

$$x_d = 4\left[\eta_2^2(0) + \frac{3}{2}L(0)\right]\delta\exp\left(-\int_0^{t_d} h\,dt\right) + 4\exp\left(-\int_0^{t_d} h\,dt\right)\int_0^{t_d}\left[K_1\exp\left(\int_0^t h\,dt\right)\right]dt + \eta_2^{-1}(0)\left[\alpha_2 + \frac{1}{4}\ln A\right]\exp\left(-\int_0^{t_d} h\,dt\right), \quad (5.6)$$

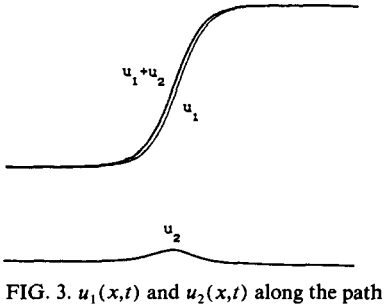


FIG. 3. $u_1(x,t)$ and $u_2(x,t)$ along the path $\theta_1 = 0$.

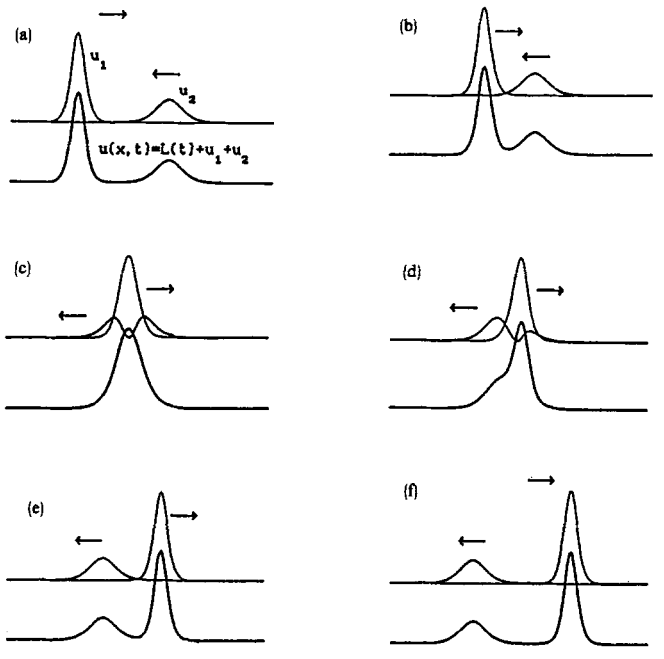


FIG. 4. Asymptotic standing two-soliton solution. (a) $t \leqslant -100$; (b) $-100 < t < t_d$; (c) $t = t_d$; (d) $t_d < t < 1$; (e) $1 < t < 100$; (f) $t \geqslant 100$.

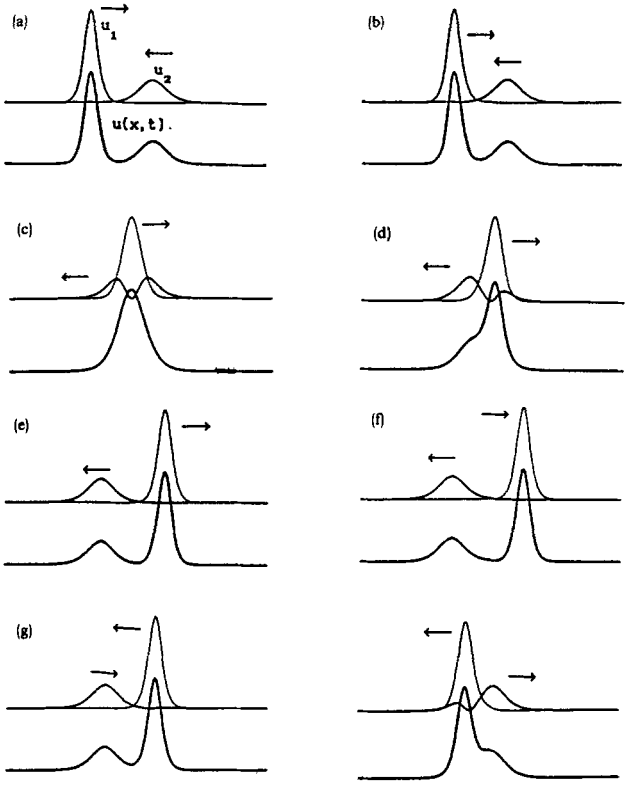


FIG. 5. Oscillating two-soliton solution. (a) $t = -\pi/2, 3\pi/2$; (b) $-\pi/2 < t < t_d$; (c) $t = t_d$; (d) t near t_d; (e) $t_d < t < \pi/2$; (f) $t = \pi/2$; (g) $\pi/2 < t < 3\pi/2$.

where

$$\delta = [\alpha_2\eta_1(0) - \alpha_1\eta_2(0)] \times(\{4\eta_1(0)\eta_2(0)[\eta_1^2(0) - \eta_2^2(0)]\})^{-1} + (\ln A/4)/\{4\eta_1(0)\eta_2(0)[\eta_1(0) + \eta_2(0)]\}. \quad (5.7)$$

The above results will now be illustrated by the following examples.

Example 5 (Asymptotically standing solitons): In Example 3, we have (4.19), (4.20), and (4.21). Here, by (5.5), (5.6), and (5.7), we find that $t_d = -0.0916$ and $x_d = -0.9155$.

Figure 4 is the graph of a two-soliton $u(x,t)$ traveling along the x-axis. It shows that (i) for $t \leqslant -100$, $u(x,t)$ is in the finite limiting position $x_{\min}$ (as $t \to -\infty$), the left wave is just the soliton $u_1(x,t)$ (except the term $-\frac{1}{3}$), the right wave is just the soliton $u_2(x,t)$; (ii) for $-100 < t < t_d$, $u(x,t)$ travels to the right, and so does $u_1(x,t)$ but $u_2(x,t)$ to the left; (iii) for t near t_d, $u(x,t)$ travels to the right but its form has considerably changed, because $u_1(x,t)$ (to right) and $u_2(x,t)$ (to left) interact at x_d. In this case, the amplitude of u_1 decreases and u_2 changes from a single-peak wave to a double-peak wave; (iv) for $t_d < t < 100$, $u(x,t)$ continues to the right and so does $u_1(x,t)$ but $u_2(x,t)$ to the left; (v) for $t \geqslant 100$, $u(x,t)$ is in the finite limiting position $x_{\max}$ (as $t \to \infty$). The left wave is just u_2, it is in the limiting position of u_2 (as $t \to \infty$). The right wave is just u_1, it is in the limiting position of u_1 (as $t \to \infty$). The two waves thus move apart and maintain a fixed separation indefinitely.

Example 6 (Oscillating solitons): In Example 4, we have (4.19), (4.20), and (4.21) also. Here, we find $t_d = -0.0918 + 2\pi m$, $m = 0, \mp 1, \mp 2, \ldots$, and $x_d = -0.9155$.

Figure 5 shows that in the interval $[-(1/2)\pi, (2/3)\pi]$, as t changes from $-(1/2)\pi$ to $(2/3)\pi$, $u(x,t)$ travels along the x axis from the finite limiting position $x_{\min}$ to the finite limiting position $x_{\max}$. At $t = -\pi/2$, the left wave is just u_1 which travels to the right. But the right wave is just u_2 which travels to the left. For t near t_d, at the point x_d, the form of $u(x,t)$ changes because u_1 and u_2 interact (similar to Example 5). At $t = \pi/2$, the left wave is u_2 but the right wave is u_1. As t changes from $\pi/2$ to $3\pi/2$, $u(x,t)$ travels back to $x_{\min}$ along the x axis until $t = 3\pi/2$; the positions of u_1 and u_2 are interchanged again. In other words, as t changes from $-\infty$ to ∞, $u(x,t)$, $u_1(x,t)$, and $u_2(x,t)$ oscillate in the finite interval $[x_{\min}, x_{\max}]$ on the x axis but never outside of it.

[1] J. Wu, R. Keolian, and I. Rudnick, Phys. Rev. Lett. **52**, 1421 (1984).
[2] R. Wei, B. Wang, Y. Mao, X. Zheng, and G. Miao, *Forced Standing Solitons and their Unique Propertties, Proceedings of the Third Asia Pacific Physics Conference*, edited by Y. W. Chan, A. F. Leung, C. N. Yang, and K. Young (World Scientific, Singapore, 1989), Vol. 2, p. 871.
[3] W. L. Chen and Y.-K Zheng, Lett. Math. Phys. **14**, 293 (1987).
[4] T. C. Au-Yeung, C. Au, and P. C. Fung, Phys. Rev. A **29**, 2703 (1984).
[5] T. P. Moloney and P. F. Nodnett, J. Phys. A: Math. Gen. **19**, L1129 (1986).
[6] S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Soliton: The Inverse Scattering Method* (Plenum, New York, 1984).

WKB theory of wave tunneling for Hermitian and nearly Hermitian vector systems of integral equations

H. J. Kull,[a] R. J. Kashuba,[b] and H. L. Berk
Institute for Fusion Studies, The University of Texas at Austin, Austin, Texas 78712

(Received 6 December 1988; accepted for publication 7 June 1989)

A general theory of wave tunneling in one dimension for Hermitian and nearly Hermitian vector systems of integral equations is presented. It describes mode conversion in terms of the general dielectric tensor of the medium and properly accounts for the forward and backward nature of the waves without regard to specific models. Energy conservation in the WKB approximation can be obtained for general Hermitian systems by the use of modified Furry rules that are similar to those used by Heading for second-order differential equations. Wave energy absorption can then be calculated perturbatively using the conservation properties of the dominant Hermitian operator. Operational graphical rules are developed to construct global wave solutions and to determine the direction of energy flow for spatially disconnected roots. In principle, these rules could be applied to systems with arbitrary mode complexity. Coupling coefficients for wave tunneling problems with up to four interacting modes are calculated explicitly.

I. INTRODUCTION

Wave tunneling is a well-known phenomenon which allows waves to penetrate into regions that would not be accessible along their ray trajectories. This effect has been particularly well studied for second-order differential equations in the quantum mechanical context.[1] More generally, tunneling can occur in systems described by higher-order differential equations and by integral equations. This is often the case in mode conversion problems in magnetized plasmas.[2] In this work, we wish to examine the general tunneling problem for Hermitian vector systems of integral equations in the Wentzel–Kramers–Brillouin (WKB) approximation. As a special case, this formulation includes Hermitian differential operators of arbitrarily high order. The theory can also be extended to nearly Hermitian systems including weak dissipation. The primary motivation for this study comes from the linear Vlasov–Maxwell theory, which leads to Hermitian operators when particle resonances are not present or neglected, e.g., for wave propagation perpendicular to the magnetic field.[3,4] As a consequence of Hermiticity, the total wave energy flow is exactly conserved.[5,6] It is therefore a basic goal of the present treatment to obtain consistency of wave tunneling rules with wave energy conservation.

In the WKB approximation, waves propagate along the inhomogeneity direction with a spatially slowly varying real wave number. Tunneling occurs across spatial regions where the wave number is complex. Across these regions, the usual WKB approximation fails and couplings occur among the various WKB waves. The mode conversion tunneling problem consists in the determination of the various coupling coefficients that arise between the ingoing and outgoing waves surrounding the tunneling region.

Previous work has been mostly based on specific model equations for tunneling regions. Mode conversion in thermal plasmas has been studied extensively using certain fourth-order differential equations with linearly varying coefficients.[2,7–9] Attempts to formulate unified descriptions of pairwise coupling have led to further model representations by second-order differential equations[10,11] and by systems of coupled first-order equations.[12–14] In the present treatment, we consider general linear systems and carefully discuss the aspect of wave energy conservation. For this purpose, we apply and extend WKB techniques developed for higher-order equations.[15–17] This approach is advantageous since the tunneling problem can be discussed in terms of the general dielectric tensor of the homogeneous medium, which is regarded as a known quantity. Furthermore, it allows one, in principle, to deal with arbitrary tunneling structures and mode complexity in slowly varying media.

The WKB formulation provides a general way of determining the energy flow direction of propagating waves. Using general properties of the dielectric tensor, we show how the energy flow direction of each wave can be obtained relative to the incident flow. It is found that the vector representation is particularly important for this problem. It determines the proper orientation of the energy flow along spatially disconnected branches by the sign change of certain cofactors of the dielectric tensor across the tunneling region. The identification of the direction of energy flow is essential to the physical interpretation of the theory, and we believe our method for identification is a new contribution to this field.

We also develop a graphical procedure based on modified Furry rules that enables one to describe the coupling between the waves in the mode conversion tunneling problem. The wave amplitudes are represented by isomorphic diagrams for the complex wave numbers, and a multiplication is defined that allows one to describe the propagation of waves by observing a set of basic operational rules.

[a] Permanent address: Institut für Angewandte Physik, Technische Hochschule, D-6100 Darmstadt, Federal Republic of Germany.
[b] Permanent address: McDonnel Douglas Corporation, Astronautics-Eastern Div., P. O. Box 516, St. Louis, MO 63166.

 0022-2488/89/112527-18$02.50

Energy conservation in the WKB theory of nondissipative systems is achieved by a modified Furry rule for the wave representation inside a tunneling region. This rule has been discussed by Heading for second-order differential equations.[18] Previously, phase integral methods have been applied for determining generalized Furry rules for higher-order equations.[15,16] Following these methods, the modified Furry rule can be derived from a bifurcation of the steepest descent path in a contour integral representation of the wave solution. This rule is shown to conserve wave energy for general tunneling structures. With weak dissipation, the bifurcation of the steepest descent path is still present, and the additional damping due to dissipation can be explicitly calculated. The general formulation of the mode coupling rules by graphical methods and its resulting energy conservation in nondissipative systems is the principal result of this work.

The paper is organized as follows. In Sec. II, the present WKB techniques are introduced. The wave diagrams and Furry rules are explained and subsequently used to demonstrate consistency with wave energy conservation and weak dissipation. In Sec. III, we use this formalism to analyze conventional tunneling through a potential barrier and make a comparison of this approximate method with an exact solution. In Sec. IV, we treat mode conversion between waves which tunnel with nonzero real parts of the wave numbers. Finally, in Sec. V, representative cases of the four wave tunneling problem are considered. In particular, a form is analyzed that describes mode conversion near ion-cyclotron frequencies.[5,9,19]

II. WKB METHOD

In this section, we define the general rules for constructing global solutions to the tunneling problem. These rules generalize the usual Furry rules and are expressed by wave diagrams that allow one to determine graphically the changes of wave amplitudes. By observing these rules, consistency with wave energy conservation is obtained in the WKB theory.

A. Local dispersion relation

We consider the propagation of waves with a frequency ω in a Vlasov–Maxwell system with spatial variations along the x direction. The electric field $\mathbf{E}(x)$ is governed by a vector system of integral equations of the general form,

$$\int dx' \, \mathbf{K}(x,x') \cdot \mathbf{E}(x') = 0. \tag{1}$$

The dependence on the constant frequency and a constant wave vector perpendicular to the inhomogeneity direction x is suppressed in the notation. First, we will neglect dissipation by assuming that the matrix elements of the kernel have the Hermiticity property, $K_{ij}(x,x') = K_{ji}^{*}(x',x)$. The effect of weak dissipation will be discussed separately in Sec. II G.

For systems with weak inhomogeneities, the WKB method determines asymptotic solutions of Eq. (1). The dielectric tensor of the medium can be introduced as the Fourier transform of the integral kernel with respect to the difference of its arguments,[15–17]

$$\mathbf{D}(x,k) = \int dq \, \mathbf{K}\left(x + \frac{q}{2}, x - \frac{q}{2}\right) \exp(-ikq). \tag{2}$$

The wave solutions are characterized by local wave numbers $k = k(x)$, which are the roots of the local dispersion relation,

$$\Lambda(x,k) = \det|D_{ij}(x,k)| = 0. \tag{3}$$

This is generally a transcendental equation and we assume that $\Lambda(z,k)$ is an entire function of the complex variables z and k. According to complex analysis, it can then be represented in the product form,[20]

$$\Lambda(z,k) = \exp\{g(k)\} k^{m} \prod_{n} \left(1 - \frac{k}{k_n}\right) \exp\{p_n(k)\}. \tag{4}$$

Here z is considered fixed, $g(k)$ represents an entire function, $p_n(k)$ are convergence producing polynomials, m is the multiplicity of the root $k = 0$, and n labels all zeros $k_n \neq 0$. The number of zeros can be zero, finite or infinite. In the following, we assume that all zeros are distinct in the asymptotic regions $|x| \to \infty$ and coalesce only pairwise elsewhere. From Eq. (4), the condition for pairwise coalescence is found to be

$$\Lambda(z,k) = \Lambda_{,k}(z,k) = 0, \quad \Lambda_{,k,k}(z,k) \neq 0, \tag{5}$$

where $\Lambda_{,k}$ denotes the partial derivative of Λ with respect to k.

If z varies in the complex plane, a branch $k = k_n(z)$ of the dispersion relation (3) becomes in general a multivalued function of z. To determine the analytic continuation of a given branch, it is convenient to consider a parametric representation $z(\tau)$, $k(\tau)$ of the zeros with a complex variable τ. In analogy with the real Hamiltonian ray equations, the functions $z(\tau)$ and $k(\tau)$ are determined by the system,

$$\frac{dz}{d\tau} = \Lambda_{,k}, \quad \frac{dk}{d\tau} = -\Lambda_{,z}. \tag{6}$$

Since $\Lambda(z,k)$ is assumed analytic, there exists a unique analytic solution to initial conditions $z(\tau_0) = z_0$, $k(\tau_0) = k_0$ at $\tau = \tau_0$. We now choose $\Lambda(z_0,k_0) = 0$ and a path in the τ plane such that $z(\tau)$ maps on a given path in the z plane. The branch $k = k(z)$ then is uniquely defined by the corresponding values of $k(\tau)$. The continuation of a given branch can therefore be reduced to the solution of an initial value problem. The WKB solutions in the complex plane depend on these trajectories and their behavior around branch points. Wave trajectories will be used in Sec. II C to represent and propagate wave solutions.

A branch point (z_B,k_B) of $k = k(z)$ occurs when $\Lambda_{,k} = 0$. Expanding about (z_B,k_B), the solution of Eq. (6) is obtained locally in the form,

$$\delta k = \delta s, \quad \delta z = -\tfrac{1}{2}(\Lambda_{,k,k}/\Lambda_{,z})\delta s^2, \tag{7}$$

with $\delta k = k - k_B$, $\delta z = z - z_B$, and $\delta s = -\Lambda_{,z}(\tau - \tau_B)$. If we propagate $\delta z \equiv |\delta z| \exp(i\psi)$ around the branch point, it follows from Eq. (7) that a phase change ψ of δz implies a phase change $\psi/2$ of δk. This rule is useful to gain a qualitative picture of the behavior of wave trajectories. Analogously, a branch point of the inversion function $z = z(k)$ occurs

when $\Lambda_{,z} = 0$ and here the phase change of ψ in δz implies a phase change of 2ψ in δk.

B. Local asymptotic solutions

The local asymptotic solutions of Eq. (1) assume the well-known form,

$$\mathbf{E}(z) = \mathbf{a}(z)\exp\{iS(z)\}. \tag{8}$$

The exponent $S(z)$ represents the rapidly varying function,

$$S(z) = S(z_0) + \int_{z_0}^{z} dz'\, k(z'). \tag{9}$$

For Hermitian operators, the solution for the components of the vector amplitude can be represented in the form,[15]

$$a_i(p) = (C_{if}/C_{ff})\, a_f(p),$$
$$a_f(p) = \{D_{,k}(p_0)/D_{,k}(p)\}^{1/2} a_f(p_0), \tag{10}$$

with the definitions,

$$\mathbf{D}\cdot\mathbf{C} = \Lambda\mathbf{I}, \quad D = \Lambda/C_{ff}, \quad p \equiv (z, k(z)). \tag{11}$$

Here $\mathbf{I}$ denotes the unit tensor, $\mathbf{C}$ the transposed matrix of the cofactors of $\mathbf{D}$, and, for simplicity, the dependence on z and $k(z)$ is expressed by a single variable p. For $\Lambda = 0$, one of the columns of $\mathbf{C}$, with an arbitrary but fixed index f, has been used to express the components of the solution vector in Eq. (10). It is then sufficient to describe the variation of the single scalar quantity a_f. We remark that the main difference between a scalar equation and a vector system lies in the appearance of the cofactor C_{ff} in the expression for D. For scalar equations, one can set $C_{ff} = 1$ and therefore $D = \Lambda$ in Eq. (10).

C. Diagrammatic representation

We now develop a graphical procedure for constructing WKB solutions of the mode coupling problem. For this purpose, a diagrammatic representation of WKB waves is introduced and basic operational rules are defined that allow one to determine mode coupling coefficients graphically.

We propagate the vector component $E_f(z)$, as defined by Eqs. (8)–(11), from p_1 to p_2 along a path in the complex plane that avoids branch points. The ratio $E_f(p_2)/E_f(p_1)$ can then be written in the form,

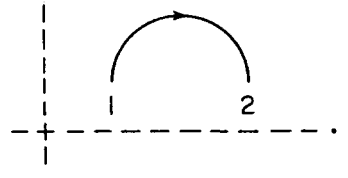
$$\langle k\,|1,2\rangle = a(k\,|1,2)e^{i\{-(1/2)\delta(k\,|1,2)\,+\,S(k\,|1,2)\}} \tag{12}$$

with

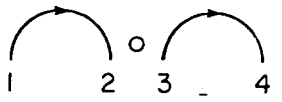
$$a(k\,|1,2) = \left|\left(\frac{D_{,k}(p_1)}{D_{,k}(p_2)}\right)^{1/2}\right|,$$

$$\delta(k\,|1,2) = \arg(D_{,k}(p))\Big|_{p_1}^{p_2}, \quad S(k\,|1,2) = S(p)\Big|_{p_1}^{p_2}.$$

This expression is uniquely determined by the path of $k = k(z)$ in the complex k plane and by the end points $p_{1,2}$. We therefore represent the WKB wave $\langle k\,|1,2\rangle$ by an isomorphic diagram,

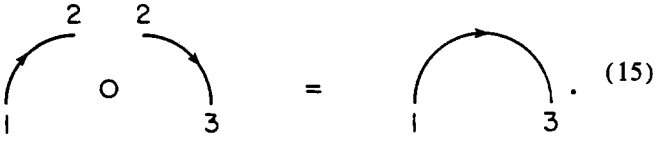
$\langle k\,|1,2\rangle \equiv$  (13)

This diagram indicates schematically the path in the k plane (solid line), the direction of propagation (arrow), and the end points (indices). The horizontal and vertical dashed lines represent the real and imaginary k axis, respectively. We now introduce a diagram product by defining,

$\langle k\,|1,2\rangle\langle k'|3,4\rangle \equiv$  (14)

For simplicity of notation, we can often suppress the coordinate lines, the end point indices, and the multiplication symbol $\circ$.

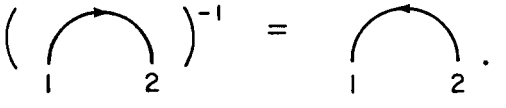
We now discuss some basic graphical operations for wave diagrams. If the wave amplitude (12) is propagated from point 1 to point 3 across an intermediate point 2, the following multiplication rule holds: $\langle k\,|1,2\rangle\langle k\,|2,3\rangle = \langle k\,|1,3\rangle$. This property is expressed graphically in the form,

 (15)

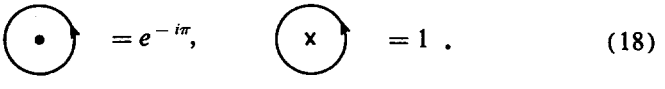
Replacing in Eq. (15) the end point 3 by the initial point 1 yields the identity,

 (16)

We can therefore define the inverse of a diagram by reversing its direction of propagation,

 (17)

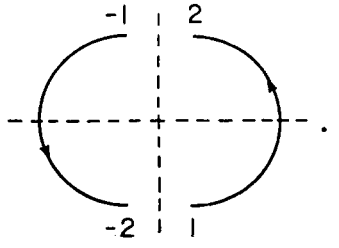
Propagation of waves around branch points is of particular importance for the mode coupling problem. Here the diagrams automatically keep track of the correct phases of the waves. We symbolize branch points, where $\Lambda_{,k} = 0$, by a dot. By propagating δk around such a branch point, we obtain from Eq. (12) that for a complete circle $\delta(k\,|1,2) = \pm 2\pi$. The plus sign corresponds to counterclockwise and the minus sign to clockwise rotation. A branch point, where $\Lambda_{,z} = 0$, is symbolized by a cross and in this case we obtain from Eq. (12) that the amplitude for a complete circle is analytic, therefore unity. Thus the phase rules for encircling branch points are expressed by the diagrams,

 (18)

To define further graphical rules, we will now assume that $D(z,k)$, defined in Eq. (11), obeys the common rela-

tions $D(z,-k)=D(z,k)$ and $D^*(z,k)=D(z^*,k^*)$. The first relation is satisfied, if the determinant Λ and the diagonal elements of the dielectric tensor are even functions of k. We are therefore dealing with media where waves with opposite signs of k propagate with opposite group velocities $v_{gr} \sim \Lambda_{,k}$. The second relation is a consequence of the reflection principle for analytic functions,[20] if $D(z,k)$ is assumed real for real arguments.

Let us now consider the product of two diagrams which are symmetric with respect to the origin,

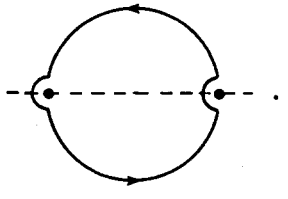


(19)

The two diagrams represent waves with wave numbers k and $-k$ and with end points: $1\equiv(k_1,z_1)$; $2\equiv(k_2,z_2)$; $-1\equiv(-k_1,z_1)$; $-2\equiv(-k_2,z_2)$. According to Eq. (12), they are related by $a(-k|-1,-2)=a(k|1,2)$, $\delta(-k|-1,-2)=\delta(k|1,2)$, $S(-k|-1,-2)=-S\times(k|1,2)$. The product diagram (19) therefore has the value

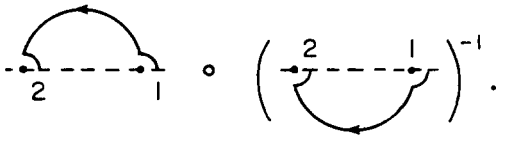
$$\langle k|1,2\rangle\langle -k|-1,-2\rangle = \left|\frac{D_{,k}(1)}{D_{,k}(2)}\right| e^{-i\delta(k|1,2)} = \frac{D_{,k}(1)}{D_{,k}(2)}. \tag{20}$$

Another diagram that occurs frequently is given by a closed loop around two branch points on the real axis,



(21)

Alternatively, using Eq. (17), this diagram can be represented by the product,

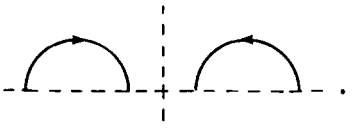


It corresponds to the ratio between two diagrams which are symmetric to the real axis. Reflection at the real axis is obtained by substituting $k\to k^*$ and $z\to z^*$. From Eq. (12), one finds the corresponding relations, $a(k^*|1,2)=a(k|1,2)$, $\delta(k^*|1,2)=-\delta(k|1,2)$, $S(k^*|1,2)=S^*(k|1,2)$. The loop diagram (21) then can be evaluated as,

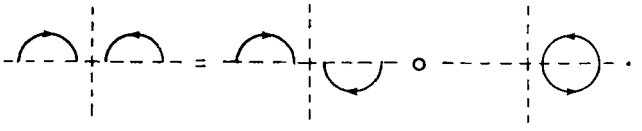
$$\frac{\langle k|1,2\rangle}{\langle k^*|1,2\rangle} = \sigma \exp i\oint k\,dz,$$
$$\sigma = e^{-i\delta(k|1,2)} = \operatorname{sgn}(D_{,k}(1)/D_{,k}(2)), \tag{22}$$

where $\operatorname{sgn}(x)=\pm 1$ denotes the sign of x, and the integration path is taken along the loop.

We finally discuss the product of two diagrams which are symmetric with respect to the imaginary axis,



Reflection at the imaginary axis $k\to -k^*$ combines the symmetries $k\to -k$ and $k\to k^*$, yielding $a(-k^*|-1,-2)=a(k|1,2)$, $\delta(-k^*|-1,-2)=-\delta(k|1,2)$, $S(-k^*|-1,-2)=-S^*(k|1,2)$. Accordingly, the waves $\langle k|1,2\rangle$ and $\langle -k^*|-1,-2\rangle$ represent complex conjugate solutions. Their product gives the squared magnitude of each diagram and can be expressed by a diagram of the form (19) times a loop diagram (21) by using Eq. (16),



(23)

We will find this representation particularly useful when evaluating energy expressions of wave solutions.

This completes the summary of the graphical identities that will be used in the subsequent sections.

D. Furry rules

In this section, phase integral methods are used to obtain an asymptotic representation of the solution inside wave tunneling regions. The result is expressed as a modified Furry rule that allows one to construct global tunneling solutions. We shall first present the rules and then justify them using a phase integral steepest descent method. The details of the derivations are presented in the Appendix.

The tunneling region is assumed to consist of two or more separated branch points x_B where the mode coupling condition $\Lambda(x_B,k_B)=\Lambda_{,k}(x_B,k_B)=0$ is satisfied. At each branch point two roots coalesce and the corresponding WKB solutions (8) fail. When passing around these points in the complex z plane, a change in the asymptotic representation can take place across Stokes' lines. These are defined by the condition,[15–18]

$$\operatorname{Im}\left(i\int_{x_B}^{z}(k_1-k_2)dz'\right)=0. \tag{24}$$

In the neighborhood of x_B, the Stokes lines form a star with three rays intersecting at x_B under an angle $2\pi/3$ (Fig. 1). On a Stokes line, the ratio between the two merging solutions acquires a purely real exponential factor. The exponentially large solution is called dominant and the exponentially small

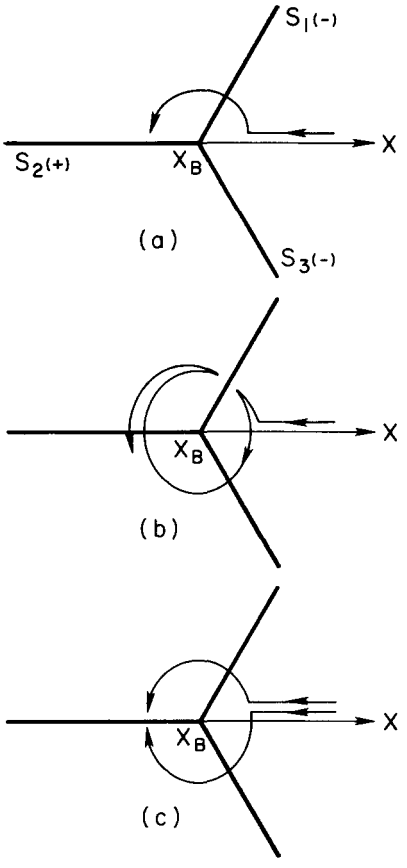
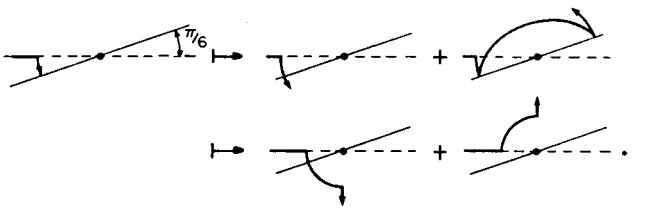


FIG. 1. Stokes lines emanating from a branch point x_B into the complex plane with one line along the real axis. The lines are labeled by the sign of $\mathrm{Im}(\delta k)$. (a) Dominant wave and path of wave propagation. (b) Subdominant wave induced by dominant wave. (c) Dominant and sudominant wave with path straightened.

solution subdominant. In the tunneling problems we consider, one of the Stokes lines is directed along the real axis. In general this requires, according to Eqs. (7) and (24), that x_B and k_B^2 are real. As shown in Fig. 1, we indicate the dominant wave on each Stokes line by the corresponding sign of $\mathrm{Im}(\delta k)$. On the real x axis, the dominant wave has $\mathrm{Im}(\delta k) > 0$ if the Stokes line is directed towards $-\infty$ and $\mathrm{Im}(\delta k) < 0$ if the Stokes line is directed towards $+\infty$.

The change of the asymptotic representation across a Stokes line is expressed by the following Furry rule: *If a dominant WKB wave is propagated across a Stokes line, one has to add a subdominant wave. The subdominant wave is the analytic continuation of the dominant wave along the path in the z plane that encircles the branch point in the opposite sense of the dominant wave.* In Fig. 1(a), we propagate the dominant wave with $\mathrm{Im}(\delta k) < 0$ in the z plane across S_1. The subdominant wave solution is obtained by propagating the wave underneath the branch point. In Fig. 1(c), we indicate the paths of propagation for the dominant and subdominant waves for $x < x_B$. The lower curve must be included in the asymptotic solution by noting that the subdominant wave is "induced" when the dominant solution (top curve) crosses S_1. More specifically, to the solution that is obtained along the path shown in Fig. 1(a), one adds an induced wave "created" at the dominant Stokes line as is shown in Fig. 1(b). When kinks of this curve are removed, the lower path of Fig. 1(c) is obtained.

This Furry rule can be readily represented by the wave diagrams in the complex k plane. According to Eq. (7), δk rotates around the branch point in the same sense as δz but only by half the angle. This gives rise to the connection formula,

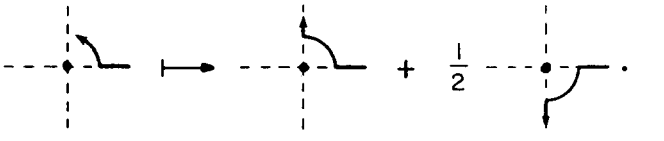


(25)

The lhs in Eq. (25) represents the wave just before crossing the Stokes line, which occurs in the complex k plane at an angle $\pi/6$ to the real axis. The middle grouping represents the dominant wave just after the crossing of the Stokes line and the subdominant wave that is obtained by backward propagation at the Stokes line. The final result represents the wave solution when the path returns to the x axis with $x < x_B$. In this last grouping kinks are straightened.

To deal with wave propagation along the Stokes line on the real axis, we now state a second modified Furry rule which has been discussed by Heading for second-order equations.[18] It is proved for higher-order equations and integral equations in the Appendix and the method of proof will only be outlined in the main text below.

We find that the modified rule can be expressed as follows: *If a wave is propagated on a Stokes line without actually crossing it, one addes to the wave one-half of the solution obtained along the subdominant path.* To illustrate this rule, suppose we propagate a wave with $\mathrm{Im}(\delta k) > 0$ on the path in Fig. 1(a) and then further along the Stokes line S_2. This wave, being dominant on S_2 without crossing it, induces one-half of the subdominant wave that corresponds to propagation on the lower curve in Fig. 1(c). In the k plane, we thus arrive at the diagram rule,



(26)

The derivation of the Stokes phenomenon described thus far can be obtained by use of phase integral methods discussed in Refs. 15 and 16. The starting point is a contour integral representation of the solution,

$$\mathbf{E}(z) = \frac{1}{\sqrt{2\pi i}} \int_C dk\, \hat{\mathbf{E}}(k) e^{ikz}, \tag{27}$$

where the contour C connects two remote regions in the k plane where the integrand vanishes. It is assumed that there exist as many independent contours as independent solu-

tions. For the vector field $\hat{\mathbf{E}}(k)$, the following asymptotic solutions have been derived in Ref. 15,

$$\hat{E}_i(k) = \left(\frac{C_{if}}{C_{ff}} \frac{1}{\sqrt{D_{,z}}}\right)\Bigg|_{z=z(k)} e^{-i\hat{S}(k)},$$

$$\hat{S}(k) = \hat{S}(k_0) + \int_{k_0}^{k} dk'\, z(k'). \tag{28}$$

This representation fails near the branch points of $z = z(k)$, but it is accurate near the branch points of $k = k(z)$ where the waves coalesce in x space. One can therefore use this representation to obtain asymptotic expansions near the mode coupling points in x space. According to the method of steepest descent, the contour C is deformed into a steepest descent path across saddle points. The saddle points follow from the condition,

$$\frac{d}{dk}\left(kz - \int_{k_0}^{k} dk'\, z(k')\right) = z - z(k) = 0, \tag{29}$$

and are coincident with the roots of the local dispersion relation at the point z. Performing the integration across the saddle at $k = k(z)$, the contribution to the contour integral is found to be

$$E_i(z) \sim \left(\frac{C_{if}}{C_{ff}} \frac{1}{\sqrt{D_{,k}}}\right)\Bigg|_{k=k(z)} e^{iS(z)}. \tag{30}$$

In this way it follows that for each saddle crossed by the contour C there is a corresponding WKB solution in x space. In the vicinity of a branch point of $k(z)$ the contour can pass across a second neighboring saddle giving rise to a subdominant solution. The method of evaluation is given in the Appendix and in this process the modified Furry rule is derived.

E. Wave energy flow

We now consider wave propagation in the asymptotic regions surrounding the tunneling structure. Generalized mode coupling coefficients are defined that express global energy conservation. Then the connection of the energy flow with respect to spatially disconnected roots is specified.

At the boundaries $|x| \to \infty$, the solution is taken as a superposition of WKB waves. We thereby assume that separate WKB solutions exist without further coupling. Specifically, there should be no mechanism for reflecting outgoing waves. In accordance with our symmetry assumptions in the local dispersion relation, the wave numbers occur in pairs of $k, -k$ and k, k^*. Outgoing wave conditions prescribe one solution of each $k, -k$ pair if k is real, and for a stable system boundedness determines one solution of the k, k^* pair if k is complex. The number of constraints from both boundaries then equals the number of independent wave solutions. In unstable systems our boundary conditions can be somewhat different. For example, a convectively unstable wave can amplify as it propagates. Then boundedness is not required and instead the amplifying wave is allowable at infinity. However, as long as the appropriate boundary conditions as to which waves can exist at infinity are determined, the wave scattering problem considered here is well defined. A discussion of how to determine the appropriate complex k solutions is well known from the theory of convective instability.[21]

For complex wave numbers, we allow only convergent solutions satisfying $x\,\mathrm{Im}(k) > 0$. In the Hermitian case, these convergent waves do not transport wave energy at infinity. For real wave numbers, we define the group velocity v_{gr}, the wave energy W, and the energy flow J by the expressions,

$$v_{gr} = \frac{\partial \omega}{\partial k} = -\frac{\Lambda_{,k}}{\Lambda_{,\omega}}, \quad W = \mathbf{E}^* \cdot \mathbf{D}_{,\omega} \cdot \mathbf{E},$$

$$\mathbf{J} = -\mathbf{E}^* \cdot \mathbf{D}_{,k} \cdot \mathbf{E}. \tag{31}$$

Asymptotically, J is conserved for each wave and obeys the familiar relation $J = v_{gr} W$, as follows also from Eqs. (10) and (11). Boundary conditions are assumed for incoming waves which have group velocities satisfying $x v_{gr} < 0$. The group velocity and the phase velocity $v_{ph} = \omega/k$ are related by

$$v_{gr} v_{ph} = -\frac{\omega}{k} \frac{\Lambda_{,k}}{\Lambda_{,\omega}} = -\frac{\Lambda_{,k^2}}{\Lambda_{,\omega^2}}, \tag{32}$$

where Λ is a function of k^2 and ω^2. The waves with opposite wave numbers always have opposite group velocities but can in general be forward waves, if the rhs is positive, or backward waves, if the rhs is negative.

Let us now assume that the region $x \to -\infty$ supports one incoming wave ($v_{gr} > 0$) and N outgoing waves ($v_{gr} < 0$) and the region $x \to \infty$ M outgoing waves ($v_{gr} > 0$). Global energy conservation requires that

$$J_i + \sum_{n=1}^{N} J_{rn} = \sum_{m=1}^{M} J_{tm}, \tag{33}$$

with

$$J_i = v_{gr,i} W_i, \quad J_{rn} = v_{gr,rn} W_{rn},$$

$$J_{tm} = v_{gr,tm} W_{tm},$$

where the indices i,r,t denote incident, reflected, and transmitted waves, respectively. Defining mode coupling coefficients $R_n = |J_{rn}/J_i|$ for the reflected waves and $T_m = |J_{tm}/J_i|$ for the transmitted waves, one can rewrite Eq. (33) in the form,

$$\sum_{n=1}^{N} \sigma_n R_n + \sum_{m=1}^{M} \sigma_m T_m = 1. \tag{34}$$

The sign of each term $\sigma_k = \mathrm{sgn}(W_k/W_i)$ is defined by the sign of the ratio of the wave energies W_k and W_i. Some of the σ_k can be negative, if negative energy waves occur. We also remark that the coupling coefficients R_n and T_m depend in general on both the wave amplitudes and the group velocities.

We now wish to specify the ingoing and outgoing waves among the various branches of the local dispersion relation on both sides of the tunneling region. For this purpose, we use a graphical procedure that determines the direction of energy propagation along each branch relative to the incident flow. Inserting the wave representation (10), (11) into the energy flow expression (31) and using the relation $\mathbf{D} \cdot \mathbf{C} = \Lambda \mathbf{I}$, one finds,

$$J = -\mathbf{E}^*\cdot\mathbf{D}_{,k}\cdot\mathbf{E} = -(\mathbf{E}^*\cdot\mathbf{D}\cdot\mathbf{E})_{,k}$$
$$= -\left(\frac{C_{jf}^* D_{ij} C_{jf}}{C_{ff}^* C_{ff}}|E_f|^2\right)_{,k} = -\left(\frac{\Lambda}{C_{ff}}|E_f|^2\right)_{,k}$$
$$= -D_{,k}|E_f|^2. \tag{35}$$

In Eq. (35), the energy flow is expressed by the amplitude of the single component E_f and by the function $D(x,k) = \Lambda(x,k)/C_{ff}(x,k)$. This function is well defined for real x and k and is zero along the branches $k = k(x)$.

In Fig. 2, we illustrate by two examples how the function $D(x,k)$ determines the relative direction of energy flow. Solid lines represent the contour lines $\Lambda(x,k) = 0$, where $k = k(x)$. The dashed line in Fig. 2(b) is a contour line $C_{ff} = 0$, where the cofactor changes sign. From these lines one can determine the sign of $D(x,k)$ in each region of the (x,k) plane relative to its sign in one particular region. For definiteness, we assume $D > 0$ in the shaded regions and $D < 0$ in the unshaded regions. According to Eq. (35), the propagation direction of wave energy is determined by the sign of the derivative of D with respect to k. From the topology of Fig. 2, it is then clear that waves propagate to the right on the branches that pass above a shaded and below an unshaded region. Likewise waves propagate to the left, if their branches lie below a shaded and above an unshaded region. For instance, in Fig. 2, k_c is a right-going and k_d a left-going wave.

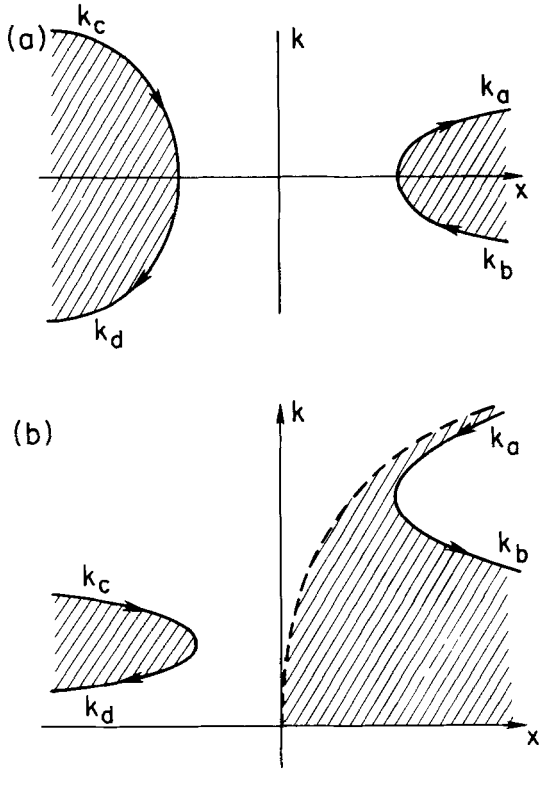


FIG. 2. Directions of energy flow. $D(x,k)$ is assumed positive in the shaded regions and changes sign across the solid contours where $\Lambda(x,k) = 0$ and across the dashed contour where $C_{ff}(x,k) = 0$. The direction of propagation on each branch is indicated by arrows. (a) Tunneling region with $C_{ff} \neq 0$. (b) Tunneling region with $C_{ff} = 0$.

The presence of a zero in the cofactor C_{ff} can allow for a change of propagation ordering across a tunneling region. This is illustrated in Fig. 2 by two examples leading to different orderings for the spatially disconnected branches. In Fig. 2(a), where no zero of C_{ff} is assumed, k_a and k_b correspond to waves propagating to the right and left, respectively. On the other hand, in Fig. 2(b), D reverses sign across the zero of C_{ff} and then k_a would be a left-going and k_b a right-going wave. Thus the understanding of the structure of the cofactor is crucial to interpreting the relative propagation direction in regions separated by a tunneling structure. A sign change of the cofactor is a particular feature of vector systems. For scalar wave equations ($C_{ff} = 1$), the cofactor is nonzero by definition and the propagation ordering would have to correspond to Fig. 2(a).

We finally wish to demonstrate the independence of our determination of propagation ordering from the particular coordinate representation that has been used. For this purpose, we show that the sign of a nonvanishing cofactor C_{ff} is independent of the index f on each real branch $k = k(x)$. Consequently, if the cofactor C_{ff} changes sign on opposite sides of the tunneling region, the same must hold for each cofactor C_{gg} corresponding to a different vector component g. For real x and k, the dielectric tensor is Hermitian, $D_{ij}(x,k) = D_{ji}^*(x,k)$, and there exist real eigenvalues λ_n and eigenvectors $\mathbf{e}_n$, determined locally by the eigenvalue problem,

$$\mathbf{D}\cdot\mathbf{e}_n = \lambda_n \mathbf{e}_n. \tag{36}$$

Along a nondegenerate branch $k = k(x)$, one of the eigenvalues has to vanish while the others are nonzero. Specifically, we set, $\lambda_1 = 0, \lambda_2 \neq 0, \lambda_3 \neq 0$, corresponding to a solution $\mathbf{E} = a\mathbf{e}_1$. Using this representation, the energy flow (31) assumes the form,

$$J = \lambda_{1,k}|a|^2. \tag{37}$$

Comparing this expression with Eq. (35) and using $\Lambda = \lambda_1\lambda_2\lambda_3$, we find,

$$C_{ff}(x,k(x)) = \lambda_2\lambda_3(|E_f|^2/|a|^2). \tag{38}$$

This representation holds on each branch, showing that C_{ff} does not change sign along the branch. Furthermore, the sign is independent of the particular component f that has been chosen to represent the wave. However, we note that C_{ff} can be zero, if a poor choice of representation is used such that $E_f = 0$.

F. Energy conservation of mode coupling rules

We now return attention to the Furry rules and discuss their consistency with energy conservation. For this purpose, we propagate a general solution around a real branch point x_B and compare the asymptotic energy flow on both its sides. Inside the tunneling region pairs of wave solutions E_f^+, E_f^- with complex conjugate wave numbers k^+, k^- give rise to an energy flow,[6]

$$J = -[D_{,k^+}(E_f^-)^* E_f^+ + \text{c.c.}], \tag{39}$$

where c.c. denotes the complex conjugate expression. Equation (39) generalizes the expression (35) for real wave

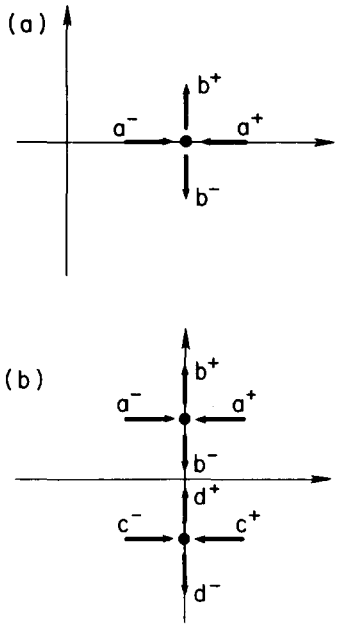


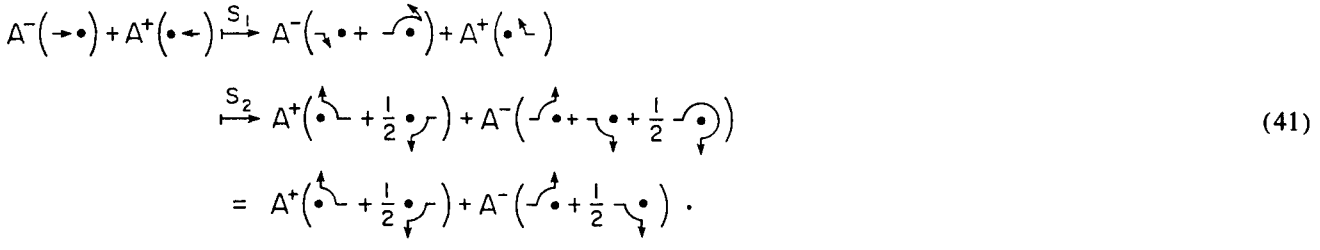
FIG. 3. Branch points (a) on the real and (b) on the imaginary axis in the k plane. The notation used for the merging roots is indicated.

numbers to the complex case. It should, however, be noticed that both asymptotic results apply in different regions. These regions are separated by the region around the branch point k_B (i.e., the point where k^+ and k^- coalesce). Since the energy flow (39) couples only complex conjugate branches, we consider first a single real branch point k_B, as in Fig. 3(a), and then a pair of complex conjugate branch points k_B and k_B^* on the imaginary axis, as in Fig. 3(b).

For real branch points k_B, two propagating waves undergo tunneling. We denote by $a^\pm$ the k values of the real branches with $\mathrm{Re}(\delta k) \lesseqgtr 0$ for $x > x_B$, and by $b^\pm$ the k values of the complex branches with $\mathrm{Im}(\delta k) \lesseqgtr 0$ for $x < x_B$, respectively. The corresponding wave solutions for the component E_f are analogously written as $A^\pm$ and $B^\pm$. Using Eqs. (35) and (39) the energy flows on both sides of x_B can be expressed in the form,

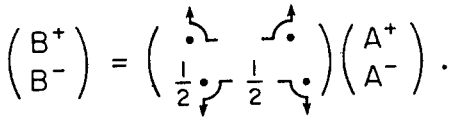
$$J(x > x_B) = -[D_{,a^+}|A^+|^2 + D_{,a^-}|A^-|^2],$$
$$J(x < x_B) = -[D_{,b^+}B^+(B^-)^* + \text{c.c.}]. \tag{40}$$

We now propagate the waves $A^\pm$ to the region $x < x_B$ on the path shown in Fig. 1(a) and obtain for the subsequent Stokes lines S_1 and S_2 the connection formulas,



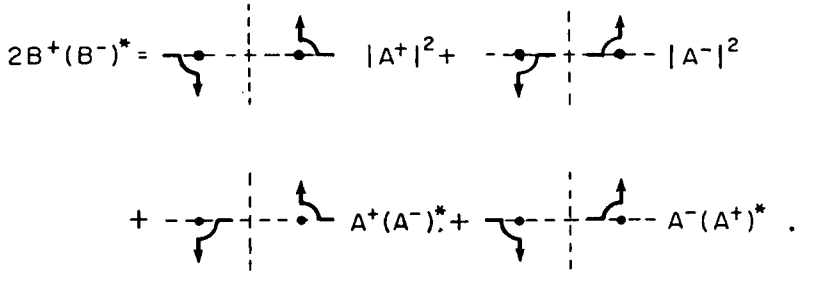
(41)

From Eq. (41), the solutions $B^\pm$ in the tunneling region are found to be



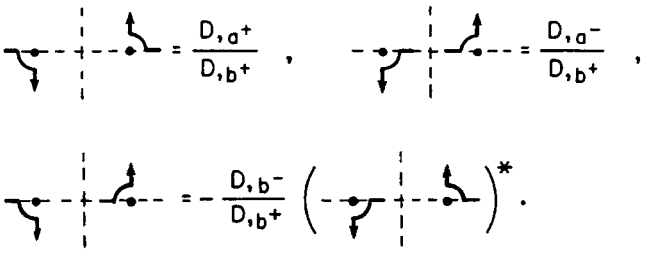
(42)

Noting that the complex conjugate diagrams are obtained by reflection about the imaginary axis [Eq. (23)], one finds with Eq. (42),



(43)

We now use Eq. (20) to obtain,



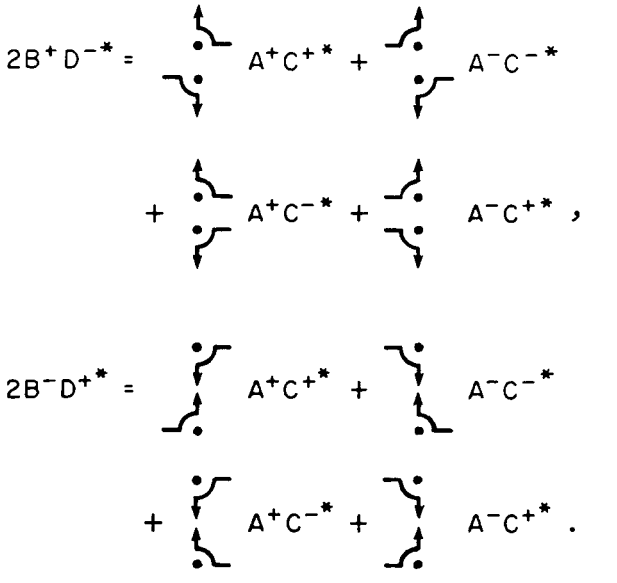
(44)

From Eqs. (40), (43), and (44) it follows that $J(x>x_B) = J(x<x_B)$. We have thus obtained energy conservation for branch points on the real axis.

Next, we consider a pair of complex conjugate branch points on the imaginary axis and choose an analogous notation [Fig. 3(b)]. In this case, the waves are tunneling on both sides of x_B and the corresponding energy flows are

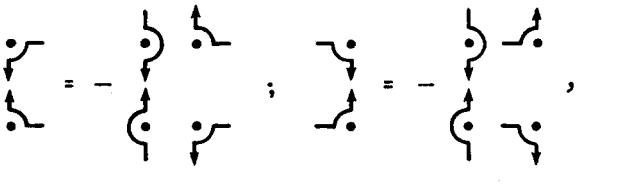
$$\begin{aligned} J(x>x_B) = & -[D_{,a^+}A^+(C^+)^* \\ & + D_{,a^-}A^-(C^-)^* + \text{c.c.}], \\ J(x<x_B) = & -[D_{,b^+}B^+(D^-)^* \\ & + D_{,b^-}B^-(D^+)^* + \text{c.c.}]. \end{aligned} \tag{45}$$

The waves $B^{\pm}$ are related to the waves $A^{\pm}$ by Eq. (42) with the branch point taken in the upper half-plane. Similarly, the waves $D^{\pm}$ are related to the waves $C^{\pm}$ by choosing the branch point in the lower half-plane. Evaluating the products in Eq. (45), one finds,



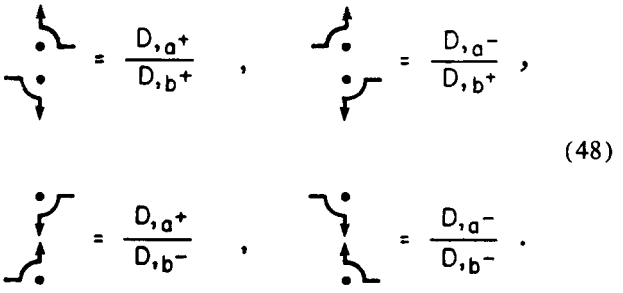
(46)

The diagrams in the last line can be written as



(47)

which leads to cancellation of the corresponding terms in the first product. The remaining diagrams can be combined to the expression for $J(x>x_B)$ by using the identities



(48)

These results demonstrate energy conservation of the modified Furry rules for any combination of branch points on the real and imaginary k axis and thereby generalize the discussion of Heading for branch points at the origin ($k_B = 0$).[18]

G. Dissipation

The present results for Hermitian systems can readily be generalized to nearly Hermitian systems with weak dissipation. If dissipation is sufficiently small, it affects only the slowly varying wave amplitudes without changing the mode coupling rules at branch points. This approximation will now be explained and thereby dissipative corrections to the coupling coefficients for tunneling regions are obtained.

For nearly Hermitian systems, the wave numbers can be calculated perturbatively by expanding about the wave numbers k^H of the associated unperturbed Hermitian system. Specifically, this expansion is written in the form,

$$\begin{aligned} & k = k^H + ik^A, \\ & \mathbf{D}(k) = \mathbf{D}^H(k^H) + ik^A\partial_k\mathbf{D}^H(k^H) + i\mathbf{D}^A(k^H), \\ & \mathbf{E} = \mathbf{E}_0 + i\delta\mathbf{E}, \end{aligned} \tag{49}$$

where H refers to the Hermitian part and A to the anti-Hermitian perturbation. Inserting Eq. (49) into the governing equation $\mathbf{D}\cdot\mathbf{E} = 0$, yields up to first order the relations,

$$\begin{aligned} & \mathbf{D}^H(k^H)\cdot\mathbf{E}_0 = 0, \\ & \mathbf{D}^H(k^H)\cdot\delta\mathbf{E} = -[k^A\partial_k\mathbf{D}^H(k^H)\cdot\mathbf{E}_0 \\ & \qquad + \mathbf{D}^A(k^H)\cdot\mathbf{E}_0]. \end{aligned} \tag{50}$$

If $\mathbf{E}_0 = \mathbf{E}_0(k^H)$ represents a solution to the Hermitian operator $\mathbf{D}^H(k^H)$, then $\mathbf{E}_0^+ = \mathbf{E}_0(k^{H*})$ will be a solution to the adjoint operator $D_{ij}^{H+}(k^H) = D_{ji}^{H*}(k^H) = D_{ij}^H(k^{H*})$. Multiplying the second equation in Eq. (50) by $(\mathbf{E}_0^+)^*$, the lhs vanishes and from the rhs there follows

$$k^A = -\frac{(\mathbf{E}_0^+)^*\cdot\mathbf{D}^A\cdot\mathbf{E}_0}{(\mathbf{E}_0^+)^*\cdot\partial_k\mathbf{D}^H\cdot\mathbf{E}_0}. \tag{51}$$

Using Eqs. (10) and (11) for the components of $\mathbf{E}_0$, one can write Eq. (51) alternatively in the form,

$$k^A(k) = -\frac{C_{if}^*(k^*)D_{ij}^A(k)C_{jf}(k)}{C_{ff}^*(k^*)\Lambda_{,k}}. \tag{52}$$

For each wave number of the unperturbed problem, the expression (52) represents a dissipative correction. For real k, it follows from the symmetry relations $D_{ij}^{A}(k) = D_{ji}^{A*}(k^*)$, $\Lambda(k) = \Lambda^*(k^*)$, $C_{jf}^*(k^*) = C_{jf}(k)$, that $k^A(k)$ is real:

$$k^{A*}(k^*) = -\frac{C_{if}(k)D_{ij}^{A*}(k^*)C_{jf}^*(k^*)}{C_{ff}(k)\Lambda_{,k}^*} = -\frac{C_{if}^*(k^*)D_{ji}^{A*}(k^*)C_{jf}(k)}{C_{ff}(k)\Lambda_{,k}^*} = -\frac{C_{if}^*(k^*)D_{ij}^{A}(k)C_{jf}(k)}{C_{ff}^*(k^*)\Lambda_{,k}} = k^A(k). \tag{53}$$

Accordingly, the propagating waves are damped by an additional slowly varying factor to the wave amplitudes,

$$\mathbf{E}(z) \sim \exp\left(-\int_{z_0}^{z} k^A dz'\right).$$

Near branch points the damping expression is integrable as follows from

$$\int_{z_0}^{z} k^A\, dz' = -\int_{z_0}^{z} \eta(z')\frac{dz'}{\Lambda_{,k}} = \int_{k(z_0)}^{k(z)} \eta(k')\frac{dk'}{\Lambda_{,z}}$$

with

$$\eta = -\frac{C_{if}^*(k^*)D_{ij}^{A}(k)C_{jf}(k)}{C_{ff}^*(k^*)}.$$

In the derivation of the mode coupling rules in the Appendix, the slowly varying amplitude is treated as a constant. To this approximation, dissipation will not alter these rules for mode coupling.

The dissipative corrections to the mode coupling coefficients can be obtained in the following way. Consider a tunneling region $x_1 < x < x_2$. In the regions of wave propagation, the energy flow of the incident, reflected and transmitted waves varies as

$$J_i(x) = J_i(x_1)\exp\left(-2\int_{x_1}^{x} k_i^A\, dx\right),$$

$$J_{rn}(x) = J_{rn}(x_1)\exp\left(-2\int_{x_1}^{x} k_{rn}^A\, dx\right),$$

$$J_{tm}(x') = J_{tm}(x_2)\exp\left(-2\int_{x_2}^{x'} k_{tm}^A\, dx\right),$$

respectively, with $k_i^A > 0$, $k_{rn}^A < 0$, and $k_{tm}^A > 0$. Analogously to the definitions (34) for Hermitian systems, we define mode coupling coefficients by

$$R_n = \left|\frac{J_{rn}(x_1)}{J_i(x_1)}\right| = \left|\frac{J_{rn}(x)}{J_i(x)}\right|\exp\left(2\int_{x}^{x_1}(k_i^A - k_{rn}^A)dx\right),$$

$$T_m = \left|\frac{J_{tm}(x_2)}{J_i(x_1)}\right| = \left|\frac{J_{tm}(x')}{J_i(x)}\right|\exp\left(2\int_{x'}^{x_1} k_i^A\, dx + 2\int_{x_2}^{x'} k_{tm}^A\, dx\right). \tag{54}$$

These definitions compensate for the damping of the propagating waves and the coupling coefficients are therefore independent of the observation points $x < x_1$ and $x' > x_2$ of the reflected and transmitted waves, respectively.

The connection formulas between the asymptotic wave representations on both sides of a tunneling region depend on certain loop integrals q. These loop integrals will be discussed for Hermitian systems in the subsequent sections. With dissipation, the loop integrals have to be evaluated with the corrected wave numbers yielding,

$$\tilde{q} = \exp\left(-\int_{x_1}^{x_2}[k^A(k) - k^A(k^*)]dx\right)q = \exp\left(-2i\int_{x_1}^{x_2}\mathrm{Im}[k^A(k)]dx\right)q.$$

Here k is the wave number of the unperturbed Hermitian problem that is propagated from $k(x_1)$ to $k(x_2)$ along the same path as defined for q. The complex conjugate wave number k^* is propagated from $k(x_2)$ to $k(x_1)$, and $k^A(k^*) = k^{A*}(k)$ according to Eq. (53).

The reflection and transmission coefficients (54) are obtained from the connection formulas by taking the squared magnitudes of the wave amplitudes. In this process, additional loop integrals are obtained as in Eq. (23). In the Hermitian case, these are identical with q; however, they are different from $\tilde{q}$ in the non-Hermitian case. Taking the squared magnitudes of amplitudes yields in the dissipative case

$$\tilde{\tilde{q}} = \exp\left(-\int_{x_1}^{x_2}[k^A(k) + k^{A*}(k)]dx\right)q = \exp\left(-2\int_{x_1}^{x_2}\mathrm{Re}[k^A(k)]dx\right)q.$$

The difference between $\tilde{q}$ and $\tilde{\tilde{q}}$ arises because the wave numbers of non-Hermitian systems no longer occur as pairs of complex conjugate roots.

The dissipative corrections can be included when the tunneling problem for the Hermitian case has been solved. As an example, the mode coupling coefficients given by Eq. (58) can be generalized to nearly Hermitian systems by the expressions,

$$R = \left|\frac{1-\tilde{q}/4}{1+\tilde{q}/4}\right|^2, \quad T = \frac{|\tilde{\tilde{q}}|}{|1+\tilde{q}/4|^2}.$$

In the following, we always assume Hermiticity and derive the mode coupling coefficients for a number of examples.

III. TUNNELING WITH ZERO REAL PART OF K

We first apply the WKB method to one of the most common tunneling problems. It consists of a tunneling region with a negative k^2 branch between two cutoff points where $k = 0$. In this case, tunneling is analogous to the quantum-mechanical problem of wave penetration through a potential barrier. We derive the general connection formula graphically and make comparison with an exact solution for second-order equations with a parabolic profile.

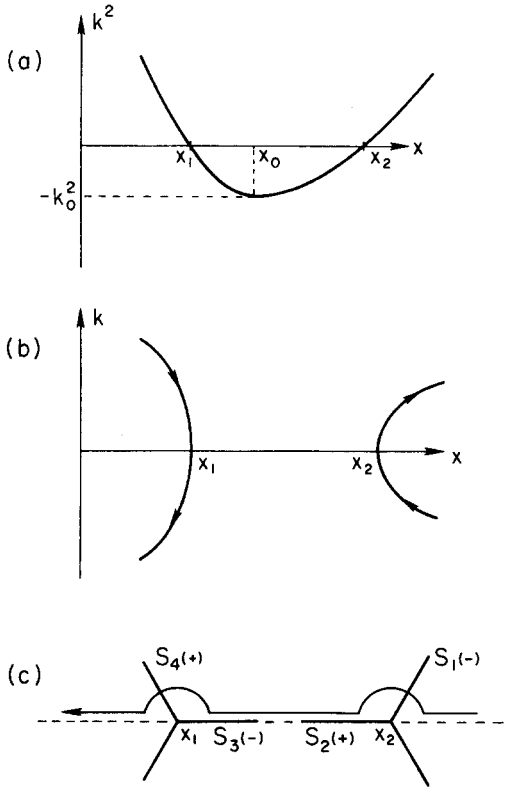
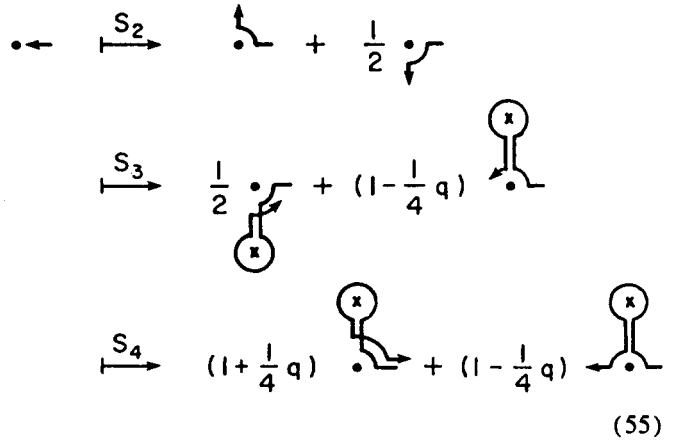


FIG. 4. Potential barrier model of wave tunneling. (a) Tunneling region with $k^2<0$. (b) Branches of $\Lambda(x,k)=0$ and energy flow directions. (c) Stokes lines.

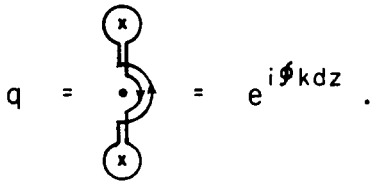
A. Graphical solution

If k^2 becomes negative for some values of x, we obtain a tunneling region as shown in Fig. 4. It has branch points of the function $k=k(x)$ at $k=0$, $x=x_{1,2}$ where mode coupling takes place. There are also branch points of the function $x=x(k)$ at $x=x_0$, $k=\pm ik_0$. For definiteness, we will assume that the k^2 branch describes forward waves with positive wave energy. We propagate the transmitted wave ($k>0$, $x>x_2$) back to the region $x<x_1$ along the path indicated in Fig. 4(c). The asymptotic representation is changed whenever a dominant wave crosses, follows, or leaves a Stokes line. The transmitted wave is subdominant on S_1, but on the subsequent lines S_2, S_3, S_4 the following waves are induced:

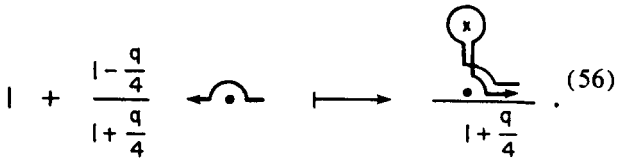


(55)

where

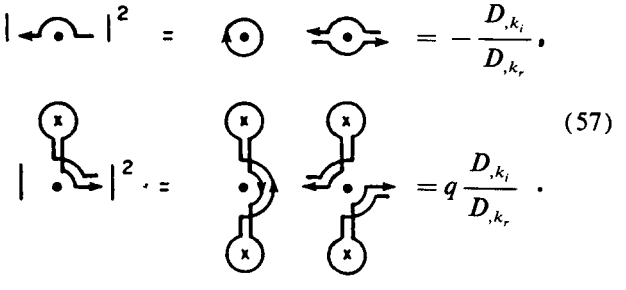


In these diagrams, the inner arcs around $k=0$ correspond to the passage around x_2 and the outer ones to the passage around x_1. The last line of Eq. (55) shows the wave solution in the region $x<x_1$, consisting of an incident wave ($k>0$) and a reflected wave ($k<0$). Dividing by the incident wave, there follows the connection formula,



(56)

To determine the mode coupling coefficients, we have to evaluate the energy flow expression (35) for each wave. This can be done graphically by use of Eq. (23) for the magnitude of wave amplitudes. For the diagrams of Eq. (56), one obtains,



(57)

With these relations, the reflection and transmission coefficients, as defined in Eq. (34), assume the form,

$$R=\frac{(1-q/4)^2}{(1+q/4)^2},\quad T=\frac{|q|}{(1+q/4)^2}. \tag{58}$$

In the expression for T, we have taken the magnitude of q to allow for later generalizations where q may become negative. It is readily seen that this result obeys energy conservation in the form $R+T=1$. This is a consequence of the modified Furry rule that introduces an appropriate correction to the reflected wave amplitude.

B. Model equation

As an example for wave tunneling with a negative branch of k^2, we consider a second-order equation with a parabolic profile,

$$\Psi''(x)+(x^2/4-a)\Psi(x)=0, \tag{59}$$

and a constant $a>0$. We specialize the general solution to the present case by setting

$$E_f=\Psi,\quad C_{ff}=1,\quad \Lambda=-k^2+x^2/4-a. \tag{60}$$

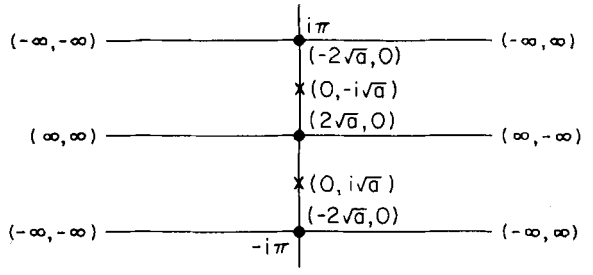


FIG. 5. Strip $-i\pi < \mathrm{Im}(\tau) < i\pi$ of the τ plane with values of (z,k) at representative points.

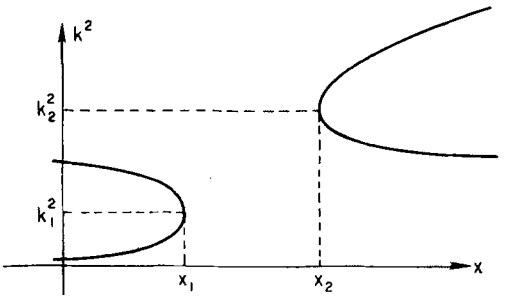


FIG. 6. Tunneling region with real branch points $k_B \neq 0$.

To evaluate the diagrams for the wave amplitudes, it is convenient to define the branches of $\Lambda(z,k) = 0$ in the parametric form,

$$z = 2\sqrt{a}\cosh(\tau), \quad k = -\sqrt{a}\sinh(\tau). \tag{61}$$

This representation is a particular solution of Eq. (6) that corresponds to the initial conditions $z(0) = 2\sqrt{a}$, $k(0) = 0$ at $\tau = 0$. The mapping (61) is 2π periodic along the imaginary τ axis and one can therefore restrict attention to the strip $-i\pi < \mathrm{Im}(\tau) < i\pi$. Some values of (z,k) corresponding to representative points in the τ plane are shown in Fig. 5.

The τ representation of the path becomes especially useful for the loop diagram q. The integration path, connecting the points $(-2\sqrt{a},0)$, $(0,i\sqrt{a})$, $(2\sqrt{a},0)$, $(0,-i\sqrt{a})$, is mapped on the section of the imaginary axis between $-i\pi$ and $+i\pi$. Noting that,

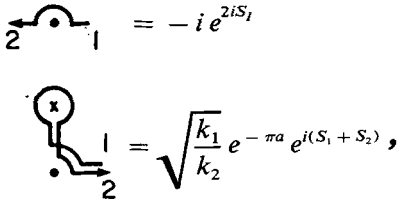
$$\int k\,dz = \int k(\tau)\frac{dz(\tau)}{d\tau}\,d\tau = a\left[\tau - \frac{1}{2}\sinh(2\tau)\right],$$

$$\oint k\,dz = 2\pi i a, \tag{62}$$

one finds,

$$q = \exp\left(i\oint k\,dz\right) = \exp(-2\pi a). \tag{63}$$

Using for the reflected wave the relations $a(k\,|1,2) = 1$, $\delta(k\,|1,2) = \pi$ and for the transmitted wave the relations $a(k\,|1,2) = \sqrt{k_1/k_2}$, $\delta(k\,|1,2) = 0$, one obtains the expressions,



(64)

where

$$S_1 = -\int_{-2\sqrt{a}}^{x_1} dx\left(\frac{x^2}{4} - a\right)^{1/2},$$

$$S_2 = \int_{2\sqrt{a}}^{x_2} dx\left(\frac{x^2}{4} - a\right)^{1/2}.$$

With Eqs. (63) and (64), the WKB connection formula (56) becomes,

$$1 - i\frac{1 - \frac{1}{4}e^{-2\pi a}}{1 + \frac{1}{4}e^{-2\pi a}}e^{2iS_1} \rightarrow \frac{e^{-\pi a}}{1 + \frac{1}{4}e^{-2\pi a}}\left(\frac{k_1}{k_2}\right)^{1/2} e^{i(S_1+S_2)}. \tag{65}$$

For comparison, we also present the exact solution for the present boundary value problem. It is given by the parabolic cylinder function $E(a,x)$ with the asymptotic representation[22]

$$1 - \frac{i}{\sqrt{1 + e^{-2\pi a}}}e^{2iS_1} \rightarrow \frac{e^{-\pi a}}{\sqrt{1 + e^{-2\pi a}}}\left(\frac{k_1}{k_2}\right)^{1/2} e^{i(S_1+S_2)}. \tag{66}$$

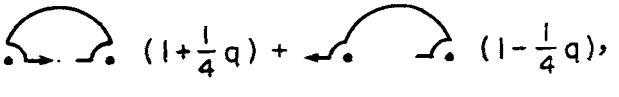
The approximate result (65) becomes valid for extended tunneling regions with separable branch points. This approximation requires $\exp(-2\pi a) \ll 1$. We note that both results (65) and (66) conserve wave energy, predict the same phases, and are in agreement up to the order $\exp(-2\pi a)$. This accuracy could not have been obtained by using the usual Furry rules.

IV. TUNNELING WITH NONZERO REAL PART OF K

We now assume branch points $k_B > 0$ on the real k axis and tunneling waves with a nonzero real part of k. In contrast to the previous case, the cofactors can change sign in these tunneling regions where k^2 is no longer real.

A. Graphical solution

In Fig. 6, we show the general structure of the tunneling region with two branch points at x_1, k_1 and x_2, k_2 which we take to be positive. We identify the incoming wave with the upper k branch in the region $x < x_1$ and assume first that the transmitted wave is given by the lower k branch in the region $x > x_2$. Analogously to the procedure in Sec. III, we propagate the transmitted wave back to the region $x < x_1$ and obtain there the asymptotic representation,



(67)

with

$$q = \quad = \sigma e^{i\oint k dz}, \quad \sigma = \operatorname{sgn}\left(\frac{W_t}{W_i}\right).$$

It has the same form as in Eq. (55). The mode coupling coefficients are still given by Eq. (58); however, with the loop diagram q as defined in Eq. (67). The conservation relation now assumes the form $R + \sigma T = 1$. If σ is negative, the reflection coefficient becomes larger than 1.

In the case where the transmitted wave is given by the upper k branch, energy conservation is obtained in exactly the same manner. It follows from Eq. (42), that all diagrams are only changed at the arcs around k_2. Accordingly, the contour in the q diagram passes around k_2 on the opposite side. Using Eq. (22), we obtain also for this case the expression (67) for the sign σ.

B. Model system

As an example of tunneling with a reversed ordering of the transmitted energy flow, we consider the crossing of two branches $a(x,k) = 0$ and $b(x,k) = 0$. A weak coupling between these branches can often be described by a matrix,[12]

$$D_{ij} = \begin{bmatrix} a(x,k) & ic \\ -ic^* & b(x,k) \end{bmatrix}, \tag{68}$$

where c denotes a small coupling constant and the functions $a(x,k)$, $b(x,k)$ have a common zero at x_0, k_0. Expanding about the crossing point x_0, k_0 up to linear order and setting $\Lambda(x,k) = 0$, one obtains

$$\tilde{k}_{1,2} = \frac{-B^+\tilde{x} \pm (B^{-2}\tilde{x}^2 - 4A|c|^2)^{1/2}}{2A}, \tag{69}$$

with $\tilde{k} = k - k_0$, $\tilde{x} = x - x_0$, $A = a_{,k}b_{,k}$ and $B^{\pm} = a_{,x}b_{,k} \pm a_{,k}b_{,x}$. Tunneling occurs if $A > 0$, yielding for the branch points x_B and the maximum imaginary part of the wave number k_m the expressions

$$\tilde{x}_B = \pm 2|c|\sqrt{A}/|B^-|, \quad k_m = |c|/\sqrt{A}. \tag{70}$$

The integral around the branch points that determines the transmission coefficient then follows to be

$$\oint k\,dz = \pi k_m |\tilde{x}_B| = 2\pi \frac{|c|^2}{|B^-|}. \tag{71}$$

This expression agrees with previous results for coupled differential equations.[12]

We now discuss the role of the cofactors in this model. From Eqs. (11) and (68) the matrix of the cofactors follows to be

$$C_{ij} = \begin{bmatrix} b(x,k) & -ic \\ ic^* & a(x,k) \end{bmatrix}. \tag{72}$$

According to the discussion following Eq. (35), the energy flow direction is determined by the zeros of one of the cofactors C_{11} or C_{22}. The equations $C_{11} = 0$, $C_{22} = 0$ describe the uncoupled branches that intersect inside the tunneling region at the crossing point. The corresponding energy flow directions are determined as in Fig. 2(b).

Outside the tunneling region, the asymptotic solution can be written in the form

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \left.\begin{bmatrix} 1 \\ C_{21}/C_{11} \end{bmatrix}\right|_{a=0} \Psi_1 + \left.\begin{bmatrix} C_{12}/C_{22} \\ 1 \end{bmatrix}\right|_{b=0} \Psi_2, \tag{73}$$

where the waves Ψ_1 and Ψ_2 correspond to the branches $a = 0$ and $b = 0$, respectively. The polarization ratios, entering this expression, are given by

$$\left.\frac{C_{12}}{C_{22}}\right|_{b=0} = \frac{-icb_{,k}}{B^-\tilde{x}},$$
$$\left.\frac{C_{21}}{C_{11}}\right|_{a=0} = \frac{-ic^*a_{,k}}{B^-\tilde{x}}, \tag{74}$$

They have apparently first been noted in Ref. 12, by iteratively solving a system of coupled wave equations. Our treatment shows that these factors arise naturally from the WKB theory of vector systems.

V. FOUR-WAVE TUNNELING PROBLEMS

In the presence of additional branch points, more than two waves can be coupled in a tunneling process. We now discuss two examples that are commonly encountered in four-wave interactions.

A. Exterior cutoff

Suppose we have the configuration of Fig. 7 with branch points at $(x_0,0)$, $(x_1, \pm k_1)$, and $(x_2, \pm k_2)$. The positive and negative branches are connected by the cutoff at x_0. In the region $x < x_0$, we assume an incoming wave with $k > 0$ and demand that no growing waves exist with $\mathrm{Im}(k) > 0$. In the region $x > x_2$, the transmitted waves are taken on the lower k branches.

Let us first construct the particular solution where only one transmitted wave ($k > 0$) is present. For this wave, the asymptotic representation in the region $x_0 < x < x_1$ is given

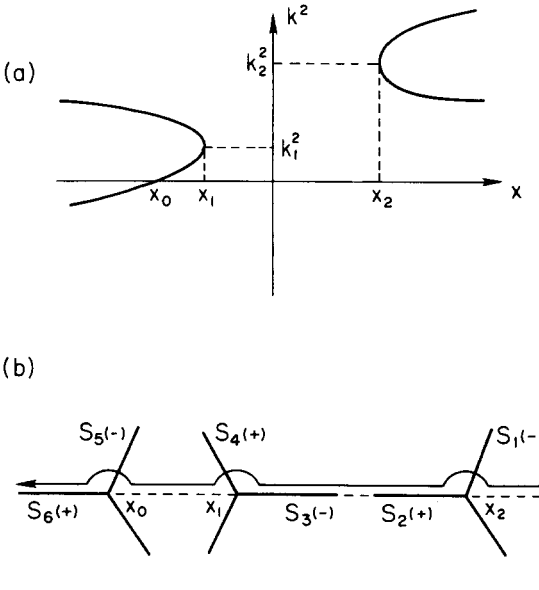
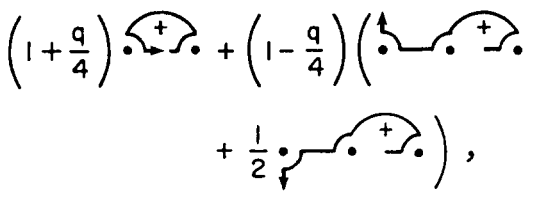


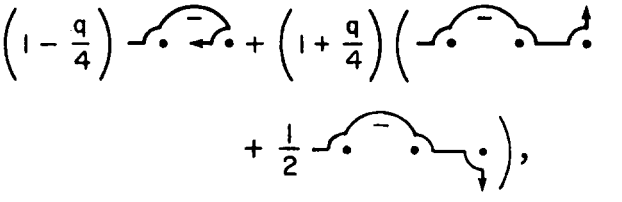
FIG. 7. Four wave tunneling problem with an exterior cutoff. (a) Tunneling configuration. (b) Schematic representation of Stokes lines.

by Eq. (67). Propagating this solution further into the region $x < x_0$, only the outgoing wave proceeds to the cutoff and induces there another wave on S_6. For $x < x_0$, the asymptotic representation then is,



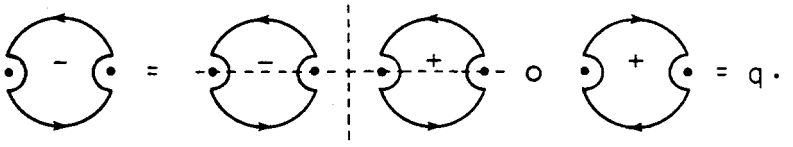
(75)

where q is given by Eq. (67) and the $+$ sign indicates that the branch points are taken on the positive k axis. The transmitted wave with $k < 0$ for $x > x_2$ leads analogously to the representation,



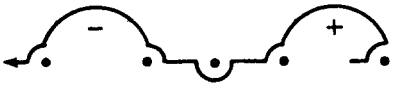
(76)

with the branch points taken on the negative k axis. We have also used Eqs. (17) and (20) to obtain the identity



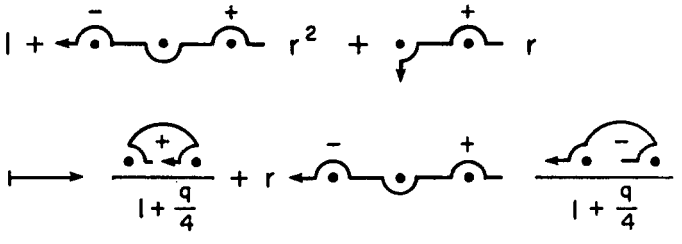
(77)

We now combine both particular solutions such that the growing waves with $\mathrm{Im}(k) > 0$ cancel. This can be achieved by adding to the first solution (75) the second solution (76) multiplied by the factor,



(78)

Dividing the result by the incident wave yields the connection formula,



(79)

The transmission coefficients T_1 and T_2 for the waves with $k > 0$ and $k < 0$, respectively, and the reflection coefficient R then are found to be,

$$T_1 = |q|/(1 + q/4)^2,$$
$$T_2 = r^2 T_1 = T_1(1 - T_1),$$
$$R = r^4 = (1 - T_1)^2. \tag{80}$$

Energy conservation holds again in the form $R + \sigma(T_1 + T_2) = 1$. This result is consistent with the physical consideration that from the incident power a fraction T_1 is transmitted to the first wave and a fraction $1 - T_1$ reflected to the cutoff. The latter energy flux can transmit a fraction $T_1(1 - T_1)$ to the second wave while a fraction $(1 - T_1)^2$ remains for the reflected wave.

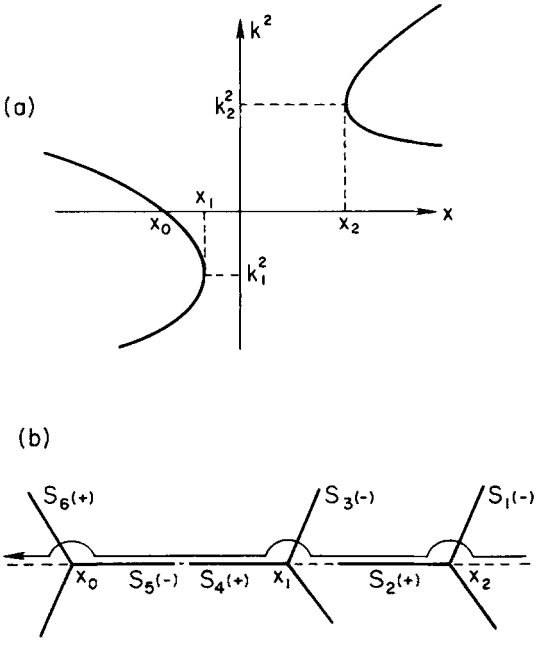
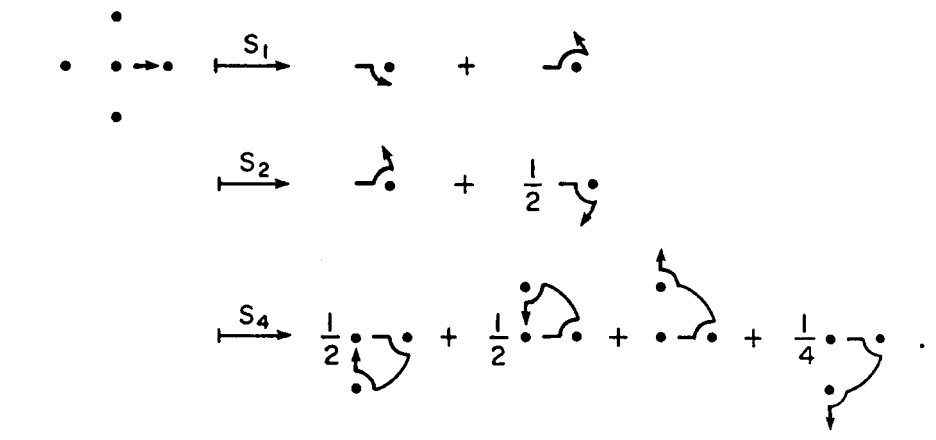


FIG. 8. Four wave tunneling problem with an additional complex branch point. (a) Tunneling configuration. (b) Corresponding Stokes lines.

B. Interior complex branch point

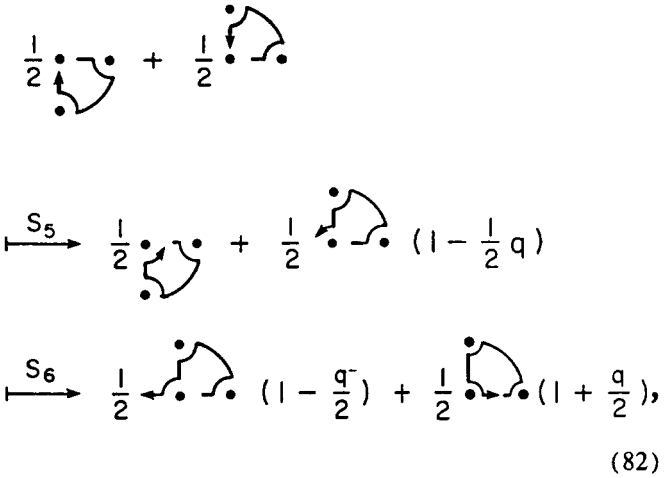
We now analyze the case where the incident wave first reaches the cutoff and tunnels then to an additional complex branch point as shown in Fig. 8. While the example in Fig. 7 could be described by two subsequent tunneling events in the form of Fig. 6, the situation in Fig. 8 is more complicated and requires an independent evaluation.

We assume the same boundary and outgoing wave conditions as before and first derive the solution corresponding to a single transmitted wave. Backward propagation of the transmitted wave with $k>0$ induces up to S_4 the following waves:



(81)

It is noted that both waves are subdominant on S_3, encircling the complex branch point counterclockwise with $\mathrm{Im}(\delta k)>0$. Proceeding to the region $x<x_0$, only the first two diagrams in the last grouping induce further waves. Their continuation yields,

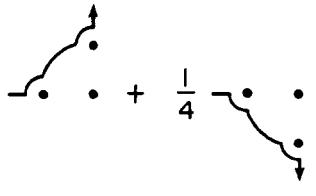


(82)

with

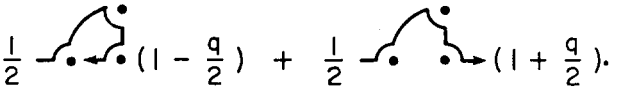
$$q = \quad = \sigma e^{i\oint k dz} \tag{83}$$

and $\sigma = \mathrm{sgn}(W_t/W_i)$. The second transmitted wave with $k<0$ leads similarly to the two tunneling waves,



(84)

and to the two propagating waves,

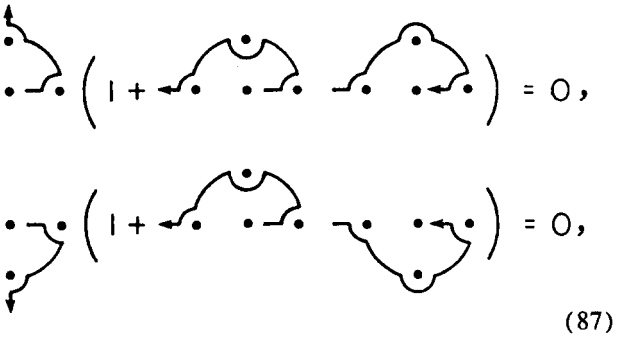


(85)

Here we used an identity analogous to Eq. (77). We now multiply the second solution by the factor,

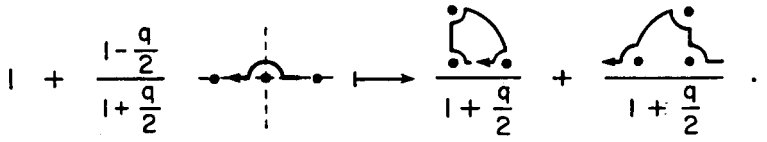
(86)

and add the product to the first solution. Then the tunneling waves cancel,



(87)

and from the propagating waves we obtain the connection formula,



(88)

The reflection and transmission coefficients now follow to be,

$$R = \frac{(1-q/2)^2}{(1+q/2)^2}, \quad T_1 = T_2 = \frac{|q|}{(1+q/2)^2}. \tag{89}$$

One can easily verify that the result (89) satisfies energy conservation, $R + \sigma(T_1 + T_2) = 1$.

The tunneling structure with an interior complex branch point occurs for mode conversion between fast Alvén waves and ion Bernstein waves in thermal plasmas at the two-ion hybrid resonance.[5,7-9] The dielectric tensor near the resonance can be approximated by the form,[23]

$$D_{ij} = \begin{bmatrix} a + \alpha n^2 & b \\ b^* & a - n^2 \end{bmatrix}, \tag{90}$$

with

$$a = 1 - \sum_j \frac{\omega_{pj}^2}{\omega^2 - \omega_{cj}^2},$$

$$b = -i \sum_j \frac{\omega_{cj}}{\omega} \frac{\omega_{pj}^2}{\omega^2 - \omega_{cj}^2},$$

$$\alpha = \sum_j \left(\frac{\omega_{pj}^2}{\omega^2 - \omega_{cj}^2} - \frac{\omega_{pj}^2}{\omega^2 - 4\omega_{cj}^2} \right) \frac{\omega^2}{\omega_{cj}^2} \frac{T}{m_j c^2},$$

$$n = \frac{ck}{\omega}.$$

where c denotes the speed of light, T the temperature, ω_p the plasma frequency, ω_c the cyclotron frequency, m the mass, and j labels the particle species. In deriving Eq. (90), dissipative processes at the cyclotron resonance frequencies have been ignored by assuming wave propagation perpendicular to the magnetic field. The singularities in Eq. (90) can be removed by a multiplication $\tilde{D}_{ij} = ND_{ij}$ with the common denominator $N = \Pi_j(\omega^2 - \omega_{cj}^2)(\omega^2 - 4\omega_{cj}^2)$, which leaves the wave number branches $n(z)$ unchanged. If $\omega_{cj}(z)$ and $\omega_{pj}(z)$ are assumed analytic, $\tilde{D}_{ij}(z,n)$ is then an analytic Hermitian form that can be treated by the present theory. From Eq. (90), the dispersion relation $\Lambda = 0$ is obtained as

$$n^4 + (1/\alpha - 1)an^2 + (1/\alpha)(|b|^2 - a^2) = 0. \tag{91}$$

The branches for the wave numbers are given by, $n^2 = -A \pm \sqrt{B}$, where

$$A = (1/\alpha - 1)(a/2),$$

$$B = A^2 + (a^2 - |b|^2)/\alpha. \tag{92}$$

There occurs a cutoff for $a^2 = |b|^2$ and further branch points at $n^2 = -A$ for $a^2 = 4\alpha|b|^2/(1+\alpha)^2$. The asymptotic branches are given by $n^2 = a(1 - |b|^2/a^2)$ for Alvén waves, and by $n^2 = -a/\alpha$ for the ion Bernstein waves. Assuming now that the parameter a varies from positive to negative values along the x axis, one obtains the general structure of Fig. 8. The contours $C_{ff} = 0$ are also easily obtained from $C_{11} = a - n^2$ and $C_{22} = a + \alpha n^2$. They describe asymptotes to the Alvén wave branches ($a > 0$) and to the Bernstein wave branches ($a < 0$) that intersect at $a = 0$. The orientation of the energy flow is therefore as illustrated in Fig. 2(b). The imaginary parts of the wave numbers in the two sections of the tunneling region can be expressed as

$$k_i = \pm \frac{\omega}{c} \begin{cases} \sqrt{A - \sqrt{B}}, & B > 0, \\ \sqrt{A + (A^2 + |B|)^{1/2}/2}, & B < 0. \end{cases} \tag{93}$$

These determine the mode coupling coefficients according to Eqs. (83) and (89) for general profiles.

VI. CONCLUSIONS

Our analysis of wave tunneling has led to transmission coefficients that are expressed by loop integrals around the branch points of the tunneling structure. These integrals are entirely determined by the local dispersion relation and therefore will not depend on the specific operator representation that has been used. This independence may explain the general agreement of various mode coupling models on the form of the transmission coefficients.[7-14]

The direction of the energy flow depends similarly on the dispersion relation, but generally also on the cofactors of the dielectric tensor. These cofactors describe the vector structure of the wave amplitudes and allow for the correct propagation ordering on the branches separated by a tunneling region. The vector representation in the present treatment is therefore appropriate for describing mode conversion with forward and backward waves without regard to specific models.

Energy conservation in the WKB theory has been obtained by the use of a modified Furry rule and dissipation has been included perturbatively. Our technique applies to extended tunneling regions where each branch point can be treated separately. When this approximation is justified, we can then describe coupling with coalescing branches both on the real and the imaginary k axis. For nonseparable branch points, different WKB techniques have been developed, which are, however, limited to specific second order-equations.[18,24]

In summary, the present treatment of wave tunneling is based on the general dielectric tensor of the medium, conserves the physical wave energy for dissipationless systems, and can be applied to arbitrary mode complexity by following simple graphical rules.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (Bonn, FRG) and by the U. S. Department of Energy, Contract No. DE-FG05-80ET53088.

APPENDIX: DERIVATION OF MODIFIED FURRY RULE

The contour integral representation (27), (28) yields for the components of the electric field,

$$E_i(z) = \int_c dk\, \hat{a}_i(k) \exp[ikz - i\hat{S}(k)], \tag{A1}$$

with

$$\hat{a}_i(k) = \frac{1}{\sqrt{2\pi i}} \left(\frac{C_{if}}{C_{ff}\sqrt{D_{,z}}} \right) \Bigg|_{z=z(k)},$$

$$\hat{S}(k) = \hat{S}(k_B) + \int_{k_T}^{k} dk'\, z(k').$$

The reader is reminded that $z(k)$ is determined from the local dispersion relation $\Lambda(z,k) = 0$ and that $dz/dk = 0$ at the branch point k_T. Assuming a sufficiently rapidly varying phase, we can approximately evaluate Eq. (A1) by expanding $z(k)$ about k_T. With $z_T = z(k_T)$, $z_T'' = z''(k_T)$, $\delta k = k - k_T$, $\delta z = z - z_T$, the two WKB wavelets that merge at $k = k_T$ can then be accurately described by the integral form,

$$E_i(z) = \exp[ik_T z - \hat{S}(k_T)] \int_c dk\, \hat{a}_i(k) \times \exp\left[i\delta z \delta k - i\frac{z_T''}{6}\delta k^3 + \cdots \right]. \tag{A2}$$

The k-space integral is basically an Airy integral if the sum is truncated after the cubic term and the amplitude $\hat{a}_i(k)$ is slowly varying.

For the Hermitian problems we are considering, z_T, k_T^2 and consequently also z_T'' are real. For definiteness, let us assume $z_T'' > 0$. The stationary phase points are at $\delta k = \pm\sqrt{2\delta z/z_T''}$ and the corresponding solutions are then propagating waves for $\delta z > 0$ and evanescent waves for $\delta z < 0$. The contours of the k-space integration are shown in Fig. 9. In this figure the shaded regions are where the integrand is divergent as $|k - k_T| \to \infty$. For $\delta z > 0$, the saddle points A ($\delta k > 0$) and B ($\delta k < 0$) are shown in Fig. 9(a). The solid contour is the steepest descent path through point A and the dashed contour, the steepest descent path through point B. By choosing one of these contours a single propagating wave with $k(z) = k_T + \delta k$ is obtained.

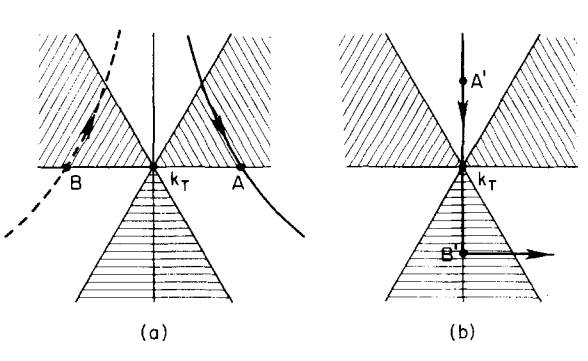


FIG. 9. Local k space contours in the vicinity of the branch point $k = k_T$. The steepest descent contour when $x - x_T > 0$, shown in (a), goes through the saddle point A and leads to a WKB propagating wave. An equivalent contour when $x - x_T < 0$ has the steepest descent contour going through the saddle points A' and B' and leads to exponentially growing and exponentially decaying waves.

Let us suppose that the solution of interest for $\delta z > 0$ corresponds to a single WKB wavelet with $\delta k > 0$. The contour of integration is then the solid contour in Fig. 9(a). For $\delta z < 0$, the saddle points are on the imaginary k axis. Equivalent to the solid contour in Fig. 9(a) is the solid contour in Fig. 9(b). The new contour goes along the directions of steepest descent through the saddle point A', where $\delta k = i\sqrt{2|\delta z|/z_T''} \equiv \delta k_1$, and the saddle point B', where $\delta k = -i\sqrt{2|\delta z|/z_T''} \equiv \delta k_2$. In approximating the contour integrals by the steepest descent method, we see that only one-half of the saddle point integral through B' is needed.

We now evaluate the contour integrals by the saddle point method. For $\delta z > 0$, the contribution from the saddle point A is given by

$$E_i(z) \approx \hat{a}_i(z,k(z)) \exp[iS(z)] \times \int_c dp \exp\left[-\frac{i}{2} z_T'' \delta k p^2 \right], \tag{A3}$$

with

$$S(z) = k_T z - \hat{S}(k_T) + \tfrac{2}{3}\delta z \delta k = S(z_T) + \int_{z_T}^{z} dz'\, k(z'),$$

and $S(z_T) = z_T k_T - \hat{S}(k_T)$. Defining $\alpha = \arg(i\delta k)$ and $p = te^{i\varphi}$, the directions of steepest descent at the saddle point satisfy the condition $\varphi = -\alpha/2 + n/\pi$. In accordance with the orientation of the solid contour in Fig. 9(a) ($\varphi = -\pi/4$) we choose $n = 0$ and find,

$$E_i(z) = \hat{a}_i(z,k(z)) e^{iS(z)} \left(\frac{2\pi}{iz_T''\delta k} \right)^{1/2} = \left(\frac{C_{if}}{iC_{ff}\sqrt{-D_{,k}}} \right) \Bigg|_{k=k(z)} e^{iS(z)}. \tag{A4}$$

Here the phase of $-D_{,k}$ is determined by the relation, $-D_{,k} = D_{,z}\, dz/dk = D_{,z} z_T'' \delta k$.

For $\delta z < 0$, there are two saddle points that the contour of integration passes through. The contribution through the point A' produces the dominant wavelet and the stationary phase contribution through the point B' one-half of the subdominant wavelet corresponding to a Stokes factor of $\frac{1}{2}$. Specifically, the steepest descent evaluation yields,

$$E_i(z) \approx \hat{a}_i|_{k=k_1} e^{iS_1} \frac{(-i)\sqrt{2\pi}}{\sqrt{|iz_T''\delta k|}} + \frac{1}{2}\hat{a}_i|_{k=k_2} e^{iS_2} \frac{\sqrt{2\pi}}{\sqrt{|iz_T''\delta k|}} = \left(\frac{C_{if}}{iC_{ff}\sqrt{-D_{,k}}} \right) \Bigg|_{k=k_1} e^{iS_1} + \frac{1}{2}\left(\frac{C_{if}}{iC_{ff}\sqrt{-D_{,k}}} \right) \Bigg|_{k=k_2} e^{iS_2}, \tag{A5}$$

where $\arg(\delta k_1) = \pi/2$, $\arg(\delta k_2) = -\pi/2$ and the subscripts 1,2 denote evaluation with $\delta k_{1,2}$, respectively. The dominant wavelet 1 is the analytic continuation in the upper-half k plane of the original WKB wavelet. The subdominant wave 2 is the analytic continuation of the wavelet 1 around

the branch point in the opposite sense of propagation. This result is just the modified Furry rule given diagrammatically in Eq. (26).

[1]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958).
[2]T. H. Stix and D. G. Swanson, in *Handbook of Plasma Physics*, edited by M. N. Rosenbluth and R. Z. Sagdeev (North-Holland, Amsterdam, 1983), Vol. 1, p. 335.
[3]H. L. Berk, R. R. Dominguez, and E. K. Maschke, Phys. Fluids **24**, 2245 (1981).
[4]R. R. Dominguez and H. L. Berk, Phys. Fluids **27**, 1142 (1984).
[5]P. L. Colestock and R. J. Kashuba, Nucl. Fusion **23**, 763 (1983).
[6]H. J. Kull, H. L. Berk, and P. J. Morrison, Phys. Fluids B **1**, 55 (1989).
[7]D. G. Swanson, Phys. Rev. Lett. **36**, 316 (1976).
[8]Y. C. Ngan and D. G. Swanson, Phys. Fluids **20**, 1920 (1977).
[9]D. G. Swanson, Phys. Fluids **28**, 2645 (1985).
[10]V. Fuchs, K. Ko, and A. Bers, Phys. Fluids **24**, 1251 (1981).
[11]V. Fuchs, A. Bers, and L. Harten, Phys. Fluids **28**, 177 (1985).
[12]R. A. Cairns and C. N. Lashmore-Davies, Phys. Fluids **26**, 1268 (1983).
[13]L. Friedland, Phys. Fluids **29**, 1105 (1986).
[14]L. Friedland, G. Goldner, and A. N. Kaufman, Phys. Rev. Lett. **58**, 1392 (1987).
[15]H. L. Berk and D. Pfirsch, J. Math. Phys. **21**, 2054 (1980).
[16]H. L. Berk, W. M. Nevins, and K. V. Roberts, J. Math. Phys. **23**, 988 (1982).
[17]J. B. Keller, Ann. Phys. **4**, 180 (1958).
[18]J. Heading, *An Introduction to Phase-Integral Methods* (Wiley, New York, 1962).
[19]S. Riyopoulos and T. Tajima, Phys. Fluids **29**, 4161 (1986).
[20]P. Henrici, *Applied and Computational Complex Analysis* (Wiley, New York, 1977).
[21]A. Bers, in *Handbook of Plasma Physics*, edited by M. N. Rosenbluth and R. Z. Sagdeev (North-Holland, Amsterdam, 1983), Vol. 1, p. 45.
[22]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[23]V. N. Oraevsky, in *Handbook of Plasma Physics*, edited by M. N. Rosenbluth and R. Z. Sagdeev (North-Holland, Amsterdam, 1983), Vol. 1, p. 243.
[24]F. L. Hinton, J. Math. Phys. **20**, 2036 (1979).

Wave splitting and the reflection operator for the wave equation in $\mathbb{R}^3$

Vaughan H. Weston
Department of Mathematics, Purdue University, West Lafayette, Indiana 47907

(Received 1 December 1988; accepted for publication 14 June 1989)

The problem of wave splitting in a nonhomogeneous medium in $\mathbb{R}^3$ is considered. Previous results for wave splitting in a planar stratified medium can be generalized to a general nonhomogeneous medium (with sufficiently smooth velocity). The wave equation is factorized into an up- and down-going wave system using certain integral and integral-differential operators. The equation for the reflection operator (which relates the up-going wave to a down-going wave) is then obtained, and certain properties of the reflection operator are deduced.

I. INTRODUCTION

One of the techniques that has been used in the time-dependent direct and inverse scattering problems associated with the one-dimensional wave equation

$$\frac{\partial^2}{\partial z^2} u(z,t) = \frac{1}{c^2(z)} \frac{\partial^2}{\partial t^2} u(z,t), \quad z,t \in \mathbb{R},$$

for a nonhomogeneous medium, is based upon the method of wave splitting.[1,2] By wave splitting we mean the decomposition of $u(z,t)$ into up-going (in the positive z direction) and down-going (in the negative z direction) waves. The importance of such splittings, in general, is that they lead to the use of invariant imbedding techniques.[3–6] Given a slab of inhomogeneous medium and a splitting one can define an associated scattering matrix. Invariant imbedding techniques then allow one to write a complex system of differential equations for the operator entries of the scattering matrix whose differentiation is with respect to the location of one of the planes of the slab. One can then deduce the behavior of the reflection operators for small time which provides a connection between up- and down-going wave fields and the properties of the medium on the edge of the slab.[1,2] The reflection operator can then be used in both direct and inverse scattering problems.

Various approaches[7–10] have been tried for extending the wave splitting to a planar stratified medium with $c = c(z)$ and $u = u(x,y,z,t)$. In particular, the approach taken by the author[10] was successful in giving rise to the form of the reflection operator and the explicit Ricatti-type integral-differential equation and initial condition that the kernel of the reflection operator must satisfy. This approach on wave splitting in a planar stratified medium was subsequently generalized to apply to the dissipative wave equation[11] (telegraph equation). Furthermore the concept of wave splitting for the wave equation was extended to the case of a nonplanar stratified medium.[12] This led to an immediate application to the inverse scattering problem associated with cylindrical geometry.[13]

In this paper it is shown that the approach[10] used for wave splitting of the wave equation in a planar stratified medium can be generalized to a general smooth nonhomogeneous medium in $\mathbb{R}^3$. (Note, Ref. 8 addresses transversely inhomogeneous environments.)

The key to the procedure for wave splitting[10] in a stratified medium is based upon the development of an up- and down-going wave condition across a planar surface based upon the use of the initial-value mixed problem (Dirichlet or Neumann boundary conditions) in a half-space. The key to extending the idea to a nonhomogeneous medium is to generalize the imbedding concept where one takes a variable plane surface bounding the nonhomogeneous medium on one side and takes the medium in the external half-space to be independent of the perpendicular Cartesian variable and yet at the same time retain continuity of the properties of the medium across the surface (i.e., if the surface is given by $x_3 = x_3^0$, then the velocity c in the external medium would depend upon x_1, x_2 only). See Figs. 1 and 2.

Section II will be devoted to the structure of the fundamental solution of the wave equation in a nonhomogeneous medium. Detailed analysis on the ray coordinate system corresponding to rays arising from a point source is given in Appendix A. This is used to develop the properties of the fundamental solution.

A further refinement of the properties of the fundamental solution for the case where the velocity c is independent of x_3 is given in Sec. III.

Based upon the results of the previous sections, the up- and down-going wave condition is developed in Sec. IV for solutions in an auxiliary space where $c = c(x_1, x_2, \alpha)$, with α being a parameter.

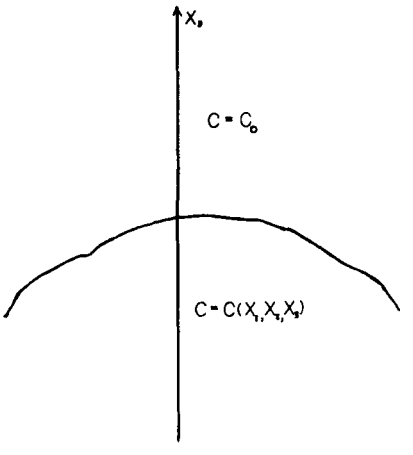


FIG. 1. Scatterer geometry. The velocity $c = c_0$ of the medium external to the scatterer is constant. The scatterer itself has variable velocity $c = c(x)$. For the direct or inverse problems referred to in the text, the sources and receivers would be located in the half-space ($x_3 >$ constant) above the scatterer.

 0022-2488/89/112545-18$02.50

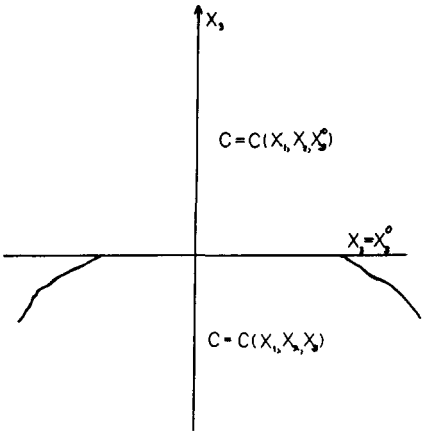


FIG. 2. Invariant imbedding geometry. For the half-space $x_3 \leqslant x_3^0$, the velocity c is the same as for scatterer geometry in Fig. 1. For $x_3 \geqslant x_3^0$, the velocity c depends upon the transverse variables x_1, x_2 only and is explicitly given by $c = c(x_1,x_2,x_3^0)$.

Using the ideas of invariant imbedding, this splitting is extended to a full nonhomogeneous medium where $c = c(x_1,x_2,x_3)$ in Sec. V. The system of equations satisfied by the up- and down-going waves is given.

The existence and form of the reflection operator relating the up-going wave to the down-going wave is given in Sec. VI. The asymptotic behavior for $t \to 0+$, of the reflection operator kernel, is also deduced there. Using the system of equations for the split system of up- and down-going waves that were developed in Sec. IV, the sought-for equation for the kernel of the reflection operator is obtained in Sec. VII. Finally, the support of the kernel of the reflection operator and its behavior at the boundary of the support is examined in Sec. VIII.

In the remainder of this paper $x = (x_1,x_2,x_3)$ and $y = (y_1,y_2,y_3)$ will denote points in $\mathbb{R}^3$, and where necessary, subscripts x and y on the Laplacian operator ∇^2 or the gradient operator ∇ will be employed to indicate the variables that are being differentiated.

II. FUNDAMENTAL SOLUTION

The form of the fundamental solution $\mathscr{E}(x,y;t)$, $x,y \in \mathbb{R}^3$, associated with the wave equation, and some of its properties will be examined here. The generalized function $\mathscr{E}(x,y;t)$ is a weak solution of the system

$$\frac{1}{c^2(x)}\frac{\partial^2}{\partial t^2}\mathscr{E} - \nabla_x^2 \mathscr{E} = \delta(x-y)\delta(t)\,, \tag{1}$$

$$\mathscr{E} = 0\,, \quad t<0\,, \tag{1'}$$

or as an alternative to systems (1), (1'), $\mathscr{E}$ is a weak solution of the initial value problem with initial conditions $\mathscr{E} = 0$, $\mathscr{E}_t = c^2(y)\delta(x-y)$ at $t = 0$.

When c is a constant, $\mathscr{E}(x,y;t)$ has the explicit form

$$\mathscr{E}(x,y;t) = \frac{\delta(t-|x-y|/c)}{4\pi|x-y|}\,,$$

where δ is the Dirac delta function. For the general case where c is a sufficiently smooth function of x, it can be shown that the fundamental solution takes the form[14]

$$\mathscr{E}(x,y;t) = \frac{\delta(t-\tau(x,y))}{\rho(x,y)} + A(x,y;t)\,, \quad 0 \leqslant t \leqslant T, \tag{2}$$

for a finite time interval. Here $\tau(x,y)$ satisfies the Eikonal equation

$$|\nabla\tau|^2 = 1/c^2(x)\,, \tag{3}$$

with asymptotic behavior as $x \to y$

$$\tau \to [1/c(y)]\,|x-y|\,. \tag{4}$$

Physically $\tau(x,y) = t$ represents "pseudospherical" wave fronts[15] diverging from the point y at time $t = 0$. The orthogonal trajectories of these wave fronts are the rays (bicharacteristics). Their equations are given in Appendix A. $\tau(x,y)$ of course represents the time of arrival of a signal travelling along the ray from the point y to x. The time constant T in the representation given by Eq. (2) for the fundamental solution is chosen, so that for the time interval $0 \leqslant t \leqslant T$, the rays originating from the point y do not intersect, i.e., there are no caustics.

The amplitude term $\rho(x,y)$ in expression (2) satisfies the equation

$$(1/\rho)\nabla_x^2\tau + 2\nabla\tau\cdot\nabla(1/\rho) = 0\,, \tag{5}$$

and its asymptotic behavior as $x \to y$ is given by

$$\rho(x,y) \to 4\pi|x-y|\,. \tag{6}$$

The second term in expression (2) satisfies the system

$$\left(\frac{1}{c^2(x)}\frac{\partial^2}{\partial t^2} - \nabla_x^2\right)A = \delta(t-\tau)\nabla_x^2\left[\frac{1}{\rho} - \frac{1}{4\pi|x-y|}\right], \tag{7}$$

$$A = 0\,, \quad t<0\,. \tag{7'}$$

In Appendix A, it is shown that a ray-coordinate system[16] (τ,θ,ϕ) centered at $x = y$, exists for a domain $0 \leqslant \tau \leqslant \tau_1$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$ and maps a region $B \subset \mathbb{R}^3$ containing the point y into the ball $0 \leqslant \tau \leqslant \tau_1$. Since τ_1 depends upon y, T is chosen so that

$$T = \min_{y\in\mathbb{R}^3} \tau_1(y)\,. \tag{8}$$

The ray system is orthogonal with metric coefficients h_τ, h_θ, h_ϕ. Employing the ray-coordinate system, Eq. (5) can be placed in the form

$$\frac{\partial}{\partial\tau}\left(\frac{S}{\rho}\right) = 0\,, \tag{9}$$

where

$$S = (h_\theta h_\phi/h_\tau)^{1/2}\,. \tag{10}$$

Using the asymptotic behavior for $\tau \to 0$ derived from Eqs. (4) and (6)

$$\rho \sim 4\pi c(y)\tau + O(\tau^2)$$

and the corresponding asymptotic behavior for S derived from Eqs. (A2) and (A9), Eq. (9) can be integrated to yield

$$\rho(x,y) = 4\pi c^{1/2}(y)\sin\theta^{-1/2}S\,. \tag{11}$$

As is pointed out in Appendix A, if c^2 is bounded from zero and is Hölder differentiable, i.e., $c^2 \in C^{(4,1)}$, then the ray coordinates $x = X(\tau,\theta,\phi)$ are three-times differentiable with respect to τ,θ,ϕ and hence $\rho(x,y)$ is twice differentiable with respect to X. To investigate the behavior of $A(x,y;t)$ set

$$\alpha(x,y;t) = \frac{1}{c(x)} \int_{-\infty}^{t} A(x,y;s)\, ds\,. \tag{12}$$

It then follows that system (7), (7′) reduces to

$$\frac{\partial^2}{\partial t^2} a - \nabla \cdot (c^2 \nabla a) - (c\nabla^2 c)a = H(t - \tau(x,y)) f(x,y)\,, \tag{13}$$

$$a(x,y;t) = 0\,, \quad t<0\,, \tag{13'}$$

where

$$f(x,y) = c(x)\nabla_x^2 \left\{ \frac{1}{\rho(x,y)} - \frac{1}{4\pi|x-y|} \right\}, \tag{14}$$

and $H(\eta)$ is the Heaviside step function.

Let the ball $0 \leqslant \tau \leqslant \tau_1$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$ in the ray coordinate system centered at y be mapped into the region B in $\mathbb{R}^3$. Let Ω be an open region containing B. We will then replace problem (13), (13′) with the Dirichlet initial-value problem in the region $\Omega \times (0 \leqslant t \leqslant T)$, with $a = 0$ on $\delta\Omega$ (the boundary of Ω). Since from Eq. (8) T is chosen so that $T \leqslant \tau_1(y)$, the boundary $\delta\Omega$ will not affect the solution. Noting that the spatial operator on the left-hand side of Eq. (13) is strongly elliptic, we can apply an existence theorem[17] on the mixed initial-value problem in the region Ω. With c sufficiently smooth so that $\rho(x,y) \in C^2$, and because of the asymptotic behavior of $\rho(x,y)$ given by Eq. (6), it follows that the right-hand side of Eq. (13) is bounded in $\Omega \times [0,T]$. It then follows from the existence theorem[17] (p. 452) that $a \in L_2((0,T); W_2^1(\Omega))$ and $\partial a/\partial t \in L_2(\Omega \times (0,T))$ with the norm for the latter term being given by

$$\left\| \frac{\partial a}{\partial t} \right\|^2 = \int_0^T \int_\Omega \left| \frac{\partial a}{\partial t} \right|^2 dx\, dt\,.$$

Since c is continuous it follows that for each fixed y,

$$A(x,y;t) \in L_2(\Omega \times (0,T))\,. \tag{15}$$

In addition, because the support of the function or the right-hand side of Eq. (13) is given by $H(t - \tau(x,y))$, it follows from the associated initial value problem (with zero initial condition) and energy integral,[18] that the support of $a(x,y;t)$ and hence $A(x,y;t)$, is contained in $H(t - \tau(x,y))$. Hence we can set

$$A(x,y;t) = H(t - \tau(x,y))A(x,y;t)\,. \tag{16}$$

The symmetry properties

$$\mathscr{E}(x,y;t-t') = \mathscr{E}(y,x;t'-t) \tag{17}$$

of the fundamental solution can be easily deduced,[14] and in particular it should be noted that $\mathscr{E}(x,y,t-t')$ satisfies the system

$$\frac{1}{c^2(y)} \frac{\partial^2 \mathscr{E}}{\partial t'^2} - \nabla_y^2 \mathscr{E} = \delta(x-y)\delta(t'-t)\,. \tag{18}$$

In addition it should be noted that $\tau(x,y)$ is symmetric,

$$\tau(x,y) = \tau(y,x)\,. \tag{19}$$

III. FUNDAMENTAL SOLUTION FOR THE AUXILIARY SPACE WHERE $c = c(x_1, x_2, \alpha)$

Before developing the wave splitting in the full nonhomogeneous medium in $\mathbb{R}^3$, we need to consider in the invariant imbedding process the special subset of the nonhomogeneous medium, where c is independent of x_3. In addition, besides being a function of x_1, x_2, c will be a differential function of a parameter α, $c = c(x_1, x_2, \alpha)$.

It will be convenient to introduce the following notation appropriate for this auxiliary space by decomposing the points in $\mathbb{R}^3$ into the components perpendicular to the x_3 axis and parallel to x_3 axis. With $x = (x_1, x_2, x_3)$, let

$$x = (\mathbf{x}, x_3), \text{ where } \mathbf{x} = (x_1, x_2).$$

The fundamental solution $\mathscr{E}^\alpha(x,y,t-t')$ associated with the equation for this space will satisfy the system

$$\frac{1}{c^2(\mathbf{x},\alpha)} \frac{\partial^2 \mathscr{E}^\alpha}{\partial t^2} - \nabla_x^2 \mathscr{E}^\alpha = \delta(\mathbf{x}-\mathbf{y})\delta(x_3 - y_3)\delta(t-t')\,, \tag{20}$$

$$\mathscr{E}^\alpha = 0\,, \quad t - t' < 0\,. \tag{20'}$$

Since c is independent of x_3 it can be shown that $\mathscr{E}(x,y;t-t')$ has the following properties: (i) translational invariance in the x_3 variable

$$\mathscr{E}^\alpha(x,y;t-t') = \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - y_3;t-t')\,, \tag{21}$$

(ii) even function of $x_3 - y_3$,

$$\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - y_3;t-t') = \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},y_3 - x_3;t-t')\,, \tag{22}$$

(iii)

$$\frac{\partial \mathscr{E}^\alpha}{\partial x_3} = -\frac{\partial \mathscr{E}^\alpha}{\partial y_3}\,. \tag{23}$$

Define $\Box_x^T$ as the transverse d'Alembertian with respect to the variables x,t,

$$\Box_x^T = \frac{1}{c^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2}\,, \tag{24}$$

with a similar definition for $\Box_y^T$

$$\Box_y^T = \frac{1}{c^2(\mathbf{y})} \frac{\partial^2}{\partial t'^2} - \frac{\partial^2}{\partial y_1^2} - \frac{\partial^2}{\partial y_2^2}\,. \tag{24'}$$

Using the result [obtained from Eq. (23)]

$$\frac{\partial^2 \mathscr{E}^\alpha}{\partial x_3^2} = \frac{\partial^2 \mathscr{E}^\alpha}{\partial y_3^2}\,,$$

together with Eq. (20) and its symmetric counterpart to Eq. (18), it can be shown that

$$\Box_x^T \mathscr{E}^\alpha(\mathbf{x},\mathbf{y};x_3 - y_3;t-t') = \Box_y^T \mathscr{E}^\alpha(\mathbf{x},\mathbf{y};x_3 - y_3;t-t')\,. \tag{25}$$

When c is independent of x_3, one can deduce some further properties of the wave fronts and in particular $\tau(x,y)$ which is used in the representation for the fundamental solution. On differentiating the Eikonal equation [Eq. (3)] with respect to x_3 and using ray coordinates (τ,θ,ϕ) centered at y (see Appendix A), one can obtain the following:

$$\nabla\tau \cdot \nabla \frac{\partial \tau}{\partial x_3} = \frac{1}{h_\tau^2} \frac{\partial}{\partial \tau}\left(\frac{\partial \tau}{\partial x_3}\right) = 0\,.$$

This implies that $\partial\tau/\partial x_3$ is constant along the ray. Using the asymptotic behavior for τ as $x \to y$, given by Eq. (4) it can be seen that

$$\frac{\partial \tau}{\partial x_3} = \frac{\cos\theta}{c(y)}\,, \tag{26}$$

where θ is associated with the local ray coordinate system centered at $x = y$. It immediately follows that the rays parameterized by $\theta = \pi/2$ lie in the $x_3 = y_3$ plane as expected. Furthermore since $\partial\tau/\partial x_3 > 0$ for $x_3 - y_3 > 0$ and $\partial\tau/\partial x_3 < 0$ for $x_3 - y_3 < 0$ it is seen that the vertical cross section of the wave fronts are concave, and that in the vertical plane, $\tau(x,y)$ achieves a maximum when $x_3 = y_3$.

Taking the second derivative of the Eikonal equation (3) with respect to x_3, one obtains

$$\left|\nabla \frac{\partial\tau}{\partial x_3}\right|^2 + \nabla\tau\cdot\nabla\frac{\partial^2\tau}{\partial x_3^2} = 0 .$$

Since $\partial\tau/\partial x_3 = 0$ in the $x_3 = y_3$ plane, it follows that

$$\frac{1}{h_\tau^2}\frac{\partial}{\partial\tau}\left(\frac{\partial^2\tau}{\partial x_3^2}\right) + \left(\frac{\partial^2\tau}{\partial x_3^2}\right)^2 = 0 , \quad x_3 = y_3 . \tag{27}$$

Since the metric coefficient $h_\tau^2 = c^2$ (Appendix A), Eq. (27) can be integrated along a ray in the $x_3 - y_3 = 0$ plane to obtain

$$\frac{\partial^2\tau}{\partial x_3^2} = \left(\int_0^\tau c^2\, d\sigma + \text{const}\right)^{-1} .$$

Using the asymptotic behavior for $\tau(x,y)$ as $x \to y$ given by Eq. (4) it can be shown that the constant is zero. Thus we have the following result:

$$\frac{\partial^2\tau}{\partial x_3^2} = \left(\int_0^\tau c^2\, d\sigma\right)^{-1} , \quad x_3 = y_3 , \tag{28}$$

where the integral is along the ray. This result will be used later on in the analysis.

IV. UP- AND DOWN-GOING WAVE CONDITION IN THE AUXILIARY SPACE

To get the up- and down-going wave condition on surfaces $x_3 = \text{const}$, in a medium where c is independent of x_3, i.e., $c = c(\mathbf{x},\alpha)$, we need to consider the mixed problem in the appropriate half-space.

Let $u(y,t)$ be the solution of the initial value, Neumann problem in the upper half space $y_3 > x_3^0$, for $0 \leqslant t' \leqslant T$,

$$\frac{1}{c^2(\mathbf{y})}\frac{\partial^2 u}{\partial t'^2} - \nabla_y^2 u = 0 , \quad y_3 > x_3^0 , \quad 0 < t' \leqslant T , \tag{29}$$

$$u(y,0) = u_t(y,0) = 0 , \quad y_3 > x_3^0 , \tag{29'}$$

$$\frac{\partial u}{\partial y_3} = v(\mathbf{y},t') , \quad y_3 = x_3^0 , \quad 0 \leqslant t' \leqslant T , \tag{30}$$

where $v(\mathbf{y},t')$ is a bounded function of compact support on the surface $y_3 = x_3^0$. Because of the support being compact and time domain being finite the problem is equivalent to the problem where the domain is a large hemisphere in the upper half space, with boundary condition $\partial u/\partial n = 0$ on the surface of the hemisphere. Since the domain here is compact, the solution[17] exists.

Set

$$\tilde{\mathscr{E}}^\alpha = \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - y_3;t - t') + \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - 2x_3^0 + y_3;t - t') . \tag{31}$$

The fundamental solution $\tilde{\mathscr{E}}^\alpha$ satisfies the system in the half-space $y_3 > x_3^0$

$$\frac{1}{c^2(\mathbf{y})}\frac{\partial^2}{\partial t'^2}\tilde{\mathscr{E}}^\alpha - \nabla_y^2\tilde{\mathscr{E}}^\alpha = \delta(\mathbf{x}-\mathbf{y})\delta(x_3 - y_3)\delta(t - t') ,$$

$$\tilde{\mathscr{E}}^\alpha = 0 , \quad t - t' < 0 \tag{32}$$

and the boundary condition

$$\frac{\partial\tilde{\mathscr{E}}^\alpha}{\partial y_3} = 0 , \quad y_3 = x_3^0 . \tag{32'}$$

Applying Green's theorem in the half-space $y_3 > x_3^0$, one obtains

$$u(x,t) = \int_0^\infty \iint_{\mathbf{R}^2} \left\{u\frac{\partial\tilde{\mathscr{E}}^\alpha}{\partial y_3} - \tilde{\mathscr{E}}^\alpha\frac{\partial u}{\partial y_3}\right\}_{y_3 = x_3^0} d\mathbf{y}\, dt'$$

$$= -2\int_0^\infty \iint_{\mathbf{R}^2} \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - x_3^0;t - t') \times v(\mathbf{y},t')\, d\mathbf{y}\, dt' , \tag{33}$$

for $x_3 > x_3^0$.

Now take the limit as $x_3 \to x_3^0$. Since $\mathscr{E}^\alpha$ has a singularity in the form $|x - y|^{-1}$, it tends to behave like a single-layer potential, and the integral given on the right-hand side of Eq. (33) is continuous across the surface $x_3 = x_3^0$. Thus we have

$$u(\mathbf{x},x_3^0;t) = -\int_0^\infty \iint_{\mathbf{R}^2} k(\mathbf{x},\mathbf{y},\alpha;t - t') \times v(\mathbf{y},t')\, d\mathbf{y}\, dt' , \tag{34}$$

where

$$k(\mathbf{x},\mathbf{y},\alpha;t) = 2\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},0;t) . \tag{35}$$

Equation (34) is the sought-for up-going wave condition relating the normal derivative of the field on the surface to itself. Because the kernel $k(\mathbf{x},\mathbf{y},\alpha;t)$ is independent of x_3 and y_3, this condition will hold on any surface $x_3 = \text{const}$.

Define the operator $\mathbf{K}_\alpha$ as the integral operator with kernel $k(\mathbf{x},\mathbf{y},\alpha;t - t')$.

With $u_n = \partial u/\partial y_3$ (the normal derivative), Eq. (34) will be placed in the form

$$u + \mathbf{K}_\alpha u_n = 0 , \quad x_3 = \text{const.} \tag{36}$$

This is the form of the up-going wave condition that we want.

By considering the mixed problem (Neumann and initial value) for the lower half space, one can obtain the down-going wave condition in a similar manner

$$u - \mathbf{K}_\alpha u_n = 0 , \quad x_3 = \text{const.} \tag{37}$$

From Eqs. (2) and (35) it is seen that $k(\mathbf{x},\mathbf{y},\alpha;t)$ has the general form

$$k(\mathbf{x},\mathbf{y},\alpha;t) = \left[2\frac{\delta(t - \tau)}{\rho} + 2H(t - \tau)A(x,y;t)\right]_{x_3 = y_3} , \tag{38}$$

where τ,ρ has the asymptotic behavior as $\mathbf{y} \to \mathbf{x}$,

$$\tau \sim |\mathbf{x} - \mathbf{y}|/c(\mathbf{x},\alpha) ,$$

$$\rho \sim 4\pi|\mathbf{x} - \mathbf{y}| .$$

For small values of t it is seen using these asymptotic results that

$$\mathbf{K}_\alpha u \sim \iint_{\mathbf{R}^2} \frac{u(\mathbf{y},t - r/c)}{2\pi r} H\left(t - \frac{r}{c}\right) d\mathbf{y} + O(t^3) ,$$

where $r=|\mathbf{x}-\mathbf{y}|$ and $c=c(\mathbf{x},\alpha)$. Using local polar coordinates this can be placed in the form

$$\mathbf{K}_\alpha u \sim \frac{1}{2\pi}\int_0^{2\pi}\int_0^{ct} u\left(\mathbf{x}+r\theta,t-\frac{r}{c}\right) dr\,d\theta + O(t^3)\,,$$

where θ is the vector with components $(\cos\theta,\sin\theta)$. Hence if $u(\mathbf{x},t)$ is a differentiable function of x_1,x_2, and t it can be shown that

$$\mathbf{K}_\alpha u \sim tc(\mathbf{x},\alpha)u(\mathbf{x},0) + (t^2/2)c(\mathbf{x},\alpha)u_t(\mathbf{x},0) + O(t^3)\,. \tag{39}$$

We will now derive an alternative form for the up- and down-going wave conditions which can lead to the existence and precise form of the inverse of $\mathbf{K}_\alpha$. (A proof for the existence of the inverse of $\mathbf{K}_\alpha$ similar to one given in Ref. 12 showing that the null space of $\mathbf{K}_\alpha$ is empty, can be obtained, but will not be given here.) The alternative form of the up- and down-going conditions will be based upon the mixed initial-value Dirichlet half-space problem.

For the up-going wave condition we will use the half-space $x_3>x_3^0$, and employ the fundamental solution

$$\begin{aligned}\tilde{\tilde{\mathscr{E}}}^\alpha(x,y;t-t') &= \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-y_3;t-t')\\ &\quad - \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-2x_3^0+y_3;t-t')\,,\end{aligned} \tag{40}$$

which satisfies the system

$$\frac{1}{c^2(y)}\frac{\partial^2}{\partial t'^2}\tilde{\tilde{\mathscr{E}}}^\alpha - \nabla_y^2\tilde{\tilde{\mathscr{E}}}^\alpha = \delta(\mathbf{x}-\mathbf{y})\delta(x_3-y_3)\delta(t-t')\,, \tag{41}$$

$$\tilde{\tilde{\mathscr{E}}}^\alpha = 0\,,\quad t-t'<0 \tag{41'}$$

for $x_3>x_3^0$ and $y_3>x_3^0$, and the boundary condition

$$\tilde{\tilde{\mathscr{E}}}^\alpha = 0\,,\quad y_3=x_3^0\,. \tag{42}$$

The solution of the mixed initial-value problem

$$\begin{aligned}&\frac{1}{c^2(\mathbf{y})}\frac{\partial^2 u}{\partial t'^2} - \nabla_y^2 u = 0\,,\quad y_3>0\,,\quad 0\leqslant t'<T,\\ &u = u_{t'} = 0\,,\quad t'=0\,,\quad y_3>x_3^0\,,\\ &u = v(\mathbf{y},t')\,,\quad y_3=x_3^0\,,\quad t'>0\,,\end{aligned} \tag{43}$$

where $v(\mathbf{y},t')$ is a twice differentiable function with respect to y_1,y_2, and t and has compact support in $\mathbb{R}^2$, can be placed in the integral form

$$u(x,t) = \int_0^\infty\int\!\!\int_{\mathbf{R}^2}\left\{\frac{\partial\tilde{\tilde{\mathscr{E}}}^\alpha}{\partial y_3}v(\mathbf{y},t')\right\}_{y_3=x_3^0} d\mathbf{y}\,dt'$$

using Green's formula. Using the result

$$\left.\frac{\partial\tilde{\tilde{\mathscr{E}}}^\alpha}{\partial y_3}\right|_{y_3=x_3^0} = -2\frac{\partial}{\partial x_3}\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-x_3^0;t-t') \tag{44}$$

this reduces to

$$\begin{aligned}u(x,t) = -2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}&\frac{\partial\mathscr{E}^\alpha}{\partial x_3}(\mathbf{x},\mathbf{y},x_3-x_3^0;t-t')\\ &\times v(\mathbf{y},t')\,d\mathbf{y}\,dt'\,.\end{aligned} \tag{45}$$

Using the identity valid when $x_3>x_3^0$, $y_3=x_3^0$,

$$\frac{\partial^2\mathscr{E}^\alpha}{\partial x_3^2} = \Box_x^T\mathscr{E}^\alpha\,,$$

it follows that on using relation (25)

$$\begin{aligned}\frac{\partial}{\partial x_3}u(x,t) &= -2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}\Box_x^T\mathscr{E}^\alpha v(\mathbf{y},t')\,d\mathbf{y}\,dt'\,,\\ &= -2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}\Box_y^T\mathscr{E}^\alpha v(\mathbf{y},t')\,d\mathbf{y}\,dt'\,.\end{aligned}$$

On integrating by parts one obtains for $x_3>x_3^0$

$$\begin{aligned}\frac{\partial}{\partial x_3}u(x,t) = &-2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-x_3^0;t-t')\\ &\times\Box_y^T v(\mathbf{y},t')\,d\mathbf{y}\,dt'\\ &-2\int\!\!\int_{\mathbf{R}^2}\frac{v(\mathbf{y},0)}{c^2(\mathbf{y})}\frac{\partial\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-x_3^0;t)\,d\mathbf{y}}{\partial t}\\ &-2\int\!\!\int_{\mathbf{R}^2}v_t(\mathbf{y},0)\frac{1}{c^2(\mathbf{y})}\\ &\times\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3-x_3^0;t)\,d\mathbf{y}\,.\end{aligned} \tag{46}$$

Since the singularity in $\mathscr{E}^\alpha$ is $|x-y|^{-1}$, the integrals are continuous in x_3. Taking the limit as $x_3\to x_3^0+$, we obtain [using the notation $\partial u/\partial x_3 = u_n$ at $x_3=x_3^0$, and replacing $v(\mathbf{y},0)$ by $u(\mathbf{y},0)$]

$$\begin{aligned}u_n(\mathbf{x},x_3^0,t) = &-\mathbf{K}_\alpha\Box^T u\\ &-2\int\!\!\int_{\mathbf{R}^2}\left\{u(\mathbf{y},x_3^0;0)\frac{\partial\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},0,t)}{\partial t}\right.\\ &\left.+u_t(\mathbf{y},x_3^0;0)\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},0;t)\right\}\frac{d\mathbf{y}}{c^2(\mathbf{y})}\end{aligned} \tag{47}$$

for $x_3=x_3^0$. An alternative form can be given by integrating the parts (treating $\mathscr{E}^\alpha$ as a distribution)

$$\begin{aligned}u_n(\mathbf{x},x_3^0,t) &= -2\Box_x^T\int_0^\infty\int\!\!\int_{\mathbf{R}}\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},0;t-t')\\ &\quad\times u(\mathbf{y},t')\,d\mathbf{y}\,dt'\\ &= -\Box^T\mathbf{K}_\alpha u\,.\end{aligned} \tag{48}$$

Expressions (47), (48) represent alternative forms of the up-going wave condition on the plane $x_3=x_3^0$. These will be rewritten for the general plane $x_3=$ constant, as follows

$$u_n + \Box^T\mathbf{K}_\alpha u = 0\,, \tag{49}$$

$$u_n + \mathbf{K}_\alpha\Box^T u + \frac{\partial}{\partial t}\mathbf{P}u + \mathbf{P}u_t = 0\,, \tag{50}$$

where

$$\mathbf{P}u = \int\!\!\int_{\mathbf{R}^2}k(\mathbf{x},\mathbf{y},\alpha;t)u(\mathbf{y},x_3;0)\frac{1}{c^2(\mathbf{y})}\,d\mathbf{y}\,. \tag{51}$$

From Eqs. (36) and (49) one can deduce the existence and form of the inverse of $\mathbf{K}_\alpha$. It is given by

$$\mathbf{K}_\alpha^{-1} = \Box^T\mathbf{K}_\alpha\,. \tag{52}$$

The range space of $\mathbf{K}_\alpha$ (domain of $\mathbf{K}_\alpha^{-1}$) will not be completely delineated here (this needs further investigation). However, it should be noted that $\mathbf{K}_\alpha$ maps $L_\infty(\mathbb{R}^2\times[0,T])$ with compact support in $\mathbb{R}^2$, into $L_\infty(\mathbb{R}^2\times[0,T])\cap[u|u(x,0)=0]$. $\mathbf{K}_\alpha^{-1}$ is a differential operator. This is seen for the special case where c is constant where in Ref. 10 an alternative form for the inverse operator is given and in

the one-dimensional case where $\mathbf{K}_\alpha^{-1}$ is given explicitly by $\mathbf{K}_\alpha^{-1} = (1/c)(\partial/\partial t)$.

When $u_n(x,t) \in C^2(\mathbb{R}^2 \times [0,T])$, Eqs. (36) and (49) yield the relation

$$u_n - \Box^T \mathbf{K}_\alpha^2 u_n = 0 .$$

This relation

$$\Box \mathbf{K}_\alpha^2 = \mathbf{I} , \tag{53}$$

where $\mathbf{I}$ is the identity operator, has been shown to be valid[19] in the space of generalized functions of slow growth in $\mathbb{R}^n$, where c is a constant.

If we take the set of functions $u(x,t) \in C^2(\mathbb{R}^2 \times [0,T])$ such that $u(x,0) = u_t(x,0) = 0$, then Eqs. (36) and (50) yield the relation

$$u - \mathbf{K}_\alpha^2 \Box^T u = 0 . \tag{53'}$$

From the conditions for up- and down-going waves we can now split the solution of the wave equation into two components as indicated by the following theorem.

Theorem: If $u(x,t)$ is a solution of

$$\frac{\partial^2 u}{c^2 \partial t^2} - \nabla^2 u = 0 , \quad x \in D ,$$
$$u = u_t = 0 , \quad \text{for} \quad t \leqslant 0, x \in D , \tag{54}$$

where $c = c(\mathbf{x},\alpha)$ in an open region D containing the plane $x_3 = \text{const}$, then u can be decomposed into an up-going wave u^+ and a down-going wave u^- as follows:

$$u = u^+ + u^- , \tag{55}$$

where

$$u^\pm = \tfrac{1}{2}[u \mp \mathbf{K}_\alpha u_n] , \tag{56}$$

and $u_n = \partial u / \partial x_3$.

Proof: We need to show that u^+ and u^- satisfy Eqs. (36) and (37), respectively. From Eq. (56) we have

$$u^\pm \pm \mathbf{K}_\alpha \frac{\partial u^\pm}{\partial x_3} = \frac{1}{2}\left[u \mp \mathbf{K}_\alpha \frac{\partial u}{\partial x_3}\right] \pm \frac{1}{2}\left[\mathbf{K}_\alpha \frac{\partial u}{\partial x_3} \mp \mathbf{K}_\alpha \frac{\partial}{\partial x_3} \mathbf{K}_\alpha \frac{\partial u}{\partial x_3}\right] .$$

Noting that $\partial/\partial x_3$ commutes with $\mathbf{K}_\alpha$, and using Eq. (53') one obtains

$$u^\pm \pm \mathbf{K}_\alpha \frac{\partial u^\pm}{\partial x_3} = \frac{1}{2}\left[u - \mathbf{K}_\alpha^2 \frac{\partial^2 u}{\partial x_3^2}\right] = \frac{1}{2}[u - \mathbf{K}_\alpha^2 \Box^T u] = \frac{1}{2}[u - \mathbf{I}u] = 0 .$$

Hence the result is shown.

V. WAVE SPLITTING IN NONHOMOGENEOUS MEDIUM IN $\mathbb{R}^3$

We will extend the wave-splitting concept of up-going and down-going waves across the surface $x_3 = \text{const}$, to the more general medium with $c = c(x_1,x_2,x_3)$. The appropriate wave splitting will be based upon the generalization of the invariant imbedding ideas developed for the one-dimensional problem.[6] We imbed the medium $c(x_1,x_2,x_3)$ into the auxiliary space $c = c(\mathbf{x},\alpha)$, where $\alpha = x_3^0$. The up-going and down-going wave splitting developed in the previous section for the auxiliary space $c(x_1,x_2,\alpha)$ will be employed on the surface $x_3 = x_3^0$, with $\alpha = x_3^0$. We will then let x_3^0 vary, and dropping the superscript zero, we have a one-parameter family of operators $\mathbf{K}_{\alpha|\alpha = x_3}$. For future analysis we will define

$$\mathbf{K} = \mathbf{K}_{\alpha|\alpha = x_3} . \tag{57}$$

The operator $\mathbf{K}$ will retain the properties of the operator $\mathbf{K}_\alpha$. The kernel of $\mathbf{K}$ is given by $k(\mathbf{x},\mathbf{y},x_3;t)$ and the inverse operator is given by

$$\mathbf{K}^{-1} = \Box^T \mathbf{K} , \tag{58}$$

where

$$\Box^T = \frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} , \tag{59}$$

and in particular the commutivity results

$$\Box^T \mathbf{K} = \mathbf{K} \Box^T \tag{60}$$

will hold for functions $u(\mathbf{x},t) \in C^2(\mathbb{R}^2 \times [0,T])$ such that $u(\mathbf{x},0) = u_t(\mathbf{x},0) = 0$.

For additional analysis, we will need the derivative of the operator $\mathbf{K}$ with respect to the parameter x_3. If $u = u(\mathbf{y},t)$ then we define

$$\left(\frac{\partial}{\partial \alpha} \mathbf{K}_\alpha\right)_{\alpha = x_3} = \mathbf{K}' u , \tag{61}$$

where

$$\mathbf{K}' u = \int_0^\infty \iint_{\mathbb{R}^2} \frac{\partial k(\mathbf{x},\mathbf{y},\alpha;t-t')}{\partial \alpha} u(\mathbf{y},t')\, d\mathbf{y}\, dt' |_{\alpha = x_3} . \tag{61'}$$

We can now apply the splitting developed in the previous section for the auxiliary space to the full space by setting

$$u^\pm = \frac{1}{2}\left(u \mp \mathbf{K} \frac{\partial u}{\partial x_3}\right) , \tag{62}$$

where u^+, u^- represent the up- and down-going wave components, respectively. It is convenient to express relation (62) in matrix form as follows

$$\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = T \begin{bmatrix} u \\ \partial u / \partial x_3 \end{bmatrix} , \quad T = \frac{1}{2}\begin{bmatrix} \mathbf{I} & -\mathbf{K} \\ \mathbf{I} & \mathbf{K} \end{bmatrix} . \tag{63}$$

Note that the inverse of T^{-1} exists and is given by

$$T^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -\mathbf{K}^{-1} & \mathbf{K}^{-1} \end{bmatrix} . \tag{64}$$

We can now derive the system of equations that must be satisfied by u^+ and u^- using the same procedure as in Refs. 1 and 2.

Let Ω be an open region in $\mathbb{R}^3$. We shall assume that u satisfies the wave equation expressed in the form

$$\frac{\partial^2 u}{\partial x_3^2} = \Box^T u , \quad x \in \Omega , \tag{65}$$

where $\Box^T$ is given by Eq. (59) and $u(x,t)$ is such that

$$u = u_t = 0 , \quad t \leqslant 0 , \quad x \in \Omega . \tag{65'}$$

Combining Eq. (65) with the identity

$$\frac{\partial u}{\partial x_3} = \frac{\partial u}{\partial x_3}$$

we obtain the system of equations for u, $\partial u/\partial x_3$ expressed in matrix form

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u \\ \partial u/\partial x_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \square^T & 0 \end{bmatrix}\begin{bmatrix} u \\ \partial u/\partial x_3 \end{bmatrix}. \tag{66}$$

This expression can be transferred into one involving u^+ and u^- by using the transformation

$$\begin{bmatrix} u \\ \partial u/\partial x_3 \end{bmatrix} = T^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix}.$$

Inserting this into expression (66) and premultiplying the resultant system by T, the following is obtained:

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = W\begin{bmatrix} u^+ \\ u^- \end{bmatrix}, \tag{67}$$

where

$$W = T\begin{bmatrix} 0 & 1 \\ \square^T & 0 \end{bmatrix}T^{-1} - T\frac{\partial T^{-1}}{\partial x_3}. \tag{68}$$

Noting that

$$-T\frac{\partial T^{-1}}{\partial x_3} = \frac{\partial T}{\partial x_3}T^{-1},$$

and

$$\frac{\partial T}{\partial x_3} = \frac{1}{2}\begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}\mathbf{K}',$$

use relation (58) for the inverse of $\mathbf{K}^{-1}$ to obtain

$$T\begin{bmatrix} 0 & 1 \\ \square_T & 0 \end{bmatrix}T^{-1} = \begin{bmatrix} -\mathbf{K}^{-1} & 0 \\ 0 & \mathbf{K}^{-1} \end{bmatrix}. \tag{68'}$$

System (67) can now be put in the explicit form

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = \begin{bmatrix} -\mathbf{K}^{-1} & 0 \\ 0 & \mathbf{K}^{-1} \end{bmatrix}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\mathbf{K}'\mathbf{K}^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix}. \tag{69}$$

The explicit form for the operator $\mathbf{K}'\mathbf{K}^{-1}$ is given in Appendix B.

VI. EXISTENCE AND FORM OF THE REFLECTION OPERATOR

Here we will show that the up-going wave u^+ is linearly related to the down-going wave u^- on a surface $x_3 = x_3^0$ (assuming of course, that there are no sources located in the lower region $x_3 \leqslant x_3^0$). The relation takes the form $u^+ = \mathbf{R}u^-$, where $\mathbf{R}$ is the reflection operator.

Let $\tilde{\mathscr{E}}(x,y;t-t')$ be the fundamental solution satisfying the system

$$\left(\frac{1}{c^2(y)}\frac{\partial^2}{\partial t'^2} - \nabla_y^2\right)\tilde{\mathscr{E}} = \delta(x-y)\delta(t-t'), \quad y_3 < x_3^0, \tag{70}$$

$$\tilde{\mathscr{E}} = 0, \quad t - t' < 0, \tag{70'}$$

and

$$\frac{\partial \tilde{\mathscr{E}}}{\partial y_3} = 0 \quad \text{on surface } y_3 = x_3^0. \tag{70''}$$

Then it can be shown in a manner similar to that used in Sec. IV, that if $u(x,t)$ is a solution of the wave equation with $c = c(x)$ in region $x_3 < x_3^0$, satisfying the initial conditions $u = u_t = 0$ at $t = 0$, and Neumann boundary condition $\partial u/\partial x_3 = u_n$ on $x_3 = x_3^0$, then the field quantity u on the surface $x_3 = x_3^0$ is related to u_n in the following manner:

$$u(\mathbf{x},x_3^0,t) = \mathbf{G}u_n, \tag{71}$$

where

$$\mathbf{G}w = \int_0^\infty \int\!\!\int_{\mathbf{R}^2} \tilde{\mathscr{E}}(\mathbf{x},x_3^0,\mathbf{y},x_3^0;t-t') \times w(\mathbf{y},t')\,d\mathbf{y}\,dt'. \tag{72}$$

Recall that in Sec. V a solution u of the wave equation with $c = c(x)$ was split into up- and down-going waves across a surface $x_3 = x_3^0$, as follows:

$$u^\pm = \tfrac{1}{2}(u \mp \mathbf{K}u_n), \quad x_3 = x_3^0,$$

where $\mathbf{K} = \mathbf{K}_\alpha$ with $\alpha = x_3^0$. Using relation (71) it then follows that

$$u^\pm = \tfrac{1}{2}(\mathbf{G} \mp \mathbf{K})u_n, \tag{73}$$

and if we define the operator $\mathbf{A}$

$$\mathbf{A} = \tfrac{1}{2}(\mathbf{G} - \mathbf{K})\mathbf{K}^{-1}, \tag{74}$$

then we see that

$$u^+ = \mathbf{A}\mathbf{K}u_n, \tag{75}$$

$$u^- = (\mathbf{I} + \mathbf{A})\mathbf{K}u_n. \tag{76}$$

In order to use Eqs. (75) and (76) to show the existence of the reflection operator $\mathbf{R}$ we need to deduce the existence of $(\mathbf{I} + \mathbf{A})^{-1}$.

From the lemma in Appendix C, we see that the solution $v(x,t) \in C(\mathbb{R}^2 \times (0,T)) \cap \{v | v(\mathbf{x},0) = 0\}$ of the equation

$$(\mathbf{G} + \mathbf{K})v = 0 \tag{77}$$

is $v = 0$. But Eq. (77) is equivalent to the equation

$$(\mathbf{I} + \mathbf{A})\mathbf{K}v = 0. \tag{78}$$

Setting $\mu = \mathbf{K}v$, it is seen that solution of the equation

$$(\mathbf{I} + \mathbf{A})\mu = 0$$

is given by $v = 0$, or $\mu = 0$. Thus the null space of $(\mathbf{I} + \mathbf{A})$ is empty. Hence $(\mathbf{I} + \mathbf{A})^{-1}$ exists.

From Eq. (76) we now can take

$$\mathbf{K}u_n = (\mathbf{I} + \mathbf{A})^{-1}u^-$$

and combining this with Eq. (75), obtain

$$u^+ = \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}u^-. \tag{79}$$

Thus we see that the reflection operator $\mathbf{R}$ exists and is given by

$$u^+ = \mathbf{R}u^-, \tag{80}$$

where $\mathbf{R} = \mathbf{A}(\mathbf{I} + \mathbf{A})^{-1}$. An alternative equation for $\mathbf{R}$ which is needed for later analysis will be derived. From Eqs. (75), (76) it is seen that

$$(\mathbf{I} + \mathbf{A})u^+ = (\mathbf{I} + \mathbf{A})\mathbf{A}\mathbf{K}u_n = \mathbf{A}(\mathbf{I} + \mathbf{A})\mathbf{K}u_n = \mathbf{A}u^-.$$

Thus from Eq. (80) we have the following equation for the reflection operator:

$$(\mathbf{I}+\mathbf{A})\mathbf{R}u^- = \mathbf{A}u^- , \tag{81}$$

where u^- is an arbitrary function belonging to $C^2(\mathbb{R}^2\times[0,T])$ such that $u^- = u_t^- = 0$ at $t=0$.

The next step is to get the explicit form of the operator **A**. From Eqs. (70), (70′), (70″) and the equation corresponding to Eq. (32) for the lower half-space $y_3 < x_3^0$, it can be shown that the difference of the fundamental solutions

$$\tilde{\mathscr{E}}(x,y;t-t') - \tilde{\mathscr{E}}^{\alpha}(x,y;t-t') ,$$

where $\alpha = x_3^0$, satisfies the system for $y_3 < x_3^0$,

$$\left(\frac{1}{c^2(\mathbf{y},\alpha)}\frac{\partial^2}{\partial t'^2} - \nabla_y^2\right)(\tilde{\mathscr{E}} - \tilde{\mathscr{E}}^{\alpha}) = \Delta c(y)\frac{\partial^2\tilde{\mathscr{E}}}{\partial t'^2}, \quad t-t'>0 ,$$

$$\tilde{\mathscr{E}} - \tilde{\mathscr{E}}^{\alpha} = 0 , \text{ for } t-t'<0 ,$$

and the boundary condition

$$\frac{\partial}{\partial y_3}(\tilde{\mathscr{E}} - \tilde{\mathscr{E}}^{\alpha}) = 0 , \quad y_3 = x_3^0 ,$$

with

$$\Delta c(y) = c^{-2}(\mathbf{y},\alpha) - c^{-2}(y) . \tag{82}$$

Hence if $u(y,t')$ satisfies the wave equation with velocity $c = c(\mathbf{y},\alpha)$, $\alpha = x_3^0$ and zero initial values $u = u_{t'} = 0$ in the lower half-space $y_3 < x_3^0$, then application of Green's theorem yields the result for $x_3 < x_3^0$,

$$\int_0^{\infty}\iint_{\mathbf{R}^2}\left\{\left(\tilde{\mathscr{E}}(x,y;t-t') - \tilde{\mathscr{E}}^{\alpha}(x,y;t-t')\right)\frac{\partial u(y,t)}{\partial y_3}\right\}_{y_3=x_3^0} dy\,dt'$$
$$= \int_0^{\infty}\iiint_{\mathbf{R}^3}\Delta c(y)\frac{\partial^2\tilde{\mathscr{E}}}{\partial t^2}(x,y;t-t')\times u(y,t')\,dy\,dt' . \tag{83}$$

Using the equivalent of Eq. (33) for the lower half-space $x_3 < x_3^0$, given by

$$u(x,t) = 2\int_0^{\infty}\iint_{\mathbf{R}^2}\frac{\partial\mathscr{E}^{\alpha}}{\partial x_3}(\mathbf{x},\mathbf{y},x_3-x_3^0;t-t')\times u(\mathbf{y},x_3^0;t')\,dy\,dt' , \tag{84}$$

one can replace $u(y,t)$ in the right-hand side of Eq. (83) by values of u on the surface $y_3 = x_3^0$. Insert expression (84) into the right-hand side of Eq. (83), interchange order of integration, and rename variables of integration. Thus the right-hand side of Eq. (83) becomes

$$2\frac{\partial^2}{\partial t^2}\int_0^{\infty}\iint_{\mathbf{R}^2}T(\mathbf{x},\mathbf{y},x_3^0;t-t')u(\mathbf{y},x_3^0,t')\,dy\,dt' , \tag{85}$$

where

$$T(\mathbf{x},\mathbf{y},x_3^0;t) = \int_0^{\infty}\iiint_{\mathbf{R}^3_-}\Delta c(z)\tilde{\mathscr{E}}(x,z;t-s)\times\frac{\partial\mathscr{E}^{\alpha}}{\partial z_3}(\mathbf{z},\mathbf{y},z_3-x_3^0;s)\,dz\,ds . \tag{86}$$

Now let $x_3 \to x_3^0$ in Eq. (83). The left-hand side of Eq. (83) becomes

$$(\mathbf{G}-\mathbf{K})u_n ,$$

where $u_n = \partial u/\partial y_3$ at $y_3 = x_3^0$. Using the down-going wave condition $u_n = \mathbf{K}^{-1}u$ associated with the lower-half space, Eq. (83) becomes when $x_3 = x_3^0$,

$$\frac{1}{2}(\mathbf{G}-\mathbf{K})\mathbf{K}^{-1}u = \frac{\partial^2}{\partial t^2}\int_0^{\infty}\iint_{\mathbf{R}^2}T(\mathbf{x},\mathbf{y},x_3^0;t-t')\times u(\mathbf{y},x_3^0;t')\,dy\,dt' . \tag{87}$$

The operator on the left-hand side of Eq. (87) is just **A**. Define the following operation:

$$T\circ u = \int_0^{\infty}\iint_{\mathbf{R}^2}T(\mathbf{x},\mathbf{y},x_3^0;t-t')u(\mathbf{y},x_3^0;t')\,dy\,dt' , \tag{88}$$

then Eq. (87) takes the form

$$\mathbf{A}u = \frac{\partial^2}{\partial t^2}(T\circ u) . \tag{89}$$

This gives us the precise form of the operator **A**.

Because of the form of the operator given by Eq. (89), it is seen from Eq. (81) that the reflection operator can be represented in the form

$$\mathbf{R}u^- = \frac{\partial^2}{\partial t^2}(R\circ u^-) , \tag{90}$$

where

$$R\circ u^- = \int_0^{\infty}\iint_{\mathbf{R}^2}R(\mathbf{x},\mathbf{y},x_3^0;t-s)u^-(\mathbf{y},x_3^0;s)\,dy\,ds . \tag{90′}$$

We will make the additional assumption on the kernel R, that

$$R\circ u^- = \frac{\partial}{\partial t}(R\circ u^-) = 0 \quad \text{at } t=0 . \tag{91}$$

Thus Eq. (81) takes the form

$$\frac{\partial^2}{\partial t^2}\left\{R\circ u^- + T\circ\frac{\partial^2}{\partial t^2}(R\circ u^-) - T\circ u^-\right\} = 0 . \tag{92}$$

Using the asymptotic behavior of T for $t\to 0$ given in Appendix D, it can be easily shown that

$$T\circ u^- = \frac{\partial}{\partial t}(T\circ u^-) = 0 \quad \text{at } t=0 .$$

Equation (92) can be integrated twice with respect to t to give

$$R\circ u^- + T\circ\frac{\partial^2}{\partial t^2}(R\circ u^-) = T\circ u^- . \tag{93}$$

Using the result that the reflection operator **R** maps u^- in $u^+ = 0$ at $t=0$, we must have in addition to Eq. (91)

$$\frac{\partial^2}{\partial t^2}(R\circ u^-) = 0 \quad \text{at } t=0 . \tag{94}$$

Using this result and the fact that u^- is an arbitrary function belonging to $C^2[\mathbb{R}^2\times(0,T)]$ such that $u = u_t = 0$ at $t=0$, it follows from Eq. (93) that the reflection kernel $R(\mathbf{x},\mathbf{y},x_3^0,t)$ satisfies the equation

$$R(\mathbf{x},\mathbf{y},x_3^0;t) + \frac{\partial^2}{\partial t^2}\int_0^{\infty}\iint_{\mathbf{R}^2}T(\mathbf{x},\mathbf{z},x_3^0;t-s)\times R(\mathbf{z},\mathbf{y},x_3^0;s)\,dz\,ds = T(\mathbf{x},\mathbf{y},x_3^0,t) . \tag{95}$$

Equation (95) is of more theoretical value than practical value. To solve for the reflection operator kernel R one needs the precise form of the fundamental solutions so as to completely specify the function T. This implies that the equation is more applicable to the direct scattering problem. In the next section an alternative equation for the reflection operator kernel will be presented which will be useful in the inverse problem.

However, Eq. (95) is useful in determining various properties of the reflection operator kernel. For instance, the asymptotic behavior of R as $t\to 0$ can be obtained from Eq. (95) using the asymptotic form of T given by Eq. (D6) in Appendix D,

$$T=\gamma H(t-|\mathbf{x}-\mathbf{y}|/c(x)),$$

where

$$\gamma(x)=-\frac{1}{8\pi c^2(x)}\frac{\partial c}{\partial x_3}.$$

Because of the nature of the support of T as $t\to 0$, one can deduce from Eq. (95) that R vanishes for $t<0$, that R has the same support [given by $H(t-|\mathbf{x}-\mathbf{y}|/c]$. In addition for nonvanishing values of R, the points $\mathbf{x}$ and $\mathbf{y}$ must be close together; thus it follows from small values of t, that R is determined by the *local* properties. Hence in obtaining the asymptotic behavior as $t\to 0+$, we may treat γ and c as a constant in Eq. (95). Also due to symmetry it follows that the dependence of R on $\mathbf{x}$ and $\mathbf{y}$ is of the form $|\mathbf{x}-\mathbf{y}|$. Thus at $t\to 0$, Eq. (95) takes the form

$$R(|\mathbf{x}-\mathbf{y}|,x_3^0;t)+\gamma\frac{\partial^2}{\partial t^2}\int\int_{R^2}\int_0^t H(t-s-|\mathbf{x}-\mathbf{z}|/c)$$
$$\times R(|\mathbf{z}-\mathbf{y}|,x_3^0;s)\,ds\,d\mathbf{z}=\gamma H(t-|\mathbf{x}-\mathbf{y}|/c) \quad (95')$$

where γ and c may be taken as constants. Since this equation is in the form of a convolution it may then be solved using the Laplace transform,

$$\mathcal{L}f=\int_0^\infty e^{-pt}f(t)\,dt$$

and Fourier transform

$$\mathcal{F}f=\int\int_{\mathbb{R}^2}e^{i\mathbf{k}\cdot\mathbf{x}}f(\mathbf{x})\,d\mathbf{x}.$$

Setting

$$\tilde{R}(\mathbf{k},p)=\mathcal{L}\mathcal{F}R,$$

and using the fact

$$\mathcal{L}\mathcal{F}H(t-|\mathbf{x}|/c)=2\pi c^2(p^2+|\mathbf{k}|^2c^2)^{-3/2},$$

Eq. (95') takes the form

$$[1+2\pi c^2\gamma p^2(p^2+|\mathbf{k}|^2c^2)^{-3/2}]\tilde{R}$$
$$=2\pi c^2\gamma(p^2+|\mathbf{k}|^2c^2)^{-3/2},$$

yielding

$$\tilde{R}=2\pi c^2\gamma(p^2+|\mathbf{k}|^2c^2)^{-3/2}-\tilde{R}_1,$$

where

$$\tilde{R}_1=-\frac{2\pi c^2\gamma}{[2\pi c^2\gamma p^2+(p^2+|\mathbf{k}|^2c^2)^{3/2}]}$$
$$+\frac{2\pi c^2\gamma}{(p^2+|\mathbf{k}|^2c^2)^{3/2}}.$$

Since the inverse Fourier transform of $\tilde{R}_1$ is bounded as follows:

$$2\pi|\mathcal{F}^{-1}\tilde{R}_1|\leqslant\left|\int_0^\infty J_0(|\mathbf{k}||\mathbf{x}|)\tilde{R}_1(|\mathbf{k}|,p)|\mathbf{k}|d|\mathbf{k}|\right|$$
$$\leqslant\int_0^\infty\tilde{R}_1(|\mathbf{k}|,p)|\mathbf{k}|d|\mathbf{k}|,$$

giving

$$|\mathcal{L}\{R(|\mathbf{x}|,x_3^0;t)-\gamma H(t-|\mathbf{x}|/c)\}|\leqslant C/p^2,$$

where

$$C=2(\pi c\gamma)^2\int_1^\infty[\eta^3+2\pi c^2\gamma\eta^{3/2}/p]^{-1}\,d\eta,$$

one can deduce the asymptotic behavior of R for $t\to 0$ from the Tauberian theorem (applied to the inverse Laplace transform).[20] However, because of the discontinuity in $R-\gamma H$, one needs to smooth or regularize it. If $\phi(\mathbf{x})$ is some test function in $C_c^0(\mathbb{R}^2)$, one can then deduce from the Tauberian theorem that

$$\int\int_{\mathbb{R}^2}[R(|\mathbf{x}-\mathbf{y}|,x_3^0;t)-\gamma H(t-|\mathbf{x}-\mathbf{y}|/c)]$$
$$\times\phi(\mathbf{y})\,d\mathbf{y}\underset{t\to 0}{\to}O(t).$$

Thus the asymptotic behavior of R for $t\to 0$ interpreted in this sense, is given by

$$R(\mathbf{x},\mathbf{y};x_3^0,t)\sim\gamma(x)H(t-|\mathbf{x}-\mathbf{y}|/c)+O(t) \quad (96)$$

where

$$\gamma(x)=-\frac{1}{8\pi c^2(x)}\frac{\partial c}{\partial x_3}.$$

In addition, estimates on the support of R can be easily obtained. This is seen as follows. Let $c(x)\leqslant c_M$ (the maximum velocity in $\mathbb{R}^3$), then the support of the fundamental solutions employed in expression (86) have the property

$$\text{Supp}\,\mathscr{E}(x,z;t-s)\subset\{z|\ |z-x|\leqslant c_M(t-s)\},$$
$$\text{Supp}\,\mathscr{E}^\alpha(\mathbf{z},\mathbf{y},z_3-x_3^0;s)\subset\{z|\ |z-y|\leqslant c_Ms\}.$$

From this it follows that

$$T(\mathbf{x},\mathbf{y},x_3^0;t)\equiv 0\ \text{if}\ c_Mt<|\mathbf{x}-\mathbf{y}|.$$

From Eq. (95) it can then be deduced that

$$R(\mathbf{x},\mathbf{y},x_3^0;t)\equiv 0\ \text{if}\ c_Mt<|\mathbf{x}-\mathbf{y}|.$$

A precise statement on the support of R will be given in Sec. VIII.

VII. EQUATION FOR THE KERNEL OF THE REFLECTION OPERATOR

In this section we will derive the equation for the reflection operator from the up- and down-going wave system. Setting

$$u^+=\mathbf{R}u^-=\frac{\partial^2}{\partial t^2}(R\circ u^-)$$

into system (69), we obtain the pair of equations

$$\frac{\partial}{\partial x_3}\left(\frac{\partial^2}{\partial t^2}(R\circ u^-)\right)+\left(\mathbf{K}^{-1}-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\right)\frac{\partial^2}{\partial t^2}(R\circ u^-)$$
$$=-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}u^-, \tag{97}$$
$$\frac{\partial u^-}{\partial x_3}-\left(\mathbf{K}^{-1}+\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\right)u^-$$
$$=-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\frac{\partial^2}{\partial t^2}(R\circ u^-). \tag{98}$$

Before combining Eqs. (97) and (98) we want to interchange the order of the operators in Eq. (97) so as to be able to integrate twice with respect to the variable t.

In what follows we will assume that $v(\mathbf{x},t)$ is a twice differentiable function of $\mathbf{x}$ and t. Note from Eq. (35) that if $v(\mathbf{x},t)$ has the property $v(\mathbf{x},0)=0$, then

$$\frac{\partial}{\partial t}\mathbf{K}_\alpha v=2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}\frac{\partial\mathscr{E}^\alpha}{\partial t}(\mathbf{x},\mathbf{y},0;t-s)v(\mathbf{y},s)\,d\mathbf{y}\,ds$$
$$=-2\int_0^\infty\int\!\!\int_{\mathbf{R}^2}\frac{\partial\mathscr{E}^\alpha}{\partial s}(\mathbf{x},\mathbf{y},0;t-s)v(\mathbf{y},s)\,d\mathbf{y}\,ds$$
$$=\mathbf{K}_\alpha\frac{\partial v}{\partial t}.$$

Setting $\alpha=x_3^0$ and using the definition for $\mathbf{K}$, we see that

$$\frac{\partial}{\partial t}\mathbf{K}v=\mathbf{K}\frac{\partial v}{\partial t},\ \text{if } v(\mathbf{x},0)=0, \tag{99}$$
$$\frac{\partial^2\mathbf{K}v}{\partial t^2}=\mathbf{K}\frac{\partial^2 v}{\partial t^2},\ \text{if } v(\mathbf{x},0)=v_t(\mathbf{x},0)=0. \tag{100}$$

One can deduce in a similar manner that

$$\frac{\partial}{\partial t}\mathbf{K}'v=\mathbf{K}'\frac{\partial v}{\partial t}\ \text{if } v(\mathbf{x},0)=0, \tag{101}$$
$$\frac{\partial^2}{\partial t^2}\mathbf{K}'v=\mathbf{K}'\frac{\partial v^2}{\partial t^2}\ \text{if } v(\mathbf{x},0)=v_t(\mathbf{x},0)=0. \tag{102}$$

It follows that if $v(\mathbf{x},0)=0$,

$$\frac{\partial}{\partial t}\mathbf{K}^{-1}v=\frac{\partial}{\partial t}\Box^T\mathbf{K}v=\Box^T\frac{\partial}{\partial t}\mathbf{K}v=\mathbf{K}^{-1}\frac{\partial v}{\partial t}, \tag{103}$$

and if $v(\mathbf{x},0)=v_t(\mathbf{x},0)=0$, then

$$\frac{\partial^2}{\partial t^2}\mathbf{K}^{-1}v=\mathbf{K}^{-1}\frac{\partial^2 v}{\partial t^2}. \tag{104}$$

Using the asymptotic behavior for $\mathbf{K}v$ for small t given by Eq. (39) and $\alpha=x_3^0$, it is seen that since $c(\mathbf{x},\alpha)=c(x)$ (after dropping the superscript 0 on x_3)

$$\mathbf{K}^{-1}v=\Box^T\mathbf{K}v=c^{-1}v_t(\mathbf{x},0)\ \text{when } t=0. \tag{105}$$

It then follows from Eq. (103) and (105) that if $v(\mathbf{x},0)=0$,

$$\frac{\partial}{\partial t}\mathbf{K}^{-1}v\Big|_{t=0}=c^{-1}v_{tt}(\mathbf{x},0). \tag{106}$$

If $v_t(\mathbf{x},0)=0=v(\mathbf{x},0)$ then it can be shown using Eq. (105) that

$$\frac{\partial}{\partial t}\mathbf{K}'\mathbf{K}^{-1}v=\mathbf{K}'\frac{\partial}{\partial t}\mathbf{K}^{-1}v=\mathbf{K}'\mathbf{K}^{-1}\frac{\partial v}{\partial t} \tag{107}$$

and if in addition $v_{tt}(\mathbf{x},0)=0$ then

$$\frac{\partial^2}{\partial t^2}\mathbf{K}'\mathbf{K}^{-1}v=\mathbf{K}'\mathbf{K}^{-1}\frac{\partial^2 v}{\partial t^2}. \tag{108}$$

Since

$$R\circ u^-=\frac{\partial}{\partial t}R\circ u^-=\frac{\partial^2}{\partial t^2}R\circ u^-=0$$

at $t=0$, we can apply these results to Eq. (97) to interchange the order of the operators to obtain the following:

$$\frac{\partial^2}{\partial t^2}\left\{\frac{\partial}{\partial x^3}(R\circ u^-)+\left(\mathbf{K}^{-1}-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\right)(R\circ u^-)\right\}$$
$$=-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}u^-. \tag{109}$$

Since it can be shown from Eqs. (39) and (61) that

$$\mathbf{K}'u\sim t\frac{\partial c}{\partial x_3}u(x,0)+O(t^2),$$

it follows from Eq. (105) that

$$\mathbf{K}'\mathbf{K}^{-1}v\sim\frac{1}{c}\frac{\partial c}{\partial x_3}tv_t(x,0)+O(t^2)$$

and

$$\mathbf{K}'\mathbf{K}^{-1}(R\circ u^-)=\frac{\partial}{\partial t}\mathbf{K}'\mathbf{K}^{-1}R\circ u^-=0\ \text{at } t=0,$$

thus Eq. (109) may be integrated twice with respect to t to give

$$\left(\frac{\partial R}{\partial x_3}\right)\circ u^-+R\circ\frac{\partial u^-}{\partial x_3}+\left(\mathbf{K}^{-1}-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\right)(R\circ u^-)$$
$$=-\frac{1}{2}\int_0^t\int_0^t\mathbf{K}'\mathbf{K}^{-1}u^-(dt)^2. \tag{110}$$

Equation (98) may now be used to eliminate $\partial u^-/\partial x_3$ from Eq. (110) to yield

$$\left(\frac{\partial R}{\partial x_3}\right)\circ u^-+\left(\mathbf{K}^{-1}-\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}\right)(R\circ u^-)$$
$$+R\circ\left(\mathbf{K}^{-1}u^-+\frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}u^-\right)$$
$$-\frac{1}{2}R\circ\mathbf{K}'\mathbf{K}^{-1}\frac{\partial^2}{\partial t^2}(R\circ u^-)$$
$$=-\frac{1}{2}\int_0^t\int_0^t\mathbf{K}'\mathbf{K}^{-1}u^-(dt)^2. \tag{111}$$

Using the fact that u^- is an arbitrary twice-differentiable function such that $u=u_t=0$ at $t=0$, the equation for the reflection operator kernel (treated as a generalized function) can be obtained from Eq. (111). To achieve this, it should be noted that the operators $\mathbf{K}u$, $\mathbf{K}'\mathbf{K}^{-1}u$ have the same form as $R\circ u$ since from Eqs. (34), (35), and (B6) it is seen that the operations can be expressed in the form

$$\mathbf{K}u=k\circ u,\quad \mathbf{K}'\mathbf{K}^{-1}u=\frac{\partial^2 m}{\partial t^2}\circ u.$$

The kernels R, k, and m, besides having compact support in $\mathbf{x},\mathbf{y}$ variables, vanish for $t<0$; hence the operators take the form of a convolution in the time variable,

$$R\circ u=\int_0^{t_+}\int\!\!\int_{\mathbf{R}^2}R(\mathbf{x},\mathbf{y};x_3;t-s)u(\mathbf{y},s)d\mathbf{y}\,ds.$$

However, it is convenient in the analysis to keep the upper limit at ∞, yet retain the idea that the kernels vanish for $t<0$. The reason for this is that the integration by parts can

be used to reduce expressions like $R\circ\partial^2 u/\partial t^2$, $R\circ\Box^T u$ to the form $\partial^2 R/\partial t^2 \circ u$, and $(\Box^T R)\circ u$, where $u = u_t = 0$ at $t = 0$. Contributions from the upper limit in the time varaible will not appear explicitly, but implicitly in generalized derivative $\partial^2 R/\partial t^2$. A similar result holds for the operations involving the kernel k.

With the above points in mind and making use of the fact the kernels of the operator $\mathbf{K}$ and $\mathbf{K}'$ are symmetric in the variables x and y, we can show the following (details left out):

$$\mathbf{K}^{-1}(R\circ u^-) = \Box^T\mathbf{K}(R\circ u^-) = (\mathbf{K}^{-1}R)\circ u^-,$$

$$R\circ(\mathbf{K}^{-1}u^-) = R\circ(\mathbf{K}\Box^T u^-) = (\Box^T\mathbf{K}R)\circ u^- = (\mathbf{K}^{-1}R)\circ u^-,$$

$$R\circ(\mathbf{K}'\mathbf{K}^{-1}u^-) = (\mathbf{K}^{-1}\mathbf{K}'R)\circ u^-,$$

$$\mathbf{K}'\mathbf{K}^{-1}(R\circ u^-) = (\mathbf{K}'\mathbf{K}^{-1}R)\circ u^-,$$

$$R\circ\left(\mathbf{K}'\mathbf{K}^{-1}\frac{\partial^2}{\partial t^2}(R\circ u^-)\right) = \left(\mathbf{K}^{-1}\mathbf{K}'R\circ\frac{\partial^2}{\partial t^2}R\right)\circ u^-,$$

and finally

$$\int_0^t\int_0^t \mathbf{K}'\mathbf{K}^{-1}u^-(dt)^2 = \int_0^\infty\int\int_{\mathbf{R}^2} m(\mathbf{x},\mathbf{y},x_3;t-s)u^-(\mathbf{y},s)d\mathbf{y}\,ds.$$

Applying these results to Eq. (111) and using the fact that u^- is arbitrary, we obtain the equation for the reflection kernel (treated as a generalized function), with x_3^0 replaced by x_3,

$$\frac{\partial R(\mathbf{x},\mathbf{y},x_3;t)}{\partial x_3} + \mathbf{K}^{-1}R(\cdot,\mathbf{y},x_3;\cdot) + \mathbf{K}^{-1}R(\mathbf{x},\cdot,x_3;\cdot) - \frac{1}{2}\mathbf{K}'\mathbf{K}^{-1}R(\cdot,\mathbf{y},x_3;\cdot) + \frac{1}{2}\mathbf{K}^{-1}\mathbf{K}'R(\mathbf{x},\cdot,x_3;\cdot) - \frac{1}{2}\left(\mathbf{K}^{-1}\mathbf{K}'R(\mathbf{x},\cdot,x_3;\cdot)\circ\frac{\partial^2}{\partial t^2}R(\cdot,\mathbf{y},x_3;\cdot)\right) = -\frac{1}{2}m(\mathbf{x},\mathbf{y},x_3;t). \tag{112}$$

Equation (112) is the sought-for "Ricatti" type equation for the kernel of the reflection operator.

As a check of its validity we will show that it reduces to the known equation for the one-dimensional case. This will be done in two steps. For the first step we will consider the special case of a stratified medium where the velocity c is independent of the transverse variables x_1, x_2.

When $c = c(x_3)$, it follows from translational invariance (in a direction perpendicular to the x_3 axis) and rotational invariance (about an axis parallel to the x_3 axis) that the reflection kernel R depends upon the transverse variables in a manner given by the following form:

$$R = R(|\mathbf{x}-\mathbf{y}|,x_3;t).$$

The operator $\mathbf{K}u$ reduces to the form

$$\mathbf{K}u = \int\int_{\mathbf{R}^2}\frac{u(\mathbf{y},t-r/c)}{2\pi r}H(t-r/c)d\mathbf{y},$$

where $r = |\mathbf{x}-\mathbf{y}|$. (Note that the corresponding operator defined in Ref. 10 differs from this by a minus sign.) From Ref. 10 and Eq. (B12), it follows that when $c = c(x_3)$,

$$\mathbf{K}'\mathbf{K}^{-1} = \mathbf{K}^{-1}\mathbf{K}' = \frac{c'(x_3)}{c^3(x_3)}\frac{\partial^2}{\partial t^2}\mathbf{K}^2,$$

and from Eq. (B6) and Eq. (A6) of Ref. 10, that

$$m(\mathbf{x},\mathbf{y},x_3;t) = \frac{c'}{2\pi c^2}\frac{H(t-r/c)}{\sqrt{c^2t^2-r^2}},$$

where $r = |\mathbf{x}-\mathbf{y}|$.

Using these results Eq. (112) can be simplified in the case where $c = c(x_3)$. First note that the second and third terms of Eq. (112) become identical, and the fourth and fifth terms cancel out. Thus Eq. (112) reduces to

$$\frac{\partial}{\partial x_3}R(|\mathbf{x}|,x_3;t) + 2\mathbf{K}^{-1}R(|\cdot|,x_3;\cdot) - \frac{1}{2}\frac{c'}{c^3}\left(\frac{\partial^2}{\partial t^2}\mathbf{K}^2R\circ R_{tt}\right) = -\frac{c'}{4\pi c^2}\frac{H(t-|\mathbf{x}|/c)}{\sqrt{c^2t^2-|\mathbf{x}|^2}}. \tag{112'}$$

The one-dimensional case can now be obtained by integrating with respect to the transverse variables over $\mathbb{R}^2$. Set

$$R(x_3;t) = \int\int_{\mathbf{R}^2}R(|\mathbf{x}|,x_3;t)d\mathbf{x}$$

and use the result

$$\int\int_{\mathbf{R}^2}\mathbf{K}R\,d\mathbf{x} = c\int_0^t R(x_3;s)ds,$$

$$\int\int_{\mathbf{R}^2}\mathbf{K}^{-1}R\,d\mathbf{x} = \frac{\partial}{c\partial t}R(x_3;t).$$

Thus Eq. (112') reduces to the one-dimensional equation

$$\frac{\partial}{\partial x_3}R(x_3;t) + \frac{2}{c}\frac{\partial}{\partial t}R(x_3;t) - \frac{c'}{2c}\int_0^t R(x_3;t-s)\times R_{tt}(x_3;s)ds = -\frac{c'}{2c}tH(t).$$

The initial conditions for $R(x_3;t)$ can be deduced from Eq. (96), yielding

$$R(x_3;0) = R_t(x_3;0) = 0,$$

$$R_{tt}(x_3,0) = -c'/4.$$

However, the reflection operator for the one-dimensional case is commonly given in the form

$$\mathbf{R}u = \int_0^t\tilde{R}(x_3;t-s)u(x_3,s)ds,$$

where $\tilde{R} = R_{tt}$. To get the equation for $\tilde{R}$, differentiate the equation for R twice with respect to t, but note that R_{tt} has a jump discontinuity at $t = 0$ ($R_{tt} = -c'/4$ for $t = 0+$, and $R_{tt} = 0$ for $t = 0-$), and the generalized derivative of this jump discontinuity cancels out the derivative of the nonhomogeneous term. Thus we end up with the standard one-dimensional equation and initial condition

$$\frac{\partial\tilde{R}}{\partial x_3} + \frac{2}{c}\frac{\partial\tilde{R}}{\partial t} - \frac{c'}{2c}\int_0^t\tilde{R}(x_3;t-s)\tilde{R}(x_3,s)ds = 0, \quad t>0 \tag{112''}$$

and $\tilde{R}(x_3;0) = -\frac{1}{4}c'$.

As a result we see that the general equation for the reflection kernel [Eq. (112)] yields the standard one-dimensional equation.

VIII. THE SUPPORT OF $R(\mathbf{x},\mathbf{y},x_3;t)$

It was pointed out in Sec. VI that for finite values of t, $0 \leqslant t \leqslant T$, $R(\mathbf{x},\mathbf{y},x_3;t) = 0$ for values of $\mathbf{x},\mathbf{y}$ such that $|\mathbf{x}-\mathbf{y}| > c_M t$, where $c_M = \max c(x)$, thus indicating then that the reflection kernel has compact suport in $\mathbb{R}^2 \times \mathbb{R}^2$. Here the precise boundary of the support region will be established, and the asymptotic behavior of R in the leading edge or neighborhood of the support boundary will be investigated. This will be achieved using the behavior of the nonhomogeneous term $m(\mathbf{x},\mathbf{y},x_3;t)$ in Eq. (112) for R.

To obtain the boundary of the support region and the behavior of m in the neighborhood of the support boundary, we need to take only the leading term for the fundamental solution $\mathscr{C}^{\alpha}$ in expression (B7) for $m(\mathbf{x},\mathbf{y},x_3^0;t)$. Thus we have

$$m(\mathbf{x},\mathbf{y},x_3^0,t) \sim \frac{\partial I}{\partial t}, \tag{113}$$

where

$$I = 4 \int\int\int_{\mathbf{R}^3_+} p(z)\, \frac{\delta(t-\tau(x,z)-\tau(z,y))}{\rho(x,z)\rho(y,z)}\, \frac{\partial \tau(y,z)}{\partial z_3}\, dz, \tag{114}$$

where $x_3 = y_3 = x_3^0$ and $\mathbb{R}^3_+$ is the half-space $z_3 > x_3^0$. τ and ρ are the wave fronts (time of arrival) and amplitude coefficients for the auxiliary space where $c = c(\mathbf{x},x_3^0)$. The points x,y will be such that $\tau(x,y) < T$. This implies that there is a single ray path which lies in the $z_3 = x_2^0$ plane going from the point x to y and vice versa. Let $\mathscr{C}_0$ represent the curve of this ray path. Since it represents the path of minimum time we have

$$\tau(x,z) + \tau(z,y) > \tau(x,y), \quad z \notin \mathscr{C}_0.$$

It immediately follows that for $t < \tau(x,y)$ then $I = 0$. This is true for the complete expression for m and not just the leading term as indicated by Eq. (113), thus we have

$$m(\mathbf{x},\mathbf{y},x_3^0;t) = 0 \quad \text{for } t < \tau(x,y). \tag{115}$$

For $t > \tau(x,y)$ let $\mathscr{S}$ be the surface given by

$$\tau(x,z) + \tau(z,y) = t, \quad z_3 \geqslant x_3^0, \tag{116}$$

And $\mathscr{A}$ its projection on the $z_3 = x_3^0$ plane with the simple closed curve $\mathscr{C}_1$ being the boundary of $\mathscr{A}$ (Fig. 3). In the limit as $t \to \tau(x,y) +$, the surface $\mathscr{S}$ approaches $\mathscr{C}_0$. Let t be slightly greater than $\tau(x,y)$, then the integral [Eq. (114) can be expressed on integrating with respect to the z_3 variable]

$$I = \int\int_{\mathscr{A}} \left[\frac{q(x,y,z)}{\rho(x,z)\rho(z,y)}\right]_{z_3 = z_3^*} dz_1\, dz_2, \tag{117}$$

where $z_3 = z_3^*$ satisfies Eq. (116), and where

$$q(x,y,z) = 4p(z)\, \frac{\partial \tau(y,z)}{\partial z_3} \left[\frac{\partial}{\partial z_3}\left[\tau(x,z) + \tau(z,y)\right]\right]^{-1}. \tag{118}$$

For $t > \tau(x,y)$, set

$$t - \tau(x,y) = \epsilon. \tag{119}$$

We want the behavior of (117) as $\epsilon \to 0 +$. The surface $\mathscr{S}$ will approach the curve $\mathscr{C}_0$, and $z_3^* \to x_3^0$. Since $\partial\tau/\partial z_3 = 0$ when $z_3 = x_3^0$ we will need to use L'Hospital's rule in the limiting expression for $q(x,y,z)$. Hence taking the second derivative with respect to z_3 and using relation (28) we obtain

$$\lim_{z \in \mathscr{S} \to z \in \mathscr{C}_0} q(x,y,z) = Q(x,y,z), \tag{120}$$

where

$$Q(x,y,z) = 4p(z)h(x,z)\left[h(x,z) + h(z,y)\right]^{-1} \tag{121}$$

with

$$h(x,z) = \int_0^{\tau(x,z)} c^2\, d\sigma. \tag{121'}$$

To complete the analysis we can use Lemma 2 in Appendix E with $\nu = 0$. Note that the integrand has a singularity at $z = y$ of order $|z-y|^{-1}$, due to the term $\rho(y,z)$. Applying the result of the lemma, we obtain the following:

$$I \underset{\epsilon \to 0}{\sim} 2(t-\tau(x,y))^{1/2} H(t-\tau(x,y)) m_0(x,y),$$

where

$$m_0(x,y) = \sqrt{2} \int_{\mathscr{C}_0} \frac{Q(x,y,z(s))\,|\psi(x,y,z(s))|^{-1/2}}{\rho(x,z(s))\rho(z(s),y)}\, ds. \tag{122}$$

Since $m(\mathbf{x},\mathbf{y},x_3^0;t)$ behaves like $\partial I/\partial t$, we have dropping the superscript 0 on x_3^0,

$$m(\mathbf{x},\mathbf{y},x_3;t) \sim (t-\tau(x,y))^{-1/2} H(t-\tau(x,y)) m_0(x,y). \tag{123}$$

Because of this singular term on the boundary of the support region, R must have the same support and its asymptotic behavior as $t \to \tau(x,y) +$ must be such as to balance the singular terms in Eq. (112). It follows then that $R(\mathbf{x},\mathbf{y},x_3;t)$ has the behavior

$$R(\mathbf{x},\mathbf{y},x_3;t) \sim H(t-\tau(x,y))\{\sqrt{t-\tau(x,y)}\, r_0(x,y) + \cdots\}, \tag{124}$$

so that

$$\frac{\partial R}{\partial x_3} \sim -\frac{1}{2} H(t-\tau(x,y)) \left\{\frac{r_0(x,y)}{\sqrt{t-\tau(x,y)}}\, \tau'(x,y) + \cdots\right\}, \tag{125}$$

where $\tau' = (\partial\tau/\partial\alpha)(x,y)|_{\alpha = x_3}$. Based upon the *assumption* that $r_0(x,y)$ has at most a singularity of order less than $|x-y|^{-1}$ at $x = y$, i.e.,

$$|x-y|\, |r_0(x,y)| < M,$$

then as will be shown, other terms in Eq. (112) will yield singularities of the order $(t-\tau(x,y))^{-1/2}$ or less.

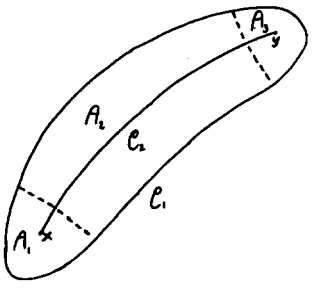


FIG. 3. Geometry for Lemma 2. The region $\mathscr{A}$ as employed in the proof of Lemma 2 is bounded by the curve $\mathscr{C}_1$ lying in the plane $z_3 = x_3^0$, and is decomposed into the three portions $\mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3$. The curve $\mathscr{C}_0$ is the ray path joining the two points x and y.

Using the behavior of m given by Eq. (123) and R given by (124) and Eq. (B6) it is seen that

$$\mathbf{K}'\mathbf{K}^{-1}R(\cdot,\mathbf{y},x_3;\cdot)\sim\frac{\partial^2}{\partial t^2}\int\int_{\mathbf{R}^2} m_0(x,z)r_0(z,y) \times I(x,y,z)dz_1\,dz_2, \tag{126}$$

where $x_3=y_3=z_3$ and where

$$I=\int_{\tau(z,y)}^{\infty}(t-s-\tau(x,z))^{-1/2}(s-\tau(z,y))^{1/2} \times H(t-s-\tau(x,z))ds \tag{127}$$

$$=\frac{\pi}{2}(t-\tau(x,z)-\tau(x,y)) \times H(t-\tau(x,z)-\tau(z,y)). \tag{127'}$$

Using the results of the lemma in Appendix E, it can be shown that the integral in expression (126) has support $H(t-\tau(x,y))$ and asymptotic behavior

$$(t-\tau(x,y))^{3/2}H(t-\tau(x,y))\tfrac{4}{3}n_0(x,y),$$

where

$$n_0(x,y)=\frac{\pi}{\sqrt{2}}\int_{\mathscr{C}_0}\frac{m_0(x,z(s))r_0(z(s),y)}{|\psi(x,y,z(s))|^{1/2}}\,ds, \tag{128}$$

where $\mathscr{C}_0$ is the ray in the plane $x_3=y_3$ from the point x to y, and $z(s)$ is a point on the ray, with s arc length. It then follows that

$$\mathbf{K}'\mathbf{K}^{-1}R(\cdot,\mathbf{y},x_3;\cdot)\sim(t-\tau(x,y))^{-1/2}H(t-\tau(x,y))n_0(x,y). \tag{129}$$

In a similar manner it can be shown that

$$\mathbf{K}^{-1}\mathbf{K}'R(\mathbf{x},\cdot,x_3;\cdot) \sim(t-\tau(x,y))^{-1/2}H(t-\tau(x,y))n_0^{+}(y,x), \tag{130}$$

where

$$n_0^{+}(y,x)=\frac{\pi}{\sqrt{2}}\int_{\mathscr{C}_0}\frac{m_0(z(s),y)r_0(x,z(s))}{|\psi(x,y,z(s))|^{1/2}}\,ds.$$

The asymptotic behavior and support of the quadratic term in Eq. (112) may now be obtained

$$\mathbf{K}^{-1}\mathbf{K}'R(\mathbf{x},\cdot,x_3,\cdot)\circ R(\cdot,\mathbf{y},x_3,\cdot) \sim\int\int_{\mathbf{R}^2} n_0(z,x)r_0(z,y)I\,dz_1\,dz_2, \tag{131}$$

where I is given by Eqs. (127), (127'). Hence using the lemma in Appendix E it can be shown that expression (131) reduces to

$$\tfrac{4}{3}(t-\tau(x,y))^{3/2}H(t-\tau(x,y))h_0(x,y),$$

where

$$h_0(x,y)=\frac{\pi}{\sqrt{2}}\int_{\mathscr{C}_0}\frac{n_0(z(s),x)r_0(z(s),y)ds}{|\psi(x,y,z(s))|^{1/2}}, \tag{132}$$

where $\mathscr{C}_0$ is the same curve as given in expression (128). Finally, we obtain

$$\frac{\partial^2}{\partial t^2}\mathbf{K}^{-1}\mathbf{K}'R\circ R\sim(t-\tau(x,y))^{-1/2}H(t-\tau(x,y))h_0(x,y). \tag{133}$$

The remaining terms in Eq. (112) involving $\mathbf{K}^{-1}R$ have a higher order singularity. This is seen as follows. Since

$$\mathbf{K}R(\cdot,\mathbf{y},x_3;\cdot)\sim 2\int\int_{\mathbf{R}^2} H(t-\tau(x,z)-\tau(z,y)) \times(t-\tau(x,z)-\tau(z,y))^{1/2}\frac{r_0(z,y)}{\rho(x,z)}\,dz_1\,dz_2,$$

use of the lemma in Appendix E yields

$$\mathbf{K}R(\cdot,\mathbf{y},x_3;\cdot)\sim(t-\tau(x,y))H(t-\tau(x,y))k_0(x,y)$$

where

$$k_0(x,y)=\sqrt{2}\,\pi\int_{\mathscr{C}_0}\frac{r_0(z(s),y)}{\rho(x,z(s))|\psi(x,y,z)|^{1/2}}\,ds. \tag{134}$$

Then using the fact that $\tau(x,y)$ satisfies the Eikonal equation

$$\mathbf{K}^{-1}R(\cdot,\mathbf{y},x_3;\cdot)\sim\Box^{T}\mathbf{K}R \sim H(t-\tau(x,y))[k_0\nabla_x^2\tau(x,y) + 2\nabla_x\tau\cdot\nabla k_0]. \tag{135}$$

Thus the terms in Eq. (112) which are singular on the boundary of the support region are given by expressions (125), (129), (130), and (133). Equating these terms in Eq. (112) we have

$$-r_0(x,y)\tau'(x,y)-n_0(x,y)+n_0^{+}(y,x) - h_0(x,y)+m_0(x,y)=0 \tag{136}$$

where $x_3=y_3$.

IX. SUMMARY

The solution of the wave equation in a smooth nonhomogeneous medium was split up into up- and down-going wave components u^{+} and u^{-}, with the decomposition given by Eqs. (62). Here the definition of an up- and down-going wave was based upon the solution in the mixed initial-value Dirichlet (or Neumann) boundary value problem for the wave equation [with velocity $c=c(x_1,x_2,x_3^0)$], in the appropriate half-space $x_3>x_3^0$ or $x_3<x_3^0$. The resulting system of equations that must be satisfied by u^{+} and u^{-} is given by Eq. (69).

The existence and form of the reflection operator $\mathbf{R}$ relating the up-going wave to the down-going wave was shown in Sec. VI. The reflection operator has the concrete form

$$\mathbf{R}u^{-}=\frac{\partial^2}{\partial t^2}R\circ u^{-},$$

where

$$R\circ u^{-}=\int_0^{\infty}\int\int_{\mathbf{R}^2}R(\mathbf{x},\mathbf{y},x_3^0;t-s)u^{-}(\mathbf{y},x_3^0;s)ds\,d\mathbf{y}.$$

In Sec. VIII it was shown that the kernel R of the reflection operator had support $H(t-\tau(x,y))$, where $\tau(x,y)$ is the time of arrival of a ray travelling from the point y to the point x in the plane $x_3=y_3=x_3^0$ with velocity $c=c(x_1,x_2,x_3^0)$. The asymptotic behavior of R as $t\to 0$ was deduced in Sec. VI and is given by Eq. (96). From this it is seen that

$$\lim_{t\to 0+}\frac{\partial^2 R}{\partial t^2}=-\frac{1}{4}\frac{\partial c}{\partial x_3}\delta(\mathbf{x}-\mathbf{y}).$$

Finally the important (Ricatti) partial differential integral equation for R was developed. It is given by Eq. (112).

The analysis used to obtain these results is valid for a finite time period T (defined in Sec. II). Some of the results may, however, be valid for a longer time.

There still remains a considerable amount of follow-up work to be done. Among the important questions that should be investigated are the following.

(i) Can the Ricatti equation (112) or an alternative form of it, and initial condition Eq. (96) for R the kernel of the reflection operator be successively used in the inverse problem as is the case for the one-dimensional problem? It will be shown in a subsequent paper that the operations $\mathbf{K}u$, $\mathbf{K}^{-1}u$, etc., can be numerically implemented. The fundamental solution (ray-tracing) representation is not used here. Instead these operations are computed using their definitions in the solution of the initial-boundary-value half-space problem as given by Eqs. (36) and (49). This is in contrast to the methods employed by Fishman[21] for the frequency domain calculations.

The main difficulty in the numerical implementation of Eq. (112) in inverse scattering lies in the taking into account of the singularity (on the support boundary) of the reflection kernel.

(ii) What scattering measurements are needed to determine R, the kernel of the reflection operator in a plane $x_3 =$ constant? This will be answered in a subsequent paper, where it will be shown that in an ideal situation (of source and receiver locations) that R can be obtained from measurements of the scattered field on a plane. For the nonideal case one would need a generalization of the deconvolution process that is used in the one-dimensional problem.

ACKNOWLEDGMENT

The research in this paper was supported by ONR on Contract N-00014-86-K-0413.

APPENDIX A: THE RAY COORDINATE SYSTEM

The rays [orthogonal trajectories to the characteristic surfaces $\tau(x,y) - t = 0$] can be represented in terms of a two-parameter family of smooth curves

$$x_i = X_i(\tau,\theta,\phi), \quad 0 \leqslant \theta \leqslant \pi, \quad 0 \leqslant \phi \leqslant 2\pi, \quad 0 \leqslant \tau \leqslant \tau_0, \tag{A1}$$

where the angles (θ,ϕ) are identified with a local spherical polar coordinate system centered at y, and where $\tau = 0$ corresponds to point $x = y$. The rays are tangent to the local radial vector at y,

$$X(\tau,\theta,\phi) = y + vc(y)\tau + O(\tau^2), \tag{A2}$$

where v is the unit vector

$$v = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{A3}$$

The rays satisfy the following differential equations[16]:

$$\frac{\partial}{\partial\tau}\left(\frac{1}{c^2}\frac{\partial X}{\partial\tau}\right) = c\nabla\frac{1}{c}. \tag{A4}$$

Using the initial condition given by Eq. (A2), Eq. (A4) may be transformed into the following system of integral equations:

$$X(\tau,\theta,\phi) = y + c(y)\tau v + c(y)v\int_0^\tau \left(\frac{c^2}{c^2(y)} - 1\right) ds - \frac{1}{2}\int_0^\tau c^2\left(\int_0^t \frac{1}{c^2}\nabla c^2\, ds\right) dt \tag{A5}$$

where of course, y is fixed.

Since v is a function of θ and ϕ, and X appears in the right-hand side of Eq. (A5) in the term involving c^2, system (A5) has the general form

$$X(\tau,\theta,\phi) = F(X(\tau,\theta,\phi)), \tag{A6}$$

where F is a vector valued function (with three components). If $c^2(x)$ is a Hölder differentiable function belonging to class $C^{(1,1)}$ and is bounded from zero, $c^2(x) \geqslant k_0 > 0$, then it can be shown (details left out) that there exist positive constants A and B such that

$$\|F(X) - F(X')\|_\infty \leqslant (A\tau + B\tau^2)\|X - X'\|_\infty,$$

where

$$\|X\|_\infty = \operatorname*{Max}_{i=1,2,3} \sup_{\Theta} |X_i(\tau,\theta,\phi)|, \tag{A7}$$

where Θ to the domain $0 \leqslant \tau < \infty$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. If τ' is the positive root of the equation $A\tau + B\tau^2 = \frac{1}{2}$, we then have for $0 \leqslant \tau \leqslant \tau'$,

$$\|F(X) - F(X')\|_\infty \leqslant \tfrac{1}{2}\|X - X'\|_\infty,$$

implying that F is a contraction operator. Hence the method of successive approximations[22]

$$X_{n+1} = F(X_n)$$

applied to Eq. (A5) starting from

$$X_0 = y + c(y)\tau v$$

will converge to a unique solution in the ball $\|X - X_0\|_\infty \leqslant r_0$ where $r_0 = 2\|X_0 - F(X_0)\|_\infty$. If τ_1 is defined as the radius of the ball in space (τ,θ,ϕ) for which a unique solution to Eq. (A5) exists, then we see that τ' yields a lower bound estimate for τ_1.

From Eq. (A4) we have

$$\frac{1}{c^2}\frac{\partial X}{\partial\tau} = \frac{v}{c(y)} - \frac{1}{2}\int_0^\tau \frac{1}{c^2}\nabla c^2\, ds; \tag{A8}$$

hence from (A4) and (A8) it can be immediately deduced that the solution X to system (A5) has the property that $\partial X/\partial\tau$, $\partial^2 X/\partial\tau^2$ are continuous in the domain $0 \leqslant \tau \leqslant \tau_1$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$.

In order for X to be twice differentiable with respect to the variables θ and ϕ one needs stronger conditions on $c(x)$. Differentiating Eq. (A5) twice with respect to θ and ϕ, it is seen that $c^2(x) \in C^{(4,1)}$ in order for X to be three-times differentiable with respect to θ and ϕ.

Since it can be deduced[16] that the Jacobian of the mapping $x = X(\tau,\theta,\phi)$ for $0 \leqslant \tau \leqslant \tau_1$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$ is greater than zero for $\tau > 0$, vanishing at $\tau = 0$, a ray coordinate system (τ,θ,ϕ) may be introduced mapping the region B in $\mathbb{R}^3$ containing the point y into the ball of radius $\tau = \tau_1$. Also since it is seen from Eq. (A2) that the ray coordinate system is locally orthogonal at $\tau = 0$, it may be deduced[16] that the ray coordinate system is orthogonal for $0 \leqslant \tau \leqslant \tau_1$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. Hence the metric coefficients h_τ, h_θ, h_ϕ can be introduced

$$h_\theta^2 = \sum_{i=1}^{3} \left(\frac{\partial X_i}{\partial \theta}\right)^2, \quad h_\phi^2 = \sum_{i=1}^{3} \left(\frac{\partial X_i}{\partial \phi}\right)^2,$$
$$h_r^2 = \sum_{i=1}^{3} \left(\frac{\partial X_i}{\partial \tau}\right)^2 = c^2. \tag{A9}$$

APPENDIX B: EVALUATION OF $(\partial \mathbf{K}_\alpha/\partial\alpha)\mathbf{K}_\alpha^{-1}$

Recall that $\mathbf{K}_\alpha$ is the integral operator with kernel $k(\mathbf{x},\mathbf{y},\alpha;t-t')$ where α is a parameter, and $(\partial/\partial\alpha)\mathbf{K}_\alpha$ is the operator with kernel $\partial k/\partial\alpha$.

From Eq. (32) it is seen that the derivative of the fundamental solution $\tilde{\mathscr{E}}^\alpha$ [defined by Eq. (31)] satisfies the system

$$\left(\frac{1}{c^2(y)}\frac{\partial^2}{\partial t'^2} - \nabla_y^2\right)\frac{\partial \tilde{\mathscr{E}}^\alpha}{\partial\alpha} = p(y)\frac{\partial^2 \tilde{\mathscr{E}}^\alpha}{\partial t'^2}, \quad x_3>0, \quad y_3>0, \tag{B1}$$

$$\frac{\partial \tilde{\mathscr{E}}^\alpha}{\partial\alpha} = 0, \quad t-t'<0, \tag{B1'}$$

and the boundary condition

$$\frac{\partial}{\partial y_3}\left(\frac{\partial \tilde{\mathscr{E}}^\alpha}{\partial\alpha}\right) = 0, \quad \text{at } y_3 = 0 \tag{B2}$$

where

$$p(\mathbf{y}) = \frac{2}{c^3}\frac{\partial c(\mathbf{y},\alpha)}{\partial\alpha}. \tag{B3}$$

Combining this with the solution of system (29), (29'), one obtains on applying Green's theorem, the following equation [corresponding to Eq. (33)]:

$$\int_0^\infty \int\int\int_{\mathbf{R}^3_+} p(\mathbf{y})\frac{\partial^2 \tilde{\mathscr{E}}^\alpha}{\partial t'^2} u(y,t')dy\,dt' = -2\int_0^\infty \int\int_{\mathbf{R}^2} \frac{\partial \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3;t-t')}{\partial\alpha} u_n(\mathbf{y},t')d\mathbf{y}\,dt' \tag{B4}$$

valid for $x_3>0$. Let $x_3\to 0$. Then in the left-hand side of Eq. (B4) use the relation [obtained from Eq. (31)]

$$\tilde{\mathscr{E}}^\alpha(x,y;t-t') = 2\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},y_3;t-t'), \quad \text{when } x_3 = 0$$

and in the right-hand side of Eq. (B4) use the relation [obtained from Eq. (35)]

$$2\frac{\partial}{\partial\alpha}\mathscr{E}^\alpha(\mathbf{x},\mathbf{y},0;t-t') = \frac{\partial}{\partial\alpha}k(\mathbf{x},\mathbf{y},\alpha;t-t').$$

Thus we have

$$\frac{\partial \mathbf{K}_\alpha}{\partial\alpha}u_n = -2\int_0^\infty \int\int\int_{\mathbf{R}^3_+} \frac{\partial^2 \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},y_3;t-t')}{\partial t'^2} \times p(\mathbf{y})u(y,t')dy\,dt. \tag{B5}$$

We want to express both sides of Eq. (B5) in terms of the value of u on the surface $y_3 = 0$. For the left-hand side of Eq. (B5) use relations (48) and (52) and for the right-hand side use relation (45). Thus we have

$$\frac{\partial \mathbf{K}_\alpha}{\partial\alpha}\mathbf{K}_\alpha^{-1}u = -4\int_0^\infty \int\int\int_{\mathbf{R}^3_+} \frac{\partial^2 \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},y_3;t-t')}{\partial t'^2} p(\mathbf{y}) \times \int_0^\infty \int\int_{\mathbf{R}^2} \frac{\partial \mathscr{E}^\alpha(\mathbf{y},\mathbf{z},y_3;t'-s)}{\partial y_3} \times u(\mathbf{z},s)d\mathbf{z}\,ds\,dy\,dt'.$$

Changing the order of integration and renaming the dummy variables of integration, this becomes

$$\left(\frac{\partial \mathbf{K}_\alpha}{\partial\alpha}\right)\mathbf{K}_\alpha^{-1}u = \frac{\partial^2}{\partial t^2}\int_0^\infty \int\int_{\mathbf{R}^2} m(\mathbf{x},\mathbf{y},\alpha;t-t') \times u(\mathbf{y},t')d\mathbf{y}\,dt', \tag{B6}$$

where

$$m(\mathbf{x},\mathbf{y},\alpha;t) = -4\int_0^\infty \int\int\int_{\mathbf{R}^3_+} \mathscr{E}^\alpha(\mathbf{x},\mathbf{z},z_3;t-s) \times p(\mathbf{z})\frac{\partial \mathscr{E}^\alpha(\mathbf{z},\mathbf{y},z_3;s)}{\partial z_3}d\mathbf{z}\,ds. \tag{B7}$$

This can be decomposed into a symmetric component

$$m_s(\mathbf{x},\mathbf{y},\alpha;t) = \tfrac{1}{2}[m(\mathbf{x},\mathbf{y},\alpha;t) + m(\mathbf{y},\mathbf{x},\alpha;t)] \tag{B8}$$

and antisymmetric component

$$m_a(\mathbf{x},\mathbf{y},\alpha;t) = \tfrac{1}{2}[m(\mathbf{x},\mathbf{y},\alpha;t) - m(\mathbf{y},\mathbf{x},\alpha;t)]. \tag{B9}$$

The symmetric component can be simplified by replacing the variable of integration s by s' in the expressions for $m(\mathbf{x},\mathbf{y},\alpha;t)$ and by $t-s'$ and in the expressions for $m(\mathbf{y},\mathbf{x},\alpha;t)$ in Eq. (B8). Then integrate by parts with respect to the z_3 variable. One then obtains

$$m_s(\mathbf{x},\mathbf{y},\alpha;t) = \frac{1}{2}\int_0^\infty \int\int_{\mathbf{R}^2} k(\mathbf{x},\mathbf{z},\alpha;t-s') \times p(\mathbf{z})k(\mathbf{z},\mathbf{y};\alpha;s')d\mathbf{z}\,ds'. \tag{B10}$$

This indicates that

$$\int_0^\infty \int\int_{\mathbf{R}^2} m_s(\mathbf{x},\mathbf{y},\alpha;t-s)u(\mathbf{y},s)d\mathbf{y}\,ds = \frac{1}{2}\mathbf{K}_\alpha p(\mathbf{K}_\alpha u). \tag{B11}$$

It can be shown that when c is independent of the spatial variables depending only on the parameter α, then the antisymmetric portion vanishes, and with p being a function of α only, the right-hand side of Eq. (B11) reduces to $\frac{1}{2}p(\mathbf{K}_\alpha)^2u$. Thus when $c = c(\alpha)$, we have

$$\frac{\partial \mathbf{K}_\alpha}{\partial\alpha}\mathbf{K}_\alpha^{-1} = \frac{1}{2}p\frac{\partial^2}{\partial t^2}(\mathbf{K}_\alpha)^2 \tag{B12}$$

a result that agrees with the calculations in Ref. 10.

APPENDIX C: LEMMA 1

Here we prove the Lemma.

Lemma 1: The solution $v(\mathbf{x},t)\in C(\mathbb{R}^2\times(0,T)) \cap\{v|v(\mathbf{x},0)=0\}$ of the equation $(\mathbf{G}+\mathbf{K})v = 0$ is $v\equiv 0$.

Consider a medium where $c = c(x_1,x_2,x_3^0)$ for $x_3>x_3^0$, otherwise $c = c(x_1,x_2,x_3)$ for $x_3<x_3^0$. Apart from the dis-

continuity in the normal derivative of c across the surface $x_3 = x_3^0$, c will be a smooth differentiable function.

Let $v(\mathbf{x},t) \in C(\mathbb{R}^2 \times (0,T))$ such that $v(x,0) = 0$. Then

$$u(x,t) = \int_0^\infty \int_{\mathbb{R}^2} \tilde{\mathscr{E}}(\mathbf{x},\mathbf{y},x_3^0;t-t')v(\mathbf{y},t')\,d\mathbf{y}\,dt', \quad x_3 < x_3^0, \tag{C1}$$

represents the solution of the wave equation with zero initial conditions $u = u_t = 0$ in the half-space $x_3 < x_3^0$, and satisfying the Neumann boundary condition $\partial u/\partial x_3 = v$ on the surface $x_3 = x_3^0$. Let $x_3 \to x_3^0 -$, then [from Eq. (72)]

$$u(\mathbf{x},x_3^0 -,t) = \mathbf{G}v. \tag{C2}$$

In addition it can be shown [from Eq. (33)] that

$$u(x,t) = -2\int_0^\infty \iint_{\mathbb{R}^2} \mathscr{E}^\alpha(\mathbf{x},\mathbf{y},x_3 - x_3^0,t-t') \times v(\mathbf{y},t')\,d\mathbf{y}\,dt', x_3 > x_3^0, \tag{C3}$$

represents the solution of the wave equation with zero initial conditions in the half-space $x_3 > x_3^0$ and Neumann boundary condition $\partial u/\partial x_3 = v$ on the surface $x_3 = x_3^0$. Let $x_3 \to x_3^0 +$, then

$$u(\mathbf{x},x^0 +;t) = -\mathbf{K}v. \tag{C4}$$

Combining Eqs. (C2) and (C4) it is seen that condition $(\mathbf{G}+\mathbf{K})v = 0$ implies that $u(\mathbf{x},x_3^0 -;t) = u(\mathbf{x},x_3^0 +;t)$, i.e., $u(x,t)$ is continuous across the surface $x_3 = x_3^0$. Since $v(x,0) = 0$, it follows from Eq. (39) that $u(x,0) = u_t(x,0) = 0$ on the surface $x_3 = x_3^0$. Since the boundary condition $\partial u/\partial x_3 = v$ on both sides implies that the normal derivative is continuous across $x = x_3^0$, we can apply the energy integral[18] to the system

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = 0, t > 0, \quad x \in \mathbb{R}^3,$$
$$u = u_t = 0, \quad t = 0, \quad x \in \mathbb{R}^3. \tag{C5}$$

Multiply Eq. (C5) by $2u_t$ to obtain the equation

$$\frac{\partial}{\partial t}\left\{\frac{1}{c^2}u_t^2 + |\nabla u|^2\right\} - 2\nabla\cdot(u_t \nabla u) = 0. \tag{C6}$$

We will take a bounded domain $\mathscr{D}_{t'}$, $t' \leqslant T$ which is homeomorphic to the cylinder $K_1 \times (0,t') = \{(x,t)\colon |\mathbf{x}| \leqslant 1, t \in (0,t')\}$ and possesses the following properties: its lower and upper bases lie in the plane $t = 0$ and $t = t'$; its lateral surface $S_{t'}$ is oriented at all points in a space-characteristic way and the normals $n = (n_1,n_2,n_3,n_4)$ to the lateral surface directed outside of $\mathscr{D}_{t'}$ form acute angles with the t axis, and satisfy the conditions

$$n_4^2 - c^2\sum_{i=1}^3 n_i^2 \geqslant 0, \quad n_4 > 0. \tag{C7}$$

Take the integral of expression (C6) over the domain $\mathscr{D}_{t'}$. Then integrate by parts (taking into account that u, $\partial u/\partial x_3$ are continuous across $x_3 = x_3^0$) to obtain

$$\int_{\Omega_{t'}}\left[\frac{1}{c^2}u_t^2 + |\nabla u|^2\right]dx + \int_{S_{t'}}\left[\left(\frac{1}{c^2}u_t^2 + |\nabla u|^2\right)n_4 - 2u_t\sum_{i=1}^3 n_i\frac{\partial u}{\partial x_i}\right]ds = 0 \tag{C8}$$

using the initial conditions $u = u_t = 0$ at $t = 0$. Here $\Omega_{t'}$ is the intersection of the domain $\mathscr{D}_{t'}$ with the plane $t = t'$. Using conditions (C7) it can be shown that the integrand in the surface integral satisfies the inequality

$$\frac{1}{n_4}\sum_{i=1}^3\left(n_4\frac{\partial u}{\partial x_i} - u_t n_i\right)^2 + \frac{u_t^2}{n_4 c^2}\left(n_4^2 - c^2\sum_{i=1}^3 n_i^2\right) \geqslant 0;$$

thus we have

$$\int_{\Omega_{t'}}\left(\frac{1}{c^2}u_t^2 + |\nabla u|^2\right)dx = 0.$$

This implies then that $u_t = \nabla u = 0$ for all values of t', $0 \leqslant t' \leqslant T$.

Now since $u \equiv 0$, $\partial u/\partial x_3 = 0$ on $x_3 = x_3^0$ it follows that $v = 0$. □

APPENDIX D: ASYMPTOTIC BEHAVIOR OF $T(\mathbf{x},\mathbf{y},x_3^0;t)$ FOR $t \to 0$

Upon examining the support of the fundamental solutions in expression (86), it is seen that nonvanishing values of $T(\mathbf{x},\mathbf{y},x_3^0;t)$ for $t \to 0$ are obtained when $\mathbf{x} \sim \mathbf{y}$, and the main contribution to the integral in expression (86) comes from values of z in the neighborhood of x. Hence the fundamental solution $\tilde{\mathscr{E}}(x,z;t)$ may be approximated by $2\mathscr{E}^\alpha(x,z;t)$ where $\alpha = x_3^0$. Thus taking the leading term for the fundamental solutions, using the asymptotic expressions given by Eqs. (4), (6) with $c(x) \sim c(y)$, we have

$$\tilde{\mathscr{E}}(x,z;t-s) \sim \frac{2\delta(t-s-|x-z|/c(x))}{4\pi|x-z|},$$

$$\frac{\partial\mathscr{E}^\alpha(\mathbf{z},\mathbf{y},z_3 - x_3^0;s)}{\partial z_3} \sim -\left\{\frac{\delta'(s-|y-z|/c(x))}{4\pi|y-z|^2 c(x)} + \frac{\delta(s-|y-z|/c(x))}{4\pi|y-z|^3}\right\}(z_3 - x_3^0),$$

$$\Delta c(z) \sim p(x)(z_3 - x_3^0),$$

where

$$p(z) = \frac{2}{c^3(z)}\frac{\partial c(z)}{\partial z_3}. \tag{D1}$$

Insert these expressions into Eq. (86) to obtain

$$T(\mathbf{x},\mathbf{y},x_3^0;t) \sim -\frac{p(x)}{8\pi^2}\left\{\frac{1}{c}\frac{\partial f_1}{\partial t} + f_2\right\}, \tag{D2}$$

where

$$f_j = \iiint_{\mathbb{R}^3_-}\delta\left(t - \frac{|x-z| + |y-z|}{c}\right) \times \frac{(z_3 - x_3^0)^2}{|x-z||y-z|^{j+1}}\,dz, \tag{D3}$$

with $c = c(x)$, and $j = 1,2$.

This integral may be evaluated using a prolate spheroidal coordinate system[23] (ξ,η,ϕ) oriented so that the axis of revolution passes through the points x and y and centered at the midpoint $(x+y)/2$. The half-space $\mathbb{R}^3_-$ is given by $1 < \xi$, $-1 < \eta < 1$, $\pi < \phi < 2\pi$. The volume element dz becomes

$$dz = \left(\frac{r}{2}\right)^3 (\xi^2 - \eta^2)\, d\xi\, d\eta\, d\phi$$

where

$$r = |x - y|/2 = |\mathbf{x} - \mathbf{y}|/2 ,$$

and

$$|x - z| = (r/2)(\xi + \eta) ,$$
$$|y - z| = (r/2)(\xi - \eta) ,$$
$$(z_3 - x_3^0) = (r/2)[(\xi^2 - 1)(1 - \eta^2)]^{1/2} \sin\phi .$$

Thus the integral in Eq. (D3) becomes

$$f_j = \int_{\pi}^{2\pi} \int_{-1}^{1} \int_{1}^{\infty} \left(\frac{r}{2}\right)^{3-j} \frac{(\xi^2 - 1)(1 - \eta^2)}{(\xi - \eta)^j} \times \delta\left(t - \frac{r}{c}\xi\right) \sin^2\phi\, d\xi\, d\eta\, d\phi , \quad \text{(D4)}$$

which reduces to

$$f_j = \frac{\pi c}{4}\left(\frac{r}{2}\right)^{2-j} \int_{-1}^{1} \frac{[(ct/r)^2 - 1](1 - \eta^2)}{((ct/r) - \eta)^j} \times d\eta\, H(t - |\mathbf{x} - \mathbf{y}|/c) .$$

It can then be shown that

$$\frac{1}{c}\frac{\partial f_1}{\partial t} + f_2 = \frac{\pi c}{2} H(t - |\mathbf{x} - \mathbf{y}|/c) . \quad \text{(D5)}$$

Thus it follows from (D1), (D2), and (D5) that as $t \to 0+$,

$$T(\mathbf{x},\mathbf{y},x_3^0;t) \sim -\frac{1}{8\pi c^2(x)} \frac{\partial c}{\partial x_3} H(t - |\mathbf{x} - \mathbf{y}|/c) . \quad \text{(D6)}$$

APPENDIX E: PROOF OF LEMMA 2

Here the auxiliary space will be considered where the velocity is given by $c = (\mathbf{x},\alpha)$ with $\alpha = x_3^0$. Throughout this section the points $x = (\mathbf{x},x_3^0)$, $y = (\mathbf{y},x_3^0)$, and $z = (\mathbf{z},x_3^0)$ with all lie on the same plane. Restrictions will be placed on the time t, and points x and y (being not too far apart) by the condition

$$0 \leqslant \tau(x,y) \leqslant T, \quad 0 \leqslant t \leqslant T . \quad \text{(E1)}$$

This implies then for the time interval under consideration that there is only one ray path from the point x to y and vice versa. Thus $\tau(x,y)$ is the time of arrival of a single ray (lying in the x_3^0 plane) going from y to x.

Define: $\mathscr{C}_0$ is the path of the ray from point y to x. Set

$$\phi(x,y,z) = \tau(x,z) + \tau(z,y) . \quad \text{(E2)}$$

It then follows that if $z \in \mathscr{C}_0$,

$$\phi(x,z,y) = \tau(x,y) \quad \text{(E3)}$$

and if $z \notin \mathscr{C}_0$,

$$\phi(x,y,z) > \tau(x,y) , \quad \text{(E3')}$$

since $\mathscr{C}_0$ is the path of minimum travel time.

For the lemma that follows it will be convenient to define n as the unit normal to the ray-path $\mathscr{C}_0$ at the point $z \in \mathscr{C}_0$ (note that n lies in the plane $z_3 = x_3^0$). Then for fixed x,y define

$$\psi(x,y,z) = (n \cdot \nabla_z)^2[\tau(x,z) + \tau(z,y)] , \quad z \in \mathscr{C}_0 . \quad \text{(E4)}$$

Because of the symmetry of τ it follows that

$$\psi(x,y,z) = \psi(y,x,z) . \quad \text{(E4')}$$

Setting

$$\epsilon = t - \tau(x,y) \quad \text{(E5)}$$

we will now prove the following lemma.

Lemma 2: Let x and y be fixed points such that $x \neq y$ and condition (E1) is satisfied. Let $g(x,y,z)$ satisfy the boundedness conditions

$$|x - z|\, |g(x,y,z)| < M_1 , \quad |y - z|\, |g(x,y,z)| < M_2 ,$$

then the integral

$$\iint_{\mathbf{R}^2} H(t - \phi(x,y,z))(t - \phi(x,y,z))^{\nu} g(x,y,z)\, dz_1\, dz_2 , \quad \text{(E6)}$$

where $0 \leqslant \nu \leqslant 1$, vanishes if $t < \tau(x,y)$, otherwise it has the asymptotic behavior as $t \to \tau(x,y) +$ given by

$$\epsilon^{\nu + 1/2} \kappa_{\nu} \int_{\mathscr{C}_0} \frac{g(x,y,z(s))}{|\psi(x,y,z(s))|^{1/2}}\, ds , \quad \text{(E7)}$$

where

$$\kappa_{\nu} = 2\sqrt{2} \int_0^1 (1 - u^2)^{\nu}\, du \quad \text{(E8)}$$

and s is the arc length along the ray path $\mathscr{C}_0$.

Proof: For time $t < \tau(x,y)$ it follows from Eq. (E3') that $t - \phi(x,y,z) < 0$, hence the integral (E5) vanishes.

For time $t \geqslant \tau(x,y)$, let $\mathscr{A}$ be the region in the $z_3 = x_3^0$ plane given by

$$\phi(x,y,z) < t$$

and denote its boundary by $\mathscr{C}_1$. The region $\mathscr{A}$ contains the curve $\mathscr{C}_0$ and as $t \to \tau(x,y)$, it shrinks to the curve $\mathscr{C}_0$.

Now break up the region $\mathscr{A}$ into three regions $\mathscr{A}_1, \mathscr{A}_2, \mathscr{A}_3$ (see Fig. 3) with $\mathscr{A}_1$ and $\mathscr{A}_3$ being the end regions each containing a length $l\epsilon$ of the arc of the curve $\mathscr{C}_0$ and the points x and y, respectively.

If z is a point on the portion of the boundary $\mathscr{C}_1$ that bounds $\mathscr{A}_2$ such that z is perpendicular to a point z^0 on the curve $\mathscr{C}_0$ in $\mathscr{A}_2$, then z satisfies the relation

$$\tau(x,z) + \tau(z,y) = \phi(x,y,z) = t ,$$
$$\tau(x,z^0) + \tau(z^0,y) + [(z - z^0)^2/2]\psi(x,y,z^0) + \cdots \simeq \tau(x,y) + \epsilon .$$

Using the relation (E3) for the point $z^0 \in \mathscr{C}_0$, we have for a point z on the boundary

$$|z - z^0| \sim (2\epsilon/|\psi(x,y,z^0)|)^{1/2} . \quad \text{(E9)}$$

For a point $z \in \mathscr{A}_2$, we have

$$(t - \phi(x,y,z))^{\nu} \sim (\epsilon - [(z - z^0)^2/2]\psi(x,y,z^0))^{\nu} .$$

Thus we obtain the following:

$$\iint_{\mathscr{A}_2} (t - \phi(x,y,z))^{\nu} g(x,y,z)\, dz_1\, dz_2 \sim \epsilon^{\nu + 1/2} \kappa_{\nu} \int_{\tilde{\mathscr{C}}_0} \frac{g(x,y,z(s))}{|\psi(x,y,z(s))|^{1/2}}\, ds \quad \text{(E10)}$$

on replacing z_0 by $z(s)$. $\tilde{\mathscr{C}}_0$ is the portion of the arc $\mathscr{C}_0$ contained in $\mathscr{A}_2$.

Since the ray path $\mathscr{C}_0$ in the vicinity of the end-point x will be directed radially away from x, it can be shown using

the asymptotic behavior of $\tau(x,z)$ as $z(s)\to x$ [given by Eq. (4)] that

$$(n\cdot\nabla_z)^2\tau(x,z)\sim 1/[c(x)|x-z|] \text{ as } z(s)\to x .$$

The other term $(n\cdot\nabla)^2\tau(z,y)$ in Eq. (E4) remains finite as $z(s)\to x$. Thus we have the asymptotic behavior

$$|\psi(x,y,z)|^{-1/2}\sim O(|x-z|^{1/2}) \text{ as } z(s)\to x .$$

Using the boundedness condition on $g(x,y,z)$ it follows

$$\int_{\mathscr{C}_0'}\frac{g(x,y,z(s))}{|\psi(x,y,z(s))|^{1/2}}\,ds\sim O(\epsilon^{1/2}) , \tag{E11}$$

where $\mathscr{C}_0'$ is the portion of the ray path of $\mathscr{C}_0$ of length $l\epsilon$ contained in the region $\mathscr{A}_1$, and with the end-point x. A similar result can be obtained for the similar integral over the ray-path $\mathscr{C}_0''$, the portion of $\mathscr{C}_0$ of length $l\epsilon$ in region $\mathscr{A}_3$ with end point at $z=y$. Thus combining these results with Eq. (E10) we obtain

$$\iint_{\mathscr{A}_2}(t-\phi(x,y,z))^\nu g(x,y,z)\,dz_1\,dz_2$$
$$\sim\epsilon^{\nu+1/2}\kappa_\nu\int_{\mathscr{C}_0}\frac{g(x,y,z(s))}{|\psi(x,y,z(s))|^{1/2}}\,ds . \tag{E12}$$

To complete the analysis we need to show that the integrals over $\mathscr{A}_1$ and $\mathscr{A}_3$ are higher order and can be neglected. We will show this for the integral over $\mathscr{A}_1$ only, since the other one can be treated in a similar manner.

For a point z on the curve $\mathscr{C}_1$ bounding $\mathscr{A}_1$ we have

$$t=\tau(x,z)+\tau(z,y)$$
$$\sim|x-z|/c(x)+\tau(x,y)-(z-x)\cdot\hat{s}/c(x) ,$$

where $\hat{s}$ is the unit tangent vector to the curve $\mathscr{C}_0$ (directed away from x). Thus we have

$$c(x)\epsilon\sim|x-z|[1-(z-x)\cdot\hat{s}/(|z-x|)] , \tag{E13}$$

which indicates that the region $\mathscr{A}_1$ can be approximated by a parabolic cap as indicated by Eq. (E13). We will use local polar coordinates (r,θ) centered at $z=x$, with $|z-x|=r$, and oriented so that $\theta=0$ corresponds to the direction $\hat{s}$. Thus the integral $\mathscr{A}_1$ has the asymptotic form

$$\left|\iint_{\mathscr{A}_1}(t-\phi(x,y,z))^\nu g(x,y,z)\,dz_1\,dz_2\right|$$
$$\leqslant M_1\iint_{\mathscr{A}_1}\left[\epsilon-\frac{r}{c}(1-\cos\theta)\right]^\nu dr\,d\theta , \tag{E14}$$

where $\mathscr{A}_1$ is given asymptotically as $\epsilon\to 0$ by

$$0\leqslant r\leqslant l\epsilon\sec\theta , \quad -\theta_0\leqslant\theta\leqslant\theta_0 ,$$
$$0\leqslant r\leqslant\epsilon c(x)/(1-\cos\theta) , \quad \text{otherwise}$$

with θ_0 defined by

$$\sec\theta_0(1-\cos\theta_0)=c(x)/l .$$

It follows that the integral (E14) is $O(\epsilon^{\nu+1})$ and hence can be neglected in comparison to the integral over $\mathscr{A}_2$.

[1] J. P. Corones, M. E. Davison, and R. J. Krueger, "Wave splittings, invariant imbedding and inverse scattering," Proceedings of the SPIE, 1983, in *Inverse Optics*, edited by A. J. Devaney (SPIE, Bellingham, WA, 1983).
[2] J. P. Corones, R. J. Krueger, and V. H. Weston, "Some recent results in inverse scattering theory," in *Inverse Problems of Acoustic and Elastic Waves*, edited by F. Santosa, Y. Poa, W. W. Symes, and C. Holland (SIAM, Philadelphia, 1984), pp. 65–81.
[3] R. Bellman and G. N. Wing, *An Introduction to Invariant Imbedding* (Wiley, New York, 1975).
[4] R. Redheffer, "On the relation of transmission-line theory to scattering and transfer," J. Math. Phys. (Cambridge, Mass) **41**, 1 (1962).
[5] J. Corones and R. J. Krueger, "Obtaining scattering kernels using invariant imbedding," J. Math. Anal. Appl. **95**, 393 (1983).
[6] J. P. Corones, M. E. Davison, and R. J. Krueger, "Direct and inverse scattering in the time domain by invariant imbedding techniques," J. Acoust. Soc. Am. **74**, 1535 (1983).
[7] J. P. Corones and R. J. Krueger, "Higher-order parabolic approximations to time-independent wave equations," J. Math. Phys. **24**, 2301 (1983).
[8] L. Fishman and J. J. McCoy, "Derivation and application of extended wave parabolic wave theories, I: The factorized Helmholtz equation," J. Math. Phys. **25**, 285 (1984).
[9] A. E. Yagle and B. L. Levy, "Layer stripping solutions of multidimensional inverse scattering problems," J. Math. Phys. **27**, 1701 (1986).
[10] V. H. Weston, "Factorization of the wave equation in higher dimensions," J. Math. Phys. **28**, 1061 (1987).
[11] V. H. Weston, "Factorization of the dissipative wave equations and inverse scattering," J. Math. Phys. **29**, 2205 (1988).
[12] V. H. Weston, "Factorization of the wave equation in a nonplanar stratified medium," J. Math. Phys. **29**, 36 (1988).
[13] K. Kreider, "Wave splitting approach to time-dependent inverse scattering for the stratified cylinder," SIAM J. App. Math. **49**, 932 (1989).
[14] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1966), Vol. 2.
[15] F. G. Friedlander, *Sound Pulses* (Cambridge U.P., Cambridge, 1958).
[16] C. O. Bloom and N. D. Kazarinoff, *Short Wave Radiation Problems in Inhomogeneous Media: Asymptotic Solutions* (Springer, New York, 1976).
[17] J. Wloka, *Partial Differential Equations* (Cambridge U.P., Cambridge, 1987).
[18] O. A. Ladyzhenskaya, *The Boundary Value Problems of Mathematical Physics* (Springer, New York, 1985).
[19] V. H. Weston, "On the square-root of the wave operator," submitted for publication.
[20] D. V. Widder, *The Laplace Transform* (Princeton U.P., Princeton, NJ, 1946).
[21] L. Fishman, "Phase Space Factorization and Functional Integral Methods in Direct and Inverse Scattering-Symbol Analysis," in *Wave Phenomena: Theoretical, Computational and Practical Aspects*, edited by H. Morris and L. Lam (Springer, New York, 1989), p. 76.
[22] L. Rall, *Computational Solution of Non-Linear Operator Equations* (Krieger, Huntington, NY, 1979).
[23] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).

One more theorem on the short-time regeneration rate

Pavel Exner
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, USSR and Nuclear Physics Institute, Czechoslovak Academy of Sciences, 25068 Řež near Prague, Czechoslovakia

(Received 28 December 1988; accepted for publication 7 June 1989)

Consider an unstable quantum system that has been found undecayed at an instant s and denote by $R(t,s)$ the rate of its regeneration into an original (undecayed) state at a later instant $t+s$. It is proved that the reduced evolution is a semigroup, i.e., there is no regeneration at all, provided $R(t,s)^{1/2}$ can be estimated by a sufficiently regular function that is nondecreasing in s and has zero derivative with respect to t at $t=0$ for every s. This generalizes the theorem of Misra and Sinha [Helv. Phys. Acta **45**, 619 (1972)] in a different direction than in a recent paper by Nishioka [J. Math. Phys. **29**, 1860 (1988)].

I. INTRODUCTION

The semigroup property of reduced evolution represents a useful tool in the quantum theory of unstable systems.[1] It is well known that it cannot hold exactly, because otherwise the corresponding total Hamiltonian H should contain the whole real axis in its spectrum.[1-4] Hence various estimates of its violation become important.

It is further known that validity of the semigroup condition is equivalent to the absence of the decayed-state regeneration: the reduced evolution operator $V_t := E_u U_t E_u$ fulfills

$$V_{t+s} - V_t V_s = E_u U_t E_d U_s E_u, \tag{1.1}$$

for any $t,s \geqslant 0$, where E_u and E_d are the projections to the state subspace of the unstable system and its orthogonal complement, respectively, and $U_t = e^{-iHt}$ is the total evolution operator (we employ the notation used in Ref. 1). The regeneration rate as a function of t and s is subjected to various restrictions. For example, Sinha has demonstrated[4] that regeneration cannot cease after a finite time: if there is a non-negative T_r such that

$$E_u U_t E_d U_s E_u = 0, \tag{1.2}$$

for all $t \geqslant 0$ and $s \geqslant T_r$, then $\sigma(H) = \mathbb{R}$.

Another restriction concerns the regeneration rate at short times, which must not be too slow unless the reduced evolution is an exact semigroup. Misra and Sinha have proved[5] the following assertion: suppose that for every ψ of some dense set D in the subspace $\mathscr{H}_u \equiv E_u \mathscr{H}$ of the unstable system[6] there is a non-negative C_ψ such that

$$\|(V_t V_s - V_{t+s})\psi\| \leqslant C_\psi t^\alpha s^\alpha, \tag{1.3}$$

with some $\alpha > 1$, holds for all $t,s \geqslant 0$, then $\{V_t : t \geqslant 0\}$ is a strongly continuous contractive semigroup, $V_t V_s = V_{t+s}$ for all $t,s \geqslant 0$.

This result has been recently generalized by Nishioka,[7] who has shown that the conclusion is preserved if one replaces the bound (1.3) by

$$\|(V_t V_s - V_{t+s})\psi\| \leqslant C_\psi t^\alpha s^\alpha (t+s)^\beta, \tag{1.4}$$

for some $\alpha > 1$, $2\alpha + \beta > 0$, and $\alpha + \beta + 1 \neq 0$. The aim of the present paper is to derive another extension of the Misra–Sinha theorem.

II. THE MAIN RESULT

We are going to prove the following assertion.

Theorem: Let $\{F(t): t \geqslant 0\}$ be a weakly continuous contractive family with $F(0) = I$ on a Hilbert space $\mathscr{K}$. Suppose there is a dense set D in $\mathscr{K}$ and a function g: $\{(t,s): 0 \leqslant t \leqslant s\} \to \mathbb{R}_+$ with the following properties: (i) $g(0,s) = 0$; (ii) there is a positive t_0 such that $g(t,s_1) \leqslant g(t,s_2)$, for a fixed $t \leqslant t_0$ and all $s_1 \leqslant s_2$; (iii) there is a function $G \in L_{\mathrm{loc}}(\mathbb{R}_+)$ such that $|t^{-1} g(t,s)| \leqslant G(s)$ holds for all sufficiently small t; and (iv) the one-sided derivative $h(s) := \partial g(t,s)/\partial t\,|_{t=0+}$ exists and equals zero for all $s > 0$; such that

$$\|[F(t)F(s) - F(t+s)]\psi\| \leqslant C_\psi g(t,s) \tag{2.1}$$

holds for every $\psi \in D$ and all $s \geqslant t \geqslant 0$. Then $\{F(t): t \geqslant 0\}$ is a strongly continuous contractive semigroup on $\mathscr{K}$.

Proof: The proof follows the same line as in Refs. 5 and 7. We take a sequence $\{\tau_i\}_{i=1}^n$ of positive numbers. Using the condition (2.1) repeatedly in combination with the triangle inequality and contractivity of the family $\{F(t): t \geqslant 0\}$, we get the estimate

$$\left\| F\left(\sum_{i=1}^n \tau_i\right)\psi - F(\tau_1)\cdots F(\tau_n)\psi \right\| \leqslant C_\psi \sum_{j=2}^n g\left(\tau_{j-1}, \sum_{i=j}^n \tau_i\right).$$

Substituting $\tau_i = t/n$, we obtain

$$\left\| F(t)\psi - F\left(\frac{t}{n}\right)^n \psi \right\| \leqslant C_\psi \sum_{k=1}^{n-1} g\left(\frac{t}{n}, \frac{t}{n}k\right).$$

For a given t, we choose n so large that $t/n \leqslant t_0$; using then the assumption (ii), we can estimate the rhs as follows:

$$\left\| F(t)\psi - F\left(\frac{t}{n}\right)^n \psi \right\| \leqslant C_\psi \frac{n}{t} \int_{t/n}^t g\left(\frac{t}{n}, s\right) ds.$$

Next we employ the assumptions (i) and (iii); the last one allows us to use the dominated convergence theorem, which yields

$$\lim_{n\to\infty} \int_{t/n}^t \frac{n}{t} g\left(\frac{t}{n}, s\right) ds = \lim_{n\to\infty} \int_0^t \frac{g(t/n, s) - g(0,s)}{t/n} \chi_{[t/n,\, t]}(s)\, ds = \int_0^t h(s)\, ds.$$

 0022-2488/89/112563-02$02.50

In view of (iv), we get finally

$$\lim_{n\to\infty} F(t/n)^n\psi = F(t)\psi, \tag{2.2}$$

for all $\psi\in D$, and, since the family $\{F(t): t\geqslant 0\}$ is uniformly bounded, this conclusion extends to all $\psi\in\mathscr{H}$. The relation (2.2) easily yields the semigroup property[1,8] and the weak continuity implies the strong one.[9]

For the regeneration rate $R_\psi(t,s) := \|(V_tV_s - V_{t+s})\psi\|^2$, which means the probability that the unstable system starting at $t=0$ in the state ψ and found decayed at s will be found undecayed again at a later instant $t+s$, we get the following.

Corollary: If there is a dense set $D\subset\mathscr{H}_u$ and a function g with the properties listed in the theorem such that

$$R_\psi(t,s)^{1/2}\leqslant C_\psi g(t,s) \tag{2.3}$$

holds for every $\psi\in D$ and all $s\geqslant t\geqslant 0$, then the reduced evolution $\{V_t: t\geqslant 0\}$ on $\mathscr{H}_u$ is a strongly continuous semigroup.

III. CONCLUDING REMARKS

Let us notice first that physically it is difficult to observe the regeneration, in particular, at short times. The reason is the same as in the case of short-time violations of the decay-law exponentiality[1,10]: the dynamics of the known decay processes is such that the interesting time region is inaccessible experimentally. Nevertheless, one cannot exclude discovery of other unstable systems (particles, nuclei, etc.) for which the semigroup approximation will not work as well, and, furthermore, the Misra–Sinha theorem and its generalizations represent interesting mathematical results.

Let us turn now to discussion of our hypotheses. First of all, the regeneration rate need not be estimated symmetrically in t,s; in fact, one has to know the function g in an octant of the (t,s) plane only. The assumption (i) is a weak one; it should be fulfilled for every reasonable estimate. As for (ii), we shall comment on it a little later, while (iii) represents a not very strong regularity requirement. The assumption (iv) is essential; it shows that *regeneration is excluded if only it starts at every instant s slowly enough.*

Our theorem generalizes the Misra–Sinha theorem; one can check easily that the bound (1.3) fulfills the hypotheses. Let us compare it further to Nishioka's result. The function g: $g(t,s) = t^\alpha s^\alpha (t+s)^\beta$ with the above indicated values of α,β fulfills the assumptions (i), (ii), and (iv), while (ii) is valid for $\alpha+\beta\geqslant 0$ only. Let us concentrate on the interesting case $\beta<0$. It was shown in Ref. 7 that the original Misra–Sinha theorem combined with the Schwarz inequality ensures the semigroup property for $2\alpha+\beta>2$. Our theorem yields a stronger result for $\beta\geqslant -2$. This case is covered by the Nishioka theorem as well as the region $\alpha>1$, $-\beta/2<\alpha<\beta$ (with exception of the half-line $\alpha+\beta+1=0$). In the last named case, however, the obtained sufficient condition represents a much weaker assertion: while for $\alpha+\beta>0$ the estimate (1.4) is a restriction actually for short times only due to the contractivity; in the other case one must check it for all times, which is considerably more difficult. Needless to say, the sufficient condition (2.1) covers a broader class of estimates than those of Refs. 5 and 7.

[1] A detailed exposition of the quantum theory of decay can be found in P. Exner, *Open Quantum Systems and Feynman Integrals* (Reidel, Dordrecht, 1985), Chaps. 1–4.

[2] D. N. Williams, Commun. Math. Phys. **21**, 314 (1971).

[3] L. P. Horwitz, J.-P. Marchand, and J. LaVita, J. Math. Phys. **12**, 2537 (1971).

[4] K. Sinha, Helv. Phys. Acta **45**, 619 (1972).

[5] B. Misra and K. Sinha, Helv. Phys. Acta **50**, 99 (1977).

[6] One can demand, of course, the condition (1.3) to be valid for a dense set in the total Hilbert space $\mathscr{H}$ as in Ref. 7. It is clear, however, that the bound in $\mathscr{H}_u^\perp$ is trivial, and, moreover, that $\{V_t: t\geqslant 0\}$ is a semigroup (fulfilling s-$\lim_{t\to 0+} V_t = I_u$) on the subspace $\mathscr{H}_u$ only.

[7] M. Nishioka, J. Math. Phys. **29**, 1860 (1988).

[8] P. Chernoff, *Product Formulas, Nonlinear Semigroups and Addition of Unbounded Operators*, Memoirs of the American Mathematical Society, No. 140 (Am. Math. Soc., Providence, RI, 1974).

[9] K. Yosida, *Functional Analysis* (Springer, Berlin, 1966), Theorem IX.1.1.

[10] J. Dolejší and P. Exner, Czech. J. Phys. B **27**, 855 (1977).

Distributions of $\hbar$-positive type and applications

Francis J. Narcowich
Department of Mathematics, Texas A&M University, College Station, Texas 77843-3368

(Received 4 February 1987; accepted for publication 28 June 1989)

States in classical mechanics are probability measures, and their Fourier transforms are continuous functions of positive type. States in the phase-space formulation of quantum mechanics are Wigner distribution functions, and their (symplectic) Fourier transforms have been characterized by Kastler [Commun. Math. Phys. **1**, 14 (1965)] and Loupias and Miracle-Sole [Commun. Math. Phys. **2**, 31 (1966); Ann. Inst. H. Poincaré **6**, 39 (1967)] as being continuous functions of "$\hbar$-positive type." In this paper, (Schwartz) distributions of $\hbar$-positive type, are defined and studied. It is shown that if such distributions are bounded on a certain sequence of test functions, then they are symplectic Fourier transforms of Wigner distribution functions. These results, are applied to a variety of problems ranging from ones involving the quantum Liouville equation to a problem in signal analysis.

I. INTRODUCTION

Some time ago, Kastler, Loupias, and Miracle-Sole [1-3] characterized the symplectic Fourier transforms of those phase-space functions that are Wigner distribution functions (WDFs)—i.e., that represent states in the Wigner–Weyl formulation of quantum mechanics.[4] Their characterization employs a class of functions similar to Bochner's functions of positive type. Indeed, Kastler termed this class the class of functions of "$\hbar$-positive type."

To describe such functions and to explain what the three workers did, we need to introduce some notation and a few definitions. For a quantum system with n spinless degrees of freedom, the phase space is $\mathbf{R}^n \times \mathbf{R}^n \approx \mathbf{R}^{2n}$; we will denote this space by Γ, and we will denote a point in Γ by $z = (q_1, q_2, \ldots, q_n; p_1, p_2, \ldots, p_n)$. The units of both the p's and q's are to be those of $\sqrt{\hbar}$. We will take the usual symplectic form on phase space to be

$$\sigma(z,z') = \sum_{j=1}^{n} q_j' p_j - p_j p_j' = z'^{T} J z,$$

$$\text{where } J = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}. \tag{1.1}$$

Here, I_n and 0_n are the $n \times n$ identity and zero matrices, respectively. When matrix notation is used, z is to be thought of as a column vector. Finally, the superscript T indicates the transpose of a matrix or vector.

If f is in $\mathcal{S}$, Schwartz space, then the *symplectic Fourier transform* of f is

$$\tilde{f}(a) \overset{\text{def}}{=} \int_{\Gamma} f(z) e^{i\sigma(a,z)}\, dz, \quad \text{where } dz = \prod_{j=1}^{n} dq_j\, dp_j. \tag{1.2}$$

Here $a = (u_1, u_2, \ldots, u_n; v_1, v_2, \ldots, v_n)$ is a point in Γ', the dual of phase space. The measure dz is the standard Liouville measure on Γ. We will also denote the Liouville measure on Γ' by da. The units of the components of a are those of $1/\sqrt{\hbar}$. The symplectic Fourier transform of f is directly related to the ordinary Fourier transform of f via $\tilde{f}(a) = \mathcal{F}(Ja)$. Using this relationship, one can invert (1.2) to get

$$f(z) = (2\pi)^{-2n} \int_{\Gamma'} \tilde{f}(a) e^{i\sigma(z,a)}\, da. \tag{1.3}$$

The *Weyl correspondence* is the famous link between doing quantum mechanics on phase space (or on its dual) and doing quantum mechanics in the usual way, with wave functions and operators on a complex Hilbert space. One may find in Ref. 5 a brief review of the Weyl correspondence, along with more references. In Secs. IV and V, we have a few more words to say about it. For present purposes, it suffices to point out that under the Weyl correspondence the operator product goes over to the *twisted convolution* on the dual of phase space, and to the *twisted product* on phase space.

Let S and T be defined, continuous, and compactly supported in Γ'. The *twisted convolution* [1] of S with T is defined by

$$S \times T(a) \overset{\text{def}}{=} (2\pi)^{-2n} \int_{\Gamma'} S(b) T(a-b) \times e^{i(\hbar/2)\sigma(b,a)}\, db. \tag{1.4}$$

Of course, one may define the twisted convolution of functions, or even distributions (see Sec. II). Here, we merely wish to point out that when S and T are in L^2, or be a Schwartz functions, or are C^∞ with compact support, then $S \times T$ will be L^2, or be a Schwartz function, or be C^∞ with compact support. Also, when $\hbar = 0$, twisted convolution reduces to ordinary convolution.

If f and g belong to $\mathcal{S}$, then the *twisted product* is given by

$$f \circ g \overset{\text{def}}{=} (\pi\hbar)^{-2n} \int_{\Gamma}\int_{\Gamma} f(z+z_1) g(z+z_2) \times e^{-i(\hbar/2)\sigma(z_1,z_2)}\, dz_1\, dz_2. \tag{1.4'}$$

Like the twisted convolution, the twisted product of functions in $\mathcal{S}$ or L^2 is still in $\mathcal{S}$ or L^2. The twisted product of functions in $\mathcal{D}$ is not, however, a function in $\mathcal{D}$. One may define the twisted product of tempered distributions, too.

States in the usual formulation of quantum mechanics are *density matrices*, i.e., trace-class operators that are non-negative and have unit trace. Such operators are well known

 0022-2488/89/112565-09$02.50

to be of the form $\hat{f} = \hat{g}^\dagger \hat{g}$, where $\hat{g}$ is a Hilbert–Schmidt operator. Since Hilbert–Schmidt operators correspond to square-integrable functions on Γ, density matrices correspond to phase-space functions of the form $f = \bar{g} \circ g(z)$, with $g \in L^2(\Gamma)$. These are, apart from normalization, the WDFs. Either by directly dealing with the Weyl correspondence, which also associates Hilbert–Schmidt operators with functions in $L^2(\Gamma')$, or by taking symplectic Fourier transforms, one sees that, again apart from normalization, a function $F = \tilde{f}$ on Γ' corresponds to a state if and only if, for some $G = \tilde{g} \in L^2(\Gamma')$,

$$F = G^\dagger \times G, \tag{1.5}$$

where, for functions on Γ', $G^\dagger(a) \equiv \bar{G}(-a)$. (Upper case letters will be used to denote functions on Γ', and lower case ones for functions on Γ. For a function like G that is the symplectic Fourier transform of a phase-space function g, we also use $\tilde{g}$ for G when it is convenient.)

What Kastler, Loupias, and Miracle-Sole did was to prove that a function F is of the form (1.5) if and only if it is a continuous function defined on Γ' such that for every finite set of points $\{a_1,...,a_m\} \subset \Gamma'$, and for every set of m complex numbers $\{\lambda_1,...,\lambda_m\}$,

$$\sum_{j,k=1}^{m} F(a_j - a_k) e^{i(\hbar/2)\sigma(a_k,a_j)} \bar{\lambda}_j \lambda_k \geqslant 0. \tag{1.6}$$

A continuous function satisfying (1.6) is said to be of *$\hbar$-positive type* (see Ref. 1, p.26). Since it is well known (see Sec. II) that functions of the form (1.5) are not only continuous, but are in L^2 and vanish at infinity, this result serves as a kind of "regularity theorem."

It also serves to clarify the connection between states in classical mechanics and states in quantum mechanics. When $\hbar = 0$ in (1.6), the inequality there reduces to the condition for a function to be of (*Bochner*) *positive type*. Such functions are, by the celebrated Bochner theorem, Fourier transforms of non-negative, finite measures on phase space. Since Fourier transforms and symplectic Fourier transforms differ in a way that is immaterial to Bochner's theorem, one can also say that positive-type functions are *symplectic* Fourier transforms of such measures. Since non-negative, finite measures on phase space are (unnormalized) classical states, the effect of letting $\hbar \to 0$ in (1.6) is to go from a characterization of *quantum mechanical states* to a characterization of *classical states*.

The purpose of this paper is to extend the result of Kastler, Loupias, and Miracle-Sole [1–3] to distributions. To do this, we first note that the condition in (1.6) is equivalent to

$$\int_{\Gamma'}\int_{\Gamma'} F(a_2 - a_1) e^{i(\hbar/2)\sigma(a_1,a_2)} \bar{\lambda}(a_2) \lambda(a_1) da_1\, da_2 \geqslant 0 \tag{1.7}$$

holding for every $\lambda(a)$ that is continuous and compactly supported on Γ'. In (1.7), set $a = a_2 - a_1$ and $b = a_2$, then use (1.4) to put (1.7) in the form

$$\int_{\Gamma'} F(a)\, \overline{\lambda^\dagger \times \lambda}\,(a) da \geqslant 0. \tag{1.8}$$

The advantage of using (1.8) rather than (1.5) is that it allows one to define *distributions* of $\hbar$-positive type, in addition to continuous functions of $\hbar$-positive type. Let $\mathscr{D}$ be the set of all compactly supported, C^∞ functions defined on $\mathbf{R}^n \times \mathbf{R}^n$, and let $\mathscr{D}$ denote the corresponding set of distributions. If $F \in \mathscr{D}'$, then we will say that F is of *$\hbar$-positive type* if (1.8) holds for every possible $\lambda \in \mathscr{D}$. When this happens, we will write $F \geqslant_\hbar 0$. Of course, if F happens to be a *tempered* distribution (i.e., $F \in \mathscr{S}'$) that is of $\hbar$-positive type when viewed as a distribution, then it is easy to see that (1.8) will actually hold for all $\lambda \in \mathscr{S}$. Under the Weyl correspondence, distributions of $\hbar$-positive type correspond to nonnegative operators (see Sec. IV). We remark that every continuous function of $\hbar$-positive type is also a distribution of $\hbar$-positive type.

Although it is not obvious, the key factor in forcing the $\hbar$-positive-type function F to have the form (1.5) is not so much F's continuity as the fact that F is *bounded* near the origin. Indeed, the main result (Theorem 4.1) of this paper is that every $\hbar$-positive-type distribution F that is "finite at the origin," in the sense that there is a "δ-function" sequence $\{D_k\}$ (see Sec. III) of compactly supported C^∞ functions for which

$$\limsup_{k\to\infty} \int_{\Gamma'} F(a)\, \overline{D_k^\dagger \times D_k}\,(a) da < \infty, \tag{1.9}$$

will be of the form (1.5). It even turns out that F being finite at the origin is a condition that can be replaced by F being represented by a function in L^∞_{loc} near the origin (Proposition 4.3), verifying the importance of *boundedness* versus *continuity*.

We originally obtained our main result to help solve a moment problem [4,5] that arises in connection with phase space formulation of quantum mechanics. In the solution of that problem, it plays a role that is analogous to the one played by the *Riesz representation theorem* in the solution of the classical moment problem. (For the readers of Ref. 5, the theorems from a preprint version of this paper are here labeled *3.3* and *4.1* instead of *3.1* and *5.2*.)

Another application of our results is to getting a "regularity theorem" for distribution solutions to the quantum Liouville equation. We briefly sketch out such a theorem in Sec. VI. This section, the last in the paper, also includes a discussion of a few potential applications to other problems involving the quantum Liouville equation and to problems in areas other than quantum mechanics; indeed, it concludes with an application to signal analysis.

The middle sections of the paper are organized this way. In Sec. II, we briefly discuss propositions related to the twisted convolution and to the twisted convolution of distributions. In Sec. III, we prove a sequence of results showing that an $\hbar$-positive-type distribution that is finite at the origin is in $L^2 \cap L^\infty$. In Sec. IV, we use the results of Sec. III and the Schrödinger representation to prove our main result, from which we draw off the corollary about boundedness mentioned earlier.

In Sec. V, we carry over our results to the *twisted product* formalism. In doing so, we define an *$\hbar$-positive tempered distribution* as the symplectic Fourier transform of an $\hbar$-positive type tempered distribution. (The relationship between the two kinds of tempered distributions is complete-

ly analogous to that between *positive* functions and functions of *positive type*.) This section also contains a few more words on the Weyl correspondence.

In closing our introductory section, we want to discuss notation that is used throughout the paper. We have already introduced the spaces $\mathscr{S}$, $\mathscr{S}'$, $\mathscr{D}$, and $\mathscr{D}'$. To avoid introducing numerous factors of $(2\pi)^n$, we define the norm and inner product on $L^2(\Gamma')$ with respect to the measure $(2\pi)^{-2n}\,da$ instead of da, and we will denote them by $\|\cdot\|_2$ and $\langle\cdot,\cdot\rangle$, respectively. Indeed, norms on all of the spaces $L^p(\Gamma')$ will be taken with respect to $(2\pi)^{-2n}\,da$; again, these norms will be denoted by $\|\cdot\|_p$. To avoid introducing extra notation, we will denote the action of a distribution T on a function S in $\mathscr{D}$ by $\langle T,\bar{S}\rangle$. A similar convention will apply for $T\in\mathscr{S}'$ and $S\in\mathscr{S}$. Finally, we let $C_0(\Gamma')$ denote the set of all functions that are continuous on Γ' and that vanish at ∞. Of course, the analogous conventions will be applied to objects defined on phase space, Γ.

II. THE TWISTED CONVOLUTION

The twisted convolution plays an important role in our discussion. In this section, we want to collect those properties that we will use later on, beginning with the ones involving L^2 and ending with ones involving distributions. Our first proposition contains a very important norm estimate that is the essential reason $L^2(\Gamma')$ is a Hilbert algebra under twisted convolution.

Proposition 2.1: If S and T are in $L^2(\Gamma')$, then the twisted convolution $S\times T$ is continuous on Γ', vanishes at infinity, and is in $L^2(\Gamma')$; that is, $S\times T\in C_0(\Gamma')\cap L^2(\Gamma')$. In addition, one has this norm estimate:

$$\|S\times T\|_2\leqslant(2\pi\hbar)^{-n/2}\|S\|_2\|T\|_2. \tag{2.1}$$

Proof: Getting $S\times T$ to be in $C_0(\Gamma')$ is a routine exercise in techniques that are standard in harmonic analysis, and so we omit the proof of this well-known fact. Showing that $S\times T$ is in L^2 and that it satisfies (2.1) is somewhat more difficult. See Refs. 2 and 6 for proof. ■

The following proposition is a collection of several well-known, useful formulas; it is an immediate consequence of the previous proposition.

Proposition 2.2: If S,T, and U are in $L^2(\Gamma')$, then we have these:

$$(S\times T)\times U=S\times(T\times U), \tag{2.2}$$

$$(S\times T)^\dagger=T^\dagger\times S^\dagger, \tag{2.3}$$

$$S\times T^\dagger(0)=\langle S,T\rangle, \tag{2.4}$$

$$S^\dagger\times S(0)=S\times S^\dagger(0)=\|S\|_2^2=\|S^\dagger\|_2^2, \tag{2.5}$$

$$S\times T\times U(0)=U\times S\times T(0)=T\times U\times S(0). \tag{2.6}$$

In addition to the elementary L^2 theory outlined above, we will also need to deal with the twisted convolution of L^1 functions.

Proposition 2.3: If S and T are in $L^1(\Gamma')$, then $S\times T$ is also in $L^1(\Gamma')$. Moreover, we have

$$\|S\times T\|_1\leqslant\|S\|_1\|T\|_1. \tag{2.7}$$

Finally, if S and T are as above, and if U is in $L^1(\Gamma')$, then (2.2) and (2.3) again hold.

Proof: The proof is identical to that used in connection with the ordinary convolution product. See Refs. 1 and 7. ■

From time to time, we will need to work with the twisted convolution of a distribution and a function in $\mathscr{D}$. If $S\in\mathscr{D}'$ and $T\in\mathscr{D}$, then, following Maillard, [8] we define $S\times T$ by

$$S\times T(a)\equiv\langle S(b),T^\dagger(b-a)e^{-(i\hbar/2)\sigma(b,a)}\rangle, \tag{2.8}$$

and $T\times S$ by

$$T\times S(a)\equiv\langle T(b),S^\dagger(b-a)e^{+(i\hbar/2)\sigma(b,a)}\rangle. \tag{2.9}$$

As in the case of the ordinary convolution (cf. Ref. 7, Chap. 27), $S\times T$ and $T\times S$ are actually C^∞ functions. We summarize their properties below. We omit the proofs, as they are quite standard.

Proposition 2.4: If one of S,T is in $\mathscr{D}'$ and the other is in $\mathscr{D}$, then $S\times T$ is a C^∞ function. Moreover, (2.3) and (2.4) still hold.

Proposition 2.5: If one of S,T, and U is in $\mathscr{D}'$ and the other two are in $\mathscr{D}$, then (2.2) and (2.6) still hold.

For results concerning twisted convolutions (and twisted products) of *tempered* distributions, we refer the reader to recent work of Gracia-Bondía and Várilly. [9]

III. REDUCTION TO THE $L^2\cap L^\infty$ CASE

In this section, we wish to show that every $\hbar$-positive-type distribution that is finite at the origin is in $L^2\cap L^\infty$. We will begin by setting down a few basic facts about $\hbar$-positive-type distributions, and by giving a precise definition for being "finite at the origin."

Recall that $F\in\mathscr{D}'$ is of $\hbar$-positive type if, for every $\lambda\in\mathscr{D}$,

$$\langle F,\lambda^\dagger\times\lambda\rangle\geqslant 0, \tag{3.1}$$

and that, when this happens, we write $F\geqslant_\hbar 0$. In addition, we say that F is *finite at the origin* if there exists a sequence $\{D_k\}$, $D_k\in\mathscr{D}$, for which these are true:

$$\text{(i) } S\times D_k\overset{k\to\infty}{\to}S \text{ in } \mathscr{D} \text{ for all } S\in\mathscr{D};$$

$$\text{(ii) } \|D_k\|_1\overset{k\to\infty}{\to}1; \tag{3.2}$$

$$\text{(iii) } \limsup_{k\to\infty}\langle F,D_k^\dagger\times D_k\rangle\equiv F^0<\infty.$$

Two remarks: First, such a sequence exists. In Sec. IV, we explicitly construct one in the course of proving Proposition 4.3. Second, it will turn out that F is in L^∞, and that F^0 in (3.2) is just the L^∞ norm of F (cf. Theorem 3.3); hence F^0 is independent of the particular sequence chosen.

Our first proposition is a list of simple, important properties involving distributions of $\hbar$-positive type. In particular, it contains a useful version of the Cauchy–Schwarz inequality.

Proposition 3.1: Let $F\in\mathscr{D}'$, and let S,T be in $\mathscr{D}$. If $F\geqslant_\hbar 0$, then the following are true:

(i) $|\langle F,S^\dagger\times T\rangle|^2\leqslant\langle F,S^\dagger\times S\rangle\langle F,T^\dagger\times T\rangle$.

(ii) If F is in L^2, then (i) holds with $S,T\in L^2$.

(iii) $F=F^\dagger$.

(iv) $S\times F\times S^\dagger\geqslant_\hbar 0$.

Proof: To get (i), let $\lambda=\alpha S+\beta T$ in (3.1); here, α and

β are arbitrary complex numbers. The non-negativity of the quadratic form then gives the desired result. Part (ii) is obvious. Part (iii) also follows from the non-negativity of the quadratic form mentioned above, for its being non-negative implies that

$$\langle F,S^\dagger\times T\rangle = \overline{\langle F,T^\dagger\times S\rangle}.$$

Choosing T to be D_k, where $\{D_k\}$ satisfies (3.2), we see that $S^\dagger\times D_k\to S^\dagger$ in $\mathscr{D}$, and so

$$\langle F,S^\dagger\rangle = \overline{\langle F,S\rangle}$$

for all $S\in\mathscr{D}$. This is, by definition, (iii) in "integral" form.

To get (iv), note that, by Propositions 2.4 and 2.5, we have

$$\begin{aligned}\langle S\times F\times S^\dagger,\lambda\times\lambda^\dagger\rangle &= S\times F\times S^\dagger\times(\lambda\times\lambda^\dagger)^\dagger(0)\\ &= F\times(S^\dagger\times\lambda^\dagger)\times(\lambda\times S)(0)\\ &= \langle F,(\lambda\times S)\times(\lambda\times S)^\dagger\rangle\geqslant 0.\end{aligned}$$

This last term is non-negative because $F\geqslant_\hbar 0$. Hence $S\times F\times S^\dagger$ satisfies (3.1) and so is of $\hbar$-positive type. ■

Our immediate goal is to show that $\hbar$-positive-type distributions that are finite at the origin are actually in L^∞. Once we have accomplished this, we will go on to show that such distributions are in L^2. To get to our immediate goal, we need the inequalities below.

Lemma 3.2: Let $F\in\mathscr{D}'$. Suppose that $F\geqslant_\hbar 0$ and that F is finite at the origin. If $S\in\mathscr{D}$, then

$$|\langle F,S\rangle|^2\leqslant F^0\langle F,S\times S^\dagger\rangle, \tag{3.3}$$

where F^0 is defined in (3.2) (iii). In addition,

$$\langle F,S\times S^\dagger\rangle\leqslant F^0\|S\|_1^2. \tag{3.4}$$

Finally,

$$|\langle F,S\rangle|\leqslant F^0\|S\|_1. \tag{3.5}$$

Proof: Let $\{D_k\}$ be a sequence in $\mathscr{D}$ satisfying (3.2). From Proposition 3.1, part (i), we have

$$|\langle F,S\times D_k\rangle|^2\leqslant\langle F,S\times S^\dagger\rangle\langle F,D_k\times D_k^\dagger\rangle.$$

Letting $k\to\infty$ gives

$$|\langle F,S\rangle|^2\leqslant\langle F,S\times S^\dagger\rangle\limsup_{k\to\infty}\langle F,D_k\times D_k^\dagger\rangle,$$

which immediately implies (3.3).

To get the second inequality, in (3.3) replace S by $S\times S^\dagger$, then rewrite the right side using Propositions 2.4 and 2.5. The result is

$$\langle F,S\times S^\dagger\rangle^2\leqslant F^0\langle S^\dagger\times F\times S,S\times S^\dagger\rangle. \tag{3.6}$$

The distribution $S^\dagger\times F\times S$ is of $\hbar$-positive type by Proposition 3.1, part (iv). By Proposition 2.4, it is also a C^∞ function. It is very easy to show that a C^∞ $\hbar$-positive-type distribution is bounded. Indeed, this is true even if the distribution is merely continuous. One need only apply (1.6) with $m=2$, $a_1=a$, and $a_2=0$ to see that this is the case. Clearly, such a function is bounded by its value at the origin; hence, in our case, we have

$$|S^\dagger\times F\times S(a)|\leqslant S^\dagger\times F\times S(0). \tag{3.7}$$

By applying Propositions 2.4 and 2.5, we may transform the right side of (3.7) so that (3.7) becomes the new inequality

$$|S^\dagger\times F\times S(a)|\leqslant\langle F,S\times S^\dagger\rangle. \tag{3.8}$$

Combining (3.6) and (3.8) yields

$$\langle F,S\times S^\dagger\rangle^2\leqslant F^0\langle F,S\times S^\dagger\rangle\|S^\dagger\times S\|_1.$$

Since, by Proposition 2.3, $\|S^\dagger\times S\|_1\leqslant\|S\|_1^2$, the last inequality implies

$$\langle F,S\times S^\dagger\rangle^2\leqslant F^0\langle F,S\times S^\dagger\rangle\|S\|_1^2.$$

When $\langle F,S\times S^\dagger\rangle=0$, (3.4) is trivially true. When $\langle F,S\times S^\dagger\rangle\neq 0$, canceling the common factor above yields (3.4). Thus (3.4) is true, in general.

To obtain (3.5), use (3.4) to replace the right side of (3.3), and then take the square root of both sides. ■

If $F\in\mathscr{D}'$ is of $\hbar$-positive type and finite at the origin as well, then (3.5) allows us to extend $\langle F,S^\dagger\rangle$ to a bounded linear functional on $L^1(\Gamma')$. Since $L^\infty(\Gamma')$ is dual to $L^1(\Gamma')$, we have that $F\in L^\infty(\Gamma')$. In addition, it is obvious from (3.5) that

$$\|F\|_\infty\leqslant F^0. \tag{3.9}$$

Let us now drop the assumption that F is finite at the origin, and replace it with the assumption that F is in L^∞. Let $\{D_k\}$ be a sequence in $\mathscr{D}$ satisfying (3.2) (i) and (3.2) (ii). From F being $L^\infty(\Gamma')$ and from Proposition 2.3, we have that

$$\langle F,D_k^\dagger\times D_k\rangle\leqslant\|F\|_\infty\|D_k\|_1^2.$$

Letting $k\to\infty$ in this inequality and using (3.2)(ii), we get

$$F^0\equiv\limsup_{k\to\infty}\langle F,D_k^\dagger\times D_k\rangle\leqslant\|F\|_\infty, \tag{3.10}$$

so (3.2) (iii) is satisfied and F is finite at the origin. Combining (3.9) and (3.10) and examining the remarks above, we see that we have proved the following.

Theorem 3.3: Let $F\in\mathscr{D}'$ and suppose that $F\geqslant_\hbar 0$. Then, F is finite at the origin if and only if $F\in L^\infty(\Gamma')$. In addition,

$$F^0=\|F\|_\infty. \tag{3.11}$$

Finally, F^0 is the same for all sequences satisfying (3.2) (i) and (3.2)(ii).

The last theorem is our first "regularity" result. For later use, we will need the corollary below.

Corollary 3.4: If F is a continuous function that is of $\hbar$-positive type, then F is finite at the origin and

$$F^0=F(0)=\|F\|_\infty. \tag{3.12}$$

Moreover, if $G\in L^2(\Gamma')$, then

$$G^\dagger\times G\geqslant_\hbar 0 \text{ and } (G^\dagger\times G)^0=\|G\|_2^2. \tag{3.13}$$

Proof: In the proof of Lemma 3.2, we pointed out that such an F must be bounded, and that $|F(a)|\leqslant F(0)=\|F\|_\infty$. Thus, Theorem 3.3 applies, and (3.12) follows from (3.11).

Consider $G^\dagger\times G$. One can use Proposition 2.2 to show that

$$\langle G^\dagger\times G,S^\dagger\times S\rangle=\|G\times S^\dagger\|_2^2.$$

Thus $G^\dagger\times G\geqslant_\hbar 0$. (Of course, we could have also obtained this result by invoking the result of Kastler, Loupias and Miracle-Sole[1-3] mentioned in the Introduction.) From Proposition 2.1, we have that $G^\dagger\times G$ is continuous, and, since $\hbar$-positive-type distributions which are continuous are

also continuous functions of $\hbar$-positive type, that $G^\dagger \times G$ is an $\hbar$-positive-type continuous function. Thus (3.12) applies. To get (3.13), we need only compute $G^\dagger \times G(0)$. Invoking Proposition 2.2 again, we have that $G^\dagger \times G(0) = \|G\|_2^2$, which yields (3.13). ■

Having shown that every $\hbar$-positive-type distribution that is finite at the origin is in L^∞, we now wish to show that such a distribution is also in L^2. To carry this out, we need the two lemmas below.

Lemma 3.5: Let $F \in \mathscr{D}'$, $F \geqslant_\hbar 0$, and F be finite at the origin. If $\phi(a)$ is a continuous function of (Bochner) positive type, then the distribution ϕF is of $\hbar$-positive type, and is finite at the origin.

Proof: By Theorem 3.3, $F \in L^\infty$. Since, as is well-known, continuous functions of positive type are bounded, ϕ is in L^∞, too. Thus the product ϕF is in L^∞, and so, again by Theorem 3.3, ϕF will be finite at the origin, provided it is of $\hbar$-positive type. Therefore, to prove the lemma, we need only show that ϕF is of $\hbar$-positive type.

Recall that Bochner's theorem gives us that ϕ is the Fourier transform of finite, non-negative measure on $\mathbf{R}^{2n}$. Because of the connection between *symplectic* Fourier transforms and ordinary ones (cf. Sec. I), we may also represent ϕ as the symplectic Fourier transform of a finite, non-negative measure on Γ. If we denote this measure by ν_ϕ, then we have

$$\phi(a) = \int_\Gamma e^{i\sigma(a,z)}\, d\nu_\phi(z). \tag{3.14}$$

If $S \in \mathscr{D}$, then, from (3.14) and an interchange of integrals, we obtain

$$\langle \phi F, S^\dagger \times S \rangle = \int_\Gamma \langle F, e^{i\sigma(z,a)} S^\dagger \times S \rangle d\nu_\phi(z). \tag{3.15}$$

By writing out the expression for $S^\dagger \times S$ and manipulating it, we can show that

$$e^{i\sigma(z,a)} S^\dagger \times S(a) = S_z^\dagger \times S_z(a),$$
$$\text{where } S_z(a) = e^{i\sigma(z,a)} S(a). \tag{3.16}$$

Substituting (3.15) into (3.16) yields

$$\langle \phi F, S^\dagger \times S \rangle = \int_\Gamma \langle F, S_z^\dagger \times S_z \rangle d\nu_\phi(z)\,, \tag{3.17}$$

which is non-negative because F being of $\hbar$-positive type implies the integrand on the left in (3.17) is non-negative. Thus we see that $\phi F \geqslant_\hbar 0$, and so our proof is complete. ■

We remark that this result is known for continuous functions of $\hbar$-positive type [4,10]. With more work, one can show that, if ϕ is C^∞ and $F \geqslant_\hbar 0$ (but not necessarily finite at the origin), then ϕF will also be of $\hbar$-positive type, although not necessarily finite at the origin.

The next lemma relates F^0 and $\|F\|_2$ for an $\hbar$-positive-type distribution that is finite at the origin.

Lemma 3.6: Let F be an $\hbar$-positive-type distribution that is finite at the origin. If F is in $L^2(\Gamma')$, then

$$\|F\|_2 \leqslant (2\pi\hbar)^{-n/2} F^0. \tag{3.18}$$

Proof: From (3.3), (2.1) and the Cauchy–Schwarz inequality for L^2, we have, for every $S \in \mathscr{D}$, that

$$|\langle F, S \rangle|^2 \leqslant (2\pi\hbar)^{-n/2} F^0 \|F\|_2 \|S\|_2^2\,. \tag{3.19}$$

Since $\mathscr{D}$ is dense in $L^2(\Gamma')$, a standard argument implies that (3.19) holds for all $S \in L^2(\Gamma')$. If $\|F\|_2 \neq 0$, then choose $S = F$ in (3.19) and divide by $\|F\|_2$ to get (3.18). If $\|F\|_2 = 0$, then (3.18) is trivially true. ■

We can now prove the main result of this section.

Theorem 3.7: If F is an $\hbar$-positive-type distribution that is finite at the origin, then $F \in L^2 \cap L^\infty$.

Proof: By Theorem 3.3, $F \in L^\infty$. We want to show that $F \in L^2$. For $m = 1,2,...$, the Gaussian $\phi_m(a) \equiv e^{-|a|^2/m}$ is a continuous function of (Bochner) positive type, because it is the Fourier transform of another (positive) Gaussian. By Lemma 3.5, $F_m(a) \equiv \phi_m(a) F(a)$ is of $\hbar$-positive type and is finite at the origin. Obviously, we also have that $F_m \in L^2 \cap L^\infty$. Thus, from Lemma 3.6, Theorem 3.3, and the form of F_m, we obtain the following norm estimate:

$$\|F_m\|_2 \leqslant (2\pi\hbar)^{-n/2} F_m^0 = (2\pi\hbar)^{-n/2} \|F_m\|_\infty$$
$$\leqslant (2\pi\hbar)^{-n/2} \|F\|_\infty. \tag{3.20}$$

We can view the norm estimate in (3.20) as being a uniform bound on the $L^1(\Gamma')$ norms of $|F_m|^2$. Since the sequence $\{|F_m(a)|^2\}$ is also increasing and converges pointwise almost everywhere to $|F(a)|^2$, the Lebesgue dominated convergence theorem applies; hence, we have that $|F|^2 \in L^1(\Gamma')$ or, equivalently, that $G \in L^2(\Gamma')$. ■

We close by remarking that it should be possible to prove Theorem 3.7 by using a harmonic oscillator basis and taking an approach similar to the one used in Ref. 9, Sec. V.

IV. THE MAIN RESULT

In the last section, we showed that every $\hbar$-positive-type distribution that is finite at the origin is in $L^2 \cap L^\infty$. We will now prove the main result of this paper.

Theorem 4.1: If F is an $\hbar$-positive-type distribution that is finite at the origin, then there exists some $G \in L^2$ for which

$$F = G^\dagger \times G \tag{4.1}$$

holds almost everywhere in Γ'. Here G may be chosen to be self-adjoint (i.e. $G^\dagger = G$), or even of $\hbar$-positive type. Conversely, if F is of the form (4.1) for some $G \in L^2$, then F is an $\hbar$-positive-type distribution that is finite at the origin.

Proof: With twisted convolution being the product, $L^2(\Gamma')$ forms a Hilbert algebra that can be faithfully represented by means of what is called the Schrödinger representation (cf. Refs. 1–3), π_ω. The underlying Hilbert space used in connection with this representation is the one that is the state space for a spinless quantum mechanical system having n degrees of freedom. In this representation, functions in L^2 correspond to Hilbert–Schmidt operators. If F_1 and F_2 are two square-integrable functions on Γ', then

$$\langle F_1, F_2 \rangle = (2\pi\hbar)^n \operatorname{Tr}[\pi_\omega(F_1)\pi_\omega(F_2)^\dagger]. \tag{4.2}$$

Using the trace formula above and noting that the $\hbar$-positive-type distribution F is in L^2, one can easily show that the operator corresponding to F, $\hat{F} = \pi_\omega(F)$, is non-negative and Hilbert–Schmidt. Now, suppose that we also have that $\hat{F}$ is trace class. If we let $\hat{G}$ be the positive square root of $\hat{F}$, then $\hat{G}$ is Hilbert–Schmidt and the operator equation $\hat{F} = \hat{G}^2$ translates into the twisted convolution product $F = G \times G$. Since the Schrödinger representation is faithful,

the G for which $\hat{G}=\pi_\omega(G)$ is in $L^2(\Gamma')$ and satisfies $G^\dagger=G$. Moreover, from (4.2) and the non-negativity of $\hat{G}$, one sees that

$$\langle G,S^\dagger\times S\rangle=\mathrm{Tr}[\hat{G}\pi_\omega(S)^\dagger\pi_\omega(S)]\geqslant 0,$$

and so G is an $\hbar$-positive-type distribution. The existence of such a G is precisely what our theorem asserts. Thus, to prove the theorem, we need only prove that $\hat{F}$ is trace class.

The first step in showing this is to derive an inequality involving the trace of the product of $\hat{F}$ with certain trace-class operators. Let $\hat{H}=\frac{1}{2}\sum_{j=1}^n(\hat{p}_j^2+\hat{q}_j^2)$ be the usual harmonic oscillator Hamiltonian, and for $\beta>0$ set

$$\hat{\Omega}_\beta=\exp(-\beta\hat{H}). \tag{4.3}$$

In Ref. 5 [cf. Eq. (2.18)], we showed that the Wigner transform of this operator has the form

$$\omega_\beta(q_1,\ldots,q_n,p_1,\ldots,p_n)=\mathrm{sech}^n\left(\frac{\hbar\beta}{2}\right)\exp\left\{-\tanh\left(\frac{\hbar\beta}{2}\right)\left(\frac{\sum_{j=1}^n(p_j^2+q_j^2)}{\hbar}\right)\right\}. \tag{4.4}$$

Taking the symplectic Fourier transform of the function in (4.4) gives us

$$\tilde{\omega}_\beta(a)\equiv\Omega_\beta(a)\equiv\left[\pi\hbar\,\mathrm{csch}\left(\frac{\hbar\beta}{2}\right)\right]^n\exp\left\{-\left[\frac{\hbar}{4}\coth\left(\frac{\hbar\beta}{2}\right)\right]a^Ta\right\}. \tag{4.5}$$

(Here, a^T is the transpose of a, which is now regarded as a column vector.) With our conventions, Ω_β is precisely $\pi_\omega^{-1}(\hat{\Omega}_\beta)$, and so the trace formula (4.2) yields

$$\langle F,\Omega_\beta\rangle=(2\pi\hbar)^n\,\mathrm{Tr}[\hat{F}\hat{\Omega}_\beta]. \tag{4.6}$$

Using (4.5), (4.6), the definition of $\langle\cdot,\cdot\rangle$, and the fact that $F\in L^\infty$, we get the following:

$$\mathrm{Tr}[\hat{F}\hat{\Omega}_\beta]\leqslant\frac{\|F\|_\infty}{\hbar^n}\mathrm{sech}^n\left(\frac{\hbar\beta}{2}\right)\leqslant\frac{\|F\|_\infty}{\hbar^n}. \tag{4.7}$$

The second step is to compute the trace in (4.7) by using an orthonormal basis of harmonic-oscillator states, $\psi_0,\psi_1,\psi_2,\ldots$. Making the computation and using (4.7), we find

$$\mathrm{Tr}[\hat{F}\hat{\Omega}_\beta]=\sum_{j=0}^\infty e^{-\lambda_j\beta}(\hat{F}\psi_j,\psi_j)\leqslant\frac{\|F\|_\infty}{\hbar^n}, \tag{4.8}$$

where $(\cdot,\cdot)$ is the inner product on the Hilbert space underlying the Schrödinger representation and λ_j is the eigenvalue of $\hat{H}$ corresponding to ψ_j. Because $\hat{F}$ is a non-negative operator, all of the terms in the series from (4.8) are non-negative. If we take the limit $\beta\downarrow 0$, then standard arguments show the resulting series is convergent and satisfies

$$\sum_{j=0}^\infty(\hat{F}\psi_j,\psi_j)\leqslant\frac{\|F\|_\infty}{\hbar^n}. \tag{4.9}$$

The last step is to observe that for a non-negative, bounded operator to be trass class, it is necessary and sufficient that it have a trace relative to *some* orthonormal basis for the underlying Hilbert space. (See Ref. 11, p. 96.) Of course, (4.9) shows that $\hat{F}$ has a finite trace relative to the orthonormal basis of harmonic oscillator states. Hence $\hat{F}$ is trace class.

The converse follows from Corollary 3.4. ■

Let us now draw off a few simple, useful corollaries. Our first corollary concerns a simple observation about the *regularity* of $\hbar$-positive-type distributions that are finite at the origin.

Corollary 4.2: If F is an $\hbar$-positive-type distribution that is finite at the origin, then $F\in C_0(\Gamma')\cap L^2(\Gamma')$.

Proof: This is an immediate consequence of Proposition 2.1 and Theorem 4.1. ■

The next result gives a useful condition that is equivalent to an $\hbar$-positive-type distribution being finite at the origin; its proof contains an example of the "δ-function sequence" described in Sec. III.

Proposition 4.3: Let F be an $\hbar$-positive-type distribution. Then, F is in L^∞_{loc} on some neighborhood $\mathcal{N}$ of the origin if and only if F is finite at the origin.

Proof: Suppose that F is in L^∞_{loc} on some neighborhood $\mathcal{N}$ of the origin. For $k=1,2,3,\ldots$, let $G_k(a)=(4k\pi)^n\times\exp(-ka^Ta)$ and, for $r>0$, let $\phi_r(x)$ be any C^∞-function such that $0\leqslant\phi_r(x)\leqslant 1$ and that

$$\phi_r(x)=\begin{cases}1, & \text{if } |x|\leqslant r/2;\\ 0, & \text{if } |x|\geqslant r.\end{cases}$$

We now fix r to be so small that $\{a\in\Gamma':|a|\leqslant 2r\}\subset\mathcal{N}$, and we define

$$D_k(a)\overset{\text{def}}{=}\phi_r(|a|)G_k(a).$$

Clearly, the D_k's form a sequence of C^∞ functions supported on $|a|\leqslant r$. It is easy to check that this sequence also satisfies the conditions (3.2)(i) and (3.2)(ii). To show that F is finite at the origin, we need to show that (3.2)(iii) holds as well. To do this, observe that the support of the twisted convolution $D_k^\dagger\times D_k$ is contained in the ball $|a|\leqslant 2r$; this is a set on which F is in L^∞, since it is a compact subset of $\mathcal{N}$. If $\|F\|_{\infty,r}$ is the L^∞ norm for F on this ball, then by Proposition 2.3 and standard estimates

$$\langle F,D_k^\dagger\times D_k\rangle\leqslant\|F\|_{\infty,r}\|D_k^\dagger\times D_k\|_1\leqslant\|F\|_{\infty,r}\|D_k\|_1^2. \tag{4.10}$$

Taking the limit superior of both sides in (4.10) and using (3.2)(ii), we find

$$\limsup_{k\to\infty}\langle F,D_k^\dagger\times D_k\rangle\leqslant\|F\|_{\infty,r}<\infty,$$

and so F is finite at the origin. The converse follows from Theorem 3.3. ■

As a help in checking whether an $\hbar$-positive-type distribution has the form (4.1), both of the equivalent conditions, being finite at the origin and being in L^∞_{loc} of the origin, are useful. Finiteness at the origin is appropriate for problems in which the $\hbar$-positive-type distribution appears in integral form, and boundedness for problems in which it appears as a function.

Combining Proposition 4.3 with Corollary 4.2 yields another "regularity" result that is analogous to the one in Corollary 4.2, but with finiteness at the origin replaced by local boundedness near the origin.

Corollary 4.4: If F is an $\hbar$-positive-type distribution that

is in L^{∞}_{loc} on some neighborhood $\mathcal{N}$ of the origin, then $F \in C_0(\Gamma') \cap L^2(\Gamma')$.

It is important to note that this corollary states that an $\hbar$-positive-type distribution that is locally bounded near the origin must be continuous on Γ' and fall to 0 as $|a| \to \infty$. It is easy to check that there are $\hbar$-positive-type distributions that are *not* locally bounded near $a = 0$, e.g., the Dirac delta function located at $a = 0$. This raises an important question: Can distributions of $\hbar$-positive type be classified according to growth at the origin? Indeed, from growth estimates near the origin can one tell whether such distributions correspond to operators belonging to various classes—compact, bounded, Hilbert–Schmidt, and so on?

V. THE CONNECTION WITH WIGNER FUNCTIONS

As we mentioned earlier, the functions defined on Γ', the dual of phase space, are related to functions defined on Γ, phase space, by the former being symplectic Fourier transforms of the latter. If $f(z)$ and $\tilde{f}(a)$ are symplectic transforms of each other, and if $\hat{f} = \pi_\omega(\tilde{f})$, then we may think of $\hat{f}$ as being the operator associated with f via the Weyl correspondence, and f the phase space function that is the pseudo-differential operator symbol of $\hat{f}$. (See Ref. 12.) Thus the Weyl correspondence gives us the following associations among the operator product, twisted convolution, and twisted product[2,12–14]:

$$\hat{f}\hat{g} \leftrightarrow \tilde{f} \times \tilde{g} \leftrightarrow f \circ g. \tag{5.1}$$

In Sec. I, we pointed out that the twisted product of two functions in $\mathcal{D}$ does *not* necessarily remain in $\mathcal{D}$. However, like the twisted convolution, the twisted product of two functions in Schwartz space, $\mathcal{S}$, *is* still in $\mathcal{S}$. For that reason, and also because both $\mathcal{S}$ and $\mathcal{S}'$ behave nicely with regard to Fourier transforms, we will confine our discussion of distributions on phase space to the *tempered* ones—i.e., to those in $\mathcal{S}'$.

We will say that a tempered distribution $f \in \mathcal{S}'(\Gamma)$ is *$\hbar$-positive* if, for every $g \in \mathcal{S}$, f satisfies

$$\int_\Gamma f(z)\bar{g} \circ g(z)\,dz \geqslant 0. \tag{5.2}$$

In addition, we will say that an $\hbar$-positive tempered distribution f is *finite on unity* if its symplectic Fourier transform $\tilde{f}$, which is obviously of $\hbar$-positive type, is finite at the origin. This is equivalent to requiring that there be a sequence of Schwartz functions $\{s_k(z)\}$ such that

$$\text{(i)} \quad g \circ s_k \overset{k\to\infty}{\to} g \text{ in } \mathcal{S}, \quad \forall g \in \mathcal{S};$$

$$\text{(ii)} \quad \|\tilde{s}_k\|_1 \overset{k\to\infty}{\to} 1; \tag{5.3}$$

$$\text{(iii)} \quad \limsup_{k\to\infty} \int_\Gamma f \bar{s}_k \circ s_k \, dz < \infty.$$

The conditions listed above are precisely the Fourier-transformed version of the corresponding ones given in Sec. III; we have left the symplectic Fourier transform in place in condition (ii) to avoid introducing an additional function space. We chose the name "finite on unity" because the sequence $\{s_k(z)\}$ tends weakly to 1 as $k \to \infty$. Thus, if the limit superior in (5.3)(iii) exists, we can assign a meaning to $\langle f, 1 \rangle$. Concerning $\hbar$-positive tempered distributions that are finite at unity, we have the following analog of Theorem 4.1.

Theorem 5.1: If f is an $\hbar$-positive tempered distribution that is finite on unity, then there exists some $g \in L^2$ for which

$$f = \bar{g} \circ g \tag{5.4}$$

holds almost everywhere in Γ. This g may be chosen to be real, or even $\hbar$-positive. Conversely, if f has the form (5.4) for some $g \in L^2$, then f is an $\hbar$-positive tempered distribution that is finite on unity.

Proof: The theorem is simply the Fourier-transformed version of Theorem 4.1. ■

We now list corollaries similar to the ones following Theorem 4.1. The first of these is again a "regularity" result.

Corollary 5.2: If f is an $\hbar$-positive tempered distribution that is finite on unity, then $f \in C_0(\Gamma) \cap L^2(\Gamma)$.

Proof: This is an immediate consequence of Theorem 5.1 and the well-known result that phase-space functions of the form (5.4) are in the space $C_0(\Gamma) \cap L^2(\Gamma)$. (See Ref. 10, Theorem 3.5.4.) ■

Our second corollary is one that is pertinent to problems in which the $\hbar$-positive tempered distribution is given as a function that has known integrability properties.

Corollary 5.3: If f is in $L^1(\Gamma)$ and is $\hbar$-positive, then f is finite on unity; it, therefore, has the form (5.4) and is in $C_0(\Gamma) \cap L^2(\Gamma)$.

Proof: The symplectic Fourier transform of f, $\tilde{f}$, is obviously of $\hbar$-positive type. Also, since $f \in L^1(\Gamma)$, a standard theorem from harmonic analysis implies that $\tilde{f} \in L^\infty$. By Theorem 3.3, $\tilde{f}$ is finite at the origin, and, by Theorem 4.1, it has the form

$$\tilde{f} = \tilde{g}^\dagger \times \tilde{g}, \quad \text{with } \tilde{g} \in L^2(\Gamma'). \tag{5.5}$$

That f is of the form (5.4) follows on taking the inverse symplectic Fourier transform of both sides (5.5). The rest of the corollary follows directly from Theorem 5.1 and Corollary 5.2. ■

There is no counterpart to Proposition 4.3 in Sec. IV. The conditions given in Corollary 5.3 do *not* constitute a set of conditions equivalent to those under which an $\hbar$-positive tempered distribution will be finite on unity. Indeed, there are certainly $\hbar$-positive tempered distributions that are functions which are not in $L^1(\Gamma)$, but which are of the form (5.4). (See Ref. 15, Eq. (2.57) and Ref. 16.) Of course, they must still be square-integrable.[13]

For a number of corollaries along the lines of our last one, we refer the interested reader to similar results we have derived in a recent paper (Ref. 5, Theorem 3.2). In that paper, these results were applied to solving a quantum mechanical moment problem.

VI. APPLICATIONS

At the end of the last section, we mentioned that our results have already been applied to solve a quantum mechanical moment problem. There are several other potential applications.

Using our results, it is possible to conclude that tempered-distribution solutions to the quantum Liouville equa-

tion are actually quantum mechanical states, provided only that the initial condition was such a state. To give a sketch of how this can be done, let us assume that the quantum Liouville equation

$$i\hbar\frac{\partial\rho^t}{\partial t}=\rho^t\circ H-H\circ\rho^t \tag{6.1}$$

generates a flow that maps $\mathscr{S}$ and $\mathscr{S}'$ into themselves. It is known [14] that, under fairly mild assumptions on the Hamiltonian H, the quantum Liouville equation will preserve the twisted product in the sense that, for $f,g\in\mathscr{S}$, if, under the flow, $f\to f^t$ and $g\to g^t$, then $f\circ g\to f^t\circ g^t$. Pick a sequence $\{s_k\}$ satisfying (5.3) and use the quantum Liouville equation to evolve it backwards in time to the sequence $\{s_k^{-t}\}$. It is easy to show that if the initial condition ρ^0 is a quantum mechanical state, then this inequality holds:

$$0\leqslant\langle\rho^0,\bar{s}_k^{-t}\circ s_k^{-t}\rangle\leqslant C\,\|s_k^{-t}\|_2^2\,. \tag{6.2}$$

Evolving the initial state and the sequence $\{s_k^{-t}\}$ for a time t turns (6.2) into

$$0\leqslant\langle\rho^t,\bar{s}_k\circ s_k\rangle\leqslant C\,\|s_k\|_2^2\,. \tag{6.3}$$

Because the first member of the sequence can be an arbitrary function in $\mathscr{S}$, the inequality (6.3) implies that ρ^t is $\hbar$-positive. In addition,

$$\|s_k\|_2^2=\tilde{s}_k^{\dagger}\times\tilde{s}_k(0)\leqslant\|\tilde{s}_k\|_1^2\to 1,$$

so from (6.3), we also have that

$$\limsup_{k\to\infty}\langle\rho^t,\bar{s}_k\circ s_k\rangle\leqslant C<\infty.$$

Hence, ρ^t is also finite on unity. By Theorem 5.1, it has the form (5.4) and is, therefore, a (possibly unnormalized) quantum mechanical state. With a bit more work, one can also show that the state is in fact properly normalized. (See Ref. 14)

Several questions arise in connection with this example. These involve the quantum Liouville equation, but in a finite domain with boundary conditions of one sort or another applied. What we have in mind is what one does if one wants to *numerically* solve the quantum Liouville equation. Such numerical solutions have recently become of interest in modeling quantum devices.[17,18] Of course, this has to be done on a bounded domain, and so boundary conditions that do not create artificial reflections must be applied (cf. Refs. 19 and 20). Is it possible to find boundary conditions that will allow the restriction of a Wigner function to a finite domain to evolve, via the quantum Liouville equation subject to boundary conditions, into the restriction of a Wigner function? In particular, do the boundary conditions derived by Ringhofer *et al.* [19] allow for this? More generally, is there a way of determining whether or not a function that is defined on some bounded domain in phase space is the restriction of a Wigner function?

Although our principal interest is in applications to quantum physics, this is not the only possible area in which our results may be applied. It should certainly be possible to use our results to determine when a time-frequency function is the Wigner distribution for an acoustical signal,[21] for example. Other possible applications are to answering similar questions about Wigner functions arising in connection with the Schrödinger-like wave equations that are used in the "parabolic" approximation to the Helmholtz equation.[22]

A specifically nonquantum mechanical application arises in signal theory. (See Chap. 8 of Schempp's book [23] for a brief introduction.) The Wigner function corresponding to a signal f, with $\int_{\mathbf{R}}|f(t)|^2\,dt=1$, is defined [23] as

$$P(f;z)\overset{\text{def}}{=}\int_{\mathbf{R}}f\left(q+\frac{t}{2}\right)\bar{f}\left(q-\frac{t}{2}\right)e^{2i\pi pt}\,dt\,. \tag{6.4}$$

This is the analog of the Wigner function for a quantum mechanical state described by a wave function f, provided one takes $\hbar=1/\pi$. One then defines the *ambiguity function* $H(f;\cdot)$ as $\tilde{P}(f;\cdot)$. A straightforward calculation then implies that

$$H(f;a)=\int_{\mathbf{R}}f\left(q+\frac{u}{4\pi}\right)\bar{f}\left(q-\frac{u}{4\pi}\right)e^{iqv}\,dq\,. \tag{6.5}$$

Schempp (Ref. 23, Theorem 8.8) uses group-theoretical methods to solve the *radar synthesis problem*: Find the image of the map $f\to H(f;\cdot)$, when f is allowed to vary over all of $\mathscr{S}(\mathbf{R})$. What he shows is that the image of this map consists precisely of all functions $F(a)\in\mathscr{S}(\Gamma')$, which are of $\hbar$-positive type and which are *extremals* in the convex set of all $\hbar$-positive type functions that are 1 at $a=0$. [Here, of course, we take $\hbar=1/\pi$. Also, it should be noted that our ambiguity function differs from Schempp's; the two are related by a scaling of the argument $a=(u,v)$.]

It is natural to ask what happens to the ambiguity function if *nonsmooth* f are allowed. Instead of approaching this question by specifying a class to which f belongs (i.e., specifying the *domain* of H), we simply start by looking at candidates for the image of H. If $F(a)$ is a distribution corresponding to some f via $F(a)=H(f;a)$, then it is reasonable to assume that F be of $\hbar$-positive type. If, in addition, F is finite at the origin, then Theorem 4.1 implies that $F=G^{\dagger}\times G$ for some $G\in L^2(\Gamma')$.

Which of these F are ambiguity functions for some f? Since F is continuous (Corollary 4.2), any F that is an ambiguity function must satisfy

$$F(0)=H(f;0)=\int_{\mathbf{R}}|f(q)|^2\,dq=1. \tag{6.6}$$

Thus the corresponding f must be square integrable. Using the Schrödinger representation, we can uniquely associate the non-negative trace-class operator $\hat{F}=\pi_{\omega}(F)$ with F. From (6.6), (4.2), and the fact that $\hat{I}$, the identity operator, corresponds to $2\pi\delta(a)$, one can easily show that $\mathrm{Tr}[\hat{F}]=\pi$. Set $\hat{\rho}=\hat{F}/\pi$, so that $\mathrm{Tr}[\hat{\rho}]=1$, making $\hat{\rho}$ a "density matrix." We now appeal to the well-known connections among Wigner functions, density matrices, pure states, and wave functions to conclude that $F(\cdot)=H(f;\cdot)$ if and only if the operator $\hat{\rho}$ is a rank 1 orthogonal projection.

Such a non-negative trace-class operator is easily characterized as having unit trace and satisfying the identity $\hat{\rho}=\hat{\rho}^2$. This, in turn, translates into $F=(1/\pi)F\times F$ for the $\hbar$-positive-type distribution F.

It is interesting to note that we can characterize all functions F such that $F(\cdot)=H(f;\cdot)$ via three simple conditions: (i) $F(0)=1$; (ii) $F^{\dagger}=F$; and (iii) $F=(1/\pi)F\times F$. The second and third conditions force F to be of $\hbar$-positive type,

and the first and third force the corresponding density matrix $\hat{\rho}$ to be a pure state—which guarantees that F be of the required form, with $f \in L^2$. If F is in $\mathcal{S}$, then, because the twisted product of two functions in $\mathcal{S}$ is still in $\mathcal{S}$, we get back Schempp's result (Ref. 23, Theorem 8.8).

ACKNOWLEDGMENTS

The author wishes to thank Professor Peter D. Lax for the many helpful and interesting conversations he had with him. In addition, he would like to thank Texas A&M's Association of Former Students for support during his recent study leave, and the Courant Institute for its hospitality during that leave.

[1]D. Kastler, "The C^*-algebras of a free boson field," Commun. Math. Phys. **1**, 14 (1965).
[2]G. Loupias and S. Miracle-Sole, "C^*-algèbres des systèms canoniques I," Commun. Math. Phys. **2**, 31 (1966).
[3]G. Loupias and S. Miracle-Sole, "C^*-algèbres des systèms canoniques II," Ann. Inst. H. Poincarè **6**, 39 (1967).
[4]F. J. Narcowich and R. R. O'Connell, "Necessary and sufficient conditions for a function defined on phase space to be a Wigner distribution," Phys. Rev. A **34**, 1 (1986).
[5]F. J. Narcowich, "The problem of moments in the phase-space formulation of quantum mechanics," J. Math. Phys. **28**, 2873 (1987).
[6]I. E. Segal, "Transforms for operators and symplectic automorphisms over a locally compact group, " Math. Scand. **13**, 31 (1963).
[7]F. Treves, *Topological Vector Spaces, Distributions, and Kernels* (Academic, New York, 1967).
[8]J.-M. Maillard, "Structure des convoluteurs à gauche des $\mathcal{S}(\mathbf{R}^{2n})$ et de symboles admissibles de la transformation de Weyl," C. R. Acad. Sci. Paris **301**, Série I, 719 (1985).
[9]J. M. Gracia-Bondía and J. C. Várilly, "Algebras of distributions suitable for phase space quantum mechanics. I," J. Math. Phys. **29**, 869 (1988).
[10]R. Howe, "Quantum mechanics and partial differential equations, " J. Funct. Anal. **38**, 188 (1980).
[11]I. C. Gohberg and M. G. Kreĭn, *Introduction to the Theory of Nonselfadjoint Operators* (American Mathematical Society, Providence, 1969).
[12]A. Grossman, G. Loupias, and E. M. Stein, "An algebra of pseudodifferential operators and quantum mechanics in phase space," Ann. Inst. Fourier (Grenoble) **18**, 343 (1968).
[13]J. C. T. Pool, "Mathematical aspects of the Weyl correspondence," J. Math. Phys. **7**, 66 (1966).
[14]F. J. Narcowich, "A Dyson-like expansion for solutions to the quantum Liouville equation," J. Math. Phys. **27**, 2502 (1986).
[15]F. J. Narcowich and S. A. Fulling, *Wigner Distribution Functions, Seminars in Mathematical Physics* (Texas A&M University, College Station, TX, 1986), No. 1.
[16]I. Daubechies, "Continuity statements and counterintuitive examples in connection with Weyl quantization," J. Math. Phys. **24**, 1453 (1983).
[17]W. R. Frensley, "Quantum transport of resonant-tunneling devices," Solid-State Electronics **31**, 739 (1988).
[18]N. C. Kluksdahl, A. M. Kriman, C. Ringhofer, and D. K. Ferry, "Quantum tunneling properties from a Wigner function study," Solid-State Electronics **31**, 743 (1988).
[19]C. Ringhofer, D. Ferry, and N. Kluksdahl, "Absorbing boundary conditions for the simulation of quantum transport phenomena," to appear in Transp. Theory Stat. Phys.
[20]B. Engquist and A. Majda, "Absorbing boundary conditions for the numerical simulation of waves," Math. Comp. **31**, 629 (1977).
[21]H. Szu, "Signal processing using bilinear and nonlinear time-frequency joint representations," in *The Physics of Phase Space, Lecture Notes in Physics*, Vol. 278, edited by Y. S. Kim and W. W. Zachary (Springer, Berlin, 1987).
[22]L. Fishman and J. J. McCoy, "Derivation and application of extended parabolic wave theories. I. The factorized Helmholtz equation," J. Math. Phys. **25**, 285 (1984); J. R. Klauder, "New asymptotics for old wave equations," Science **239**, 760 (1988).
[23]W. Schempp, *Harmonic Analysis on the Heisenberg Nilpotent Lie Group*, Pitman Research Notes in Mathematics, No. 147 (Longman, Harlow, Essex, 1986).

Bound states in curved quantum waveguides

P. Exner[a] and P. Seba[b]
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141980 Dubna, Union of Soviet Socialist Republics

(Received 20 November 1987; accepted for publication 31 May 1989)

A free quantum particle living on a curved planar strip Ω of a fixed width d with Dirichlet boundary conditions is studied. It can serve as a model for electrons in thin films on a cylinder-type substrate, or in a curved quantum wire. Assuming that the boundary of Ω is infinitely smooth and its curvature decays fast enough at infinity, it is proved that a bound state with energy below the first transversal mode exists for all sufficiently small d. A lower bound on the critical width is obtained using the Birman–Schwinger technique.

I. INTRODUCTION

During the early days of quantum mechanics, descriptions of a particle motion confined to a fixed subset of configuration space were the most useful textbook illustrations. The situation has changed substantially with the advance of thin-film physics and microelectronics. Today there are many possibilities of how to construct metallic or semiconductor structures, in which the electron motion is essentially two-dimensional as in the thin films, or one-dimensional as in the so-called quantum wires,[1] and free or quasifree.[2]

From this point of view, studies of a quantum particle motion in a layer or a tube with some boundary conditions (and more complicated sets: branching tubes, sandwiched layers, etc.) are of a great physical interest. Recall that analogous problems have been studied extensively in other fields, particularly in radar physics and acoustics. Some results concerning stationary scattering problems are easily translated from here to the situations, when one is interested in the Schrödinger equation instead of the wave equation. On the other hand, there are effects which have no counterpart in classical waveguides. In the present paper, we are going to study one of them.

We shall be interested in a free quantum motion on a curved planar strip with a Dirichlet boundary.[3] Such a system can be viewed as a model of two physical situations: a thin film on a substrate of an open-cylinder form (so the third dimension may be separated) or a curved quantum wire on a planar substrate (where the "vertical" dimension is separated). The method employed in this paper extends easily to more general cases of an arbitrary curved layer or tube[4] in $\mathbb{R}^3$; we are going to discuss it in a separate publication.

There is one more motivation for our study. The motion on a thin strip can be modeled by motion on a curve. In this case one must ask, of course, how this one-dimensional dynamics will be influenced by the curvature of the prescribed "path." Several authors[5–8] found independently that the problem was formally equivalent to the motion on a straight line under the influence of a potential which is nonpositive and proportional to the squared curvature; this implied that bound states may exist on such a curve.[9] The method was the same in all cases, starting with a strip of a finite width and limiting the latter to zero. The trouble is that the procedure was performed on a purely formal level, with subtraction of infinite quantities representing the transversal-mode energies in the limit. Moreover, it is not clear, what happens with the bound states for a small but finite width of the strip.

Our aim here is to discuss this problem on a rigorous level. We assume that the strip has an infinitely smooth boundary and its curvature decays rapidly enough at infinity, roughly speaking as $|s|^{-3/2-\epsilon}$, where s is the natural longitudinal coordinate on the strip. Both these assumptions may be weakened, but not too much. It can be seen easily that the present method works for a C^4 boundary. On the other hand, we will give examples suggesting that the results might not be valid for the "bookcover" strip whose boundary is C^1 only.

As for the decay restriction imposed on the curvature, our main result (Theorem 4.1) is derived under a weaker assumption, roughly speaking the $|s|^{-1-\epsilon}$ decay. The above-mentioned hypothesis is needed for proof of the finer estimate presented in Sec. V, but it might be weakened with some additional effort. On the other hand, the character of spectrum could be completely different if we relax the decay restriction at all. A particularly interesting case is represented by a periodic strip which we comment on briefly in the conclusions.

Let us finally review the contents of this paper. In the next section, we describe the problem. Motivated by the above physical considerations, we use the term *quantum waveguide* to describe free motion of a Schrödinger particle on a strip. We collect there also some assumptions which will be used further on. In Sec. III, we reformulate our problem using the natural curvilinear coordinates to investigate a unitarily equivalent operator of a more complicated structure which acts, however, on a straight strip. Estimating this operator and using the minimax principle, we find that for a sufficiently small width of the strip there is at least one bound state below the first transversal-mode energy. We obtain also an estimate of the critical width, but a relatively poor one. In Sec. V, we improve it using the Birman–Schwinger technique. Some examples are presented; in conclusion we mention the physical relevance of the results.

[a] On leave of absence from Nuclear Physics Institute, Czechoslovak Academy of Sciences, 25068 Řež near Prague, Czechoslovakia.
[b] On leave of absence from Nuclear Centre, Charles University, V. Holešovičkách 2, 18000 Prague, Czechoslovakia.

 0022-2488/89/112574-07$02.50

II. QUANTUM WAVEGUIDES

In order to describe a quantum particle whose motion is confined to a region Ω of the configuration space, one has to specify its interaction with the boundary. Here we assume infinitely hard walls, i.e., we choose the Hamiltonian to be

$$H_\Omega = (\hbar^2/2m)\Delta_D, \tag{2.1}$$

where m is the (effective) mass and Δ_D is Dirichlet Laplacian on $L^2(\Omega)$, i.e., the Friedrichs extension of the map $\psi \mapsto \Delta\psi$ defined on $C_0^\infty(\Omega)$.[10]

As we have said, Ω is assumed to be a curved planar strip of a width d (Fig. 1). Its points are described by the curvilinear coordinates s,u as follows:

$$\begin{aligned} x &= a(s) - ub'(s), \\ y &= b(s) + ua'(s), \end{aligned} \tag{2.2}$$

where a,b are smooth functions that characterize the reference curve $\Gamma = \{(a(s),b(s)): s\in\mathbb{R}\}$. We assume

$$a'(s)^2 + b'(s)^2 = 1, \tag{2.3}$$

so s is the arc length of Γ, while u means the distance of the point (x,y) from Γ. It is useful to introduce the *signed curvature* $\gamma(s)$ of Γ,

$$\gamma(s) = b'(s)a''(s) - a'(s)b''(s), \tag{2.4}$$

named so because $|\gamma(s)|$ represents curvature of the reference curve at s. It is easy to see that $|\gamma(s)| = (a''(s)^2 + b''(s))^{1/2}$. More generally,

$$c(s,u) = \frac{|\gamma(s)|}{|1 + u\gamma(s)|} \tag{2.5}$$

is the curvature of the fixed u curve at the point (s,u), so the strip width d must be restricted by the requirement

$$d\gamma(s) > -1. \tag{2.6}$$

In fact, it is sufficient to know the function γ only, since a,b can be reconstructed from the relations

$$\begin{aligned} a(s) &= a(s_0) + \int_{s_0}^{s} \cos\left(\int_{s_0}^{s_1} \gamma(s_2)ds_2\right)ds_1, \\ b(s) &= b(s_0) + \int_{s_0}^{s} \sin\left(\int_{s_0}^{s_1} \gamma(s_2)ds_2\right)ds_1; \end{aligned} \tag{2.7}$$

the correspondence is unique up to Euclidean transformations of the plane. In what follows, we shall therefore characterize Ω by the function γ and the width d.

For obvious reasons, Ω will be called *quantum waveguide*, or briefly *Q-guide*. Let us collect the hypotheses that will be used in the following considerations.

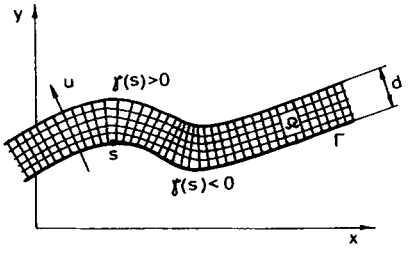


FIG. 1. A curved quantum waveguide.

Assumption 2.1: (a) Ω has an infinitely smooth boundary, i.e., $\gamma\in C^\infty(R)$; (b) the functions γ, γ', and γ'' are bounded; (c) as a strengthening of (2.6), we assume existence of $\gamma_-\in(-d^{-1},0]$ such that $\gamma(s)\geqslant\gamma_-$ for all $s\in\mathbb{R}$. In the present paper, we restrict out attention to the Q-guides which are curved substantially only within a bounded region. The requirements on the decay of γ at infinity may be formulated, e.g., as follows: (d) the functions γ^2 and γ'' belong to $L(\mathbb{R},(1+s)ds)$; (e) the functions γ and γ' belong to $L^2(\mathbb{R},(1+s^2)ds)$; (f) $\gamma\in L(\mathbb{R})$. As an additional assumption, we shall be concerned only with the nontrivial case of *curved Q-guides*, i.e., such that $\gamma(s)$ is nonzero for some s.

A Q-guide is called *simply bent* if γ does not change sign in $\mathbb{R}$. Without loss of generality, we may consider $\gamma(s)\geqslant 0$ for all $s\in\mathbb{R}$, the case of a nonpositive γ being reduced to the present one by taking the other boundary as the reference curve. If γ is sign-changing, we say Ω is *multiply bent*. It is clear that $\int_{s_1}^{s_2}\gamma(s)ds$ is the angle between the tangent vectors to Γ at the points s_1 and s_2. Hence the assumption (f) gives us the possibility to define the *overall bending*

$$\beta(\gamma) = \int_{\mathbb{R}} \gamma(s)ds \tag{2.8}$$

of Ω. Notice that $\beta(\gamma)$ may exist even if Γ has no asymptotes. If it has, we say that Ω is *asymptotically straight*; it is true, e.g., if $\gamma(s) = O(|s|^{-2-\epsilon})$ as $|s|\to\infty$.

The curvilinear coordinates s,u are locally orthogonal so the metrics in Ω are expressed with respect to them through a diagonal metric tensor, $dx^2 + dy^2 = g_{ss}\,ds^2 + g_{uu}\,du^2$. One finds easily

$$g_{ss} \equiv g \equiv g(s,u) = (1 + u\gamma(s))^2, \tag{2.9a}$$

$$g_{uu} = 1. \tag{2.9b}$$

We shall also need the Jacobian

$$\frac{\partial(x,y)}{\partial(s,u)} = 1 + u\gamma(s) = g^{1/2}, \tag{2.10}$$

which shows that the transition to the curvilinear coordinates represents an isometric map of $L^2(\Omega)$ to $L^2(\mathbb{R}\times[0,d],g^{1/2}\,ds\,du)$.

III. THE HAMILTONIAN

Let us return now to the operator (2.1). Our aim in this section is to prove the following assertion:

Theorem 3.1: Assume (a). Then H_Ω is unitarily equivalent to Friedrichs extension H of the operator H_0 defined on $L^2(\mathbb{R}\times[0,d])$ by

$$H_0\psi = -\frac{\hbar^2}{2m}\frac{\partial}{\partial s}g^{-1}\frac{\partial\psi}{\partial s} + \frac{\partial^2\psi}{\partial u^2} + V(s,u)\psi, \tag{3.1a}$$

with the domain $D(H_0) = C_0^\infty(\mathbb{R}\times[0,d])$, where

$$V(s,u) = \frac{\hbar^2}{2m}\left(\frac{1}{2}g^{-3/2}\frac{\partial^2\sqrt{g}}{\partial s^2} - \frac{5}{4}g^{-2}\left(\frac{\partial\sqrt{g}}{\partial s}\right)^2 - \frac{1}{4}g^{-1}\left(\frac{\partial\sqrt{g}}{\partial u}\right)^2\right). \tag{3.1b}$$

The Friedrichs extension contains the set

$$D=\{\psi:\ \psi \text{ is } C^\infty,\ \psi(s,0)=\psi(s,d)=0 \text{ for all } s\in\mathbb{R},\ H\psi\in L^2\} \tag{3.2}$$

in its domain and acts on $\psi\in D$ according to (3.1).

Proof: Translating the action of H_Ω on $C_0^\infty(\Omega)$ into curvilinear coordinates, one should replace Laplacian by the Laplace–Beltrami operator relative to the given metric tensor (g_{ij}),

$$H_\Omega\psi=-\frac{\hbar^2}{2m}g^{-1/2}\frac{\partial}{\partial s^i}g^{1/2}g^{ij}\frac{\partial\psi}{\partial s^j},$$

where $s^1=s$, $s^2=u$, and furthermore, $g=\det(g_{ij})=1+u\gamma(s)$, which justifies the notation (2.9a). Since $(g^{ij})=\mathrm{diag}(g^{-1},0)$, we have

$$H_\Omega\psi=-\frac{\hbar^2}{2m}\left(g^{-1/2}\frac{\partial}{\partial s}g^{-1/2}\frac{\partial\psi}{\partial s}+g^{-1/2}\frac{\partial}{\partial u}g^{1/2}\frac{\partial\psi}{\partial u}\right), \tag{3.3}$$

where ψ belongs to the subset $C_0^\infty(\mathbb{R}\times[0,d])$ of $L^2(\mathbb{R}\times[0,d],g^{1/2}\,ds\,du)$. Alternately, the relation (3.3) can be checked by a straightforward differentiation. In order to get an operator on $L^2(\mathbb{R}\times[0,d])$, one has to pass from ψ to $g^{-1/4}\psi$. The sought unitary operator U: $L^2(\Omega)\to L^2(\mathbb{R}\times[0,d])$ is therefore given by

$$(U\psi)(s,u)=g^{1/4}\psi(x,y), \tag{3.4}$$

where the points (x,y) and (s,u) are related by (2.2); it is important that U maps $C_0^\infty(\Omega)$ onto $C_0^\infty(\mathbb{R}\times[0,d])$.

Let us show that $H=UH_\Omega U^{-1}$. First we define the quadratic form q_0 on $C_0^\infty(\mathbb{R}\times[0,d])\times C_0^\infty(\mathbb{R}\times[0,d])$ by

$$q_0(\varphi,\psi)=\frac{\hbar^2}{2m}\int_{\mathbb{R}\times[0,d]}\left(g^{-1}\frac{\overline{\partial\varphi}}{\partial s}\frac{\partial\psi}{\partial s}+\frac{\overline{\partial\varphi}}{\partial u}\frac{\partial\psi}{\partial u}+V(s,u)\bar\varphi\psi\right)ds\,du. \tag{3.5}$$

It is not difficult to see that the form q_0 is associated with the operator H_0, i.e., that

$$q_0(\varphi,\psi)=(\varphi,H_0\psi)=(\varphi,UH_\Omega U^{-1}\psi) \tag{3.6}$$

holds for all $\varphi,\psi\in C_0^\infty(\mathbb{R}\times[0,d])$. Using (3.5), we get a relation between the form q_0 and the closed symmetric form q_Ω associated with H_Ω, namely,

$$q_0(\varphi,\psi)=q_\Omega(U^{-1}\varphi,U^{-1}\psi) \tag{3.7}$$

for all $\varphi,\psi\in C_0^\infty(\mathbb{R}\times[0,d])$, where q_Ω is the standard Dirichlet form[10] on Ω. In order to get Friedrichs extension of the operator H_0, we have to close the form q_0; then H is the self-adjoint operator associated with $\overline{q_0}$ by the first representation theorem.[11] It follows immediately from (3.7) that $D(\overline{q_0})=U^{-1}D(q_\Omega)$ and

$$\overline{q_0}(\varphi,\psi)=q_\Omega(U^{-1}\varphi,U^{-1}\psi) \tag{3.8}$$

for $\varphi,\psi\in D(\overline{q_0})$. The relation $H=UH_\Omega U^{-1}$ is then a simple consequence of (3.8) and the first representation theorem. It remains to prove the relation $D\subset D(H)$, which follows from the established unitary equivalence together with the inclusion

$$D_\Omega\equiv\{f\in L^2(\Omega):\ f\in C^\infty(\Omega) \text{ and } f=0 \text{ on } \partial\Omega\}\subset D(H),$$

which can be checked using an argument similar to Ref. 10, Proposition 1. ■

We notice also the scaling property of the Hamiltonian. Consider the rescaled strip Ω_ϵ of width ϵd and the reference curve corresponding to the signed curvature γ_ϵ: $\gamma_\epsilon(s)=\epsilon^{-1}\gamma(s/\epsilon)$, and denote by H_ϵ the corresponding Hamiltonian. Then a simple calculation yields the following proposition.

Proposition 3.2: H_ϵ is unitarily equivalent to $\epsilon^{-2}H$.

Before proceeding further, let us express the potential (3.1b) in terms of the signed curvature. Substituting from (2.9a), we get

$$V(s,u)=\frac{\hbar^2}{2m}\left\{-\frac{\gamma^2}{4(1+u\gamma)^2}+\frac{u\gamma''}{2(1+u\gamma)^3}-\frac{5}{4}\frac{u^2\gamma'^2}{(1+u\gamma)^4}\right\}. \tag{3.9}$$

IV. EXISTENCE OF BOUND STATES

It may not be easy to find the bound states of H directly by solving the partial differential equation $H\psi=E\psi$. However, H is bounded from below so one may employ the minimax principle.[12] It is sufficient to find a suitable self-adjoint operator H_+ such that $H\leqslant H_+$ in the form sense on a common core; then the relation $E_k\leqslant E_k^{(+)}$ holds between the eigenvalues of H and H_+ arranged in the growing order with respect to multiplicity, and $E_\infty\leqslant E_\infty^{(+)}$ between the lower edges of the essential spectrum. In particular, if $E_\infty=E_\infty^{(+)}$ and H_+ has an eigenvalue below E_∞, the same is true for H.

Theorem 4.1: Let Ω be a curved Q-guide fulfilling the assumptions (a)–(d). Then

(i) $\sigma_{\mathrm{ess}}(H)=[E_\infty,\infty)$, where $E=\hbar^2\pi^2/2md^2$ and d is the width of the strip;

(ii) there is a positive d_0 such that for each $d<d_0$, H has at least one bound state in $[0,E_\infty)$.

Proof: Consider the operators

$$H_s^\pm=-\frac{\hbar^2}{2m(1+d\gamma_\mp)^2}\frac{d^2}{ds^2}+V_\pm(s) \tag{4.1a}$$

with

$$D(H_s^\pm)=\{f:\ f\in C^\infty(\mathbb{R}),\ H_s^\pm f\in L^2\}, \tag{4.1b}$$

where $\gamma_+=\sup_s\gamma(s)$ and $\gamma_-=\inf_s\gamma(s)$ [both are finite according to (b) and (c)] and $V_\pm(s)$ are suitable upper and lower bounds to the potential (3.9), and furthermore

$$H_u=-\frac{\hbar^2}{2m}\frac{d^2}{du^2} \tag{4.2a}$$

with

$$D(H_u)\equiv D_u=\{g:\ g\in C^\infty[0,d],\ g(0)=g(d)=0\}. \tag{4.2b}$$

In view of (b), the potentials $V_\pm$ can be chosen to be bounded so $H_s^\pm$ are e.s.a. on

$$D_s=\{f:\ f\in C^\infty,\ f''\in L^2\}. \tag{4.1c}$$

Then the operators

$$H_{\pm} = H_s^{\pm} \otimes I + I \otimes H_u \tag{4.3a}$$

are e.s.a. on $D_s \otimes D_u$. Since this domain is easily seen to be contained in D, it represents a common form core for H and $H_{\pm}$, and

$$(\psi, H_- \psi) \leqslant (\psi, H\psi) \leqslant (\psi, H_+ \psi) \tag{4.3b}$$

holds for all $\psi \in D_s \otimes D_u$ according to construction. Hence the minimax principle may be applied.

Let us now specify the potentials $V_{\pm}$. We choose

$$V_+(s) = \frac{\hbar^2}{2m}\left\{ -\frac{\gamma(s)^2}{4(1+d\gamma_+)^2} + \frac{d\gamma''_+(s)}{2(1+d\gamma_-)^3}\right\}, \tag{4.4a}$$

$$V_-(s) = \frac{\hbar^2}{2m}\left\{ -\frac{\gamma(s)^2}{4(1+d\gamma_-)^2} - \frac{d\,|\gamma''(s)|}{2(1+d\gamma_-)^3} - \frac{5}{4}\frac{d^2\gamma'(s)^2}{(1+d\gamma_-)^4}\right\}, \tag{4.4b}$$

where γ''_+ is the positive part of γ'' and $\gamma_{\pm}$ have been defined above. According to (d), the functions γ^2 and γ''_+ are integrable, and since they are bounded due to (b), they belong to L^2. Furthermore, the integrability of γ'' implies $\gamma'(s) \to 0$ as $s \to \pm\infty$. Since γ and γ'' are square-integrable, a simple integration by parts shows that γ' is again square-integrable. It means that the functions $V_{\pm}$ belong to L^2, and consequently,[13] the essential spectrum of the operators $H_s^{\pm}$ equals $[0,\infty)$. On the other hand, the spectrum of H_u is purely discrete and equal to $\{\hbar^2\pi^2k^2/2md^2 : k = 1,2,...\}$.

The spectrum of the operators (4.3a) is then[14] none other than the "sum" of the spectra of $H_s^{\pm}$ and H_u. In particular, the essential spectrum of both the operators $H_{\pm}$ has its bottom at $E = \hbar^2\pi^2/2md^2$, and the same is, of course, true for $\sigma_{\mathrm{ess}}(H)$. This proves the assertion (i). Moreover, it is clear that H_+ has an eigenvalue below E_∞ iff the operator H_s^+ has a negative eigenvalue. Since V_+ fulfills the decay condition

$$\int_{\mathbb{R}} V_+(s)(1+|s|)\,ds < \infty$$

according to (d), such an eigenvalue exists iff[15]

$$\int_{\mathbb{R}} V_+(s)\,ds < 0. \tag{4.5a}$$

Obviously, this is true for small enough d. ■

The proof not only establishes existence of the critical width d_0, but allows us also to estimate it from below. For simplicity, we restrict ourselves to the case of a simply bent Q-guide. Then $\gamma_- = 0$ so the condition (4.5a) yields

$$\frac{d}{2}\int_{\mathbb{R}} \gamma''_+(s)\,ds < \frac{1}{4(1+d\gamma_+)^2}\int_{\mathbb{R}} \gamma(s)^2\,ds. \tag{4.5b}$$

Solving the resulting cubic equation for d, we arrive at the following proposition.

Proposition 4.2: For a simply bent Q-guide fulfilling (a)–(d), the critical width d_0 obeys

$$d_0 \geqslant d_+ \equiv (2/3\gamma_+)\{\mathrm{ch}(\tfrac{1}{3}\ln(z+\sqrt{z^2-1})) - 1\}, \tag{4.6a}$$

where

$$z = 1 + \frac{27}{4}\gamma_+\left(\int_{\mathbb{R}} \gamma''_+\,ds\right)^{-1}\int_{\mathbb{R}} \gamma^2\,ds. \tag{4.6b}$$

As an illustration we evaluate d_+ for a few cases listed below. To obtain comparable results, we choose all of them in such a way that the maximal curvature is $\gamma_+ = \alpha/\rho$ and the overall bending $\beta = \pi\alpha$.

Example 4.3: (i) $\gamma(s) = \alpha\rho/(\rho^2+s^2)$, which is not asymptotically straight; (ii) Gaussian curvature, $\gamma(s) = (\alpha/\rho)\exp(-s^2/\pi\rho^2)$; (iii) a generalization of the previous case, $\gamma(s) = (\alpha/\rho)\exp(-\mu s^{2n})$ with $\mu = \{\Gamma(1/2n)/\pi\rho n\}^{2n}$.

The ratio $\alpha d_+/\rho$ of the estimated critical width to the minimal curvature radius is plotted on Fig. 2. In case (iii), the estimate is getting worse with growing n. Recall that our method requires the boundary of Ω to be smooth enough, otherwise problems with the domains will appear when using the unitary transformation (3.4). The curves (iii) approach with growing n the "bookcover" case[8]

$$\gamma(s) = \begin{cases} \alpha/\rho \cdots |s| \leqslant \frac{1}{2}\pi\rho, \\ 0 \cdots |s| > \frac{1}{2}\pi\rho, \end{cases} \tag{4.7}$$

where the boundary is only C^1. Hence the behavior of the estimate is interesting; it may suggest that the lack of smoothness is more than a mathematical nuisance here. For a nonsmooth γ, the potential (3.9) contains contact-interaction terms supported by the discontinuity points of g, and this contact interaction might prevent existence of a bound state.

Remark 4.4: In the first two cases we have plotted the ratio $\alpha d_+/\rho$ also for the overall bending $\beta > \pi$ which is, strictly speaking, not possible under our assumptions. As mentioned in the Introduction, coiled quantum wires should be modeled by a curved tube in the three-dimensional space. If the coil is nearly planar, however, one may describe it approximately in the present formalism, with $\mathbb{R}^2$ replaced by a multisheeted Riemannian surface. The above considerations translate to this situation immediately and the results then make sense for an arbitrary β.

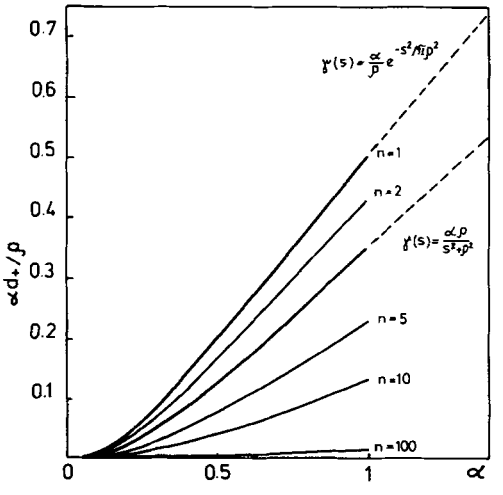


FIG. 2. The estimate (4.6).

V. ANOTHER BOUND ON THE CRITICAL WIDTH

From the viewpoint of existence, Theorem 4.1 solves the problem posed in the Introduction. The bound (4.6) is relatively poor, however, since we have obtained it estimating H roughly by the operator (4.3a). In the present section, we are going to derive a better bound under slightly stronger assumptions about the decay of γ at infinity. We employ the Birman–Schwinger technique, which has yielded already the condition (4.5a). First we derive a useful technical result concerning the operator

$$H_\lambda = -\Delta_D + \lambda V \tag{5.1}$$

on $L^2(\mathbb{R}) \otimes L^2(0,d)$, where Δ_D is the Dirichlet Laplacian.

Lemma 5.1: Suppose that $V: \mathbb{R}\times[0,d]\to\mathbb{R}$ is bounded and measurable, and such that

$$\int_{\mathbb{R}\times[0,d]} (1+x^2)V(x,y)\,dx\,dy < \infty .$$

Then H_λ with $\lambda > 0$ has at least one bound state $E(\lambda)$ below the bottom of essential spectrum, $E(\lambda) < (\pi/d)^2$, iff

$$\int_{\mathbb{R}}\int_0^d V(x,y)\sin^2\left(\frac{\pi y}{d}\right)dx\,dy < 0. \tag{5.2a}$$

Moreover,

$$E(\lambda) = \frac{\pi^2}{d^2} - \frac{\lambda^2}{d^2}\left(\int_{\mathbb{R}}\int_0^d V(x,y)\sin^2\left(\frac{\pi y}{d}\right)dx\,dy\right)^2 + O(\lambda^3) \tag{5.2b}$$

holds for small λ.

Proof: According to the Birman–Schwinger principle,[15,16] $E(\lambda)$ is a bound state of H_λ iff the operator λK_α has the eigenvalue -1 for $\alpha^2 = E(\lambda)$. Here K_α is the integral operator with the kernel

$$K_\alpha(x,y;x',y') = |V(x,y)|^{1/2} R_0(\alpha;x,y;x',y')\, V(x',y')^{1/2},$$

where $R_0(\alpha;\cdot,\cdot,)$ is the kernel of $R_0(\alpha) = (-\Delta_D - \alpha^2)^{-1}$ and $V^{1/2} := |V|^{1/2}\,\mathrm{sgn}\,V$. To express the resolvent $R_0(\alpha)$, we employ the orthogonal decomposition

$$L^2(\mathbb{R})\otimes L^2(0,d) = \bigoplus_{n=1} L^2(\mathbb{R})\otimes\{\chi_n\} \tag{5.3a}$$

with

$$\chi_n(y) = \sqrt{\frac{2}{d}}\sin\left(\frac{\pi n}{d}y\right). \tag{5.3b}$$

One can check easily that $-\Delta_D$ is e.s.a. on

$$D = \{\psi:\ \psi(x,y) = \sum_{n=1}^{N} f_n(x)\chi_n(y),\ f_n\in C_0^\infty(\mathbb{R})\}$$

and that

$$-\Delta_D\psi = \sum_{n=1}^{N}(-f_n'' + \varkappa_n^2 f_n)\chi_n$$

holds for all $\psi\in D$, where

$$\varkappa_n = \pi n/d \tag{5.4a}$$

and $\varkappa_n^2$ is the nth transversal mode energy. Hence $-\Delta_D \upharpoonright D$ is reduced by the projections P_n to the subspaces $L^2(\mathbb{R})\otimes\{\chi_n\}$, and the same is true for its closure. We come to the following conclusion: the operator $-\Delta_D$ is expressed as the orthogonal sum

$$-\Delta_D = \bigoplus_{n=1}^{\infty} P_n\left[\left(-\frac{d^2}{dx^2} + \varkappa_n^2\right)\otimes I\right]P_n \tag{5.4b}$$

with

$$D(-\Delta_D) = \{\psi:\ \psi(x,y) = \sum_{n=1}^{\infty} f_n(x)\chi_n(y),\ f_n\in AC^2(\mathbb{R}), \times \sum_{n=1}^{\infty}\|f_n\|^2 < \infty\}. \tag{5.4c}$$

The resolvent is then also reduced, $P_n R_0(\alpha) = R_0(\alpha)P_n$, and its kernel is given by

$$R_0(\alpha;x,y;x',y') = \sum_{n=1}^{\infty}\chi_n(y) r_n(\alpha;x,x')\chi_n(y'), \tag{5.5a}$$

where $r_n(\alpha;\cdot,\cdot)$ is the kernel of $(-d^2/dx^2 + \varkappa_n^2 - \alpha^2)^{-1}$. We are interested in the energies $E\in[0,\varkappa_1^2)$, i.e., in the case when $\alpha\in[0,\varkappa_1)$. The operators $r_n(\alpha)$ are then bounded for $n = 1,2,\ldots$ since $\varkappa_n^2 \geqslant \varkappa_1^2 > \alpha^2$, and their kernels are well known. It yields

$$R_0(\alpha;x,y;x',y') = \sum_{n=1}^{\infty}\chi_n(y)\frac{\exp[-k_n(\alpha)|x-x'|]}{2k_n(\alpha)}\chi_n(y'), \tag{5.5b}$$

where

$$k_n(\alpha) = \sqrt{\varkappa_n^2 - \alpha^2}. \tag{5.5c}$$

Next we divide the kernel (5.5b) into two parts, one of them being regular as α approaches $\varkappa_1$, and the other singular. Returning to the original operator $\varkappa_\alpha$, we can write it as

$$K_\alpha = M_\alpha + L_\alpha, \tag{5.6a}$$

where the two operators are given by the kernels

$$M_\alpha(x,y;x',y') = |V(x,y)|^{1/2}\sum_{n=2}^{\infty}\chi_n(y)\frac{\exp[-k_n(\alpha)|x-x'|]}{2k_n(\alpha)} \times\chi_n(y')V(x',y')^{1/2} + |V(x,y)|^{1/2}\chi_1(y) \times\frac{\exp-1[-k_1(\alpha)|x-x'|]}{2k_1(\alpha)} \times\chi_1(y')V(x',y')^{1/2} \tag{5.6b}$$

and

$$L_\alpha(x,y;x',y') = [1/2k_1(\alpha)]|V(x,y)|^{1/2} \times\chi_1(y)\chi_1(y')V(x',y')^{1/2}. \tag{5.6c}$$

The first part of the operator M_α can be written as $|V|^{1/2}A_\alpha V^{1/2}$. The operators $|V|^{1/2}$ and $V^{1/2}$ are bounded by assumption, and furthermore, A_α is the orthogonal sum of the operators whose norms are $(\varkappa_n^2 - \alpha^2)^{-1}$. Hence

$$\|A_\alpha\| \sup\{|\varkappa_n^2 - \alpha^2|^{-1}:\ n = 2,3,\ldots\} = (\varkappa_2^2 - \alpha^2)^{-1} < d^2/3\pi^2,$$

and the norm is bounded uniformly with respect to α. Let us

denote the remaining part of M_α for a moment as B_α. We have

$$|(B_\alpha\psi)(x,y)|\leqslant\frac{1}{2}|V(x,y)|^{1/2}\int_{\mathbb{R}\times[0,d]}|x-x'||V(x',y')|^{1/2}\times|\psi(x',y')|dx'\,dy'$$
$$\leqslant\frac{1}{2}|V(x,y)|^{1/2}\|\psi\|\left(\int_{\mathbb{R}\times[0,d]}|x-x'|^2\times|V(x',y')|dx'\,dy'\right)^{1/2}$$

so

$$\|B_\alpha\psi\|^2\leqslant\frac{1}{4}\|\psi\|^2\int_{\mathbb{R}\times[0,d]}dx\,dy|V(x,y)|\times\int_{\mathbb{R}\times[0,d]}|x-x'|^2|V(x',y')|dx'\,dy'.$$

Using the inequality $|x-x'|^2\leqslant(1+x^2)(1+x'^2)$, we find

$$\|B_\alpha\psi\|\leqslant\frac{1}{2}\|\psi\|\int_{\mathbb{R}\times[0,d]}(1+x^2)|V(x,y)|dx,\,dy,$$

where the integral is finite by assumption. It means that $\|M_\alpha\|$ may be estimated by a constant independent of α. Furthermore, L_α is a well-defined rank-one operator, because V is integrable due to the assumption, and therefore the functions $|V|^{1/2}\chi_1$ and $V^{1/2}\chi_1$ belong to $L^2(\mathbb{R}\times[0,d])$.

For a fixed $\alpha\in[0,\varkappa_1)$ and each sufficiently small $\lambda>0$, we have $\|\lambda M_\alpha\|<1$ so

$$(1+\lambda K_\alpha)^{-1}=[1+\lambda(1+\lambda M_\alpha)^{-1}L_\alpha]^{-1}\times(1+\lambda M_\alpha)^{-1}$$

and λK_α has the eigenvalue -1 iff the same is true for $\lambda(1+\lambda M_\alpha)^{-1}L_\alpha$. This operator, however, is of rank one, so it has just one nonzero eigenvalue which we denote as $\xi(\lambda)$; the corresponding eigenvector is $\varphi:=(1+\lambda M_\alpha)^{-1}|V|^{1/2}\chi_1$. Using the explicit form of L_α, we find easily

$$\xi(\lambda)=\frac{\lambda}{2k_1(\alpha)}\int_{\mathbb{R}\times[0,d]}V(x,y)^{1/2}\chi_1(y)\times[(1+\lambda M_\alpha)^{-1}|V|^{1/2}\chi_1](x,y)dx\,dy. \quad (5.7)$$

We are interested in the behavior of $\xi(\lambda)$ for small λ. It holds $(1+\lambda M_\alpha)^{-1}=1-\lambda M_\alpha(1+\lambda M_\alpha)^{-1}$ and

$$\left|\int V^{1/2}\chi_1\lambda M_\alpha(1+\lambda M_\alpha)^{-1}|V|^{1/2}\chi_1\,dx\,dy\right|\leqslant\lambda\,\|V^{1/2}\chi_1\|^2\|M_\alpha(1+\lambda M_\alpha)^{-1}\|$$

so

$$\xi(\lambda)=[\lambda/2k_1(\alpha)][A+O(\lambda)], \quad (5.8a)$$

where

$$A=\int_{\mathbb{R}\times[0,d]}V\chi_1^2\,dx\,dy. \quad (5.8b)$$

As mentioned above, a bound state corresponds to $\xi(\lambda)=-1$, i.e., $k_1(\alpha)=-\frac{1}{2}\lambda A+O(\lambda^2)$; it exists for small enough λ iff $k_1(\alpha)>0$, i.e., $A<0$. The corresponding energy is $E(\lambda)=\varkappa_1^2-k_1(\alpha)^2$; substituting for $k_1(\alpha)$, we arrive at the formula (5.2b). ■

Now we are ready to prove the following theorem.

Theorem 5.2: Let Ω be a curved Q-guide fulfilling the assumptions (a)–(e). Then a bound state $E\in[0,E_\infty)$ exists if

$$\int_{\mathbb{R}\times[0,d]}\left[-\frac{\gamma(s)^2}{(1+u\gamma(s))^2}+\frac{u^2\gamma'(s)^2}{(1+u\gamma(s))^4}\right]\times\sin^2\left(\frac{\pi u}{d}\right)ds\,du<0. \quad (5.9)$$

Proof: As in Theorem 4.1, we estimate H from above, now by the e.s.a. operator

$$H_+=-\frac{\hbar^2}{2m}\left(\frac{1}{(1+d\gamma_-)^2}\frac{d^2}{ds^2}+\frac{d^2}{du^2}\right)+V(s,u) \quad (5.10)$$

defined on $D_s\otimes D_u$. The potential V fulfills the required assumptions under (a)–(e), so Lemma 5.1 may be applied. Furthermore, $\gamma(s)\to0$ as $|s|\to\infty$, so the part containing the second derivative may be integrated by parts in s; it yields the condition (5.9). ■

As an illustration, let us return to the particular case treated in Proposition 4.2.

Corollary 5.3: For a simply bent Q-guide fulfilling (a)–(e), the critical width obeys

$$d_0\geqslant d_+\equiv(1/2\gamma_+)(\sqrt{1+4\gamma_+\sqrt{z}}-1), \quad (5.11a)$$

where

$$z=\left(\int_{\mathbb{R}}\gamma^2\,ds\right)^{-1}\int_{\mathbb{R}}\gamma^2\,ds. \quad (5.11b)$$

Proof: Denoting the integral of the square bracket in (5.9) over s as $I(u)$, we can write the condition as

$$\int_0^d I(u)\sin^2\left(\frac{\pi u}{d}\right)du<0.$$

Since $I(\cdot)$ is a smooth function, $I(0)=-\int_{\mathbb{R}}\gamma^2\,ds<0$, it is sufficient to find d_+ such that $I(u)\leqslant0$ for $u\in[0,d_+]$. For a simply bent Ω, we have

$$I(u)\leqslant-\frac{1}{(1+u\gamma_+)^2}\int_{\mathbb{R}}\gamma^2\,ds+u^2\int_{\mathbb{R}}\gamma'^2\,ds;$$

it yields an equation for d_+ which is solved by (5.11). ■

Example 5.4: (i)–(iii) The same as in Examples 4.3, (iv) $\gamma(s)=16\alpha\rho^3(s^2+4\rho^2)^{-2}$. More generally, the curves: $\gamma(s)=(\alpha/\rho)(\mu\rho)^{2n}(s^2+\mu^2\rho^2)^{-n}$ approach the Gaussian curve as $n\to\infty$; we plot the results for $n=1,2$ only.

The critical ratio $\alpha d_+/\rho$ in dependence on the overall bending is shown on Fig. 3. As in the previous section, we observe how the estimate spoils when γ changes steeply in some region.

VI. CONCLUSIONS

The most important question is whether the results are actually physically interesting. In order to observe bound states of this type, one must be able to fabricate sufficiently small curved Q-guides, because

(i) the size of the curved part must be small enough so that the free-particle model is applicable,

(ii) according to our results, the Q-guide thickness must be smaller than some critical value,

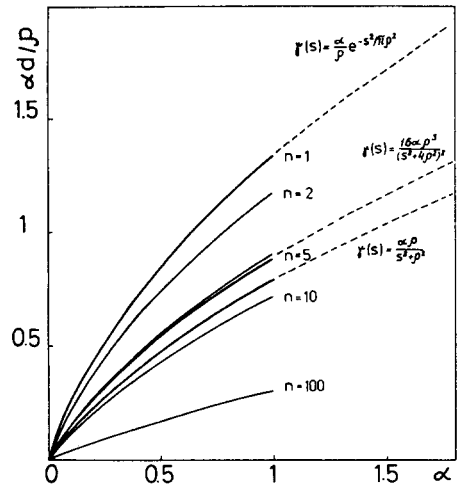


FIG. 3. The estimate (5.11).

(iii) except for that, the device must be sufficiently small to make the energy gap between the bound state and the bottom of the essential spectrum measurable (cf. Proposition 3.2).

The first requirement means[2] that the size must be $\lesssim 10^3$Å, which is achievable with today's technologies. The characteristic thickness of thin films or quantum wires is of the order of 10^2 Å, and therefore our examples suggest that such bound states may exist. The requirement (iii) will also be fulfilled if the gap is not substantially smaller than the first transversal-mode energy itself; recall that the transversal-mode energy intervals in thin semiconductor films are actually measured[17] being of the order of $(10^{-2}$–$10^{-1})$ eV.

It would certainly be useful to know the values of bound-state energies. For a rectangular Q-guide, this problem will be solved by another method in a separate publication.[18] In a general case, one might try to find at least the threshold behavior of the bound states. For a Q-guide with a smooth boundary and the thickness slightly less than critical, a natural guess is motivated by the formula (5.2b)

$$E \approx \frac{\hbar^2\pi^2}{2md^2}\left\{1-\left[(2\pi)^{-1}\int_{\mathbb{R}\times[0,d]}\left(\frac{\gamma^2}{(1+u\gamma)^2}-\frac{\gamma'^2}{(1+u\gamma)^4}\right)\sin^2\left(\frac{\pi u}{d}\right)ds\,du\right]^2\right\}. \tag{6.1}$$

We have no estimates, however, for the remainder term in such an expansion; the standard coupling-constant-threshold arguments do not work here, because the potential is not linear with respect to $(d-d_0)$.

Our last remark concerns the case of periodically curved Q-guides which is physically very interesting; it can model, e.g., a thin film over a periodically scratched substrate. Motivated by the above considerations, we formulate the following conjecture.

Conjecture 6.1: Let Ω be a planar Q-guide fulfilling the assumptions (a)–(c). The spectrum of H is absolutely continuous; for all sufficiently small d it contains at least one band below the first transversal-mode energy.

Let us mention that similar effects are known in the classical waveguide theory too[19]; in the quantum case, however, such a band should exist even if the waveguide cross section remains constant.

Note added in proof: We have learned recently that efforts similar to those described in this paper were experimentally meausred.[20]

ACKNOWLEDGMENTS

The authors are indebted to Ya. A. Smorodinsky, P. Šťovíček, and J. Tolar for stimulating discussions.

[1]H. Sakaki, in *Proceedings of the International Symposium on Foundations of Quantum Mechanics in the Light of New Technology*, edited by S. Kamefuchi *et al.* (Physical Society of Japan, Tokyo, 1984), pp. 94–110.
[2]This is clear for metals. In an ideally clean semiconductor, the electron would move as a free particle with some effective mass, which may differ substantially from the actual electron mass. The mean free path is finite, of course, due to impurities of the material. It can be made, however, of the order of μm, i.e., comparable with the size of modern microelectronic devices.
[3]The results cannot be modified directly for the Neumann case, since in the "straightening" transformation (3.4) the Neumann condition on Ω is not transformed into the Neumann condition on $\mathbb{R}\times[0,d]$.
[4]It is possible, at least in principle, to coil a quantum wire around a thin cylindrical rod.
[5]J. Marcus, J. Chem. Phys. **45**, 4493 (1966).
[6]H. Koppe and H. Jensen, Sitzungber. Heidelberger Akad. Wiss., p. 127 (1971); H. Jensen and H. Koppe, Ann. Phys. **63**, 586 (1971).
[7]D. Picca, Lett. Nuovo Cimento **34**, 449 (1982).
[8]R. C. T. da Costa, Phys. Rev. A **23**, 1982 (1981); R. C. T. da Costa, Phys. Rev. A. **25**, 2893 (1982); R. C. T. da Costa, Lett. Nuovo Cimento **36**, 393 (1983).
[9]The potential is proportional to $\hbar^2$ so the corresponding effects are purely quantal. The result has an important philosophical aspect, namely that the behavior of a constrained quantum particle can be influenced substantially by the geometrical properties of its configuration manifold in a way that no quantization procedure can be predicted.
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, IV. Analysis of Operators* (Academic, New York, 1978), Sec. XIII. 5.
[11]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966), Sec. 6.2.1.
[12]See Ref. 10, Sec. XIII.1.
[13]See Ref. 10, Sec. XIII.4.
[14]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, I. Functional Analysis* (Academic, New York, 1972), Theorem VIII.33.
[15]B. Simon, Ann. Phys. **97**, 279 (1976); R. Blanckenbecler, M. L. Goldberger, and B. Simon, Ann. Phys. **108**, 69 (1977).
[16]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971), Sec. III.2.
[17]A. Harwitt and J. S. Harris, Appl. Phys. Lett. **50**, 685 (1987).
[18]P. Exner, P. Šeba, and P. Šťovíček (in preparation).
[19]A. Boström, IEEE Trans. Microwave Theory Tech. **31**, 752 (1983).
[20]G. Timp *et al.*, Phys. Rev. Lett. **60**, 2081 (1988).

Scattering of charged particles

R. O. Barrachina[a] and J. H. Macek
University of Tennessee, Knoxville, Tennessee 37996-1200 and Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

(Received 4 October 1988; accepted for publication 28 June 1989)

Different methods of avoiding the known difficulties of the Coulomb potential scattering theory are reviewed. Mulherin and Zinnes' [J. Math. Phys. **11**, 1402 (1976)] "distorted" free waves and van Haeringen's [J. Math. Phys. **17**, 995 (1976)] Coulomb asymptotic states are considered. The equivalence of both approaches on the energy shell is shown. Actually the possibility of deriving the first method within van Haeringen's formalism by means of a distorted wave procedure is demonstrated.

I. INTRODUCTION

The scattering theory of charged particles does not fit into formal scattering theory.[1] Even the simplest case of two structureless particles that interact via a Coulomb potential $V(r) = \alpha/r$ is exceptional in many different ways: Although the Schrödinger equation can be solved explicitly for this system, and a cross section extracted using a time independent analysis, there are important difficulties that stem from the infinite range of the potential in a conventional time-dependent approach and in the Lippman–Schwinger time-independent theory. Actually, the Coulomb potential falls off too slowly at large distances to satisfy the asymptotic conditions of the usual scattering theory. Since the pioneering work by Dollard,[2] many different ways of circumventing this difficulty have been proposed. The formalism of Dettmann[1] employs wave packets and the Lippman–Schwinger equation to obtain a convergent expression for the transition probability even though the transition matrix element itself diverges logarithmically. The formalism of Dollard[2] and of Belkic–Gayet–Salin[3] seeks a convergent expression for the T-matrix element by including the long range distortion due to the Coulomb potential. The resulting expression for the T matrix employing these "Coulomb asymptotic conditions" has the distorted wave form and has been treated in the literature as a particular type of distorted wave amplitude.[4] Van Haeringen[5] also includes the effect of the long range Coulomb potential but does so in a way that avoids a distorting potential yet obtains a convergent expression for the T matrix. In Mulherin and Zinnes' paper[6] a Coulomb potential scattering theory is developed by introducing "distorted" asymptotic states of the eikonal form. However, it has been usually stressed that this procedure does not represent a distorted wave formalism,[7] but that the eikonal phases represent a basic requirement that the wave functions in a time-independent scattering theory of charged particles should meet. In van Haeringen's procedure the scattering theory of charged particles is developed using an alternate generalization of the usual asymptotic free states. With these asymptotic waves the time-independent scattering theory is formallly reconstructed[5] by a limiting process.

This work shows that the alternative asymptotic states are not contradictory, but represent equivalent translations of the same idea. We show by explicit evaluation that both procedures give the well-known Coulomb scattering amplitude. We further show that the expression for the T matrix in the representation of Mulherin and Zinnes derives from that of van Haeringen. Actually, we will demonstrate that Mulherin and Zinnes' asymptotic states *do* represent distorted waves in the sense of conventional distorted wave theory applied to van Haeringen's more fundamental Coulomb asymptotic states.

II. COULOMB ASYMPTOTIC STATES

In the time-dependent theory of potential scattering the essential properties of the collision are associated with the isometric Moller operators

$$\Omega \pm = \underset{t \to \mp \infty}{\text{s-lim}} \exp[(i/\hbar)\mathbb{H}t]\exp[-(i/\hbar)\mathbb{H}_0 t]\,. \quad (1)$$

Here $\mathbb{H} = \mathbb{H}_0 + V$ and $\mathbb{H}_0 = -\hbar^2\Delta/2m$ are the full and free Hamiltonians, respectively. However, when a Coulomb interaction is involved, the strong limit (1) does not exist. The reason is that the Coulomb potential $V(r) = \alpha/r$ falls off so slowly as $r \to \infty$ that it continues to influence the particles even as they move far apart. According to Dollard,[2] a "renormalization" term has to be included in Eq. (1) as follows:

$$\Omega^c \pm = \underset{t \to \mp \infty}{\text{s-lim}} \exp[(i/\hbar)\mathbb{H}t]\exp\left[-(i/\hbar)(\mathbb{H}_0 t \pm \alpha\sqrt{m/2}\,\mathbb{H}_0^{-1/2}\log(4\mathbb{H}_0|t|/\hbar))\right]. \quad (2)$$

Here m is the reduced mass of the interacting particles. This limit exists and formally defines the improper Coulomb stationary scattering states

$$|\mathbf{k} \pm\rangle = \Omega^c \pm |\mathbf{k}\rangle\,. \quad (3)$$

These states are eigenvectors of the full Hamiltonian $\mathbb{H}$,

$$\mathbb{H}|\mathbf{k} \pm\rangle = (k^2/2m)|\mathbf{k} \pm\rangle\,. \quad (4)$$

In coordinate space they are given by

$$\langle \mathbf{r}|\mathbf{k} \pm\rangle = h^{-3/2}\Gamma(1 \pm in)\exp(-\pi n/2)\exp((i/\hbar)\mathbf{k}\cdot\mathbf{r}) \times {}_1F_1[\mp in, 1, \pm(i/\hbar)(kr \mp \mathbf{k}\cdot\mathbf{r})]\,, \quad (5)$$

Here n is Sommerfeld's parameter $n = m\alpha/hk$. On the other hand, the momentum representation can be evaluated in terms of the matrix element $\langle \mathbf{p}|V|\mathbf{k} \pm\rangle$

[a] Permanent address: Centro Atómico Bariloche, Comisión Nacional de Energía Atómica, 8400 Bariloche, Río Negro, Argentina.

 0022-2488/89/112581-04$02.50

$$\langle \mathbf{p}|\mathbf{k}\pm\rangle = h^{-3/2}\int \exp\left(-\frac{i}{\hbar}\mathbf{p}\cdot\mathbf{r}\right)\langle \mathbf{r}|\mathbf{k}\pm\rangle d\mathbf{r}$$

$$= -\frac{\hbar}{\alpha}\frac{d}{d\epsilon}h^{-3/2}\int \exp\left(-\frac{i}{\hbar}\mathbf{p}\cdot\mathbf{r}\right)\frac{\alpha}{r}$$

$$\times e^{-\epsilon r/\hbar}\langle \mathbf{r}|\mathbf{k}\pm\rangle d\mathbf{r}$$

$$= -\frac{\hbar}{\alpha}\frac{d}{d\epsilon}\langle \mathbf{p}|V\epsilon|\mathbf{k}\pm\rangle , \tag{6}$$

where the limit $\epsilon\to 0$ is understood.

Now $\langle \mathbf{p}|V\epsilon||\mathbf{k}\pm\rangle$ can be evaluated by means of Nordsieck's[8] method;

$$\langle \mathbf{p}|V\epsilon|\mathbf{k}\pm\rangle = \frac{\alpha}{\pi h}\frac{\Gamma(1\pm in)}{e^{\pi n/2}}\frac{[p^2-(k\pm i\epsilon)^2]^{\pm in}}{[|\mathbf{p}-\mathbf{k}|^2+\epsilon^2]^{\pm in+1}} . \tag{7}$$

Therefore,

$$\langle \mathbf{p}|\mathbf{k}\pm\rangle = \frac{\epsilon\Gamma(2\pm in)}{\pi^2 e^{\pi n/2}}\frac{[p^2-(k\pm i\epsilon)^2]^{\pm in}}{[|\mathbf{p}-\mathbf{k}|^2+\epsilon^2]^{2\pm in}}$$

$$-\frac{n}{\pi^2}\frac{\Gamma(1\pm in)}{e^{\pi n/2}}(k\pm i\epsilon)$$

$$\times\frac{[p^2-(k\pm i\epsilon)^2]^{-1\pm in}}{[|\mathbf{p}-\mathbf{k}|^2+\epsilon^2]^{1\pm in}} . \tag{8}$$

Following van Haeringen,[5] we note that the second term on the right-hand side is just

$$\langle \mathbf{p}|\mathbb{G}_0((k\pm i\epsilon)^2/2m)V\epsilon|\mathbf{k}\pm\rangle .$$

Therefore we have obtained a Lippmann–Schwinger (LS) equation for $|\mathbf{k}\pm\rangle$

$$|\mathbf{k}\pm\rangle = |\mathbf{k}\pm\epsilon\rangle + \mathbb{G}_0((k\pm i\epsilon)^2/2m)V\epsilon|\mathbf{k}\pm\rangle , \tag{9}$$

which looks very similar to the usual LS equation except for a Coulomb-modified improper free state,

$$\langle \mathbf{p}|\mathbf{k}\pm\epsilon\rangle = \frac{\epsilon\Gamma(2\pm in)}{\pi^2 e^{\pi n/2}}\frac{[p^2-(k\pm i\epsilon)^2]^{\pm in}}{[|\mathbf{p}-\mathbf{k}|^2+\epsilon^2]^{2\pm in}} , \tag{10}$$

taking the place of the free particle plane waves. This Coulomb asymptotic state (CAS) represents a generalization of the usual free asymptotic state for the case of Coulomb interactions. Within this procedure the on-shell transition matrix is evaluated between Coulomb asymptotic states[5] to obtain

$$\langle \mathbf{p}-\epsilon|\mathbb{T}\left(\frac{(k+i\epsilon)^2}{2m}\right)|\mathbf{k}+\epsilon\rangle$$

$$= \langle \mathbf{p}-\epsilon|V\epsilon + V\epsilon\mathbb{G}_c\left(\frac{(k+i\epsilon)^2}{2m}\right)V\epsilon|\mathbf{k}+\epsilon\rangle$$

$$= \langle \mathbf{p}-\epsilon|V\epsilon|\mathbf{k}+\rangle$$

$$= \frac{\alpha}{\pi h}\frac{\Gamma(1\pm in)}{\Gamma(1\mp in)}\frac{1}{|\mathbf{p}-\mathbf{k}|^2}\left[\frac{4k^2}{|\mathbf{p}-\mathbf{k}|^2}\right]^{\pm in} , \tag{11}$$

which is the proper Coulomb amplitude.

III. MULHERIN AND ZINNES' ASYMPTOTIC STATES

Mulherin and Zinnes' formulation of Dollards' theory employs a particular form of asymptotic waves deduced from the behavior of Coulomb waves in coordinate space. The asymptotic form of the Coulomb stationary state $\langle \mathbf{r}|\mathbf{k}\pm\rangle$, given by Eq. (5), through terms of order $\hbar/(kr\mp\mathbf{k}\cdot\mathbf{r})$ is

$$\langle \mathbf{r}|\mathbf{k}\pm\rangle \approx h^{-3/2}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{r}}\left[\frac{kr\mp\mathbf{k}\cdot\mathbf{r}}{\hbar}\right]^{\pm in}$$

$$+ h^{-3/2}ne^{\pm(i/\hbar)kr}\frac{\Gamma(1\pm in)}{\Gamma(1\mp in)}$$

$$\times\left[\frac{kr\mp\mathbf{k}\cdot\mathbf{r}}{\hbar}\right]^{-(1\pm in)} . \tag{12}$$

By analogy with the wave function for short-range potentials, the asymptotic form of $\langle \mathbf{r}|\mathbf{k}+\rangle$ is interpreted as an "incident" plane wave plus a spherically spreading scattered wave. On this basis Mulherin and Zinnes introduce the asymptotic states

$$\langle \mathbf{r}|\mathbf{k}\infty\pm\rangle \approx h^{-3/2}e^{(i/\hbar)\mathbf{k}\cdot\mathbf{r}}\left[\frac{kr\mp\mathbf{k}\cdot\mathbf{r}}{\hbar}\right]^{\pm in} . \tag{13}$$

These states are eigenvectors of a modified "free" Hamiltonian $\mathbb{H}_\infty = \mathbb{H}_0 + \mathbb{U}$,

$$\mathbb{H}_\infty|\mathbf{k}\infty\pm\rangle = k^2/2m|\mathbf{k}\infty\pm\rangle , \tag{14}$$

where the action of the distortion potential $\mathbb{U}$ on the asymptotic state $|\mathbf{k}\infty\pm\rangle$ is defined by

$$\langle \mathbf{r}|\mathbb{U}|\mathbf{k}\infty\pm\rangle$$

$$= \langle \mathbf{r}|\mathbb{H}_\infty - \mathbb{H}_0|\mathbf{k}\infty\pm\rangle$$

$$= (k^2/2m + \hbar^2\Delta/2m)\langle \mathbf{r}|\mathbf{k}\infty\pm\rangle$$

$$= (\alpha/r)[1-(\hbar n/(kr\mp\mathbf{k}\cdot\mathbf{r}))]\langle \mathbf{r}|\mathbf{k}\infty\pm\rangle . \tag{15}$$

These asymptotic states satisfy the following Lippmann–Schwinger equation:

$$|\mathbf{k}\pm\rangle = |\mathbf{k}\infty\pm\rangle$$

$$+ \mathbb{G}((k\pm i\epsilon)^2/2m)(V-\mathbb{U})|\mathbf{k}\infty\pm\rangle . \tag{16}$$

Now the time-independent scattering theory can be formally reconstructed if the phases introduced by the Coulomb potential are retained in the asymptotic states, and the potential and "free" Hamiltonian are consistently distorted. For instance, the on-shell $\mathbb{T}$-matrix element reads

$$\langle \mathbf{p}\infty-|\mathbb{T}_0((k+i\epsilon)^2/2m)|\mathbf{k}\infty+\rangle = \langle \mathbf{p}\infty-|(V-\mathbb{U})+(V-\mathbb{U})\mathbb{G}((k+i\epsilon)^2/2m)(V-\mathbb{U})|\mathbf{k}\infty+\rangle$$

$$= \langle \mathbf{p}\infty-|(V-\mathbb{U})|\mathbf{k}+\rangle .$$

The matrix element is evaluated using the coordinate space representation:

$$\langle \mathbf{p}\infty\mp|(V-\mathbb{U})|\mathbf{k}\pm\rangle = nph^{-3/2}\int\left[\frac{pr\mp\mathbf{p}\cdot\mathbf{r}}{\hbar}\right]^{\pm inp-1}\exp\left(-\frac{i}{\hbar}\mathbf{p}\cdot\mathbf{r}\right)\frac{\alpha}{r}e^{-\epsilon r/\hbar}\langle \mathbf{r}|\mathbf{k}\pm\rangle d\mathbf{r}.$$

By using an integral representation[9] of $[(pr \mp \mathbf{p}\cdot\mathbf{r})/\hbar]^{\pm inp-1}$ we obtain

$$\langle \mathbf{p}\infty \mp |(V-U)|\mathbf{k}\pm\rangle = \frac{n_p}{\Gamma(1\mp inp)}\int_0^\infty \lambda^{\mp inp} h^{-3/2}\int e^{-(i/\hbar)(1\mp i\lambda)\mathbf{p}\cdot\mathbf{r}}\frac{\alpha}{r}e^{-(\epsilon+p\lambda)r/\hbar}\langle \mathbf{r}|\mathbf{k}\pm\rangle d\mathbf{r}\, d\lambda$$

$$= \frac{np\alpha}{\Gamma(1\mp inp)}\int_0^\infty \lambda^{\mp inp}\frac{1}{\pi h}\frac{\Gamma(1\pm in_k)}{e^{(\pi/2)n_k}k^2}\frac{[p^2-(k\pm i\epsilon)^2\mp 2ip(p+k\pm i\epsilon)\lambda]^{\pm in_k}}{\{(\mathbf{p}-\mathbf{k})^2+\epsilon^2+2[\epsilon p\mp i\mathbf{p}\cdot(\mathbf{p}-\mathbf{k})]\lambda\}^{1\pm in_k}}\,d\lambda$$

$$= \frac{\alpha}{\pi h}\Gamma(1\pm in_k)\Gamma(1\pm in_p)[2\mathbf{p}\cdot(\mathbf{p}-\mathbf{k})\pm 2i\epsilon p]^{-1}\left[\frac{2p(p+k\pm i\epsilon)}{(\mathbf{p}-\mathbf{k})^2+\epsilon^2}\right]^{\pm in_k}$$

$$\times\left[\frac{2\epsilon p\mp 2i\mathbf{p}\cdot(\mathbf{p}-\mathbf{k})}{(\mathbf{p}-\mathbf{k})^2+\epsilon^2}\right]^{\pm i(n_p-n_k)}{}_2F_1\left[\mp in_k,\pm in_p;1;1-\left(\frac{p-k\mp i\epsilon}{2p}\right)\left(\frac{2\mathbf{p}\cdot(\mathbf{p}-\mathbf{k})\pm 2i\epsilon p}{(\mathbf{p}-\mathbf{k})^2+\epsilon^2}\right)\right].$$

Upon taking the limits as $\epsilon\to 0$ and $p\to k$ we obtain the on-shell amplitude

$$\langle \mathbf{p}\infty \mp |(V-U)\mathbf{k}\pm\rangle = \frac{\alpha}{\pi h}\frac{\Gamma(1\pm in)}{\Gamma(1\mp in)}\frac{1}{|\mathbf{p}-\mathbf{k}|^2}\left[\frac{4k^2}{|\mathbf{p}-\mathbf{k}|^2}\right]^{\pm in}, \tag{17}$$

which, again, is the correct Coulomb amplitude. We see that within this prescription the $\mathbb{T}$-matrix element is evaluated using "distorted" waves rather than plane waves. However, it has been usually stressed[7] that this procedure does not represent a distorted wave formalism, but a consistent way of avoiding the known difficulties of the Coulomb potential, i.e., the plane wave, is modified only by a phase factor. However, the potentials must also be consistently modified; thus as in Eq. (17) we may regard Eq. (12) as a particular form of distorted wave.

IV. RELATION BETWEEN THE TWO APPROACHES

In order to examine the relation between the Mulherin and Zinnes asymptotic states and the CAS, we evaluate the Fourier transform of Eq. (13),

$$\langle \mathbf{p}|\mathbf{k}\infty\pm\rangle = h^{-3/2}\int e^{-(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\langle \mathbf{r}|\mathbf{k}\infty\pm\rangle d\mathbf{r}$$

$$= -\hbar\frac{d}{d\epsilon}h^{-3/2}\int \exp\left(-\frac{i}{\hbar}\mathbf{p}\cdot\mathbf{r}\right)\frac{1}{r}e^{-\epsilon r/\hbar}\langle \mathbf{r}|\mathbf{k}\infty\pm\rangle d\mathbf{r}$$

$$= -\hbar\frac{d}{d\epsilon}h^{-3}\int\frac{1}{r}e^{-(i/\hbar)(\mathbf{k}-\mathbf{p}+i\epsilon\hat{r})\cdot\mathbf{r}}\left[\frac{kr\mp\mathbf{k}\cdot\mathbf{r}}{\hbar}\right]^{\pm in}d\mathbf{r}$$

$$= -\frac{1}{\pi}\frac{d}{d\epsilon}\Gamma(1\pm in)2^{\pm in}\frac{[\epsilon k\mp i\mathbf{k}\cdot(\mathbf{k}-\mathbf{p})]^{\pm in}}{[|\mathbf{k}-\mathbf{p}|^2+\epsilon^2]^{\pm in}}$$

$$= \frac{\epsilon}{\pi^2}\Gamma(2\pm in)\frac{[2\epsilon k\mp i2\mathbf{k}\cdot(\mathbf{k}-\mathbf{p})]^{\pm in}}{[|\mathbf{k}-\mathbf{p}|^2+\epsilon^2]^{2\pm in}}\mp i\frac{nk}{\pi^2}\Gamma(1\pm in)\frac{[2\epsilon k\mp i2\mathbf{k}\cdot(\mathbf{k}-\mathbf{p})]^{\pm in-1}}{[|\mathbf{k}-\mathbf{p}|^2+\epsilon^2]^{1\pm in}}. \tag{18}$$

As $\epsilon\to 0$ the first term on the right-hand side of Eq. (18) becomes a distribution function with support on $\mathbf{p}=\mathbf{k}$, which is precisely the Coulomb asymptotic state CAS. On the other hand, the second term is easily shown to equal

$$\langle \mathbf{p}|\mathbb{G}_0((k\pm i\epsilon)^2/2m)\mathbb{U}|\mathbf{k}\pm\rangle.$$

Therefore a Lippmann–Schwinger equation for $|\mathbf{p}\infty\pm\rangle$ is obtained,

$$|\mathbf{k}\infty\pm\rangle = |\mathbf{k}\pm\epsilon\rangle + \mathbb{G}_0((k\pm i\epsilon)^2/2m)\mathbb{U}|\mathbf{k}\infty\pm\rangle. \tag{19}$$

The LS equation (19) lets us rewrite the $\mathbb{T}$-matrix element $\langle \mathbf{p}\mp\epsilon|V\epsilon|\mathbf{k}\pm\rangle$ in the following form:

$$\langle \mathbf{p}\mp\epsilon|V\epsilon|\mathbf{k}\pm\rangle = \langle \mathbf{p}\infty\mp|V\epsilon|\mathbf{k}\pm\rangle - \langle \mathbf{p}\infty\mp|\mathbb{U}\mathbb{G}_0((k\pm i\epsilon)^2/2m)V|\mathbf{k}\pm\rangle. \tag{20}$$

By using the Lippmann–Schwinger equation (9) we rewrite the rhs of Eq. (20);

$$\langle \mathbf{p}\mp\epsilon|V\epsilon|\mathbf{k}\pm\rangle = \langle \mathbf{p}\infty\mp|(V\epsilon-\mathbb{U})|\mathbf{k}\pm\rangle + \langle \mathbf{p}\infty\mp|\mathbb{U}|\mathbf{k}\pm\epsilon\rangle. \tag{21}$$

This is the standard two-potential formula of the distorted-wave theory.[10] The second term on the rhs of Eq. (21) vanishes on the energy shell in the limit as $\epsilon\to 0$. Actually

$$\langle \mathbf{p}\infty\mp|\mathbb{U}|\mathbf{k}\pm\epsilon\rangle = \int\langle \mathbf{p}\infty\mp|\mathbb{U}|\mathbf{q}\rangle\cdot\langle \mathbf{q}|\mathbf{k}\pm\epsilon\rangle d\mathbf{q} = \int\langle \mathbf{p}\infty\mp|\mathbb{U}|\mathbf{q}\rangle\frac{\epsilon}{\pi^2}\frac{\Gamma(2\pm in)}{e^{\pi n/2}}\frac{[q^2-(k\pm i\epsilon)^2]^{\pm in}}{[|\mathbf{q}-\mathbf{k}|^2+\epsilon^2]^{2\pm in}}d\mathbf{q} \tag{22}$$

is absolutely dominated by

$$|\langle \mathbf{p}\infty \mp |\mathbb{U}|\mathbf{k}\pm\epsilon\rangle| \leqslant \frac{\epsilon}{\pi^2}\frac{\Gamma(2\pm in)}{e^{\pi n/2}}\int |\langle \mathbf{p}\infty \mp |\mathbb{U}|\mathbf{q}\rangle| \frac{1}{[|\mathbf{q}-\mathbf{k}|^2+\epsilon^2]^2}\,d\mathbf{q} = \frac{1}{\pi^2}\frac{|\Gamma(2\pm in)|}{e^{\pi n/2}}\int_0^\infty \frac{z^2}{[z^2+1]^2}\xi(\epsilon z)dz \tag{23}$$

with

$$\xi(\mathbf{q}) = \int |\langle \mathbf{p}\infty \mp |\mathbb{U}|\mathbf{q}+\mathbf{k}\rangle| d\hat{q} = \frac{\alpha\Gamma(1\pm in)}{\pi h e^{-\pi n/2}} q \int \frac{q+2\hat{q}\cdot\hat{k}}{|\mathbf{p}-\mathbf{q}-\mathbf{k}|^2+\epsilon^2}\frac{1}{|2\epsilon k + 2i\mathbf{p}\cdot(\mathbf{p}-\mathbf{k}-\mathbf{q})|}\,d\mathbf{q}. \tag{24}$$

We see that for $\epsilon\to 0$ the integrand of Eq. (23) vanishes pointwise. Furthermore it is bounded by the integrable function $\|\xi\|z^2/(z^2+1)^2$. Therefore the matrix element $\langle \mathbf{p}\infty \mp |\mathbb{U}|\mathbf{k}\pm\epsilon\rangle$ also vanishes in the limit as $\epsilon\to 0$.

V. CONCLUSIONS

We have shown that Mulherin and Zinnes' asymptotic states can be obtained within van Haeringen's formalism by means of a simple distorted wave procedure: Both approaches are equivalent. Nevertheless, more fundamental issues concerning the relationship between these Coulomb time-independent schemes and experimental cross sections remain to be investigated.

When the Moller operator $\Omega^{\pm}$ in Eq. (1) exists, it can be applied to a state $|\Phi\rangle$ to yield $|\Phi\pm\rangle = \Omega^{\pm}|\Phi\rangle$. Here, $|\Phi\pm\rangle$ is to be thought of as describing the actual state at $t=0$ which evolves from (or to) the freely moving prepared (detected) asymptotic state $\exp[-(i/\hbar)\mathbb{H}_0 t]|\Phi\rangle$, in an idealization of a real scattering experiment. On the other hand, when a Coulomb potential is involved, the state[6,7]

$$\exp\left[-\frac{i}{\hbar}\mathbb{H}_\infty t\right]\Phi = h^{-3/2}\int e^{(i/\hbar)\mathbf{k}\cdot\mathbf{r}}\left[\frac{kr\mp\mathbf{k}\cdot\mathbf{r}}{\hbar}\right]^{\pm in} \times e^{(ik^2/2m)/\hbar}\tilde{\Phi}(\mathbf{k})d\mathbf{k} \tag{25}$$

does not have a straightforward physical interpretation. Actually $\exp[-(i/\hbar)\mathbb{H}_\infty t]$ is not an isometry.[11] Thus the usual identification of such a state at large times with the prepared (or detected) beam of projectiles is not obvious. Further discussion would be needed to relate the time-independent schemes discussed here to actual experiments when Coulomb interactions are involved.[1]

ACKNOWLEDGMENTS

Support for this research by the National Science Foundation under NSF Grant No. PHYS-8602988 is gratefully acknowledged.

[1] K. Dettmann, Z. Phys. **244**, 86 (1971).
[2] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[3] Dz. Belkic, R. Gayet, and A. Salin, Phys. Rep. **56**, 279 (1979).
[4] J. Macek, Phys. Rev. A **37**, 2365 (1988).
[5] H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[6] D. Mulherin and I. I. Zinnes, J. Math. Phys. **11**, 1402 (1970).
[7] R. Gayet, Proceedings of the VIII ICPEAC, Belgrade, 1973, edited by B. Cobic and M. V. Kupera (Institute of Physics, Belgrade, 1973), p. 237.
[8] A. Nordsieck, Phys. Rev. **93**, 785 (1954).
[9] J. N. Gau and J. Macek, Phys. Rev. A **10**, 522 (1974).
[10] J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[11] J. D. Dollard, J. Math. Phys. **16**, 391 (1975).

On modeling discontinuous media. Three-dimensional scattering

P. C. Sabatier
Laboratoire de Physique Mathématique, Université des Sciences et Techniques du Languedoc, 34060 Montpellier Cedex, France

(Received 3 March 1989; accepted for publication 24 May 1989)

A realistic model for the propagation of scalar waves in a medium must take into account discontinuities of the medium parameters, which may act as "hard" reflectors; discontinuities of their gradient, which may act as "soft" reflectors; and continuous variations of parameters. The direct scattering problem is presented here. The "mixed potential–impedance" equation that characterizes the model depends on two arbitrary parameters: one of them would be the potential if the equation reduces to the Schrödinger equation; the other one would be related to the impedance if the equation describes more "classical" waves. After the mathematical tools are constructed (Green's functions, etc.), a rigorous three-dimensional scattering theory is described. It encompasses the quantum three-dimensional scattering theory and the theory of acoustical scattering (for instance) by systems of regular surfaces of arbitrary shape. The main integral equations of the quantum scattering theory are generalized. Scattering amplitudes due to reflectors and scattering amplitudes due to diffuse scattering after reflectors have been taken into account are defined and constructed. Born and quadratic approximations are discussed: the explicit formulas corresponding to the scattering by discontinuities and those corresponding to diffuse scattering are not reducible to each other exactly (i.e., unless filtering and errors are allowed). The results can also be used to described rigorously the three-dimensional scattering by the "wave-equation" in the frequency domain—and in particular the response to an impulsive localized source. Further generalizations are in progress.

I. INTRODUCTION

A number of problems arise in the study of wave propagation through discontinuous media. We studied them in the one-dimensional case[1] on a model equation, the impedance equation, and a slightly generalized one, the mixed (impedance–potential) equation. The three-dimensional mixed (impedance–potential) equation is studied in the present paper. It bridges the gap between quantum wave propagation and acoustical wave propagation. The price to be paid is of a complicated mathematical structure, which is here fully analyzed with working assumptions sufficient for all physical purposes. A complete scattering theory is provided, together with the standard Born and second-order approximations.

Throughout the paper, we deal with a set of ordered domains Ω_i and surfaces S_i of class C^2, with $S = U_i S_i$, such that each domain is finite but Ω_{N+1}, which extends to infinity in all directions, and $i \neq j \Rightarrow S_i \cap S_j = \emptyset$. The "topological aspect" of such surfaces is that of spherelike surfaces, each S_i enclosing a ball-like domain

$$\overline{D}_i = \sum_{k=0}^{k=1} \overline{\Omega}_k$$

(see Fig. 1). There is one unit normal vector at each point of S_i: $\nu(x)$ pointing into the $+$ side, which for S_i is the side of Ω_{i+1}. We are interested in the equation

$$[\alpha^{-2} \operatorname{div} \alpha^2 \operatorname{grad} + k^2 - V]\varphi(\mathbf{k},x) = 0, \quad x \in \mathbb{R}^3, \tag{1.1}$$

where $V(x)$ is a real function, belonging to class $\mathscr{V}$ of "potentials" for which the scattering problem described by (1.1) in the case $\alpha = 1$ is meaningful—for instance, the Rollnick class:

$$\int dr \int dr' |V(r)V(r')|\,|r-r'|^{-2} < \infty, \tag{1.2}$$

$\mathbf{k},x \in \mathbb{R}^3$, $k^2 = |\mathbf{k}\cdot\mathbf{k}|$, the impedance factor α is everywhere twice differentiable with $\alpha^{-1}\Delta\alpha \in \mathscr{V}$, except at the surfaces $S_0, S_1, \ldots, S_N$. At a point x of a surface S_i, $\alpha(x)$ and $\nu\cdot\operatorname{grad}\alpha(x)$ may be undefined but their limits from the $(+)$ and $(-)$ sides are defined and described by the "transmission factor" t, the "reflection factor" r, and the "slope factor" $\bar{s}$:

$$[t(x)]^{-1} = \frac{1}{2}\left[\frac{\alpha(x^+)}{\alpha(x^-)} + \frac{\alpha(x^-)}{\alpha(x^+)}\right], \tag{1.3a}$$

$$\frac{r(x)}{t(x)} = \frac{1}{2}\left[\frac{\alpha(x^+)}{\alpha(x^-)} - \frac{\alpha(x^-)}{\alpha(x^+)}\right], \tag{1.3b}$$

$$\frac{\bar{s}(x)}{t(x)} = \frac{1}{2}\,\nu\cdot\left[\frac{\operatorname{grad}\alpha(x^-)}{\alpha(x^+)} - \frac{\operatorname{grad}\alpha(x^+)}{\alpha(x^-)}\right]. \tag{1.3c}$$

These values are called the problem "singular data."

The first results we obtained in this three-dimensional case[2] are summarized by the following theorem.

Theorem I: Let $\beta(x) > 0$, such that $\alpha(x)/\beta(x)$ is everywhere continuous, $\beta(x)$ is twice differentiable at each regular point, and

$$\tfrac{1}{2}[(\alpha \operatorname{grad}\beta - \beta \operatorname{grad}\alpha)\cdot\nu]_{x^-}^{x^+} = \alpha(x^-)\beta(x^+)p(x). \tag{1.4}$$

It can be proved[1,2] that the function $\Phi = \alpha\varphi/\beta$ satisfies the equation

$$[\beta^{-2}\operatorname{div}\beta^2\operatorname{grad} + k^2 - V + (\alpha\beta)^{-1}\operatorname{div}(\alpha\operatorname{grad}\beta - \beta\operatorname{grad}\alpha)]\Phi = 0, \tag{1.5}$$

at each regular point, and that its continuity conditions

 0022-2488/89/112585-14$02.50

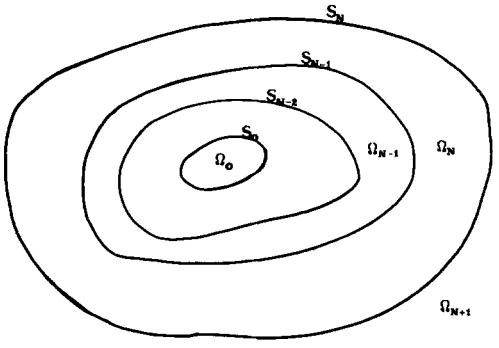


FIG. 1.

across a singular surface at x are

$$\Phi(x^+) = \Phi(x^-), \tag{1.6a}$$

$$[\nu\cdot\beta^2 \operatorname{grad}\Phi]_{x^-}^{x^+} = -2p(x)\Phi(x)\beta(x^-)\beta(x^+). \tag{1.6b}$$

The impedance factors α and β have the same transmission and reflection factors but not the same slope factor:

$$\bar{s}_\beta(x) - \bar{s}_\alpha(x) + p(x)t(x) = 0. \tag{1.7}$$

Corollaries of Theorem I are the two equivalence rules.

Standard equivalence: Assume

$$\begin{aligned} &\beta\in C^2(\mathbb{R}^3\backslash S), \qquad \alpha/\beta\in C(\mathbb{R}^3),\\ &p(x) = 0, \qquad x\in S,\\ &\alpha\Delta\beta - \beta\Delta\alpha = 0, \quad x\in\mathbb{R}^3\backslash S. \end{aligned} \tag{1.8}$$

Then α and β have the same singular data, the function $\alpha\varphi/\beta$ satisfies Eq. (1.1) with β instead of α, and the scattering problems of (1.1) with α or β are identical to each other.

Schrödinger equivalence: Here φ is a solution of (1.1) if and only if $\psi = :\alpha\varphi$ is a solution of the chain of Schrödinger equations

$$\begin{aligned} &\Delta\psi + (k^2 - V - \alpha^{-1}\Delta\alpha)\psi = 0, \quad x\in\mathbb{R}^3\backslash S,\\ &\frac{\psi}{\alpha}, \alpha\frac{\partial\psi}{\partial\nu} - \psi\frac{\partial\alpha}{\partial\nu} \text{ continuous/} S. \end{aligned} \tag{1.9}$$

Hence the *true* data of our problem are a "potential"

$$W = V + \alpha^{-1}\Delta\alpha, \quad x\in\mathbb{R}^3\backslash S,$$

and the "singular data" (1.3). Another way to see it is by introducing a "singular data function."

Singular data functions: We define such a function $\sigma(x)$, if it exists, by assuming that it is positive and harmonic inside the shells, and such that α/σ is everywhere continuous. Information on the normal derivatives on S may be inserted into σ or they remain in the boundary conditions with the equation for $\Phi = \alpha\varphi/\sigma$. Hence σ carries on the information on the "hard singular data" r_n and t_n, and does not depend on α—or V—inside the shells Ω_i. Accordingly, the equation for Φ shows σ only in the differential operator, and Φ is continuous through $\mathbb{R}^3$:

$$\begin{aligned} &(\sigma^{-2}\operatorname{div}\sigma^2\operatorname{grad} + k^2 - V - \alpha^{-1}\Delta\alpha)\Phi = 0, \quad x\in\mathbb{R}^3\backslash S,\\ &\Phi, \sigma^2\left[\frac{\partial\Phi}{\partial\nu} - \left(\frac{1}{\alpha}\frac{\partial\alpha}{\partial\nu} - \frac{1}{\sigma}\frac{\partial\sigma}{\partial\nu}\right)\Phi\right] \text{continuous}\backslash S. \end{aligned} \tag{1.10}$$

Singular data functions were very useful[1] in the one-dimensional case because it was always possible to construct a piecewise constant $\sigma(x)$, or a $\sigma(x)$ that carries on the information on the slope factors (in our three-dimensional case it would mean that $\sigma\,\partial\alpha/\partial\nu - \alpha\,\partial\sigma/\partial\nu$ is continuous through S). The first of these extreme cases generally does not exist in the three-dimensional case, and it is only with a supplementary (physical) assumption that we can use a piecewise constant $\sigma(x)$. The second one does not afford so much simplification.

The present paper contains an essentially complete scattering theory (direct problem) of Eq. (1.1). In Sec. II, we construct our tools, the Green's functions of chains of Schrödinger–Helmholtz equations. In Sec. III, the scattering amplitude is constructed and it is shown that two parts can be identified: the one due to "diffuse scattering" is related to $V + \alpha^{-1}\Delta\alpha$ inside the shells, and the one due to "reflectors" is related to the singular data. In Sec. IV, a first- and second-order perturbation theory is described. A few other problems are discussed in Sec. V.

Parts of the results contained in the present paper were prepublished in the proceedings of various seminars.[3–6] Gathering and completing these results in the present paper can be also considered as achieving a "marriage," that of the scattering theory of the Helmholtz equation, as presented, for instance, by Colton and Kress,[7] and the scattering theory of the Schrödinger equation, as presented, for instance, by Newton.[8]

II. GREEN'S FUNCTIONS FOR CHAINS OF SCHRÖDINGER EQUATIONS

We study $\Psi\in C^2(\mathbb{R}^3\backslash S)$, which satisfies the Schrödinger equation inside each Ω_i,

$$(\Delta + k^2 - W)\Psi(x) = 0, \quad x\in\mathbb{R}^3\backslash S, \tag{2.1}$$

and the continuity conditions

$$[\Psi(x)/a(x)]_{x\in S_i^+} = \lambda_i[\Psi(x)/a(x)]_{x\in S_i^-}, \tag{2.2}$$

$$\left[a(x)\frac{\partial\Psi(x)}{\partial\nu_x} - b(x)\Psi(x)\right]_{x\in S_i^+} = \mu_i\left[a(x)\frac{\partial\Psi(x)}{\partial\nu_x} - b(x)\Psi(x)\right]_{x\in S_i^-}, \tag{2.3}$$

where $\operatorname{Im} k\geqslant 0$, $W(x)$ is locally integrable, and where λ_i and μ_i are real positive *numbers*, the functions $a(x) > 0$, $b(x)$ real, belong to $C(\mathbb{R}^3\backslash S)$ and go to finite limits as $x\to S$, but can have jumps through the surfaces S. We call the problem (2.1)–(2.3) a chain of Schrödinger equations, or, simply, a Schrödinger chain. If $\lambda_i = \mu_i = 1$, as we shall see later, the "current" is conserved. If, in addition, $a(x)$ is continued inside domains Ω_i by a positive function $\alpha(x)\in C^2(\mathbb{R}^3\backslash S)$, which can have jumps through the surfaces S_i, together with its normal derivative $\partial\alpha(x)/\partial\nu(x)$ but in such a way that everywhere $b(x) = \partial\alpha(x)/\partial\nu(x)$, this "conserved current" Schrödinger chain is equivalent to the problem

$$[\alpha^{-2}\operatorname{div}\alpha^2\operatorname{grad} + k^2 - W + \alpha^{-1}\Delta\alpha]\Phi = 0, \quad x\in\mathbb{R}^3, \tag{2.4}$$

provided $\Phi(x) = \Psi(x)/\alpha(x)$. Equation (2.4) is the impedance equation, the corresponding chain is the impedance chain, and α is the impedance factor. Another special case is $W = 0$. We call the corresponding chain a Schrödinger–Helmholtz chain. All the cited chains are homogeneous. A

right-hand side to (2.1) may also be added to show inhomogeneous chains. In the following, we are interested in solutions of chains that satisfy, in addition, the Sommerfeld condition:

$$\langle x/|x|,\text{grad}\,\Psi(x)\rangle - ik\Psi(x) - o(|x|^{-1})$$
$$\text{as}\quad |x|\to\infty \quad \text{uniformly for all directions } x/|x|. \tag{2.5}$$

So as to study the general chain (2.1)–(2.3) and (2.5), or an inhomogeneous one, the main tool is obviously the Green's function that corresponds to the Schrödinger–Helmholtz chain (2.1'), (2.2), (2.3), and (2.5):

$$\Delta\Psi + k^2\Psi = 0, \quad x\in\mathbb{R}^3\backslash S. \tag{2.1'}$$

The present section shows a construction of this Green's function in the "impedance case," and $k\in\mathbb{R}$.

A. Homogeneous chains

For a solution of the Helmholtz equation in Ω_{N+1}, the Sommerfeld condition (2.5) (see Ref. 7, p. 70) implies in particular that $\Psi(x) = O(|x|^{-1})$ as $|x|\to\infty$. It physically means that $\Psi(x)$ is a pure "outgoing" function.

Conditions (2.2) and (2.3) imply that the total current on the (+) side of a surface S_i is proportional to the total current on the (−) side:

$$\int_{S_i^+}\left(\overline{\Psi}\frac{\partial\Psi}{\partial\nu} - \Psi\frac{\overline{\partial\Psi}}{\partial\nu}\right)ds = \lambda_i\mu_i\int_{S_i^-}\left(\overline{\Psi}\frac{\partial\Psi}{\partial\nu} - \Psi\frac{\overline{\partial\Psi}}{\partial\nu}\right)ds. \tag{2.6}$$

We claim that *a function $\Psi(x)$ that satisfies (2.1'), (2.2), (2.3), and (2.5) must vanish.*

Proof: Inside each domain Ω_i, the Green's theorem (Ref. 7, p. 68) yields

$$\int_{S_i^+} - \int_{S_{i+1}^-}\left(\overline{\Psi}\frac{\partial\Psi}{\partial\nu} - \Psi\frac{\partial\overline{\Psi}}{\partial\nu}\right)ds(x)$$
$$= \int_{\Omega_{i+1}}(\overline{\Psi}\Delta\Psi - \Psi\Delta\overline{\Psi})dx = 0. \tag{2.7}$$

A similar argument inside Ω_0 implies that the total current on S_0^- vanishes. Using (2.6) and (2.7) it follows that

$$\int_{S_N^+}\left(\overline{\Psi}\frac{\partial\Psi}{\partial\nu} - \Psi\frac{\partial\overline{\Psi}}{\partial\nu}\right)ds = \text{Im}\int_{S_N^+}\Psi\frac{\partial\overline{\Psi}}{\partial\nu}\,ds = 0. \tag{2.8}$$

Remember

$$\overline{D}_N = \sum_{k=0}^{N}\overline{\Omega}_k.$$

A known result[7] on the Helmholtz equation is that any function Ψ that satisfies (2.1') in $\mathbb{R}^3\backslash\overline{D}_N$ and the condition (2.5) for a given (real) k also satisfies

$$\lim_{R\to\infty}\int_{|y|=R}\left\{\left|\frac{\partial\Psi}{\partial\nu}\right|^2 + k^2|\Psi|^2 + 2k\,\text{Im}\left(\Psi\frac{\overline{\partial\Psi}}{\partial\nu}\right)\right\}ds = 0. \tag{2.9}$$

Equation (2.8) implies that the last term in (2.9) vanishes, and thus Eq. (2.9) implies that

$$\lim_{R\to\infty}\int_{|y|=R}|\Psi|^2\,ds = 0.$$

It follows from the Rellich lemma (Ref. 7, p. 77) that Ψ identically vanishes in $\mathbb{R}^3\backslash\overline{D}_N$. Hence Ψ and $\partial\Psi/\partial\nu$ vanish on S_N^+, and, thanks to (2.2) and (2.3), on S_N^-, so that Ψ satisfies (2.1') throughout $\mathbb{R}^3\backslash\overline{D}_{N-1}$ and hence vanishes therein according to the Rellich lemma. Continuing the process yields $\Psi = 0$. Q.E.D.

B. A Green's function for the impedance chain

For $\alpha(x)$, the domains Ω_i and the surfaces S_i as defined above, we seek $G(x,y)$ such that

$$x\to G(x,y)\in C^2(\mathbb{R}^3\backslash S, x\neq y); \tag{2.10}$$
$$\Delta G(x,y) + k^2G(x,y) = -\delta(x-y); \tag{2.11}$$
$$G(x,y)/\alpha(x) \quad\text{and}\quad G(x,y)\frac{\partial\alpha(x)}{\partial\nu_x} - \alpha(x)\frac{\partial G(x,y)}{\partial\nu_x}$$
are continuous through S,
and the Sommerfeld condition holds; (2.12a)
$$\text{if}\quad x\to S,\quad y\to G(x,y)\quad\text{and}\quad y\to\frac{\partial G(x,y)}{\partial\nu_x}$$
remain locally integrable. (2.12b)

These conditions obviously imply that for $x\neq y$, $x\to G(x,y)$ satisfies (2.1'), and that if $\Phi(x,y)$ is the standard Helmholtz–Green function,[7]

$$\Phi(x,y) = [4\pi|x-y|]^{-1}\exp[ik|x-y|]. \tag{2.13}$$

Then $x\to G(x,y) - \Phi(x,y)$ is a solution of the Helmholtz equation (2.1') that satisfies the Sommerfeld condition. We recall by the way that two functions $g(x)$ and $h(x)$ that satisfy the Sommerfeld condition are such that

$$\lim_{R\to\infty}\int_{|x|=R}\left[g(x)\frac{\partial}{\partial\nu}h(x) - h(x)\frac{\partial}{\partial\nu}g(x)\right]ds = 0. \tag{2.14}$$

The function $G(x,y)$ is defined uniquely by (2.11) and (2.12) since the difference of two solutions would be a solution of the problem (2.1'), (2.2), (2.3), and (2.5). If it exists, it is symmetric with respect to the exchange x, y: using the solution of (2.11), completed by (2.12), and the solution $G(x,z)$ of the same equation, with z instead of y, completed by (2.12), we calculate, by means of Green's formula,

$$\int_{x\in\Sigma\Omega_i}[G(x,z)\Delta G(x,y) - G(x,y)\Delta G(x,z)]dx$$
$$= \lim_{R\to\infty}\int_{|x|=R}\left[G(x,z)\frac{\partial}{\partial\nu_x}G(x,y) - G(x,y)\frac{\partial}{\partial\nu_x}G(x,z)\right]ds_x$$
$$- \int_{S_N^+}\left[G(x,z)\frac{\partial}{\partial\nu_x}G(x,y) - G(x,y)\frac{\partial}{\partial\nu_x}G(x,z)\right]ds_x + \int_{S_N^-}\left[G(x,z)\frac{\partial}{\partial\nu_x}G(x,y) - G(x,y)\frac{\partial}{\partial\nu_x}G(x,z)\right]ds_x$$

$$+\cdots+\left(\int_{S_0^+}-\int_{S_0^-}\right)\left[G(x,z)\frac{\partial}{\partial\nu_x}G(x,y)-G(x,y)\frac{\partial}{\partial\nu_x}G(x,z)\right]ds_x=0, \tag{2.15}$$

and it follows from (2.11) that the left-hand side is also equal to $-G(y,z)+G(z,y)$. Q.E.D.

It remains to construct the solution of (2.11) and (2.12). For this we need to study the problem operators.

C. Surface operators

We first define the operators S_{ij}, K_{ij}, K'_{ij}, and T_{ij} by

$$(S_{ij}f)(x)=2\int_{S_j}ds(z)\Phi(x,z)f(z),\quad x\in S_i, \tag{2.16}$$

$$(K_{ij}f)(x)=2\int_{S_j}ds(z)\frac{\partial\Phi(x,z)}{\partial\nu_z}f(z),\quad x\in S_i, \tag{2.17}$$

$$(K'_{ij}f)(x)=2\int_{S_j}ds(z)\frac{\partial\Phi(x,z)}{\partial\nu_x}f(z),\quad x\in S_i, \tag{2.18}$$

$$(T_{ij}f)(x)=2\frac{\partial}{\partial\nu_x}\int_{S_j}ds(z)\frac{\partial\Phi(x,z)}{\partial\nu_z}f(z),\quad x\in S_i. \tag{2.19}$$

Each of these operators maps a space of functions defined on S_j, with values in $\mathbb{C}$, say $E(S_j)$, into a space of functions defined on S_i, with values in $\mathbb{C}$, say $F(S_i)$. Setting the couples E, F must be done in a different way for **S**, **K**, **K'**, and **T**, and may also depend on whether $i=j$ or $i\neq j$. Since the surfaces are unconnected, a "nondiagonal" operator is the value of a single or double layer potential, or its derivative, at a point that does not belong to the layer, and it is not difficult to prove by means of the Ascoli theorem that S_{ij}, K_{ij}, K'_{ij}, and T_{ij} are compact operators from $C(S_j)$ to $C(S_i)$. As for the "diagonal operators," they are nothing but the "one surface" operators already studied in the literature.[7] Hence S_{ii}, K_{ii}, and K'_{ii} are compact in $C(S_i)$ and $C^{0,\alpha}(S_i)$, for $0<\alpha<1$. But T_{ii} is an unbounded operator, defined on a set $\mathcal{M}(S_i)$ containing $C^{1,\alpha}(S_i)$ and which is mapped into $C(S_i)$. In fact, the reader can make at this point the choice $E=C^{1,\alpha}$, and assume that not only the four operators, for any choice of i and j, but also their operator products, are mappings into the space of continuous functions on the last surface, and are compact unless a diagonal element of **T** is involved. In the study of integral equations to be given below, the diagonal elements of **T** will be eliminated, and we shall set $E(S)=F(S)=C(S)$. We have called S_{ij}, etc., "operator elements" and we have implicitly defined their "operator product," e.g., S_{ki}, T_{ij}, by

$$(S_{ki}T_{ij}f)(x)=4\int_{S_i}\Phi(x,t)ds(t)\frac{\partial}{\partial\nu_t}\int_{S_j}\frac{\partial\Phi(t,y)}{\partial\nu_y}f(y)ds(y)$$
$$[\forall f\in E(S_j),\quad x\in S_k], \tag{2.20}$$

Now, an "operator element" like S_{ij} applies to any f defined on S_j—say, $f_j(x)$. We define a function $f(x)$ on S by the equalities

$$f(x)=f_j(x),\quad x\in S_j,\quad j=0,1,\ldots,N. \tag{2.21a}$$

If each $f_j\in E(S_j)$, $f\in E(S)$. We define a "surface operator" **S**, or **K**, etc., as a mapping from $E(S)$ to $F(S)$ such that if $f_j(x)$ is the restriction of f to $x\in S_j$,

$$(\mathbf{S}f)(x)=\sum_j\int_{S_j}\Phi(x,t)f_j(t)ds(t),\quad x\in S. \tag{2.21b}$$

With these definitions, the operator product of two surface operators, say, **ST**, is the operator such that, for any $f\in E(S)$, $x\in S_k$, $k=0,1,\ldots,N$,

$$(\mathbf{S}\mathbf{T}f)(x)=\left(\sum_i S_{ki}(Tf)\quad(x\in S_i)\right)(x)$$
$$=\left(\sum_i\sum_j S_{ki}T_{ij}f_j\right)(x). \tag{2.20'}$$

Since N is finite **S**, **K**, and **K'** are compact mappings from $E(S)$ to $F(S)$, whereas **T** is not, because of its "diagonal" elements.

D. Relations between the surface operators

In the standard case of one surface, there are simple relations between **T** and the other operators. They are more complicated in our case, but their proofs are elementary and need not be reproduced in detail. They proceed through repeated use of Green's identity and of jump discontinuity formulas. For example, (2.24d) below is obtained by summing over i the element

$$Q_{kij}=4\frac{\partial}{\partial\nu_x}\int_{S_i}\left[\frac{\partial\Phi(x,t)}{\partial\nu_t}g_j(t)-\Phi(x,t)\frac{\partial g_j(t)}{\partial\nu_t}\right]ds(t), \tag{2.22}$$

where $x\in S_k$ and

$$g_j(t)=\int_{S_j}\frac{\partial\Phi(t,y)}{\partial\nu_y}f_j(y)ds(y) \tag{2.23a}$$

is a continuous solution of the Helmholtz equation across any element of S except S_j, through which the jump discontinuity formula is

$$g_j(t^+)-g_j(t)=g_j(t)-g_j(t^-)=\tfrac{1}{2}f_j(t),\quad t\in S_j. \tag{2.23b}$$

The cases $k>i$, $k=i$, $k<i$, and, for each one, $i<j$, $i=j$, $i>j$, are successively evaluated (see Ref. 4, Appendix). As an example, for $k>i>j$, we obtain the self-explanatory formulas

$$Q_{kij}=4\frac{\partial}{\partial\nu_x}\int_{S_i}\left[\frac{\partial\Phi(x,t)}{\partial\nu_t}g_j(t)-\Phi(x,t)\frac{\partial}{\partial\nu_t}g_j(t)\right]ds(t)$$
$$=\cdots$$
$$=4\frac{\partial}{\partial\nu_x}\int_{S_j}\left[\frac{\partial\Phi(x,t)}{\partial\nu_t}g_j(t^-)-\Phi(x,t)\frac{\partial}{\partial\nu_t}g_j(t^-)\right]$$
$$\times ds(t)+4\frac{\partial}{\partial\nu_x}\int_{S_j}\frac{\partial\Phi(x,t)}{\partial\nu_t}f_j(t)ds(t)$$
$$=4\frac{\partial}{\partial\nu_x}g_j(x)\quad(k>i>j). \tag{2.22'}$$

Quite similar methods apply to deriving all the other cases. The final results are

$$\mathbf{SK}' - \mathbf{KS} = -\dot{\mathbf{S}}, \tag{2.24a}$$

$$\mathbf{KK} - \mathbf{ST} = \mathbf{1} + \dot{\mathbf{K}}, \tag{2.24b}$$

$$\mathbf{K}'\mathbf{K}' - \mathbf{TS} = \mathbf{1} - \dot{\mathbf{K}}', \tag{2.24c}$$

$$\mathbf{TK} - \mathbf{K}'\mathbf{T} = \dot{\mathbf{T}}, \tag{2.24d}$$

where $\dot{\mathbf{M}}$ is obtained from the matrix $\mathbf{M}$ made of operator elements M_{ij} by the ansatz

$$\dot{M}_{kj} = \begin{cases} 2(k-j-1)M_{kj}, & \text{for } j \leqslant k-1, \\ 0, & \text{for } j = k, \\ 2(k+1-j)M_{kj}, & \text{for } j \geqslant k+1. \end{cases} \tag{2.25}$$

The dotted matrices involve only "nondiagonal" elements. Therefore $\dot{\mathbf{T}}$ is compact on $C(S)$.

E. Potentials

The right-hand sides of (2.16)–(2.19) can be continued for any $x \in \mathbb{R}^3 \setminus S$ as simple and double layer potentials and their derivatives. Well known results apply[7] and readily show that the single layer potential [(2.16)–(2.21)] and the double layer potential [(2.17)–(2.21)] define solutions of the Helmholtz equation in $\mathbb{R}^3 \setminus S$ that satisfy the Sommerfeld condition (2.5). In addition, it is well known that the single layer potential is continuous throughout $\mathbb{R}^3$ but the normal derivative has known discontinuities when crossing the surface where it is defined. The reverse is true for the double layer potential. These results will be used to construct our Green's function.

F. Construction of the Green's function

We can choose between constructing a solution of (2.11) and (2.12) directly, by writing down $G(x, y)$ as the sum $\Phi(x, y)$ + a single layer potential + a double layer potential, or using the same ansatz to solve the inhomogeneous equation

$$\Delta u(x) + k^2 u(x) = -f(x), \tag{2.26a}$$

completed by the conditions

$$\frac{u}{\alpha} \quad \text{and} \quad u\frac{\partial \alpha}{\partial \nu} - \alpha \frac{\partial u}{\partial \nu} \text{ continuous/}S,$$
$$u(x) \text{ is Sommerfeld outgoing.} \tag{2.26b}$$

Then $G(x, y)$ is identified as the resolvent kernel of this problem. We use this second approach and write down

$$u(x) = F(x) + \sum_{j=0}^{N} \int_{S_j} ds(z) \frac{\partial \Phi(x,z)}{\partial \nu_z} \psi_j(z) + \sum_{j=0}^{N} \int_{S_j} ds(z) \Phi(x,z) \varphi_j(z), \tag{2.27}$$

where

$$F(x) = \int_{\mathbb{R}^3} \Phi(x, y) f(y)\, dy. \tag{2.28}$$

The potential representations satisfy the Helmholtz equation and the Sommerfeld condition; since $F(x)$ satisfies (2.26) and the Sommerfeld condition, so does $u(x)$ also. Now the continuity conditions through S_i imply that, for $x \in S_i$, the following equalities must be satisfied, where the index i reminds us that $x \in S_i$, $i = 0,1,2,\ldots,N$, and the notations introduced above are used:

$$[\psi_i(x)(\alpha^-(x) + \alpha^+(x))]_{x \in S_i} = \{(\alpha^+(x) - \alpha^-(x))[2F(x) + (\mathbf{K}\psi)(x) + (\mathbf{S}\varphi)(x)]\}_{x \in S_i}, \tag{2.29}$$

$$[\varphi_i(x)(\alpha^-(x) + \alpha^+(x)) + \psi_i(x)(\alpha'^-(x) + \alpha'^+(x))]_{x \in S_i}$$
$$= \{(\alpha^+(x) - \alpha^-(x))[2F'(x) + (\mathbf{T}\psi)(x) + (\mathbf{K}'\varphi)(x)] - (\alpha'^+(x) - \alpha'^-(x))[2F(x) + (\mathbf{K}\psi)(x) + (\mathbf{S}\varphi)(x)]\}_{x \in S_i}. \tag{2.30}$$

In these formulas, the prime used in α' and F' stands for the normal derivative at $x \in S$; and φ_i and ψ_i are, respectively, the restrictions of φ and ψ to $x \in S_i$. Equations (2.29) and (2.30) must be satisfied for all i's. It is convenient to rewrite them by introducing the functions on S, or "multiplicative operators" β, β', and γ, hereafter defined as

$$\beta(x) = \frac{\alpha^+(x) - \alpha^-(x)}{\alpha^+(x) + \alpha^-(x)}, \qquad \beta'(x) = \frac{\alpha'^+(x) - \alpha'^-(x)}{\alpha^+(x) + \alpha^-(x)}, \tag{2.31a}$$

$$\gamma(x) = \frac{\alpha'^+(x) + \alpha'^-(x)}{\alpha^+(x) + \alpha^-(x)}. \tag{2.31b}$$

Sufficient assumptions on β, β', and γ are that they belong to $C^{1,\alpha}(S)$. Hence, with this condensed notation, the continuity conditions through S yield the following equations for the functions φ $(x \in S)$ and ψ $(x \in S)$:

$$\psi = 2\beta F + \beta \mathbf{S}\varphi + \beta \mathbf{K}\psi, \tag{2.32}$$

$$\varphi + \gamma\psi = 2\beta F' - 2\beta' F + \beta \mathbf{T}\psi - \beta' \mathbf{K}\psi + \beta \mathbf{K}'\varphi - \beta' \mathbf{S}\varphi. \tag{2.33}$$

If the system obtained from (2.32) and (2.33) by setting $F = F' = 0$ has a solution ψ^0, φ^0, then inserting it into (2.27) yields a solution $\Psi^0(x)$ of the homogeneous chain (2.1'), (2.2), (2.3), and (2.5) (with $\lambda_i = \mu_i = 1$). We know from Sec. V that Ψ^0 necessarily vanishes, and using jump relations across layers shows that so do ψ^0 and φ^0. Hence if the system (2.32) and (2.33) has a solution, it is unique.

The domains in the following derivations are easier to check if one assumes *a priori* β, β', γ, and ψ in $C^{1,\alpha}(S)$, φ in $C^{0,\alpha}(S)$ $(0 \leqslant \alpha \leqslant 1)$ so that $\mathbf{T}\psi$, $\mathbf{TK}\varphi$, etc., make sense. But the final equations and results make sense for $\alpha = 0$ $[C^{0,0}(S) = C(S), C^{1,0}(S) = C^1(S)]$. So to obtain the solution, we may try solving (2.32) in terms of ψ and insert the result into (2.32). However, $(\mathbf{1} - \beta\mathbf{K})^{-1}$ exists as a bounded operator if and only if 1 is not eigenvalue of $\beta\mathbf{K}$. Since $\mathbf{K}$ is

bounded, there is at least a range of values of β small enough (say, $\|\beta\| \leqslant \|\mathbf{K}\|^{-1}$) where 1 cannot be an eigenvalue. On the other hand, if 1 is an eigenvalue of $\beta\mathbf{K}$, Eq. (3.32) yields ψ if and only if $\beta(2F + \mathbf{S}\varphi)$ is orthogonal to the n-dimensional null space of the adjoint of $(1 - \beta\mathbf{K})$. Then ψ itself contains n parameters corresponding to the null space $N(\mathbf{1} - \beta\mathbf{K})$; inserting it into (2.33) can give φ up to n parameters, which are eventually determined by the condition on $\beta(2F + \mathbf{S}\varphi)$. This process is complicated. For the sake of simplicity, we shall not study it here. Except in the special case with Assumption B below, we assume that

$$N(\mathbf{1} - \beta\mathbf{K}) = 0. \tag{2.34}$$

Since $\mathbf{K}$ is compact, the Fredholm alternative holds for (2.32) considered as an equation for ψ; $(\mathbf{1} - \beta\mathbf{K})^{-1}$ is a bounded operator and it enables us to construct ψ from φ. Reinserting the result into (2.34), we meet the operator

$$\mathbf{U} = \mathbf{T}(\mathbf{1} - \beta\mathbf{K})^{-1}\beta\mathbf{S}, \tag{2.35}$$

which needs special study, because it involves the product of an unbounded operator and a compact one. First, notice that if $\mathbf{1} - \beta\mathbf{K}$ is inversible, the null space of the adjoint operator $\mathbf{1} - \mathbf{K}'\beta$ is zero, so that if v did exist such that $v = \beta\mathbf{K}'v$, $\mathbf{K}'v$ would vanish and hence so would $\mathbf{v}$. Therefore $\mathbf{1} - \beta\mathbf{K}'$, too, is inversible.

Using (2.23) and (2.24), we obtain, after elementary calculations,

$$\mathbf{U} = (\mathbf{1} - \beta\mathbf{K}')^{-1}\{[\mathbf{T},\beta]\mathbf{S} - \beta(\mathbf{1} - \mathbf{K}'\mathbf{K}' - \dot{\mathbf{K}}') + ([\mathbf{T},\beta]\mathbf{K} + \beta\dot{\mathbf{T}})(\mathbf{1} - \beta\mathbf{K})^{-1}\beta\mathbf{S}\}. \tag{2.36}$$

Let us show that $[\mathbf{T},\beta]$ is a bounded operator. The only elements that need a proof are the "diagonal ones," which read, for $x \in S_k$ [and applied to $f_k(x)$],

$$\begin{aligned} D_k(x) &= \frac{\partial}{\partial \nu_x}\int_{S_k} \frac{\partial \varphi(x,t)}{\partial \nu_t}\beta(t) f_k(t)\,ds(t) \\ &\quad - \beta(x)\int_{S_k} \frac{\partial \varphi(x,t)}{\partial \nu_t} f_k(t)\,ds(t) \\ &= \left(\frac{\partial}{\partial \nu_x}\beta(x)\right)\int_{S_k} \frac{\partial \varphi(x,t)}{\partial \nu_t} f_k(t)\,ds(t) \\ &\quad + \frac{\partial}{\partial \nu_x}\int_{S_k} \frac{\partial \varphi(x,t)}{\partial \nu_t}[\beta(t) - \beta(x)] f(t)\,ds(t). \end{aligned} \tag{2.37}$$

The first term in the last right-hand side of (2.37) is the product of $\partial\beta(x)/\partial\nu_x$ by $(K_{kk}f_k)(x)$ and thus defines a bounded operator. The boundedness of the next one follows from the lemma (2.10) of Ref. 7 (after a few easy calculations). Q.E.D.

Going back to $\mathbf{U}$ in (2.35), and since the product of a compact operator by a bounded operator is compact, we see that $\mathbf{U} + \beta\mathbf{1}$ is compact:

$$\mathbf{U} = -\beta\mathbf{1} + \mathbf{C}, \tag{2.38}$$

where C is complicated, but trivially written down from (2.36). Inserting the solution ψ of (2.32) into (2.33) and taking (2.38) into account yield

$$(\mathbf{1} + \beta^2)\varphi = \mathbf{A}F + \mathbf{B}\varphi, \tag{2.39}$$

where

$$\mathbf{A}F = 2\beta F' + [2\beta\mathbf{T}(\mathbf{1} - \beta\mathbf{K})^{-1}\beta - 2\beta' - 2\beta'\mathbf{K} \times (\mathbf{1} - \beta\mathbf{K})^{-1}\beta - 2\gamma(\mathbf{1} - \beta\mathbf{K})^{-1}\beta]F, \tag{2.40}$$

$$\mathbf{B} = -\gamma(\mathbf{1} - \beta\mathbf{K})^{-1}\beta\mathbf{S} + \beta\mathbf{C} - \beta'\mathbf{K}(\mathbf{1} - \beta\mathbf{K})^{-1}\beta\mathbf{S} + \beta\mathbf{K}' - \beta'\mathbf{S}, \tag{2.41}$$

$(\mathbf{1} + \beta^2)^{-1}\mathbf{B}$ is compact, the Fredholm alternative holds, and the homogeneous solution yields a solution of (2.1'), (2.2), (2.3), and (2.5) and therefore vanishes. Hence φ can be constructed in $C(S)$ as

$$\varphi = [(1 + \beta^2)\mathbf{1} - \mathbf{B}]^{-1}\mathbf{A}F. \tag{2.42}$$

G. Simplifications

Construction of the Green's function is much simplified in two cases.

(a) Assumption A. $\alpha(x)$ is continuous throughout $\mathbb{R}^3$: Then one can use (2.27) with $\psi = 0$, and β vanishes identically. The remaining equation (2.33) reduces to

$$\varphi = -\beta'(2F + \mathbf{S}\varphi). \tag{2.43}$$

With our assumptions on α, $\beta'\mathbf{S}$ is compact on $C(S)$, the homogeneous equation has only the zero solution because of our results on homogeneous chains, and

$$\varphi = -2(\mathbf{1} + \beta'\mathbf{S})^{-1}\beta'F =: -2(\mathbf{1} - \beta'\mathbf{R})\beta'F. \tag{2.44}$$

It follows from (2.27) and (2.28) that the Green's function is then

$$\begin{aligned} G(x,y) &= \Phi(x,y) - 2\sum_{k=0}^{N}\int_{S_k} ds(z)\Phi(x,z)\beta'(z)\Phi(z,y) \\ &\quad + 2\sum_{k=0}^{N}\int_{S_k} ds(z)\Phi(x,z)\beta'(z) \\ &\quad \times \sum_{j=0}^{N}\int_{S_j} ds(t)R(z,t)\beta'(t)\Phi(t,y), \end{aligned} \tag{2.45}$$

where $R(z,t)$ is the kernel of $\mathbf{R}$.

(b) Assumption B: For each surface, the "relative discontinuity" $\alpha^+(x)/\alpha^-(x)$ does not depend on the position of x on the surface.

In this case, it is possible to construct, as in the one-dimensional case,[1] a piecewise constant "singular data" function, say, $\sigma(x)$ equal to $\sigma_{N+1} = 1$ and $x \in \Omega_{N+1}$, and

$$\sigma_i = \sigma_{i+1}\alpha(x \in S_i^-)\backslash\alpha(x \in S_i^+), \tag{2.46}$$

in each domain Ω_i $(i = N, N-1, \ldots, 0)$. We can solve the problem made of Eq. (2.26) completed by the Sommerfeld condition and the impedance continuity conditions (2.11) and (2.11) by setting $v = u/\sigma$, $g = f/\sigma$ and solving instead the problem

$$\Delta v(x) + k^2 v(x) = -g(x), \quad x \in \mathbb{R}^3\backslash S, \tag{2.47a}$$

$$v(x) \quad \text{and} \quad \sigma^2\left[\frac{\partial v}{\partial \nu} - \frac{v}{\alpha}\frac{\partial \alpha}{\partial \nu}\right] \quad \text{continuous,}$$

Sommerfeld condition. (2.47b)

The analysis of Sec. II A applies to the homogeneous form of (2.47) and guarantees the uniqueness of the solution, which

can be constructed by making use of (2.27), with $\psi = 0$, solving the Fredholm equation

$$\varphi = 2\tilde{\beta}G' - 2\tilde{\beta}'G + \tilde{\beta}\mathbf{K}'\varphi - \tilde{\beta}'\mathbf{S}\varphi, \tag{2.48}$$

where

$$G(x) = \int_{\mathbb{R}^3} \Phi(x,y)g(y)dy$$

and, for each $x \in S$,

$$\begin{aligned} \tilde{\beta}(x) &= \frac{(\sigma^+)^2 - (\sigma^-)^2}{(\sigma^+)^2 + (\sigma^-)^2}, \\ \tilde{\beta}'(x) &= \left((\sigma^+)^2 \frac{1}{\alpha^+}\frac{\partial \alpha^+}{\partial \nu} - (\sigma^-)^2 \frac{1}{\alpha^-}\frac{\partial \alpha^-}{\partial \nu}\right) \\ &\quad \times [(\sigma^+)^2 + (\sigma^-)^2]^{-1}. \end{aligned} \tag{2.49}$$

The Fredholm alternative guarantees that $(1 + \tilde{\beta}'\mathbf{S} - \tilde{\beta}\mathbf{K}')^{-1}$ exists as a bounded operator. It yields

$$v(x) = \int_{\mathbb{R}^3} g(x,y)g(y)dy.$$

This resolvent $g(x,y)$ (which is, of course, not symmetric), readily yields the Green's function of the original problem as the product

$$G(x,y) = \sigma(x)g(x,y)/\sigma(y).$$

Thus we see that our problem is much simpler than the general three-dimensional problem. Yet, this case is more common in physical problems. From the mathematical point of view, it is the direct generalization of the one-dimensional case.

H. Consequences of the "standard equivalence"

The "standard equivalence" between impedance factors α, which lead to the same function $\Delta\alpha/\alpha$ inside the domains Ω_i and the same singular data on the S_i, leads us to the question: how can a "representative" of this class be constructed, uniquely determined, for instance, by the condition that it goes to 1 and $|x| \to \infty$? The answer is that we have to construct α^0 such that

$$\Delta\alpha^0 - \alpha_0(\alpha^{-1}\Delta\alpha) = 0, \tag{2.50a}$$

$$\frac{\alpha^0}{\alpha} \quad \text{and} \quad \alpha^0 \frac{\partial \alpha}{\partial \nu} - \alpha \frac{\partial \alpha^0}{\partial \nu} \quad \text{continous through } S, \tag{2.50b}$$

$$\alpha^0 - 1 \text{ is Sommerfeld outgoing}, \tag{2.50c}$$

It is a limit case of the problem (3.2) below (for $k = 0$) and is solved in a similar way.

Going further into the "standard equivalence," we can see that the Green's function should not be modified if α is replaced by $\tilde{\alpha}$, which satisfies with α condition (2.50b). This invariance is related to that of the chains when continuity conditions are mixed or when, equivalently, a part of Eq. (2.32) is added to Eq. (2.33). It follows that β' can be replaced by $\beta' + q\beta/\alpha^+\alpha^-$, γ by $\gamma - q/\alpha^-\alpha^+$, where $q(x)$ is any function continuous through S. A clever choice yields the "standard" coefficients, completely expressed in terms of the singular data (1.3):

$$\begin{aligned} &\bar{\beta} = (1 - t + r)/(1 + t + r), \\ &\bar{\beta}' = -\bar{s}/t, \quad \bar{\gamma} = -\bar{\beta}\bar{\beta}', \end{aligned} \tag{2.51}$$

where we used the overbar to denote the standard coefficients, instead of a superscript zero, to mean that these two reductions are independent. Notice also that the standard coefficients (2.51) are defined only on S and may not correspond to an impedance factor continued in $\mathbb{R}^3 \setminus S$.

III. THE THREE-DIMENSIONAL IMPEDANCE SCATTERING

A. The fundamental integral equation

We come back to Eq. (1.1), with $\alpha(x) \in C^2$ on the Ω_i's and jumping together with its normal derivative through the surfaces S_i. We are looking for $\varphi(\mathbf{k},x)$ satisfying

$$\begin{aligned} &(\alpha^{-2} \operatorname{div} \alpha^2 \operatorname{grad} + k^2 - V)\varphi(\mathbf{k},x) = 0, \quad x \in \mathbb{R}^3, \\ &\alpha\varphi - \exp[i\mathbf{k}\cdot x] \quad \text{outgoing as } |x| + \infty. \end{aligned} \tag{3.1}$$

The "outgoing condition" is that precisely settled by the Sommerfeld condition (2.5). Now, let $\psi = \alpha\varphi$. As in (1.9), ψ satisfies

$$\begin{aligned} &\Delta\psi + k^2\psi - (V + \alpha^{-1}\Delta\alpha)\psi = 0, \quad (x \in \mathbb{R}^3 \setminus S), \\ &(\alpha^+)^{-1}\psi^+ = (\alpha^-)^{-1}\psi^-, \quad x \in S, \\ &\alpha^+ \frac{\partial \psi^+}{\partial \nu} - \psi^+ \frac{\partial \alpha^+}{\partial \nu} = \alpha^- \frac{\partial \psi^-}{\partial \nu} - \psi^- \frac{\partial \alpha^-}{\partial \nu}, \quad x \in S, \\ &\psi(\mathbf{k},x) - \exp[i\mathbf{k}\cdot x] \quad \text{outgoing} \quad (|x| \to \infty). \end{aligned} \tag{3.2}$$

So as to construct ψ, we shall use the Green's function $G(x,y)$ defined by (2.10)–(2.12). Let R be large enough so that $\overline{D}_N \subset B(0,R) = \{x: |x| < R\}$. Applying the Green's theorem to ψ and G in $\omega(R) = B(0,R) \setminus \overline{D}_N$, we get

$$\begin{aligned} &\int_{\omega(R)} [\Delta_y \psi(k,y)G(x,y) - \psi(k,y)\Delta_y G(x,y)]dy \\ &= \int_{\omega(R)} G(x,y)[V(y) + \alpha^{-1}(y)\Delta\alpha(y)]\psi(\mathbf{k},y)dy \\ &\quad + \begin{cases} \psi(\mathbf{k},x), & \text{if } x \in \omega(R), \\ 0, & \text{otherwise,} \end{cases} \\ &= \int_{|y|=R} \left(\frac{\partial \psi}{\partial \nu_y} G - \psi \frac{\partial G}{\partial \nu_y}\right) ds(y) \\ &\quad - \int_{S_N^+} \left(\frac{\partial \psi}{\partial \nu_y} G - \psi \frac{\partial G}{\partial \nu_y}\right) ds(y). \end{aligned} \tag{3.3}$$

The integral on $|y| = R$, say $I(R)$, can be divided into two parts:

$$\begin{aligned} I(R) &= I_1(R) + I_2(R) \\ &= \int_{|y|=R} \left[\frac{\partial(\psi - e^{i\mathbf{k}\cdot y})}{\partial \nu_y} G - (\psi - e^{i\mathbf{k}\cdot y})\frac{\partial G}{\partial \nu_y}\right] ds(y) \\ &\quad + \int_{|y|=(R)} \left[\frac{\partial e^{i\mathbf{k}\cdot y}}{\partial \nu_y} G(x,y) - e^{i\mathbf{k}\cdot y} \right. \\ &\quad \left. \times \frac{\partial G(x,y)}{\partial \nu_y}\right] ds(y). \end{aligned} \tag{3.4}$$

If we apply the Green's theorem to $\exp[i\mathbf{k}\cdot x]$ and $G(x,y)$ in $\omega(R)$ we obtain

$$\begin{aligned} I_2(R) &= \int_{S_N^+} \left[\frac{\partial e^{i\mathbf{k}\cdot y}}{\partial \nu_y} G(x,y) - e^{i\mathbf{k}\cdot y}\frac{\partial G}{\partial \nu_y}(x,y)\right] \\ &\quad + \begin{cases} e^{i\mathbf{k}\cdot x}, & \text{if } x \in \omega(R), \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.5}$$

Since $\psi(\mathbf{k},y)-\exp[i\mathbf{k}\cdot y]$ and $G(x,y)$ satisfy the Sommerfeld condition, $I_1(R)$ goes to zero as $R\to\infty$. Hence letting $R\to\infty$ in (3.4) yields, for any $x\in\Omega_{N+1}$,

$$\psi(\mathbf{k},x)=\psi_{in}(\mathbf{k},x)-\int_{\Omega_{N+1}}G(x,y)\left[V(y)+\frac{\Delta\alpha(y)}{\alpha(y)}\right]\psi(\mathbf{k},y)dy-\int_{S_N^+}\left(\frac{\partial\psi(\mathbf{k},y)}{\partial\nu_y}G(x,y)-\psi(\mathbf{k},y)\frac{\partial G(x,y)}{\partial\nu_y}\right)ds(y),\tag{3.6}$$

where $\psi_{in}(\mathbf{k},x)$ is defined by the formula

$$\psi_{in}(\mathbf{k},x)=\int_{S_N^+}\left[\frac{\partial e^{i\mathbf{k}\cdot y}}{\partial\nu_y}G(x,y)-e^{i\mathbf{k}\cdot y}\frac{\partial G(x,y)}{\partial\nu_y}\right]ds(y)+\begin{cases}e^{i\mathbf{k}\cdot x}, & \text{if } x\in\Omega_{N+1},\\ 0, & \text{otherwise.}\end{cases}\tag{3.7}$$

Now, in each domain Ω_i, the Green's theorem yields

$$\begin{aligned}&-\int_{\Omega_i}G(x,y)\left[V(y)+\frac{\Delta\alpha(y)}{\alpha(y)}\right]\psi(k,y)dy\\&+\int_{S_{i+1}^-}\left[\frac{\partial\psi(\mathbf{k},y)}{\partial\nu_y}G(x,y)-\psi(\mathbf{k},y)\frac{\partial G(x,y)}{\partial\nu_y}\right]ds(y)\\&-\int_{S_i^+}\left[\frac{\partial\psi(\mathbf{k},y)}{\partial\nu_y}G(x,y)-\psi(\mathbf{k},y)\frac{\partial G(x,y)}{\partial\nu_y}\right]ds(y)\\&=\begin{cases}\psi(\mathbf{k},x), & \text{if } x\in\Omega_i,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}\tag{3.8}$$

and the continuity conditions (2.12) imply that $(\partial\psi/\partial\nu)G-\psi\,\partial G/\partial\nu$ has the same value on S_i^+ and S_i^-. Adding (3.6) and the formulas (3.8) written for $i=(N-1),(N-2),\ldots,0$, we obtain the scattering integral equation

$$\psi(\mathbf{k},x)=\psi_{in}(\mathbf{k},x)-\int_{\mathbf{R}^3\setminus S}G(x,y)\times\left[V(y)+\frac{\Delta\alpha}{\alpha}\right]\psi(\mathbf{k},y)dy,\quad\in\mathbf{R}^3\setminus S.\tag{3.9}$$

It generalizes the usual scattering integral equation to our problem, and reduces to it when there are no discontinuities. Then, according to (3.7), ψ_{in} reduces to $\exp[i\mathbf{k}\cdot x]$, and $G(x,y)$ reduces to $\Phi(x,y)$ given by (2.13), with $k=|\mathbf{k}|$. Hence there are two steps in our way of processing, and they have a different physical meaning.

First step: Construct ψ_{in} and $G(x,y)$ from a knowledge of the singular surfaces and the singular data only—this can be called the step of reflectors (or hard scatterers).

Second step: Construct ψ by solving the integral equation (3.9), where the "diffuse scatterers" represented by $V(x)$ and $\alpha^{-1}\Delta\alpha$ appear for the first time—this is the step of diffuse scatterers.

B. Properties of the incident wave function

The argument used above also shows that ψ_{in} is the solution of

$$(\Delta+k^2)\psi_{in}=0,\quad x\in\mathbf{R}^3\setminus S,$$

$$\frac{\psi_{in}}{\alpha}\quad\text{and}\quad\alpha\frac{\partial\psi_{in}}{\partial\nu}-\psi_{in}\frac{\partial\alpha}{\partial\nu}\quad\text{continuous/}S,\tag{3.10}$$

$$\psi_{in}-\exp[i\mathbf{k}\cdot x]\quad\text{outgoing at}\quad\infty.$$

Hence ψ_{in} is the wave function that corresponds to the physical problem of impedance scattering by an impedance factor $\bar\sigma(x)$, which would be the (piecewise harmonic) solution of (3.10) with $k=0$. To find the behavior of ψ_{in} as $|x|\to\infty$, we apply the Green's theorem to ψ_{in} and $\Phi(x,y)$ in $\omega(R)$. For $x\in\omega(R)$, we readily obtain

$$\begin{aligned}\psi_{in}(\mathbf{k},x)=\int_{|y|=R}&\left[\frac{\partial\psi_{in}(\mathbf{k},y)}{\partial\nu_y}\Phi(x,y)-\psi_{in}(\mathbf{k},y)\frac{\partial}{\partial\nu_y}\Phi(x,y)\right]ds(y)\\&-\int_{S_N^+}\left[\frac{\partial\psi_{in}(\mathbf{k},y)}{\partial\nu_y}\Phi(x,y)-\psi_{in}(\mathbf{k},y)\frac{\partial}{\partial\nu_y}\Phi(x,y)\right]ds(y).\end{aligned}\tag{3.11}$$

Since $\psi_{in}-e^{i\mathbf{k}\cdot x}$ is outgoing, as R goes to infinity, the integral over the sphere tends to $e^{i\mathbf{k}\cdot x}$, giving the representation

$$\psi_{in}(\mathbf{k},x)=\exp[i\mathbf{k}\cdot x]-\int_{S_N^+}\left[\frac{\partial\psi_{in}(\mathbf{k},y)}{\partial\nu_y}\Phi(x,y)-\psi_{in}(\mathbf{k},y)\frac{\partial\Phi(x,y)}{\partial\nu_y}\right]ds(y),\quad x\in\Omega_{N+1}.\tag{3.12}$$

Now using (2.13), the derived formula

$$\frac{\partial\Phi(x,y)}{\partial\nu_y}=\frac{\langle\nu(y),x-y\rangle}{|x-y|^2}(1-ik|x-y|)\Phi(x,y)\tag{3.13}$$

and the asymptotic behavior

$$\Phi(x,y)=(4\pi|x|)^{-1}\exp[ik(|x|-\hat x\cdot y)]+o(|x|)^{-1},\quad|x|\to\infty,\tag{3.14}$$

we obtain, for $|x|\to\infty$,

$$\psi_{in}(\mathbf{k},x)=\exp[i\mathbf{k}\cdot x]-(4\pi)^{-1}\frac{e^{ik|x|}}{|x|}A_0(|\mathbf{k}|\hat x,\mathbf{k})+o(|x|)^{-1},\tag{3.15}$$

with $\hat x=x/|x|$, we used $|\mathbf{k}|$ in A_0 for the length of $\mathbf{k}$ instead of k to be more clear, and

$$A_0(\xi,\mathbf{k})=\int_{S_N^+}e^{-ix\cdot\xi}\left[\frac{\partial\psi_{in}}{\partial\nu_x}(\mathbf{k},x)+i\xi\cdot\nu(x)\psi_{in}(\mathbf{k},x)\right]ds(x).\tag{3.16}$$

This "first step" scattering amplitude gives information not only on S_N but also, by means of ψ_{in}, on all the reflectors.

It is interesting to see that $\psi_{in}(\mathbf{k},x)$ is also related to the asymptotic behavior of $G(x,y)$. So as to prove it, we apply Green's theorem to $G(x,y)$ and $\Phi(x,y)$ in $z\in\omega(R)$, with x,y fixed $(x\neq y)$, x in $\omega(R)$. We get

$$\int_{\omega(R)}[\Phi(x,z)\Delta_zG(z,y)-G(z,y)\Delta\Phi(x,z)]dz$$

$$= \begin{cases} G(x,y) - \Phi(x,y), & \text{if } y \notin \overline{D}_N, \\ G(x,y), & \text{if } y \in D_N, \end{cases}$$

$$= \left(\int_{|z|=R} - \int_{S_N^+} \right) \left[\Phi(x,z) \frac{\partial G}{\partial \nu_z}(z,y) - \frac{\partial \Phi}{\partial \nu_z}(x,z) G(z,y) \right] ds(z). \tag{3.17}$$

Letting $R \to \infty$, and since G and Φ are outgoing, we obtain, for any $x \in \Omega_{N+1}$,

$$G(x,y) = -\int_{S_N^+} \left[\Phi(x,z) \frac{\partial G}{\partial \nu_z}(z,y) - \frac{\partial \Phi(x,z)}{\partial \nu_z} G(z,y) \right] ds(z) + \begin{cases} \Phi(x,y), & \text{if } y \notin \overline{D}_N, \\ 0, & \text{if } y \in D_N. \end{cases} \tag{3.18}$$

As $|x| \to \infty$, inserting (3.14), this yields

$$G(x,y) \sim \frac{1}{4\pi} \frac{e^{ik|x|}}{|x|} \psi_{in}(-|\mathbf{k}|\hat{x}, y) \quad (|x| \to \infty). \tag{3.19}$$

C. The full scattering amplitude

Substituting (3.19) in (3.9) yields

$$\psi(\mathbf{k},x) \sim \psi_{in}(\mathbf{k},x) - \frac{1}{4\pi} \frac{e^{ik|x|}}{|x|} A_1(|\mathbf{k}|\hat{x},\mathbf{k}) + o(|x|^{-1}) \quad \text{as} \quad |x| \to \infty, \tag{3.20}$$

where

$$A_1(\xi,\mathbf{k}) = \int_{\mathbf{R}^3 \setminus S} \psi_{in}(-\xi,x) \left[V(x) + \frac{\Delta\alpha}{\alpha}(x) \right] \psi(\mathbf{k},x) dx. \tag{3.21}$$

Combining (3.15) and (3.20) finally yields the asymptotic representation of $\psi(k,x)$:

$$\psi(\mathbf{k},x) = e^{i\mathbf{k}\cdot x} - \frac{1}{4\pi} \frac{e^{ik|x|}}{|x|} A(|\mathbf{k}|\hat{x},\mathbf{k}) + o(|x|^{-1}), \tag{3.22}$$

where

$$A(\xi,\mathbf{k}) = A_0(\xi,\mathbf{k}) + A_1(\xi,\mathbf{k}). \tag{3.23}$$

Here A_0 is the scattering amplitude due to reflectors only; A_1 is the scattering amplitude due to diffuse scattering, *the reflectors being present*; A is the full scattering amplitude, the quantity to be measured. It is easy to see now that these results trivially reduce to the usual ones for the Schrödinger problem when no reflector is present. Notice also that the generalization to $\mathbb{R}^n$ with $n \geqslant 3$, is trivial.

IV. PERTURBATION THEORY

Linear and higher-order approximations are obtained in the problem by expanding results along "small" parameters, which may be either the relative jumps through the singular surfaces measured by β, β', γ, or the diffuse scattering parameter $V + \Delta\alpha/\alpha$, or both. Hence we obtain three kinds of approximation: the approximations of reflectors scattering, no diffuse scattering being present; the approximations of full scattering (reflectors plus diffuse); and the approximations of diffuse scattering, the scattering amplitude A_0 and the reflectors' Green's function being first calculated exactly. More precisely, we introduce the norms

$$\|\beta\| = \operatorname{Sup}_{x \in S}(|\beta|, |\beta'|, |\gamma|), \tag{4.1}$$

$$\|W\| = \int_{\mathbf{R}^3 \setminus S} \left| V(x) + \frac{\Delta\alpha}{\alpha}(x) \right| dx, \tag{4.2}$$

$$w = \operatorname{Sup}(\|\beta\|, \|W\|). \tag{4.3}$$

The first approximations are those of small $\|\beta\|$, $\|W\| = 0$; the second ones of small w; and the third ones of small $\|W\|$. We shall give the linear approximation in all cases, but the quadratic one only in the cases where it is reasonably simple, viz., the continuous case (Assumption A of Sec. II G) and the specially discontinuous case (Assumption B of Sec. II G). When the "standard coefficients" $\overline{\beta}, \overline{\beta}'$, and $\overline{\gamma}$ are used, $\overline{\gamma}$ is $O(\|\beta\|^2)$ so that further simplifications occur.

A. Approximations of $G(x,y)$ and $A_0(\xi,\mathbf{k})$

The Neumann series expansion of $(\mathbf{I} - \beta\mathbf{K})^{-1}$, where $\mathbf{K}$ is a compact operator and β a multiplicative operator, certainly converges for $\|\beta\|$ small enough. We readily derive, from Eqs. (2.35)–(2.42),

$$\psi = (\mathbf{I} + \beta\mathbf{K})(2\beta F + \beta\mathbf{S}\psi) + (\|\beta\|^3), \tag{4.4}$$

$$\mathbf{A}F = 2\beta F' - 2[\beta' + \gamma\beta - \beta\mathbf{T}\beta + \beta'\mathbf{K}\beta] \times F + O(\|\beta\|^3), \tag{4.5}$$

$$\mathbf{B} = \beta\mathbf{K}' - \beta'\mathbf{S} + O(\|\beta\|^2), \tag{4.6}$$

$$\varphi \sim (\mathbf{I} + \mathbf{B})\mathbf{A}F + O(\|\beta\|^2). \tag{4.7}$$

So to justify that $\beta\mathbf{T}\beta F$ is $O(\|\beta\|^2)$, in spite of the unboundedness of $\mathbf{T}$, it is sufficient to insert the representation (2.28) and a decomposition of $\mathbf{T}$ as in (2.36). Notice that $\mathbf{C}$ being $O(\|\beta\|)$ does not appear in (4.6) and (4.7), so that the linear and quadratic approximations are much simpler than the third-order one. Let us now give the linear approximation of G and A_0 in more detail. Inserting (2.28) into (4.4) and (4.7), and the results into (2.27), keeping the linear terms, and identifying $G(x,y)$, we obtain

$$G(x,y) = \Phi(x,y) + 2\sum_{j=0}^{N} \int_{S_j} ds(z)\beta(z) \frac{\partial}{\partial \nu_z} \times [\Phi(x,z)\Phi(z,y)] - 2\sum_{j=0}^{N} \int_{S_j} ds(z)\beta'(z) \times \Phi(x,z)\Phi(z,y) + O(\|\beta\|^2). \tag{4.8}$$

Let us now use the asymptotic behaviors (3.13) and (3.14) of Φ and derivatives as $|x| \to \infty$. Inserting them in (4.8), and comparing the asymptotic behavior of (4.8) to (3.19), we readily derive

$$\psi_{in}(\mathbf{k},y) = \exp(i\mathbf{k}\cdot y) + 2\sum_{j=0}^{N} \int_{S_j} ds(z) \times \beta(z) \frac{\partial}{\partial \nu_z} [\exp(i\mathbf{k}\cdot z)\Phi(z,y)] - 2\sum_{j=0}^{N} \int_{S_j} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z)\Phi(z,y) + O(\|\beta\|^2). \tag{4.9}$$

Finally, the asymptotic behavior of (4.9) as $|y| \to \infty$ yields A_0 by means of (3.15):

$$A_0(|\mathbf{k}|\hat{y},\mathbf{k}) = -2\sum_{j=0}^{N}\int_{S_j} ds(z) \times \beta(z)\frac{\partial}{\partial \nu_z}[\exp(i\mathbf{k}\cdot z - i|\mathbf{k}|\,\hat{y}\cdot z)] + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z - i|\mathbf{k}|\,\hat{y}\cdot z) + O(\|\beta\|^2). \tag{4.10}$$

Notice that if the standard coefficients are used, the term containing $\gamma(z)$ can be suppressed in (4.5), since it would be $O(\|\beta\|^2)$. The quadratic approximation is nevertheless complicated in the general case. When the simplifying assumptions A or B hold, it is easy to write it down. We only give the results.

For Assumption A,

$$G(x,y) = \Phi(x,y) - 2\sum_{k=0}^{N}\int_{S_k} ds(z)\beta'(z)\Phi(x,z)\Phi(z,y) + 4\sum_{k=0}^{N}\int_{S_k} ds(z)\beta'(z)\Phi(x,z)\sum_{j=0}^{N}\int_{S_j} ds(t)\beta'(t)\Phi(z,t)\Phi(t,y) + O(\|\beta\|^3), \tag{4.11}$$

$$\psi_{in}(\mathbf{k},y) = \exp(i\mathbf{k}\cdot y) - 2\sum_{h=0}^{N}\int_{S_h} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z)\Phi(z,y) + 4\sum_{h=0}^{N}\int_{S_h} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z)\sum_{j=0}^{N}\int_{S_j} ds(t)\beta'(t)\Phi(z,t)\Phi(t,y) + O(\|\beta\|^3), \tag{4.12}$$

$$A_0(|\mathbf{k}|\,\hat{y},\mathbf{k}) = 2\sum_{h=0}^{N}\int_{S_h} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z - i|\mathbf{k}|\,\hat{y}\cdot z) - 4\sum_{h=0}^{N}\int_{S_h} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z)\sum_{j=0}^{N}\int_{S_j} ds(t)\beta'(t)\Phi(z,t)\exp(-i|\mathbf{k}|\,\hat{y}\cdot t) + O(\|\beta\|^3). \tag{4.13}$$

For Assumption B,

$$G(x,y) = \frac{\sigma(x)}{\sigma(y)}\left\{\Phi(x,y) + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\Phi(x,z)\Psi(z,y) + 4\sum_{h=0}^{N}\int_{S_h} ds(z)\Phi(x,z)\sum_{j=0}^{N}\int_{S_j} ds(t)\Psi(t,z)\Psi(z,y) + O(\|\beta\|^3)\right\}, \tag{4.14}$$

where

$$\Psi(z,y) = \tilde{\beta}(z)\frac{\partial\Phi(z,y)}{\partial\nu_z} - \tilde{\beta}'(z)\Phi(z,y), \tag{4.15}$$

$$\sigma(y)\psi_{in}(\mathbf{k},y) = \exp(i\mathbf{k}\cdot y) + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\exp(i\mathbf{k}\cdot z)\Psi(z,y) + 4\sum_{h=0}^{N}\int_{S_h} ds(z)\exp(i\mathbf{k}\cdot z)\sum_{j=0}^{N}\int_{S_j} ds(t)\Psi(z,t)\Psi(t,y) + O(\|\beta\|^3), \tag{4.16}$$

$$A_0(|\mathbf{k}|\hat{y},\mathbf{k}) = -2\sum_{j=0}^{N}\int_{S_j} ds(z)\exp(i\mathbf{k}\cdot z)\tilde{\Psi}(z,|\mathbf{k}|\hat{y}) - 4\sum_{h=0}^{N}\int_{S_h} ds(z)\exp(i\mathbf{k}\cdot z)\sum_{j=0}^{N}\int_{S_j} ds(t)\Psi(z,t)\tilde{\Psi}(t,|\mathbf{k}|\hat{y}) + O(\|\beta\|^3), \tag{4.17}$$

where

$$\tilde{\Psi}(z,|\mathbf{k}|\hat{y}) = \tilde{\beta}(z)\frac{\partial}{\partial\nu_z}\exp(-i|\mathbf{k}|\hat{y}\cdot z) - \tilde{\beta}'(z)\exp(-i|\mathbf{k}|\hat{y}\cdot z). \tag{4.18}$$

Needless to say, the approximate formulas (4.11) and (4.14) must be equivalent, up to $O(\|\beta\|^2)$, to the formula (4.8), conveniently reduced by means of Assumption A or B. This is obvious as for the couple (4.11) and (4.18). It can also be proved directly for the couple (4.14) and (4.8) after using the Green's theorem repeatedly, and relating the coefficient $\sigma(x)/\sigma(y)$ to $\tilde{\beta}$ and $\tilde{\beta}'$.

B. Approximations of $A(\xi,k)$

In the following, we shall assume that there is no solution of the homogeneous form of (3.9) on the real k axis or at $k = 0$, i.e., that there is no real exceptional point or zero energy bound state (part of this assumption can be proved if $V + \Delta\alpha/\alpha$ is sufficiently "short range" but we shall not give this proof here).

It follows from (3.21) and (3.23) that second-order approximations of A_1, i.e., estimates with remainders $O(w^3)$, can be obtained after inserting into (3.21) an estimate of $\psi(\mathbf{k},x)$ with remainder $O(w^2)$. This estimate is readily obtained from (3.9) as

$$\psi(\mathbf{k},x) = \psi_{in}(\mathbf{k},x) + \delta\psi_{in}(\mathbf{k},x) + O(w^2), \tag{4.19}$$

$$\delta\psi_{in}(\mathbf{k},x) = -\int_{\mathbf{R}^3\setminus S} G_0(x,y)\left[V(y) + \frac{\Delta\alpha}{\alpha}(y)\right]\psi_{in}(\mathbf{k},y)dy = O(w), \tag{4.20}$$

where $G_0(x,y)$ is the zeroth-order approximation of $G(x,y)$, i.e., $\Phi(x,y)$, and ψ_{in} is given in (4.9), (4.12), or (4.16). The results are for the general case,

$$\begin{aligned}\psi(\mathbf{k},x) = {} & \exp(i\mathbf{k}\cdot x) + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta(z)\frac{\partial}{\partial\nu_z}[\exp(i\mathbf{k}\cdot z)\Phi(z,x)] \\ & - 2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta'(z)\exp(i\mathbf{k}\cdot z)\Phi(z,x) - \int_{\mathbf{R}^3\setminus S} dz\left[V(z) + \frac{\Delta\alpha}{\alpha}(z)\right]\exp(i\mathbf{k}\cdot z)\Phi(z,x) + O(w^2).\end{aligned} \tag{4.21}$$

When (4.21) is inserted into (3.21), it yields the part A_1 of the scattering amplitude due to diffuse scattering up to $O(w^3)$. Since we have calculated the part A_0 due to reflectors up to $O(w^2)$ only in (4.10), the full scattering amplitude is given, up to $O(w^2)$, by the generalized Born formula:

$$\begin{aligned}A(\mathbf{k}',\mathbf{k}) = {} & -2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta(z)\frac{\partial}{\partial\nu_z}\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] \\ & + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta'(z)\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] + \int_{\mathbf{R}^3\setminus S} dz\left[V(z) + \frac{\Delta\alpha}{\alpha}(z)\right]\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] + O(w^2).\end{aligned} \tag{4.22}$$

Assumption A: Inserting (4.21) into (3.21) and taking into account (4.13), we obtain the full scattering amplitude up to $O(w^3)$:

$$\begin{aligned}A(\mathbf{k}',\mathbf{k}) = {} & 2\sum_{h=0}^{N}\int_{S_h} ds(z)\beta'(z)\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] + \int_{\mathbf{R}^3\setminus S} dz\left[V(z) + \frac{\Delta\alpha}{\alpha}(z)\right]\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] \\ & - 4\sum_{h=0}^{N}\sum_{j=0}^{N}\int_{S_h} ds(z)\beta'(z)\int_{S_j} ds(t)\beta'(t)\exp[i(\mathbf{k}\cdot z - \mathbf{k}'\cdot t)]\Phi(z,t) \\ & - 2\sum_{j=0}^{N}\int_{S_j} ds(z)\beta'(z)\int_{\mathbf{R}^3\setminus S} dt\left[V(t) + \frac{\Delta\alpha}{\alpha}(t)\right]\{\exp[i(\mathbf{k}\cdot t - \mathbf{k}'\cdot z)] + \exp[i(\mathbf{k}\cdot z - \mathbf{k}'\cdot t)]\}\Phi(z,t) \\ & - \int_{\mathbf{R}^3\setminus S} dz\left[V(z) + \frac{\Delta\alpha}{\alpha}(z)\right]\int_{\mathbf{R}^3\setminus S} ds\left[V(t) + \frac{\Delta\alpha}{\alpha}(t)\right]\exp[i(\mathbf{k}\cdot z - \mathbf{k}'\cdot t)]\Phi(z,t) + O(w^3).\end{aligned} \tag{4.23}$$

Notice in this formula the self-interference terms and the cross-interference term between diffuse scattering and (soft) reflectors. Notice also that $\beta'(z)$ can be replaced by its "standard value," which is simply $-\bar{s}(z)$.

Assumption B: From (4.16) and (4.22) we derive

$$\begin{aligned}\psi(k,x) = {} & [\sigma(x)]^{-1}\exp(i\mathbf{k}\cdot x) + 2\sum_{j=0}^{N}\int_{S_j} ds(z)\exp(i\mathbf{k}\cdot z)\Psi(z,x) \\ & - \int_{\mathbf{R}^3\setminus S}\Phi(x,y)\left[V(y) + \frac{\Delta\alpha}{\alpha}(y)\right]\exp(i\mathbf{k}\cdot y)dy + O(w^2),\end{aligned} \tag{4.24}$$

and, inserting this result into (3.21), we derive A_1, which can be added to A_0, as given by (4.17), in order to derive the quadratic approximation of A:

$$\begin{aligned}A(\mathbf{k}',\mathbf{k}) = {} & -2\sum_{j=0}^{N}\int_{S_j} ds(z)\exp(i\mathbf{k}\cdot z)\tilde{\Psi}(z,\mathbf{k}') + \int_{\mathbf{R}^3\setminus S} dz\left[V(z)\frac{\Delta\alpha}{\alpha}(z)\right]\sigma^{-2}(z)\exp[i(\mathbf{k}-\mathbf{k}')\cdot z] \\ & - 4\sum_{h=0}^{N}\int_{S_h} ds(z)\exp(i\mathbf{k}\cdot z)\sum_{j=0}^{N}\int_{S_j} ds(t)\Psi(z,t)\tilde{\Psi}(t,\mathbf{k}') \\ & + 2\sum_{j=0}^{N}\int_{\mathbf{R}^3\setminus S} dx\left[V(x) + \frac{\Delta\alpha}{\alpha}(x)\right]\sigma^{-1}(x)\int_{S_j} ds(z)\Psi(z,x)\{\exp[i(\mathbf{k}\cdot z - \mathbf{k}'\cdot x)] + \exp[i(\mathbf{k}\cdot x - \mathbf{k}'\cdot z)]\} \\ & - \int_{\mathbf{R}^3\setminus S} dx\left[V(x) + \frac{\Delta\alpha}{\alpha}(x)\right]\sigma^{-1}(x)\int_{\mathbf{R}^3\setminus S} dz\left[V(z) + \frac{\Delta\alpha}{\alpha}(z)\right]\sigma^{-1}(z)\exp[i(\mathbf{k}\cdot z - \mathbf{k}'\cdot x)]\Phi(z,x),\end{aligned} \tag{4.25}$$

where Ψ and $\tilde{\Psi}$, as written in terms of "standard coefficients," i.e., in the case of Assumption B,

$$\tilde{\beta}(z) = r(z), \quad +\beta'(z) = -\bar{s}(z), \tag{4.26}$$

are

$$\Psi(z,y) = r(z)\frac{\partial \Phi(z,y)}{\partial \nu_z} + \bar{s}(z)\Phi(z,y), \tag{4.27a}$$

$$\tilde{\Psi}(z,k) = r(z)\frac{\partial}{\partial \nu_z}\exp[-i\mathbf{k}\cdot z] + \bar{s}(z)\exp[-i\mathbf{k}\cdot z]. \tag{4.27b}$$

One easily checks that (4.25) reduces to (4.23) for $\sigma = 1$, since $\tilde{\beta}'$ is then β'. Checking that the first order of (4.25) reduces to (4.22) when Assumption B holds requires a repeated use of Green's theorem, with estimates of $\sigma(x)$ in terms of its values on the S_i's. Again, notice the self-interference and the cross-interference terms of hard reflectors [represented by $r(z)$], soft reflectors [represented by $\bar{s}(z)$], and diffuse scattering [represented by $V(z) + (\Delta\alpha/\alpha)(z)$]. Notice also that $\sigma = 1 + N_0(w)$, so that σ^{-1} can be dropped in the quadratic terms if N is small.

C. Approximations of $A_1(\xi,k)$ only

There are a few cases where ψ_{in} and $G(x,y)$ can be exactly calculated and the results can be used to derive approximately the additional term $A_1(\xi,\mathbf{k})$. The most general of these special cases is that of spherical surfaces $|x| = R_0, R_1, \ldots, R_N$, between which $\sigma(x)$ is constant. The Green's function $G(x,y)$ is then equal to $\sigma(x)g(x,y)/\sigma(y)$, where $g(x,y)$ is the solution of the problem:

$$\Delta g(x,y) + k^2 g(x,y) = -\delta(x-y),$$

$$g(x,y), \quad \sigma^2(x)\frac{\partial g(x,y)}{\partial \nu_x} \tag{4.28}$$

continuous functions of x through S,

$g(x,y)$ purely outgoing.

For a fixed "point" y, we can use $0y$ as the axis of a system of spherical coordinates ($\cos\theta = \hat{x}\cdot\hat{y}$) and expand $g(x,y)$ along the spherical harmonics in such a way that it solves (4.28). This implies that the expansion coefficients are functions g_l of $|x|$ and $|y|$ (which are, respectively, the radial coordinates of the "points" x and y):

$$g(x,y) = (4\pi k\,|x|\,|y|)^{-1} \times \sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta)g_l(|x|,|y|). \tag{4.29}$$

The g_l's satisfy the following constraints for any $y \in \mathbb{R}^3 \setminus S$:

$$\left[\frac{d^2}{d|x|^2} + k^2 - \frac{l(l+1)}{|x|^2}\right]g_l(|x|,|y|) = 0, \quad |x| \neq R_0, R_1, \ldots, R_N; \tag{4.30}$$

$|x| \to g_l(|x|,|y|)$ continuous in $\mathbb{R}^3$,

$|x| \to \sigma^2(|x|)\frac{d}{d|x|}g_l(|x|,|y|)$ continuous in $\mathbb{R}^3 \setminus \{|x| = |y|\}$,

$|x| \to g_l(|x|,|y|)$ outgoing,

$$\left[\frac{d}{d|x|}g_l(|x|,|y|)\right]_{|x|=|y|^-} - \left[\frac{d}{d|x|}g_l(|x|,|y|)\right]_{|x|=|y|^+} = 1. \tag{4.31}$$

Let $u_l(|k|,|x|)$ be the solution of (4.30) such that

$$\frac{u_l(|k|,|x|)}{((\pi/2)|k|\,|x|)^{1/2}\mathcal{J}_{l+1/2}(|k|\,|x|)} \to 1 \quad \text{as} \quad |x| \to 0,$$

and let $f_l(|k|,|x|)$ be the solution of (4.30) that goes to $\exp[i(|k|\,|x| - l\pi/2)]$ as $|x| \to \infty$. We readily obtain

$$4\pi|x|\,|y|g(x,y) = \sum_{l=0}^{N}(2l+1)P_l(\cos\theta)(w_l)^{-1}\sigma^2(|y|) \times \begin{cases} u_l(|k|,|x|)f_l(|k|,|x|), & x<y, \\ u_l(|k|,|y|)f_l(|k|,|x|), & x>y, \end{cases} \tag{4.32}$$

where w_l is the number (independent of $|y|$) defined by

$$w_l = \sigma^2(|y|)[u_l'(|k|,|y|)f_l(|k|,|y|) - u_l(|k|,|y|)f_l'(|k|,|y|)]. \tag{4.33}$$

Going now to the general representation of $G(x,y)$ in an arbitrary reference system centered at 0, we obtain

$$G(x,y) = \frac{1}{4\pi|x|\,|y|}\sum_{l=0}^{\infty}(2l+1)P_l[\cos\theta_x\cos\theta_y + \sin\theta_x\sin\theta_y\cos(\varphi_x - \varphi_y)] \times \frac{\sigma(|x|)\sigma(|y|)}{w_l} \times \begin{cases} u_l(|k|,|x|)f_l(|k|,|y|), & |x|<|y|, \\ u_l(|k|,|y|)f_l(|k|,|x|), & |x|>|y|. \end{cases} \tag{4.34}$$

The other case where the discontinuous problem, with piecewise constant $\sigma(x)$, is easily dealt with, is the cylindrical problem: in some way the limit case of the previous one as $R_0, R_1, R_2, \ldots, R_N$ go to infinity. However, this problem is not a particular case of the previous one and the scattering amplitude can be defined only for angles that are not $\pi/2$.

In all these cases, the approximations for A are easily derived by expanding (3.21). In particular, the Born approximation is only

$$A_1(\xi,\mathbf{k}) = \int_{\mathbb{R}^3\setminus S} dx\,\psi_{in}(-\xi,x)\left[V(x) + \frac{\Delta\alpha}{\alpha}(x)\right] \times \psi_{in}(\mathbf{k},x), \tag{4.35}$$

where ψ_{in} is obtained from (4.34) by using (3.19).

We reserve more detailed studies for applications.

V. MISCELLANEOUS RESULTS

The perturbation theory presented in Sec. IV corresponds to series of powers of singular and diffuse data that converge if w is small enough and our assumptions are satisfied, but the rate of convergence may depend heavily on k, especially at low energies. Thus a study of Born series (especially zero energy ones) should include additional information on the discrete spectrum and the analytical properties in

k. Here we show a few results on these points, without any attempt to be more complete. In fact, all the topics presented in this paper are only meant to give a clear exposition of the direct problem, but more general and more complete studies would be needed to prepare an interesting study of the inverse problem. For instance, we should have studied in Sec. II piecewise C^1 surfaces rather than C^2 surfaces (angular points), and we should have worked in $H^{-1/2}(S)$ instead of $C(S)$, so as to get larger domains for the scattering operators. This generality, which requires much heavier mathematics, is not in the scope of the present paper. We shall finish the paper with a sketch for further generalizations.

A. Bound states

We can make a study quite similar to a previous one[9] valid for $\alpha \in C^2$. We basically assume that there is in the set of all standard-equivalent impedance factors one that is everywhere bounded above and bounded below by two positive numbers $0 < a, < \alpha^2 < b$. Now k is considered in (1.1) as a complex number, and we seek eigenvalues, i.e., values of k allowing a solution φ such that $|\varphi|^2$ and $|\text{grad}\,\varphi \cdot \text{grad}\,\varphi|$ belong to $L^1(\mathbb{R}^3)$. Let k be such a value and k^* its complex conjugate. Writing down (1.1) and its conjugate, and since α and V are real, we obtain, in the ball $\omega(R)$ $[\overline{D}_N \subset \omega(R)]$,

$$\int_{\omega(R)} [\varphi^* \,\text{div}\,\alpha^2\,\text{grad}\,\varphi - \varphi\,\text{div}\,\alpha^2\,\text{grad}\,\varphi^*]dx + (k^2 - k^{*2})\int_{\omega(R)} \alpha^2 \varphi\varphi^* dx = 0. \tag{5.1}$$

Because of our assumption on α, $\alpha^2\varphi\varphi^*$ and $\alpha^2|\varphi|\,|\text{grad}\,\varphi|$ are in $L^1(\mathbb{R}^3)$. Using the Gauss theorem and letting $R \to \infty$ yields

$$\lim_{R\to\infty} \int_{|x|=R} \alpha^2\left(\varphi \frac{\partial \varphi^*}{\partial \nu} - \varphi^* \frac{\partial \varphi}{\partial \nu}\right) ds = 0$$

$$= 4\,\text{Re}\,k\,\text{Im}\,k \int_{\mathbb{R}^3} \alpha^2 \varphi\varphi^* dx. \tag{5.2}$$

Hence the discrete spectrum contains only real values of k^2. We have ruled out by assumption the real values of k ("exceptional points"). As for the imaginary ones (bound states), if $k = i\kappa$ is one of them, we readily obtain, from (1.1),

$$\int_{\omega(R)} \varphi\,\text{div}(\alpha^2\,\text{grad}\,\varphi)dx - \kappa^2 \int_{\omega(R)} \alpha^2 \varphi^2 dx - \int_{\omega(R)} V\alpha^2\varphi^2 = 0. \tag{5.3}$$

Again, using the Gauss theorem and letting $R \to \infty$ yield

$$\lim_{R\to\infty} \int_{|x|=R} \alpha^2 \varphi\,\text{grad}\,\varphi \cdot ds = 0$$

$$= \kappa^2 \int_{\mathbb{R}^3} \alpha^2\varphi^2 dx + \int_{\mathbb{R}^3} \alpha^2\,\text{grad}\,\varphi \cdot \text{grad}\,\varphi\, dx + \int_{\mathbb{R}^3} V\alpha^2\varphi^2 dx. \tag{5.4}$$

Hence, if $V \geqslant 0$, there cannot be any true bound state [$\kappa > 0$, $\varphi \in L^2(\mathbb{R}^3)$, $|\text{grad}\,\varphi| \in L^2(\mathbb{R}^3)$]. If V is somewhere attractive, it is easy to give conditions on its size that discard bound states. The method and results closely follow Ref. 9 and we shall not reproduce them here—neither shall we give the proof that there is no exceptional point, i.e., no solution of the homogeneous version of (3.9), if $V + \alpha^{-1}\Delta\alpha$ is short range; the proof requires heavier mathematics.

B. Analytic properties of $G(x, y)$ as a function of k

In Secs. II–IV, k was meant as the length k (and sometimes noted $|\mathbf{k}|$). However, k is a number, we can consider it in $\mathbb{C}$ and study the corresponding continuation of $G(x, y)$. The only case that is reasonably simple is that of Assumption B(or Assumption A, since it is a particular case of B). When Assumption B is valid, it follows from (2.47)–(2.49) that

$$\varphi = 2\mathbf{R}[rG' + \bar{s}G], \tag{5.5}$$

where we used the standard coefficients (4.26) and $\mathbf{R}$ is the operator $(\mathbf{1} - r\mathbf{K}' - \bar{s}\mathbf{S})^{-1}$. Let $\mathscr{R}(x, y)$ be the kernel of $\mathbf{R}$ in $C(S)$. We readily obtain

$$\sigma(y)G(x, y) = \sigma(x)\Phi(x, y) + 2\sum_{j=0}^{N}\sum_{h=0}^{N} \int_{S_j} ds(z)\Phi(x, y) \times \int_{S_h} ds(t)\mathscr{R}(z,t) \times \left[r(t)\frac{\partial \Phi(t, y)}{\partial \nu_t} + \bar{s}(t)\Phi(t, y)\right]. \tag{5.6}$$

Now $\Phi(x, y)$ is defined as an holomorphic function for any finite $k \in \mathbb{C}$, and it has the properties we used to study homogeneous chains for any k with $\text{Im}\,k \geqslant 0$. It follows from the Fredholm alternative that, for $\text{Im}\,k \geqslant 0$, $\mathscr{R}(z,t)$ is a holomorphic function of k, and thus so is $G(x, y)$.

C. Generalizations and applications

The mixed impedance–potential equation (1.1) is sufficiently general to describe the propagation of almost all scalar waves when it is studied in the frequency domain. The two independent parameters α^2 and $k^2 - V$ are both functions of x and can be used to describe separately two independent local properties of the medium. For instance, it is easy to relate (1.1) to the acoustical equation

$$\lambda\,\text{div}\,\rho^{-1}\,\text{grad}\,P + \omega^2\rho = 0 \tag{5.7}$$

(density ρ, Lamé parameter λ, frequency ω, pressure P) by setting $\alpha^2 = \rho^{-1}, c^2 = \lambda/\rho$, and $k^2 - V = \omega^2/c^2$. Notice that our assumptions on V allow the velocity c to be discontinuous (or more singular) even inside the "regular domains" Ω.

Also, the scattering theory can very easily be modified to accommodate, for instance, geophysical problems where a source is located at a point $a \in \Omega_{N+1}$. If such a source is purely impulsive, and of strength A, we replace problem (3.2) by

$$\begin{cases} \Delta\psi + k^2\psi - (V + \alpha^{-1}\Delta\alpha)\psi = -B\delta(x-a), \\ \dfrac{\psi}{\alpha} \quad \text{and} \quad \alpha\dfrac{\partial\psi}{\partial\nu} - \psi\dfrac{\partial\alpha}{\partial\nu} \quad \text{continuous/}S, \\ \psi \text{ is Sommerfeld outgoing as } |x| \to \infty, \end{cases} \tag{5.8}$$

where $B = A\alpha(0)$. This problem is, in fact, that of the general Green's function for the mixed impedance–potential equation. It is convenient to set $H(x,a) = BG(x,a)$, so that $\chi = :\psi - H$ is the solution of the problem

$$\begin{aligned} &\Delta\chi + k^2\chi - (V + \Delta/\alpha)\psi = 0, \\ &\frac{\chi}{\alpha} \quad \text{and} \quad \alpha\frac{\partial\chi}{\partial\nu} - \chi\frac{\partial\alpha}{\partial\nu} \quad \text{continuous/}S, \\ &\chi \text{ is Sommerfeld outgoing as } |x| \to \infty. \end{aligned} \tag{5.9}$$

The analysis of (5.9) is quite similar to that of Sec. III and yields the integral equation

$$\begin{aligned} \chi(x;a) = &-\int_{\mathbf{R}^3\setminus S} G(x,y)\left[V(y) + \left(\frac{\Delta\alpha}{\alpha}\right)(y)\right]H(y,a)\,dy \\ &-\int_{\mathbf{R}^3\setminus S} G(x,y)\left[V(y) + \left(\frac{\Delta\alpha}{\alpha}\right)(y)\right]\chi(y;a)\,dy. \end{aligned} \tag{5.10}$$

The problem where Ω_{N+1} is limited by a plane (physically, the Earth's surface, say) where ψ must vanish can be easily dealt with: setting $x = (x_0, x_\perp)$, the plane being $x_0 = 0$, the reflection principle shows that the response is

$$\psi(x_0,x_\perp;a_0,a_\perp) - \psi(-x_0,x_\perp;-a_0,a_\perp).$$

Notice also that if Ω_{N+1} is a medium with $V = 0$ and constant α, the values of the response ψ can be related to those on S_N^+ by a formula quite similar to (3.18) and can thus be related to the quantities introduced in the scattering problem. More detailed results are given elsewhere.[10] A similar analysis could deal with normal mode problems by generalizing a former study[11] of the author.

Vectorial cases, in particular, those of wave equations in rigid media and those of electromagnetic wave theory, are in progress. Applications to evaluate the replacement of discontinuities distribution by inhomogeneous continuous media for interpreting the data will be published soon.

ACKNOWLEDGMENTS

Sections II and III gave me (in the summer of 1988) the opportunity for a very helpful collaboration with A. Nachman, who studied the mathematical structure of a somewhat related inverse problem. Several discussions with B. de Facio, who recently tried to generalize these results to electromagnetic problems, are also gratefully acknowledged. Finally I would like to thank Mrs. Cellier for her typing of a manuscript partly made out of broken parts of lectures.

This research was done in association with CNRS URA 768 No. D07680, under Program No. G02640.

[1] P. C. Sabatier and B. Dolveck-Guilpard, "On modeling discontinuous media. One-dimensional approximations," J. Math. Phys. **29**, 861 (1988).

[2] P. C. Sabatier, "Remark on the three-dimensional mixed impedance potential equation," Inverse Problems **3**, L83–L86 (1987); "Corrigendum," Inverse Problems **4**, L1 (1988).

[3] P. C. Sabatier, "For an impedance scattering theory," in *Non linear Evolutions*, edited by J. Léon (World Scientific, Singapore, 1988), pp. 723–745.

[4] P. C. Sabatier, "Three-dimensional impedance scattering theory," in *Electromagnetic and Acoustic Scattering. Detection and Inverse Problem*, edited by C. Bourrely, P. Chiapetta, and B. Torresani (World Scientific, Singapore, 1989).

[5] P. C. Sabatier, "Green's function for chains of Schrödinger–Helmholtz equations," in *Proceedings of the 4th International Conference on Control of Distributed Parameter Systems*, edited by F. Kappel, K. Kunisch, and W. Schappacher (Pitman, New York, 1989).

[6] P. C. Sabatier, "Perturbation theory in impedance scattering," in *Proceedings of the NATO Workshop on Direct and Inverse Methods in Radar Polarimetry*, edited by W. Boerner (Reidel, Dordrecht, 1989).

[7] D. Colton and R. Kress, *Integral Equation Method in Scattering Theory* (Wiley–Interscience, New York, 1983).

[8] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, Berlin, 1982), 2nd ed.

[9] A. Martin and P. C. Sabatier, "Impedance, zero energy wave function, and bound states," J. Math. Phys. **18**, 1623 (1977).

[10] P. C. Sabatier, "Impedance Scattering Theory and Geophysical Problems," to be published in *Model Optimization in Exploration Geophysics. Vol. 4: Proceedings of the Seventh International Mathematics Geophysics Seminar held at the Free University of Berlin*, 8–11 February 1989, edited by A. Vogel (Vieweg, Braunschweig, in press).

[11] P. C. Sabatier, "Spectral and scattering inverse problems," J. Math. Phys. **19**, 2410 (1978).

Approximation of Schrödinger eigenvalues and eigenfunctions by canonical perturbation theory: The periodically driven quantum rotator

Franco Nardini
Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato, 5, I-40127, Bologna, Italy

(Received 27 October 1988; accepted for publication 7 June 1989)

A first-order approximate formula for eigenvalues and eigenfunctions of the Schrödinger operator of the periodically driven quantum rotator is given by means of the classical perturbation theory and an estimate for the remainder is provided; in the nonresonant case the remainder is of order 2 while in the resonant case the remainder turns out to be of order 3/2. In the first case, to what extent the procedure can be interated to higher orders is also discussed.

I. INTRODUCTION

Recently Graffi and Paul[1] have shown that classical perturbation theory by the Hamilton–Jacobi method is an effective and powerful tool in quantum perturbation theory and in semiclassical quantum mechanics. Namely, they have shown that, in the case of the entire holomorphic perturbation of the nonresonant harmonic oscillator, the Rayleigh–Schrödinger perturbation series for the quantum eigenvalues and the classical Birkhoff normal form are generated by the same algorithm. By the same device, Graffi, Paul, and Silverstone[2,3] have given a criterion for predicting a threshold for the occurrence of avoided crossing in the periodically driven quantum rotator which is deeply related to the Chirikov resonance overlapping criterium (see Ref. 4, Sec. 4).

The aim of this paper is to exploit this method in the calculation of approximate eigenvalues and eigenvectors of the Schrödinger operator

$$T_\varepsilon = -\frac{\hbar^2}{2}\frac{\partial^2}{\partial\alpha^2} + \varepsilon V(\alpha,\beta) - i\hbar\omega\frac{\partial}{\partial\beta}. \tag{1.1}$$

acting in $L^2(T^2)$, where T^2 is the two-dimensional torus and the perturbating potential has the form

$$V(\alpha,\beta) = v(\alpha)\cos\beta, \tag{1.2}$$

with

$$v(\alpha) = \sum_{-\infty}^{+\infty}{}_r\, v_r e^{ir\alpha}. \tag{1.3}$$

We assume that the function (1.3) has an analytic continuation in the sense that there exists a constant $\xi > 0$ such that the Laurent series

$$\sum_{-\infty}^{+\infty}{}_r\, v_r z^r \tag{1.4}$$

converges in the complex neighborhood of the one-torus

$$C_\xi = \{z\in\mathbb{C}:\ e^{-\xi} < |z| < e^{\xi}\}. \tag{1.5}$$

The interest in the eigenvalues and eigenfunctions of the operator (1.1) lies in their close connection with the solution of the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = H_\varepsilon(t)\psi,$$

where $H_\varepsilon(t)$ is the time-dependent Hamiltonian of the periodically driven quantum rotator

$$H_\varepsilon(t) = -\frac{\hbar^2}{2}\frac{\partial^2}{\partial\alpha^2} + \varepsilon V(\alpha,\omega t) \tag{1.6}$$

(see Ref. 5, §1 and Ref. 2, §1).

The following lemma represents the foundation on which the forthcoming proofs rely.

Lemma 1.1: (See Ref. 2, §3, step 1; see also Ref. 1, §III.) Let

$$\varphi(\alpha,\beta) = e^{(i/\hbar)W(\alpha,\beta)}. \tag{1.7}$$

Then φ is an eigenfunction for the operator T_ε with eigenvalue E,

$$T_\varepsilon\varphi = E\varphi, \tag{1.8}$$

if and only if the function W solves the equation

$$\frac{1}{2}\left(\frac{\partial W}{\partial\alpha}\right)^2 + \omega\frac{\partial W}{\partial\beta} + \varepsilon V(\alpha,\beta) - \frac{i\hbar}{2}\frac{\partial^2 W}{\partial\alpha^2} = E. \tag{1.9}$$

Remark: Observe that, apart from the term $-(i\hbar/2)(\partial^2 W/\partial\alpha^2)$, Eq. (1.9) is precisely the Hamilton–Jacobi equation for the classical Hamiltonian

$$\mathcal{T}_\varepsilon(A,B,\alpha,\beta) = \tfrac{1}{2}A^2 + \omega B + \varepsilon V(\alpha,\beta), \quad (A,B,\alpha,\beta)\in T^*T^2 \tag{1.10}$$

(T^*T^2 is the cotangent bundle of the two-torus). Hamiltonian (1.10) is the classical counterpart of the Schrödinger operator T_ε [(1.1)] and, on the other hand, it is the quasienergy representation of the time-dependent Hamiltonian of the periodically driven rotator

$$\mathcal{H}_\varepsilon(A,\alpha,t) = \tfrac{1}{2}A^2 + \varepsilon V(\alpha,\omega t), \tag{1.11}$$

which in turn corresponds to the quantum operator $H_\varepsilon(t)$ [(1.6)]. The representation (1.7) and Eq. (1.9) are the bridge between the stationary Schrödinger equation (1.8) and classical perturbation theory.

The paper is divided into two sections besides this Introduction.

In the following section we treat the nonresonant case; that is, the case of quantum numbers n such that the oscillation frequency of the angle α of the unperturbed classical system $\mathcal{T}_0(A,B,\alpha,\beta)$ corresponding to the action in $A = n\hbar$ is not a rational multiple of the frequency ω of the angle β. In this case we solve Eq. (1.9) to first order in ε, setting

$$W = W_0 + W_1, \quad E = E_0 + E_1,$$

 0022-2488/89/112599-08$02.50

and requiring that W_j and E_j satisfy Eq. (1.9) to order ε^j, for $j=0,1$ (see Proposition 2.1 for a precise statement). Then we examine the dependence of W_1 on $\hbar$ in Proposition 2.2 and the possibility of iterating the procedure in Proposition 2.3. These results are related to those obtained by Herczynski[6] of the perturbation of nonresonant rotators.

In the last section we pass to the resonant case, i.e., when the frequency A of the angle α is a rational multiple of the frequency ω of the angle β. In this case we work out only the first-order approximate solution of (1.9), $W=W_0+W_1$ (for the precise statement, see Theorem 3.1). The calculus is much more involved since we must separate the nonresonant component of W_1, which can be treated along the lines of the nonresonant case, and the resonant one, which gives rise to a Mathieu equation; for details we refer to the proof of Theorem 3.1.

II. THE NONRESONANT CASE

In this section we examine the nonresonant case; that is, [see (2.1)] the oscillation frequency of the nonperturbed classical system, $\mathscr{H}_0$[(1.11)], is not a rational multiple of the frequency of the forcing term of the perturbed one, $\mathscr{H}_\varepsilon$. In Proposition 2.1 we compute the solution of (1.9) to the first order in ε, while in Proposition 2.2 we express this solution by means of a series, the terms of which are of order j with respect to $\hbar$ as $\hbar\to 0$, for $j=0,1,\ldots$.

Proposition 2.1: Let $A,B\in\mathbb{R}$ such that

$$Ar+\omega\neq 0, \quad \forall r\in\mathbb{Z}. \tag{2.1}$$

Then for every $m,n\in\mathbb{Z}$ and $\hbar>0$ such that

$$n\hbar=A, \quad m\hbar=B, \tag{2.2}$$

and

$$|Ar\pm\omega+(\hbar/2)r^2|\geqslant d, \quad \forall r\in\mathbb{Z} \tag{2.3}$$

(here d denotes an r-independent positive constant), there exist an approximate eigenvalue E and an approximate eigenfunction φ: $T^2\to\mathbb{C}$ in the form

$$E=\tfrac{1}{2}A^2+\omega B, \tag{2.4}$$

$$\varphi(\alpha,\beta)=e^{(i/\hbar)W(\alpha,\beta,E,\varepsilon)}, \tag{2.5}$$

such that

$$T_\varepsilon\varphi=E\varphi+R, \tag{2.6}$$

with

$$\|R(\alpha,\beta)\|_{L^2(T^2)}\leqslant C\varepsilon^2\|\varphi\|_{L^2(T^2)}, \tag{2.7}$$

where C is a constant depending only on the potential V and on the constant d in (2.3); we observe that despite the dependence of d on $\hbar$ the constant C may be chosen independent of $\hbar$ if we assume a condition slightly stronger than (2.1). We refer to the remark after the proof.

Proof: In order to compute the approximate eigenfunction (2.5) and the approximate eigenvalue (2.4) satisfying (2.6) and (2.7) we look for a constant E and a function W that solve Eq. (1.9) approximately; we write

$$E=E_0+E_1 \tag{2.8}$$

and

$$W(\alpha,\beta,E,\varepsilon)=W_0(\alpha,\beta,E_0)+W_1(\alpha,\beta,E,\varepsilon), \tag{2.9}$$

where E_0 and W_0, E_1 and W_1 solve (1.9) to order 0 and 1 in ε, respectively. Namely, E_0 and W_0 solve (1.9) as $\varepsilon=0$,

$$\frac{1}{2}\left(\frac{\partial W_0}{\partial\alpha}\right)^2+\omega\frac{\partial W_0}{\partial\beta}-\frac{i\hbar}{2}\frac{\partial^2 W_0}{\partial\alpha^2}=E_0, \tag{2.10}$$

while E_1 and W_1 solve

$$\frac{\partial W_0}{\partial\alpha}\frac{\partial W_1}{\partial\alpha}+\omega\frac{\partial W_1}{\partial\beta}+\varepsilon V(\alpha,\beta)-\frac{i\hbar}{2}\frac{\partial^2 W_1}{\partial\alpha^2}=E_1. \tag{2.11}$$

Looking for solutions of (2.10) in separated variables and requiring periodicity in α and β for the function $e^{(i/\hbar)W_0}$, it is easy to obtain (see Ref. 2, §III, step 1)

$$E_0=\tfrac{1}{2}(n\hbar)^2+\omega m\hbar, \tag{2.12}$$

$$W_0(\alpha,\beta,E_0)=n\hbar\alpha+m\hbar\beta. \tag{2.13}$$

Obviously the assumptions of Proposition (2.1) lead us to consider values for m, n, and $\hbar$ that satisfy (2.2) and (2.3).

In order to solve (2.11) we expand the function W_1 into a Fourier series

$$W_1(\alpha,\beta,E,\varepsilon)=\sum_{-\infty}^{+\infty}{}'\left(w_{r,1}(E,\varepsilon)e^{i(r\alpha+\beta)}+w_{r,-1}(E,\varepsilon)e^{i(r\alpha-\beta)}\right). \tag{2.14}$$

Keeping in mind (1.2) and (1.3) it is easy to see that the function (2.14) solves Eq. (2.11) if and only if

$$w_{r,1}(E,\varepsilon)=\varepsilon\frac{i}{2}\frac{v_r}{Ar+\omega+(\hbar/2)r^2}, \quad r\in\mathbb{Z}, \tag{2.15}$$

$$w_{r,-1}(E,\varepsilon)=\varepsilon\frac{i}{2}\frac{v_r}{Ar-\omega+(\hbar/2)r^2}$$

and

$$E_1=\varepsilon\frac{1}{(2\pi)^2}\int_{T^2}V(\alpha,\beta)\,d\alpha\,d\beta=0. \tag{2.16}$$

The form of the Fourier coefficients (2.15) and the assumption (2.3) yield the uniform convergence in T^2 of the series in the rhs of (2.14); moreover, the function W_1 can be analytically continued to the whole complex neighborhood C_ξ [(1.5)] of the torus as the convergence of the Laurent series (1.4) in C_ξ implies the convergence of the series obtained by replacing the coefficients v_r by $w_{r,\pm 1}$.

Inserting the rhs of (2.13) and (2.14) into that of (2.9) and substituting into (2.5) we get

$$(T_\varepsilon\varphi)(\alpha,\beta)=E\varphi(\alpha,\beta)+\frac{1}{2}\left(\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right)^2\varphi(\alpha,\beta), \tag{2.17}$$

where E is given by (2.8), (2.12), and (2.16). Formula (2.17) is precisely (2.6) with

$$R(\alpha,\beta,E,\varepsilon)=\frac{1}{2}\left(\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right)^2\varphi(\alpha,\beta). \tag{2.18}$$

Obviously

$$\|R(\alpha,\beta,E,\varepsilon)\|_{L^2(T^2)}\leqslant\frac{1}{2}\max_{T^2}\left|\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right|^2\|\varphi\|_{L^2(T^2)}. \tag{2.19}$$

In order to estimate the factor

$$\max_{T^2}\left|\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right|^2$$

in the rhs of (2.19), we again exploit the Fourier expansion (2.14) for the function W_1 and the form of its coefficients [(2.15)]:

$$\max_{T^2}\left|\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right|^2 \leqslant\left(\max_{T^2}\left|\varepsilon\sum_{-\infty}^{+\infty}{}_r\left(\frac{rv_r}{Ar+\omega+(\hbar/2)r^2}e^{i(r\alpha+\beta)}+\frac{rv_r}{Ar-\omega+(\hbar/2)r^2}e^{i(r\alpha-\beta)}\right)\right|\right)^2. \tag{2.20}$$

Since the series (1.4) converges in the neighborhood C_ξ [(1.5)], for every $a\in(0,\xi)$, there exists a constant $C>0$ such that

$$|v_r|\leqslant Ce^{-a|r|},\quad \forall r\in\mathbb{Z}. \tag{2.21}$$

Inequality (2.7) can now be easily proved estimating the modulus of the series in the rhs of (2.20) by the series of the moduli and then using (2.21) and the assumption (2.3).

This concludes the proof the Proposition (2.1).

Remark: It has been remarked by Herczynski[7] that assumptions (2.1) and (2.3) can be replaced by the following Diophantine condition on the couple (A,ω): there exist two positive constants C and γ such that

$$|Ar+\omega s|^{-1}\leqslant C(|r|+|s|)^\gamma, \tag{2.22}$$

for every $r,s\in\mathbb{Z}$ with $(r,s)\neq(0,0)$. In fact, taking into account (2.2) the condition (2.22) allows us to estimate the denominator in the rhs of (2.20) as follows:

$$|Ar\pm\omega+(\hbar/2)r^2|^{-1}\leqslant C[1/(\hbar^{\gamma+1})]|r|^{2\gamma}, \tag{2.23}$$

for every $r\in\mathbb{Z}$. Since the maximum of the lhs of (2.23) is attained for r in a neighborhood of $+2n\pm(\omega A)$, (2.23) and (2.21) allow us to prove that the constant C in the rhs of (2.7) can be chosen independent of $\hbar$, at least if $\hbar$ is sufficiently small. For the proof of (2.23) we refer to Ref. 6, Sec. 2, Lemma 1.

The explicit expression (2.15) of the Fourier coefficients of the function (2.14) shows that W_1 is the product of ε times a function depending on the coordinates α and β and the parameters A and $\hbar$.

The problem of the dependence on $\hbar$ of this function is of considerable importance (see Ref. 1); in the following proposition we give a deeper insight into this problem.

Proposition 2.2: The function $W_1(\alpha,\beta,E,\varepsilon)$ can be given the form

$$W_1(\alpha,\beta,E,\varepsilon)=\varepsilon f(\alpha,\beta,E,\hbar)+\varepsilon\sum_{j=1}^{\infty}\hbar^j g_j(\alpha,\beta,E,\hbar)+\varepsilon R(\alpha,\beta,E,\hbar), \tag{2.24}$$

where

$$R(\alpha,\beta,E,\hbar)=o(\hbar^n),\quad \text{as } \hbar\to0,\quad \forall n\in N; \tag{2.25}$$

the functions f, g_j, and the remainder R have analytic continuation to the neighborhood C_ξ, where the same property is true for the potential $V(\alpha,\beta)$ [(cf. (1.3)–(1.5)]; moreover, f and g_j converge to functions of the same kind as $\hbar\to0$.

Proof: We first examine how the generic Fourier coefficient (2.15) of the function (2.14) depends on $\hbar$. For $r\in\mathbb{Z}$ with $|Ar+\omega|>(\hbar/2)r^2$, $w_{r,1}$ can be written in the form

$$w_{r,1}=\varepsilon\frac{i}{2}\frac{v_r}{Ar+\omega}+\varepsilon\frac{i}{2}\frac{v_r}{Ar+\omega}\sum_{j=1}^{\infty}\hbar^j\left(\frac{-r^2}{2(Ar+\omega)}\right)^j; \tag{2.26}$$

on the other hand, for $r\in\mathbb{Z}$ with $|Ar+\omega|<(\hbar/2)r^2$, we have

$$w_{r,1}=\varepsilon i\frac{v_r}{\hbar r^2}\frac{1}{1+2[(Ar+\omega)/(\hbar r^2)]}. \tag{2.27}$$

By the assumption (2.1), if $\hbar$ is small enough, the interval

$$[(A+\sqrt{A^2-2\omega\hbar})/\hbar,(A-\sqrt{A^2+2\omega\hbar})/\hbar]$$

does not contain an integer; we shall suppose that this condition is met throughout the following and this ensures that (2.26) holds if $-r_-(\hbar)<r<r_+(\hbar)$, while (2.27) holds elsewhere. Here we have set

$$r_+(\hbar)=(A+\sqrt{A^2+2\omega\hbar})/\hbar,\qquad r_-(\hbar)=(A+\sqrt{A^2-2\omega\hbar})/\hbar. \tag{2.28}$$

By the very same assumption we find the following formulas for $w_{r,-1}$ analogous to (2.26) and (2.27), respectively:

$$w_{r,-1}=\varepsilon\frac{i}{2}\frac{v_r}{Ar-\omega}+\varepsilon\frac{i}{2}\frac{v_r}{Ar-\omega}\sum_{j=1}^{\infty}\hbar^j\left(\frac{-r^2}{2(Ar-\omega)}\right)^j, \tag{2.29}$$

for $r\in\mathbb{Z}$, with $-r_+(\hbar)<r<r_-(\hbar)$, and

$$w_{r,-1}=\varepsilon i\frac{v_r}{\hbar r^2}\frac{1}{1+2(Ar-\omega)/(\hbar r^2)}, \tag{2.30}$$

for $r\in\mathbb{Z}$, with $r\notin(-r_+(\hbar),r_-(\hbar))$.

Set now

$$f(\alpha,\beta,A,\hbar)=\frac{i}{2}\left(\sum_{-r_-(\hbar)<r<r_+(\hbar)}\frac{v_r}{Ar+\omega}e^{i(r\alpha+\beta)}+\sum_{-r_+(\hbar)<r<r_-(\hbar)}\frac{v_r}{Ar-\omega}e^{i(r\alpha-\beta)}\right) \tag{2.31}$$

and

$$g_j(\alpha,\beta,A,\hbar)=\frac{i}{2}\left(\sum_{-r_-(\hbar)<r<r_+(\hbar)}\frac{v_r}{Ar+\omega}\left(\frac{-r^2}{2(Ar+\omega)}\right)^j e^{i(r\alpha+\beta)}+\sum_{-r_+(\hbar)<r<r_-(\hbar)}\frac{v_r}{Ar-\omega}\left(\frac{-r^2}{2(Ar-\omega)}\right)^j e^{i(r\alpha-\beta)}\right). \tag{2.32}$$

Now $r_\pm(\hbar)\to+\infty$ as $\hbar\to0$ by definitions (2.28); this and the convergence properties of the series (1.4) entail the analyticity for the functions f and g_j in the neighborhood C_ξ

and for their limits as $\hbar \to 0$ as well.

Finally

$$R(\alpha,\beta,E,\hbar) = i \sum_{\substack{r\in\mathbb{Z} \\ r\notin(-r_-(\hbar),r_+(\hbar))}} \frac{v_r}{\hbar r^2} \frac{1}{1+2(Ar+\omega)/(\hbar r^2)} e^{i(r\alpha+\beta)} + i \sum_{\substack{r\in\mathbb{Z} \\ r\notin(-r_+(\hbar),r_-(\hbar))}} \frac{v_r}{\hbar r^2} \frac{1}{1+2(Ar-\omega)/(\hbar r^2)} e^{i(r\alpha-\beta)}. \tag{2.33}$$

In order to prove the estimate (2.25) we first go back to the asymptotic behavior (2.21) of the coefficients of the series (1.4). Inserting (2.21) in (2.33) we estimate the modulus of $R(\alpha,\beta,E,\hbar)$ by

$$|R(\alpha,\beta,E,\hbar)| \leqslant 2 \sum_{\substack{r\in\mathbb{Z} \\ r\notin(-r_-(\hbar),r_-(\hbar))}} \frac{C}{\hbar} \frac{e^{-a|r|}}{r^2};$$

on the other hand, $r_-(\hbar) > A/\hbar$ and therefore the above estimate gives

$$|R(\alpha,\beta,E,\hbar)| \leqslant 4 \frac{C}{\hbar} e^{-aA/\hbar} \sum_{r=1}^{\infty} \frac{1}{r^2}. \tag{2.34}$$

Obviously (2.34) yields (2.25).

The above two propositions show clearly that the function W[(2.9)] and the approximate eigenvalue E[(2.8)] solve (1.9) up to the term $\frac{1}{2}(\partial W_1/\partial\alpha)^2$; this term is of order ε^2 and has analytic continuation to the complex neighborhood C_ξ [(1.5)]. Hence one can wonder whether the procedure leading to the proof of Proposition 2.1 can be iterated. The answer is negative, as can be easily seen setting

$$E = E_0 + E_1 + E_2 \tag{2.35}$$

and

$$W(\alpha,\beta) = W_0(\alpha,\beta) + W_1(\alpha,\beta) + W_2(\alpha,\beta); \tag{2.36}$$

and requiring that E_j and W_j be solutions of (1.9) to order j in ε, for $j = 0,1,2$. This means that E_j and W_j solve the equation

$$\frac{\partial W_0}{\partial\alpha}\frac{\partial W_j}{\partial\alpha} + \omega\frac{\partial W_j}{\partial\beta} - \frac{i\hbar}{2}\frac{\partial^2 W_j}{\partial\alpha^2} + S_j(\alpha,\beta,E,\varepsilon) = E_j, \quad j = 0,1,2, \tag{2.37}$$

where obviously

$$S_0(\alpha,\beta,E,\varepsilon) = 0, \quad S_1(\alpha,\beta,E,\varepsilon) = \varepsilon V(\alpha,\beta),$$
$$S_2(\alpha,\beta,E,\varepsilon) = \frac{1}{2}\left(\frac{\partial W_1}{\partial\alpha}(\alpha,\beta)\right)^2.$$

From (2.14) and (2.15) it is easy to see that the function $S_2(\alpha,\beta,E,\varepsilon)$ can be expanded into the Fourier series

$$S_2(\alpha,\beta,E,\varepsilon) = \sum_{-\infty}^{+\infty} {}_r(s_{r,2}(E,\varepsilon)e^{i(r\alpha+2\beta)} + s_{r,0}(E,\varepsilon)e^{ir\alpha} + s_{r,-2}(E,\varepsilon)e^{i(r\alpha-2\beta)}). \tag{2.38}$$

Now (2.38) suggests we look for a solution of (2.37) for $j = 2$ of the form

$$W_2(\alpha,\beta,E,\varepsilon) = \sum_{-\infty}^{+\infty} {}_r(w_{r,2}^{(2)}(E,\varepsilon)e^{i(r\alpha+2\beta)} + w_{r,0}^{(2)}(E,\varepsilon)e^{ir\alpha} + w_{r,-2}^{(2)}(E,\varepsilon)e^{i(r\alpha-2\beta)}. \tag{2.39}$$

Inserting (2.38) and (2.39) into (2.37) we obtain

$$(Ar+\omega s+(\hbar/2)r^2)w_{r,s}^{(2)}(E,\varepsilon) = -s_{2,s}(E,\varepsilon), \tag{2.40}$$

for every $(r,s)\neq(0,0)$, with $r\in\mathbb{Z}$ and $s = -2,0,2$, while

$$E_2 = s_{0,0}(E,\varepsilon). \tag{2.41}$$

From (2.40) we could obtain formulas analogous to (2.15) if and only if

$$(Ar+\omega s+(\hbar/2)r^2)\neq 0, \text{ for every } (r,s)\neq(0,0), \text{ with } r\in\mathbb{Z} \text{ and } s = -2,0,2,$$

but this is false, at least for $r = -2n$ and $s = 0$ [we recall (2.2)].

If we assume that the Diofantine condition (2.22) holds, we can solve Eq. (2.37) for $j = 2$ only in the following approximate sense: define the function $W_2(\alpha,\beta,E,\varepsilon)$ by (2.38) where the coefficients $w_{r,s}^{(2)}(E,\varepsilon)$ are given by (2.40) for every $r\in\mathbb{Z}$ and $s = -2,0,2$ with $(r,s)\neq(0,0)$ and $(r,s)\neq(-2n,0)$ and set $w_{0,0}^{(2)}(E,\varepsilon) = w_{-2n,0}^{(2)}(E,\varepsilon) = 0$; then $W_2(\alpha,\beta,E,\varepsilon)$ solves (2.37) except for the term $s_{-2n,0}^{(2)}(E,\varepsilon)e^{-2ni\alpha}$. By the analyticity of the known term $S_2(\alpha,\beta,E,\varepsilon)$ in a complex neighborhood of the two-torus it is easy to see that there exists a constant $k>0$ such that [cf. (2.21)]

$$s_{-2n,0}(E,\varepsilon) = O(e^{-kn}), \quad \text{as } n\to\infty.$$

Keeping in mind (2.2), we see that the function $W_2(\alpha,\beta,E,\varepsilon)$ satisfies

$$\frac{\partial W_0}{\partial\alpha}\frac{\partial W_2}{\partial\alpha} + \omega\frac{\partial W_2}{\partial\beta} - \frac{i\hbar}{2}\frac{\partial^2 W_2}{\partial\alpha^2} + S_2(\alpha,\beta,E,\varepsilon) = E_2 + O(e^{-k/\hbar}), \quad \text{as } \hbar\to 0, \tag{2.42}$$

where E_2 is given by (2.41) and k is some positive constant. Eq. (2.42) is the counterpart of Eq. (15) in Ref. 6.

Now if $W_2(\alpha,\beta,E,\varepsilon)$ and E_2 are as above and $W(\alpha,\beta)$, E, and ϕ are defined by (2.36), (2.37), and (1.7), respectively, then E is an approximate eigenvalue and ϕ is an approximate eigenfunction of the operation T_ε in the sense that they satisfy

$$T_\varepsilon\varphi = E\varphi + R, \tag{2.43}$$

with

$$\|R(\alpha,\beta)\|_{L^2(T^2)} \leqslant \varepsilon^3 C\|\varphi\|_{L^2(T^2)} + \varepsilon^2 Ce^{-k/\hbar}. \tag{2.44}$$

Under the Diophantine condition (2.22) this procedure can be obviously iterated by setting

$$E = E_0 + E_1 + \ldots + E_k \tag{2.45}$$

and

$$W(\alpha,\beta) = W_0(\alpha,\beta) + W_1(\alpha,\beta) + \ldots + W_k(\alpha,\beta), \tag{2.46}$$

and requiring that $W_j(\alpha,\beta)$ be an approximate solution of (2.37) for $j = 0,1,\ldots,k$ in the sense of (2.42). We state this

result in the following proposition.

Proposition 2.3: Under the above assumptions the function (1.7) with $W(\alpha,\beta)$ given by (2.46) is an approximate eigenfunction relative to the approximate eigenvalue E [(2.45)] in the sense that

$$T_\varepsilon \varphi = E\varphi + R,$$

with

$$\|R(\alpha,\beta)\|_{L^2(T^2)} \leqslant \varepsilon^{k+1} C \|\varphi\|_{L^2(T^2)} + \varepsilon^2 C e^{-k/\hbar}.$$

At least for the particular potential (1.2) the above proposition provides a (perhaps unsatisfactory) answer to the question raised by Graffi, Paul, and Silverstone (see Ref. 2, Sec. III, Step 2) of whether Eqs. (2.37) could be solved recursively.

III. THE RESONANT CASE

We come now to the resonant case; that is, the case in which the oscillation frequency of the unperturbed system is a rational multiple of the frequency of the forcing perturbative term. The following theorem is analogous to Proposition 2.1. However, in this case the first-order approximation has no trivial extension to higher orders in ε, which parallels Theorem 2.3 of the preceding section. In fact, in this case the remainder turns out to be of order $\varepsilon^{3/2}$ and qualitatively different from the potential $V(\alpha,\beta)$ which appears in the equation valid to first order in ε [compare (3.7) and (3.44) with (2.7) and (2.18), respectively].

The reason for this is to be found in the technique of the proof, which is reminiscent of the canonical perturbation theory introduced by Poincaré in the resonant case (see Ref. 8, Chap XIX, §199–203); precisely we split the function W_1 [see (2.9)] and the perturbation $V(\alpha,\beta)$ into a nonresonant part and a resonant one (see step 1 of the following proof). The nonresonant part satisfies a linear equation (see step 2) while the resonant one is required to satisfy a nonlinear equation (see step 3 (3.22)): as in the classical case the fractional power of ε arises from this equation. We remark also that Eq. (3.22) coincides with the equation in Ref. 8, Chap. XIX, §199, p. 316, but for the (quantum) term

$$-i\frac{\hbar}{2}\frac{p^2}{4}\frac{\partial^2 S}{\partial \mu^2}.$$

Theorem 3.1: Let $A,B\in\mathbb{R}$ such that there exists $p\in\mathbb{Z}$ such that

$$Ap + \omega = 0. \tag{3.1}$$

Then for every $m,n\in\mathbb{Z}$ and $\hbar > 0$ with

$$n\hbar = A, \quad m\hbar = B, \tag{3.2}$$

and

$$|Ar \pm \omega + (\hbar/2)r^2| \geqslant d > 0, \quad \forall r\in\mathbb{Z} \quad \text{with} \quad r\neq p, \tag{3.3}$$

there exist an approximate eigenvalue E and an approximate eigenfunction φ: $T^2\to\mathbb{C}$ of the form

$$\varphi(\alpha,\beta) = e^{(i/\hbar)\hat{W}(\alpha,\beta,E,\varepsilon)}\varphi_1(\alpha,\beta) \tag{3.4}$$

and

$$E = E_0 + E_1 \tag{3.5}$$

such that

$$E_0 = \frac{1}{2}A^2 + \omega B \quad [\text{c.f. } (2.12)], \tag{3.6}$$

$$E_1 = \varepsilon V + p\hbar\left[\sqrt{\varepsilon V}/\log(\varepsilon V)\right] r(p,\hbar,\varepsilon) + O(\hbar^2), \quad \text{as } \hbar\to 0 \quad [cf.(2.16)],$$

and

$$T_\varepsilon \varphi = E\varphi + R,$$

with

$$\|R\|_{L^2(T^2)} \leqslant \varepsilon^{3/2} C(1 + o(1)) \max_{T^2}|\varphi_1| \, \|e^{(i/\hbar)\hat{W}}\|_{L^2(T^2)} + \varepsilon^2 C \|\varphi\|_{L^2(T^2)}, \tag{3.7}$$

where $o(1)$ is to be meant as $\hbar\to 0$. [The function $\hat{W}(\alpha,\beta)$ is quite similar to the function W appearing in the approximate eigenfunction (2.5) in the nonresonant case, while the function φ_1 is a Mathieu function; for the precise definitions see (3.35) and (3.36)]. The constant V in the first term of the rhs of (3.6) denotes the absolute value of the Fourier coefficient v_p of the potential $V(\alpha,\beta)$ relative to the resonant mode $e^{\pm i(p\alpha+\beta)}$ [see (1.2) and (1.3)]. The function $r(p,\hbar,\varepsilon)$ in the second term in the rhs of (3.6) is bounded as $\varepsilon\to 0$ and $\hbar\to 0$. As for the constants in the rhs of (3.7), they depend on the $L^2(T^2)$ norm of the potential $V(\alpha,\beta)$ and on (3.3).

Remark: Also in this case the constant C_1 in the second term of the rhs of (3.7) can be chosen independent of $\hbar$ if $\hbar$ is sufficiently small since the minimum of the absolute value of the rhs of (3.15) is attained in a neighborhood of $-2n-p-\frac{1}{2}p^2/n$ and is of order $\hbar^{-1}$ as $\hbar\to 0$; again the dependence of the constant d in (3.3) on $\hbar$ does not imply the same dependence for C_1.

Proof: We divide the proof into six steps.

Step 1 (Separation of the resonant and nonresonant components): We proceed as in Ref. 2, § III, step 2, and write the potential $V(\alpha,\beta)$ as a sum of two terms:

$$V(\alpha,\beta) = \hat{V}(\alpha,\beta) + \tilde{V}(\alpha,\beta).$$

Then $\hat{V}(\alpha,\beta)$ is obtained by summing up all the nonresonant modes,

$$\hat{V}(\alpha,\beta) = \sum_{\substack{r\in\mathbb{Z}\\ r\neq p}} \left(\frac{v_r}{2} e^{i(r\alpha+\beta)} + \frac{v_{-r}}{2} e^{-i(r\alpha+\beta)}\right), \tag{3.8}$$

while the term $\tilde{V}(\alpha,\beta)$ consists of the resonant part,

$$\tilde{V}(\alpha,\beta) = (v_p/2)\, e^{i(p\alpha+\beta)} + (v_{-p}/2)\, e^{-i(p\alpha+\beta)}. \tag{3.9}$$

We modify the method of Proposition 2.1 for finding a solution for (1.9) of the form (1.7) by writing the function W_1 [see (2.9)] as a sum of a nonresonant term $\hat{W}_1$ and a resonant one $\tilde{W}_1$; namely,

$$E = E_0 + \hat{E}_1 + \tilde{E}_1 \tag{3.10}$$

and

$$W(\alpha,\beta,E,\varepsilon) = W_0(\alpha,\beta,E_0) + \hat{W}_1(\alpha,\beta,E,\varepsilon) + \tilde{W}_1(\alpha,\beta,E,\varepsilon), \tag{3.11}$$

where E_0 and W_0 are still given by (2.12) and (2.13), respectively, while

$$\hat{W}_1(\alpha,\beta,E,\varepsilon) = \sum_{\substack{r\in\mathbb{Z}\\ r\neq p}} (\hat{w}_{r,1}(E,\varepsilon)e^{i(r\alpha+\beta)} + \hat{w}_{-r,-1}(E,\varepsilon)e^{-i(r\alpha+\beta)}) \quad (3.12)$$

and

$$\tilde{W}_1(\alpha,\beta,E,\varepsilon) = \sum_{j\in\mathbb{Z}} \tilde{w}_{jpj}(E,\varepsilon)\, e^{ij(p\alpha+\beta)}. \quad (3.13)$$

The functions $\tilde{W}(\alpha,\beta)$ and $\hat{W}(\alpha,\beta)$ are, respectively, the counterparts of the terms $\sqrt{\mu}\, S_1$ and $\mu\, S_2$ that appear in the development of the generating function introduced in Ref. 8, Chap. XIX, §204, Eq. (2).

Step 2 (Equation for the nonresonant components): As for $\hat{E}_1$ and $\hat{W}_1$, we require that they satisfy an equation analogous to (2.11),

$$\frac{\partial W_0}{\partial\alpha}\frac{\partial\hat{W}_1}{\partial\alpha} + \omega\frac{\partial\hat{W}_1}{\partial\beta} - i\frac{\hbar}{2}\frac{\partial^2\hat{W}_1}{\partial\alpha^2} + \varepsilon\hat{V}(\alpha,\beta) = \hat{E}_1. \quad (3.14)$$

From (3.9) it is clear that the function (3.12) satisfies Eq. (3.14) if and only if

$$\hat{w}_{r,1}(E,\varepsilon) = \varepsilon\frac{i}{2}\frac{v_r}{Ar+\omega+(\hbar/2)r^2}, \quad r\in\mathbb{Z} \text{ with } r\neq p, \quad (3.15)$$

$$\hat{w}_{r,-1}(E,\varepsilon) = \varepsilon\frac{i}{2}\frac{v_r}{Ar-\omega+(\hbar/2)r^2}, \quad r\in\mathbb{Z} \text{ with } r\neq -p,$$

and

$$\hat{E}_1 = \varepsilon\frac{1}{(2\pi)^2}\int_{T^2}\hat{V}(\alpha,\beta)d\alpha\, d\beta = 0, \quad (3.16)$$

where A is the constant defined in (3.2)

It is worth noticing that (3.15) and (3.16) parallel (2.15) and (2.16), respectively; thus we can conclude also in this case that the function $\hat{W}$ has analytic continuation to the complex neighborhood C_ξ [(1.5)] by (3.3) and the assumption on the series (1.3).

Step 3 (equation for the resonant components): In order to derive an equation for the resonant components $\tilde{E}_1$ and $\tilde{W}_1$ we first observe that requiring that they satisfy an equation analogous to (3.14) would lead to a solution like (3.15) with $r=p$; in this case, however, the denominator in the rhs of (3.15) would be simply $(\hbar/2)p^2$ [by (3.1)] and hence infinitesimal as $\hbar\to 0$. The way out from this shortcoming is taking into account the quadratic term $(\partial\tilde{W}_1/\partial\alpha)^2$ arising in (1.9) under substitution of (3.11); hence we require that $\tilde{E}_1$ and $\tilde{W}_1$ satisfy the equation

$$\frac{\partial W_0}{\partial\alpha}\frac{\partial\tilde{W}_1}{\partial\alpha} + \omega\frac{\partial\tilde{W}_1}{\partial\beta} + \varepsilon\tilde{V}(\alpha,\beta) - \frac{i\hbar}{2}\frac{\partial^2\tilde{W}_1}{\partial a^2} + \frac{1}{2}\left(\frac{\partial\tilde{W}_1}{\partial\alpha}\right)^2 = \tilde{E}_1. \quad (3.17)$$

On the other hand, the sum of the first and second term in the lhs of (3.17) is zero by (3.1) and (2.13) and the equation reduces to

$$\varepsilon\tilde{V}(\alpha,\beta) - \frac{i\hbar}{2}\frac{\partial^2\tilde{W}_1}{\partial\alpha^2} + \frac{1}{2}\left(\frac{\partial\tilde{W}_1}{\partial\alpha}\right)^2 = \tilde{E}_1. \quad (3.18)$$

We start computing a solution for (3.18) by noticing that

$$\tilde{V}(\alpha,\beta) = V\cos(p\alpha+\beta+\gamma),$$

where $V=|v_p|$ and $\gamma=\arg v_p$ [see (3.9)]. With no loss of generality we can suppose $\gamma=0$ and set

$$\tilde{V}(\alpha,\beta) = V\cos(p\alpha+\beta). \quad (3.19)$$

Now we use a relevant property of Eq. (3.18) highlighted in Ref. 2, §III, step 4; namely, the known term (3.19) and the unknown (3.13) depend only on the classical "slow" variable. Precisely we perform the change of variable

$$\mu = (p\alpha+\beta)/2 \quad (3.20)$$

and set

$$S(\mu) = \tilde{W}_1(\alpha,\beta), \quad (3.21)$$

where S is a periodic function with period π. Inserting (3.19)–(3.21), Eq. (3.18) becomes

$$\frac{1}{2}\left(\frac{p}{2}\frac{\partial S}{\partial\mu}\right)^2 - i\frac{\hbar}{2}\frac{p^2}{4}\frac{\partial^2 S}{\partial\mu^2} + \varepsilon V\cos(2\mu) = \tilde{E}_1. \quad (3.22)$$

We follow once more Ref. 2, §III, step 5, and set

$$U = e^{(i/\hbar)S}. \quad (3.23)$$

Thus we transform (3.22) again into a Schrödinger equation, since S solves (3.22) if and only if U [(3.23)] solves

$$-\frac{1}{2}\left(\frac{p\hbar}{2}\right)^2\frac{\partial^2 U}{\partial\mu^2}(\mu) + \varepsilon V\cos(2\mu)U(\mu) = \tilde{E}_1 U(\mu). \quad (3.24)$$

Equation (3.24) is a Mathieu equation which we write in the canonical form (see Ref. 9, Chap. 2, §2.21)

$$\frac{\partial^2 U}{\partial\mu^2}(\mu) + (\lambda - 2q\cos(2\mu))U(\mu) = 0, \quad (3.25)$$

where

$$\lambda = 8\tilde{E}_1/(\hbar^2p^2) \quad (3.26)$$

and

$$q = 4(\varepsilon V/(\hbar^2p^2)). \quad (3.27)$$

In accordance with the meaning of the function U [see, (3.21) and (3.23)], we look for a periodic solution of (3.25) with period π. It is well known (Ref. 9, Chap. 2,§2.2) (see, also, Ref. 10, part I, Chap. II) that, for every q, Eq. (3.25) has two sequences of eigenvalues

$$\lambda = a_{2m}(q), \quad m = 0,1,2,...,$$

and

$$\lambda = b_{2m}(q), \quad m = 1,2,3,...,$$

and two corresponding sequences of periodic solutions with period π: the first one made up of even functions

$$U(\mu) = ce_{2m}(\mu,q), \quad m = 0,1,2,...,$$

and the second one made up of odd functions

$$U(\mu) = se_{2m}(\mu,q), \quad m = 1,2,3,... .$$

Step 4 (correction of the eigenvalue to first order in ε): Since we are looking for an approximate eigenvalue of the form (3.5) and $\tilde{E}_1$ plays the role of first-order correction (with respect to the perturbation size), we must require that

in some sense $\tilde{E}_1 \sim \varepsilon V$ as $\hbar \to 0$; this entails that the parameters λ [(3.26)] and $2q$ [(3.27)] of the Mathieu equation (3.25) are near to each other in the sense that they belong to the region IV of Ref. 11, Chap. I, §1.2 and Chap. II, §4.1, (4.1). Therefore we exploit the formulas [Ref. 11, Chap. II, §4, (4.13b)] with $n = 2m$, keeping in mind that they hold under the condition that the quantity

$$r_{\pm}(m,\varepsilon,\hbar) = (2m \pm \tfrac{1}{4})\pi - \sqrt{16q} \tag{3.28}$$

is small.

We start examining (3.28) with the plus sign; substituting (3.27) in (3.28) and choosing m such that the rhs of (3.28) is minimum we find

$$4\frac{\sqrt{\varepsilon V}}{\pi\hbar p} - \frac{5}{8} < m \leqslant 4\frac{\sqrt{\varepsilon V}}{\pi\hbar p} + \frac{3}{8} \tag{3.29}$$

and correspondingly

$$\lambda = a_{2m}\left(4\frac{\varepsilon V}{\hbar^2 p^2}\right) = 8\frac{\varepsilon V}{\hbar^2 p^2} + 16\sqrt{\frac{\varepsilon V}{\hbar^2 p^2}}\,\frac{r_+(m,\varepsilon,\hbar)}{\log(\varepsilon V/\hbar^2 p^2) + 8\log 2 - \Gamma'(\frac{1}{4})/\Gamma(\frac{1}{4})} + O(1), \tag{3.30}$$

as $\hbar \to 0$. In this case the solution of (3.25) is the even Mathieu function

$$U(\mu) = ce_{2m}(\mu, 4(\varepsilon V/(\hbar^2 p^2))). \tag{3.31}$$

If we choose the minus sign in (3.28), we must replace (3.29) with

$$4\frac{\sqrt{\varepsilon V}}{\pi\hbar p} - \frac{3}{8} < m \leqslant 4\frac{\sqrt{\varepsilon V}}{\pi\hbar p} + \frac{5}{8} \tag{3.32}$$

and correspondingly

$$\lambda = b_{2m}\left(4\frac{\varepsilon V}{\hbar^2 p^2}\right) = 8\frac{\varepsilon V}{\hbar^2 p^2} + 16\sqrt{\frac{\varepsilon V}{\hbar^2 p^2}}\,\frac{r_-(m,\varepsilon,\hbar)}{\log(\varepsilon V/\hbar^2 p^2) + 8\log 2 - \Gamma'(\frac{3}{4})/\Gamma(\frac{3}{4})} + O(1), \tag{3.33}$$

as $\hbar \to 0$. In this case the solution of (3.25) is the odd Mathieu function

$$U(\mu) = se_{2m}(\mu, 4\varepsilon V/(\hbar^2 p^2)). \tag{3.34}$$

Conditions (3.29) and (3.32) yield that $|r_{\pm}(m,\varepsilon,\hbar)| < \pi$ and correspond to the quantization condition of the slow action introduced in Ref. 2, § III, step 5.

We conclude the computation of the approximate eigenvalue (3.5) and the approximate eigenfunction (3.4) by setting

$$\hat{W}(\alpha,\beta) = W_0(\alpha,\beta) + \hat{W}_1(\alpha,\beta) \tag{3.35}$$

and

$$\varphi_1(\alpha,\beta) = U((p\alpha + \beta)/2), \tag{3.36}$$

where W_0 is given by (2.13), $\hat{W}_1$ by (3.12) and (3.15), and finally U by either (3.31) or (3.34), according to the fact that either (3.29) or (3.32) holds. The three terms of the approximate eigenvalue E[(3.10)] are given by (2.12), (3.16), and (3.26) with λ provided by either (3.30) or (3.33).

Step 5 (proof of (3.6) and computation of the reminder): We are now ready to undertake the proof of (3.6). Set

$$\hat{T}_\varepsilon = T_0 + \varepsilon\hat{V}, \quad \tilde{T}_\varepsilon = T_0 + \varepsilon\tilde{V}, \tag{3.37}$$

where $\hat{V}$ and $\tilde{V}$ are, respectively, the nonresonant component (3.8) of the potential $V(\alpha,\beta)$ and the resonant one (3.9). We express $T_\varepsilon\varphi$ with the aid of the pair of operators introduced in (3.37):

$$\begin{aligned} T_\varepsilon\varphi &= T_\varepsilon(e^{(i/\hbar)(W_0 + \hat{W}_1)}\varphi_1) \\ &= (\hat{T}_\varepsilon e^{(i/\hbar)(W_0 + \hat{W}_1)})\varphi_1 + (\tilde{T}_\varepsilon\varphi_1)e^{(i/\hbar)(W_0 + \hat{W}_1)} \\ &\quad - \hbar^2\frac{\partial}{\partial\alpha}e^{(i/\hbar)(W_0 + \hat{W}_1)}\frac{\partial\varphi_1}{\partial\alpha}. \end{aligned} \tag{3.38}$$

For the computation of the first term in the rhs of (3.38) we go back to (2.17) and notice that a similar analysis yields

$$\begin{aligned} &(\hat{T}_\varepsilon e^{(i/\hbar)(W_0 + \hat{W}_1)})\varphi_1 \\ &= E_0 e^{(i/\hbar)(W_0 + \hat{W}_1)}\varphi_1 + \frac{1}{2}\left(\frac{\partial\hat{W}_1}{\partial\alpha}\right)^2\varphi_1 e^{(i/\hbar)(W_0 + \hat{W}_1)}, \end{aligned} \tag{3.39}$$

where E_0 is given by (3.6). As for the other term in the rhs of (3.38) we observe that

$$\begin{aligned} &(\tilde{T}_\varepsilon\varphi_1)e^{(i/\hbar)(W_0 + \hat{W}_1)} - \hbar^2\frac{\partial}{\partial\alpha}e^{(i/\hbar)(W_0 + \hat{W}_1)}\frac{\partial\varphi_1}{\partial\alpha} \\ &= -\frac{\hbar}{2}\left[\frac{\partial^2}{\partial\alpha^2}U\left(\frac{p\alpha+\beta}{2}\right) - 2\frac{\varepsilon V}{\hbar}\cos(p\alpha+\beta)U\left(\frac{p\alpha+\beta}{2}\right)\right]e^{(i/\hbar)(W_0 + \hat{W}_1)} \\ &- \left\{i\omega\hbar\frac{\partial}{\partial\beta}U\left(\frac{p\alpha+\beta}{2}\right) + i\hbar\frac{\partial W_0}{\partial\alpha}\frac{\partial}{\partial a}U\left(\frac{p\alpha+\beta}{2}\right)\right\}e^{(i/\hbar)(W_0 + \hat{W}_1)} + i\hbar\frac{\partial\hat{W}_1}{\partial\alpha}\frac{\partial}{\partial\alpha}U\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0 + \hat{W}_1)}. \end{aligned}$$

By the resonance condition the term within the curly brackets in the rhs of (3.40) vanishes, while the term within the square brackets can be computed by means of (3.24) and (3.26):

$$(\tilde{T}_\varepsilon\varphi_1)e^{(i/\hbar)(W_0+\hat{W}_1)}+\hbar^2\frac{\partial}{\partial\alpha}e^{(i/\hbar)(W_0+\hat{W}_1)}\frac{\partial\varphi_1}{\partial\alpha}$$
$$=\frac{\hbar^2p^2}{8}\lambda U\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0+\hat{W}_1)}+i\frac{\hbar p}{2}\frac{\partial\hat{W}_1}{\partial\alpha}U'\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0+\hat{W}_1)},$$

where λ is given by either (3.30) or (3.33) since U has been chosen accordingly to either (3.31) or (3.34) under the conditions (3.29) and (3.32), respectively.

Now inserting (3.39) and (3.41) in (3.38) and using (3.30) and (3.33) again we get

$$T_\varepsilon\varphi=(E_0+\tilde{E}_1)\varphi+R(\alpha,\beta,E,\varepsilon,\hbar), \tag{3.42}$$

where E_0 is given by (2.12).

$$\tilde{E}_1=\varepsilon V\pm 2\hbar p\sqrt{\varepsilon V}\,\frac{r_\pm(m,\varepsilon,\hbar)}{\log(\varepsilon V)-2\log(\hbar p)+8\log 2-\Gamma'((2\mp1)/4)/\Gamma((2\mp1)/4)}+O(\hbar^2),\quad\text{as}\quad\hbar\to0, \tag{3.43}$$

and

$$R(\alpha,\beta,E,\varepsilon,\hbar)=\frac{1}{2}\left(\frac{\partial\hat{W}_1}{\partial\alpha}\right)^2\varphi+i\frac{\hbar p}{2}\frac{\partial\hat{W}_1}{\partial\alpha}U'\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0+\hat{W}_1)}. \tag{3.44}$$

Clearly (3.43) proves (3.6).

Step 6 (estimate of the remainder): We start the estimate of the remainder (3.44) by observing that the term

$$\left\|\left(\frac{\partial\hat{W}_1}{\partial\alpha}\right)^2\varphi\right\|_{L^2(T^2)}$$

can be treated just like (2.19).

As for the second term in the rhs of (3.44) we obviously obtain

$$\left\|i\frac{\hbar p}{2}\frac{\partial\hat{W}_1}{\partial\alpha}U'\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0+\hat{W}_1)}\right\|_{L^2(T^2)}\leqslant\frac{1}{2}\hbar p\max_{T^2}\left|\frac{\partial\hat{W}_1}{\partial\alpha}\right|\max_{[0,\pi]}|U'|\,\|e^{(i/\hbar)(W_0+\hat{W}_1)}\|_{L^2(T^2)}. \tag{3.45}$$

Then we estimate the term $\max_{[0,\pi]}|U'|$ by the following well known interpolation inequality: there exists a constant $C>0$ such that, for every $\delta\in(0,1)$,

$$\max_{[0,\pi]}|U'|\leqslant C(1/\delta)\max_{[0,\pi]}|U|+\delta\max_{[0,\pi]}|U''|. \tag{3.46}$$

Using Eq. (3.25) and substituting from (3.27), (3.30), and (3.33) we can estimate the second term in the rhs of (3.46) by

$$\delta\max_{[0,\pi]}|(2q\cos(2\mu)-\lambda)U(\mu)|\leqslant\delta\left|8\frac{\varepsilon V}{\hbar^2p^2}+8\frac{\varepsilon V}{\hbar^2p^2}(1+o(1))\right|\max_{[0,\pi]}|U|,$$

as $\hbar\to0$. Inserting this into (3.46) we get

$$\max_{[0,\pi]}|U'|\leqslant(C\delta^{-1}+\delta16\frac{\varepsilon V}{\hbar^2p^2}(1+o(1)))\max_{[0,\pi]}|U|. \tag{3.47}$$

Keeping in mind that the term

$$\max_{T^2}\left|\frac{\partial\hat{W}_1}{\partial\alpha}\right|$$

can be estimated as (2.20), we now choose $\delta=\hbar p/\sqrt{\varepsilon V}$ and insert (3.47) into (3.45); this yields

$$\left\|i\frac{\hbar p}{2}\frac{\partial\hat{W}_1}{\partial\alpha}U'\left(\frac{p\alpha+\beta}{2}\right)e^{(i/\hbar)(W_0-\hat{W}_1)}\right\|_{L^2(T^2)}\leqslant C\varepsilon\sqrt{\varepsilon V}\max_{[0,\pi]}|U|(1+o(1))\|e^{(i/\hbar)(W_0+\hat{W}_1)}\|_{L^2(T^2)}.$$

This concludes the proof of (3.7).

ACKNOWLEDGMENTS

The author gratefully acknowledges Graffi's stimulating advice and J. Herczynski's constructive criticism, which led to many improvements to the paper.

This paper was partially supported by M.P.I., fondi 40% and 60%.

[1] S. Graffi and T. Paul, "The Schrödinger equation and canonical perturbation theory," Commun. Math. Phys. **108**, 25 (1987).

[2] S. Graffi, T. Paul, and H. J. Silverstone, "Resonance overlapping in classical mechanics and avoided crossing in quantum mechanics," Phys. Rev. A **37**, 3835 (1988).

[3] S. Graffi, T. Paul, and H. J. Silverstone "Classical resonance overlapping and quantum avoided crossings," Phys. Rev. Lett. **59**, 255 (1987).

[4] B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," Phys. Rep. **52**, 263 (1979).

[5] K. Yajima, "Resonances for the *A-C*-Stark effect," Commun. Math. Phys. **87**, 331 (1982).

[6] J. Herczynski, "Semiclassical limit for perturbation of non-resonant rotators," Preprint of the Uniwersytet Warszawski, 1988.

[7] J. Herczynski, (oral communications).

[8] H. Poincaré, *Méthodes nouvelles des la mechanique céleste* (Gauthier-Villars, Paris, 1893).

[9] J. Meixner and F. W. Schäfke, *Mathieusche Functionen und Sphäroidfunctionen* (Springer, Berlin, 1954).

[10] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Clarendon, Oxford, 1947).

[11] R. Langer, "The solution of the Mathieu equation with complex variable and at least one parameter large," Trans. Am. Math. Soc. **36**, 637 (1934).

Stationary axisymmetric fluid solutions of the Einstein–Born–Infeld equations

Nora Bretón B.

Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del I. P. N. Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 16 December 1988; accepted for publication 14 June 1989)

A new class of Einstein's field equations with the energy-momentum tensor of a perfect fluid with nonlinear electromagnetic field aligned along the Debever–Penrose directions is determined. The solution obtained is a nonlinear electromagnetic generalization of the fluid NUT-$\tilde{B}(+)$ [Newman–Unti–Tamburino and Carter $B(+)$] solution.

I. INTRODUCTION

Recently several axisymmetric fluid solutions have been determined,[1,2] and it is well known that this class of solutions has great interest in cosmological problems.[3] On the other hand, nonlinear electrodynamics is interesting in cosmological theories as a simple classical model to explain vacuum polarization processes—a possible influence of the mechanism of the evolution of the early universe. In this paper, a metric that can be considered as produced by a perfect fluid with nonlinear electromagnetic field of the Born–Infeld type is given. It may be remarked that there are very few known exact solutions to the Einstein–Born–Infeld (EBI) equations, and that even in the flat space-time, the solutions to the Born–Infeld equations are rather scarce.

This paper is organized as follows. Section II outlines some facts about the nonlinear electrodynamics of the Born–Infeld type[4,5] in terms of the null tetrad formalism. Section III contains the EBI equations for a general type D metric with fluid in the case when the double Debever–Penrose vectors are aligned along the eigenvectors of the nonlinear electromagnetic field. In Sec. IV the stationary axisymmetric fluid metric with nonlinear electromagnetic field is derived, and Sec. V contains some conclusions.

II. THE EINSTEIN–BORN–INFELD EQUATIONS

Here we give a very concise exposition of the basic facts about the nonlinear electrodynamics in terms of the null tetrad formalism of Debney–Kerr–Schild,[6] according to which the metric is given by

$$g = 2e^1 \otimes e^2 + 2e^3 \otimes e^4, \qquad e^2 = \bar{e}^1, \quad e^3 = \bar{e}^3, \quad e^4 = \bar{e}^4, \tag{2.1}$$

where the $e^a \in \Lambda^1$ have to fulfill the first Cartan structure equations

$$de^a = e^b \wedge \Gamma^a{}_b, \tag{2.2}$$

with $\Gamma^a{}_b \in \Lambda^1$ satisfying the second structure equations

$$d\Gamma^a{}_b + \Gamma^a{}_5 \wedge \Gamma^5{}_b = \tfrac{1}{2} R^a{}_{bcd} e^c \wedge e^d. \tag{2.3}$$

The Riemann curvature components $R^a{}_{bcd}$ may be replaced by the Weyl conformal tensor components and the components of the traceless Ricci tensor

$$C_{ab} := R_{ab} - \tfrac{1}{4} g_{ab} R, \qquad R_{ab} := R^q{}_{abq}, \quad R = R^a{}_a. \tag{2.4}$$

In the Born–Infeld nonlinear electrodynamics the electromagnetic field is described by two skew-symmetric tensor fields F_{ab} and P_{ab} which must fulfill the equations

$$F^{ab}{}_{;a} = 0, \quad P^{ab}{}_{;a} = 4\pi j_b, \tag{2.5}$$

where $j_b = \rho u_b$ denotes the electromagnetic current density. We shall use the convention $\check{F}^{ab} = \frac{1}{2}\epsilon_{abcd} F^{cd}$, where ϵ_{abcd} is the Levi-Civita symbol with $\epsilon_{1234} = 1$. The fields F_{ab} and P_{ab} are related by the "constitutive equations"

$$F_{ab} = \left(\frac{\partial \mathscr{H}}{\partial P}\right) P_{ab} + \left(\frac{\partial \mathscr{H}}{\partial \check{Q}}\right) \check{P}_{ab}, \tag{2.6}$$

where $\mathscr{H} = b^2 - (b^4 - 2b^2 P + \check{Q})^{1/2}$ is the Born–Infeld structural function, b is a positive real constant, and

$$P = \tfrac{1}{4} P_{ab} P^{ab}, \quad \check{Q} = \tfrac{1}{4} P_{ab} \check{P}^{ab} \tag{2.7}$$

are the invariants of P_{ab}; P is real and $\check{Q}$ is pure imaginary. In the limit when $b \to \infty$ one recovers the linear theory.

III. EINSTEIN EQUATIONS COUPLED TO NONLINEAR ELECTROMAGNETIC FIELD

In this paper, we are concerned with solutions to the Einstein equations,

$$R_{ab} - \tfrac{1}{2} g_{ab} R = -T_{ab}, \tag{3.1}$$

$$T_{ab} = (\not{p} + \epsilon) u_a u_b + \not{p} g_{ab} - 8\pi E_{ab}, \tag{3.2}$$

$$u_a u^a = -1, \quad \not{p} + \epsilon \neq 0, \tag{3.3}$$

where the energy-momentum tensor of the nonlinear electromagnetic field is given by

$$4\pi E_{ab} = \mathscr{H}_P(-P_{as} P_b{}^s + g_{ab} P) + (P\mathscr{H}_P + \check{Q}\mathscr{H}_{\check{Q}} - \mathscr{H}) g_{ab}, \tag{3.4}$$

and u_a is the (timelike) fluid four-velocity, $\not{p}$ is the fluid pressure, and ϵ is the energy density.

The system of field equations is completed by the Maxwell–Faraday equations (2.5). Furthermore, we are assuming that there exists a known equation of state such that the energy density is given by $\epsilon = \epsilon(n,s)$, where n denotes the particle density and s the entropy density of a fluid point; the fluid flow may be characterized by (1) a Eulerian velocity

 0022-2488/89/112607-04$02.50

$u^\mu(x^\nu)$ defined for every event $P(x^\nu)$ situated in its world trajectory, (2) the specific volume $V=1/n$, and (3) the phenomenologic temperature T and pressure p encountered. These hydrodynamic variables are measured with respect to a local rest frame. The velocity is normalized such that $u_\mu u^\mu = -1$. Since the fluid is isotropic and frictionless, entropy is conserved. Therefore

$$u^\alpha s_{,\alpha} = 0$$

and

$$(nu^\alpha)_{;\alpha} = 0. \tag{3.5}$$

The first and second laws of thermodynamics,

$$d\epsilon = nT\,ds + [(\epsilon + p)/n]dn, \tag{3.6}$$

are postulated.

IV. DERIVATION OF THE SOLUTION

We shall consider the Carter metric[7]

$$g(-1,0) = (1/H^2)\{(N/P)dx^2 + PN\,dy^2 + (N/Q)dy^2 - Q/N(-d\tau + Md\sigma)^2\}, \tag{4.1}$$

with $H=H(x,y)$, $P=P(x)$, $Q=Q(y)$, $N=N(y)$, and $M=2lx$. We choose the null tetrad in the form

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = (1/\sqrt{2}H)\{(\sqrt{N/P})dx \pm i\sqrt{PN}\,d\sigma\},$$
$$\left.\begin{matrix} e^3 \\ e^4 \end{matrix}\right\} = \frac{1}{\sqrt{2}H}\{(\sqrt{N/Q})dy \pm (\sqrt{Q/N})(-d\tau + M\,d\sigma)\}. \tag{4.2}$$

The connection one-forms for the tetrad (4.2) are given by[2]

$$\Gamma_{42} = H\sqrt{\frac{Q}{2N}}\left[\frac{Hy}{H} - \frac{Ny}{2N} - \frac{i}{2}\frac{Mx}{N}\right]e^1 - H\left(\frac{P}{2N}\right)^{-1}\left[\frac{Hx}{H}\right]e^3 =: -Ze^1 + Ae^3,$$
$$\Gamma_{31} = -Ze^2 + Ae^4, \tag{4.3}$$
$$\Gamma_{12} + \Gamma_{34} = \frac{H}{2}\left(\frac{P}{2N}\right)^{-1}\left[2\frac{Hx}{H} - \frac{Px}{P}\right](e^1 - e^2) + \frac{H}{2}\left(\frac{Q}{2N}\right)^{-1}\left[2\frac{Hy}{H} - \frac{Qy}{Q} + \frac{Ny}{N} + i\frac{Mx}{N}\right](e^3 - e^4),$$

and the nonvanishing components of the traceless Ricci tensor C_{ab} read as

$$C_{11} = C_{22} = (H^2P/2N)[-2(Hxx/H)],$$
$$C_{33} = C_{44} = -\frac{H^2Q}{2N}\left[2\frac{Hyy}{H} - \frac{Nyy}{N} + \frac{1}{2}\frac{(Mx)^2 + (Ny)^2}{N^2}\right],$$
$$C_{13} = C_{24} = \frac{H}{2N}\sqrt{PQ}\left[-2Hxy + \frac{Ny}{N}Hx + \frac{i}{N}\left(-MxHx + \frac{Mxx}{2}H\right)\right], \tag{4.4}$$
$$C_{12} = \frac{1}{4}\frac{H^2}{N}\left[Pxx - 2Px\left(\frac{Hx}{H}\right) - Qyy + 2Qy\left(\frac{Hy}{H} + \frac{Ny}{N}\right) - P\left(-2\frac{Hxx}{H}\right)\right] + Q\left[2\frac{Hyy}{H} - 4\frac{Ny}{N}\frac{Hy}{H} - \frac{3}{2}\frac{(Mx)^2 + (Ny)^2}{N^2}\right].$$

Aligning the eigenvectors of F_{ab} (and P_{ab}) in the directions of the DP vector e^3 and e^4, the only nonvanishing components of F_{ab} are F_{12} and F_{34} (P_{12} and P_{34}), and the two-form ω assumes the form

$$\omega = (F_{12} + P_{34})e^1 \wedge e^2 + (F_{34} + P_{12})e^3 \wedge e^4. \tag{4.5}$$

In the Born–Infeld nonlinear electromagnetism

$$\mathscr{H} = b^2 - (b^4 - 2b^2P + \check{Q})^{1/2}, \tag{4.6}$$

we parametrize the nonzero components of P_{ab} as

$$P_{12} = ib[1 - \exp(-2\nu)]^{1/2}\sin\varphi = ib\sin\varphi G^{(-)},$$
$$P_{34} = b[\exp(2\nu) - 1]^{1/2}\cos\varphi = b\cos\varphi G^{(+)}, \tag{4.7}$$

with ν, φ real functions. In this parametrization,

$$F_{12} = i\exp(\nu)b\sin\varphi G^{(-)},$$
$$F_{34} = \exp(-\nu)b\cos\varphi G^{(+)}. \tag{4.8}$$

Considering the field equations (3.1)–(3.4) with $u^\mu = \delta^\mu{}_\tau H/\sqrt{Q}$, and assuming that $u^4 = u_3 = -u_4 = -u^3 = 1/\sqrt{2}$, the trace of the momentum-energy tensor is

$$T = 3\not{\epsilon} - \epsilon + (1/\pi)(P\mathscr{H}_P + \check{Q}\mathscr{H}_{\check{Q}} - \mathscr{H}), \tag{4.9}$$

or in the (4.8) parametrization

$$T = 3\not{\epsilon} - \epsilon + (b^2/\pi)(\cosh\nu - 1), \tag{4.10}$$

the non-null components of the fluid part of the momentum–energy tensor are

$$T_{12} = \not{\epsilon}, \quad T_{33} = T_{44} = \tfrac{1}{2}(\not{\epsilon} + \epsilon), \quad T_{34} = \tfrac{1}{2}(\not{\epsilon} - \epsilon). \tag{4.11}$$

The nonvanishing components of the traceless Ricci tensor are

$$C_{33} = C_{44} = -\tfrac{1}{2}(\not{\epsilon} + \epsilon),$$
$$C_{12} = -C_{34} = \tfrac{1}{4}(\not{\epsilon} + \epsilon) - 2E_{34} + \tfrac{1}{4}\,\mathrm{Tr}\,E_{ab}. \tag{4.12}$$

By considering that in the BI nonlinear electromagnetism[8]

$$\left.\begin{matrix} 8\pi E_{12} \\ 8\pi E_{34} \end{matrix}\right\} = \pm\mathscr{H}_P(P_{12}^2 - P_{34}^2) + 2(P\mathscr{H}_P - \check{Q}\mathscr{H}_{\check{Q}} - \mathscr{H}), \tag{4.13}$$

Eqs. (4.12) take the form

$$-C_{12} = C_{34} = \tfrac{1}{4}(\not{\epsilon} + \epsilon) - (b^2/4\pi)\sinh\nu. \tag{4.14}$$

From Eqs. (4.12) and (4.4) for the traceless Ricci tensor components, one obtains the energy-pressure equations

$$\not{r} + \epsilon = H^2 Q[(2\ddot{H}/H) + 2l^2] \tag{4.15}$$

and

$$\begin{aligned} C_{12} &= -\tfrac{1}{4}(\not{r} + \epsilon) + (b^2/4\pi)\sinh\nu \\ &= (-H^2/4)\{-2\epsilon_0 - \ddot{Q} + 2Q(\dot{H}/H) \\ &\quad - Q[(2\ddot{H}/H) - 6l^2]\}, \end{aligned} \tag{4.16}$$

where we have made the assumption that $P = a + dx - \epsilon_0 x^2$, with a,d constants, and $\epsilon_0 = (-1,0,1)$ and $H = H(y)$, $\dot{H} = dH(y)/dy$, etc.

By substituting Eq. (4.15) in Eq. (4.16), one arrives at

$$\begin{aligned} &(b^2/4\pi)\sinh\nu + (H^2/4)\{-2\epsilon_0 - \ddot{Q} + 2Q(\dot{H}/H) \\ &\quad - 4Q[(\ddot{H}/H) - l^2]\} = 0. \end{aligned} \tag{4.17}$$

The right-hand side of Eq. (4.17) depends only on the variable y, so the left-hand side must depend only on y, i.e., $\nu = \nu(y)$. Furthermore, Eq. (2.5) must be fulfilled, these equations are the electromagnetic current equations, and there is a requirement that F_{ij} be a curl; in tensorial language we have

$$\begin{aligned} &P^{\mu\nu}{}_{;\nu} = 4\pi j^{\mu} \\ &\text{and} \\ &F_{\mu\nu} = A_{\mu;\nu} - A_{\nu;\mu}. \end{aligned} \tag{4.18}$$

The solution also must satisfy the continuity equation $j^i{}_{;i} = 0$. We suppose that the vectorial potential depends only on x and y and also we are aligning the eigenvectors of F_{ij} with the DP vectors; this condition amounts to

$$\begin{aligned} &F_{\sigma x} = iNF_{12}/H^2 = A_{\sigma,x}, \\ &F_{\tau x} = 0 = A_{\tau,x}, \\ &F_{\tau y} = -F_{34}/H^2 = A_{\tau,y}, \\ &F_{\sigma y} = MF_{34}/H^2 = A_{\sigma,y}. \end{aligned} \tag{4.19}$$

From these equations, expressions (4.8), and the integrability conditions for $A_{\mu\nu}$, $A_{\sigma,xy} = A_{\sigma,yx}$, we obtain the condition

$$\begin{aligned} &4le^{-\nu}\cos\phi\sinh\nu \\ &\quad = 4\dot{H}\sin\phi\sinh\nu/H \\ &\qquad - 2\dot{\phi}\cos\phi\sinh\nu - \dot{\nu}e^{\nu}\sin\phi, \end{aligned} \tag{4.20}$$

while the electromagnetic currents are given by

$$\begin{aligned} &j^x = 0 = j^y = j^{\sigma} = 0, \\ &j^{\tau} = -lbe^{cy/2}[cy/(cy+1)]^{1/2}/2\pi. \end{aligned} \tag{4.21}$$

Therefore, solutions to Eqs. (4.17) and (4.20) complete the nonlinear electromagnetic generalization to the fluid NUT-$B(+)$ solution.

We shall consider the case when $\phi = \pi/2$; then Eq. (4.20) turns out to be

$$4\dot{H}\sinh\nu/H - \dot{\nu}e^{\nu} = 0. \tag{4.22}$$

One solution to (4.22) is $H = e^{cy/4}$ and $\nu = \ln(cy+1)^{1/2}$, $c = \text{const}$. Then (4.17) amounts to

$$\begin{aligned} &b^2 e^{-cy/2}[(e^{cy}+1)^{1/2} - (e^{cy}+1)^{-1/2}]/2\pi - 2\epsilon_0 \\ &\quad = \ddot{Q} - c\dot{Q}/2 + Q(c^2/4 - 4l^2). \end{aligned} \tag{4.23}$$

This is a second-order linear inhomogeneous differential equation for $Q = Q(y)$, and the general solution to (4.23) is given by

$$Q(y) = c_1 u_1(y) + c_2 u_2(y) + u_p(y), \tag{4.24}$$

where u_1, u_2 are two solutions of the associated homogeneous equation, c_1, c_2 are arbitrary constants, and u_p is any particular solution of the inhomogeneous equation.

One solution is $u_1(y) = e^{ky}$ with $k = (c/4) \pm \frac{1}{4}[-3c^2 + 64l^2]^{1/2}$, for this solution $l \neq 0$; and $u_2 = De^{-ky + cy/2}$, $D = \text{const}$. Given u_1 and u_2, u_p can be determined:

$$\begin{aligned} u_p = \frac{1}{D}\int_{y_0}^{y} F(\xi)\exp\left[\int p(\xi)d\xi\right] \\ \times [u_2(y)u_1(\xi) - u_1(y)u_2(\xi)]d\xi, \end{aligned} \tag{4.25}$$

with

$$F(\xi) = b^2 e^{-cy/2}[(e^{cy}+1) - (e^{cy}+1)]/2\pi - 2\epsilon_0$$

and $p(\xi) = c/2$. This leads to integrals of the form $\int e^{A\xi}(e^{c\xi}+1)^{1/2}\,d\xi$ which can be integrated numerically in the interval $-1 < y < 1$. From Eq. (4.15) we see that the condition for $\not{r} + \epsilon \geqslant 0$ is $Q \geqslant 0$, so this solution has a physical meaning.

The equations of motion (Lorentz equations) for the charged fluid are satisfied in the example presented.

For the pressure of the fluid one obtains the explicit expression from the scalar curvature

$$\begin{aligned} \not{r} = -\frac{b^2}{4\pi}(\cosh\nu(y) - 1) + \frac{H^2}{4} \\ \times\left\{-\epsilon_0 + \ddot{Q} - \dot{Q}\left(6\frac{\dot{H}}{H}\right) - 4Q\left(\frac{\ddot{H}}{H} - 3\left(\frac{\dot{H}}{H}\right)^2\right)\right\}, \end{aligned} \tag{4.26}$$

while from the Bianchi identities, one obtains

$$\begin{aligned} &\not{r}_x = 0, \\ &\not{r}_y = -\frac{b^2}{4\pi}\sinh\nu(y) + \frac{1}{4}\partial_y\left\{H^2\left[-\epsilon_0 + \ddot{Q}\right.\right. \\ &\qquad \left.\left. - \dot{Q}\left(6\frac{\dot{H}}{H}\right) - 4Q\left(\frac{\ddot{H}}{H} - 3\left(\frac{\dot{H}}{H}\right)^2\right)\right]\right\}, \end{aligned} \tag{4.27}$$

which hold for $\not{r}$ from Eq. (4.26).

We can see from Eq. (4.26) that the pressure is affected by the nonlinear electromagnetic field and that in the limit of linear field the expression seems to be of the same form as that in absence of electromagnetic field.[2]

V. CONCLUSIONS

A nonlinear electromagnetic generalization of the NUT-$\tilde{B}(+)$ metric with fluid is presented here. The field equations are reduced to a pair of nonlinear differential equations for three functions of one variable. This fact is the result of the alignment of the nonlinear electromagnetic field with the DP directions.

The cosmological constant λ can be incorporated into an energy-momentum tensor of the perfect fluid type by substituting $(\not{r} - \lambda)$ for $\not{r}$, and $(\epsilon + \lambda)$ for ϵ; of course this may violate the condition $\epsilon > 0$.

This solution requires a deeper analysis in order to find the corresponding topology. It is also of interest to investigate the possibility of regarding this solution as an interior metric.

ACKNOWLEDGMENTS

The author acknowledges Dr. G. F. Torres del Castillo for helpful discussions.

The author also acknowledges support received from the Sistema Nacional de Investigadores (Mexico).

[1]E. Herlt, Gen. Relativ. Gravit. **20**, 635 (1988).
[2]A. García Díaz and J. Téllez Ulloa, "Stationary axisymmetric rotating perfect fluid type D solutions," CINVESTAV preprint, 1988.
[3]C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[4]J. F. Plebañski, "Lectures on nonlinear electrodynamics," monograph of the Niels Bohr Institute, Nordita, Copenhagen, 1968.
[5]S. Alarcón, A. L. Dudley, and J. F. Plebañski, J. Math. Phys. **22**, 2835 (1981).
[6]G. C. Debney, R. P. Kerr, and A. Schild, J. Math Phys. **10**, 1842 (1969).
[7]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[8]D. A. García, I. H. Salazar, and J. F. Plebañski, Nuovo Cimento **84**, 65 (1984).

A new approach to Bianchi V models

Carles Bona
Departament de Física, Universitat de les Illes Balears, E-07071, Palma de Mallorca, Spain
Pere Palou
Institut Politécnic de Formació Profesional, E-07071, Palma de Mallorca, Spain

(Received 4 April 1989; accepted for publication 12 July 1989)

Spatially homogeneous perfect fluid space-times pertaining to the Bianchi V class are considered from the point of view of the evolution (3 + 1) formalism of general relativity. It is shown that the metric can be written in diagonal form in the nontilted case. The necessary and sufficient condition for the metric of a Bianchi V model to be diagonalizable is given.

I. INTRODUCTION

Spatially homogeneous models have been widely considered in cosmology as generalizations of the standard Friedmann–Robertson–Walker space-times. The (spatial) homogeneity means that there is an isometry group whose orbits are spacelike three-dimensional surfaces. This provides a privileged foliation of the space-time and this suggests that a 3 + 1 splitting of the four-dimensional geometry may be suitable for hypersurface-homogeneous models.

In the case of Bianchi V models,[1] the isometry group is generated by three Killing vectors m such that

$$[m_1,m_2]=0 \quad [m_1,m_3]=m_1, \quad [m_2,m_3]=m_2, \tag{1}$$

and the line element is given by

$$ds^2 = -dt^2+\gamma_{ab}(t)\omega^a\omega^b, \tag{2}$$

where the vectors m_a and the one-forms ω^a can be written in an adapted coordinated system as follows:

$$m_1=\frac{\partial}{\partial x}, \quad m_2=\frac{\partial}{\partial y}, \quad m_3=x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}-\frac{\partial}{\partial z}, \tag{3}$$

$$\omega^1=e^z\,dx, \quad \omega^2=e^z\,dy, \quad \omega^3=dz. \tag{4}$$

The Bianchi V line element (2) is said to be diagonal if the off-diagonal terms in the matrix of coefficients γ_{ab} vanish. This is usually taken as a simplifying hypothesis in constructing explicit solutions.[2,3] We will show that the metric of a Bianchi V perfect fluid model is always diagonalizable in the nontilted case.

II. A NEW COORDINATE SYSTEM

A direct calculation starting from (2) shows that the spatial hypersurfaces (the group orbits) are locally constant curvature manifolds. Space-times admitting a constant curvature foliation have been considered in other works.[4] On the other hand, it is clear that there exist local coordinate systems [different from the one used in (3) and (4)] in which this geometrical property of the space metric is more transparent. We will look for a coordinate transformation that carries the line element (2) to the form

$$ds^2 = -dt^2+\gamma_{ij}(dx^i+\beta^i\,dt)(dx^j+\beta^j\,dt) \quad (i,j=1,2,3) \tag{5}$$

(3 + 1 decomposition) with the space metric γ_{ij} being of the form

$$\gamma_{ij}=R(t)^2\,\mathrm{diag}(e^{2z},e^{2z},1). \tag{6}$$

This can be achieved in three steps.

(i) The transformation

$$x\to x+e^{-z}f(t), \quad y\to y+e^{-z}g(t), \tag{7}$$

where $f(t)$ and $g(t)$ must be chosen in order to cancel the two off-diagonal terms containing dz. The resulting coefficient of dz^2 is to be identified with $R(t)^2$.

(ii) The rotation in the $x-y$ plane that cancels the $dx\,dy$ term so that the space metric is written in diagonal form:

$$\gamma_{ij}=\mathrm{diag}\{A(t)^2e^{2z}, B(t)^2e^{2z}, R(t)^2\}. \tag{8}$$

(iii) The two separate dilations of the new x and y axis such that the final form (6) is obtained.

Of course, these transformations introduce new off-diagonal terms that must be taken into account by the shift vector β^i. It will be of the form

$$\begin{aligned}\beta^x &= a(t)x+b(t)y+h_1(t)e^{-z},\\ \beta^y &= b'(t)x+c(t)y+h_2(t)e^{-z},\\ \beta^z &= 0,\end{aligned} \tag{9}$$

where the six coefficients are related with (the time derivatives of) the five transformation parameters f, g, etc. The presence of a redundant parameter is related with the invariance of the space metric (6) under time-dependent rotations in the $x-y$ plane. This can be used to get $b(t)=b'(t)$ in (9) so that the final form of the four-dimensional metric (5) contains the same number of arbitrary functions [five in β^i plus $R(t)$ in γ_{ij}] as the original form (2).

Note that all the transformations we have made preserve both the z coordinate and the form of the third Killing vector in (3). The first two Killing vectors will now appear as time-dependent combinations of $\partial/\partial x$ and $\partial/\partial y$, which are not by themselves Killing vectors in the generic case.

III. NONTILTED MODELS

A homogeneous model is said to be nontilted if the normal n to the group orbits is an eigenvector of the Ricci ten-

 0022-2488/89/112611-03$02.50

sor. Allowing for the constraint equations in the evolution (3 + 1) formalism,[5] this condition can be written

$$D_j K^j{}_i - \partial_t(K^j{}_j) = 0, \tag{10}$$

where D is the covariant derivative operator associated to the three-dimensional metric γ_{ij} and K_{ij} is the extrinsic curvature of the spatial hypersurfaces; in our case

$$K_{ij} = -\tfrac{1}{2}[\partial_t - L_{(\beta)}](\gamma_{ij}), \tag{11}$$

where L is the three-dimensional Lie derivative operator.

A straightforward computation starting from (6) and (9) shows that the vector equation (10) is equivalent to the three algebraic conditions

$$h_1(t) = h_2(t) = 0, \quad c(t) = -a(t), \tag{12}$$

so that the nontilted Bianchi V metric depend only on the three functions $a(t)$, $b(t)$,and $R(t)$. The spatial components of the Ricci tensor can now be computed from the evolution equations in the 3 + 1 formalism[5]:

$$\begin{aligned} &R_{11} = R^2 e^{2z}[-2/R^2 - (\partial_t a + 3\theta a) + \partial_t\theta + 3\theta^2], \\ &R_{12} = -R^2 e^{2z}(\partial_t b + 3\theta b), \quad R_{13} = 0, \\ &R_{22} = R^2 e^{2z}[-2/R^2 + (\partial_t a + 3\theta a) + \partial_t\theta + 3\theta^2], \\ &R_{23} = 0, \quad R_{33} = R^2[-2/R^2 + \partial_t\theta + 3\theta^2], \end{aligned} \tag{13}$$

where we have taken for short

$$\theta(t) = \frac{\partial_t R}{R}. \tag{14}$$

In the case of a perfect fluid model, the algebraic structure of the Ricci tensor is

$$R_{\mu\nu} = (\rho + p)u_\mu u_\nu + [(\rho - p)/2]g_{\mu\nu} \quad (\mu,\nu = 0,1,2,3), \tag{15}$$

So that in the nontilted case $u_1 = 0$ and Eqs. (13) read

$$\partial_t a + 3\theta a = \partial_t b + 3\theta b = 0, \tag{16}$$

$$(\rho - p)/2 = -2/R^2 + \partial_t\theta + 3\theta^2, \tag{17}$$

where the energy density ρ is given by the scalar constraint equation[5]

$$\rho = -3/R^2 + 3\theta^2 - (a^2 + b^2), \tag{18}$$

so that the pressure p can be computed from (17):

$$p = 1/R^2 - 2\,\partial_t\theta - 3\theta^2 - (a^2 + b^2). \tag{19}$$

Both equations (16) are easily integrated allowing for (14):

$$a(t) = k_1/R^3, \quad b(t) = k_2/R^3, \tag{20}$$

where k_1 and k_2 are arbitrary constants. If we now substitute (20) and (12) into the expression (9) for the shift vector, we obtain

$$\begin{aligned} &\beta^x = (k_1 x + k_2 y)/R^3, \\ &\beta^y = (k_2 x - k_1 y)/R^3, \quad \beta^z = 0, \end{aligned} \tag{21}$$

and it is clear that we can perform a (time-independent!) rotation in the $x - y$ plane in order to obtain

$$\beta^x = kx/R^3, \quad \beta^y = -ky/R^3, \quad \beta^z = 0, \tag{22}$$

where we have noted $k^2 = k_1^2 + k_2^2$. We can collect our results as follows.

Theorem 1: The general form of the metrics of nontilted homogeneous perfect fluids of Bianchi class V is given by

$$ds^2 = -dt^2 + R(t)^2\{e^{2z}[(dx + kx/R^3\,dt)^2 + (dy - ky/R^3\,dt)^2] + dz^2\}, \tag{23}$$

where k is constant and $R(t)$ is an arbitrary function of time. The energy density and the pressure of the fluid are given, respectively, by

$$\begin{aligned} &\rho = 3\theta^2 - 3/R^2 - k^2/R^6, \\ &p = -2\,\partial_t\theta - 3\theta^2 + 1/R^2 - k^2/R^6. \end{aligned} \tag{24}$$

Corollary: The metrics of the nontilted Bianchi V perfect fluids are diagonalizable: they can be written in the form

$$ds^2 = -dt^2 + R(t)^2\{e^{2z}[e^{-2\lambda(t)}dx^2 + e^{2\lambda(t)}\,dy^2] + dz^2\}, \tag{25}$$

where the two functions $\lambda(t)$ and $R(t)$ are related one to another by

$$R(t)^3\,\partial_t\lambda = k \quad \text{(constant)}. \tag{26}$$

IV. TILTED MODELS

The situation is much more complicated in the tilted case, because one cannot start from the algebraic conditions (12) and the general form (9) of the shift vector (with $b' = b$) must be used. We will restrict ourselves to the case in which the spatial velocity of the fluid is directed along the z axis [remember that our z coordinate coincides with the original one in (2)]. It follows from the constraint equations that the first two equations (12) (but not the third one) also hold true in that case,

$$h_1(t) = h_2(t) = 0. \tag{27}$$

By making $R_{12} = 0$ and $R_{22} = R_{33}$, we obtain, respectively,

$$\partial_t b + b(3\theta - 2a_+) = 0, \tag{28a}$$

$$\partial_t a_- + a_-(3\theta - 2a_+) = 0, \tag{28b}$$

where we have noted

$$a_+ = (a + c)/2, \quad a_- = (a - c)/2, \tag{29}$$

and this will be enough for us. Of course, the evolution equation for a_+ (which was zero in the nontilted case) and the expressions for ρ and p could also be obtained, but they are rather complicated and we will not use them in what follows.

Let us introduce now the auxiliary function $\lambda_+(t)$ such that

$$\partial_t\lambda_+ = a_+(t), \tag{30}$$

so that Eqs. (27) can be easily integrated:

$$a_- = k_1/R^3\,e^{2\lambda_+}, \quad b = k_2/R^3\,e^{2\lambda_+}, \tag{31}$$

where k_1 and k_2 are constants. Allowing for (27) and (31), the components of the shift vector β^i can be expressed now:

$$\beta^x = a_+(t)x + (k_1 x + k_2 y)/R^3 e^{2\lambda_+},$$
$$\beta^y = a_+(t)y + (k_2 x - k_1 y)/R^3 e^{2\lambda_+}, \tag{32}$$
$$\beta^z = 0,$$

and again we can perform a time-independent rotation such that

$$\beta^1 = [a_+(t) + k/R^3 e^{2\lambda_+}]x,$$
$$\beta^2 = [a_+(t) - k/R^3 e^{2\lambda_+}]y, \tag{33}$$
$$\beta^3 = 0,$$

which amounts to taking $k_2 = 0$ ($b(t) = 0$).

The form (33) of the shift vector implies that the four-dimensional metric is diagonalizable. To see this, let us introduce the function $\lambda_-(t)$,

$$\partial_t \lambda_- = a_-(t) = k/R^3 e^{2\lambda_+} \tag{34}$$

and let us consider the coordinate transformation

$$x \to x \exp[-(\lambda_+ + \lambda_-)], \quad y \to y \exp[-(\lambda_+ - \lambda_-)], \tag{35}$$

which puts the space-time metric into diagonal form. Our results can be stated as follows.

Theorem 2: The metrics of Bianchi V models in which the spatial velocity of the fluid is directed along the z axis are diagonalizable. They can be written in the form

$$ds^2 = -dt^2 + R(t)^2\{\exp[2z - 2\lambda_+(t)] \times [e^{-2\lambda_-(t)} dx^2 + e^{2\lambda_-(t)} dy^2] + dz^2\}. \tag{36}$$

The nontilted models correspond to the case $\lambda_+ = \text{const}$.

Corollary: The necessary and sufficient condition for the metric (2) of a Bianchi V perfect fluid model to be diagonalizable is that the spatial velocity of the fluid be directed along the z axis.

Proof: Theorem 2 states that the condition is sufficient. To see that it is necessary, one can start with the general diagonal form (36) and verify, by using the constraint equations, that the tilting is along the z axis.

V. CONCLUDING REMARKS

A lot of work has been done by many authors in the nontilted case,[2] starting with diagonal metrics as a simplifying hypothesis. In Sec. III, we have shown that the results obtained by these authors do actually apply to general nontilted Bianchi V models.

The situation is slightly different in the tilted case, where most of the work has been done under the assumption of local rotational symmetry. This means that the space-time admits a fourth Killing vector field that can be written in the notation of Sec. I,

$$m_4 = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}, \tag{37}$$

so that, in this particular case, it is known that the metric can be put into diagonal form.[6]

It is easy to see that this case corresponds to making our assumption (27) plus the supplementary condition

$$c(t) = a(t), \tag{38}$$

and the metric can be put into the diagonal form (36) with $\lambda_- = \text{const}$. In Sec. IV, we have extended this result by giving a necessary and sufficient condition for a Bianchi V metric to be diagonalizable, which does not assume the condition (38).

ACKNOWLEDGMENT

The authors acknowledge financial support from the CAICYT of Spain under Project No. PB87-0583.

[1] See for instance, F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, J. Math. Phys. **9**, 497 (1968).

[2] R. A. Matzner, Astrophys. J. **157**, 1085 (1969); G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969); V. A. Ruban, Sov. Phys. JETP **45**, 629 (1977); B. K. Nayak, Gen. Relativ. Gravit. **15**, 1067 (1983); B. K. Nayak and G. B. Bhuyan, *ibid.* **18**, 79 (1986).

[3] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967); D. L. Farnsworth, *ibid.* **8**, 2315 (1967); J. M. Stewart and G. F. R. Ellis, *ibid.* **9**, 1072 (1968); I. S. Shikin, Commun. Math. Phys. **26**, 24 (1972); A. R. King and G. F. R. Ellis, *ibid.* **31**, 209 (1973); C. B. Collins, *ibid.* **39**, 131 (1974).

[4] C. Bona and B. Coll, J. Math. Phys. **26**, 1583 (1983); **29**, 327 (1987).

[5] See for instance, R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation*, edited by L. Witten (Wiley, New York, 1962).

[6] See the works of Farnsworth and King and Ellis in Ref. 3.

Killing spinors and separability of Maxwell's equations

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, 72000 Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 24 January 1989; accepted for publication 7 June 1989)

It is shown that the separation constant not related to the space-time symmetries, which appears in the solution of the source-free Maxwell equations on a type-D vacuum background with cosmological constant, can be defined in an invariant way as the eigenvalue of a differential operator made out of a two-index Killing spinor, with the eigenfunctions being the separable solutions of the Maxwell equations.

I. INTRODUCTION

In a recent paper[1] it has been shown that the separable solutions of the source-free Maxwell equations on a type-D vacuum space-time can be characterized by means of a differential operator constructed from a two-index Killing spinor. This differential operator provides, at the same time, a covariant definition of the Starobinsky constant, which arises from the differential relations that connect the separated functions (the so-called Teukolsky–Starobinsky identities).

Each separated function that depends on one of the non-ignorable coordinates satisfies an ordinary differential equation that involves a separation constant whose existence is not related to the space-time symmetries (this separation constant is analogous to Carter's "fourth constant" found in the case of the Hamilton–Jacobi equation in the Kerr background). In Ref. 2, by using the expressions for the electromagnetic field in terms of Debye potentials, it has been shown that in the specific case of the Kerr metric, which is of type D, the separation constant mentioned above can be characterized in a covariant way in terms of the Killing–Yano tensor admitted by the metric.

Killing spinors and Killing–Yano tensors have been investigated previously in connection with the separability of the Hamilton–Jacobi equation,[3] the Dirac equation,[4,5] the Weyl neutrino equation,[6] and Maxwell's equations.[7] (See also Refs. 8 and 9.) The existence of a Killing–Yano tensor imposes additional conditions to those implied by the existence of a two-index Killing spinor; whereas all the type-D vacuum metrics admit a two-index Killing spinor,[3] not all of them admit a Killing–Yano tensor. Nevertheless, in the case of massless fields, or massless particles, the separability of the corresponding equations is associated with the existence of a two-index Killing spinor.

The aim of this paper is to show that for all the type-D vacuum metrics with a cosmological term, the separation constant referred to above, appearing in the solution of the source-free Maxwell equations, can be defined in a covariant way as the eigenvalue of a certain differential operator made out of the existing two-index Killing spinor, with the eigenfunctions being the separable solutions of Maxwell's equations.

This fact can be proven by using the explicit expression for the complete solution of the source-free Maxwell equations obtained by means of the procedure given in Ref. 1. Table I contains the complete solution of Maxwell's equations on all the type-D vacuum metrics, following the notation of Ref. 10. (The separability of certain components of the electromagnetic spinor on all the type-D vacuum metrics was originally established in Refs. 11 and 12.)

II. COVARIANT CHARACTERIZATION OF THE SEPARATION CONSTANTS

The separable solutions of the source-free Maxwell equations involve three separation constants, denoted as k, l, and A in Refs. 1 and 10; the separation constants k and l are related to the two-dimensional Abelian isometry group admitted by the type-D vacuum metrics and, apart from a factor i, they can be defined in a covariant way as the eigenvalues of the Lie derivatives with respect to the Killing vectors ∂_u and ∂_v, respectively, where u and v are the ignorable coordinates.[10]

The separation constant A, by contrast, is not related to space-time symmetries; it turns out that A is the eigenvalue of a second-order differential operator constructed from the two-index Killing spinor, L_{AB}, which a type-D vacuum metric possesses. In fact, a straightforward computation, making use of the ordinary differential equations satisfied by the separated functions, the Teukolsky–Starobinsky identities,[10] and the expressions given in Table I, shows that if φ_{AB} is a separable solution of the source-free Maxwell equations then

$$\nabla_{C'(B}X_{D)}{}^{C'} = A\varphi_{BD}, \tag{1}$$

where

$$X_{AB'} \equiv \tfrac{1}{2} L_{B'C'}\{\phi\bar{\phi}^{-1}\nabla_M{}^{C'}(\phi^{-1}\bar{\phi}L^{MD}\varphi_{DA}) - L_{AD}\varphi^{DM}\bar{\phi}^{-1}\nabla_M{}^{C'}\bar{\phi}\} \tag{2}$$

and

$$L_{B'C'} \equiv \overline{L_{BC}}, \quad \phi^{-2} = -\tfrac{1}{2} L_{AB}L^{AB}.$$

With respect to a frame such that $L_{00} = 0 = L_{11}$, using the source-free Maxwell equations in order to eliminate φ_1 in favor of φ_0 and φ_2, which have simpler expressions, one finds that the components of $X_{AB'}$, defined in (2), are given by

$$X_{00'} = \tfrac{1}{2}(\phi\bar{\phi})^{-1}[(D - 2\rho)\varphi_1 + (\bar{\delta} - 2\alpha + 3\pi + 2\bar{\tau})\varphi_0]$$
$$= (\phi\bar{\phi})^{-1}(\bar{\delta} - 2\alpha + 2\pi + \bar{\tau})\varphi_0,$$

 0022-2488/89/112614-02$02.50

TABLE I. Complete solution of the source-free Maxwell equations. The standard components of the electromagnetic spinor with respect to the null tetrads defined in Ref. 10 are given by $\varphi_0 = \Phi_0 e^{i(ku+lv)}$, $\varphi_1 = (1/\sqrt{2})\phi\Phi_1 e^{i(ku+lv)}$, $\varphi_2 = (1/2)\phi^2\Phi_2 e^{i(ku+lv)}$. Here B is the Starobinsky constant.[10]

Metric	Φ_0	Φ_1	Φ_2
B,gS,g^*S,gC	$\phi R_{+1}S_{+1}$	$B^{-1}\phi\mathscr{D}_0 R_{-1}\mathscr{L}_1 S_{+1}$	$\phi R_{-1}S_{-1}$
$pC\bar{B}(+)$	$R_{+1}S_{+1}$	$B^{-1}(\mathscr{D}_0-\phi)R_{-1}\mathscr{L}_1 S_{+1}$	$R_{-1}S_{-1}$
$pC\bar{B}(-)$	$R_{+1}S_{+1}$	$B^{-1}\mathscr{D}_0 R_{-1}(\mathscr{L}_1 - i\phi\sqrt{P})S_{+1}$	$R_{-1}S_{-1}$
pGS	$R_{+1}S_{+1}$	$B^{-1}2iqR_{-1}(\mathscr{L}_1 - i\phi\sqrt{P})S_{+1}$	$R_{-1}S_{-1}$
pCA	$R_{+1}S_{+1}$	$B^{-1}(\mathscr{D}_0 R_{-1}\mathscr{L}_1 S_{+1} - \phi R_{-1}\mathscr{L}_1 S_{+1} - i\phi\sqrt{P}S_{+1}\mathscr{D}_0 R_{-1})$	$R_{-1}S_{-1}$
pDM	$R_{+1}S_{+1}$	$B^{-1}2iqR_{-1}(\mathscr{L}_1 - i\phi\sqrt{P})S_{+1}$	$R_{-1}S_{-1}$
pPD	$(1-xy)R_{+1}S_{+1}$	$B^{-1}\phi[(y+ix)\mathscr{D}_0 R_{-1}\mathscr{L}_1 S_{+1} - R_{-1}\mathscr{L}_1 S_{+1} - i\sqrt{P}S_{+1}\mathscr{D}_0 R_{-1}]$	$(1-xy)R_{-1}S_{-1}$

$$
\begin{aligned}
X_{01'} &= -\tfrac{1}{2}(\phi\bar{\phi})^{-1}[(\delta-2\tau)\varphi_1 \\
&\quad + (\Delta-2\gamma+3\mu-2\bar{\mu})\varphi_0] \\
&= -(\phi\bar{\phi})^{-1}(\Delta-2\gamma+2\mu-\bar{\mu})\varphi_0, \\
X_{10'} &= \tfrac{1}{2}(\phi\bar{\phi})^{-1}[(\bar{\delta}+2\pi)\varphi_1 + (D+2\epsilon-3\rho+2\bar{\rho})\varphi_2] \\
&= (\phi\bar{\phi})^{-1}(D+2\epsilon-2\rho+\bar{\rho})\varphi_2, \\
X_{11'} &= -\tfrac{1}{2}(\phi\bar{\phi})^{-1}[(\Delta+2\mu)\varphi_1 \\
&\quad + (\delta+2\beta-3\tau-2\bar{\pi})\varphi_2] \\
&= -(\phi\bar{\phi})^{-1}(\delta+2\beta-2\tau-\bar{\pi})\varphi_2. \qquad (3)
\end{aligned}
$$

And, using the fact that $\kappa=\sigma=\lambda=\nu=0$, one has

$$
\begin{aligned}
&\nabla_{C'0}X_0{}^{C'} = (D-\epsilon+\bar{\epsilon}-\bar{\rho})X_{01'} - (\delta-\beta-\bar{\alpha}+\bar{\pi})X_{00'}, \\
&2\nabla_{C'(0}X_{1)}{}^{C'} \\
&\quad = (D+\epsilon+\bar{\epsilon}+\rho-\bar{\rho})X_{11'} \\
&\qquad + (\bar{\delta}-\alpha+\bar{\beta}-\pi-\bar{\tau})X_{01'} \\
&\qquad - (\delta+\beta-\bar{\alpha}+\tau+\bar{\pi})X_{10'} \\
&\qquad - (\Delta-\gamma-\bar{\gamma}-\mu+\bar{\mu})X_{00'}, \\
&\nabla_{C'1}X_1{}^{C'} = (\bar{\delta}+\alpha+\bar{\beta}-\bar{\tau})X_{11'} - (\Delta+\gamma-\bar{\gamma}+\bar{\mu})X_{10'}. \qquad (4)
\end{aligned}
$$

Making use of Eqs. (3) and (4), together with the equations given in Ref. 10 and Table I, one can readily verify that Eq. (1) holds.

III. CONCLUDING REMARKS

The validity of Eq. (1) implies that the operator acting on φ_{AB} on the left-hand side of Eq. (1) maps a solution of the source-free Maxwell equations into another solution. The field given by $\varphi_0 = 0 = \varphi_2$, $\varphi_1 = \text{const}$ (ϕ^2) is a solution of the source-free Maxwell equations that does not have the form given in Table I and, as can be easily seen from Eq. (3), satisfies Eq. (1) with $A = 0$. (This solution is precisely the electromagnetic field corresponding to the type-D solutions of the Einstein–Maxwell equations with a non-null electromagnetic field, whose principal directions coincide with those of the conformal curvature.)

In the case of the Schwarzschild metric, where the separated functions $S_{\pm 1}$ are spin-weighted spherical harmonics, the separation constant A has the value $j(j+1)$, with j being an integer greater than 0.

[1] G. F. Torres del Castillo, J. Math. Phys. **29**, 971 (1988).
[2] E. G. Kalnins, W. Miller, Jr., and G. C. Williams, J. Math. Phys. **30**, 2360 (1989).
[3] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 256 (1970).
[4] B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[5] N. Kamran and R. G. McLenaghan, Bull. Cl. Sci. Acad. R. Belg. **70**, 596 (1984).
[6] N. Kamran and R. G. McLenaghan, Lett. Math. Phys. **7**, 381 (1983); J. Math. Phys. **25**, 1019 (1984).
[7] G. F. Torres del Castillo, Proc. R. Soc. London Ser. A **400**, 119 (1985).
[8] B. P. Jeffryes, Proc. R. Soc. London Ser. A **392**, 323 (1984).
[9] G. F. Torres del Castillo, J. Math. Phys. **27**, 1583 (1986).
[10] G. F. Torres del Castillo, J. Math. Phys. **29**, 2078 (1988).
[11] A. L. Dudley and J. D. Finley, III, J. Math. Phys. **20**, 311 (1979).
[12] N. Kamran and R. G. McLenaghan, in *Gravitation and Geometry: A Volume in Honor of I. Robinson*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987).

Special conformal Killing vector space-times and symmetry inheritance

A. A. Coley
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia B3H 4H8, Canada

B. O. J. Tupper
Department of Mathematics and Statistics, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada

(Received 6 June 1988; accepted for publication 21 June 1989)

Viscous heat-conducting fluid and anisotropic fluid space-times admitting a special conformal Killing vector (SCKV) are studied and some general theorems concerning the inheritance of the symmetry associated with the SCKV are proved. In particular, for viscous fluid space-times it is shown that (i) if the SCKV maps fluid flow lines into fluid flow lines, then all physical components of the energy-momentum tensor inherit the SCKV symmetry; or (ii) if the Lie derivative along a SCKV of the shear viscosity term $\eta\sigma_{ab}$ is zero then, again, we have symmetry inheritance. All space-times admitting a SCKV and satisfying the dominant energy condition are found. Apart from the vacuum *pp*-wave solutions, which are the only vacuum solutions that can admit a SCKV, the energy-momentum tensor associated with these space-times is shown to admit at least one null eigenvector and can represent either a viscous fluid with heat conduction or an anisotropic fluid. No perfect fluid space-times can admit a SCKV. These SCKV space-times and, also, space-times admitting a homothetic vector are used to illustrate the symmetry inheritance theorems.

I. INTRODUCTION

Homothetic vectors (HV's) and conformal Killing vectors (CKV's) have been studied at length by various authors. Cahill and Taub[1] and Taub[2] have discussed perfect fluid solutions which are self-similar, i.e., admit a HV. Wainwright and Yaremovich[3] have studied charged perfect fluids and McIntosh[4] has made a general study of HV's in general relativity, with an emphasis on vacuum and perfect fluid space-times. Herrera and co-workers[5] have studied CKV's, with particular reference to perfect fluids and anisotropic fluids; Mason and Tsamparlis[6] have investigated spacelike CKV's; and Maartens *et al.*[7] have made a study of CKV's in anisotropic fluids, in which they are particularly concerned with special conformal Killing vectors (SCKV's).

In this article we are principally interested in imperfect fluids (i.e., viscous, heat-conducting fluids) and, to a lesser extent, anisotropic fluids. The energy-momentum tensor for an imperfect fluid is

$$T_{ab} = \mu u_a u_b + p h_{ab} - 2\eta\sigma_{ab} + q_a u_b + q_b u_a \,, \quad (1.1)$$

where μ is the energy density, p is the isotropic pressure, q^a is the heat flux vector relative to the four-velocity u^a, $\eta(\geqslant 0)$ is the shear viscosity coefficient, $h_{ab} = g_{ab} + u_a u_b$ is the projection tensor, and σ_{ab} is the shear tensor. The energy-momentum tensor for an anisotropic fluid is

$$T_{ab} = \mu u_a u_b + p_{\|} n_a n_b + p_{\perp} p_{ab} \,, \quad (1.2)$$

where n_a is a unit spacelike vector orthogonal to u_a; p_{ab} is the projection tensor onto the two-plane orthogonal to u^a and n^a; and $p_{\|}, p_{\perp}$ denote the pressures parallel to and perpendicular to n^a, respectively.

The effect of a HV on space-times corresponding to the energy-momentum tensor (1.1) has been discussed by Hall and Negm,[8] but only in the case when one of η and q_a is zero. In fact, there has been no systematic study of CKV's, HV's, or even Killing vectors (KV's) in fluids of type (1.1), although fluids of type (1.2) were discussed in Refs. 5 and 7; in the latter reference the SCKV discussion was confined to the case $\mu + p_{\|} \neq 0$.

We shall consider space-times that admit a CKV ξ^a, i.e.,

$$\mathcal{L}_\xi g_{ab} = 2\psi g_{ab} \,, \quad (1.3)$$

where $\mathcal{L}_\xi$ signifies the Lie derivative along ξ^a and $\psi(x^a)$ is the conformal factor. If $\psi_{;ab} = 0$, but $\psi_{,a} \neq 0$, then ξ^a is a SCKV; when ψ is a constant, ξ^a is a HV and $\psi = 0$ corresponds to a KV. Although our ultimate aim is to study the properties of proper CKV's (i.e., CKV's that do not degenerate into SCKV's or HV's in imperfect fluids, in this article we shall confine our attention to the simpler SCKV's and HV's.

In Sec. II, the Lie derivatives of the various kinematical quantities are calculated and the results applied to the energy-momentum tensor (1.1) and, also, to (1.2) in the special case $\mu + p_{\|} = 0$. In Sec. III, we define what we mean by *symmetry inheritance* for a SCKV and prove a number of theorems on the inheritance of the symmetry of a SCKV by the physical components of the energy-momentum tensor (1.1). In particular, we prove the results that if either $\mathcal{L}_\xi u_a = \psi u_a$ or $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$, the symmetry of ξ^a is inherited by all physical quantities. We conclude Sec. III by discussing symmetry inheritance by an anisotropic fluid in the particular case $\mu + p_{\|} = 0$, which was omitted from a similar discussion in Ref. 7.

In Sec. IV we find all space-times (irrespective of the field equations that they satisfy) which admit a proper SCKV, i.e., a SCKV for which ψ is not constant, and which satisfy the dominant energy condition. We find that (i) *there*

 0022-2488/89/112616-10$02.50

are no perfect fluid space-times admitting a proper SCKV; (ii) *anisotropic fluid space-times admitting a proper SCKV must satisfy* $\mu + p_{\|} = 0, p_{\perp} = 0$; and (iii) *there do exist imperfect fluid space-times admitting a proper SCKV.*

Result (i) invalidates that part of Ref. 5 in which it was assumed that perfect fluid space-times admitting SCKV's do exist and result (ii) invalidates a result in Ref. 7 since it shows that SCKV anisotropic fluids are not compatible with the assumption $\mu + p_{\|} \neq 0$ made in Ref. 7. The space-times admitting a proper SCKV form a very restricted class in that they must admit either two null eigenvectors, or a repeated null vector, of the energy-momentum tensor. Because of the limited number of proper SCKV solutions, the work described here on symmetry inheritance has, perhaps, its greatest relevance in the study of HV's, but is couched in the language of CKV's because of our intention to extend the work to proper CKV's. In Sec. V, we illustrate the theorems of Sec. III with examples of both the SCKV and HV and in Sec. VI we make some concluding remarks.

II. KINEMATICAL AND DYNAMICAL RESULTS

In order to discuss the effect on the various kinematical quantities of the Lie derivative along a CKV, we first note the following result proved by Maartens *et al.*,[7] namely, if X^a is any unit vector (timelike or spacelike) and ξ^a is a CKV satisfying (1.3), then

$$\mathcal{L}_\xi X^a = -\psi X^a + Y^a, \tag{2.1}$$

$$\mathcal{L}_\xi X_a = \psi X_a + Y_a, \tag{2.2}$$

where Y^a is some vector orthogonal to X^a, i.e., $X^a Y_a = 0$. Applying this result to the fluid velocity vector u^a, we have

$$\mathcal{L}_\xi u^a = -\psi u^a + v^a, \tag{2.3}$$

$$\mathcal{L}_\xi u_a = \psi u_a + v_a, \tag{2.4}$$

where v^a is a spacelike vector with $u_a v^a = 0$. Note that $u_a \mathcal{L}_\xi u^a = -u^a \mathcal{L}_\xi u_a = \psi$. If $v^a = 0$, i.e., $\mathcal{L}_\xi u^a = -\psi u^a$, then fluid flow lines are mapped into fluid flow lines by the action of ξ^a.

We first consider imperfect fluids with T_{ab} of the form (1.1). The heat flux vector q^a is not a unit vector and if we define Q to be the magnitude of q^a, i.e., $q^a q_a = Q^2$, then by an argument similar to that used in establishing (2.1) it can be shown that

$$\mathcal{L}_\xi q^a = (Q^{-1}\mathcal{L}_\xi Q - \psi)q^a + w^a, \tag{2.5}$$

where $w_a q^a = 0$. Note that if ξ^a is a HV, i.e., ψ is a constant and if we require a self-similar solution (which does not automatically follow for an imperfect fluid), the dimensional requirements[9] imply that $\mathcal{L}_\xi Q = -2\psi Q$ and $w^a = 0$.

Since $u_a q^a = 0$, it follows that

$$-u_a \mathcal{L}_\xi q^a = -u^a \mathcal{L}_\xi q_a = q_a \mathcal{L}_\xi u^a = q^a \mathcal{L}_\xi u_a \equiv \Delta, \tag{2.6}$$

which serves as the definition of the scalar quantity Δ. Equation (2.6) implies that

$$v_a q^a = -u_a w^a = \Delta. \tag{2.7}$$

If $\{_b{}^a{}_c\}$ is the metric affinity of g_{ab}, then[10]

$$\mathcal{L}_\xi \{_b{}^a{}_c\} = \delta^a_b \psi_{,c} + \delta^a_c \psi_{,b} - g_{bc}\psi^a \tag{2.8}$$

and (2.4) and (2.8) give

$$\mathcal{L}_\xi u_{a;b} = \psi u_{a;b} + v_{a;b} + g_{ab}\psi_{,c}u^c - \psi_{,a}u_b, \tag{2.9}$$

$$\mathcal{L}_\xi \Theta = -\psi\Theta + v^a{}_{;a} + 3\psi_{,a}u^a, \tag{2.10}$$

$$\mathcal{L}_\xi h_{ab} = 2\psi h_{ab} + 2u_{(a}v_{b)}, \tag{2.11}$$

where $\Theta = u^a{}_{;a}$ is the expansion scalar for the fluid velocity congruence.

Recalling that the shear tensor is defined by

$$\sigma_{ab} = \tfrac{1}{2}(u_{a;c}h^c{}_b + u_{b;c}h^c{}_a) - \tfrac{1}{3}\Theta h_{ab} \tag{2.12}$$

and using (2.9)–(2.13), after a long calculation we obtain

$$\mathcal{L}_\xi \sigma_{ab} = \psi\sigma_{ab} - \tfrac{1}{3}h_{ab}v^c{}_{;c} - \tfrac{2}{3}\Theta u_{(a}v_{b)} + \dot{v}_{(a}u_{b)} + v_{(a;b)} + \dot{u}_{(a}v_{b)} + u_{(a}u_{b);c}v^c, \tag{2.13}$$

where the overdot indicates the covariant derivative in the direction of the fluid flow, i.e., $\dot{X} = X_{;a}u^a$. Note that $\mathcal{L}_\xi \sigma_{ab}$ is explicitly independent of the derivatives of ψ. Also, $\sigma_{ab}g^{ab} = 0$, $\sigma_{ab}u^b = 0$, and $g^{ab}\mathcal{L}_\xi \sigma_{ab} = 0$, but

$$u^b \mathcal{L}_\xi \sigma_{ab} = -\sigma_{ab}v^b. \tag{2.14}$$

Note, also, that if $\mathcal{L}_\xi u_a = \psi u_a$, i.e., $v_a = 0$, then $\mathcal{L}_\xi \sigma_{ab} = \psi\sigma_{ab}$.

Turning now to dynamical results we note that if ξ^a is a CKV satisfying (1.3), then[7]

$$\mathcal{L}_\xi R_{ab} = -2\psi_{;ab} - g_{ab}\Box\psi, \tag{2.15}$$

$$\mathcal{L}_\xi R = -2\psi R - 6\Box\psi, \tag{2.16}$$

$$\mathcal{L}_\xi G_{ab} = 2g_{ab}\Box\psi - 2\psi_{;ab}, \tag{2.17}$$

where $\Box\psi \equiv g^{ab}\psi_{;ab}$. We consider Einstein's field equations in the form

$$G_{ab} + \Lambda g_{ab} = T_{ab} \tag{2.18}$$

and so find

$$\mathcal{L}_\xi T_{ab} = 2(\Box\psi + \Lambda\psi)g_{ab} - 2\psi_{;ab}. \tag{2.19}$$

We take T_{ab} to be of the form (1.1) and, by taking the Lie derivative of (1.1) with respect to ξ^a, we obtain

$$\begin{aligned}\mathcal{L}_\xi \mu(u_a u_b) + \mathcal{L}_\xi p(h_{ab}) + 2\psi(\mu u_a u_b + p h_{ab}) \\ + 2(\mu + p)v_{(a}u_{b)} - 2\sigma_{ab}\mathcal{L}_\xi \eta - 2\eta\mathcal{L}_\xi \sigma_{ab} \\ + 2(Q^{-1}\mathcal{L}_\xi Q + 2\psi)u_{(a}q_{b)} + 2q_{(a}v_{b)} \\ + 2u_{(a}w_{b)} = 2(\Box\psi + \Lambda\psi)g_{ab} - 2\psi_{;ab},\end{aligned} \tag{2.20}$$

where $\mathcal{L}_\xi \sigma_{ab}$ is given by (2.13) and we have used (2.4) and (2.5).

For the remainder of this article we shall confine our attention to HV's and SCKV's, i.e., we assume that $\psi_{;ab} = 0$. We shall also assume that $\Lambda = 0$; this is not a particularly restrictive assumption since the replacements $\mu \to \mu + \Lambda$, $p \to p - \Lambda$ will reproduce the effects of including Λ. Thus (2.19) becomes $\mathcal{L}_\xi T_{ab} = 0$ and we focus our attention on (2.20) with zero rhs.

Contracting (2.20) in turn with $u^a u^b$, h^{ab}, $u^a h^b{}_c$, $h^{ac}h^{bd} - \tfrac{1}{3}h^{ab}h^{cd}$, q^b, $q^a u^b$, and $q^a q^b$, and simplifying we obtain

$$\mathcal{L}_\xi \mu + 2\psi\mu + 2\Delta = 0\,, \tag{2.21}$$

$$\mathcal{L}_\xi p + 2\psi p + \tfrac{2}{3}\Delta = 0\,, \tag{2.22}$$

$$2\eta\sigma_{ab}v^b = w_a - \Delta u_a + (\mu + p)v_a + (Q^{-1}\mathcal{L}_\xi Q + 2\psi)q_a\,, \tag{2.23}$$

$$\mathcal{L}_\xi(\eta\sigma_{ab}) = 2\eta\sigma_{c(a}u_{b)}v^c + q_{(a}v_{b)} - \tfrac{1}{3}\Delta h_{ab}\,, \tag{2.24}$$

$$2q^b\mathcal{L}_\xi(\eta\sigma_{ab}) = \tfrac{1}{3}\Delta q_a + Q^2 v_a + [(\mu + p)\Delta + Q\mathcal{L}_\xi Q + 2\psi Q^2]u_a\,, \tag{2.25}$$

$$2\eta\sigma_{ab}v^b q^a = Q(\mathcal{L}_\xi Q + 2\psi Q) + (\mu + p)\Delta\,, \tag{2.26}$$

$$2q^a q^b \mathcal{L}_\xi(\eta\sigma_{ab}) = \tfrac{4}{3}\Delta Q^2\,. \tag{2.27}$$

The case in which T_{ab} is given by (1.2) has been discussed in Ref. 7. However, Ref. 7 assumed that $\mu + p_\parallel \neq 0$. As we shall see, space-times admitting a SCKV and satisfying the field equations for an anisotropic fluid must have $\mu = -p_\parallel = \frac{1}{2}R$ and $p_\perp = 0$, so that the energy-momentum tensor is limited to the form

$$T_{ab} = \tfrac{1}{2}R(u_a u_b - n_a n_b)\,. \tag{2.28}$$

For a SCKV we have $\mathcal{L}_\xi T_{ab} = 0$, $\mathcal{L}_\xi R = -2\psi R$ and since n_a is a unit vector,

$$\mathcal{L}_\xi n_a = \psi n_a + m_a\,, \tag{2.29}$$

where $m_a n^a = 0$. From Eqs. (2.29) and (2.4), the Lie derivative of (2.28) yields

$$0 = \tfrac{1}{2}R(v_a u_b + v_b u_a - m_a n_b - m_b n_a)\,.$$

Since R cannot be zero for a nonvacuum solution, this implies that

$$v_a = \Sigma n_a, \quad m_a = \Sigma u_a\,, \tag{2.30}$$

where $\Sigma = -u^a m_a = n^a v_a$. It follows that either $v_a = m_a = 0$ or v_a, m_a are parallel to u_a, n_a, respectively. Thus the result that $v_a = m_a = 0$, given in Ref. 7 and based on the assumption $\mu + p_\parallel \neq 0$, is not necessarily true.

III. SYMMETRY INHERITANCE

If a perfect fluid space-time is self-similar, i.e., admits a HV ξ^a, the density, pressure, and fluid velocity must satisfy

$$\mathcal{L}_\xi \mu = -2\psi\mu, \quad \mathcal{L}_\xi p = -2\psi p, \quad \mathcal{L}_\xi u_a = \psi u_a\,,$$

and we say that these quantities inherit the space-time symmetry defined by ξ^a. In contrast, if a space-time admitting a HV satisfies Einstein's field equations with T_{ab} given by (1.1), then in general, the symmetry is not inherited by the dynamical and kinematical quantities appearing in T_{ab}. However, if we impose self-similarity on the complete solution, dimensional considerations[9] will imply that the following set of equations will hold:

$$\mathcal{L}_\xi \mu = -2\psi\mu, \quad \mathcal{L}_\xi p = -2\psi p, \quad \mathcal{L}_\xi u_a = \psi u_a\,,$$
$$\mathcal{L}_\xi q_a = -\psi q_a, \quad \mathcal{L}_\xi \sigma_{ab} = \psi\sigma_{ab}, \quad \mathcal{L}_\xi \eta = -\psi\eta\,. \tag{3.1}$$

For a SCKV, there is no self-similarity unless the SCKV is in fact a HV. However, we will now make the following definition.

Definition: If the space-time solution of Einstein's field equations with T_{ab} given by (1.1) admits a SCKV ξ^a given by (1.3), then the solution will be said to inherit the symmetry corresponding to ξ^a if the set of equations (3.1) holds.

In this section we investigate the conditions under which an imperfect fluid given by (1.1) will inherit the symmetries corresponding to a SCKV ξ^a. We also comment on the symmetry inheritance for an anisotropic fluid (1.2), thus extending the work of Maartens *et al.*[7] to a crucial case which they omitted. Throughout this investigation we require that the fluid satisfies the dominant energy condition.

We consider a number of possible cases.

Case 1: In Ref. 7 it is shown that when $q^a = 0$, Eqs. (3.1) will hold provided that $\mu + p \neq 0$ and either $\mathcal{L}_\xi u_a = \psi u_a$ (i.e., $v_a = 0$) or $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$. We now complete this result by considering the exceptional case $\mu + p = 0$. Equation (2.23) becomes $2\eta\sigma_{ab}v^b = 0$, so that σ_{ab} must be of the form $\sigma_{ab} = \sigma(x_a x_b - y_a y_b)$, where x_a, y_a are orthogonal unit spacelike vectors which are also orthogonal to u_a and v_a. By applying the dominant energy condition to the resulting T_{ab}, we find that $\eta\sigma = 0$, so that the fluid degenerates into a perfect fluid with $\mu + p = 0$. Hence, Eqs. (3.1) hold for an imperfect fluid and we have the following theorem.

Theorem 1: If $q^a = 0$ and if either $\mathcal{L}_\xi u_a = \psi u_a$ or $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$, the symmetries of a SCKV ξ^a are inherited.

Case 2: Suppose that $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$. Contracting Eq. (2.25) with q^a yields $\frac{4}{3}\Delta Q^2 = 0$, so that either $\Delta = 0$ or $Q = 0$. The latter case immediately leads to inheritance from Theorem 1, so we consider $\Delta = 0$. Equation (2.25) then becomes

$$Q^2 v_a + Q(\mathcal{L}_\xi Q + 2\psi Q)u_a = 0\,,$$

which implies that $v_a = 0$ and $\mathcal{L}_\xi Q + 2\psi Q = 0$. Equation (2.23) then implies that $w_a = 0$, and so we have the following theorem.

Theorem 2: If $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$, the symmetries of a SCKV ξ^a are inherited.

Corollary 1: If $\eta\sigma_{ab} = 0$, the symmetries of a SCKV ξ^a are inherited.

Case 3: Suppose that $\mathcal{L}_\xi u_a = \psi u_a$, i.e., fluid flow lines are mapped onto fluid flow lines. In this case $v_a = 0$, $\Delta = 0$, and Eq. (2.24) becomes $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$, so that Theorem 2 leads to the following theorem.

Theorem 3: If $\mathcal{L}_\xi u_a = \psi u_a$, the symmetries of a SCKV ξ_a are inherited.

Corollary 2: If a SCKV ξ^a is parallel to u^a, then the symmetries of ξ^a are inherited.

Theorems 2 and 3 are the primary results in this paper; they are new results which generalize the work of Ref. 7 to viscous fluids with nonzero heat conduction. Theorems 2 and 3 are definitive in that they give a complete characterization of the SCKV inheritance problem for the fluid (1.1). Note that from Theorems 1 and 2, the vanishing of the shear viscosity is sufficient to ensure inheritance, whereas the vanishing of the heat conduction is not sufficient. However, as we shall see in Case 4, there are conditions on q^a which will ensure inheritance.

Case 4: Suppose that q^a is an eigenvector of the shear tensor, i.e.,

$$2\eta\sigma_{ab}q^b = \lambda q_a \,. \tag{3.2}$$

Equation (3.2) implies that, geometrically, q^a and u^a span a timelike invariant two-space of T_{ab}[8] and, physically, there exist no shear velocities between neighborhood surface elements orthogonal to the direction of the heat flux.[11] Taking the Lie derivative of (3.2) and using Eqs. (2.21)–(2.24) we obtain

$$[Q(\mathcal{L}_\xi Q + 2\psi Q) - (\mu + p)\Delta]u_a - \Delta q_a + Q^2 v_a - \lambda w_a + 2\eta\sigma_{ab}w^b = 0 \,, \tag{3.3}$$

$$\mathcal{L}_\xi \lambda + 2\lambda\psi = \tfrac{4}{3}\Delta \,. \tag{3.4}$$

Relations (3.3) and (3.4) do not imply inheritance. However, if we make the additional assumption

$$\mathcal{L}_\xi q_a = -\psi q_a \,, \tag{3.5}$$

i.e., $w_a = 0$, $\Delta = 0$, and $\mathcal{L}_\xi Q = -2\psi Q$, then (3.3) shows that $v_a = 0$, so that from Theorem 3, we have complete inheritance.

Conversely, if we first assume (3.5), the Lie derivative of (2.23) contracted with $q^a v^b$ leads to

$$v_b v^b Q^2 + 2\eta\sigma_{ab}q^a \mathcal{L}_\xi v^b = (\mu + p)q^a \mathcal{L}_\xi v_a \tag{3.6}$$

and since (3.5) implies that $q^a \mathcal{L}_\xi v_a = 0$, this shows that the imposition of (3.2) leads to inheritance. On the other hand, when $v_a \neq 0$, i.e., if we have noninheritance, then (3.6) shows that q^a cannot be an eigenvector of σ_{ab}. Thus we have proved the following theorem.

Theorem 4: If q_a is an eigenvector of σ_{ab} and, also, $\mathcal{L}_\xi q_a = -\psi q_a$, then the symmetries of a SCKV ξ^a are inherited. If either of these conditions does not hold, then the symmetries are not inherited. Furthermore, if the symmetries are not inherited q^a cannot be an eigenvector of σ_{ab}.

Theorem 4 is a new result since it requires $q^a \neq 0$, a situation that has not been investigated for SCKV's.

It should be emphasized that the results given here apply not only to SCKV's, but also to HV's. Indeed, when ψ is a constant, Theorems 2 and 3 give the definitive conditions under which the physical quantities constituting the energy-momentum tensor (1.1) can inherit the self-similar symmetry associated with a HV.

These results apply also to KV's ($\psi = 0$) and are again new since no investigation has been made of KV's in a viscous, heat-conducting fluid. The major results for KV's may be summarized in the following theorem.

Theorem 5: If ξ^a is a KV of space-time satisfying Einstein's field equations for an imperfect fluid with the energy-momentum tensor (1.1), then the necessary and sufficient condition for the symmetry defined by ξ^a to be inherited by the physical quantities associated with the fluid, i.e., for the quantities

$$\mathcal{L}_\xi \mu = \mathcal{L}_\xi p = \mathcal{L}_\xi u_a = \mathcal{L}_\xi \eta = \mathcal{L}_\xi \sigma_{ab} = 0$$

to hold, is that $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$ or, equivalently, $\mathcal{L}_\xi u_a = 0$.

We now turn to the case of an anisotropic fluid given by (1.2) and, in particular, to the special case $\mu + p_\parallel = 0$, $p_\perp = 0$, for which T_{ab} is given by (2.28). Since $\mu = -p_\parallel = \frac{1}{2}R$, it follows that $\mathcal{L}_\xi \mu = -2\psi\mu$ and $\mathcal{L}_\xi p_\parallel = -2\psi p_\parallel$, so that μ and $p_\parallel$ always inherit the symmetry of the SCKV ξ^a. However, as shown in Sec. II, v_a and m_a [as defined by (2.29)] are either zero or nonzero and parallel to n_a, u_a, respectively, so that the symmetry of the SCKV may or may not be inherited. Maartens *et al.*[7] showed that in general, v_a is given by

$$v_a = 2\omega_{ab}\xi^b + \alpha\dot{u}_a - \alpha_{,b}h^b{}_a \,, \tag{3.7}$$

where $\alpha = -\xi_a u^a$ and ω_{ab} is the vorticity tensor. Complete inheritance occurs when $v_a = 0$; two special cases of this are when ξ^a is parallel to u^a and, also, when ξ^a is orthogonal to u^a and the fluid is vorticity-free, i.e., $\xi^a u_a = 0$ and $\omega_{ab} = 0$. Hence we have the following theorem.

Theorem 6: For an anisotropic fluid of the form (1.2) with $\mu + p_\parallel = 0$, $p_\perp = 0$, the symmetries of a SCKV ξ^a are inherited if and only if expression (3.7) for v_a is zero.

IV. SPACE-TIMES ADMITTING A PROPER SCKV

We now turn to the problem of determining those space-times that admit a proper SCKV. A SCKV is defined by (1.3) with $\psi_{;ab} = 0$. This implies that $\psi_{,a}$ is a covariantly constant, hypersurface orthogonal, geodesic vector, resulting in considerable simplification of the space-time metric.[12] It also implies that $\psi_{,a}$ is globally timelike, globally spacelike, or globally null. We consider these three possibilities in turn.

A. The vector $\psi_{,a}$ timelike

We can choose coordinates in which

$$\psi_{,a} = (-1,0,0,0) \tag{4.1}$$

and the metric is of the form

$$ds^2 = -dt^2 + g_{\alpha\beta}(x^\gamma)dx^\alpha dx^\beta \equiv -dt^2 + d\Omega^2 \,, \tag{4.2}$$

where $\alpha, \beta, \gamma = 1, 2, 3$.

Equation (4.1) implies that $\psi = -t$ and the metric (4.2) implies that $\{_j{}^i{}_k\} = 0$ if any of $i, j, k = 0$. Equations (1.3) take the form

$$\xi_{0,0} = t \,, \tag{4.3}$$

$$\xi_{0,\alpha} + \xi_{\alpha,0} = 0 \,, \tag{4.4}$$

$$\xi_{\alpha;\beta} + \xi_{\beta;\alpha} = -2tg_{\alpha\beta} \,. \tag{4.5}$$

Integrating (4.3) yields

$$\xi_0 = \tfrac{1}{2}t^2 + A(x^\gamma) \,, \tag{4.6}$$

where A is a scalar function and (4.4) and (4.6) yield

$$\xi_{\alpha,0} = -A_{,\alpha} \,,$$

i.e.,

$$\xi_\alpha = -A_{,\alpha}t + B_\alpha(x^\gamma) \,, \tag{4.7}$$

where B_α is a vector function. Substituting (4.7) into (4.5) yields

$$-2A_{;\alpha\beta}t + B_{\alpha;\beta} + B_{\beta;\alpha} = -2tg_{\alpha\beta}$$

and equating coefficients of t we obtain

$$B_{\alpha;\beta} + B_{\beta;\alpha} = 0 \,, \tag{4.8}$$

so that B_α is a KV of the three-dimensional space and

$$A_{;\alpha\beta} = g_{\alpha\beta} \,. \tag{4.9}$$

Petrov[13] quotes a result due to Sinyukov,[14] namely that

if a V_n admits a vector field ϕ_α satisfying $\phi_{\alpha;\beta} = \rho g_{\alpha\beta}$, where ρ is a nonzero scalar function, a system of coordinates exists in which the metric takes the form

$$ds_n^2 = g_{11}\,(dx^1)^2 + (1/g_{11})\Gamma_{pq}(x^2,\ldots,x^n)dx^p\,dx^q\,, \quad (4.10)$$

where $p, q \neq 1$; $g_{11} = [2\int\rho(x^1)dx^1 + C]^{-1}$; and ρ is now an arbitrary function of x^1 only. Applying the result (4.10) to Eq. (4.9), in which $\phi_\alpha = A_{,\alpha}$ and $\rho = 1$, we find that for the three-dimensional metric $d\Omega^2$, $g_{11} = (2x^1 + C)^{-1}$ and the transformation $\sqrt{2x^1 + C} \to x$ yields

$$d\Omega^2 = dx^2 + x^2\Gamma_{AB}(x^C)dx^A\,dx^B\,,$$

where A, B, C take the values 2, 3. The two-dimensional metric $\Gamma_{AB}\,dx^A dx^B$ can be transformed into $dy^2 + f^2(y,z)dz^2$, so that the space-time metric takes the final form

$$ds^2 = -dt^2 + dx^2 + x^2[dy^2 + f^2(y,z)dz^2]\,. \quad (4.11)$$

After excluding linear combinations with the KV admitted by this metric, we find that only one SCKV exists, namely

$$\xi^a = (-\tfrac{1}{2}t^2 - \tfrac{1}{2}x^2, -tx, 0, 0)\,. \quad (4.12)$$

which is timelike.

B. The vector $\psi_{,a}$ spacelike

We can choose coordinates in which

$$\psi_{,a} = (0,1,0,0)\,, \quad (4.13)$$

so that $\psi = x$ and the metric will be of the form

$$ds^2 = dx^2 + g_{\alpha\beta}(x^\gamma)dx^\alpha\,dx^\beta\,,$$

where, in this case, $\alpha, \beta, \gamma = 0, 2, 3$. Following precisely the same argument as in the timelike case, we obtain two possible solutions, namely

$$ds^2 = dx^2 - dt^2 + t^2[dy^2 + g^2(y,z)dz^2] \quad (4.14)$$

and

$$ds^2 = dx^2 + dy^2 + y^2[-dt^2 + h^2(t,z)dz^2]\,. \quad (4.15)$$

However, the metric (4.15) does not satisfy the dominant energy condition and so will be discarded.

The metric (4.14) admits only one proper SCKV, namely

$$\xi^a = (xt, \tfrac{1}{2}x^2 + \tfrac{1}{2}t^2, 0, 0)\,; \quad (4.16)$$

this SCKV is spacelike.

C. The vector $\psi_{,a}$ null

Since $\psi_{,a}$ is a gradient vector and a null KV, it follows that we have a generalized *pp*-wave space-time[15] with a metric of the form

$$ds^2 = P^{-2}(dx^2 + dy^2) - 2\,du(dv - m\,dx + H\,du)\,, \quad (4.17)$$

where H, P, and m are arbitrary functions of u, x, and y only. We label the coordinates $(u,v,x,y) \equiv (x^0, x^1, x^2, x^3)$ and then the null KV $k^a = \psi^{,a}$ is given by $k^a = (0,1,0,0)$, i.e., $k_a = (-1,0,0,0)$, so that

$$\psi = -u\,. \quad (4.18)$$

When $R = 0$, it can be shown that the general metric admitting a covariantly constant null gradient vector is (4.17) with $P = 1$ and $m = 0$,[15] i.e.,

$$ds^2 = dx^2 + dy^2 - 2\,du\,dv - 2H\,du^2\,. \quad (4.19)$$

However, in general, the imperfect and anisotropic fluids considered in this article will not have zero Ricci scalar, so we will use the metric (4.17). We require those metrics of this form which admit a SCKV.

The nonzero components of the Ricci tensor for the metric (4.17) are

$$\begin{aligned}
R_{00} &= P^2(H_{xx} + H_{yy} + m_{ux} + \tfrac{1}{2}m_y^2P^2) \\
&\quad + 2P^{-2}(PP_{uu} - 2P_u^2)\,, \\
R_{02} &= -\tfrac{1}{2}m_{yy}P^2 - m_yPP_y + P^{-2}(PP_{ux} - P_uP_x)\,, \\
R_{03} &= \tfrac{1}{2}m_{xy}P^2 + m_yPP_x + P^{-2}(PP_{uy} - P_uP_y)\,, \\
R_{22} &= R_{33} = P^{-2}(PP_{xx} + PP_{yy} \\
&\quad - P_x^2 - P_y^2) = \tfrac{1}{2}P^{-2}R\,.
\end{aligned} \quad (4.20)$$

Recalling (4.18), the SCKV equations are

$$\begin{aligned}
\xi_{0,0} &= (H_u + mm_uP^2 + mH_xP^2)\xi_1 \\
&\quad + (m_u + H_x)P^2\xi_2 + H_yP^2\xi_3 + 2Hu\,, \\
\xi_{0,1} + \xi_{1,0} &= 2u\,, \\
\xi_{0,2} + \xi_{2,0} &= 2(H_x - mP^{-1}P_u)\xi_1 \\
&\quad - 2P^{-1}P_u\xi_2 - m_yP^2\xi_3 - 2mu\,, \\
\xi_{0,3} + \xi_{3,0} &= 2(H_y + \tfrac{1}{2}m_ymP^2)\xi_1 \\
&\quad + m_yP^2\xi_2 - 2P^{-1}P_u\xi_3\,, \\
\xi_{1,1} &= \xi_{1,2} + \xi_{2,1} = \xi_{1,3} + \xi_{3,1} = 0\,, \\
\xi_{2,2} &= -(P^{-3}P_u + mP^{-1}P_x + m_x)\xi_1 \\
&\quad - P^{-1}P_x\xi_2 + P^{-1}P_y\xi_3 - uP^{-2}\,, \\
\xi_{2,3} + \xi_{3,2} &= -(m_y + 2mP^{-1}P_y)\xi_1 \\
&\quad - 2P^{-1}P_y\xi_2 - 2P^{-1}P_x\xi_3\,, \\
\xi_{3,3} &= -(P^{-3}P_u - mP^{-1}P_y)\xi_1 \\
&\quad + P^{-1}P_x\xi_2 - P^{-1}P_y\xi_3 - uP^{-2}\,.
\end{aligned} \quad (4.21)$$

Solving equations (4.21), we find that the most general form of the SCKV when $R \neq 0$ is

$$\begin{aligned}
\xi^a = [&-(u^2 + \alpha u + \beta), \alpha v - D(u,x,y) + (2H + m^2P^2) \\
&\times(u^2 + \alpha u + \beta) + mP^2B(u,x,y)\,, \\
&mP^2(u^2 + \alpha u + \beta) + P^2B(u,x,y), P^2C(u,x,y)]\,,
\end{aligned} \quad (4.22)$$

where α and β are arbitrary constants and B, C, and D are three functions satisfying the differential equations

$$\begin{aligned}
D_u &= (H_u + mm_uP^2 + mH_xP^2)(u^2 + \alpha u + \beta) \\
&\quad + (m_u + H_x)P^2B + H_yP^2C + 2Hu\,, \\
B_x &= -(P^{-3}P_u + mP^{-1}P_x + m_x)(u^2 + \alpha u + \beta) \\
&\quad - P^{-1}P_xB + P^{-1}P_yC - uP^{-2}\,, \\
C_y &= -(P^{-3}P_u - mP^{-1}P_x)(u^2 + \alpha u + \beta)
\end{aligned}$$

$$+P^{-1}P_xB-P^{-1}P_yC-uP^{-2}, \tag{4.23}$$

$$D_x+B_u=2(H_x-mP^{-1}P_u)(u^2+\alpha u+\beta)$$
$$-2P^{-1}P_uB-m_yP^2C-2mu,$$
$$D_y+C_u=2(H_y+\tfrac{1}{2}m_ymP^2)(u^2+\alpha u+\beta)$$
$$+m_yP^2B-2P^{-1}P_uC,$$
$$B_y+C_x=-(m_y+2mP^{-1}P_y)(u^2+\alpha u+\beta)$$
$$-2P^{-1}P_yB-2P^{-1}P_xC.$$

By eliminating α,β,B,C, and D from Eqs. (4.23), an expression connecting H, P, m, and their derivatives will be obtained which delineates those members of the general set of space-times with the metric (4.17) which admit a SCKV.

In the special case of the metric (4.19), i.e., when $R=0$, the SCKV is of the form

$$\xi^a=[-(u^2+\alpha u+\beta),\alpha v-\tfrac{1}{2}x^2-\tfrac{1}{2}y^2$$
$$+J_ux+K_uy+L(u),-ux+\gamma y$$
$$+J(u),-uy-\gamma x+K(u)], \tag{4.24}$$

where α,β, and γ are arbitrary constants and J,K, and L are arbitrary functions of u only. In order to admit a SCKV, the function H in the metric must satisfy

$$H_u(u^2+\alpha u+\beta)+H_x(ux-\gamma y-J)$$
$$+H_y(uy+\gamma x-K)$$
$$+2H(u+\alpha)-J_{uu}x-K_{uu}y+L_u=0. \tag{4.25}$$

Thus far in this section, we have found all space-times admitting a SCKV irrespective of the field equations that they satisfy. These are the space-times with metrics given by (4.11), (4.14), (4.15), and those metrics (4.17) that satisfy Eqs. (4.23). Of these, (4.15) and some members of the set (4.17) do not satisfy the dominant energy condition; we shall confine our attention only to those space-times that do satisfy this condition.

The integrability conditions for the existence of a covariantly constant vector $\psi_{,a}$ are

$$R^a{}_{bcd}\psi_{,a}=0; \tag{4.26}$$

by contraction, this implies

$$T^a{}_b\psi_{,a}=-\tfrac{1}{2}R\psi_{,b}, \tag{4.27}$$

so that $\psi_{,a}$ is an eigenvector of the energy-momentum tensor. In the cases of the metrics (4.11) and (4.14), by calculating the Einstein tensor, each of these solutions possesses a timelike and a spacelike eigenvector in the tx plane which have the same eigenvalues, so that there are two independent null eigenvectors in the tx plane, namely

$$k_a=(1/\sqrt{2})(-1,1,0,0),\quad l_a=(1/\sqrt{2})(1,1,0,0), \tag{4.28}$$

where we have normalized the null vectors to satisfy $k_al^a=1$. Furthermore, there exist two spacelike eigenvectors in the yz plane, each of which has a zero eigenvalue. Hence, it follows that for the two metrics (4.11) and (4.16), T_{ab} is of Segré type $\{(1,1)(11)\}$[16] and can be written in the form

$$T_{ab}=-\tfrac{1}{2}R(k_al_b+k_bl_a). \tag{4.29}$$

In the case of the space-times with the metric (4.17), Eq. (4.27) shows that $\psi_{,a}$ is a null eigenvector of the energy-momentum tensor. Since we also require T_{ab} to satisfy the dominant energy condition, this implies[16] that T_{ab} must be either of Segré type $\{(1,1)11\}$ or $\{2,11\}$ and so can be written, respectively, in the forms

$$T_{ab}=-A(k_al_b+k_bl_a)+Cx_ax_b+Dy_ay_b \tag{4.30}$$

or

$$T_{ab}=-A(k_al_b+k_bl_a)+Bk_ak_b$$
$$+Cx_ax_b+Dy_ay_b, \tag{4.31}$$

where k_a, l_a are null vectors with $k_al^a=1$ and x_a, y_a are mutually orthogonal unit spacelike vectors which are also orthogonal to k_a and l_a. The quantities A, B, C, and D are scalar functions of the coordinates. The eigenvectors are k_a $(\equiv\psi_{,a})$, x_a, and y_a in the second case and, additionally, l_a in the first case.

From (4.20), the nonzero components of the Einstein tensor are G_{00}, G_{01} $(=\frac{1}{2}R)$, G_{02}, and G_{03}. Equating G_{ab} with T_{ab} given by the more general expression (4.31) and using the fact that $k^a=(0,1,0,0)$, so that $l_1=1, x_1=y_1=0$, we find that

$$A=\tfrac{1}{2}R,\quad C=D=0,$$

so that T_{ab} is given by

$$T_{ab}=-\tfrac{1}{2}R(k_al_b+k_bl_a)+Bk_ak_b, \tag{4.32}$$

which includes the form (4.29) when $B=0$. Thus T_{ab} is either of Segré type $\{(1,1)(11)\}$ or $\{2(11)\}$ and we have proved the following theorems.

Theorem 7: A space-time that admits a SCKV and satisfies the dominant energy condition has an energy-momentum tensor which admits two independent spacelike eigenvectors with zero eigenvalues and, also, admits either two independent null eigenvectors with the same eigenvalue or a repeated null eigenvector, i.e., the Segré type of the energy-momentum tensor is either $\{(1,1)(11)\}$ or $\{2(11)\}$. The energy-momentum tensor has the form (4.32), where $B=0$ or $B\neq0$ according to whether T_{ab} is $\{(1,1)(11)\}$ or $\{2(11)\}$, respectively.

Theorem 8: There exist no perfect fluid space-times which admit a SCKV.

Having established the Segré type of the energy-momentum tensor, we shall investigate the field equations that are satisfied by the SCKV space-times.

For the metric (4.11), the only nonzero components of the Einstein tensor are

$$G^0_0=G^1_1=x^{-2}(1+f^{-1}f_{yy}) \tag{4.33}$$

and, assuming a comoving velocity $u^a=(1,0,0,0)$, the metric satisfies the field equations for an anisotropic fluid, with T_{ab} given by (1.2). We find that

$$\mu=-p_{\|}=-x^{-2}(1+f^{-1}f_{yy}),\quad p_\perp=0, \tag{4.34}$$

and we must have $1+f^{-1}f_{yy}<0$ for the dominant energy

condition to be satisfied. Alternatively, if we assume a noncomoving velocity of the form

$$u^a = (\cosh\phi, \sinh\phi, 0, 0)\,, \tag{4.35}$$

where $\phi = \phi(t,x)$, then the metric satisfies the viscous fluid field equations, with T_{ab} given by (1.1), with

$$\mu = -3p = 2\eta X = -x^{-2}(1 + f^{-1}f_{yy}), \quad Q = 0\,, \tag{4.36}$$

where

$$X = (\phi_t - x^{-1})\sinh\phi + \phi_x \cosh\phi\,. \tag{4.37}$$

Note that for $\mu > 0$, $\eta \geqslant 0$, we must have $1 + f^{-1}f_{yy} < 0$ and $X \geqslant 0$.

The space-time with the metric (4.14) has similar properties; the nonzero components of the Einstein tensor are

$$G^0_0 = G^1_1 = -t^{-2}(1 - f^{-1}f_{yy}) \tag{4.38}$$

and the field equations for a comoving anisotropic fluid are satisfied, with

$$\mu = -p_{\|} = t^{-2}(1 - f^{-1}f_{yy}), \quad p_\perp = 0\,, \tag{4.39}$$

so that $1 - f^{-1}f_{yy} > 0$ for $\mu > 0$. The space-time (4.14) also satisfies the viscous fluid field equations with u^a of the form (4.35) and

$$\mu = -3p = 2\eta X = t^{-2}(1 - f^{-1}f_{yy}), \quad Q = 0\,, \tag{4.40}$$

where

$$X = \phi_t \sinh\phi + (\phi_x - t^{-1})\cosh\phi \tag{4.41}$$

and we must have $1 - f^{-1}f_{yy} < 0$ and $X \geqslant 0$.

Thus each of the space-times (4.11) and (4.14) may represent an infinite set of viscous fluid solutions depending on the choice of the "tilt function" $\phi(t,x)$ and, in the particular case when $\phi = 0$, i.e., when u^a is comoving, the viscous fluid solution degenerates into the anisotropic solution given by (4.34) or (4.39).

We now investigate those members of the set of space-times with the metric (4.17), if any, which satisfy the field equations for an anisotropic fluid. Since the rhs of (1.2) is obviously diagonalizable, those solutions for which T_{ab} is of the type $\{2(11)\}$ cannot represent an anisotropic fluid since a T_{ab} of this Segré type is not diagonalizable. Hence, the only possibility for a SCKV space-time to satisfy the field equations with T_{ab} given by (1.2) is for the energy-momentum tensor to be of the form (4.29), i.e., we must have

$$\mu u_a u_b + p_{\|} n_a n_b + p_\perp (x_a x_b y_a y_b) = -\tfrac{1}{2}R(k_a l_b + k_b l_a)\,, \tag{4.42}$$

where x_a, y_a are two mutually orthogonal spacelike unit vectors in the two-space orthogonal to that of u_a and n_a.

Contracting (4.42) with $l^b u^a$ we obtain

$$-\mu u_b l^b = -\tfrac{1}{2}R u^a l_a$$

and since u_a is timelike and l_a is null, $u^a l_a \neq 0$, so that

$$\mu = \tfrac{1}{2}R\,. \tag{4.43}$$

Contracting in turn with $l^b n^a$ and $k^b n^a$ we obtain

$$p_{\|} l^b n_b = -\tfrac{1}{2}R l_a n^a, \quad p_{\|} k^b n_b = -\tfrac{1}{2}R k_a n^a\,,$$

so that either $p_{\|} = -\frac{1}{2}R$ or $l^b n_b = k^b n_b = 0$. However, if we contract (4.42) with $n^a n^b$ we obtain

$$p_{\|} = -R k_a n^a l_b n^b\,,$$

so that $l^b n_b = k^b n_b = 0$ implies that $p_{\|} = 0$. Hence, we have two possibilities, namely

$$\text{(i)}\ p_{\|} = -\tfrac{1}{2}R, \quad l^b n_b \neq 0, \quad k^b n_b \neq 0$$

or

$$\text{(ii)}\ p_{\|} = 0, \quad l^b n_b = k^b n_b = 0\,.$$

These are two distinct possibilities since we discard the case when $R = 0$, which implies a vacuum solution.

Noting that the contraction of (4.42) with g^{ab} yields

$$\mu - p_{\|} - 2p_\perp = R \tag{4.44}$$

and taking into account (4.43), possibility (i) leads to $p_\perp = 0$, while (ii) implies that $p_\perp = -\frac{1}{4}R$. Contracting (4.42) with $k^b x^a$ we obtain $p_\perp k^b x_b = -\frac{1}{2}R k^a x_a$ and since $p_\perp = -\frac{1}{2}R$ is not a possibility, it follows that $k^a x_a = 0$. Similarly, we can show that $k^a y_a = l^a x_a = l^a y_a = 0$, so that k^a, l^a lie entirely in the two-plane of u^a and n^a. Since k^a and l^a are null vectors this implies that $k^a n_a$ and $l^a n_a$ cannot be zero, so that (i) is the only possibility. Hence, we have proved the following theorem.

Theorem 9: If a space-time satisfies the field equations for an anisotropic fluid, with T_{ab} given by (1.2), and also admits a SCKV, then, necessarily,

$$\mu = -p_{\|} = \tfrac{1}{2}R, \quad p_\perp = 0\,, \tag{4.45}$$

i.e., the energy-momentum tensor is of the form

$$T_{ab} = \tfrac{1}{2}R(u_a u_b - n_a n_b)\,. \tag{4.46}$$

As we have seen, the assumption $\mu + p_{\|} \neq 0$ cannot hold for an anisotropic fluid admitting a SCKV and thus Theorem 9 invalidates some of the results given in Ref. 7, in which this specific assumption was made.

We now show that the general class of metrics (4.17), satisfying conditions (4.23) for the existence of a SCKV, does indeed contain space-times with energy-momentum tensors of the forms (1.1) and (1.2). In the case of the viscous fluid, it is known that the conformally flat null electrovac space-time, which is a special case of the simpler metric (4.19) with $H = f(u)(x^2 + y^2)$, where f is an arbitrary function of u, can be interpreted as representing a viscous fluid.[17] Here we note that the space-time (4.17), with

$$P = u^{-1}e^{x^2 + y^2}, \quad m = 0, \quad H = P^2 + 1\,, \tag{4.47}$$

admits the timelike SCKV $\xi^a = (-u^2, u^2, 0, 0)$ and satisfies the viscous fluid field equations, with T_{ab} given by (1.1), with

$$\begin{aligned}
&\mu = 4P^2(P^2+1)^{-1}[1 + 2P^2(1 + x^2 + y^2)]\,,\\
&P = \tfrac{4}{3}P^2(P^2+1)^{-1}[2P^2(x^2+y^2) - 1]\,,\\
&\eta = 3\times 2^{1/2}(P^2+1)^{1/2}[2P^2(x^2+y^2) - 1]u\,,\\
&Q = 4P^4(P^2+1)^{-1}(1 + 2x^2 + 2y^2)\,,\\
&u^a = [2^{-1/2}(P^2+1)^{-1/2}, 0, 0, 0]\,,\\
&q_a = Q\,[0, 2^{-1/2}(P^2+1)^{-1/2}, 0, 0]\,,
\end{aligned} \tag{4.48}$$

where, in order that $\eta \geqslant 0$, the solution is confined to that region of space-time for which $2P^2(x^2+y^2) \geqslant 1$, i.e., $2(x^2+y^2)e^{2(x^2+y^2)} \geqslant u^2$. Note that this restriction also ensures $p \geqslant 0$.

To represent an anisotropic fluid, T_{ab} must be of Segré type $\{(1,1)(11)\}$; this will be the case for the metric (4.17) if the following contribution holds:

$$RR_{00} = 2P^2(R_{02}^2 + R_{03}^2), \quad R \neq 0, \tag{4.49}$$

where the Ricci tensor components are given by (4.20). As an example, condition (4.49) is satisfied, with $R_{00} = R_{02} = R_{03} = 0$, by

$$P = u^{-1}e^{x^2+y^2}, \quad m = 0, \quad H_{xx} + H_{yy} = 0; \tag{4.50}$$

and the space-time satisfies the field equations for an anisotropic fluid with $R = 8P^2$,

$$\begin{aligned} u^a &= (2^{-1/2}H^{-1/2},0,0,0), \\ n^a &= (2^{-1/2}H^{-1/2}, -2^{1/2}H^{1/2},0,0), \end{aligned} \tag{4.51}$$

and $\mu, p_{\parallel}$, and $p_{\perp}$ given by Eq. (4.45). It is interesting to note that among this class of solutions given by Eq. (4.50), there are several possible behaviors for the SCKV ξ^a. For example, the following three choices for H satisfying (4.50):

$$\text{(i)} \quad H = \ln(x^2+y^2), \tag{4.52}$$

$$\text{(ii)} \quad H = u^{-2}\ln(x^2+y^2), \tag{4.53}$$

$$\text{(iii)} \quad H = u^{-2}, \tag{4.54}$$

lead to the following SCKV:

$$\text{(i)} \quad \xi^a = [-u^2, u^2\ln(x^2+y^2),0,0], \tag{4.55}$$

$$\text{(ii)} \quad \xi^a = (-u^2,0,0,0), \tag{4.56}$$

$$\text{(iii)} \quad \xi^a = (-u^2,2,0,0), \tag{4.57}$$

respectively. In case (i), ξ^a is null; in case (ii), ξ^a is timelike and parallel to u^a; and in case (iii), ξ^a is spacelike and parallel to n^a. Note that these space-times also admit a viscous fluid interpretation with a noncomoving velocity, as in the case of solutions (4.11) and (4.14).

Finally, we note that the space-times (4.11) and (4.14) contain no nontrivial vacuum solutions since when $R_{ab} = 0$, the space-times are flat. However, the metric (4.19) does contain vacuum space-times, namely the vacuum *pp*-wave solutions which satisfy the condition

$$H_{xx} + H_{yy} = 0. \tag{4.58}$$

Thus we have the following theorem.

Theorem 10: The only vacuum space-times admitting a proper SCKV are the *pp*-wave solutions of the form (4.19) with (4.58), which also satisfy condition (4.25).

V. EXAMPLES OF INHERITANCE PROPERTIES

Having found all space-times admitting a SCKV and satisfying the dominant energy condition, we now illustrate the theorems of Sec. III by investigating the inheritance properties of these solutions.

All the SCKV space-times can be intepreted as representing either viscous or anisotropic fluids, or both of these, except for the vacuum plane-wave solutions contained in the metric (4.19). These SCKV space-times may also admit other physical interpretations, such as a null electromagnetic field [the general null electrovac conformally flat space-time is contained in (4.19)] and a perfect fluid with an electromagnetic field (it was shown in Ref. 7 that an anisotropic fluid may be so interpreted). However, none of the SCKV space-times can be interpreted as a perfect fluid solution and none can be interpreted as a non-null electrovac solution. We are concerned here only with the viscous and anisotropic fluid interpretations; possible electromagnetic interpretations and their properties will be investigated elsewhere. We note that all spacetimes admitting a SCKV admit at least one null eigenvector, so that they form a very restricted set when interpreted as fluid space-times since, in general, neither the viscous fluid energy-momentum tensor (1.1) nor the anisotropic fluid energy-momentum tensor (1.2) admit a null eigenvector. For example, the FRW models, which have been shown to be solutions of the viscous fluid field equations,[18] do not admit a null eigenvector, in general. Furthermore, while the FRW models do not admit a SCKV, the $k = 0$ models with the scale factor $R(t) = t^a$ admit a HV and thus will provide us further illustrative examples of the inheritance theorems.

We first consider the viscous fluid solutions. The solution (4.11), in its viscous fluid form given by (4.35)–(4.37), has $q_a = 0$; thus from Eqs. (2.21) and (2.22) it follows that $\mathcal{L}_\xi \mu + 2\psi\mu = 0$ and $\mathcal{L}_\xi p + 2\psi p = 0$, a fact that is easily confirmed by calculating $\mathcal{L}_\xi \mu$ and $\mathcal{L}_\xi p$ with respect to the SCKV (4.12). Upon calculating $\mathcal{L}_\xi u^a$ and $\mathcal{L}_\xi(\eta\sigma_{ab})$ we find that

$$v^a = V(\sinh\phi, \cosh\phi, 0, 0), \tag{5.1}$$

$$\begin{aligned} \mathcal{L}_\xi(\eta\sigma_{00}) &= \tfrac{1}{3}Vx^{-2}\sinh 2\phi(1 + f^{-1}f_{yy}), \\ \mathcal{L}_\xi(\eta\sigma_{01}) &= -\tfrac{1}{3}Vx^{-2}\cosh 2\phi(1 + f^{-1}f_{yy}), \\ \mathcal{L}_\xi(\eta\sigma_{11}) &= \tfrac{1}{3}Vx^{-2}\sinh 2\phi(1 + f^{-1}f_{yy}), \end{aligned} \tag{5.2}$$

where

$$V = \tfrac{1}{2}(t^2 + x^2)\phi_t + tx\phi_x - x \tag{5.3}$$

and all other components of $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$. It follows that if $v^a = 0$, then $V = 0$ and so $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$ and vice versa. Thus the SCKV symmetry is inherited if either $\mathcal{L}_\xi u^a = -\psi u^a$ or $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$, in accordance with Theorems 1–3. Note that the condition $V = 0$ for inheritance implies that only those viscous models whose tilting velocity components satisfy this condition can inherit the SCKV symmetry.

The solution given by (4.14) and (4.39) to (4.41) behaves in a similar fashion. On the other hand, the solution given by (4.17), (4.47), and (4.48) cannot inherit the symmetry of the SCKV.

In order to illustrate the inheritance theorems in the case of a viscous fluid with nonzero heat conduction we turn to the FRW models. These models have an energy-momentum tensor of the Segré type $\{1,(1\ 1\ 1)\}$ and are thus commonly regarded as perfect fluid solutions. However, they can satisfy Einstein's field equations with an energy-momentum tensor of the form (1.1).[18] In such viscous fluid solutions, the four-velocity is necessarily tilting. While FRW models of any curvature can satisfy the viscous fluid field equations, we

will consider here only $k = 0$ models which admit a HV and, in particular, the Einstein–de Sitter model.

The known viscous fluid solutions with the Einstein–de Sitter metric falls into two classes. One class, known as *radial solutions*, is obtained by writing the metric in spherical polar coordinates and taking the four-velocity to have a nonzero radial component, while the second class, known as *axial solutions*, is obtained by using cylindrical polar coordinates and taking the four-velocity to have an axial component in the z direction. For our example we shall consider only the radial case in which the metric has the form

$$ds^2 = -dt^2 + t^{4/3}(dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\Phi^2) \tag{5.4}$$

and the four-velocity has the components

$$u^a = (\cosh\phi, t^{-2/3}\sinh\phi, 0, 0)\,, \tag{5.5}$$

where $\phi = \phi(t,r)$. The field equations then give the solution in the form

$$\begin{aligned} &\mu = \tfrac{4}{3}t^{-2}\cosh^2\phi, \quad p = \tfrac{4}{9}t^{-2}\sinh^2\phi, \\ &\eta X = -\tfrac{2}{3}t^{-2}\sinh^2\phi, \\ &q_a = \tfrac{4}{3}t^{-2}\sinh\phi\cosh\phi \\ &\qquad \times(\sinh\phi, -t^{2/3}\cosh\phi, 0, 0)\,, \end{aligned} \tag{5.6}$$

where

$$X = \phi_t\sinh\phi + t^{-2/3}\phi_r\cosh\phi - r^{-1}t^{-2/3}\sinh\phi\,. \tag{5.7}$$

The metric (5.4) admits the HV

$$\xi^a = (t, \tfrac{1}{3}r, 0, 0)\,, \tag{5.8}$$

corresponding to $\psi = 1$. Calculating $\mathcal{L}_\xi u^a$, $\mathcal{L}_\xi q_a$, and $\mathcal{L}_\xi(\eta\sigma_{ab})$ we find that

$$v^a = A(\sinh\phi, t^{-2/3}\cosh\phi, 0, 0)\,, \tag{5.9}$$

$$\begin{aligned} w_a = \tfrac{4}{3}At^{-2}[&(2\cosh^2\phi \\ &+ \sinh^2\phi)\cosh\phi, t^{2/3}(\cosh^2\phi \\ &+ 2\sinh^2\phi)\sinh\phi, 0, 0]\,, \end{aligned} \tag{5.10}$$

$$\begin{aligned} &\mathcal{L}_\xi(\eta\sigma_{00}) = 4A\coth\phi\eta\sigma_{00}\,, \\ &\mathcal{L}_\xi(\eta\sigma_{01}) = A(3\coth\phi + \tanh\phi)\eta\sigma_{01}\,, \\ &\mathcal{L}_\xi(\eta\sigma_{11}) = 2A(\coth\phi + \tanh\phi)\eta\sigma_{11}\,, \\ &\mathcal{L}_\xi(\eta\sigma_{22}) = 2A\coth\phi\eta\sigma_{22}\,, \\ &\mathcal{L}_\xi(\eta\sigma_{33}) = 2A\coth\phi\eta\sigma_{33}\,, \end{aligned} \tag{5.11}$$

where

$$A = t\phi_t + \tfrac{1}{3}r\phi_r \tag{5.12}$$

and all other components of $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$. In addition, we find that $\mathcal{L}_\xi\mu = -2\psi\mu + 2A\mu\tanh\phi$ and $\mathcal{L}_\xi p = -2\psi p + 2A\mu\coth\phi$. It follows that if $v^a = 0$, i.e., $A = 0$, we have complete inheritance, in accordance with Theorem 3. Furthermore, in this model q_a is an eigenvector of σ_{ab}, so that if $w^a = 0$, i.e., $A = 0$, we again have complete inheritance, thus illustrating Theorem 4.

Note that the inheritance condition $t\phi_t + \frac{1}{3}r\phi_r = 0$ implies that $\phi = \phi(\chi, \theta, \Phi)$, where

$$\chi = t^{1/3}r^{-1} \tag{5.13}$$

is the self-similar variable associated with the space-time (5.4). In fact, any viscous fluid $k = 0$ FRW model with $R(t) = t^a$ will inherit the symmetry of the HV admitted by such space-times if and only if the local Minkowskian components of the four-velocity are functions of the self-similar variable associated with the HV (as well as other coordinates not appearing in the self-similar variable), i.e., if and only if they are self-similar solutions. The perfect fluid solutions, which have comoving four-velocity, i.e., $\phi = 0$, are trivially self-similar and so inherit the symmetry.

Turning now to anisotropic fluid solutions, we first consider the solution (4.11) and its anisotropic fluid form given by (4.34). Using the SCKV ξ^a given by (4.12) we find

$$\begin{aligned} &\mathcal{L}_\xi u^a = -\psi u^a + v^a, \quad v^a = (0, x, 0, 0)\,, \\ &\mathcal{L}_\xi n^a = -\psi n^a = m^a, \quad m^a = (x, 0, 0, 0)\,. \end{aligned} \tag{5.14}$$

Note that u^a and n^a do not inherit the SCKV symmetry and that, in accordance with (2.30), v^a and m^a are indeed parallel to n^a and u^a, respectively, thus illustrating the apparent contradiction, mentioned earlier, with a result of Ref. 7.

The solution (4.14), in the form (4.39), and with the SCKV ξ^a given by (4.16), leads to expressions (5.15), but with v^a and m^a given by $v^a = (0, -t, 0, 0)$ and $m^a = (-t, 0, 0, 0)$.

The inheritance behavior of the solutions given by the metric (4.17) and Eqs. (4.50)–(4.57) is as follows: Solution (i) is noninheriting with $v^a = un^a$ and $m^a = uu^a$; solution (ii), in which ξ^a is parallel to u^a, obviously must inherit, i.e., $v^a = w^a = 0$; and solution (iii), in which ξ^a is orthogonal to u^a, is also an inheriting solution. The reason for inheritance in case (iii) is that since $\xi_a u^a = 0, \alpha = 0$ in Eq. (3.7) and, for this solution, the vorticity tensor $\omega_{ab} = 0$. Hence, Eq. (3.7) implies that $v_a = 0$ and, from (2.30), $m_a = 0$.

VI. CONCLUSION

This work consists essentially of two parts. In one part, namely Sec. IV, we found all space-times which admit a SCKV and satisfy the dominant energy condition. None of these space-times can represent a perfect fluid and the only vacuum solutions are given by the *pp*-wave metric. However, in general, these SCKV space-times can represent either viscous heat-conducting fluids or a special case of anisotropic fluids. In the second part, largely Sec. III, we derived theorems concerning the inheritance of the symmetries associated with a SCKV ξ^a by the physical components of a viscous imperfect fluid and also by those of the only type of anisotropic fluid that can admit a SCKV. The main results of Sec. III show that in the viscous fluid case, the SCKV symmetries are completely inherited if and only if either of the equivalent statements $\mathcal{L}_\xi(\eta\sigma_{ab}) = 0$ or $\mathcal{L}_\xi u^a = -\psi u^a$ (i.e., fluid flow lines are mapped conformally) is true. These results also apply to the symmetries associated with HV's and KV's.

Various subcases of the general imperfect fluid source are also covered by these results. Apart from an imperfect fluid ($\eta = q^a = 0$), which cannot admit a SCKV and for which the results are already known for HV's and KV's, these include a viscous fluid with no heat conduction ($q^a = 0$), the results for which are already known[7]; the

models (4.11) and (4.14) are examples of such fluid space-times admitting a SCKV. Another subcase is that of a heat-conducting perfect fluid ($\eta\sigma_{ab}=0$) which will always inherit the symmetry; this result is the generalization to nonzero heat conduction of a known result.[7]

The inheritance of symmetry results presented here can be extended to generalizations of the energy-momentum tensor (1.1). For example, the actions of KV's and HV's on an electromagnetic field and on an electromagnetic field with perfect fluid are well known[3]; an investigation of the effect of SCKV's on an electromagnetic field with imperfect fluid would be a logical extension. Such fields have been the subject of a number of cosmological investigations[18,19] and, in the same way as it has been shown that FRW models can be interpreted as electromagnetic field plus imperfect fluid models,[18] so, also, can some of the SCKV models found in Sec. IV be interpreted as such models.

Another possible extension is to multifluid models and, in particular, to two-fluid models. Models in which one fluid is a radiation perfect fluid, representing the cosmic microwave background, and the second fluid is either a perfect or an imperfect fluid, representing the galactic matter, have been studied extensively.[20] Again, as in the case of FRW models, the SCKV models of Sec. IV can also be interpreted as two-fluid models. However, whether we consider the case when one fluid is a radiation perfect fluid or the case of two general imperfect fluids, the expression for T_{ab} contains too many physical variables for the field equations to provide information on the inheritance properties of the separate physical quantities, although the inheritance theorems of Sec. III can be applied formally to suitable summed quantities. In the case when the four-velocities of the separate fluids are not parallel, the question of the symmetry inheritance is not well posed.

As stated earlier, our intended goal is the study of proper CKV's. Such a study is the natural mathematical generalization of work that has been done previously. Also, CKV's are of more physical interest than SCKV's. The results in this work will be useful in the proposed investigation. Moreover, some of the points made in this article serve to motivate the further study of CKV's and illustrate the potential problems inherent in such an investigation.

We have shown that there are very few space-times admitting SCKV's. In particular, there exist no SCKV's in FRW space-times. However, it is known that there do exist proper CKV's in FRW models[21] (nine in general) including the simple timelike CKV $\xi^a = R(\partial/\partial t)$. This indicate the greater physical significance in the study of proper CKV's.

In the case of a proper CKV with $\psi_{;ab}\neq 0$, it can be seen from Eqs. (2.19) that $\mathcal{L}_\xi T_{ab}$ is no longer zero. By studying the analog of Eq. (2.24), it can be shown that the equation for $\mathcal{L}_\xi(\eta\sigma_{ab})$ now includes the term $\psi_{;ab}$ on the rhs and cannot be shown to be zero when $\psi_{;ab}\neq 0$. Thus in the case, it is impossible for the physical quantities to satisfy (3.1); consequently, the symmetries cannot be inherited in the sense defined in Sec. III. In particular, it can be shown that even in the case of a perfect fluid source, a conformal motion will not, in general, map fluid flow conformally (i.e., $\mathcal{L}_\xi u^a \neq -\psi u^a$). Clearly, one of the starting points of future research is a notion of what is actually meant by symmetry inheritance in space-times admitting CKV's and what modifications are required to equations such as (3.1) for proper CKV's.

ACKNOWLEDGMENTS

We would like to thank Jack Gegenberg and Bob McKellar for helpful comments on Sec. IV. We are particularly grateful to Graham Hall for his interest in this work.

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada through operating grants to each author.

[1] M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[2] A. H. Taub, in *General Relativity: Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Oxford U.P., London, 1972).
[3] J. Wainwright and P. E. A. Yaremovich, Gen. Relativ. Gravit. **7**, 345, 596 (1976).
[4] C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 199 (1976).
[5] L. Herrera, J. Jiménez, L. Leal, J. Ponce de León, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).
[6] D. P. Mason and M. Tsamparlis, J. Math. Phys. **26**, 2881 (1985).
[7] R. Maartens, D. P. Mason, and M. Tsamparlis, J. Math. Phys. **27**, 2987 (1986).
[8] G. S. Hall and D. A. Negm, Int. J. Theor. Phys. **25**, 405 (1986).
[9] D. M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[10] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1955).
[11] C. A. Kolassis, N. O. Santos, and D. Tsoubelis, Class. Quantum Grav. **5**, 1329 (1988).
[12] H. Stephani, *General Relativity* (Cambridge U.P., Cambridge, 1982).
[13] A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[14] N. S. Sinyukov, *Scientific Annual*, Odessa University (1957).
[15] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge, U.P., Cambridge, 1980).
[16] G. S. Hall, Arabian J. Sci. Eng. **9**, 88 (1984).
[17] B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).
[18] A. A. Coley and B. O. J. Tupper, Astrophys. J. **280**, 26 (1984).
[19] J. B. Benton and B. O. J. Tupper, Phys. Rev. D **33**, 3534 (1986).
[20] A. A. Coley and B. O. J. Tupper, J. Math. Phys. **27**, 406 (1986).
[21] R. Maartens and S. D. Maharaj, Class. Quantum Grav. **3**, 1005 (1986).

Linear perturbations of plane polarized plane waves. I. The absence of purely incoming perturbations

Basilis C. Xanthopoulos
Department of Physics, University of Crete and Research Center of Crete, Iraklion, Crete, Greece

(Received 9 November 1988; accepted for publication 26 April 1989)

The linearized perturbations of plane polarized plane gravitational waves coupled with electromagnetic waves are studied. The dependence of the perturbations on the two spatial directions spanning the planar wave fronts and on advanced time is analyzed in normal modes (since the background depends only on retarded time). The linearized Newman–Penrose equations (Ricci and Bianchi identities, commutation relations, and Maxwell equations) are written explicitly for the general perturbations and without imposing any gauge condition. Then, making the assumption that there are no outgoing perturbations—i.e., that the perturbations do not depend on advanced time—the surprising result that there are no nontrivial purely incoming perturbations is obtained.

I. INTRODUCTION

In our studies, in collaboration with Chandrasekhar, of collisions of plane gravitational waves,[1–4] possibly coupled with electromagnetic or hydrodynamic waves, we have established some interesting results predicted by general relativity: impulsive plane gravitational waves could be coupled only with shock electromagnetic waves[2]; the collision of plane symmetric null dust might result in the formation of a massive fluid with a stiff equation of state[3]; and the collision of plane gravitational and /or electromagnetic waves might result in the formation of a Cauchy horizon—as opposed to the formation of a three-dimensional spacelike curvature singularity—to the future of which timelike or spacelike, two-dimensional or three-dimensional curvature singularities,[4,5] or even no curvature singularities at all,[6] may develop.

All these conclusions were obtained under the assumption of exact plane symmetry and they are supported by exact solutions of the field equations describing them. However, one is left wondering whether and which of these conclusions will continue to hold under the weaker, and closer to reality, assumption of almost plane symmetry.

Yurtsever[7,8] has already undertaken an investigation of the stability of the formation of Cauchy horizons in the collisions of plane gravitational waves. By considering only perturbations that preserve the planar symmetry and by allowing for arbitrary initial data in the characteristic surfaces emanating from the collision he has shown that near the Cauchy horizon (of the background) the space-time resembles a Kasner space-time and that in the generic case the Cauchy horizon is transformed to a three-dimensional spacelike curvature singularity.

Yurtsever's analysis is restricted in two respects. First, he perturbs the plane waves with plane symmetric perturbations. Thus, in particular, the unsatisfactory property of the plane waves that they have infinite extent (and, therefore, that they carry infinite energy) is preserved. Were the perturbations nonplanar, one could envision constructing, by a suitable superposition of their normal modes, perturbations of compact spatial support corresponding to finite total energy. And, second, Yurtsever assumes arbitrary characteristic initial data for the interaction region of the collision, without considering whether all of them are compatible with the incoming waves, prior to the collision. However, our analysis,[9] with Chandrasekhar, of the perturbations of the Bell–Szekeres[10] space-time has demonstrated that this may well be a very tricky issue. In our analysis we found that in region II of the space-time (see, for instance, Fig. 1 of Ref. 2 for the standard picture of the space-times describing collisions of plane waves), which describes one of the waves bound for the collision, there are no nonplanar perturbations that do not depend on advanced time and that the general perturbations—i.e., those consisting of a mixture of incoming and outgoing waves—become singular even before the collision. This analysis suggested that[9] "any formulation of the problem of the future time-development of pre-assigned initial values on the null boundary at $u = 0$, without an adequate assessment of the implications of the space-time in Region II of the postulated waves in this region, should be viewed with skepticism."

The present investigation was motivated by the impasse we reached in Ref. 9 when matching the perturbations of the Bell–Szekeres space-time in the two regions, before and after the collision. Soon it became clear that perturbations of (single) plane gravitational waves might have interesting (and surprising) properties. In this paper we show that the "peculiarity" of the Bell–Szekeres space-time, i.e., the absence of purely incoming linear perturbations, is a property of the most general space-time representing plane polarized plane gravitational and electromagnetic waves. It should be clarified that it is crucial for our reductions that the perturbations we are considering do not respect the planar symmetry, and that they are squared integrable. There are nontrivial plane symmetric perturbations of plane waves and their complete determination will be the subject of a future communication.

In the first sections of the paper we formulate the problem and we derive the linearized Newman–Penrose equations for the general perturbations of plane polarized plane waves in a framework more general than that required for obtaining the conclusion of Sec. VI. More particularly, in the

 0022-2488/89/112626-12$02.50

analysis of Secs. III–V we include dependence of the perturbations on advanced time as well, to avoid the repetition of rederiving such large sets of equations for the most general perturbations we shall be considering in subsequent communications. In addition, the linearized equations have been obtained without the imposition of any gauge condition, tetrad or coordinate. Because of the high symmetry of the background (it admits five Killing fields), all but three of the spin coefficients (λ, μ, γ) and all but two of the Weyl and Maxwell scalars (Ψ_4, ϕ_2) vanish in the background. Thus the possible choices of gauge conditions are rather restrictive (see Sec. V) and the choices of gauge that can be imposed for different kinds of perturbations (independent of advanced time, plane symmetric, general) do not coincide.

Finally, throughout the paper we consider the perturbations of plane gravitational waves coupled with electromagnetic waves. Since it is the custom in the literature to give only the plane pure gravitational waves, in Appendix A we compare the different expressions for the space-time metric without the imposition of the Einstein vacuum equations, for the Einstein–Maxwell waves. In Appendix B we solve the perturbation equations of the Bell–Szekeres space-time in a gauge invariant manner.

II. THE UNPERTURBED SPACE-TIME

The background space-time, whose perturbations we shall be considering, will be the space-time of region II (or III) in the standard picture for colliding plane waves. It describes one of the plane gravitational waves (possibly coupled with an electromagnetic wave or with null dust) bound for the collision. *We shall be making the assumption that the two spacelike Killing fields $(\partial/\partial x^1)$ and $(\partial/\partial x^2)$, which span the wave front of the propagating waves, are hypersurface orthogonal.* Those space-times are described by a metric of the form (Ref. 2, Sec. 8, Eq. 150)

$$\begin{aligned} ds^2 &= U^2[(dx^0)^2 - (dx^3)^2] \\ &\quad - (1-v^2)[\chi^{-1}(dx^1)^2 + \chi(dx^2)^2] \\ &= U^2(du)(dv) - (1-v^2)[\chi^{-1}(dx^1)^2 + \chi(dx^2)^2], \end{aligned} \tag{2.1}$$

where

$$v = x^0 - x^3, \quad u = x^0 + x^3 \tag{2.2}$$

are retarded and advanced time, respectively, and

$$U = U(v), \quad \chi = \chi(v). \tag{2.3}$$

(For comparison with the notation of Sec. 8 of Ref. 2 note that $U = e^{\nu}$.)

A suitable Newman–Penrose null tetrad for the metric (2.1) is

$$\begin{array}{lllll} & \underline{x^0} & \underline{x^3} & \underline{x^1} & \underline{x^2} \\ l_i = \dfrac{1}{\sqrt{2}} & (U & -U & 0 & 0), \\ n_i = \dfrac{1}{\sqrt{2}} & (U & U & 0 & 0), \\ m_i = -\dfrac{\sqrt{1-v^2}}{\sqrt{2}} & (0 & 0 & \chi^{-1/2} & -i\chi^{1/2}), \end{array} \tag{2.4}$$

$$\bar{m}_i = -\frac{\sqrt{1-v^2}}{\sqrt{2}} \quad (0 \quad 0 \quad \chi^{-1/2} \quad i\chi^{1/2}).$$

In addition we shall be using the contravariant components of this null tetrad in the null coordinates (u, v, x^1, x^2). They are

$$\begin{array}{lllll} & \underline{u} & \underline{v} & \underline{x^1} & \underline{x^2} \\ l^i = \sqrt{2}U^{-1} & (1 & 0 & 0 & 0), \\ n^i = \sqrt{2}U^{-1} & (0 & 1 & 0 & 0), \\ m^i = \dfrac{1}{\sqrt{2}\sqrt{1-v^2}} & (0 & 0 & \chi^{1/2} & -i\chi^{-1/2}), \\ \bar{m}^i = \dfrac{1}{\sqrt{2}\sqrt{1-v^2}} & (0 & 0 & \chi^{1/2} & i\chi^{-1/2}). \end{array} \tag{2.5}$$

The directional derivative operators associated with this tetrad are

$$\begin{aligned} D = (l) &= \frac{\sqrt{2}}{U}\frac{\partial}{\partial u}, \quad \Delta = (n) = \frac{\sqrt{2}}{U}\frac{\partial}{\partial v}, \\ \delta = (m) &= \frac{1}{\sqrt{2}\sqrt{1-v^2}}\left(\chi^{1/2}\frac{\partial}{\partial x^1} - i\chi^{-1/2}\frac{\partial}{\partial x^2}\right), \\ \delta^* = (\bar{m}) &= \frac{1}{\sqrt{2}\sqrt{1-v^2}}\left(\chi^{1/2}\frac{\partial}{\partial x^1} + i\chi^{-1/2}\frac{\partial}{\partial x^2}\right). \end{aligned} \tag{2.6}$$

It is straightforward to calculate the spin coefficients and the curvature scalars of this tetrad (Ref. 2, Sec. 8). One finds that

(a) the only nonvanishing spin coefficients are

$$\lambda = -\frac{1}{U\sqrt{2}}\frac{\chi'}{\chi}, \quad \mu = -\frac{\sqrt{2}}{U}\frac{v}{1-v^2}, \quad \gamma = -\frac{U'}{\sqrt{2}U^2}, \tag{2.7}$$

where the prime denotes differentiation with respect to v; and

(b) the curvature of the metric (2.1) has two degrees of freedom. The two independent, nonvanishing curvature scalars are described by

$$M + L = \frac{1}{U^2}\left[-\frac{2}{(1-v^2)^2} + \frac{4v}{1-v^2}\frac{U'}{U} + \frac{\chi'^2}{2\chi^2}\right], \tag{2.8}$$

$$\begin{aligned} M - L &= U^{-2}\left[(\ln x)'' - 2(\ln x)'\left(\frac{U'}{U} + \frac{v}{1-v^2}\right)\right] \\ &= \frac{1}{(1-v^2)}\left[\frac{(1-v^2)x'}{xU^2}\right]'. \end{aligned} \tag{2.9}$$

Note that $M - L$ and $M + L$ provide all the information about the Weyl and the Ricci curvature, respectively. In the standard Newman–Penrose notation and for the null tetrad (2.4)

$$\Psi_4 = M - L, \tag{2.10}$$

while all the remaining Weyl scalars vanish. Thus, in particular, all space-times described by the metric (2.1) are of Petrov type N. Moreover

(i) when $M + L = 0$, the Ricci curvature vanishes and the metric (2.1) is a solution of the vacuum Einstein equations;

(ii) when $M + L < 0$, the metric (2.1) is a solution the Einstein–Maxwell (electrovacuum) equations and the only nonvanishing Maxwell scalar is

$$\Phi_{22} = \phi_2\phi_2^* = -(M + L). \tag{2.11}$$

The case $M+L<0$ can equally well be interpreted as a solution of the Einstein equations coupled with null dust [energy-momentum tensor $T_{ab}=\epsilon u_a u_b$, $u_0^2-u_3^2=0$, and $2\epsilon u_0^2=-(M+L)$, see Ref. 3, Sec. 12, but for a different convention].

(iii) When $M+L>0$, the metric (2.1) does not satisfy a physically acceptable energy condition.

In summary, the metric (2.1) describes plane polarized plane waves that are purely gravitational when $L+M=0$, gravitational waves coupled with electromagnetic waves or with null dust when $L+M<0$, and it is physically unacceptable when $L+M>0$.

In linearizing the Newman–Penrose equations in Sec. IV we shall need the action of the derivative operator Δ on the nonvanishing spin coefficients. We readily find that

$$\begin{aligned}\Delta\lambda &= U^{-2}[(\ln U)'(\ln\chi)' - (\ln\chi)''],\\ \Delta\mu &= 2U^{-2}\left[\frac{v}{1-v^2}(\ln U)' - \frac{1+v^2}{(1-v^2)^2}\right],\\ \Delta\gamma &= U^{-1}(U^{-1})'' = \frac{2U'^2}{U^4} - \frac{U''}{U^3}.\end{aligned} \tag{2.12}$$

Before closing this section we would like to mention that the tetrad (2.4) we shall be using in the present paper is slightly different from the null tetrad used (in region II) in our previous studies, with Chandrasekhar, of collisions of plane waves. A characteristic of the tetrad (2.4) is that it is symmetrical in l and n and in the null coordinates u and v. On the other hand, the tetrad used in Refs. 2–5, and 9 is obtained from the symmetrical tetrad by rescaling the vector l by a suitable scalar factor and the vector n by the inverse factor (so that $l\cdot n=1$ is preserved). The asymmetrical tetrad was adopted in Refs. 2–5 because it leads to a continuous matching of the obtained curvature scalars Ψ_4 and Φ_{22} of region II with their values obtained in the interaction region I. The same asymmetrical null tetrad was adopted in the study of the perturbations of the Bell–Szekeres space-time (Ref. 9) because it implies, in addition to the continuity of Ψ_4 and Φ_{22}, that the only nonvanishing spin coefficient in the background is μ. It turns out that this particularly nice property of the Bell–Szekeres space-time does not generalize to all the metrics of the form (2.1) and there is no advantage whatever to choosing the asymmetrical null tetrad. In this paper we shall be using the symmetrical null tetrad given by Eqs. (2.4) or (2.5) exclusively.

III. PERTURBATIONS: THE SETUP

The perturbation analysis of the plane polarized plane waves will be performed by using the Newman–Penrose formalism (for a review, see Ref. 11, Sec. 8; this book will be referred to hereafter as M. T.) In this formalism the unknowns are (a) the spin coefficients $\kappa, \sigma, \lambda, \nu, \rho, \mu, \tau, \pi, \epsilon, \gamma, \alpha$, and β; (b) the Weyl and the Maxwell scalars $\Psi_0, \Psi_1, \Psi_2, \Psi_3, \Psi_4$ and ϕ_0, ϕ_1, ϕ_2; and (c) the null vectors l, n, m, and $\bar{m}$.

For the perturbed quantities, (a') we shall denote by $\mu^{(1)}$, $\gamma^{(1)}$, and $\lambda^{(1)}$ the perturbations of the spin coefficients that do not vanish in the background, while keeping the same symbol $\kappa, \sigma, \nu, \rho, \tau, \pi, \epsilon, \alpha$, and β for the perturbations of those that do vanish in the background; (b') we shall denote, similarly, by $\Psi_4^{(1)}$ and $\phi_2^{(1)}$ the perturbations of those Weyl and Maxwell scalars that do not vanish in the background and by $\Psi_0, \Psi_1, \Psi_2, \Psi_3, \phi_0$, and ϕ_1 the perturbations of those that vanish in the background; and (c') the perturbations of the null vectors l, n, m, and $\bar{m}$ will be projected in the unperturbed tetrad according to

$$\begin{aligned}l^{(1)} &= A_1^1 l + A_2^1 n + A_3^1 m + A_4^1 \bar{m},\\ n^{(1)} &= A_1^2 l + A_2^2 n + A_3^2 m + A_4^2 \bar{m},\\ m^{(1)} &= A_1^3 l + A_2^3 n + A_3^3 m + A_4^3 \bar{m},\\ \bar{m}^{(1)} &= A_1^4 l + A_2^4 n + A_3^4 m + A_4^4 \bar{m},\end{aligned} \tag{3.1}$$

and they are described by the 4×4 matrix (A_j^i). Obviously, complex conjugation in A_j^i would interchange the indices 3 and 4.

By linearizing the relation

$$g_{ab} = 2l_{(a}n_{b)} - 2m_{(a}\bar{m}_{b)}, \tag{3.2}$$

it is straightforward to find the perturbation of the metric itself:

$$\begin{aligned}g_{ab}^{(1)} = {} & 2A_1^2 l_a l_b + 2A_2^1 n_a n_b + 2(A_1^1 + A_2^2) l_{(a}n_{b)}\\ & - 2(A_3^3 + A_4^4) m_{(a}\bar{m}_{b)} - 2A_3^4 m_a m_b - 2A_4^3 \bar{m}_a \bar{m}_b\\ & + 2(A_3^2 - A_1^4) l_{(a}m_{b)} + 2(A_4^2 - A_1^3) l_{(a}\bar{m}_{b)}\\ & + 2(A_3^1 - A_2^4) n_{(a}m_{b)} + 2(A_4^1 - A_2^3) n_{(a}\bar{m}_{b)}.\end{aligned} \tag{3.3}$$

The associated perturbations of the metric coefficients are given explictly by

$$\begin{aligned}&g^{uu(1)} = 4U^{-2}A_1^2, \quad g^{vv(1)} = 4U^{-2}A_2^1, \quad g^{uv(1)} = 2U^{-2}(A_1^1 + A_2^2),\\ &g^{x^1x^1(1)} = -\frac{\chi}{1-v^2}(A_3^3 + A_4^4 + A_4^3 + A_3^4), \quad g^{x^2x^2(1)} = \frac{1}{\chi(1-v^2)}(A_4^3 + A_3^4 - A_3^3 - A_4^4),\\ &g^{ux^1(1)} = \frac{\chi^{1/2}}{U\sqrt{1-v^2}}(A_3^2 + A_4^2 - A_1^3 - A_1^4), \quad g^{ux^2(1)} = \frac{i}{U\chi^{1/2}\sqrt{1-v^2}}(A_4^2 - A_3^2 + A_1^4 - A_1^3),\\ &g^{vx^1(1)} = \frac{\chi^{1/2}}{U\sqrt{1-v^2}}(A_3^1 + A_4^1 - A_2^3 - A_2^4), \quad g^{x^1x^2(1)} = \frac{i(A_3^4 - A_4^3)}{1-v^2},\\ &g^{vx^2(1)} = \frac{i}{U\chi^{1/2}\sqrt{1-v^2}}(A_4^1 - A_3^1 + A_2^4 - A_2^3).\end{aligned} \tag{3.4}$$

Because the background space-time (2.1) admits the three commuting Killing vectors $(\partial/\partial x^1)$, $(\partial/\partial x^2)$, and $(\partial/\partial u)$, we Fourier analyze the x^1, x^2, and u dependence of the perturbations. Thus for all the perturbed quantities we assume that the x^1, x^2, and u dependences are of the form

$$e^{i(k_1x^1+k_2x^2+k_3u)}, \tag{3.5}$$

where k_1, k_2, and k_3 are real constants. In what follows we shall suppress the common phase factor (3.5) from all the perturbed quantities while keeping the same symbols—introduced in (a'), (b'), and (c') above—to describe the amplitudes of the perturbations. Obviously, these amplitudes would depend on the coordinate v and the parameters k_1, k_2, and k_3 which characterize the different modes.

Because all the perturbations have the dependence (3.5) on x^1, x^2, and u, the directional derivatives (2.6) acting on all perturbed quantities become

$$\begin{aligned} &D\to\frac{\sqrt{2}}{U}ik_3, \quad \Delta\to\frac{\sqrt{2}}{U}\frac{d}{dv}, \\ &\delta\to(1/\sqrt{2}\sqrt{1-v^2})(ik_1\chi^{1/2}+k_2\chi^{-1/2}), \\ &\delta^*\to(1/\sqrt{2}\sqrt{1-v^2})(ik_1\chi^{1/2}-k_2\chi^{-1/2}). \end{aligned} \tag{3.6}$$

Expressions (3.6) should be contrasted with the same directional derivatives acting on background quantities, in which case they read

$$D=0, \quad \Delta=\frac{\sqrt{2}}{U}\frac{d}{dv}, \quad \delta=\delta^*=0. \tag{3.7}$$

Finally we shall add some clarifying remarks about the "nonstandard" notion of complex conjugation we shall be using in our analysis. The same notion of complex conjugation was invoked in M.T., Chap. 9, in the treatment of the perturbations of the Kerr space-time, and, more recently in Ref. 9, in the treatment of the perturbations of the Bell–Szekeres space-time. However, although extensively used, this "nonstandard" notion of complex conjugation has never been elucidated in writing.

Let **B** and **Γ** stand for two typical Newman–Penrose perturbed quantities that generally will be complex. The Newman–Penrose formalism will provide for them linear differential equations, some involving **B** and **Γ** and some **B** and **Γ***. Then **B** and **Γ** will be Fourier analyzed, picking the phase factors (3.5). For the phase factors to drop out of the equations involving **B**, **Γ** and **B**, **Γ*** we have to assume that

$$\begin{aligned} &\mathbf{\Gamma}(v,x^1,x^2,u)=\mathbf{\Gamma}(v,k_1,k_2,k_3)e^{i(k_1x^1+k_2x^2+k_3u)} \\ &\text{and} \\ &\mathbf{\Gamma}^*(v,x^1,x^2,u)=\mathbf{\Gamma}^*(v,k_1,k_2,k_3)e^{i(k_1x^1+k_2x^2+k_3u)}. \end{aligned} \tag{3.8}$$

Therefore, the complex conjugation in the amplitudes would require that

$$k_1\to k_1, \quad k_2\to k_2, \quad k_3\to k_3, \quad i\to i; \tag{3.9}$$

this is what Chandrasekhar privately refers to as the *dishonest* complex conjugation. In some of the Newman–Penrose equations, however, **B** and **Γ** will be acted on by the complex differential operators δ and δ^* [(2.6)]. The conjugation between δ and δ^* is an *honest* complex conjugation, and it is effective by the change $i\to -i$. Since x^1 and x^2 dependence appears only in the phase factors (3.5) we shall have

$$\begin{aligned} &\delta\sim\left(\chi^{1/2}\frac{\partial}{\partial x^1}-i\chi^{-1/2}\frac{\partial}{\partial x^2}\right)=\mathbf{X}^+, \\ &\delta^*\sim\left(\chi^{1/2}\frac{\partial}{\partial x^1}+i\chi^{-1/2}\frac{\partial}{\partial x^2}\right)=\mathbf{X}^-, \end{aligned} \tag{3.10}$$

where we have introduced the notation

$$\mathbf{X}^+=(ik_1\chi^{1/2}+k_2\chi^{-1/2}), \quad \mathbf{X}^-=(ik_1\chi^{1/2}-k_2\chi^{-1/2}). \tag{3.11}$$

Hence, in all the equations, complex conjugation on k_1 will always be dishonest while on k_2 there will be a combination of an honest and a dishonest conjugation. Similarly, because the operator D, Eq. (2.6), is real, the conjugation on k_3 will always be dishonest.

The upshot of the previous remarks is that complex conjugation on the amplitudes of the perturbed quantities will be effected by the changes

$$k_1\to k_1, \quad k_2\to -k_2, \quad k_3\to k_3, \quad i\to i \quad ! \tag{3.12}$$

IV. THE LINEARIZED NEWMAN–PENROSE EQUATIONS

We have already mentioned that the space-time (2.1) we shall be considering describes plane polarized plane gravitational waves, possibly coupled with electromagnetic waves or with null dust. We shall analyze the perturbations only in the Einstein–Maxwell case, i.e., when $L+M<0$. Then the pure gravitational case follows as a particular case while the case of null dust can be obtained similiarly.

The set of the Newman–Penrose equations that ought to be considered are the 36 Ricci identities, the 16 Bianchi identities, the 8 Maxwell equations, and the 24 commutation relations, i.e., 84 equations altogether, when we are counting real equations as one and complex equations as two. These equations (before the linearization) can be obtained, for instance, from M. T., pp. 46 and 47, Eqs. (310) for the Ricci identities, pp. 49 and 50, Eqs. (321) and (321') for the Bianchi identities, and p. 52, Eqs. (330)–(333) for the Maxwell equations. For the linearization of the commutation relations [M.T., p. 45, Eqs. (303)–(306)], which is more involved, we adopt the procedure described in M. T., Sec. 84, pp. 448–450, Eqs. (112)–(115), where we enumerate the null vectors as $i=1,2,3,4$ in the order l, n, m, $\bar{m}$, respectively.

For the background value of the Maxwell scalar ϕ_2 we shall ignore a physically irrelevant phase factor (which could depend on v) and we shall choose

$$\phi_2=\sqrt{\Phi_{22}}=\sqrt{-(L+M)}=H \text{ (say)}. \tag{4.1}$$

Since $\Phi_{mn}=\phi_m\phi_n^*$ and since ϕ_0 and ϕ_1 vanish in the background we readily find that Φ_{00}, Φ_{01}, Φ_{11} will continue to vanish to first order,

$$\Phi_{00}^{(1)}=\Phi_{01}^{(1)}=\Phi_{11}^{(1)}=0, \tag{4.2}$$

while

$$\begin{aligned} &\Phi_{02}^{(1)}=H\phi_0, \quad \Phi_{20}^{(1)}=H\phi_0^*, \quad \Phi_{12}^{(1)}=H\phi_1, \\ &\Phi_{21}^{(1)}=H\phi_1^*, \quad \Phi_{22}^{(1)}=H(\phi_2^{(1)}+\phi_2^{*(1)}). \end{aligned} \tag{4.3}$$

By using Eqs. (2.12) as well it is straightforward to obtain the following linearized equations: *Linearized Ricci identities,*

(a) $2ik_3\rho\sqrt{1-v^2} - U\kappa\mathbf{X}^- = 0,$

(b) $2ik_3\sigma - \frac{U}{\sqrt{1-v^2}}\kappa\mathbf{X}^+ = U\Psi_0\sqrt{2},$

(c) $2U^{-2}(U^2k)' - 2ik_3\tau + \Psi_1 U\sqrt{2} = 0,$

(d) $2ik_3\alpha - \frac{U}{\sqrt{1-v^2}}\epsilon\mathbf{X}^- - \kappa(\ln\chi)' - \kappa^*(\ln U)' = 0,$

(e) $-2ik_3\beta + \frac{U}{\sqrt{1-v^2}}\epsilon\mathbf{X}^+ + \kappa\left[\ln\left(\frac{U}{1-v^2}\right)\right]' + \Psi_1 U\sqrt{2} = 0,$

(f) $2U^{-3/2}(U^{3/2}\epsilon)' + \epsilon^*(\ln U)' - 2ik_3\gamma^{(1)} - A\,{}^1_2\sqrt{2}(U^{-1})'' + U\Psi_2\sqrt{2} = 0,$

(g) $-\frac{U\pi\mathbf{X}^-}{\sqrt{1-v^2}} + 2ik_3\lambda^{(1)} - \sigma^*[\ln(1-v^2)]' + (\rho+\epsilon^*-3\epsilon)(\ln\chi)' - HU\phi_0^*\sqrt{2} - \sqrt{2}A\,{}^1_2\left(\frac{\chi'}{\chi U}\right)' = 0,$

(h) $-\frac{U\pi\mathbf{X}^+}{\sqrt{1-v^2}} + 2ik_3\mu^{(1)} + \sigma(\ln\chi)' + \frac{2v}{1-v^2}(\rho^* - \epsilon - \epsilon^*)$
$- U\Psi_2\sqrt{2} + \frac{2\sqrt{2}A\,{}^1_2}{U}\left[\frac{v}{1-v^2}(\ln U)' - \frac{1+v^2}{(1-v^2)^2}\right] = 0,$

(i) $\frac{2}{\sqrt{1-v^2}}(\pi\sqrt{1-v^2})' - \frac{2v}{1-v^2}\tau^* - (\pi^*+\tau)(\ln\chi)' - 2ik_3\nu + \Psi_3 U\sqrt{2} + HU\phi_1^*\sqrt{2} = 0,$

(j) $\frac{2U}{1-v^2}[(1-v^2)U^{-1}\lambda^{(1)}]' - (\mu^{(1)} + \mu^{(1)*} + 3\gamma^{(1)} - \gamma^{(1)*})(\ln\chi)' - \sqrt{2}A\,{}^2_2\left(\frac{\chi'}{U\chi}\right)' - \frac{\nu U\mathbf{X}^-}{\sqrt{1-v^2}} + \Psi_4^{(1)}U\sqrt{2} = 0,$

(k) $\rho\mathbf{X}^+ - \sigma\mathbf{X}^- + \Psi_1\sqrt{2}\sqrt{1-v^2} = 0,$ (R)

(l) $\frac{U}{\sqrt{1-v^2}}(\alpha\mathbf{X}^+ - \beta\mathbf{X}^-) + \frac{2v}{1-v^2}\rho - \sigma(\ln\chi)' + (\rho-\rho^*)(\ln U)' + U\Psi_2\sqrt{2} = 0,$

(m) $\frac{U}{\sqrt{1-v^2}}(\lambda^{(1)}\mathbf{X}^+ - \mu^{(1)}\mathbf{X}^-) - \sqrt{2}A\,{}^3_2\left(\frac{\chi'}{U\chi}\right)' + \frac{2v}{1-v^2}(\alpha+\beta^*) + (\alpha^*-3\beta)(\ln\chi)'$
$+ \Psi_3 U\sqrt{2} - HU\phi_1^*\sqrt{2} - \frac{2\sqrt{2}A\,{}^4_2}{U}\left[\frac{v}{1-v^2}(\ln U)' - \frac{1+v^2}{(1-v^2)^2}\right] = 0,$

(n) $-\frac{2U}{1-v^2}[(1-v^2)U^{-1}\mu^{(1)}]' + \frac{\nu U\mathbf{X}^+}{\sqrt{1-v^2}} + (\lambda^{(1)} + \lambda^{(1)*})(\ln\chi)'$
$+ \frac{2v}{1-v^2}(\gamma^{(1)} + \gamma^{(1)*}) - HU\sqrt{2}(\phi_2^{(1)} + \phi_2^{(1)*}) - \frac{2\sqrt{2}A\,{}^2_2}{U}\left[\frac{v}{1-v^2}(\ln U)' - \frac{1+v^2}{(1-v^2)^2}\right] = 0,$

(o) $\frac{U\gamma^{(1)}\mathbf{X}^+}{\sqrt{1-v^2}} - 2\beta' + \frac{2v}{1-v^2}(\beta+\tau) + (\tau - \alpha^* - \beta)(\ln U)' + \alpha(\ln\chi)'$
$+ A\,{}^3_2\sqrt{2}(U^{-1})'' - HU\phi_1\sqrt{2} = 0,$

(p) $2U^{-1}(1-v^2)^{-1/2}[\sigma U\sqrt{1-v^2}]' - \rho(\ln\chi)' + HU\phi_0\sqrt{2} - \frac{\tau U\mathbf{X}^+}{\sqrt{1-v^2}} = 0,$

(q) $2U^{-1}(1-v^2)^{-1/2}[\rho U\sqrt{1-v^2}]' - \sigma(\ln\chi)' + U\Psi_2\sqrt{2} - \frac{\tau U\mathbf{X}^-}{\sqrt{1-v^2}} = 0,$

(r) $2\alpha' + (\alpha + \beta^* - \tau^*)(\ln U)' - \frac{2v}{1-v^2}\alpha - (\tau+\beta)(\ln\chi)' - A\,{}^4_2(U^{-1})''\sqrt{2} + U\Psi_3\sqrt{2} - \frac{U\gamma^{(1)}\mathbf{X}^-}{\sqrt{1-v^2}} = 0;$

Linearized Bianchi identities,

(a) $2ik_3\Psi_1\sqrt{1-v^2} - \Psi_0 U\mathbf{X}^- = 0,$

(b) $\Psi_0(\ln\chi)' + \frac{U\Psi_1\mathbf{X}^-}{\sqrt{1-v^2}} - 2ik_3\Psi_2 = 0,$

(c) $-2\Psi_1(\ln\chi)' + \frac{U}{\sqrt{1-v^2}}(H\phi_0^*X^+ - \Psi_2X^-) + \kappa U(M-L)\sqrt{2}$
$+ 2ik_3(\Psi_3 - H\phi_1^*) - \kappa^* UH^2\sqrt{2} = 0,$

$$
\begin{aligned}
&\text{(d)}\ -\frac{2}{\sqrt{1-v^2}}(H\phi_0^*\sqrt{1-v^2})'+\frac{U(\Psi_3+H\phi_1^*)\mathbf{X}^-}{\sqrt{1-v^2}}+3\Psi_2(\ln\chi)'\\
&\quad -2ik_3\Psi_4^{(1)}+U\sqrt{2}\left[H^2\sigma^*-4\epsilon(M-L)-A\,{}_2^1(\Delta(M-L))\right]=0,\\
&\text{(e)}\ 2U^{-2}(1-v^2)^{-1/2}\left[\Psi_0U^2\sqrt{1-v^2}\right]'+2ik_3H\phi_0-\frac{\Psi_1U\mathbf{X}^+}{\sqrt{1-v^2}}=0,\qquad\text{(B)}\\
&\text{(f)}\ 2U^{-2}(1-v^2)^{-1/2}[\Psi_1U(1-v^2)]'+H\phi_0\mathbf{X}^--\Psi_2\mathbf{X}^+=0,\\
&\text{(g)}\ -2(1-v^2)^{-3/2}[(1-v^2)^{3/2}\Psi_2]'-2ik_3H(\phi_2^{(1)}+\phi_2^{(1)*})+H(\ln\chi)'\phi_0^*\\
&\quad+\frac{U}{\sqrt{1-v^2}}(\Psi_3+H\phi_1^*)\mathbf{X}^++U\sqrt{2}\left[H^2(\rho^*-2\epsilon-2\epsilon^*)-A\,{}_2^1(\Delta H^2)+\sigma(M-L)\right]=0,\\
&\text{(h)}\ \frac{2}{1-v^2}[(1-v^2)HU^{-1}\phi_1^*]'-\frac{2}{(1-v^2)^2}[(1-v^2)^2\Psi_3U^{-1}]'+(4\beta-\tau)(M-L)\sqrt{2}\\
&\quad+(\tau^*-2\alpha-2\beta^*)H^2\sqrt{2}+\sqrt{2}A\,{}_2^3(\Delta(M-L))\\
&\quad-\sqrt{2}A\,{}_2^4(\Delta H^2)-2HU^{-1}\phi_1(\ln\chi)'+\frac{1}{\sqrt{1-v^2}}[\Psi_4^{(1)}\chi^+-H(\phi_2^{(1)}+\phi_2^{(1)*})\chi^-]=0,
\end{aligned}
$$

where $M-L=\Psi_4$ (for the background) is given by Eq. (2.9);

Linearized Maxwell equations,

$$
\begin{aligned}
&\text{(a)}\ 2ik_3\phi_1-\frac{U\phi_0\mathbf{X}^-}{\sqrt{1-v^2}}+HU\kappa\sqrt{2}=0,\\
&\text{(b)}\ 2ik_3\phi_2^{(1)}+(2\epsilon-\rho)HU\sqrt{2}-\phi_0(\ln\chi)'+UA\,{}_2^1(\Delta H)\sqrt{2}-\frac{\phi_1U\mathbf{X}^-}{\sqrt{1-v^2}}=0,\\
&\text{(c)}\ -2U^{-2}(\phi_0U\sqrt{1-v^2})'+\phi_1X^++\sigma H\sqrt{2}\sqrt{1-v^2}=0,\\
&\text{(d)}\ -\frac{2}{U(1-v^2)}[(1-v^2)\phi_1]'+\frac{\phi_2^{(1)}\mathbf{X}^+}{\sqrt{1-v^2}}+(2\beta-\tau)H\sqrt{2}+A\,{}_2^3(\Delta H)\sqrt{2}=0.
\end{aligned}
\qquad\text{(M)}
$$

Recall that λ,μ, and γ are the only nonvanishing spin coefficients in the background, given by Eqs. (2.7) (note they are all real!). By using Eqs. (109) from M. T., p. 449, we find that the only nonvanishing structure constants $C\,{}_k^{ij}$ in the background are

$$
C\,{}_3^{32}=C\,{}_4^{42}=-\frac{\sqrt{2}}{U}\frac{U}{1-v^2},\quad C\,{}_1^{21}=-\frac{U'\sqrt{2}}{U^2},\quad C\,{}_4^{32}=C\,{}_3^{42}=-\frac{1}{U\sqrt{2}}(\ln\chi)',\qquad(4.4)
$$

together with those obtained from the antisymmetry of $C\,{}_k^{ij}$ in the two upper indices. As in M. T., the different equations will be characterized by the triplet of indices $\binom{ij}{k}$. When pairs of complex conjugate equations arise, only one equation will be included. We obtain for the

linearized commutation relations,

$$
\begin{aligned}
&\binom{21}{3},\quad\text{(a)}\quad 2U^{-1}\sqrt{1-v^2}\left(\frac{UA\,{}_3^1}{\sqrt{1-v^2}}\right)'+A\,{}_4^1(\ln\chi)'-2ik_3A\,{}_3^2+(\tau^*+\pi)U\sqrt{2}=0,\\
&\binom{43}{1},\quad\text{(b)}\quad \mathbf{X}^-A\,{}_1^3-\mathbf{X}^+A\,{}_1^4=(\mu^{(1)*}-\mu^{(1)})\sqrt{2}\sqrt{1-v^2},\\
&\binom{43}{2},\quad\text{(c)}\quad \mathbf{X}^-A\,{}_2^3-\mathbf{X}^+A\,{}_2^4=(\rho^*-\rho)\sqrt{2}\sqrt{1-v^2},\\
&\binom{31}{2},\quad\text{(d)}\quad X^+A\,{}_2^1-2U^{-1}\sqrt{1-v^2}ik_3A\,{}_2^3=\kappa\sqrt{2}\sqrt{1-v^2},\\
&\binom{31}{4},\quad\text{(e)}\quad U\mathbf{X}^+A\,{}_4^1-2ik_3A\,{}_4^3\sqrt{1-v^2}=A\,{}_2^1(\ln\chi)'\sqrt{1-v^2}-\sigma U\sqrt{2}\sqrt{1-v^2},\\
&\binom{32}{1},\quad\text{(f)}\quad 2U(A\,{}_1^3U^{-1}\sqrt{1-v^2})'=U\mathbf{X}^+A\,{}_1^2+A\,{}_1^4(\ln\chi)'\sqrt{1-v^2}+\nu^*U\sqrt{2}\sqrt{1-v^2},\\
&\binom{32}{4},\quad\text{(g)}\quad 2(A\,{}_4^3)'+(A\,{}_2^2+A\,{}_3^3-A\,{}_4^4)(\ln\chi)'-\frac{U\mathbf{X}^+A\,{}_4^2}{\sqrt{1-v^2}}+\lambda^{(1)*}U\sqrt{2}=0,\\
&\binom{21}{1},\quad\text{(h)}\quad (A\,{}_1^1)'-ik_3A\,{}_1^2-A\,{}_2^2(\ln U)'=\frac{U}{\sqrt{2}}(\gamma^{(1)}+\gamma^{(1)*}),\\
&\binom{21}{2},\quad\text{(i)}\quad U^{-1}(UA\,{}_2^1)'-ik_3A\,{}_2^2=\frac{U}{\sqrt{2}}(\epsilon+\epsilon^*),
\end{aligned}
$$

$$
\begin{aligned}
&\binom{43}{3}, \quad \text{(j)} \quad && \frac{U}{\sqrt{1-v^2}}\left(X^- A^3_3 - X^+ A^4_3\right) = A^3_2 (\ln\chi)' - \frac{2v}{1-v^2} A^4_2 + (\alpha - \beta^*) U\sqrt{2}\,, \qquad \text{(C)}\\
&\binom{31}{1} + \binom{32}{2}, \quad \text{(k)} \quad && -\frac{2}{U\sqrt{1-v^2}}\left(A^3_2 U\sqrt{1-v^2}\right)' + A^4_2 (\ln\chi)' - 2ik_3 A^3_1 + \frac{UX^+}{\sqrt{1-v^2}}\left(A^1_1 + A^2_2\right)\\
&&& \quad = (\tau - \pi^*) U\sqrt{2}\,,\\
&\binom{31}{3} + \binom{41}{4}, \quad \text{(l)} \quad && \frac{U}{\sqrt{1-v^2}}\left(A^1_3 X^+ + A^1_4 X^-\right) - 2ik_3\left(A^3_3 + A^4_4\right) + (\rho + \rho^*) U\sqrt{2} = \frac{4v}{1-v^2} A^1_2\,,\\
&\binom{32}{3} + \binom{42}{4}, \quad \text{(m)} \quad && \frac{U}{\sqrt{1-v^2}}\left(A^2_3 X^+ + A^2_4 X^-\right) - 2\left(A^3_3 + A^4_4\right)' - \frac{4v}{1-v^2} A^2_2 = \left(\mu^{(1)} + \mu^{(1)*}\right) U\sqrt{2}\,,\\
&\binom{31}{1} - \binom{32}{2}, \quad \text{(n)} \quad && \frac{2U}{\sqrt{1-v^2}}\left(A^3_2 U^{-1}\sqrt{1-v^2}\right)' + \frac{UX^+}{\sqrt{1-v^2}}\left(A^1_1 - A^2_2\right) - 2ik_3 A^3_1 - A^4_2 (\ln\chi)'\\
&&& \quad = [2(\alpha^* + \beta) - (\tau + \pi^*)] U\sqrt{2}\,,\\
&\binom{41}{4} - \binom{31}{3}, \quad \text{(o)} \quad && \frac{U}{\sqrt{1-v^2}}\left(A^1_4 X^- - A^1_3 X^+\right) + 2ik_3\left(A^3_3 - A^4_4\right) = [2(\epsilon - \epsilon^*) - (\rho - \rho^*)] U\sqrt{2}\,,\\
&\binom{42}{4} - \binom{32}{3}, \quad \text{(p)} \quad && \frac{U}{\sqrt{1-v^2}}\left(A^2_4 X^- - A^2_3 X^+\right) + 2\left(A^3_3 - A^4_4\right)' + 2\left(A^3_4 - A^4_3\right)(\ln\chi)'\\
&&& \quad = [2(\gamma^{(1)} - \gamma^{(1)*}) - (\mu^{(1)} - \mu^{(1)*})] U\sqrt{2}\,.
\end{aligned}
$$

In the following, the different linearized equations will be characterized by (R.), (B.), (M.), and (C.) standing, respectively, for the Ricci, the Bianchi, the Maxwell, and the commutation relation equations, where after the period a latin letter specifies the particular equation.

V. GAUGE CONDITIONS

In investigations of perturbations via the Newman–Penrose formalism one must be aware of, and take advantage of, the ten degrees of the available gauge freedom. The six (real) degrees of tetrad gauge freedom have their origin in the null tetrad rotations of types I, II, and III, which preserve the Newman–Penrose equations (see M.T., Sec. 8g, p. 53). On the other hand, the four (real) degrees of the coordinate gauge arise from the freedom in identifying the points of the unperturbed and the perturbed manifolds.

The (finite) changes of the Newman–Penrose scalars under the tetrad rotations are given in M.T., Sec. 8g, Eqs. (342)–(347). The rotations are "measured" by the complex scalars a and b and the real scalars A and θ. The absence of any rotation is given by the choice

$$a = b = \theta = 0, \quad A = 1\,. \tag{5.1}$$

We shall be interested in infinitesimal tetrad rotations. In accordance with the rotation conventions adopted in Sec. III, we shall describe by a, b, θ, and $\mathscr{A}$ the linearized changes of a, b, θ, and A, respectively, where $A = 1 + \mathscr{A}$. Next we linearize the relevant equations [M.T., Sec. 8g, Eqs. (343), (345), and (347)] around the "background" values (5.1) and use that the only nonvanishing spin coefficients in the unperturbed space-time are λ, μ, and γ. We find that the only spin coefficients subject to nonzero changes are, for type I,

$$
\begin{aligned}
&\pi \to \pi + Da^*\,,\\
&\lambda^{(1)} \to \lambda^{(1)} + \delta^* a^*\,,\\
&\mu^{(1)} \to \mu^{(1)} + \delta a^*\,,\\
&\nu \to \nu + a\lambda + a^*(\mu + 2\gamma) + \Delta a^*\,;
\end{aligned}
$$

for type II,

$$
\begin{aligned}
&\pi \to \pi + b^*\mu + b\lambda\,,\\
&\tau \to \tau + 2b\gamma - \Delta b\,,\\
&\beta \to \beta + b(\mu + \gamma)\,,\\
&\alpha \to \alpha + b^*\gamma + b\lambda\,,\\
&\sigma \to \sigma - \delta b\,,\\
&\rho \to \rho - \delta^* b\,,\\
&\kappa \to \kappa - Db\,;
\end{aligned}
\tag{5.2}
$$

and for type III,

$$
\begin{aligned}
&\lambda^{(1)} \to \lambda^{(1)} + (\mathscr{A} - 2i\theta)\lambda\,,\\
&\mu^{(1)} \to \mu^{(1)} + \mathscr{A}\mu\,,\\
&\gamma^{(1)} \to \gamma^{(1)} + \mathscr{A}\gamma - \tfrac{1}{2}\Delta\mathscr{A} + (i/2)\Delta\theta\,,\\
&\epsilon \to \epsilon - \tfrac{1}{2} D\mathscr{A} + (i/2) D\theta\,,\\
&\alpha \to \alpha + (i/2)\delta^*\theta - \tfrac{1}{2}\delta^*\mathscr{A}\,,\\
&\beta \to \beta + (i/2)\delta\theta - \tfrac{1}{2}\delta\mathscr{A}\,.
\end{aligned}
$$

(The spin coefficients not included in the above transformations remain invariant to the respective infinitesimal tetrad rotations.)

In the next section we shall be considering the u-independent perturbations. Therefore, the available tetrad rotations should be restricted to be independent of u as well:

$$D a = D b = D\mathscr{A} = D\theta = 0\,. \tag{5.3}$$

Hence, in addition to the spin coefficients omitted from Eqs. (5.2), now π, κ, and ϵ also remain invariant to rotations of types I, II, and III, respectively.

In considering which gauge conditions to impose on the

perturbations of the spin coefficients, our guiding principle is to seek conditions that can be imposed independently of the order in which the different rotations are performed. Such conditions take advantage of one of the rotations and they are not affected by the other two rotations. In particular, we shall impose the tetrad gauge conditions

$$\nu = 0, \quad \rho = 0, \quad \gamma^{(1)} = 0, \tag{5.4}$$

by making rotations of types I, II, and III, respectively. (The only other possible choices could have been $\tau = 0$ or $\sigma = 0$, instead of $\rho = 0$, again by suitable rotations of type II.)

VI. THE *u*-INDEPENDENT PERTURBATIONS

In this section we shall establish the main result of the paper; that is, there are no nontrivial u-independent perturbations of the metric (2.1) that do not respect the plane symmetry.

That the perturbations are u independent means that

$$k_3 = 0 \tag{6.1}$$

in the equations of Sec. IV. That the perturbations are not plane symmetric means that at least one of k_1 and k_2 in Eq. (3.5) is nonzero:

$$k_1^2 + k_2^2 \neq 0. \tag{6.2}$$

In addition, as we have explained in Sec. V, we shall impose the gauge conditions

$$\nu = \rho = \gamma^{(1)} = 0. \tag{6.3}$$

Our objective is to solve the system of equations (R), (B), (M), and (C) of Sec. IV, subject to the conditions (6.1)–(6.3).

The first step in the solution is easy. By using successively Eqs. (R.a), (R.b), (R.c), (R.d), (R.k), (M.a), (M.c), (R.p), (R.q), (R.f), (R.g), (R.i), and (C.e) we obtain, in this order, that

$$\kappa = 0, \quad \Psi_0 = 0, \quad \Psi_1 = 0, \quad \epsilon = 0, \quad \sigma = 0, \quad \phi_0 = 0,$$
$$\phi_1 = 0, \quad \tau = 0, \quad \Psi_2 = 0, \quad A_2^1 = 0, \quad \pi = 0,$$
$$\Psi_3 = 0, \quad A_4^1 = 0 \quad (\text{hence also } A_3^1 = 0),$$

while Eqs. (R.e), (B.a), (B.b), (B.e), (R.h), (B.c), (B.d), (B.f), (B.g), (M.b), (C.a), (C.d), (C.i), (C.l), and (C.o) are identically satisfied. The easy step has revealed, therefore, that

$$\nu = \rho = \gamma^{(1)} = \kappa = \epsilon = \sigma = \tau = \pi = 0,$$
$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0, \quad \phi_0 = \phi_1 = 0, \tag{6.4}$$
$$A_2^1 = A_3^1 = A_4^1 = 0.$$

Next, we proceed with the more involved part of the solution in which, gradually, we shall be imposing additional gauge conditions compatibly with the four (real) degrees of coordinate gauge freedom.

A coordinate gauge invariant treatment of the problem would involve solving the remaining of Eqs. (R), (B), (M), and (C) under the conditions (6.4). The solution is expected to involve four (real) arbitrary functions, which could be consistently chosen to be zero. The absence of nontrivial u-independent perturbations then would mean that all the perturbations should vanish when the four arbitrary functions are chosen to be zero. Since, however, the system of equations under consideration remains formidable even after the conditions (6.4), gradually we shall be imposing coordinate gauge conditions compatibly with the equations of the system.

First, we shall impose

$$A_4^3 = A_3^4 = 0 \tag{6.5}$$

and

$$A_2^3 = A_2^4 = 0. \tag{6.6}$$

Since A_4^3 is complex (and A_3^4 is its complex conjugate), the condition (6.5) takes care of two degrees of gauge freedom. On the other hand, Eq. (C.c) gives

$$(\mathrm{X}^+ A_2^4)^* = \mathrm{X}^+ A_2^4, \tag{6.7}$$

which implies that A_2^4 is real and, therefore, the condition (6.6) takes care of only one degree of gauge freedom.

Equation (R.l) gives $\alpha \mathrm{X}^+ = \beta \mathrm{X}^-$ from which we obtain that

$$\begin{aligned} (\alpha^* + \beta)\mathrm{X}^- &= (\alpha + \beta^*)\mathrm{X}^+, \\ (\alpha^* - \beta)\mathrm{X}^- &= -(\alpha - \beta^*)\mathrm{X}^+. \end{aligned} \tag{6.8}$$

By adding and subtracting Eq. (R.o) to the complex conjugate of Eq. (R.r) we obtain

$$\begin{aligned} &2(\alpha^* - \beta)' - [2v/(1-v^2)](\alpha^* - \beta) \\ &\quad = -(\alpha - \beta^*)(\ln \chi)', \\ &2(\alpha^* + \beta)' + 2(\alpha^* + \beta)(\ln U)' \\ &\quad - [2v/(1-v^2)](\alpha^* + \beta) = (\alpha + \beta^*)(\ln \chi)'. \end{aligned} \tag{6.9}$$

By combining Eqs. (6.8) and (6.9), and using that

$$(\ln \mathrm{X}^\pm)' = (\mathrm{X}^\mp / 2\mathrm{X}^\pm)(\ln \chi)', \tag{6.10}$$

we find that Eqs. (6.9) can be integrated. They give

$$\alpha^* + \beta = \mathrm{B}\mathrm{X}^+ / U\sqrt{1-v^2}, \quad \alpha^* - \beta = i\Gamma \mathrm{X}^+ / \sqrt{1-v^2}, \tag{6.11}$$

where B and Γ are integration constants. In fact, by comparing Eqs. (6.8) and (6.11) we conclude that B *and* Γ *are real constants*. Hence

$$\begin{aligned} \alpha &= \frac{\mathrm{X}^-}{2U\sqrt{1-v^2}}(\mathrm{B} - i\Gamma U), \\ \beta &= \frac{\mathrm{X}^+}{2U\sqrt{1-v^2}}(\mathrm{B} - i\Gamma U). \end{aligned} \tag{6.12}$$

We should remark that with the gauge conditions (6.5) and (6.6) and the solutions (6.12) we have taken care of Eqs. (C.c), (R.l), (R.o), and (R.r).

To proceed further, we shall use at this point the residual (tetrad) gauge freedom that remains after the imposition of the gauge conditions

$$\nu = \rho = \gamma^{(1)} = 0, \tag{6.13}$$

imposed by making tetrad rotations of types I, II, and III, respectively. We shall explain the ideas in some detail for rotations of type III.

With a view to the transformations (5.2) we wonder whether we can make, for instance, α and β vanish by the allowed rotations of type III, i.e., by those rotations that preserve the solution obtained up to the present stage. Since

$\lambda^{(1)}$ and $\mu^{(1)}$ have not been determined so far, they remain outside of these considerations.

The vanishing of $\gamma^{(1)}$ is preserved by those infinitesimal rotations of type III that satisfy

$$\mathscr{A}\gamma - \tfrac{1}{2}\Delta\mathscr{A} + (i/2)\Delta\theta = 0\,. \tag{6.14}$$

By using Eqs. (2.7) and (3.6), this condition reads

$$\frac{\partial}{\partial v}(\mathscr{A} - i\theta) = -\mathscr{A}(\ln U)'\,. \tag{6.15}$$

The vanishing of ϵ is automatically preserved (the rotations considered respect the plane symmetry!) while the requirement that α or β vanish leads to the condition

$$\mathscr{A} - i\theta = (\sqrt{2}/U)(\mathrm{B} - i\Gamma U)\,. \tag{6.16}$$

It is remarkable that Eqs. (6.15) and (6.16) admit the solution

$$\mathscr{A} = \mathrm{B}\sqrt{2}/U, \quad \theta = \Gamma\sqrt{2}\,, \tag{6.17}$$

for which α and β are made zero,

$$\alpha = \beta = 0\,, \tag{6.18}$$

while preserving the vanishing of ϵ and $\gamma^{(1)}$. We adopt this choice.

From Eqs. (M.d) and (B.h) we next obtain that

$$\phi_2^{(1)} = 0, \quad \Psi_4^{(1)} = 0\,. \tag{6.19}$$

The fact that [see Eqs. (6.4) and (6.19)] we have established that all the perturbations of the curvature scalars vanish to first order is a good enough argument for the u-independent perturbations to be neutral to first order. However, we shall complete the proof that all the perturbations are trivial by considering the remaining equations and using the remaining gauge freedom.

From Eqs. (C.j), (C.k), and (C.n) we obtain that

$$A_1^1 = A_2^2 = A_3^3 = A_4^4\,, \tag{6.20}$$

while Eq. (C.h) is identically satisfied.

Equations (C.m) and (C.p) give (and they are equivalent to)

$$\mu^{(1)} = A_3^2\mathrm{X}^+/\sqrt{2}\sqrt{1-v^2}\,; \tag{6.21}$$

Eq. (C.g) gives

$$\lambda^{(1)} = A_3^2\mathrm{X}^-/\sqrt{2}\sqrt{1-v^2}\,. \tag{6.22}$$

Using the solutions (6.21) and (6.22), Eqs. (R.j) and (R.n) lead to

$$(U^{-1}\sqrt{1-v^2}\,A_3^2)'\mathrm{X}^+ = U^{-1}\sqrt{1-v^2}\,A_4^2(\mathrm{X}^-)' \tag{6.23}$$

and

$$(U^{-1}\sqrt{1-v^2}\,A_3^2)'\mathrm{X}^- = U^{-1}\sqrt{1-v^2}\,A_4^2(\mathrm{X}^+)'\,, \tag{6.24}$$

respectively.

Next we take advantage of the residual gauge freedom of rotations of type I [Eqs. (5.2)]. We readily conclude that the vanishing of π is automatically preserved; that $\mu^{(1)}$ can be made zero by performing a rotation with

$$a^* = -A_3^2\,; \tag{6.25}$$

that the choice (6.25) leads to the vanishing of $\lambda^{(1)}$, compatibly with Eqs. (6.21) and (6.22); and that the preservation of the vanishing of ν leads, after some computations, to the condition (6.23).

The upshot of these conclusions is that by using the residual gauge freedom we can perform a type I rotation with $a = -A_4^2$ and have

$$\mu^{(1)} = \lambda^{(1)} = 0 \tag{6.26}$$

and, therefore,

$$A_3^2 = A_4^2 = 0\,. \tag{6.27}$$

There are only two equations that have not been considered so far, Eqs. (C.b) and (C.f). By virtue of Eq. (6.26), Eq. (C.b) reads

$$(A_1^3\mathrm{X}^-)^* = A_1^3\mathrm{X}^-\,. \tag{6.28}$$

Equation (6.28) shows that we can choose

$$A_1^3 = A_1^4 = 0 \tag{6.29}$$

by using one degree of gauge freedom, the last remaining degree of coordinate gauge freedom. Equation (C.f) then is used to show that $A_1^2 = 0$; and with this we have established the neutrality of the space-time (2.1) to the u-independent first-order perturbations.

We would like to point that, contrary to the type I and III rotations, there is no residual gauge freedom of type II rotations. For instance, the preservation of the vanishing of σ under type II rotations immediately gives [Eq. (5.2)] that $b = 0$. This remark shows that in order to establish the neutrality of the u-independent perturbations (and the vanishing of the perturbations of all spin coefficients, Weyl and Maxwell scalars, and the null tetrad vectors) we have to use every bit of the available gauge freedom and all 84 Newman–Penrose equations!

VII. CONCLUDING REMARKS

The conclusion of the present paper took us by surprise, to say the least: Consider the most general metric for plane polarized plane gravitational and electromagnetic waves (the family depends on two arbitrary real functions). It is a pure radiative metric, of Petrov type N, and it depends only on retarded time v. Conside its linearized perturbations, which do not respect plane symmetry but do respect causality, i.e., they do not depend on advanced time, but only on retarded time. In other words, we consider perturbations that destroy the planar symmetry but with no mixing of incoming and outgoing waves. In general relativity, this situation would not be accepted and the only linear perturbations compatible with the field equations are the trivial, pure gauge perturbations. It should be clarified at this point that nowhere in our analysis have we demanded explicitly any regularity conditions from the perturbations. The only requirement on the perturbations (in addition to the assumption that they do not depend on advanced time) is that their dependence on the two spatial directions (spanning the wave fronts of the waves) can be Fourier analyzed. The convergence of the Fourier integrals, of course, does demand implicitly a certain asymptotic falloff and regularity of the perturbations.

What does general relativity tell us? Shall we be careful

in using normal mode analysis in perturbation theory? Is the set of normal modes considered incomplete, and are physically relevant perturbations lost? Are the perturbations of plane waves only a second-order effect? Has our surprising conclusion anything to do with the fact that plane waves are an idealization, carrying infinite energy? Was the restriction to consider only plane polarized plane waves too severe and can the conclusion, therefore, not be generalized to the general plane waves? Or, is there another "no hair" theorem (like the black holes or the cosmological no hair theorems) waiting its discovery?

Note added in proof: In collaboration with B. G. Schmidt, we have constructed a nontrivial, u-dependent vacuum linearized perturbation of plane gravitational waves. The perturbation is not squared integrable! It appears, therefore, that the absence of purely incoming perturbations is due to the assumption that their u-dependence can be analyzed in normal modes. The implications to the customary way of studying perturbations should be obvious. In an alternative investigation, we have been able to determine the general solution of the u-dependent perturbations of Einstein–Maxwell plane polarized plane waves. These results will be reported in subsequent communications.

ACKNOWLEDGMENTS

This work was initially started as a collaboration with Dr. V. Ferrari, of the University of Rome. The author would like to thank her, as well as Professor S. Chandrasekhar of the University of Chicago, for useful discussions.

APPENDIX A: ALTERNATIVE DESCRIPTIONS OF PLANE POLARIZED PLANE WAVES

The form of the metric

$$ds^2 = U^2(v)(du)(dv) - (1-v^2)[\Psi^{-1}(v)(dx)^2 + \Psi(v)(dy)^2], \quad \text{(A1)}$$

whose linear perturbations we have studied, is isometric with the most common expression

$$ds^2 = (du)(dv_1) - F^2(v_1)(dx)^2 - G^2(v_1)(dy)^2, \quad \text{(A2)}$$

usually used for the description of plane polarized plane waves. The identification is obvious,

$$F(v_1) = [(1-v^2)\Psi^{-1}(v)]^{1/2}, \quad dv_1 = U^2(v)dv,$$
$$G(v_1) = [(1-v^2)\Psi(v)]^{1/2}. \quad \text{(A3)}$$

Then

$$\frac{\ddot{F}}{F} + \frac{\ddot{G}}{G} = U^{-4}\left[-\frac{2}{(1-v^2)^2} + \frac{4v}{1-v^2}\frac{U'}{U} + \frac{(\Psi')^2}{2\Psi^2}\right], \quad \text{(A4)}$$

where the dot and the prime denote differentiations with respect to v_1 and v, respectively. Thus the metric (A2) is

(a) a vacuum solution when $\ddot{F}/F + \ddot{G}/G = 0$;

(b) a solution of the Einstein–Maxwell electrovacuum equations or a solution of the Einstein equations coupled with null dust when $\ddot{F}/F + \ddot{G}/G < 0$.

(c) By the Einstein equations, the associated energy-momentum tensor would not satisfy any physical energy condition when $\ddot{F}/F + \ddot{G}/G > 0$.

For the metric (A2) the only nonvanishing (tensor) components of the Riemann tensor are

$$R_{vxvx} = F\ddot{F}, \quad R_{vyvy} = G\ddot{G}, \quad \text{(A5)}$$

of the Weyl tensor are

$$C_{vxvx} = -\frac{F}{2G}(F\ddot{G} - G\ddot{F}), \quad C_{vyvy} = \frac{G}{2F}(F\ddot{G} - G\ddot{F}), \quad \text{(A6)}$$

of the Ricci tensor is

$$R_{vv} = -\left(\frac{\ddot{F}}{F} + \frac{\ddot{G}}{G}\right), \quad \text{(A7)}$$

while the scalar curvature vanishes identically.

The essential difference between expressions (A1) and (A2) for the line element is that the vacuum Einstein equations lead to *first-order* differential equations for the metric (A1) and to *second-order* equations for the metric (A2).

By making the coordinate transformation

$$x = \frac{X}{F(v_1)}, \quad y = \frac{Y}{G(v_1)},$$
$$u = \tilde{U} - X^2\frac{\dot{F}}{F} - Y^2\frac{\dot{G}}{G}, \quad v_1 = V, \quad \text{(A8)}$$

the metric (A2) becomes

$$ds^2 = (d\tilde{U})(dV) - (dX)^2 - (dY)^2 - \left(X^2\frac{\ddot{F}}{F} + Y^2\frac{\ddot{G}}{G}\right)(dV)^2. \quad \text{(A9)}$$

In the vacuum case we set

$$\ddot{F}/F = -\ddot{G}/G = h(V) \quad \text{(A10)}$$

and the metric (A9) becomes

$$ds^2 = (d\tilde{U})(dV) - (dX)^2 - (dY)^2 + h(V)(Y^2 - X^2)(dV)^2; \quad \text{(A11)}$$

this is the standard form (Ref. 12) for plane polarized plane waves with amplitude $h(V)$.

APPENDIX B: GAUGE INVARIANT PERTURBATIONS OF THE BELL–SZEKERES SPACE-TIME

In this appendix we give the complete *gauge invariant* solution of the u-independent perturbations of the plane wave solution of the Einstein–Maxwell equations for the Bell–Szekeres space-time. The background metric

$$ds^2 = 4(du)(dv)/\sqrt{1-v^2} - (1-v^2)[(dx^1)^2 + (dx^2)^2] \quad \text{(B1)}$$

is a particular case of the metric (2.1) considered in the main body of the paper. It corresponds to the choices

$$U = 2(1-v^2)^{-1/4}, \quad \chi = 1, \quad \text{(B2)}$$

and the relevant perturbation equations can be readily obtained from Eqs. (R), (B), (M), and (C) of Sec. IV by making these choices. Here we present the general solution of the resulting equations *without* imposing *any gauge condition*. The motivation for performing the massive reductions involved and for including the result in an appendix was to

understand how the ten degrees of gauge freedom would appear explicitly in the solution. It is remarkable that they do not appear (as we were expecting) as five complex free functions but as six complex functions subject to two reality conditions!

From Eqs. (2.8), (2.9), and (4.1) we find that

$$M-L=0,\ M+L=-1/2\sqrt{1-v^2},\qquad H=1/\sqrt{2}(1-v^2)^{1/4}; \tag{B3}$$

thus, in particular, the space-time (B1) is conformally flat.

We leave unspecified the spin coefficients perturbations β, ρ, and $\lambda^{(1)}$. In addition, we set

$$A^3_4=F,\quad A^4_3=F^*,\quad A^3_1=G,\qquad A^4_1=G^*,\quad A^3_2=J,\quad A^4_2=J^*, \tag{B4}$$

where F, G, and J are arbitrary complex functions of v. The obtained solution of the perturbation equations is

$$\Psi_0=\Psi_1=\Psi_2=\Psi_3=\Psi_4^{(1)}=0,\quad \phi_0=0,\quad \kappa=\epsilon=0,$$

$$\sigma=\frac{ik_1+k_2}{ik_1-k_2}\rho,\quad \pi=\frac{v}{(1-v^2)^{1/4}}\frac{\rho^*}{ik_1+k_2},\quad \tau=\frac{\sqrt{1-v^2}}{ik_1-k_2}[(1-v^2)^{1/4}\rho]',$$

$$\alpha=\frac{ik_1-k_2}{ik_1+k_2}\beta-\frac{v(5\rho-\rho^*)}{4(1-v^2)^{1/4}(ik_1+k_2)},\quad \nu=\frac{[(1-v^2)^{5/4}\lambda^{(1)}]'}{(ik_1-k_2)\sqrt{1-v^2}},$$

$$\mu^{(1)}=\frac{ik_1+k_2}{ik_1-k_2}\lambda^{(1)}+\frac{2+v^2}{2\sqrt{2}(1-v^2)}\frac{J^*}{ik_1-k_2}+\frac{v}{(1-v^2)^{1/4}}\left[\frac{\beta}{ik_1+k_2}+\frac{\beta^*}{ik_1-k_2}\right]+\frac{5v^2\rho+(3v^2-4)\rho^*}{4(k_1^2+k_2^2)\sqrt{1-v^2}},$$

$$\gamma^{(1)}=\frac{(1-v^2)^{3/4}}{ik_1+k_2}\beta'+\frac{5v\sqrt{1-v^2}}{4(k_1^2+k_2^2)}\rho'+\frac{v^2(5\rho^*-27\rho)+16\rho}{16(k_1^2+k_2^2)\sqrt{1-v^2}}$$
$$-\frac{v}{4(1-v^2)^{1/4}}\left[\frac{3\beta}{ik_1+k_2}-\frac{\beta^*}{ik_1-k_2}\right]+\frac{(v^2+2)}{8\sqrt{2}(1-v^2)}\frac{J}{ik_1+k_2},$$

$$\phi_1=-\frac{(1-v^2)^{1/4}\rho}{ik_1-k_2},\quad \phi_2^{(1)}=-\frac{2v\rho}{k_1^2+k_2^2}-\frac{1}{(ik_1+k_2)}\left[\frac{vJ}{2\sqrt{2}\sqrt{1-v^2}}+2\beta(1-v^2)^{1/4}\right],$$

$$A^1_1=\sqrt{2}\sqrt{1-v^2}\left[\frac{\beta}{ik_1+k_2}+\frac{\beta^*}{ik_1-k_2}\right]+\frac{v\sqrt{2}(1-v^2)^{1/4}}{k_1^2+k_2^2}(\rho+\rho^*)$$
$$+\frac{v}{4(1-v^2)^{1/4}}\left[\frac{J}{ik_1+k_2}+\frac{J^*}{ik_1-k_2}\right],\quad A^1_2=0,$$

$$A^1_3=-\frac{\sqrt{2}\sqrt{1-v^2}}{ik_1+k_2}\rho^*,\quad A^2_1=\left[\frac{(1-v^2)^{3/4}}{ik_1+k_2}G-\frac{\sqrt{2}(1-v^2)^{5/4}}{(ik_1+k_2)^2}\lambda^{(1)*}\right]',$$

$$A^2_2=-\sqrt{2}\sqrt{1-v^2}\left[\frac{\beta}{ik_1+k_2}+\frac{\beta^*}{ik_1-k_2}\right]+\frac{1}{2}(1-v^2)^{1/4}\left\{\sqrt{1-v^2}\left[\frac{J}{ik_1+k_2}+\frac{J^*}{ik_1-k_2}\right]\right\}'$$
$$-\frac{(1-v^2)^{3/2}}{\sqrt{2}(k_1^2+k_2^2)}\left[\frac{(\rho+\rho^*)}{(1-v^2)^{1/4}}\right]',$$

$$A^2_3=\frac{\sqrt{1-v^2}}{ik_1-k_2}[\sqrt{2}\lambda^{(1)}+(1-v^2)^{1/4}F^{*\prime}],$$

$$A^3_3=\frac{ik_1+k_2}{ik_1-k_2}F^*-\frac{v}{(1-v^2)^{1/4}}\frac{J^*}{ik_1-k_2}+\frac{v(1-v^2)^{1/4}(5\rho-\rho^*)}{2\sqrt{2}(k_1^2+k_2^2)}+\sqrt{2}\sqrt{1-v^2}\left[\frac{\beta}{ik_1+k_2}-\frac{\beta^*}{ik_1-k_2}\right], \tag{B5}$$

provided that

$$\frac{J}{ik_1+k_2}-\frac{J^*}{ik_1-k_2}=\frac{\sqrt{2}\sqrt{1-v^2}(\rho-\rho^*)}{k_1^2+k_2^2} \tag{B6}$$

and

$$\frac{G}{ik_1+k_2}-\frac{G^*}{ik_1-k_2}=\sqrt{2}\sqrt{1-v^2}\left[\frac{\lambda^{(1)*}}{(ik_1+k_2)^2}-\frac{\lambda^{(1)}}{(ik_1-k_2)^2}\right]. \tag{B7}$$

[1] S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).

[2] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985).

[3] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **402**, 37 (1985).

[4] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **408**, 175 (1986).

[5]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **410**, 311 (1987).
[6]T. Papacostas and B. C. Xanthopoulos, J. Math. Phys. **30**, 97 (1989).
[7]U. Yurtsever, Phys. Rev. D **36**, 1662 (1987).
[8]U. Yurtsever, Phys. Rev. D **37**, 2803 (1988); **38**, 1706 (1988); **38**, 1731 (1988).
[9]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **420**, 93 (1988).
[10]P. Bell and P. Szekeres, Gen. Relativ. Gravit. **5**, 275 (1974).
[11]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[12]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).

A possible generalization of the concept of symmetry in analytical mechanics

D. R. Grigore
Department of Theoretical Physics, Central Institute of Physics, P. O. Box MG-6, Bucharest, Romania

(Received 17 November 1987; accepted for publication 24 May 1989)

A theorem of Chung [Proc. R. Soc. Edinburgh, Sec. A **LXII**, 237 (1947)] suggests a possible generalization of the symmetry concepts in classical mechanics. It is shown that the theory of Kostant–Souriau–Kirillov can be adapted easily to this more general case. The theory is illustrated with a number of examples.

I. INTRODUCTION

According to a theorem of Lee Hwa Chung (see e.g., Refs 1–3) the natural mathematical objects that describe the symmetries of the symplectic manifold (X, Ω) (Ω is the symplectic form of X), are the diffemorphisms $\varphi: X \to X$ which verify the condition:

$$\varphi^* \Omega = a\,\Omega. \tag{1.1}$$

Here $\varphi^*\,\Omega$ is the pullback of Ω, and $a \in \mathbb{R} \setminus \{0\}$. In fact, this theorem asserts that if φ sends Hamiltonian vector fields into Hamiltonian vector fields, then φ verifies (1.1). The hypothesis of the theorem is a natural definition for the concept of symmetry. Moreover an analog reasoning is valid also in quantum mechanics (see Ref. 4, Vol. I, Sec. 8 and Vol. II, Sec. 3): One can say that the Chung theorem is the analog of the Wigner theorem in quantum mechanics, as remarked in Ref. 5.

Usually, one considers only transformations with $a = \pm 1$. If $X = \mathbb{R}^{2n}$, this can be justified as follows.[5] Let $x_1 \ldots, x_{2n}$ be a system of coordinates in X. Define $\varphi_0: X \to X$ by $\varphi_0(x) = \sqrt{|a|}x$ and $\Psi \equiv \varphi \circ \varphi_0^{-1}$. One can verify immediately that

$$\Psi^* \Omega = b\Omega,$$

where $b = \pm 1$. If we interpret the transformation φ_0 as a convenient choice of the units of measure then one is justified in considering only canonical ($a = 1$) and anticanonical ($a = -1$) transformation.

A priori it is not clear if the same scaling argument goes in general so it would be safer to define a symmetry by (1.1).

Another problem appears when one considers not only a single symmetry transformation, but a family of symmetry transformations which has a group structure. To be more specific, let G be a Lie group acting smoothly on X. If $G \ni g \to \varphi_g \in \mathrm{Diff}(X)$ is the action of G, then according to (1.1) we can call G a symmetry group of (X, Ω) if for any $g \in G$, one has $a_g \in \mathbb{R} \setminus \{0\}$, such that

$$\varphi_g^* \Omega = a_g \Omega. \tag{1.2}$$

Even in the case $X = \mathbb{R}^{2n}$, one cannot hope to eliminate all the a_g' s for any Lie group G by the scaling procedure mentioned above. So it is interesting to have a criterion to decide when the group G permits only $ag = 1$, $\forall g$, and in the opposite case to generalize the theory of Kostant, Souriau, and Kirillov (see, e.g., Ref. 6) as much as possible.

Connected with the more general concept of symmetry (1.2) introduced above, is the question of physical equivalence. Suppose that (X_1, Ω_1) and (X_2, Ω_2) are two symplectic manifolds and that G is a symmetry group in the sense of (1.2). We now ask in which case the two symplectic manifolds describe the same physical situation. Guided by the preceding discussion, we think that a reasonable definition is the following: there exist a diffeomorphism φ: $X_1 \to X_2$, which is also a G-morfism, i.e., commutes with the action of G and there exists $a \in \mathbb{R} \setminus \{0\}$ such that we have

$$\varphi^* \Omega_2 = a\Omega_1. \tag{1.3}$$

The concept of physical equivalence is usually considered in a more restricted sense, specifically with $a = 1$ (see, e.g., Ref. 7). Let us note that in quantum mechanics one has a concept of physical equivalence (see Ref. 8, p. 157) that is more in the spirit of (1.3). These problems are treated in Sec. II.

In Sec. III we study some examples of transitive systems with symmetries, and in Sec. IV we draw some conclusions and suggest further developments.

II. THE GENERAL THEORY

(A) From (1.2) and the fact that $g \to \phi_g$ is an action of G, i.e., satisfies (i) $\phi_g \circ \phi_g, = \phi_{gg}$, (ii) $\phi_e = Id$, it follows that $g \to a_g$ is a one-dimensional real representation of G. Because ϕ is a smooth action, $g \to a_g$ is a smooth function.

Let Lie $G \equiv T_e(G)$ be the Lie algebra of G. One knows that to $g \to a_g$ there corresponds a one-dimensional real representation $\dot{a}$ of Lie G, i.e., $\dot{a}$: Lie $G \to \mathbb{R}$ verifies:

$$[\dot{a}(\xi), \dot{a}(\eta)] = \dot{a}([\xi, \eta]), \quad \forall\, \xi, \eta \in \text{Lie } G.$$

But the left-hand side is zero so we get the condition

$$\dot{a}([\xi, \eta]) = 0, \quad \forall\, \xi, \eta \in \text{Lie } G. \tag{2.1}$$

We have the following.

Theorem: Let G be a simply connected Lie group. Then in (1.2) we have $a_g = 1$, $\forall\, g$ iff [Lie G, Lie G] = Lie G.

Proof: $\Leftarrow$ If we have [Lie G, Lie G] = Lie G, then (2.1) shows that $\dot{a} = 0$, which implies that $a_g = 1$, $\forall\, g \in G$.

$\Rightarrow$ Suppose that [Lie G, Lie G] is strictly included in Lie G. Let $\eta_1, \ldots, \eta_r$ be a basis in Lie G such that $\eta_1, \ldots, \eta_{r'}$, $(r' < r)$ is a basis in [Lie G, Lie G]. Then define $\dot{a}(\alpha_1 \eta_1 + \cdots + \alpha_r \eta_r) = \rho_{r'+1}\alpha_{r'+1} + \cdots + \rho_r \alpha_r$ with $\rho_{r'+1}, \cdots, \rho_r \in \mathbb{R}$; $\dot{a}$ verifies (2.1) so it is a nontrivial representation of Lie G, which gives a corresponding nontrivial representation of G, because G is simple connected. ■

Remarks: (1) In the language of the theory of Lie algebras cohomology (see, for instance, Ref. 6), the condition [Lie G, Lie G] = Lie G is equivalent to H^1 (Lie G, $\mathbb{R}$) = 0. So H^1 (Lie G, $\mathbb{R}$) = 0 is a sufficient condition to have $a_g = 1$, $\forall\, g \in G$.

 0022-2488/89/112638-08$02.50

A great number of groups of interest in physics verifies this condition: the semisimple Lie groups [the assertion that $H^1(\text{Lie } G, \mathbb{R}) = 0$ is one of the Whitehead lemmas, see Refs. 6 and 9]. This is also true for the special Euclidian group and the proper orthochronous Poincaré group (see, e.g., Refs. 3 and 6).

(2) If G is not simply connected it is possible that the only one-dimensional real representation of G is the trivial one, even if $H^1(\text{Lie } G\,\mathbb{R}) \neq 0$, e.g., $G = \mathrm{U}(1)$. From the proof of the theorem it follows that if G is not simply connected, then the condition $H^1(\text{Lie } G, \mathbb{R}) \neq 0$ is only necessary for the existence of a nontrivial, real, one-dimensional representation of G. We note that there exist Lie groups that are not simply connected, and have nontrivial, real, one-dimensional representations, e.g., the proper Galilei group.

(B) Suppose now that the Lie group G is such that there exists a nontrivial, real, one-dimensional representation. We will show that the general theory of construction of transitive actions can be adapted to this situation. We follow the presentation used in Ref. 6 and indicate only the necessary modifications.

(1) If (X, Ω) is a symplectic manifold, and $g \to \phi_g$ is a symmetry of (X, Ω) in the sense of (1.2), define as in Ref. 6 for any $x \in X$, $\psi_x : G \to M$ by

$$\psi_x(g) \equiv \phi_g(x)$$

and let ω_x be the two differential form on G, given by

$$\omega_x \equiv \psi_x^* \Omega. \tag{2.2}$$

Then one can verify immediately that ω_x is closed ($d\omega_x = 0$) and that ω_x verifies that identity

$$l_g^* \omega_x = a_g \omega_x, \quad \forall\, g \in G, \tag{2.3}$$

where l_g is the left multiplication with g in G.

(2) Let ω be a closed r-form on G, which satifies

$$l_g^* \omega = a_g \omega, \quad \forall\, g \in G. \tag{2.4}$$

First notice that ω is completely determined by ω_e. In fact (2.4) gives us

$$\omega_g(\xi_1, \ldots, \xi_r) = a_g \omega_e(l_{g^{-1}*}\xi_1, \ldots, l_{g^{-1}*}\xi_r), \tag{2.5}$$

for $\forall\, g \in G$ and $\forall\, \xi_1, \ldots, \xi_r \in T_g(G)$.

Let us translate the condition $d\omega = 0$ into an equivalent condition on $\omega_e \in \Lambda^r$ Lie G. If $x_0, \ldots, x_r \in \mathfrak{X}(G)$ ($\equiv$ the set of vector fields on G), one has the following formula (see, e.g., Ref. 10, p. 36)

$$\begin{aligned}(d\omega)(X_0, X_1, \ldots, X_r) &= \frac{1}{r+1} \sum_{0 \le i \le r} (-1)^i X_i(\omega(X_0, \ldots, \hat{X}_i, \ldots, X_r)) \\ &+ \frac{1}{r+1} \sum_{0 \le i < j \le r} (-1)^{i+j} \omega([X_i, X_j], \\ &X_0 \cdots \hat{X}_i \cdots \hat{X}_j \cdots X_r)\end{aligned} \tag{2.6}$$

(with the Bourbaki convention $\Sigma_{i \in \phi} = 0$). The condition $d\omega = 0$ is equivalent to

$$(d\omega)_g(\xi_0, \ldots, \xi_r) = 0, \tag{2.7}$$

$\forall\, g \in G$, $\forall\, \xi_0, \ldots, \xi_r \in T_g(G)$. We denote $\eta_i \equiv l_{g^{-1}*}\xi_i$ for $i = 0, \ldots, r$. Evidently $\eta_i \in T_e(G) = \text{Lie } G$ for any i. If $\eta \in \text{Lie } G$, let $Y_\eta \in \mathfrak{X}(G)$ be defined by

$$(Y_\eta)_g \equiv l_{g*}\eta. \tag{2.8}$$

Then (2.7) is equivalent with

$$d\omega(Y_{\eta_0}, \ldots, Y_{\eta_r})_g = 0, \quad \forall\, g \in G, \forall\, \eta_0, \ldots, \eta_r \in \text{Lie } G.$$

If we use (2.6) this relation is equivalent with

$$\sum_{0 \le i \le r} (-1)^i Y_{\eta_i}(\omega(Y_{\eta_0}, \ldots, Y_{\eta_i}, \ldots, Y_{\eta_r}) + \sum_{0 \le i < j \le r} (-1)^{it+j} \omega(Y_{[\eta_i, \eta_j]}, Y_{\eta_0}, \ldots, \hat{Y}_{\eta_i}, \ldots, \hat{Y}_{\eta_j}, \ldots)_g = 0,$$

for $\forall\, g \in G$, $\forall\, \eta_0, \ldots, \eta_r \in \text{Lie } G$.

But by definition

$$\begin{aligned}Y_{\eta_0}(\omega(Y_{\eta_1}, \ldots, Y_{\eta_r}))_g &= \frac{d}{dt} \omega_{(\exp t\eta_0)g}((Y_{\eta_1})_{(\exp t\eta_0)g}, \ldots, (Y_{\eta_r})_{(\exp t\eta_0)g})\big|_{t=0} \\ &\overset{(2.7)}{=} \frac{d}{dt} \omega_{(\exp t\eta_0)g}(l_{(\exp t\eta_0)g*}\eta_1, \ldots, l_{(\exp t\eta_0)g*}\eta_r)\big|_{t=0} \\ &= \frac{d}{dt}(l^*_{(\exp t\eta_0)g}\,\omega)_e(\eta_1, \ldots, \eta_r)\big|_{t=0} \\ &\overset{(2.4)}{=} \frac{d}{dt} a_{\exp t\eta_0} a_g \omega_e(\eta_1, \ldots, \eta_r)\big|_{t=0} \\ &= \dot{a}(\eta_0) a_g \omega_e(\eta_1, \ldots, \eta_r).\end{aligned}$$

Using this relation in (2.9) we get the identity

$$\begin{aligned}&\sum_{0 \le i \le r} (-1)^i \dot{a}(\eta_i)\omega_e(\eta_0, \ldots, \hat{\eta}_i, \ldots, \eta_r) \\ &+ \sum_{0 \le i < j \le 0} (-1)^{ij} \\ &\omega_e([\eta_i \eta_j], \eta_0, \ldots, \hat{\eta}_1, \ldots, \hat{\eta}_j, \ldots, \eta_r) = 0,\end{aligned} \tag{2.10}$$

for $\forall\, \eta_1, \ldots, \eta_r \in \text{Lie } G$.

Remark: Let us reformulate the result with the help of the theory of Lie algebra cohomology (see Refs. 6 and 9). Let L be a Lie algebra, V a vector space, and ρ a representation of L in V. Denote by $C^r(L, V)$ the V-valued completely antisymmetric r-linear forms on L (for $r \in \mathbb{N}$), $C^0(L, V) \equiv V$, $C^{-1}(L, V) \equiv 0$. The elements of $C^r(L, V)$ are

called r-cochains with values in V. Define the cobord operator $\delta = C^r(L, V) \to C^{r+1}(L, V)$ by the following formula:

$$\delta c(\eta_0, \ldots, \eta_r) \equiv \sum_{0<i<j<r} \rho(\eta_i) c(\eta_0, \ldots, \hat{\eta}_i, \ldots, \eta_r) + \sum_{0<i<j<r} (-1)^{ij} c([\eta_i, \eta_j], \eta_0, \ldots, \hat{\eta}_i, \ldots, \hat{\eta}_j, \ldots, \eta_r). \tag{2.11}$$

Then one can prove that $\delta^2 = 0$ (see Ref. 9). In $C^r(L, V)$ there are two remarkable subspaces: $Z^r_\rho(L, V) = \operatorname{Ker} \delta$ (the r-cocycles) and $B^r_\rho(L, V) = \operatorname{Im} \delta$ (r-cobords). Because $\delta^2 = 0$, one has $B^r_\rho(L, V) \subset Z^r_\rho(L, V)$. The subspaces $H^r_\rho(L, V) = Z^r_\rho(L, V)/B^r_\rho(L, V)$ are the r-cohomology groups of L with respect to ρ. If ρ is zero, the index ρ is omitted. So we can formulate the result above as follows:

$d\omega = 0$ iff $\omega_e \in Z^2_a(\text{Lie } G, \mathbb{R})$.

Let us give the formulas for two particular cases that we will need in the following: if $c \in Z^2_\rho(L, V)$ we have from (2.11)

$$\rho(\eta_1)c(\eta_2, \eta_3) + \text{cycl} - \{c([\eta_1, \eta_2], \eta_3) + \text{cycl}\} = 0 \tag{2.12}$$

if $c \in B^2_\rho(L, V)$, i.e., $c = \delta b$ with $b \in c^1(L, V)$, we have

$$c(\eta_1, \eta_2) = \rho(\eta_1)b(\eta_2) - \rho(\eta_2)b(\eta_1) - b([\eta_1, \eta_2]). \tag{2.13}$$

(3) From (1) and (2) it follows that we have a map $\psi_\Omega : X \to Z^2_a(\text{Lie } G, \mathbb{R})$ given by

$$\psi_\Omega(X) \equiv (\omega_x)_e. \tag{2.14}$$

As in Ref. 6 it can be proved that ψ_Ω verifies

$$\psi_\Omega(\phi_g(x)) = a_g \operatorname{Ad}^\#_g \psi_\Omega(x), \quad \forall g \in G, \tag{2.15}$$

where $g \to \operatorname{Ad}^\#_g$ if the coadjoint action of G in Λ^2 Lie G. Let H_x denote the stability subgroup of the point $x \in X$, and G_σ the stability subgroup of the point $\sigma \in Z^2_a(\text{Lie } G, \mathbb{R})$ with respect to the action $g \to a_g \operatorname{Ad}^\#_g$ of G. Then from (2.15) it follows that $H_x \subset G_{\psi_\Omega(x)}$.

(4) Let $x \in X$. Denote $\sigma \equiv \psi_\Omega(x)$ and

$$h_\sigma \equiv \{\xi \in \text{Lie } G \mid \sigma(\xi, \eta) = 0, \ \forall \eta \in \text{Lie } G\}.$$

Then from (2.12) it follows immediatley that h_σ is a Lie subalgebra of Lie G.

(5) Suppose that G acts transitively on X. Then one knows that X is a homogeneous space of the form G/H, where $H \subset G$ is a closed subgroup of G. One can prove as in Ref. 6 that H^0 ($\equiv$ the connected component of the identity in H) is equal with H_σ ($\equiv$ the unique connected Lie subgroup of G associated with the Lie subalgebra h_σ), where $\sigma = \psi_\Omega(H)$. It follows that G/H_σ is a covering manifold for X.

(6) Suppose that $\sigma \in Z^2_a(\text{Lie } G, \mathbb{R})$ is such that H_σ is a closed subgroup of G. Let ω be the unique closed two-form on G which verifies (2.3) and $\omega_e = \sigma$. Then using the reduction principle (see Ref. 6, theorem 25.2) it follows that there exists a unique symplectic form Ω on G/H such that $\pi^* \Omega = \omega$. (Here π: $G \to G/H_\sigma$ is the canonical projection which is in this case a submersion.)

By a simple computation one can establish the formula

$$\pi^*(a_g^{-1} \phi_g^* \Omega) = \omega, \quad \forall g \in G.$$

Because π is a submersion, π^* is injective, and we get that Ω satisfies (1.2).

(7) Let G/H_1 and G/H_2 be two homogeneous manifolds for the group G, Ω_1 and Ω_2 the corresponding symplectic forms on G/H_1 and G/H_2, respectively, and φ: $G/H_1 \to G/H_2$ a G-diffeomorphism which satisfies (1.3). Because φ is a G-morphism it can be proved that H_1 and H_2 are conjugated subgroups of G, and φ is of the form $\varphi(gH_1) = gg_0H_2$. Then one gets easily:

$$a\psi_{\Omega_1}(H_1) = a_{g_0} \operatorname{Ad}^\#_{g_0} \psi_{\Omega_2}(H_2). \tag{2.16}$$

(8) One knows that if V is a vector space and T a linear representation of a Lie group in V, then T factorize to the projective space of V. Let us denote by $P(Z^2_a(\text{Lie } G, \mathbb{R}))$ the projective space of $Z^2_a(\text{Lie } G, \mathbb{R})$ and by $g \to \operatorname{Ad}^\natural_g$ the factorized of the action $g \to a_g \operatorname{Ad}^\#_g$ of G in $Z^2_a(\text{Lie } G, \mathbb{R})$. Then (2.16) gives

$$\widehat{\psi_{\Omega_1}(H_1)} = \operatorname{Ad}^\natural_{g_0} \widehat{\psi_{\Omega_2}(H_2)}, \tag{2.17}$$

where $\hat{\sigma} \in PZ^2_a(\text{Lie } G, \mathbb{R})$ is the equivalence class of $\sigma \in Z^2_a(\text{Lie } G, \mathbb{R})$.

(9) Let $\mathcal{O} \in PZ^2_a(\text{Lie } G, \mathbb{R})$ be an orbit with respect to the action $g \to \operatorname{Ad}^\natural_g$ of G. We say that $\mathcal{O}$ is a regular orbit if there exists $\Sigma \in \mathcal{O}$ and $\sigma \in \Sigma$ such that H_σ is a closed subgroup of G. [Because of (7), it follows that $\forall \Sigma \in \mathcal{O}$, $\forall \sigma \in \Sigma$, H_σ is a closed subgroup of G.]

Summing up the preceding line of arguments, we can formulate the main result of this paper.

Theorem: Let G be a Lie group. Then, up to covering, the transitive nontrivial symplectic manifolds on which G acts according to (1.2) are parametrized by the regular orbits of $(PZ^2_a(\text{Lie } G, \mathbb{R}))$.

(10) From (3) and (5) we can see that the transitive symplectic manifolds on which G acts according to (1.2) are of the form G/H, where the closed subgroup H verifies that: (i) there exists $\sigma \in Z^2_a(\text{Lie } G, \mathbb{R})$, such that H_σ is a closed subgroup of G; (ii) $H^0 = H_\sigma$; (iii) $H \subset G_\sigma$.

We have now the following useful corollary.

Corollary: Suppose that the Lie group G satisfies the following condition: if $H \subset G$ is a closed subgroup of G which satisfies (i)–(iii), then $H = H_\sigma$. Then all the transitive symplectic manifolds of G are of the form G/H_σ, where $\hat{\sigma}$ belongs to a regular orbit in $(PZ^2_a(\text{Lie } G, \mathbb{R}))$.

Remark: One can prove that, in general, all the transitive symplectic manifolds on which G acts according to (1.2) are of the form G/H where the closed subgroup H verifies (i)–(iii). In the examples in Sec. III we will give only the maximal manifolds G/H_σ.

III. EXAMPLES

We illustrate the theory developed in Sec. II with some examples.

A. The one-dimensional Newton group[11]

We identify this group with real 3×3 matrices of the form

$$(\eta, a, v) \equiv \begin{pmatrix} \operatorname{ch}\eta & \operatorname{sh}\eta & 0 \\ \operatorname{sh}\eta & \operatorname{ch}\eta & 0 \\ a & v & 1 \end{pmatrix}, \quad \eta\, a, v \in \mathbb{R},$$

with the law of composition induced by the matrix multiplication. The Lie algebra of this group can be identified with the real linear space of 3×3 real matrices of the form:

$$(t, x, u) \equiv \begin{pmatrix} 0 & t & 0 \\ t & 0 & 0 \\ x & u & 0 \end{pmatrix}, \quad t, x, u \in \mathbb{R},$$

with the Lie bracket

$$[(t, x, u), (t', x', u')] = (0, ut' - u't, xt' - x't).$$

The real one-dimensional representation of this Lie algebra are of the form

$$\dot{a}(t, x, u) = \rho t, \quad \rho \in \mathbb{R}\backslash\{0\},$$

and they induce representations of the group.

I. We take the following basis in Lie G.

$$h \equiv \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$k \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then the cocycle condition (2.12) $\delta c\ (h, p, k) = 0$, gives $c(p, k) = 0$, so the most general element in $Z^2_{\dot{a}}$ (Lie G, $\mathbb{R}$) is of the form: $\lambda\, c_1 + \sigma\, c_2$, where

$$c_1((t, x, u), (t', x', u')) = xt' - x't,$$

$$c_2((t, x, u), (t', x', u')) = ut' - u't.$$

It follows that the projective space ($PZ^2_{\dot{a}}$ (Lie G, $\mathbb{R}$)) is formed by the elements:

$$(a)\,\Sigma^{(1)}_{\epsilon} \equiv \{\mu(c_1 + \epsilon\, c_2) \,|\, \mu \in \mathbb{R}\backslash\{0\}\}, \quad |\epsilon| < 1,$$

$$(b)\,\Sigma^{(2)}_{\epsilon} \equiv \{\mu(c_2 + \epsilon c_1) \,|\, \mu \in \mathbb{R}\backslash\{0\}\}, \quad |\epsilon| < 1,$$

$$(c)\,\Sigma^{(3),(4)} \equiv \{\mu(c_1 \pm c_2) \,|\, \mu \in \mathbb{R}\backslash\{0\}\}.$$

II. By a simple computation one gets

$$\mathrm{Ad}^{\natural}_{\eta, a, v}\, \Sigma^{(1),(2)}_{\epsilon} = \Sigma^{(1),(2)}_{(\epsilon + \operatorname{th}\eta)/(1 + \epsilon \operatorname{th}\eta)},$$

so, choosing η conveniently we can arrange it so that $\mathrm{Ad}^{\natural}_{\eta, a, v}\, \Sigma^{(1),(2)}_{\epsilon} = \Sigma^{(1),(2)}_{0}$. So, $\{\Sigma^{(1),(2)}_{\epsilon} \,|\, |\epsilon| < 1\}$, are two distinct orbits. Also $\Sigma^{(3),(4)}$ are left invariant by $\mathrm{Ad}^{\natural}$.

III. (a) $h_{c_1} = \{(0, 0, u) \,|\, u \in \mathbb{R}\}$

$$\Rightarrow H_{c_1} = \{(0, 0, v) \,|\, v \in \mathbb{R}\}$$

is a closed subgroup.

(b) $h_{c_2} = \{(0, x, 0) \,|\, x \in \mathbb{R}\} \Rightarrow H_{c_2} = \{(0, a, 0) \,|\, a \in \mathbb{R}\}$

is a closed subgroup.

(c) $h_{c_1 \pm c_2} = \{(0, x, \mp x)\} \Rightarrow H_{c_1 \pm c_2} = \{(0, a, \mp a)\}$

is a closed subgroup.

IV. (a) From (2.5) we get

$$\omega_{\eta, a, v} = e^{\rho\eta}\, da \wedge d\eta.$$

If we identify X with the set $\{(\eta, a, 0) \,|\, \eta, a \in \mathbb{R}\}$ then by the reduction principle the symplectic form is

$$\Omega_{\eta, a} = e^{\rho\eta}\, da \wedge d\eta.$$

The action of the group is

$$\phi_{\eta', a', v'}(\eta, a) = (\eta + \eta', a + v' \operatorname{ch}\eta + a' \operatorname{sh}\eta).$$

(b) From (2.5) we get

$$\omega_{\eta, a, v} = e^{\rho\eta}\, dv \wedge d\eta.$$

We identify X with the set $\{(\eta, 0, v) \,|\, \eta, v \in \mathbb{R}\}$. Then the symplectic form is

$$\Omega_{\eta', v} = e^{\rho\eta}\, d v \wedge d\eta$$

and the action of the groups is

$$\phi_{\eta', a', v'}(\eta, v) = (\eta + \eta', a' \operatorname{sh}\eta + v' \operatorname{ch}\eta + v).$$

(c), (d) The form is

$$\omega_{\eta, a, v} = e^{\rho\eta}(da \pm dv) \wedge d\eta.$$

We identify the space X with the set $\{(t, \alpha, \pm\alpha)\}$. Then, the action of the group is

$$\phi_{\eta, a, v}(t, \alpha) = (\eta + t, \alpha + e^{\pm t}(a \pm v)),$$

and the symplectic form is

$$\Omega_{t,\alpha} = e^{\rho t} d\alpha \wedge d\eta.$$

B. The one-dimensional Poincaré group[11]

This group can be identified with 3×3 real matrices of the form

$$(\chi, \eta, a) \equiv \begin{pmatrix} \operatorname{ch}\chi & \operatorname{sh}\chi & \eta \\ \operatorname{sh}\chi & \operatorname{th}\chi & a \\ 0 & 0 & 1 \end{pmatrix}, \quad \chi, \eta, a \in \mathbb{R}$$

with the composition law induced by matrix multiplication. The Lie algebra of this group can be identified with the real linear space of 3×3 matrices of the form:

$$(v, t, x) \equiv \begin{pmatrix} 0 & v & t \\ v & 0 & x \\ 0 & 0 & 0 \end{pmatrix}, \quad v, t, x \in \mathbb{R}$$

with the Lie bracket

$$[(v, t, x), (v', t', x')] = (0, vx' - v'x, vt' - v't).$$

The real one-dimensional representations of this Lie algebra are of the form:

$$\dot{a}(v, t, x) = \rho\, v, \quad \rho \in \mathbb{R}\backslash\{0\}.$$

and they induce representations of the group.

I. We take the following basis in the Lie algebra:

$$h = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then the two-cocycle condition (2.12) $\delta c(h,p,k) = 0$ gives $c(h,p) = 0$, so the most general two-cocycle is of the form $\mathrm{H}c_1 + \mathrm{P}c_2$, where

$$c_1((v,t,x),(v',t',x')) = xv' - x'v,$$

$$c_2((v,t,x),(v',t',x')) = v't - vt'.$$

It follows that the projective space $P(Z^2_{\dot{a}}$ (Lie G,)) is formed from elements of the form

$$(a)\ \Sigma^{(1)}_{\epsilon} \equiv \{\mu(c_1 + \epsilon c_2) \,|\, \mu \in /\{0\}\}, \quad |0| < 1,$$

$$(b)\ \Sigma^{(2)}_{\epsilon} \equiv \{\mu(c_2 + \epsilon c_1) \,|\, \mu \in /\{0\}\}, \quad |0| < 1,$$

$$(c)\ \Sigma^{(3),(4)} \equiv \{\mu(c_1 + c_2) \,|\, \mu \in /\{0\}\}, \quad |0|.$$

II. By a simple computation one finds that

$$\mathrm{Ad}^{\natural}_{\chi,\eta,a}\,\Sigma^{(1),(2)}_{\epsilon} = \Sigma^{(1),(2)}_{(\epsilon - \mathrm{th}\,\chi)/(1+\epsilon\,\mathrm{th}\,\chi)}.$$

For a convenient χ, we have $\mathrm{Ad}^{\natural}_{\chi,\eta,a}\Sigma^{(1),(2)}_{\epsilon} = \Sigma^{(1),(2)}_{0}$ so that $\{\Sigma^{(1),(2)}_{\epsilon} \mid |\epsilon| < 1\}$ are two distinct orbits.

Also $\Sigma^{(3)}$ and $\Sigma^{(4)}$ are left invariant by $\mathrm{Ad}^{\natural}$.

III. (a) $h_{c_1} = \{(0,t,0)\,|\,t\in\mathbb{R}\}$; then $H_{c_1} = \{(0,\eta,0)\,|\,\eta\in\mathbb{R}\}$ is a closed subgroup.

(b) $h_{c_2} = \{(0,0,x)\,|\,x\in\mathbb{R}\}$; then $H_{c_2} = \{(0,0,a)\,|\,a\in\mathbb{R}\}$ is a closed subgroup.

(c),(d) $h_{c_1\pm c_2} = \{(0,t,\mp t)\}$; then $H_{c_1\pm c_2} = \{(0,\eta,\pm\eta)\}$ is a closed subgroup.

It follows that all four orbits are regular.

IV. (a) From (2.5) we find the form ω on the group

$$\omega_{\chi,\eta,a} = e^{\rho\chi}(\mathrm{ch}\,\chi\,da - \mathrm{sh}\,\chi\,d\eta)\wedge d\chi.$$

From the composition law we get that

$$(\chi,\eta,a) = (\chi,0,a')(0,\eta',0)$$

with $a' = a - \chi\,\mathrm{th}\,\chi$, $\eta' = \eta/\mathrm{ch}\,\chi$. If we take as coordinates χ, a', η', then ω becomes

$$\omega_{\chi,\eta',a'} = \mathrm{ch}\,\chi e^{\rho\chi}\,da'\wedge d\chi.$$

We can identify X with the set $\{(\chi,0,a)\,|\,\chi,a\in\mathbb{R}\}$, and the reduction principle gives the symplectic form

$$\Omega_{\chi,a} = \mathrm{ch}\,\chi e^{\rho\chi}\,da\wedge d\chi.$$

After some computations we find the action of the group to be

$$\phi_{\chi',\eta',a'}(\chi,a) = (\chi+\chi', a' + a\,\mathrm{ch}\,\chi' - (\eta' + \mathrm{sh}\,\chi' a)\mathrm{th}(\chi+\chi')).$$

(b) From (2.5) the form ω on the group is

$$\omega_{\chi,\eta,a} = e^{\rho\chi}(\mathrm{ch}\,\chi\,d\eta - \mathrm{sh}\,\chi\,da)\wedge d\chi.$$

As in case (a), we observe first that from the composition law of the group we have

$$(\chi,\eta,a) = (\chi,\eta',0)(0,0,a')$$

with $\eta' = \eta - a\,\mathrm{th}\,\chi$, $a' = a/\mathrm{ch}\,\chi$. If we take as coordinates χ, η', a' then ω becomes

$$\omega_{\chi,\eta',a'} = e^{\rho\chi}\,\mathrm{ch}\,\chi\,d\eta'\wedge d\chi.$$

We identify X with the set $\{(\chi,\eta,0)\,|\,\chi,\eta,\in\mathbb{R}\}$, and the reduction principle gives the following symplectic form:

$$\Omega_{\chi,\eta} = e^{\rho\chi}\,\mathrm{ch}\,\chi\,d\eta\wedge d\chi.$$

As in case (a) we find easily the action of the group:

$$\phi_{\chi',\eta',a'}(\chi,\eta) = (\chi+\chi', \eta' + \eta\,\mathrm{ch}\,\chi' - (a' + \mathrm{sh}\,\chi'\eta)\mathrm{th}(\chi+\chi'))$$

(c),(d) The form ω is

$$\omega = \pm e^{(\rho\mp 1)\chi}(d\eta \pm da)\wedge d\chi.$$

We identify $X\simeq\{(v,\alpha,\pm\alpha)\}$. Then, the action of the group is

$$\phi_{\chi,\eta,a}(v,\alpha) = (\chi + v, \eta \pm a + e^{\pm\chi}\alpha)$$

and the asymptotic form

$$\Omega_{v,\alpha} = \pm e^{(\rho\mp 1)\chi}d\alpha\wedge d\chi.$$

C. The one-dimensional Galilei group[11]

We identify this group with 3×3 real matrices of the form

$$(v,\eta,a) \equiv \begin{pmatrix} 1 & v & a \\ 0 & 1 & \eta \\ 0 & 0 & 1 \end{pmatrix}, \quad v,\eta,a\in\mathbb{R}$$

with the matrix multiplication as the composition law.

The Lie algebra of the group can be identified with the linear space of 3×3 real matrices of the form:

$$(u,t,x) = \begin{pmatrix} 0 & u & x \\ 0 & 0 & t \\ 0 & 0 & 0 \end{pmatrix}, \quad u,t,x\in\mathbb{R}$$

with the Lie bracket

$$[(u,t,x),(u',t',x')] = (0,0,ut' - u't).$$

The real one-dimensional representations of this Lie algebra are of the form:

$$\dot{a}(u,t,x) = \xi u + \rho t, \quad \xi^2 + \rho^2 \neq 0, \quad \xi,\rho\in\mathbb{R}$$

and give the one-dimensional real representation of the group.

I. We take the following basis in the Lie algebra:

$$h = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then the two-cocycle identity (2.12) $\delta c(h,p,g) = 0$ gives $\rho c(p,g) + \xi c(h,p) = 0$. It follows that there exists $\lambda\in\mathbb{R}$ such that $c(p,g) = \lambda\xi$ and $c(h,p) = -\lambda\rho$. So, the most general element of Z^2_{a} (Lie $G,\mathbb{R}$) is of the form $\lambda c_1 + \sigma c_2$, where

$$c_1((u,t,x),(u',t',x')) = \xi(xu' - x'u) - \rho(tx' - t'x),$$

$$c_2((u,t,x),(u',t',x')) = tu' - t'u.$$

The elements of $P(Z^2_{\mathrm{a}}$ (Lie $G,\mathbb{R}))$ are then

(a) $\Sigma^{(1)} = \{\sigma c_2\,|\,\sigma\in\mathbb{R}\setminus\{0\}\}$,

(b) $\Sigma^{(2)}_{\epsilon} = \{\lambda(c_1 + \epsilon c_2)\,|\,\lambda\in\mathbb{R}\setminus\{0\}\}$.

II. $\Sigma^{(1)}$ is invariant with respect to the $\mathrm{Ad}^{\natural}$ action, and

$$\mathrm{Ad}^{\natural}_{v,\eta,a}\Sigma^{(2)}_{\epsilon} = \Sigma^{(2)}_{\epsilon - (\xi v + \rho\eta)}$$

For a convenient choice of v and η we have $\mathrm{Ad}^{\natural}_{v,\eta,a}\,\Sigma^{(2)}_{\epsilon} = \Sigma^{(2)}_{0}$, so $\{\Sigma^{(2)}_{\epsilon}\,|\,\epsilon\in\mathbb{R}\}$ is a whole orbit.

III. (a) $h_{c_2} = \{(0,0,x)\,|\,x\in\mathbb{R}\}$; then $H_{c_2} = \{(0,0,a)\,|\,a\in\mathbb{R}\}$ is a closed subgroup.

(b) $h_{c_1} = \{(u,t,0)\,|\}u + \rho t = 0\}\Rightarrow$

$H_{c_1} = \{(v,\eta,0)\,|\}v + \rho\eta = 0\}$, a closed subgroup.

IV. (a) From (2.5) ω is

$$\omega_{v,\eta,a} = e^{\xi v + \rho\eta}\,d\eta\Lambda dv.$$

We identify X with $\{(v,\eta,0)\,|\,v,\eta\in\mathbb{R}\}$. The symplectic form is

$$\Omega_{v,\eta} = e^{\xi v + \rho\eta}\,d\eta\Lambda dv$$

and the action of the group is

$$\phi_{v',\eta',a'}(v,\eta) = (v + v', \eta + \eta').$$

It is convenient to work in new variables:

$$\begin{cases} \alpha \equiv \xi v + \rho\eta, \\ \beta \equiv \xi\eta - \rho v. \end{cases}$$

Then, the form ω is

$$\omega_{\alpha,\beta,a} = e^{\alpha}(da\wedge d\alpha + [(\zeta\alpha-\rho\beta)/(\zeta^2+\rho^2)]d\alpha\wedge d\beta).$$

In these new coordinates $H_{c_1} = \{(0,\beta,0)\}$ so we can identify $X \simeq \{(\gamma,0,b)\}$. The symplectic form on X is then

$$\Omega_{\gamma,b} = e^{\gamma} db\wedge d\gamma.$$

The action of the group on X is

$$\phi_{\alpha,\beta,a}(\gamma,b) = (\alpha+\gamma, a+b-\gamma\beta/(\zeta^2+\rho^2) + \zeta\alpha(\rho\gamma-\zeta\beta)/(\zeta^2+\rho^2)^2).$$

D. The special Galilei group

As in Ref. 6 we realize the special Galilei $\mathscr{G}^{\uparrow}_{+}$ group as 5×5 real matrices of the form

$$(R,v,\eta,a) \equiv \begin{pmatrix} R & v & a \\ 0 & 1 & \eta \\ 0 & 0 & 1 \end{pmatrix}, \quad R\in SO(3), \quad v,a\in\mathbb{R}, \ \eta\in\mathbb{R},$$

with the matrix multiplication as the composition law.

The Lie algebra Lie $\mathscr{G}^{\uparrow}_{+}$ can be thought of as the linear space of 5×5 matrices of the form

$$(a,u,t,x) = \begin{pmatrix} a & u & x \\ 0 & 0 & t \\ 0 & 0 & 0 \end{pmatrix},$$

where a is a real antisymmetric 3×3 matrix, $u,x\in\mathbb{R}^3$, $t\in\mathbb{R}$. The Lie bracket is

$$[(a,u,t,x),(a'u',t',x')] = ([a,a'],au'-a'u,0,ax'-a'x+t'u-tu').$$

The real one-dimensional representations of Lie $\mathscr{G}^{\uparrow}_{+}$ are of the form

$$\dot{a}(a,u,t,x) = \rho t, \quad \rho\in\mathbb{R}\setminus\{0\}$$

and give one-dimensional real representations of $\mathscr{G}^{\uparrow}_{+}$.

I. We follow closely the line of argument from Ref. 3. First we take the following basis in *Lie* $\mathscr{G}^{\uparrow}_{+}$:

$$j_i = \begin{pmatrix} l_i & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p_i = \begin{pmatrix} 0 & 0 & e_i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$g_i = \begin{pmatrix} 0 & e_i & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad h = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

where $i = 1,2,3$, $\{e_i\}_{i=1}^{3}$ is the natural orthonormal basis in $\mathbb{R}^3$, and l_i are the 3×3 matrices given by $(l_i)_{jk} \equiv -\epsilon_{ijk}$. ($\epsilon_{ijk}$ is the completely antisymmetric tensor of rank 3, defined with the convention $\epsilon_{123} = 1$.)

We divide the study into seven steps.

(i) Denote $C_{ik} = c(j_i,j_k)$. C is a 3×3 real antisymmetric matrix. Define $B\in C^1$ (Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) by $b(a,u,t,x) = \frac{1}{2}\,\mathrm{Tr}\,aC$, and $c^1\in Z^2_{\dot{a}}$ (Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) by $c^1 = c+\delta b$. Then it is easy to establish that $c^1(j_i,j_k) = 0$.

(ii) Denote $C^1_{ij} = c^1(j_i,p_j)$. From the cocycle identity $\delta c^1(j_i,j_j,p_k) = 0$ one gets $C^1_{ij} = -c^1_{ji}$. Now we can define $B^1\in\mathbb{R}^3$ by $B^1_i = \frac{1}{2}\sum_{j,k=1}^{3}\epsilon_{ijk}c^1_{jk}$, $b^1\in C^1$ (Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) by $b(a,u,t,x) = x\cdot B^1$, and $c^2\in Z^2_{\dot{a}}$ (Lie $\mathscr{G}_{+},\mathbb{R}$) by $c^2 = c^1+\delta b^1$. Then it is easy to verify that $c^2(j_i,p_j) = 0$ and $c^2(j_i,j_k) = 0$.

(iii) From the cocycle identity $\delta c^2(p_i,p_j,j_k) = 0$, one gets also $c^2(p_i,p_j) = 0$.

(iv) Denote $C^2_{ij} = C^2(j_i,g_j)$. From the cocycle identity $\delta c^2(j_i,j_j,g_k) = 0$ one gets, $C^2_{ij} = -C^2_{ji}$. As before, define $B^2\in\mathbb{R}^3$ by $B^2_i \equiv \frac{1}{2}\sum_{j,k=1}^{3}\epsilon_{ijk}C^2_{jk}$, $b^2\in C^1$ (Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) by $b^2(a,u,t,x) = u\cdot B^3$, and $c^3\in Z^2_{\dot{a}}$ (Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) by $c^3 = c^2+\delta b^2$.

Then one can verify that $c^3(j_i,j_j) = 0$, $c^3(j_i,p_j) = 0$, $c^3(p_i,p_j) = 0$, and $c^3(j_i,g_j) = 0$.

(v) By some computation it can be established that

$$\delta c^3(g_i,g_k,j_k) = 0 \Rightarrow c^3(g_i,g_j) + 0,$$

$$\delta c^3(j_i,g_j,h) = \Rightarrow c^3(j_i,h) = 0,$$

$$\delta c^3(j_i,p_j,h) = 0 \Rightarrow c^3(p_i,h) = 0.$$

(vi) Denote $C^3_{ij} = c^3(g_i,p_j)$. Then from $\delta c^3(j_i,g_j,p_k) = 0$ it follows that there exists $m\in\mathbb{R}$ such that $C^3_{ij} = m\delta_{ij}$.

(vii) From $\delta c^3(h,g_i,p_j) = 0$ and the fact that $\rho\neq 0$, it follows now that $m = 0$.

In conclusion we have proved that $c^3 = 0$. It follows that $c = \delta d$, where $d\in C^1$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$), i.e., $c\in B^2_{\dot{a}}$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$).

Remark: In the language of Lie algebra cohomology we have proved that $H^2_{\dot{a}}$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) $= 0$. This can be contrasted with H^2(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) $\neq 0$. Indeed for $\dot{a} = 0$, $\delta c^3(h,g_i,p_j) = 0$ is trivially satisfied and we get from (vi):

$$c^3((a,u,t,x),(a',u',t',x')) = m(u\cdot x' - u'\cdot x), \tag{3.1}$$

i.e., the result first obtained by Bargmann.[12] Now, the most general element $d\in C^1$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) = (Lie $\mathscr{G}^{\uparrow}_{+}$)* is of the form: (τ,P,G,H) with τ real antisymmetric 3×3 matrix, $P,G,\in\mathbb{R}^3$, $H\in\mathbb{R}$, and is given by

$$(\tau,P,G,H)(a,u,t,x) \equiv \tfrac{1}{2}\,\mathrm{Tr}\,\tau a + P\cdot x + G\cdot u + Ht. \tag{3.2}$$

Using (3.2) one can see easily that if $C_{\tau,P,G,H} = \delta(\tau,P,G,H)$ then we have

$$C_{\tau,P,G,H}((a,u,t,x),(a',u',t',x')) = \tfrac{1}{2}\mathrm{Tr}\,\tau(\rho(ta'-t'a) - [a,a']) + P\cdot[\rho(tx'-t'x) - ax' + a'x + tu' - tu'] + G\cdot[tu' - t'u - au' + a'u].$$

From this relation one can see that H drops out, and that the elements of $Z^2_{\dot{a}}$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$) are indexed by τ,P,G. We denote by $C_{\tau,P,G}$ the corresponding two-cocycle. The elements of $P(Z^2_{\dot{a}}$(Lie $\mathscr{G}^{\uparrow}_{+},\mathbb{R}$)) are the following ones:

(a) $\Sigma^{(1)}_{\tau_0} = (\{sC_{\tau_0,0,0} \mid s\in\mathbb{R}\setminus\{0\}\})$

and τ_0 verifies $\mathrm{Tr}\,\tau_0^2 = 2$.

In the following we identify $(\mathbb{R}^3)^*$ with $\mathbb{R}^3$ naturally and $\Lambda\mathbb{R}^3$ with the 3×3 real antisymmetric matrices by the formula

$$A_{u\Lambda w} \equiv (w\cdot v)u - (v\cdot u)w.$$

Then we have

(b) $\Sigma^{(2)}_{\tau_0,G} \equiv \{\lambda C_{\tau_0,0,G} \mid \lambda\in\mathbb{R}\setminus\{0\}\}$

and τ_0, G verify: $|G| = 1$, $G\Lambda\tau_0 = 0$.

(c) $\Sigma^{(3)}_{\tau_0,G} \equiv \{\lambda C_{\tau_0,0,G} \mid \lambda\in\mathbb{R}\setminus\{0\}\}$

and τ_0, G verify $|G| = 1$, $G \Lambda \tau_0 \neq 0$.

(d) $\Sigma^{(4)}_{\tau_0,P,G} \equiv \{\lambda C_{\tau_0,P,G} | \lambda \in \mathbb{R} \setminus \{0\}\}$,

where $|P| = 1$ and $P \Lambda G = 0$.

(e) $\Sigma^{(5)}_{\tau_0,P,G} \equiv \{\lambda C_{\tau_0,P,G} | \lambda \in \mathbb{R} \setminus \{0\}\}$,

where $|P| = 1$ and $P \Lambda G \neq 0$.

II. (a) Because

$$\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(1)}_{\tau_0} = \Sigma^{(1)}_{R\tau_0 R^{-1}},$$

choosing R conveniently we have $\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(1)}_{\tau_0} = \Sigma^{(1)}_{l_3}$. So $\{\Sigma^{(1)}_{\tau_0} | \mathrm{Tr}\, \tau_0^2 = 2\}$ is an orbit.

(b) We have

$$\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(2)}_{\tau_0,G} = \Sigma^{(2)}_{R\tau_0 R^{-1} + v^t RG - G^t R^t v, RG}. \tag{3.3}$$

Choosing, R, v, and η conveniently we can get $\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(2)}_{\tau_0,G} = \Sigma^{(2)}_{0,e_3}$, so $\{\Sigma^{(2)}_{\tau_0,G} | |G| = 1, G \Lambda \tau_0 = 0\}$ is an orbit.

(c) From (3.3) we can prove that by choosing R, v, and η conveniently we get $\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(3)}_{\tau_0,G} = \Sigma^{(3)}_{sl_3,e_3}$ with $s \in \mathbb{R}_+$. So the orbits ar parametrized by $s \in \mathbb{R}_+$

(d)

$$\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(4)}_{\tau_0,P,G} = \Sigma^{(4)}_{R\tau_0 R^{-1} + a^t RP - P^t R^t a + v^t RG - G^t R^t v, RP, R(G + \eta P)}. \tag{3.4}$$

Choosing η, R, and a conveniently we get $\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(4)}_{\tau_0,P,G} = \Sigma^{(4)}_{sl_3,e_3,0}$, where $s \in \mathbb{R}_+$. So the orbits are indexed by $s \in \mathbb{R}_+$.

(e) From (3.4) we can get by taking suitable R, η, a, and v, $\mathrm{Ad}^{\natural}_{R,v,\eta,a} \Sigma^{(5)}_{\tau_0,P,G} = \Sigma^{(5)}_{0,e_3,ge_1}$, with $g \in \mathbb{R}_+$. So the orbtis are indexed by $g \in \mathbb{R}_+$.

III. (a) $h_{C_{l_3,0,0}} = \{(0,u,0,x) | u,x \in \mathbb{R}^3\}$; then $H_{C_{l_3,0,0}} = \{(1,v,0,a) | a,v \in \mathbb{R}^3\}$ is a closed subgroup of $\mathscr{G}^{\uparrow}_{+}$.

(b) $h_{C_{0,0,e_3}} = \{(\alpha l_3,0,0,x) | x \in \mathbb{R}^3, \alpha \in \mathbb{R}\}$; then $H_{C_{0,0,e_3}} = \{(R(\varphi,e_3),0,0,a) | a \in \mathbb{R}^3, \varphi \in [0,2\pi)\}$, where $R(\varphi,\nu)$ is the rotation of angle φ around the vector ν in the direct sense. $H_{C_{0,0,e_3}}$ is a closed subgroup of $\mathscr{G}^{\uparrow}_{+}$.

(c) $h_{C_{sl_3,0,e_3}} = \{a \cdot l, -sa, 0, x\}$; then $H_{C_{sl_3,0,e_3}} = \{R(\varphi,\nu), -s\varphi v, 0, a\}$ is a closed subgroup of $\mathscr{G}^{\uparrow}_{+}$.

(d) $h_{C_{sl_3,e_3,0}} = \{(\alpha l_3, -s\rho(\alpha + \beta)e_3, 0, \beta e_3)\}$;

then

$$H_{C_{sl_3,e_3,0}} = \{R(e_3,\varphi), -s\rho(\varphi + v)e_3, 0, ve_3)\}.$$

is a closed subgroup of $e\mathscr{G}^{\uparrow}_{+}$.

(e) $h_{C_{0,e_3,ge_1}} = \{(1, v_2 e_2 + v_3 e_3, 0, v_3(ge_1 - (1/\rho)e_3))\}$;

then

$$H_{C_{0,e_3,ge_1}} = \{(0, \gamma_2 e_2 + \gamma_3 e_3, 0, \gamma_3(ge_1 - (1/\rho)e_3))\}$$

is a closed subgroup of $\mathscr{G}^{\uparrow}_{+}$.

It follows that all the orbits are regular.

IV. The formulas for the sympletic forms are very complicated and not very illuminating so will not be given.

E. Physical equivalence

For the purpose of clarifying the notion of physical equivalence we analyze now the transitive action of $\mathscr{G}^{\uparrow}_{+}$ in the case $ag = 1$, $\forall g \in G$. From (3.1) we get the most general element from $Z^2(\mathrm{Lie}\, \mathscr{G}^{\uparrow}_{+}, \mathbb{R})$ as:

$$C_{\tau,P,G,m}((a,u,t,x),(a',u',t',x')) = m(u \cdot x_1 - u' \cdot x) - \tfrac{1}{2} \mathrm{tr}\, \tau[a,a'] - P \cdot (ax' - a'x - tu' + t'u) - G \cdot (au' - a'u).$$

The elements in $P(Z^2(\mathrm{Lie}\, G_+, \mathbb{R}))$ are then of the following form:

(a) $\Sigma^{(1)}_{\tau_0,P,G} = \{\lambda C_{\tau_0,P,G,0} | \lambda \in \mathbb{R} \setminus \{0\}\}$,

with $\mathrm{Tr}\, \tau_0^2 = 2$,

(b) $\Sigma^{(2)}_{\tau_0,P,G} = \{\lambda C_{\tau_0,P,G,1} | \lambda \in \mathbb{R} \setminus \{0\}$.

The orbits in the first case are analyzed symilarily to the orbits in Sec. III D. In the second case, one can prove that there is an orbit generated by the element $\Sigma^2_{0,0,0}$, and a family of orbits parametrized by $s \in \mathbb{R}_+$ and generated by $\Sigma^{(2)}_{sl_3,0,0,1}$.

Recalling that s can be interpreted as the value of the spin and m as the mass of the system, it follows that in classical physics, the physical situations are indexed by the ratio s/m.

Remark: Similar results hold for the proper orthochronous Poincaré group.

IV. CONCLUSIONS

We have generalized the concept of symmetry in classical mechanics in the spirit of Lee Hwa Chung theorem. If this theorem is the analog of the Wigner theorem in quantum mechanics, the analysis from this paper can be thought of as some generalization of the Bargmann study of projective representations in quantum mechanics.[12] The condition $H^1(\mathrm{Lie}\, G, \mathbb{R}) = 0$ from the theorem in Sec. II is to be compared with the condition of Bergmann $H^2(\mathrm{Lie}\, G, \mathbb{R}) = 0$.

The results from Sec. III E can be interpreted by remarking that in classical physics one does not have a preferred unit of mass.

We lack a physical interpretations of the parameters characterizing the representation $g \mapsto a_g$. It would be interesting to find the analog of this phenomenon in Lagrangian theory (which is probably connected with the fact that if L is a Lagrangian and $a \in \mathbb{R} \setminus \{0\}$, then L and aL give the same Euler–Lagrange equations). This will be done in a subsequent paper.

Also it must be noted that, although one of the motivation for enlarging the concept of symmetry was a closer analogy with quantum mechanics, it is not evident that there are quantum analogs of the classical systems described in Secs. III A–III D.

From the technical point of view, it would be convenient to have some regularity criteria of the type of Kostand, Souriau, and Chu (see Ref. 6). We make a last comment connected with the existence of the momentum map. The usual definition[6], is the following wing. Let X_ξ be the vector field associated with $\xi \in \mathrm{Lie}\, G$. Then it is easy to prove that

$$d i_{X_\xi} \Omega = \dot{a}(\xi)\Omega.$$

If $\dot{a}(\xi) = 0$ then there exists at least locally $f_\xi \in \mathscr{F}(X)$ such that

$$i_{X_\xi} \Omega = df_\xi. \tag{4.1}$$

Then one defines the moment map $\Phi: X \to (\text{Lie } G)^*$ by the formula

$$\phi(x)(\xi) = f_\xi(x).$$

We see that in the general case we cannot define f_ξ by (4.1) so there is no momentum map in the general case.

ACKNOWLEDGMENT

The author wishes to express his thanks to T. O. Popp for reading the manuscript and suggesting many improvements.

Note added in proof. After the completion of this paper, the author became aware that the generalization of the concept of symmetry from Sec. I, has also been proposed by G. M. Marle[13] but was exploited in the spirit of Souriau [*Structure des Systems Dynamiques* (Dunod, Paris, 1970)].

[1] L. H. Chung, Proc. R. Soc. Edinburgh, Sect. A, **LXII**, 237 (1947).
[2] F. Gantmacher, *Lectures in Analytical Mechanics* (Mir, Moscow, 1970).
[3] E. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[4] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand Reinhold, New York, 1968), Vol. I.
[5] M. Pauri and G. M. Prosperi, J. Math. Phys. **16**, 1503 (1975).
[6] V. Guillemin and S. Sternberg *Symplectic Techniques in Physics* (Cambridge U. P., London, 1984).
[7] M. Pauri and G. M. Prosperi, J. Math. Phys. **7**, 366 (1966).
[8] V. S. Varadaranjan, *Geometry of Quantum Theory* (Van Nostrand Reinhold, New York, 1970), Vol. II.
[9] M. Goto and F. D. Grosshans *Semisimple Lie Algebras* (Dekker, New York, 1978).
[10] S. Kobayashi and K. Nomizu *Foundation of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[11] J. M. Levy-Leblond, Commun. Math. Phys. **12**, 64 (1969).
[12] V. Bargmann, Ann, Math. **59**, 1 (1954).
[13] G. M. Marle, in *Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, The Netherlands, 1976).

Localizability and covariance in analytical mechanics

D. R. Grigore
Department of Theoretical Physics, Central Institute of Physics, Department of Theoretical Physics, P. O. Box MG-6, Bucharest, Romania

(Received 17 November 1987; accepted for publication 24 May 1989)

Two familiar concepts in quantum mechanics, namely localizability and covariance, are investigated in the framework of classical mechanics. The analysis of such well-known problems in quantum mechanics as localizability of elementary relativistic particles gives the same result in classical mechanics as well.

I. INTRODUCTION

In quantum mechanics one describes localizable systems in the following way.[1] Suppose the configuration space of the system is the manifold Q.

In quantum mechanics an observable is a projector-valued measure (see Ref. 2). One supposes that Q has a natural Borel structure β, and one has a projection-valued measure $\beta \ni E \mapsto P_E \in \mathcal{P}(\mathcal{H})$, where $\mathcal{H}$ is the Hilbert space of the problem and $\mathcal{P}(\mathcal{H})$ is the set of orthogonal projectors in $\mathcal{H}$, such that the states in the range of P_E are interpreted as localized in E. A more interesting situation is when a symmetry Lie group G acts on Q. Then according to the usual description of the symmetries in quantum mechanics one has a projective unitary representation of G acting in $\mathcal{H}$: $g \mapsto U_g$. A natural compatibility condition is then to require that $\forall E \in \beta$, $\forall g \in G$ one has

$$U_g P_E U_g^{-1} = P_{g \cdot E} . \tag{1.1}$$

Here $Q \ni q \mapsto g \cdot q \in Q$ is the action of G on Q and $g \cdot E$ is the image of $E \in \beta$ under g. This relation has a simple physical interpretation: if one has two observers connected by the transformation $g \in G$, then if the first one sees the system in E, the second one must see the system in $g \cdot E$. One usually says in this case that the system is covariant with respect to the group G. The couple (U,P) satisfying to (1.1) is called a system of imprimitivity. A system of imprimitivity is called irreducible if the only subspace in $\mathcal{H}$ invariant with respect to all U's and all P's are $\mathcal{H}$ and $\{0\}$. Two systems of imprimitivity (U,P) and (U',P') are called equivalent if there exists a unitary operator W such that $U'_g = W U_g W^{-1}$ $(\forall g \in G)$, and $P'_E = W P_E W^{-1}$ $(\forall E \in \beta)$. There are two interesting problems in quantum mechanics connected with these concepts: (1) Given the G-space Q, find all the associated systems of imprimitivity; (2) Suppose one has a unitary representation U of the proper Euclidian group SE(3); then find all the projector-valued measures P based on $\mathbb{R}^3$ such that (U,P) is a system of imprimitivity. In particular, U can be obtained by appropriately restricting an irreducible projective representation of a larger group containing SE(3) as a subgroup (e.g., the proper Galilei group $\mathcal{G}^{\uparrow}_{+}$, or the proper orthochronous Poincare group $\mathcal{P}^{\uparrow}_{+}$). The first problem for $Q = \mathbb{R}^n$ and $G = T(n)$ ($\equiv$ the translation group in n dimensions acting naturally on Q) is essentially solved by the Stone–von Neumann theorem.[1] The second problem was proposed for $G = \mathcal{P}^{\uparrow}_{+}$ by Newton and Wigner[3] and rigorously stated and proved by Wightman[4] (see also Ref. 1 where the case $G = \mathcal{G}^{\uparrow}_{+}$ is treated). The results of this analysis are the following: Elementary relativistic particles of nonzero mass are localizable; on the contrary, elementary relativistic particles of zero mass are localizable only for zero helicity. Elementary Galilean particles of nonzero mass are also localizable.[1]

For more recent work on this subject see Refs. 5, 6 and literature cited there.

In this paper we want to investigate the similar problems in analytical mechanics. We adopt the following point of view: (i) A classical system is described by a symplectic manifold (M,ω); (ii) for a system covariant with respect to a group G, the group must act on M by canonical transformations; (iii) for an elementary system, M must be homogeneous with respect to the action of G (Refs. 7–13).

We have to find the analogs of the concepts of localizability and covariance in this framework. We start from the observation that the basic motivation for the introduction of the notion of a symplectic manifold is the fact that for a given configuration space Q, the cotangent bundle $T^*(Q)$ has a natural symplectic structure such that the time evolution is a canonical transformation. If we do not want to lose this starting point we think that (i) above must be supplemented by the requirement that M must be fibered over a base manifold of the type $T^*(Q)$, together with some other natural compatibility conditions which will be described in detail in Sec. II. In this way one gets a natural notion of localizability, i.e., for every state $m \in M$ one has a configuration $x(m) \in Q$.

A related problem was studied in Refs. 9–11 where one demands essentially that for every state $m \in M$ one has an associated universe line, i.e., a straight line in $\mathbb{R}^4$ (= the Minkowski space). Also, in Ref. 12, (i)–(iii) above are supplemented by demanding the existence of an evolution space, which is essentially a quadridimensional generalization of the notion of configuration space. Another related definition of localizability can be found in Ref. 14.

In Sec. III we investigate the analogs of Stone–von Neumann and Wightman theorems. From the physical point of view it is interesting to note that in analytical mechanics the problem of localizability has the same answer as in quantum mechanics. The analysis produces well-known formulas from the physical literature[15–17] in a more systematic fashion, and exhibiting the underlying structure of localizability.

The last section is devoted to some comments.

 0022-2488/89/112646-07$02.50

II. LOCALIZABILITY AND COVARIANCE IN ANALYTICAL MECHANICS I

A. Localizability

In analytical mechanics one describes a system by a symplectic manifold (M,ω); ω is the symplectic form on the manifold M ($\equiv$ the phase space). For a particle with configuration space Q, the usual prescription of the Hamiltonian formalism is to take $M(Q)\equiv T^*(Q)$ ($\equiv$ the cotangent bundle over Q) and $\omega(Q)$ given in local coordinates by

$$\omega(Q)_{q,p}=\sum_i dp_i\wedge dp_i\,,\quad q\in Q\,,\quad p\in T_q^*(Q)\,; \tag{2.1}$$

$(q_i,\ p_i)$ is a chart around $(q,p)\in M(Q)$. Equivalently, the Poisson bracket $\{\ ,\ \}_Q$ is

$$\{f_1,f_2\}_Q=\sum_i\frac{\partial f_1}{\partial q_i}\frac{\partial f_2}{\partial p_i}-(1\leftrightarrow 2)\,. \tag{2.2}$$

One can take this as a definition of localizability, i.e., the system (M,ω) is called localizable if there exists Q such that $M\simeq T^*(Q)$ and ω is given by (2.1). We think that this point of view artificially excludes systems with spin, so we propose the following generalization.

Definition: We call the system (M,ω) localizable if there exists a nontrivial manifold Q such that M is a fiber bundle over $T^*(Q)$ and the symplectic structure verifies the following condition: Let $\{\ ,\ \}$ be the Poisson bracket associated with ω, and π: $M\to T^*(Q)$ the canonical projection of the bundle. If $f_1,f_2\in\mathcal{F}[T^*(Q)]$ [$\equiv$ smooth functions on $T^*(Q)$] then we demand that

$$\{f_1\circ\pi\,,\ f_2\circ\pi\}=\{f_1,f_2\}_Q\circ\pi\,, \tag{2.3}$$

where $\{\ ,\ \}_Q$ is given by (2.2).

Remarks: (1) Q is not uniquely determined by this condition.

(2) For $M=T^*(Q)$ we recover the previous definition.

(3) In the general case, the fibers of M describe internal degrees of freedom.

(4) Relation (2.3) tells, roughly speaking, that if two observables do not depend on the internal degrees of freedom, then their Poisson bracket must be calculated as in the simple case $M=T^*(Q)$ in local coordinates.

(5) A related definition of localizability can be found in Ref. 14, namely that M is fibered over Q and the fibers are Lagrange submanifolds.

B. Covariance

We turn now to the notion of covariance. Suppose as in Sec. I that G is a Lie group which acts smoothly on Q: $G\times Q\ni g,\ q\to g\cdot q\in Q$. Suppose also that G is a symmetry group of (M,ω), i.e., acts canonically on M: $G\ni g\to\phi_g\in\mathrm{Can}(M)$. Denote by π_0: $T^*(Q)\to Q$ the canonical projection in $T^*(Q)$ and define x: $M\to Q$ by $x\equiv\pi_0\circ\pi$. Then we think that the proper equivalent of covariance with respect to G is to require that x is a G morphism, i.e., commutes with the action of the group:

$$x\circ\phi_g(m)=g\cdot x(m)\,,\quad\forall m\in M\,,\ \forall g\in G\,. \tag{2.4}$$

Because in analytical mechanics the observables are elements of $\mathcal{F}(M)$ ($\equiv$ the smooth real-valued functions defined on M), one can think of x: $M\to Q$ as a Q-valued observable. If $m\in M$ is a state of the system then $x(m)\in Q$ is interpreted as the configuration of the system in the state m.

Definitions: If the G space Q is given we call the symplectic manifold (M,ω) a (classical) system of imprimitivity based on Q with respect to G, if G acts canonically on M (M,ω) is localizable in the sense of A, and x: $M\to Q$ verifies the condition (2.4). Two systems of imprimitivity (M,ω) and (M',ω') are called equivalent if there exists a G diffeomorphysm α: $M\to M'$ such that $\alpha^*\omega'=\omega$, and $\pi=\pi'\circ\alpha$.

C. Localizability on $\mathbb{R}^n$

Let us study the analog of problem (1) from Sec. I, in the case $Q=\mathbb{R}^n$, $G=T(n)$, i.e., the condition appearing in the statement of the Stone–von Neumann theorem. The problem is to classify, up to equivalence, the systems of imprimitivity based on $\mathbb{R}^n$, with respect to $T(n)$. We indicate a possible solution based on an analogy with the Weyl system. Let $\{\ ,\ \}$ be the Poisson bracket on M associated with ω. Let $p_1,\ldots,p_n$: $\mathbb{R}^n\to\mathbb{R}$ be the canonical projections in $\mathbb{R}^n$, and define $x_i\in\mathcal{F}(M)$, $i=\overline{1,n}$ by $x_i\equiv p_i\circ x$. Then from (2.2) and (2.3) it follows that

$$\{x_i,x_j\}=0\,,\quad i,j=\overline{1,n}\,. \tag{2.5}$$

We denote by ψ^i: $\mathbb{R}\to\mathrm{Can}(M)$ the one-parameter group of canonical transformation generated by x_i ($i=\overline{1,n}$). Because of (2.5), one knows that ψ^i commute among themselves.

Let now $T(n)\ni a\to\phi_a\in\mathrm{Can}(M)$ be the action of $T(n)$ on the phase space M. From (2.4) we have

$$x\circ\phi_a(m)=x(m)+a\,,\quad\forall a\in\mathbb{R}^n\,,\quad\forall m\in M\,. \tag{2.6}$$

From this, one can prove that ϕ commutes with ψ^i for any $i=\overline{1,n}$. Then

$$\mathbb{R}^n\times\mathbb{R}^n\ni a,b\to\phi_a\circ\psi^1_{b_1}\circ\cdots\circ\psi^n_{b_n}\in\mathrm{Can}(M) \tag{2.7}$$

is a canonical action of $T(2n)$ on M. We are reminded now that the Stone–von Neumann theorem is proved by showing that a general system of imprimitivity is a direct sum of irreducible ones and then classifying the irreducible systems of imprimitivity. If we want to implement this line of argument in analytical mechanics we must first find the equivalent of irreducibility. First, we recall that the irreducibility of a system of imprimitivity is equivalent with the irreducibility of the corresponding projective representation of $T(2n)$, i.e., of the Weyl system associated with (U,P). Second, in Refs. 8–12 it is argued that the concept of elementary of a classical system must be modeled mathematically by demanding the transitivity of the action of the symmetry group of the system on the phase space. So, in some sense, the classical counterpart of an irreducible linear representation in quantum mechanics is the transitivity of the group action on the phase space in classical mechanics.

Guided by this argument we say the (classical) system of imprimitivity based on $\mathbb{R}^n$ with respect to $T(n)$ from above is irreducible if the action (2.7) is transitive. This action can be called the (classical) Weyl system of imprimitivity of the system. Two systems of imprimitivity (M,ω) and (M',ω') based on $\mathbb{R}^n$ with respect to $T(n)$ are equivalent if

there exists a diffeomorphism $\alpha: M \to M'$ such that $\alpha^*\omega' = \omega$ and α commutes with the action of the (classical) Weyl system. (This follows easily from the definition of equivalence given in Sec. II B.)

Unfortunately there seems to exist no natural analog to the notion of a direct sum of systems of imprimitivity, so in Sec. III we will analyze only (classical) irreducible systems of imprimitivity based on $\mathbb{R}^n$ with respect to $T(n)$.

D. Localizability for relativistic systems

The classical counterpart of problem (2) in Sec. I is the following one. One has a symmetry group G which includes SE(3) as a subgroup and acts canonically on (M,ω). By restriction we have a canonical action of SE(3) on (M,ω). Then we want to decide if the system admits a notion of localizability in the sense of A with $Q = \mathbb{R}^3$, and also is covariant with respect to SE(3) in the sense (2.4); SE(3) is supposed to act naturally on $\mathbb{R}^3$ by the following formula:

$$(R,a)q = Rq + a\,. \tag{2.8}$$

Here $R \in SO(3)$, $q,a \in \mathbb{R}^3$. In this case (2.4) becomes

$$x \circ \phi_{R,a}(m) = Rx(m) + a\,, \quad \forall m \in M\,, \quad \forall (R,a) \in SE(3)\,. \tag{2.9}$$

Let us indicate now one possible way to reformulate this condition, which is important in practical applications.

Because $H^1(\text{Lie SE}(3)) = 0$ and $H^2(\text{Lie SE}(3)) = 0$, one knows (see e.g., Ref. 13) that there exists the moment map, i.e., for any $\xi \in \text{Lie SE}(3)$, there exists $f_\xi \in \mathcal{F}(M)$ which is the Hamiltonian function for the one-parameter subgroup of canonical transformation $t \to \phi_{\exp t\xi}$.

In particular, if $l_i \in \text{Lie SE}(3)$ is the infinitesimal generator of the rotation around the axis $e_i \in \mathbb{R}^3$, and $p_i \in \text{Lie SE}(3)$ is the infinitesimal generator of the translation along e_i, then we have from (2.9) and the definition of the Poisson bracket:

$$\{x_i,P_j\} = \delta_{ij}\,,$$
$$\{J_i,x_j\} = \sum_{k=1}^{3} \epsilon_{ijk} x_k\,, \quad (i,j = 1,2,3)\,, \tag{2.10}$$

where $P_i \equiv f_{p_i}$ and $J_i \equiv f_{l_i}$. Also we have as in Sec. II C, (2.5).

In this infinitesimal form the problem (2) was recognized and solved in Ref. 16 for $G = \mathcal{G}^\uparrow_+$, and in Ref. 17 for $\mathcal{P}^\uparrow_+$ in the hypothesis that the canonical action $g \to \phi_g$ corresponds to particles of nonzero mass. Also in Ref. 17 it is argued that if in M one has an action of the full Poincare group $\mathcal{P}$ (with the elements in $\mathcal{P}^\downarrow$ realized by anticanonical transformation), then (2.9) must be supplemented with

$$x \circ \phi_{I_s} = -x\,, \quad x \circ \phi_{I_t} = x \tag{2.11}$$

(I_s is the spatial inversion, and I_t the temporal inversion).

The method used in Refs. 16, 17 is to determine if there exist solutions for x_i of the system (2.5) + (2.10). Then one tries to see if one can find a set of coordinates on M such that x and P are among them. In this case, the localizability emerges. On the contrary, if (2.5) + (2.10) has no solution, then the system is not localizable.

III. LOCALIZABILITY AND COVARIANCE IN ANALYTICAL MECHANICS II

A. Homogeneous symplectic manifolds

To analyze further the two problems stated in Sec. II we need the theory of Kostant–Souriau–Kirillov. We essentially follow Ref. 13. Let G be a Lie group and Lie $G \equiv T_e(G)$ the associated Lie algebra. Using standard cohomological notations, let $\sigma \in Z^2(\text{Lie } G,\mathbb{R})$. Then one knows that $h_\sigma \subset \text{Lie } G$ given by

$$h_\sigma \equiv \{\xi \in \text{Lie } G \mid i_\xi \sigma = 0\}$$

is a Lie subalgebra of Lie G. Let $H_\sigma \subset G$ be the associated connected Lie subgroup of G. If H_σ is a closed subgroup of G, then $\sigma \in Z^2(\text{Lie } G,\mathbb{R})$ is called regular. If $\sigma \in Z^2(\text{Lie } G,\mathbb{R})$ is a regular element, then G/H_σ can be given a differential structure. Moreover, using the so-called reduction principle, one can obtain an unique symplectic form on G/H_σ which is invariant with respect to the natural action of G on G/H_σ. Let $\Sigma \subset Z^2(\text{Lie } G,\mathbb{R})$ be an orbit with respect to the coadjoint action of G in $\Lambda^2(\text{Lie } G)^*$. Then if $\sigma_0 \in \Sigma$ is regular, any $\sigma \in \Sigma$ is regular; in this case the orbit Σ is called regular. Moreover if $\sigma_1,\sigma_2 \in \Sigma$, Σ regular, then G/H_{σ_1} and G/H_{σ_2} are diffeomorphic symplectic manifolds. For $\sigma \in Z^2(\text{Lie } G,\mathbb{R})$, $G_\sigma \subset G$ denotes the stability subgroup of σ with respect to the coadjoint action. Let M be a homogeneous manifold, i.e., $M \simeq G/H$ with $H \subset G$ a closed subgroup of G. Then there exists a regular $\sigma \in Z^2(\text{Lie } G,\mathbb{R})$ such that $H_\sigma \subset H \subset G_\sigma$, and G/H_σ is a covering of M. It follows that, up to covering, the symplectic homogeneous manifolds of G are parametrized by regular orbits of the coadjoint action of G in $\Lambda^2(\text{Lie } G)^*$. Of course, it is sufficient to take a single element σ from every regular orbit.

B. Localizability on $\mathbb{R}^n$

With the theory outlined in the above subsection we can now easily solve the problem in Sec. II C. As found there, we must classify the transitive actions of the classical Weyl system, i.e., of the auxiliary $T(2n)$ group. First we have to study $Z^2(\text{Lie } T(2n),\mathbb{R})$. We identify Lie $T(2n)$ with $T(2n)$. Then every two-cochain $\sigma \in \Lambda^2 T(2n)$ is a two-cocycle. The group $T(2n)$ acts trivially on $\Lambda^2 T(2n)$, so every two-cochain is an orbit. From A it follows that the homogeneous symplectic manifolds of $T(2n)$ are of the form $M = (T(2n)/h_\sigma,\hat{\sigma})$, where $\hat{\sigma}$ is the factorized of σ to the factor space $T(2n)/h_\sigma$ by the reduction principle. By a simple dimensionality argument we see that M can be a fiber bundle over $T^*(\mathbb{R}^n) \simeq \mathbb{R}^{2n}$ only if σ is nondegenerate. So, the symplectic homogenenous manifolds of $T(2n)$, which are of interest for the problem Sec. II C, are of the form $(\mathbb{R}^{2n},\sigma)$ where σ is an arbitrary nondegenerate two-form on $\mathbb{R}^n$. From (2.3) it follows that σ must be equal with $\omega(Q)$ given by (2.1), with $q_1,\ldots,q_n$, $p_1,\ldots,p_n$ as global coordinates on $\mathbb{R}^{2n}$. The canonical action of $T(n)$ on $\mathbb{R}^{2n}$ is

$$\phi_a(q,p) = (q + a,p)\,.$$

We have also

$$x(q,p) = q\,.$$

C. Coadjoint orbits

To analyze the problem in Sec. II D for $G=\mathscr{P}^{\uparrow}_{+}$ we need a particular but important case of the analysis in A, namely when $H^1(\text{Lie } G,R)=0$ and $H^2(\text{Lie } G,R)=0$. In particular, for $G=\mathscr{P}^{\uparrow}_{+}$ this is true, as a consequence of Whitehead lemmas.[13] In this case, it is known (see Ref. 13, Sec. 26) that there exists the moment map. Also, one knows (see Ref. 13, Prop. 25.2), that in this case, G/H_σ covers an orbit from (Lie G)* with respect to the coadjoint action of G. The symplectic structure on such orbits can be given as follows.

Let C^k_{ij} be the constants of structure of the Lie group G, with respect to the basis $\{\xi_i\}_{i=1}^r$ from Lie G. Let $\{\beta_i\}_{i=1}^r$ be the dual basis from (Lie G)*. Then one has in (Lie G)* a Poisson structure, which is generally degenerate, given by the formula

$$\{F_1,F_2\}(\beta)=\sum_{i,j,k=1}^{r} C^k_{ij}\beta_k\frac{\partial F_1}{\partial\beta_i}\frac{\partial F_2}{\partial\beta_j}. \tag{3.1}$$

By restriction to an orbit $\mathscr{O}\subset(\text{Lie } G)^*$, this Poisson structure becomes nondegenerate and gives the symplectic structure of $\mathscr{O}$. We denote it by $\{\ ,\ \}_{\mathscr{O}}$. This can be proved as follows. One knows (see Ref. 13, Prop. 24.1) that for $H^1(\text{Lie } G,\mathbb{R})=0$ and $H^2(\text{Lie } G,\mathbb{R})$, one can choose the Hamiltonian functions f_ξ such that

$$\{f_\xi,f_\eta\}=f_{[\xi,\eta]},\quad \forall\xi,\eta\in\text{Lie } G. \tag{3.2}$$

Then (3.1) is valid for $F_1=f_\xi$ and $F_2=f_\eta$, as can be shown by direct computation. Using now the properties of the Poisson bracket, (3.1) is extended to polynomials from $\mathscr{F}(\mathscr{O})$, and after that to all $\mathscr{F}(\mathscr{O})$ by continuity and density. So the Poisson bracket on $\mathscr{O}$ can be found as follows. If $f_1,f_2\in\mathscr{F}(\mathscr{O})$, take any smooth extensions to (Lie G)*, F_1, and F_2, respectively. Then

$$\{f_1,f_2\}(\beta)=\sum_{i,j,k=1}^{r} C^k_{ij}\beta_k\frac{\partial F_1}{\partial\beta_i}(\beta)\frac{\partial F_2}{\partial\beta_j}(\beta). \tag{3.3}$$

This form of the symplectic structure can be found also in Refs. 15–17.

D. Elementary systems for $\mathscr{P}^{\uparrow}_{+}$

For $G=\mathscr{P}^{\uparrow}_{+}$, the results are the following[13,15,17]: We identify $\mathscr{P}^{\uparrow}_{+}$ with 5×5 real matrices of the form

$$(\Lambda,a)\equiv\begin{pmatrix}\Lambda & a\\ 0 & 1\end{pmatrix},\quad \Lambda\in\mathscr{L}^{\uparrow}_{+},\quad a\in\mathbb{R}^4.$$

Then Lie $\mathscr{P}^{\uparrow}_{+}$ can be identified with the real subspace of 5×5 real matrices of the form

$$(A,x,u,t)\equiv\begin{pmatrix}0 & u^t & t\\ u & A & x\\ 0 & 0 & 0\end{pmatrix},$$

where $t\in\mathbb{R}$, $u,x\in\mathbb{R}^3$, and A is a real 3×3 antisymmetric matrix. We take in Lie $\mathscr{P}^{\uparrow}_{+}$ the following basis:

$$j_i\equiv(l_i,0,0,0),\quad k_i\equiv(0,0,e_i,0),$$
$$p_i\equiv(0,e_i,0,0),\quad h\equiv(0,0,0,1),$$

where $(l_i)_{jk}\equiv-\epsilon_{ijk}$ and $(e_i)_j=\delta_{ij}$. Then (Lie $\mathscr{P}^{\uparrow}_{+}$)* can be identified with $\mathbb{R}^{10}\cong\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}$ because a generic element in (Lie $\mathscr{P}_{+}$)* is

$$\langle(J,P,K,H),\ (\alpha\cdot j,\ x,u,t)\rangle = J\cdot\alpha+P\cdot x+K\cdot u+Ht.$$

Here "$\cdot$" is the Euclidian scalar product. Computing the coadjoint action one can prove that there are only two functionally independent Cazimir invariants, namely H^2-P^2 and $W_0^2-W^2$, where $(W_0,W)\in\mathbb{R}^4$ is the Pauli–Liubanski quadrivector given by $W_0\equiv J\cdot P$, $W=HJ+P\Lambda K$ (Λ is the vector product); for $H^2\geqslant P^2$, sign H is also an invariant. From the physical point of view we are interested only in this case, when the orbits are the following.[13]

(I) $\mathscr{O}^{\pm}_{m,s}\equiv\{(J,P,K,H)\in\mathbb{R}^{10}\,|\,H^2-P^2=m^2,\ W^2-W_0^2=m^2s^2,H\gtrless0\}$.

Here $m\in\mathbb{R}_+$ and $s\in\mathbb{R}_+\cup\{0\}$; m is called the mass and s is called the spin of the system.

(II) $\mathscr{O}^{\pm}_{c}\equiv\{(J,P,K,H)\in\mathbb{R}^{10}\,|\,H^2=P^2,\ W^2-W_0^2=c,\ H\gtrless0\}$

with $c\neq0$.

Remark: In fact $c>0$.

(III) $\Omega^{\pm}_{h}\equiv\{(J,P,K,H)\in\mathbb{R}^{10}\,|\,H^2=P^2,\ P\neq0,\ W=hP,\ H\gtrless0\}$;

$h\in\mathbb{R}$ is called helicity.

In all cases the Poisson bracket is obtained according to (3.1) by restricting, to the desired orbit, the following degenerate Poisson structure: if F_1,F_2: $\mathbb{R}^{10}\to\mathbb{R}$ are smooth, then

$$\begin{aligned}\{F_1,F_2\} &= \sum_{i,j,l=1}^{3}\epsilon_{ijl}J_l\left(\frac{\partial F_1}{\partial J_i}\frac{\partial F_2}{\partial J_j}-\frac{\partial F_1}{\partial K_i}\frac{\partial F_2}{\partial K_j}\right)\\ &+\sum_{i,j,l=1}^{3}\epsilon_{ijl}\left[K_l\left(\frac{\partial F_1}{\partial J_i}\frac{\partial F_2}{\partial K_j}-(1\leftrightarrow2)\right)\right.\\ &\left.+P_l\left(\frac{\partial F_1}{\partial J_i}\frac{\partial F_2}{\partial P_j}-(1\leftrightarrow2)\right)\right]\\ &+H\sum_{i=1}^{3}\left(\frac{\partial F_1}{\partial K_i}\frac{\partial F_2}{\partial P_i}-(1\leftrightarrow2)\right)\\ &+\sum_{i=1}^{3}P_i\left(\frac{\partial F_1}{\partial K_i}\frac{\partial F_2}{\partial H}-(1\leftrightarrow2)\right).\end{aligned} \tag{3.4}$$

E. Localizability for Poincaré covariant systems

We turn now to the question of localizability. Up to a point, the analysis is identical with that in Ref. 17 and will be done briefly. The cases I and II can be analyzed simultaneously. We observe first that for I and II, $H\neq0$ so we have

$$J=(1/H)(W-P\Lambda K). \tag{3.5}$$

This suggests to take as independent coordinates

$$(W,P,K,H)\in\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}.$$

Then we have

$$\begin{aligned}(\text{I})\ \mathscr{O}^{\pm}_{m,s}=\{&(W,P,K,H)\\ &\times\in\mathbb{R}^{10}\,|\,H^2-P^2=m^2,\\ &W^2-(P\cdot W)^2/H^2=m^2s^2,\ H\gtrless0\}.\end{aligned}$$

(II) $\mathcal{O}_c^{\pm} \equiv \{(W,P,K,H)$
$\times \in \mathbb{R}^{10} | H^2 = P^2,$
$W^2 - (P\cdot W)^2/H^2 = c, \quad H \gtrless 0\}$.

The Poisson bracket (3.4) becomes in this new coordinates

$$\{F_1,F_2\} = \sum_{i,j,l=1}^{3} \epsilon_{ijl}\left[\left(HW - \frac{P\cdot W}{H}P\right)_l \frac{\partial F_1}{\partial W_i}\frac{\partial F_2}{\partial W_j} - \frac{1}{H}(W - P\Lambda K)_l \frac{\partial F_1}{\partial K_i}\frac{\partial F_2}{\partial K_j}\right] + \sum_{i=1}^{3}\left(H\frac{\partial F_1}{\partial K_i}\frac{\partial F_2}{\partial P_i} - \frac{P\cdot W}{H}\frac{\partial F_1}{\partial W_i}\frac{\partial F_2}{\partial K_i} - P_i\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial K_i}\right) - (1 \leftrightarrow 2). \tag{3.6}$$

We look now for three observables $x_1,x_2,x_3 \in \mathcal{F}(\mathcal{O})$ such that (2.5) + (2.10) are fulfilled. It is sufficient to find three smooth functions $X_1,X_2,X_3:\mathbb{R}^{10}\to\mathbb{R}$ such that we have

$$\{X_i,X_j\} = 0, \quad \{X_i,P_j\} = \delta_{ij};$$
$$\{J_i,X_j\} = \sum_{k=1}^{3} \epsilon_{ijk}X_k, \tag{3.7}$$

for $i,j = 1,2,3$. Here J is given by (3.5). If we find X then we can take $x \equiv X|_{\mathcal{O}}$.

The idea is to exploit first the linear relations from (3.7); namely the last two ones. The result is the following.[17] From the second relation (3.7) it follows easily that

$$X(W,P,K,H) = (K/H) + \Lambda(W,P,H), \tag{3.8}$$

where $\Lambda: \mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}\to\mathbb{R}^3$ is smooth. Then one observes that the vectors P, W and $P\Lambda W$ form a basis in $\mathbb{R}^3$ if P and W are such that (W,P,K,H) is on $\mathcal{O}_{m,s}^{\pm}$ or $\mathcal{O}_c^{\pm}$. So we can write Λ in the form:

$$\Lambda(W,P,H) = Pa_1(P,W,H) + Wa_2(P,W,H) + P\Lambda Wa_3(P,W,H),$$

where a_1,a_2,a_3: $\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}\to\mathbb{R}$ are smooth. Then by a tedious computation one can prove that the last relation (3.7) is equivalent with

$$\sum_{j,k=1}^{3}\epsilon_{ijk}\left(P_j\frac{\partial a_e}{\partial P_k} + W_j\frac{\partial a_e}{\partial W_k}\right) = 0, \quad i,e = 1,2,3.$$

This can be shown to be equivalent with the following global condition:

$$a_i(RP,RW,H) = a_i(P,W,H), \quad R\in SO(3), \quad i = 1,2,3.$$

So the functions a_i are constant on the orbits of SO(3) acting naturally on $\mathbb{R}^3\times\mathbb{R}^3$. But it is easy to prove that we have three functionally independent invariants: $P^2, W^2, P\cdot W$, which determine the orbit structure. It follows that the functions a_i depend on P and W only by these invariants. But on the orbits $\mathcal{O}_{m,s}^{\pm}$ and $\mathcal{O}_c^{\pm}$ only one is independent, e.g., $(P\cdot W/H) \equiv \alpha$. So we look for a Λ of the form

$$\Lambda(P,W,H) = PA_1(H,\alpha) + WA_2(H,\alpha) + P\Lambda WA_3(H,\alpha), \tag{3.9}$$

where $A_i: \mathbb{R}^2\to R$ are smooth. We still have the first relation (3.7). One can guess rather easily a solution of this relation for $\mathcal{O}_{m,s}^{\pm}$ as in Ref. 17. Because we want to study also $\mathcal{O}_c^{\pm}$, we proceed more systematically. If we use (3.8), (3.9), and the Poisson bracket (3.6), we get after a long but straightforward computation that

$$\{X_i,X_j\} = \sum_{k=1}^{3}\epsilon_{ijk}(u_1P + u_2W + u_3P\Lambda W)_k,$$

where u_1,u_2,u_3: $\mathbb{R}^2\to\mathbb{R}$ are given by

$$u_1 \equiv \frac{1}{H}\left[2\alpha A_3 + \alpha H\frac{\partial A_3}{\partial H} + (\alpha^2+c)\frac{\partial A_3}{\partial\alpha}\right] - H\alpha A_2\frac{\partial A_1}{\partial\alpha} - (\alpha^2+c)A_2\frac{\partial A_2}{\partial\alpha} + [\alpha^2m^2 - c(H^2-m^2)]A_3\frac{\partial A_3}{\partial\alpha} + \alpha(m^2A_3 - A_2^2),$$
$$u_2 \equiv -\frac{1}{H^3} - \frac{1}{H}\left[2HA_3 + (H^2-m^2)\frac{\partial A_3}{\partial H} + \alpha H\frac{\partial A_3}{\partial\alpha}\right] + (H^2-m^2)A_2\frac{\partial A_1}{\partial\alpha} + \alpha HA_2\frac{\partial A_2}{\partial\alpha} + HA_2^2,$$
$$u_3 \equiv -\frac{1}{H}\left(\frac{\partial A_1}{\partial\alpha} - \frac{\partial A_2}{\partial H}\right) + A_3\left[(H^2-m^2)\frac{\partial A_1}{\partial\alpha} + \alpha H\frac{\partial A_2}{\partial\alpha} + HA_2\right].$$

Because P, W and $P\Lambda W$ form a basis in $\mathbb{R}^3$, the first relation (3.7) is equivalent with

$$u_i = 0, \quad i = 1,2,3. \tag{3.10}$$

We want to eliminate $\partial A_1/\partial\alpha$ from the equation $u_3 = 0$. This is possible if

$$H(H^2-m^2)A_3 \neq 1.$$

For $m \neq 0$, this is evident: $A_3 = (1/H(H^2-m^2))$ is not defined for $H = \pm m$. For $m = 0$, we can have in principle $A_3 = (1/H^3)$. Let us investigate if this solution is possible. The system (3.10) becomes

$$\frac{\alpha}{H^3} + \alpha H^2A_2\frac{\partial A_1}{\partial\alpha} + (\alpha^2+c)HA_2\frac{\partial A_2}{\partial\alpha} + \alpha HA_2^2 = 0,$$
$$A_2\left(H\frac{\partial A_1}{\partial\alpha} + \alpha\frac{\partial A_2}{\partial\alpha} + A_2\right) = 0, \tag{3.11}$$
$$\alpha\frac{\partial A_2}{\partial\alpha} + A_2 + H\frac{\partial A_2}{\partial H} = 0.$$

We cannot have $A_2 = 0$, because the first equation would be contradicted. So we can get from the second equation

$$\frac{\partial A_1}{\partial\alpha} = -\frac{1}{H}\left(\alpha\frac{\partial A_2}{\partial\alpha} + A_2\right).$$

Inserting in the first equation (3.11) we get

$$\frac{\partial A_2^2}{\partial \alpha} = -\frac{2\alpha}{cH^4} \Leftrightarrow A_2^2 = -\frac{\alpha^2}{cH^4} + f(H),$$

where f is a smooth real function of H. Inserting this expression in the last equation (3.11) we get $f(H) = (ct/H^2)$. So we have

$$A_2^2 = \frac{ct}{H^2} - \frac{\alpha^2}{cH^4}.$$

But $c>0$ so we get a contradiction. We have proved that $H(H^2 - m^2)A_3 \neq 1$ so we can get from $u_3 = 0$,

$$\frac{\partial A_1}{\partial \alpha} = -\frac{H^2(\alpha(\partial A_2/\partial \alpha) + A_2)A_3 + \partial A_2/\partial H}{H(H^2 - m^2)A_3 - 1}. \quad (3.12)$$

Inserting this formula in $u_1 = 0$ and $u_2 = 0$ we get after some computations the following system:

$$\frac{\partial I}{\partial \beta} = -\frac{1}{H}(\beta m^2 - cH^2 + cm^2) \times [H(H^2 - m^2)A_3 - 1]\frac{\partial A_3}{\partial \beta} - \frac{F}{2H^2},$$

$$\frac{\partial I}{\partial H} = -\frac{1}{H}(\beta m^2 - cH^2 + cm^2) \times [H(H^2 - m^2)A_3 - 1]\frac{\partial A_3}{\partial H} - \frac{G}{H^3}.$$

Here $\beta \equiv \alpha^2$ and we have denoted

$$I \equiv \tfrac{1}{2}(\beta m^2 - cH^2 + cm^2)A_2^2,$$

$$F \equiv m^2H^2(H^2 - m^2)A_3 - 2mHA_3 - 1,$$

$$G \equiv \beta(m^2H^4A_3^2 + m^2HA_3 + 1) - c[2H^4(H^2 - m^2)A_3^2 - H(H^2 + m^2)A_3 - 1]. \quad (3.13)$$

A straightforward computation shows that the condition of integrability of Frobenius is satisfied for this system for any smooth function A_3. So it remains to decide if A_3 can be chosen such that the solution I is of the form (3.13). The simplest thing would be to have $A_2 = 0$. Then, from (3.12) we see that we can take also $A_1 = 0$. For $A_2 = 0$, the system (3.13) becomes

$$\frac{\partial A_3}{\partial \beta} = -\frac{F}{2HX},$$

$$\frac{\partial A_3}{\partial H} = -\frac{G}{H^2X}. \quad (3.14)$$

Again the Forbenius condition of integrability is satisfied. It is easy to see that we can get two solutions independent of β:

$$(A_3)_{\pm} = \mp(1/mH(H \pm m)). \quad (3.15)$$

These are smooth functions for $H>0$ and $H<0$, respectively, for $m \neq 0$. Collecting these results we have from (3.9) the following solution of (3.7) for the case I:

$$X_{\pm}(W,P,K,H) = K/H \mp (P \wedge W/mH(H \pm m)). \quad (3.16)$$

To prove the localizability in the sense of Sec. II A one must exhibit also the fiber bundle structure. We proceed as follows.[15] We define the orbital angular moment:

$$L_{\pm} \equiv X_{\pm} \wedge P,$$

and observe that for $s = 0$, $L_{\pm} = J$. Taking as new variables X and P we get $\mathcal{O}^{\pm}_{m,0} \cong \mathbb{R}^3 \times \mathbb{R}^3$, and by computing the Poisson bracket we get

$$\{f_1, f_2\} = \sum_{i=1}^{3} \frac{\partial f_1}{\partial X_i}\frac{\partial f_2}{\partial P_i} - (1 \leftrightarrow 2),$$

so in this case $M = T^*(Q)$ (see Sec. II A). If $s>0$, then we define the intrinsic angular moment:

$$S_{\pm} \equiv J - L_{\pm}.$$

We get easily that $S_{\pm}^2 = s^2$. If we take as new independent coordinates X, P, and S, we have

$$\mathcal{O}^{\pm}_{m,s} \equiv \{(X,P,S) \in \mathbb{R}^9 | S^2 = s^2\},$$

and the Poisson bracket is

$$\{f_1, f_2\} = \sum_{i=1}^{3} \frac{\partial F_1}{\partial X_i}\frac{\partial F_2}{\partial P_i} - (1 \leftrightarrow 2) + \sum_{i,j,k=1}^{3} \epsilon_{ijk}S_k \frac{\partial F_1}{\partial S_i}\frac{\partial F_2}{\partial S_j},$$

where F_1, F_2 are two smooth extensions of f_1 and f_2, respectively. In this case π: $M \to T^*(Q)$ is simply $\pi(X,P,S) = (X,P)$. The formulas for the action of $\mathcal{P}^{\uparrow}_{+}$ on M in these two cases can be found, e.g., in Ref. 15. So in case (I), the system is localizable.

For the case (II), we return to the system (3.13). Putting $m = 0$, we get

$$\frac{\partial I}{\partial \beta} = cH(H^3A_3 - 1)\frac{\partial A_3}{\partial \beta} + \frac{1}{2H^2},$$

$$\frac{\partial I}{\partial H} = c\frac{H^3A_3 - 1}{H^2}\left(H^3\frac{\partial A_3}{\partial H} + 3H^3A_3 - \frac{H^3A_3 - 1}{H}\right) - \frac{\beta}{H^3}.$$

We denote $U \equiv (H^3A_3 - 1)^2$ and the system becomes

$$\frac{\partial I}{\partial \beta} = \frac{c}{2H^2}\frac{\partial U}{\partial \beta} + \frac{1}{2H^2},$$

$$\frac{\partial I}{\partial H} = \frac{c}{2H^2}\frac{\partial U}{\partial H} - \frac{cU}{H^3} - \frac{\beta}{H^3}. \quad (3.17)$$

From the first equation we get

$$I = \frac{cU}{2H^2} + \frac{\beta}{2H^2} + g(H),$$

where g is a smooth real function of H. The second equation of (3.17) gives then $g' = 0$. So, (3.17) has the solution

$$I = \frac{cU}{2H^2} + \frac{\beta}{2H^2} + ct.$$

If we use now (3.13) we get

$$\alpha^2 + ct\,H^2 = -c[(H^3A_3 - 1)^2 + H^4A_2^2],$$

so we must have $\alpha^2 + ct\,H^2 < 0$, which is absurd. So, in case (II), the system (3.7) has no solution and as a consequence is not localizable.

It remains for us to analyze case (III). We see that $\Omega_h^{\pm} \simeq \mathbb{R}^3 \times \mathbb{R}^3$ and we can take as independent coordinates P and K. The Poisson bracket is then

$$\{f_1,f_2\}=H\sum\frac{\partial f_1}{\partial K_i}\frac{\partial f_2}{\partial P_i}-(1\leftrightarrow 2)$$

$$-\frac{1}{H}\sum_{i,j,l=1}^{3}\epsilon_{ijl}(hP-P\Lambda K)_l\frac{\partial f_1}{\partial K_i}\frac{\partial f_2}{\partial K_j}\,. \quad (3.18)$$

We analyze again the system (3.7). From the second equation we get

$$X(P,K)=(K/H)+\Lambda(P)\,. \quad (3.19)$$

Then the third equation of (3.7) gives, after a short computation,

$$\sum_{l,m=1}^{3}\epsilon_{ilm}P_m\frac{\partial\Lambda_j}{\partial P_l}=\sum_{l=1}^{3}\epsilon_{ijl}\Lambda_l\,,\quad i,j=1,2,3\,.$$

From this system it follows easily that Λ is of the form

$$\Lambda(P)=PV(P^2)\,. \quad (3.20)$$

It remains to exploit the first relation of (3.7). With the help of (3.19) and (3.20) we get

$$\{X_i,X_j\}=-h/H^3\sum_{k=1}^{3}\epsilon_{ijk}P_k\,;$$

so for $h\neq 0$ the system is not localizable. For $h=0$, we can take $V=0$, so from (3.19) we get the solution of (3.7):

$$X=K/H\,. \quad (3.21)$$

In independent coordinates X, P we can see that in this case $M=T^*(\mathbb{R}^3)$, so the system is localizable.

Remark: This case is the limit $m\to 0$ of case I for $s=0$. The final result is the following.

Theorem: The elementary relativistic particles of nonzero mass $\mathcal{O}^{\pm}_{m,s}$ are localizable for any spin. The elementary relativistic particles of zero mass $\mathcal{O}^{\pm}_{c},\mathcal{O}_h$ are not localizable, except $\Omega_0^{\pm}$.

Let us note finally that recently relation (3.21) has been proposed as a solution for the localizability of all relativistic particles, including the photon[5,6] (see also Ref. 18).

IV. CONCLUSIONS

We have proposed definitions for covariance and localizability in analytical mechanics similar with the ones in quantum mechanics. The results obtained are in correspondence with those in quantum mechanics, which seems to indicate that the definitions proposed are realistic. We remark in connection with this parallelism another one. Because the Newton–Wigner concept of localizability excludes the photons, there is a definition of localizability which is more general,[19] and which essentially abandons the requirement that the quantum position operators commute. In our scheme it would mean abandoning the requirement (2.3), and consequently the relation (2.5). From our proof it follows easily that, in analytical mechanics, the situation is the same. If we abandon (2.5) then in all cases (I), (II), (III) from Sec. III we have localizability.

These are a number of problems which deserves further attention. First, we are reminded that in quantum mechanics there exists another case of a localizable system of zero mass besides those of zero helicity, which corresponds to a reducible representation of $\mathcal{P}^{\uparrow}_{+}$ (see Ref. 1, Corollary 12.17). It would be interesting to search for a classical analog of this system which would be associated with a nontransitive action of $\mathcal{P}^{\uparrow}_{+}$. Second, one knows a related problem with that of localizability, namely the construction of an evolution space.[12,13] All cases (I)–(III) investigated in Sec. III admit such an evolution space. It would be interesting to investigate if there is an exact relationship between the two concepts, namely if localizability implies always the existence of an evolution space. Finally, it would be desirable to prove the theorem in Sec. III without the aid of the infinitesimal relations (2.10), more in the spirit of the Mackey analysis of systems of imprimitivity.

ACKNOWLEDGMENT

The author wishes to thank the referee for pointing some interesting and relevant literature on the subject.

[1] V. S. Varadarajan, *Geometry of Quantum Theory* (von Nostrand Reinhold, New York, 1970), Vol. II.
[2] V. S. Varadarajan, *Geometry of Quantum Theory* (von Nostrand Reinhold, New York, 1968), Vol. I.
[3] T. D. Newton and E. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
[4] A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[5] H. Bacry, "Localizability and Space," *Lecture Notes in Physics*, Vol. 308 (Springer, Berlin, 1988).
[6] H. Bacry, Ann. Inst. H. Poincaré **49**, 245 (1988).
[7] J. M. Souriau, Compt. Rend. Acad. Sci. B **263**, 1191 (1966).
[8] H. Bacry, Commun. Math. Phys. **5**, 97 (1967).
[9] R. Arens, Commun. Math. Phys. **21**, 125 (1971).
[10] R. Arens, Commun. Math. Phys. **21**, 139 (1971).
[11] R. Arens, J. Math. Phys. **12**, 2415 (1971).
[12] J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).
[13] V. Guillemin and S. Stenberg, *Symplectic Techniques in Physics* (Cambridge U.P., Cambridge, 1984).
[14] H. P. Kuntzle, in *Geometrie Symplectique et physique mathematique*, edited by J. M. Souriau, CNRS publ. 237 (1974).
[15] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics. A Modern Perspective* (Wiley, New York, 1974).
[16] M. Pauri and G. M. Prosperi, J. Math. Phys. **9**, 1146 (1968).
[17] M. Pauri and G. M. Prosperi, J. Math. Phys. **16**, 1503 (1975).
[18] H. Bacry, J. Phys. A **14**, L73 (1981).
[19] W. Amrein, thesis, Geneva (1968); Helv. Phys. Acta **42**, 149 (1969).

Symmetries in the Lagrangian formalism

D. R. Grigore
Department of Theoretical Physics, Central Institute of Physics, P. O. Box MG-6, Bucharest, Romania

(Received 17 November 1987; accepted for publication 24 May 1989)

The analysis of Levy-Leblond [Commun. Math. Phys. **12**, 64 (1969)] is generalized for Lagrangian systems with symmetry. It is proved that this analysis goes through practically unchanged, and then some examples are analyzed in detail.

I. INTRODUCTION

The purpose of this paper is to generalize the study of the Lagrangian systems with a Lie group of symmetry presented in Ref. 1. To be more specific we must give some details. We study only systems with a finite number of degrees of freedom. Let $x_1,...,x_n$ be the real coordinates describing the configuration of the system. The system is called Lagrangian if one has a real smooth function L on time t, x's, and the velocities $v_1,...,v_n \in \mathbb{R}$, such that an evolution in time $\mathbb{R} \ni t \to q(t)$ is given by the usual Euler–Lagrange equations:

$$\frac{d}{dt}\frac{\partial L}{\partial v_i}(t,q,\dot{q}) = \frac{\partial L}{\partial x_i}(t,q,\dot{q}), \quad i = 1,...,n. \tag{1.1}$$

To describe the notion of symmetry it is convenient to treat t and x on the same footing with the following trick. Associate with L a new Lagrangian $\tilde{L}$ with $n+1$ degrees of freedom denoted $\xi_0,...,\xi_n$, and velocities $\rho_0,...,\rho_n$ by the formula

$$\tilde{L}(\xi_0,...,\xi_n,\rho_0,...,\rho_n) \equiv \rho_0 L\left(\xi_0, (\xi_1,...,\xi_n), \left(\frac{\rho_1}{\rho_0},...,\frac{\rho_n}{\rho_0}\right)\right). \tag{1.2}$$

Then, an evolution $\tau \mapsto q_0(\tau),...,q_n(\tau)$ given by the Euler–Lagrange equation of this system is nothing but an evolution for L given in parametrized form.

One can define now a symmetry to be a smooth map $\xi_0,...,\xi_n \overset{\varphi}{\to} \xi'_0,...,\xi'_n$ that maps solutions of the Euler–Lagrange equations for (1.2) into solutions of the same equations.

It is easy to prove now that φ is a symmetry if the following identity holds:

$$\tilde{L}(\varphi(\xi),\hat{\varphi}(\rho)) = a\tilde{L}(\xi,\rho) + \sum_{i=0}^{n} \frac{\partial \Lambda_\varphi}{\partial \xi_i}(\xi)\rho_i. \tag{1.3}$$

Here

$$\hat{\varphi}(\rho)_i \equiv \sum_{j=0}^{n} \frac{\partial \varphi_i}{\partial \xi_j}(\xi)\rho_j,$$

Λ_φ is a smooth real function depending on φ, and $a \in \mathbb{R}$ (remark that we must have $a \neq 0$, because in the opposite case L would give trivial equations of motion).

In Ref. 1 the concept of symmetry is based on relation (1.3) with $a = 1$. In this paper we investigate the more general case $a \neq 1$.

In Sec. II we establish that the analysis in Ref. 1 goes through practically unchanged in our more general case.

In Sec. III we give some details on group cohomology needed for practical applications.

In Sec. IV we treat some examples. The last section is devoted to conclusions and comments. Some computations are made in the Appendix.

II. LAGRANGIAN SYSTEMS WITH LIE GROUPS OF SYMMETRY

Suppose we have a Lie group G acting smoothly on the variables ξ, i.e., for any $g \in G$, one has a map $\xi \overset{\varphi_g}{\mapsto} \xi'$ such that $g,\xi \to \varphi_g(\xi)$ is smooth and we also have

(i) $\varphi_e = \mathrm{Id}$,

(ii) $\varphi_{g_1} \circ \varphi_{g_2} = \varphi_{g_1 g_2}, \quad \forall g_1,g_2 \in G.$

By virtue of the discussion in the Introduction one can call G a group of symmetry for the Lagrangian L if for any $g \in G$, there exists $a_g \in \mathbb{R} \setminus \{0\}$ depending smoothly on G and one also has a smooth real valued function Λ depending on g and ξ such that

$$\tilde{L}(\varphi_g(\xi),\hat{\varphi}_g(\rho)) = a_g \tilde{L}(\xi,\rho) + \sum_{i=0}^{n} \frac{\partial \Lambda}{\partial \xi_i}(g,\xi)\rho_i. \tag{2.1}$$

Remark: From (1.2) it follows that we have

$$\tilde{L}(\xi,\alpha\rho) = \alpha \tilde{L}(\xi,\rho), \quad \forall \alpha \in \mathbb{R}. \tag{2.2}$$

[In fact, $\tilde{L}$ is of the form (1.2) *iff* this condition is true.]

We denote by $\hat{\rho}$ the direction of ρ. Then, the action of G on (ξ,ρ) variables,

$$(\xi,\rho) \overset{\Phi_g}{\mapsto} (\varphi_g(\xi),\hat{\varphi}_g(\rho)),$$

factorizes to an action $g \to \hat{\Phi}_g$ on $(\xi,\hat{\rho})$ variables. As remarked in Ref. 2, from (2.1), L is determined on every orbit of G in $(\xi,\hat{\rho})$ variables, if one knows it in one point of the orbit. In particular, if G acts transitively in $(\xi,\hat{\rho})$ variables then L is determined by its value at a point ξ_0.

In this case we can give an explicit formula useful in applications. From the hypothesis of transitivity of the G action $\hat{\Phi}$, we infer that for any ξ,v there exists $g \in G$ and $\rho_0 \in \mathbb{R} \setminus \{0\}$ (depending on ξ and v) such that

$$\Phi_g(\xi_0,\rho_0,0) = (\xi,1,v).$$

Then from (2.1) we have

$$\tilde{L}(\xi,1,v) = a_g \tilde{L}(\xi_0,\rho_0,0) + \frac{\partial \Lambda}{\partial \xi_0}(g,\xi_0)\rho_0,$$

or, in t,x variables,

$$L(t,x,v) = \rho_0\left[a_g c + \frac{\partial \Lambda}{\partial t}(g,t_0,x_0)\right], \tag{2.3}$$

where we have denoted $L(\xi_0,0) \equiv c$ and $\xi_0 = (t_0,x_0)$. We

 0022-2488/89/112653-11$02.50

return now to the study of (2.1). By some computations one can show that, for any $g_1,g_2\in G$, one has

$$a_{g_1}a_{g_2}=a_{g_1g_2} \tag{2.4}$$

and the functions $i\Lambda(g_1,g_2)$, given by

$$i\Lambda(g_1,g_2)(\xi)$$
$$\equiv\Lambda(g_1,g_2\cdot\xi)+a_{g_1}\Lambda(g_2,\xi)-\Lambda(g_1g_2,\xi), \tag{2.5}$$

do not depend on ξ. [From now on, we use the simpler notation $g\cdot\xi\equiv\varphi_g(\xi)$.] One must suppose the domain of values of ξ to be connected.

Because $a_g\neq 0$, $\forall g\in G$, from (2.4) it follows that we have $a_e=1$, so $g\mapsto a_g$ is a smooth real one-dimensional representation of the Lie group G. Thus, at least at the formal level, we have a problem related to the one studied in a preceding paper.[3] We will return to this point later.

Also from (2.1) one has $\Lambda(e,\xi)=\text{const}$. We remark now that the function Λ is determined by (2.1) up to a function of g. It is convenient to eliminate this arbitrariness as in Ref. 1: choose a reference point ξ_0 and demand

$$\Lambda(g,\xi_0)=0,\quad \forall g\in G. \tag{2.6}$$

We denote by Z_a the vector space of smooth real functions depending on g and ξ, such that $i\Lambda(g_1,g_2)$ does not depend on ξ for any $g_1,g_2\in G$, and one has also (2.6). The elements of Z_a are called gauge functions in Ref. 1. It is easy to verify that if λ is a real smooth function depending on ξ, then Λ, given by

$$\Lambda(g,\xi)=a_g[\lambda(\xi)-\lambda(\xi_0)]-[\lambda(g\cdot\xi)-\lambda(g\cdot\xi_0)], \tag{2.7}$$

is an element of Z_a. We denote by B_a the subset of Z_a formed by elements of this form, and remark that B_a is a linear subspace of Z_a. Two gauge functions differing by an element in B_a are called equivalent. We denote $H_a=Z_a/B_a$; H_a is also a vector space.

We now denote the vector space of Lagrangians $\tilde{L}$ admitting G as a group of symmetry in the sense (2.1) by $\mathscr{L}$. We recall that for Lagrangians of the form

$$\sum_{i=0}^{n}\frac{\partial\lambda}{\partial\xi_i}(\xi)\rho_i$$

(λ is a smooth real function on ξ), the Euler–Lagrange equations are identities. We denote by $\mathscr{L}_0$ the subset of $\mathscr{L}$ formed by Lagrangians of this form. Here $\mathscr{L}_0$ is a linear subspace of $\mathscr{L}$, so we can form the vector space $\mathscr{H}=\mathscr{L}/\mathscr{L}_0$. We now remark that we have a natural linear operator from $\mathscr{H}$ into H_a. Indeed, we have a linear map from $\mathscr{L}$ into Z_a, which associates with $\tilde{L}\in\mathscr{L}$ the function Λ in (2.1) with the appropriate normalization condition (2.6). It is easy to see that this linear operator maps $\mathscr{L}_0$ into B_a, so it factorizes to a linear operator $S:\mathscr{H}\to H_a$. (Because Lagrangians differing by a multiplicative constant give the same Euler–Lagrange equations, they must not be considered distinct: thus from the physical point of view we need the projective space $\hat{\mathscr{H}}$.) Thus the study of the Lagrangian systems with symmetry reduces to the study of the map S. In the case where G acts transitively in variables $(\xi,\hat{\rho})$, the remark following (2.1) shows that S is surjective.

First we must study H_a. We follow the spirit of Ref. 1 with some important clarifications from Ref. 2.

First we remind the reader of some standard cohomology notions.[4] Let G and K be Lie groups, K Abelian, and a: $G\to$ Aut K a smooth homomorphism of G into the group of the automorphisms of K. Let, for any $n\in\mathbb{N}$, $C^n(G,K)$ be the set of smooth maps

$$c:\underbrace{G\times\cdots\times G}_{n\text{ times}}\to K.$$

We also define $C^0(G,K)\equiv K$ and $C^{-1}(G,k)\equiv\{0\}$ (0 is the neutral element in K; we use additive notation for the composition law in K). The elements of $C^n(G,K)$ are called K-valued cochains of dimension n. In what follows we will always suppose the cochains to be normed in the sense

$$c(g_1,\ldots,g_{i-1},e,g_{i+1},\ldots,g_n)=0, \tag{2.8}$$

for any $i=1,\ldots,n$. One defines next the cobord operator δ: $C^n(G,K)\mapsto C^{n+1}(G,K)$ by the formula

$$\delta c(g_1,\ldots,g_{n+1})$$
$$=a_{g_1}c(g_2,\ldots,g_{n+1})$$
$$+\sum_{1\le i\le n}(-1)^i c(g_1,\ldots,g_{i-1},g_ig_{i+1},g_{i+2},\ldots,g_{n+1})$$
$$+(-1)^{n+1}c(g_1,\ldots,g_n). \tag{2.9}$$

We will need some particular cases: for $c\in C^0(G,K)$,

$$\delta c(g)=a_gc-c; \tag{2.9'}$$

for $c\in C^1(G,K)$,

$$\delta c(g_1,g_2)=a_{g_1}c(g_2)-c(g_1,g_2)+c(g_1); \tag{2.9''}$$

for $c\in C^2(G,K)$,

$$\delta c(g_1,g_2,g_3)=a_{g_1}c(g_2,g_3)-c(g_1g_2,g_3)$$
$$+c(g_1,g_2g_3)-c(g_1,g_2). \tag{2.9'''}$$

Now denote

$$Z_a^n(G,K)\equiv\{c\in C^n(G,K)\,|\,\delta c=0\}$$

[the elements of $Z_a^n(G,K)$ are called K-valued cocycles of dimension n, relative to a], and

$$B_a^n(G,K)\equiv\{c\in C^n(G,K)\,|\,\exists b\in C^{n-1}(G,K)$$
$$\text{such that } c=\delta b\}$$

[the elements of $B_a^n(G,K)$ are called K-valued cobords of dimension n, relative to a].

Because $\delta^2=0$, one has $B_a^n(G,K)\subset Z_a^n(G,K)$ and we can define the nth cohomology group of G with coefficients in K relative to a, by $H_a^n(G,K)\equiv Z_a^n(G,K)/B_a^n(G,K)$. We remark that if K is a vector space, then $Z_a^n(G,K)$, $B_a^n(G,K)$, and $H_a^n(G,K)$ are also vector spaces. If a is the trivial homomorphism, the index a is omitted.

In what follows we need an auxiliary cohomology group. Let $H\subset G$ be a subgroup of G. We define, for $n\in\mathbb{N}$,

$$C^n(G\,|H,K)$$
$$\equiv\{c\in C^n(G,K)\,|\,c(g_1,\ldots,g_{n-1},g_nh)$$
$$=c(g_1,\ldots,g_n),\ \forall h\in H,\forall g_1,\ldots,g_n\in G\},$$
$$C^0(G\,|H,K)\equiv\{0\},\quad C^{-1}(G\,|H,K)=\{0\}.$$

Because δ maps $C^n(G|H,K)$ into $C^{n+1}(G|H,K)$, one can define $Z_a^n(G|H,K)$, $B_a^n(G|H,K)$, and $H_a^n(G|H,K)$ by analogy with the similar definitions above.

We can proceed now with the analysis as in Ref. 1. We denote by H the subgroup of stability of the reference point ξ_0, and suppose that the manifold on which G acts is a connected homogeneous G-space, i.e., is of the form G/H. Some of the results hold for any K and some only for $K=\mathbb{R}^n$ ($n\in\mathbb{N}$).

(1) Suppose the bundle $(G/H,G,\pi)$ is trivial. Then H_a is isomorphic with $H_a^2(G|H,\mathbb{R})$.

Proof: The proof is similar to the one in Ref. 1 and is based on simple computations, which are omitted.

(i) Let $A\in H_a$, and $\Lambda\in A$. Then the map

$$G\times G\ni g_1,g_2\mapsto i\Lambda(g_1,g_2)\in\mathbb{R}$$

is an element of $Z_a^2(G|H;\mathbb{R})$. Moreover, if $\Lambda\in B_a$, this map is an element of $B_a^2(G|H;\mathbb{R})$. So we have by factorization a map

$$\hat{i}: H_a\mapsto H_a^2(G|H,\mathbb{R}).$$

(ii) Conversely, let $C\in H_a^2(G|H,\mathbb{R})$, and $c\in C$. From the triviality of the bundle $(G/H,G,\pi)$ we infer the existence of a smooth cross section $\sigma: G/H\to G$ such that

$$\sigma(\xi)\xi_0=\xi,\quad \sigma(\xi_0)=e. \tag{2.10}$$

We now define $jc: G\times G/H\to\mathbb{R}$ by the formula

$$jc(g,\xi)\equiv c(g,\sigma(\xi)). \tag{2.11}$$

Then one can show, using (2.9‴) and (2.10), that $jc\in Z_a$. Moreover, if $c\in B_a^2(G/H,\mathbb{R})$ the $jc\in B_a$. So we have a natural map $\hat{j}: H_a^2(G|H,\mathbb{R})\to H_a$.

(iii) By direct computation one shows that $i\circ j=\mathrm{Id}$ and $j\circ i=\mathrm{Id}$. It follows that $\hat{i}\circ\hat{j}=\mathrm{Id}$, $\hat{j}\circ\hat{i}=\mathrm{Id}$. So $\hat{i}$ is the desired isomorphism. ■

(2) Let $c\in Z_a^2(G,K)$. Then $c\in Z_a(G|H,K)$ *iff* $c|G\times H=0$.

Proof: $\Rightarrow$: By definition

$$c(g_1,g_2h)=c(g_1,g_2),\quad \forall g_1,g_2\in G,\quad \forall h\in H.$$

Take $g_1=g$, $g_2=e$; we get $c(g,h)=c(g,e)=0$.

$\Leftarrow$: Take in the cocycle identity $\delta c=0$, $g_3=h\in H$. If we use $c|G\times H=0$, we get $c(g_1,g_2h)=c(g_1,g_2)$. ■

(3) Let $c\in Z_a^2(G|H,K)$. Then $c\in B_a^2(G|H,K)$ *iff* $c=\delta b$, where $b: G\to K$ verifies $b|H=0$.

Proof: $\Rightarrow$: By definition $c\in B_a^2(G|H,K)$ means that $c=\delta b$, where $b: G\to K$ verifies $b(gh)=b(g)$, $\forall g\in G$, $\forall h\in H$. Take $g=e$. Then $b(h)=b(e)=0$, because of (2.8).

$\Leftarrow$: By hypothesis $c=\delta b$, i.e., according to (2.9″),

$$c(g_1,g_2)=a_{g_1}b(g_2)-b(g_1,g_2)+b(g_1),$$

and $b|H=0$. We take $g_1=g\in G$, $g_2=h\in H$. But $c(g,h)=0$, according to (2). So we get

$$a_g b(h)-b(gh)+b(g)=0.$$

Because $b|H=0$, we get $b(gh)=b(g)$, i.e., $b\in C^1(G|H,K)$. It follows that $c=\delta b\in B_a^2(G|H;K)$. ■

We need now another cohomology structure based on cochains defined for $n\in\mathbb{N}$:

$$C^n(G,H,K)\equiv\{c\in C^n(G,K)\,|\,c|\underbrace{H\times\cdots\times H}_{n\text{ times}}=0\},$$

$$C^0(G,H,K)\equiv\{0\},\quad C^{-1}(G,H,K)\equiv\{0\}.$$

These groups of cohomology were introduced in Ref. 2 and correspond to the notion of H superequivalence from Ref. 1.

Because δ maps $C^n(G,H,K)$ into $C^{n+1}(G,H,K)$, then $Z_a^n(G,H,K)$, $B_a^n(G,H,K)$, and $H_a^n(G,H,K)$ can be defined similarly to the preceding definitions.

We now have a central lemma, suggested by Proposition 2 of Ref. 1, and found explicitly in Ref. 2.

(4) Let the bundle $(G/H,G,\pi)$ be trivial, and let $C\in H_a^2(G,H,K)$. Then there exists $c\in C$ such that $c\in Z_a^2(G|H,K)$.

Proof: (i) If $\sigma: G/H\to G$ is the cross section defined in (1), define now $s: G\to G$ by $s\equiv c\circ\pi$, and $\chi: G\to H$ by $\chi(g)\equiv s(g)^{-1}g$. One then proves[1]

$$s(gh)=s(g),\quad \forall g\in G,\quad \forall h\in H,$$

$$s|H=e,$$

$$\chi(gh)=\chi(g)h,\quad \forall g\in G,\quad \forall h\in H,$$

$$\chi(h)=h.$$

(ii) Let $\bar{c}\in C$. Define $b: G\to K$ by

$$b(g)\equiv\bar{c}(s(g),\chi(g)).$$

We have for any $h\in H$,

$$b(h)=\bar{c}(s(h),\chi(h))=\bar{c}(e,h)=0,$$

so $b\in C^1(G,H,K)$. Then define $c\in Z_a^2(G,H,R)$ by

$$c\equiv\bar{c}+\delta b.$$

One can easily verify that $\forall g\in G$, $\forall h\in H$,

$$c(g,h)=\bar{c}(s(g)\chi(g),h)-\bar{c}(s(g),\chi(g)h)+\bar{c}(s(g),\chi(g)).$$

Using the cocycle identity (2.9‴) we get

$$c(g,h)=a_{s(g)}\bar{c}(\chi(g),h)=0,$$

because by definition $\bar{c}|H\times H=0$. It follows that $c|G\times H=0$. Now applying (2), we get that $c\in Z_a^2(G|H,K)$. ■

(5) In the hypothesis of (4), in every cohomology class from $H_a^2(G,H,K)$ there exists one and only one cohomology class from $H_a^2(G|H,K)$.

Proof: In every cohomology class from $H_a^2(G,H,K)$ there exists at least one cohomology class from $H_a^2(G|H,K)$ by virtue of (4). We now prove the unicity. Let $C\in H_a^2(G,H,K)$ and $c,c'\in C\cap Z_a^2(G|H,K)$. Because $c,c'\in C$, it follows that $c'-c\in B_a^2(G,H,K)$, i.e., then $b\in C^1(G,H,K)$ exists such that $c'-c=\delta b$. By definition $b|H=0$. Applying (3) we get that $c'-c\in B_a^2(G|H,K)$, i.e., c and c' belong to the same cohomology class in $H_a^2(G|H,K)$. ■

As immediate corollaries we have the following.

(6) If the bundle $(G/H,G,\pi)$ is trivial, then

$$H_a^2(G|H,K)\simeq H_a^2(G,H,K).$$

(7) For the same conditions as above, we have

$$H_a\simeq H_a^2(G,H,\mathbb{R}).$$

Proof: From (1) and (6) for $K=\mathbb{R}$. ∎

So, we are left with the problem of studying the second cohomology group $H_a^2(G,H,\mathbb{R})$. This can be done by using the long cohomology sequence as in Ref. 2 but we prefer a simpler approach in the spirit of Ref. 1.

We remark first that if $H\subset G$ is a subgroup and r: $C^n(G,K)\to C^n(H,K)$ is the restriction application, then r commutes with δ. As a consequence we have $r(B_a^n(G,K))\subset B_{a|H}^n(H,K)$ and $r(Z_a^n(G,K))\subset Z_{a|H}^n(H,K)$. We call a cohomology class $C\in H_a^2(G,K)$ admissible if it contains at least one cohomology class from $H_a^2(G,H,K)$. We take $K=\mathbb{R}^n$. Then we have the following proposition.

(8) Let G be a paracompact Lie group, and $C\in H_a^2(G,\mathbb{R}^n)$ be an admissible cohomology class. Then the cohomology classes from $H_a^2(G,H,\mathbb{R}^n)$ included in C are in one to one correspondence with $Z_{a|H}^2(H,\mathbb{R}^n)/r(Z_a^1(G,\mathbb{R}^n))$.

Proof: Let $\mathscr{C}_0\in H_a^2(G,H,\mathbb{R}^n)$, $\mathscr{C}_0\subset C$ be fixed in the following.

(i) Let $\mathscr{C}$ be any cohomology class from $H_a^2(G,H,\mathbb{R}^n)$ included in C. We take $c\in\mathscr{C}$ and $c_0\in\mathscr{C}_0$. Because $c,c_0\in C$, $c-c_0\in B_a^2(G,\mathbb{R}^n)$, i.e., there exists $b\in C^1(G,\mathbb{R}^n)$, such that $c-c_0=\delta b$. But, on the other hand, $c-c_0\in Z_a^2(G,H,\mathbb{R}^n)$, so by (2), $(c-c_0)|H=0$. It follows that $b|H\in Z_{a|H}^1(H,\mathbb{R}^n)$. We will show now that if we take another choice for c and c_0, we modify $b|H$ by an element in $r(Z_a^1(G,\mathbb{R}^n))$. Indeed, let $c'\in\mathscr{C}$ and $c_0'\in\mathscr{C}_0$. We have as before $c'-c_0'=\delta b'$, where $b'|H\in Z_{a|H}^1(H,\mathbb{R}^n)$. On the other hand, $c,c'=\in\mathscr{C}$ implies the existence of $\bar{b}\in C^1(G,H,\mathbb{R}^n)$, such that $c-c'=\delta\bar{b}$. Similarly, there exists $\bar{b}_0\in C^1(G,H,\mathbb{R})$ such that $c_0-c_0'=\delta\bar{b}_0$. It follows that $\delta(\bar{b}-\bar{b}_0-b+b')=0$, i.e., $\bar{b}-\bar{b}_0-b+b'\in Z_a^1(G,\mathbb{R}^n)$. Because $(\bar{b}-\bar{b}_0-b+b')|H=b'|H-b|H$ it follows that $b'|H-b|H\in r(Z_a^1(G,\mathbb{R}^n))$.

So we can define a map that associates with every cohomology class from $H_a^2(G,H,\mathbb{R}^n)$ an element from

$$Z_{a|H}^1(H,\mathbb{R}^n)/r(Z_a^1(G,\mathbb{R}^n)).$$

(ii) Conversely, let $B\in Z_{a|H}^1(H,\mathbb{R}^n)/r(Z_a^1(G,\mathbb{R}^n))$ and $b\in B$. Take $\bar{b}$: $G\to\mathbb{R}^n$ to be any smooth extension of b ($\bar{b}$ exists according to Th. 5.7, Chap. I in Ref. 5). Let $c_0\in\mathscr{C}_0$ and define $c\equiv c_0+\delta\bar{b}$. Because $\bar{b}|H=b\in Z_{a|H}^1(H,\mathbb{R}^n)$, it follows that $c\in C^2(G,H,\mathbb{R}^n)$. Because $\delta c=\delta c_0=0$ we have, moreover, that $c\in Z_a^2(G,H,\mathbb{R}^n)$. We will show that if we take another choice of $b\in B$ and $c_0\in\mathscr{C}_0$, then we modify c by an element in $B_a^2(G,H,\mathbb{R}^n)$. Indeed, let $b'\in B$ and $c_0'\in\mathscr{C}_0$. Let $\bar{b}'$: $G\to\mathbb{R}^n$ be a smooth extension of b' and $c'\equiv c_0'+\delta\bar{b}'$. Again $c'\in Z_a^2(G,H,\mathbb{R}^n)$. On the other hand, we have $c-c'=c_0-c_0'+\delta(\bar{b}-\bar{b}')$. But $c_0',c_0\in\mathscr{C}_0$ implies the existence of $b_1\in C^1(G,H,\mathbb{R}^n)$ such that $c_0'-c_0=\delta b_1$. So we have $c-c'=\delta(\bar{b}-\bar{b}'-b_1)$. But

$$(\bar{b}-\bar{b}'-b_1)|H=b-b-b_1|H=0,$$

so $\bar{b}-\bar{b}-b_1\in C^1(G,H,\mathbb{R}^n)$. It follows that $c-c'\in B_a^2(G,H,\mathbb{R}^n)$.

We have defined in this way a map that associates with every element in $Z_{a|H}^1(H,\mathbb{R}^n)/r(Z_a^1(G,\mathbb{R}^n))$ a cohomology class from $H_a^2(G,H,\mathbb{R}^n)$ contained in C.

(iii) It is easy to see that the maps defined in (i) and (ii) are inverse to each other. This finishes the proof. ∎

We can formulate the following theorem.

Main theorem: Let G be a paracompact Lie group, $H\subset G$ a closed subgroup such that the bundle $(G/H,G,\pi)$ is trivial, and σ: $G/H\to G$ a smooth cross section. In every admissible cohomology class $C\in H_a^2(G,\mathbb{R})$, we fix $\mathscr{C}_0\in H_a^2(G,H,\mathbb{R})$ and $c_0\in\mathscr{C}_0$. In every class $B\in Z_{a|H}^1(H,\mathbb{R})/r(Z_a^1(G,\mathbb{R}))$ we fix $b\in B$. Then every gauge function for the homogeneous G space G/H is equivalent to one of the form

$$\Lambda(g,\xi)=c_0(g,\sigma(\xi))+a_{\sigma(g\cdot\xi)}b(\gamma(g,\xi))-a_{\sigma(g\cdot\xi_0)}b(\gamma(g,\xi_0)),\tag{2.12}$$

where γ: $G\times G/H\to H$ is given by

$$\gamma(g,\xi)\equiv c(g\cdot\xi)^{-1}gc(\xi).\tag{2.13}$$

Moreover, in this way we obtain one and only one representative from every equivalence class in H_a.

Proof: Let $A\in H_a$. Then, according to (7), we can find a unique $\mathscr{C}\in H_a^2(G,H,\mathbb{R})$. Let $C\in H_a^2(G,\mathbb{R})$, the admissible class that contains $\mathscr{C}$. Then, to the couple $\mathscr{C}_0$ and $\mathscr{C}$ ($\mathscr{C}_0$ is fixed by hypothesis) we can attach, according to (8), a unique element $B\in Z_{a|H}^1(H,\mathbb{R})/r(Z_a^1(G,\mathbb{R}))$. Let $b\in B$ be the element fixed by the hypothesis and $\bar{b}$: $G\to\mathbb{R}$ be any smooth extension of b. Define $\bar{c}=c_0-\delta\bar{b}$. Then, according to (8), $[\bar{c}]=\mathscr{C}$. With the help of (4), we now find $c\in\mathscr{C}$ such that $c\in Z_a^2(G|H,\mathbb{R})$. Then the preceding argument shows that $\Lambda'\equiv jc\in A$.

It remains for us to put this program into action. By some computations one can find that

$$\begin{aligned}\Lambda'(g,\xi)=&c_0(g,\sigma(\xi))+a_{gs(\sigma(\xi))}\bar{b}(\chi(\sigma(\xi)))\\&-a_{s(g\sigma(\xi))}\bar{b}(\chi(g\sigma(\xi)))+a_{s(g)}\bar{b}(\chi(g))\\&+a_g b(s\sigma(\xi))-\bar{b}(s(g\sigma(\xi)))+\bar{b}(s(g)).\end{aligned}$$

The sum of the last three terms is in B_a (take $\lambda=\bar{b}\circ s\circ\sigma$), so we can take $\Lambda''\in A$ as

$$\begin{aligned}\Lambda''(g,\xi)=&c_0(g,\sigma(\xi))+a_{gs(\sigma(\xi))}\bar{b}(\chi(\sigma(\xi)))\\&-a_{s(g\sigma(\xi)))}\bar{b}(\chi(g\sigma(\xi)))+a_{s(g)}\bar{b}(\chi(g))\end{aligned}$$

But we have

$$\begin{aligned}\bar{b}(\chi(g\sigma(\xi)))&=\bar{b}(\chi(\sigma(g\cdot\xi))\gamma(g,\xi))\\&=a_{\chi(\sigma(g\cdot\xi))}\bar{b}(\gamma(g,\xi))+\bar{b}(\chi(\sigma(g\cdot\xi))).\end{aligned}$$

(The last equality follows from the cocycle identity for $\bar{b}$.) So, Λ'' can be rewritten as

$$\begin{aligned}\Lambda''(g,\xi)=&c_0(g,\sigma(\xi))-a_{\sigma(g\cdot\xi)}b(\gamma(g,\xi))\\&+a_{s(g)}b(\chi(g))+a_{gs(\sigma(\xi))}b(\chi(\sigma(\xi)))\\&-a_{s(\sigma(g\cdot\xi))}b(\chi(\sigma(g\cdot\xi))).\end{aligned}$$

The last two terms can be written as $a_g[\lambda(\xi)-\lambda(\xi_0)]-\lambda(g\cdot\xi)$, where $\lambda(\xi)\equiv a_{s(\sigma(\xi))}b(\chi(\sigma(\xi)))$, so we can take $\Lambda\in A$ as

$$\begin{aligned}\Lambda(g,\xi)=&c_0(g,\sigma(\xi))-a_{\sigma(g\cdot\xi)}b(\gamma(g,\xi))\\&+a_{s(g)}b(\chi(g))-a_{s(\sigma(g,\xi_0))}b(\chi(\sigma(g\cdot\xi_0))).\end{aligned}$$

Again applying the cocycle identity for b, this expression can be proved to coincide with (2.12). ■

Remark: The last term in (2.13) is constant so it can be left out in (2.1).

III. THE COHOMOLOGY OF LIE GROUPS

Section II shows that if we want to find the Lagrangians admitting a Lie group of symmetry, we must be able to describe first $Z^1_a(G,\mathbb{R})$ and $H^2_a(G,\mathbb{R})$. Fortunately for Lie groups, one can relate these objects to corresponding ones for the Lie algebra associated to the Lie group, which reduces the problems to purely algebraic ones. So we now remind the reader of the basic notions of Lie algebra cohomology.[6] Let L be a Lie algebra and V a vector space. Let, for any $n\in\mathbb{N}$, $C^n(L,V)$ be the set of n - linear completely antisymmetric maps c: $\underbrace{L\times\cdots\times L}_{n \text{ times}}\to V$. We define also $C^0(L,V)\equiv V$, $C^{-1}(L,V)=\{0\}$. The elements of $C^n(L,V)$ are called V cochains of dimension n. Let f: $L\to \mathrm{End}(V)$ be a linear representation of L in V. Then one defines the cobord operator δ: $C^n(L,V)\to C^{n+1}(L,V)$ by

$$\delta c(X_0,\ldots,X_n)$$
$$=\sum_{0\le i\le n}(-1)^i f(X_i)c(X_0,..,\hat{X}_i,\ldots,X_n)$$
$$+\sum_{0<i<j<n}(-1)^{i+j}c([X_i,X_j],X_0,..,\hat{X}_i,..,\hat{X}_j,\ldots,X_n). \tag{3.1}$$

Now denote

$$Z^n_f(L,V)\equiv\{c\in C^n(L,V)\,|\,\delta c=0\}$$

and

$$B^n_f(L,V)\equiv\{c\in C^n(L,V)\,|\,\exists b\in C^{n-1}(L,V) \text{ such that } c=\delta b\}.$$

The elements of $Z^n_f(L,V)$ are called cocycles and those of $B^n_f(L,V)$ cobords. Because $\delta^2=0$, $B^n_f(L,V)$ is a linear subspace of $Z^n_f(L,V)$ and one can define the nth space of cohomology of L relative to f to be $H^n_f(L,V)=Z^n_f(L,V)/B^n_f(L,V)$. If f is the trivial representation, the index f is omitted. We need some particular examples of (3.1).

For $c\in C^0(L,V)\equiv V$,

$$\delta c(X)=f(X)c.$$

For $c\in C^1(L,V)$,

$$\delta c(X,Y)=f(X)c(X)-f(Y)c(X)-c([X,Y]).$$

For $c\in C^2(L,V)$,

$$\delta c(X,Y,Z)=f(X)c(Y,Z)+\text{cycl.}-(c[[X,Y],Z]+\text{cycl.}).$$

We suppose now that G is connected and simply connected. If a: $G\to \mathrm{Aut}\,V$ is a representation of G in V, denote by $\dot{a}$: Lie $G\to\mathrm{End}(V)$ the corresponding representation of the Lie algebra Lie G of the group G.

(A) The connection between $Z^1_a(G,V)$ and $Z^1_f(\text{Lie } G,V)$ can be described as follows.[4] If $c\in C^1(G,V)$ then we define, for any g, an operator T_g acting in $V+\mathbb{R}$ and given in obvious matrix notation by

$$T_g=\begin{pmatrix}a_g & c(g)\\ 0 & 1\end{pmatrix}. \tag{3.2}$$

Then $c\in Z^1_a(G,V)$ *iff* $g\mapsto T_g$ is a representation of G in $V+\mathbb{R}$. Analogously if $c\in C^1(L,V)$, where L is a Lie algebra, we define for any $X\in L$ an operator $t(X)$ acting in $V+\mathbb{R}$ and given by

$$t(X)=\begin{pmatrix}f(X)c(X)\\ 0 \quad 0\end{pmatrix}. \tag{3.3}$$

Again $c\in Z^1_f(L,V)$ *iff* $X\mapsto t(X)$ is a representation of L in $V+\mathbb{R}$.

Since for connected and simple connected Lie groups there is a one to one correspondence between representations of the group and representations of its Lie algebra, we have, by virtue of the preceding remarks, a one to one correspondence between $Z^1_a(G,V)$ and $Z^1_{\dot{a}}(\text{Lie } G,V)$.

We remark (although we will not need this in the following) that it is easy to prove that this correspondence factorizes to a one to one correspondence between $H^1_a(G,V)$ and $H^1_{\dot{a}}(\text{Lie } G,V)$.

(B) According to a result of Hochschild,[7] a similar result also holds for the second cohomology groups; namely, if G is a connected and simple connected Lie group and K is a conected Lie group, with a and $\dot{a}$ as above, then there is a one to one correspondence between $H^2_a(G,K)$ and $H^2_{\dot{a}}(\text{Lie } G,\text{Lie } K)$. In our case, $K=V$, and we can also take Lie $K=V$.

We give only the idea of the proof because it is needed in the practical computations. This idea is to use the so-called extensions of Lie groups and of Lie algebras, respectively.

(1) Let G and K be Lie groups. Then, an extension of K by G is a triplet (H,i,j), where H is a Lie group, and the short sequence

$$0\to K\xrightarrow{i} H\xrightarrow{j} G\to 0$$

is exact. If K is Abelian, then one has a natural homomorphism a: $G\to\mathrm{Aut}\,K$ defined by

$$i(a(g)k)=hi(k)h^{-1},$$

where $h\in H$ is any element verifying $j(h)=g$. The extension is called "associated" with a. Two extensions of K by G, (H,I,j) and (H',i',j'), are called equivalent if there exists a group isomorphism q: $H\to H'$ such that $q\circ i=i'$ and $j'\circ q=j$. If K is Abelian, then two equivalent extensions are associated with the same homomorphism a: $G\to\mathrm{Aut}\,K$. We denote by $\mathrm{Ext}(G,K,a)$ the set of equivalence classes of extensions of the Abelian group K by G associated with a. The first step in proving the needed result is establishing a one to one correspondence between $\mathrm{Ext}(G,K,a)$ and $H^2_a(G,K)$. We will need explicitly the map from $H^2_a(G,K)$ in $\mathrm{Ext}(G,K,a)$. Let $c\in Z^2_a(G,K)$. We define on the set $H=K\times G$ the following composition law:

$$(k_1,g_1)(k_2,g_2)=(k_1+a(g_1)k_2+c(g_1,g_2),g_1g_2). \tag{3.4}$$

Then one can prove that H is a Lie group. We denote by H_c this Lie group. If i: $K\to H_c$ and j: $H_c\to G$ are given by

$i(k) = (k,e)$ and $j(k,g) = g$, then (H_c,i,j) is an extension of K by G associated with a. We have a map from $Z_a^2(G,K)$ into the set of extensions of K by G associated with a. One proves easily that this map factorizes to a map from $H_a^2(G,K)$ into $\mathrm{Ext}(G,K,a)$. The construction of the inverse of this map is much more difficult, and can be found in Ref. 7.

(2) Now we give the corresponding construction for Lie algebras.[6] Let L and V be two Lie algebras over the same field P. An extension of V by L is a triplet (E,α,β), where E is a Lie algebra over P, and the short sequence

$$0 \to V \xrightarrow{\alpha} E \xrightarrow{\beta} L \to 0$$

is exact. If V is Abelian, then one has a natural representation $f: L \to \mathrm{End}(V)$ defined by

$$\alpha(f(X)v) = [e,\alpha(v)],$$

where $e \in E$ is any element such that $\beta(e) = X$. The extension is called "associated" with f. Two extensions of V by G, (E,α,B) and (E',α',β'), are called equivalent if there exists a Lie algebra isomorphism $\theta: E \to E'$ such that $\theta \circ \alpha = \alpha'$ and $\beta' \circ \theta = \beta$. If V is Abelian, then equivalent extensions are associated with the same representation $f: L \to \mathrm{End}(V)$. We denote by $\mathrm{Ext}(L,V,f)$ the set of equivalence classes of extensions of the Abelian Lie algebra V by L, associated with f. The second step in proving the main result is to establish a one to one correspondence between $\mathrm{Ext}(L,V,f)$ and $H_f^2(L,V)$. As in (1), we will need explicitly the map from $H_f^2(L,V)$ in $\mathrm{Ext}(L,V,f)$. Let $c \in Z_f^2(L,V)$. We define on the set $E = V \times L$ the following composition law:

$$[(v_1,X_2),(v_2,X_2)] = (f(X_1)v_2 - f(X_2)v_1 + c([X_1,X_2]),[X_1,X_2]). \quad (3.5)$$

Then, one can prove that $[\ ,\ ]$ is a Lie bracket, so $(E,[\ ,\ ])$ is a Lie algebra denoted by E_c. If $\alpha: V \to E_c$ and $\beta: E_c \to L$ are given by $\alpha(v) = (v,e)$ and $\beta(v,X) = X$, then (E_c,α,β) is an extension of V by L associated with f.

So we have a map from $Z_f^2(L,V)$ into the set of extensions of V by L associated with f. One can prove easily that this map factorizes to a map from $H_f^2(L,V)$ into $\mathrm{Ext}(L,V,f)$. The construction of the inverse of this map is relatively easy to do and can be found in Ref. 6.

(3) The last step of the proof of the main result is to establish a one to one correspondence between $\mathrm{Ext}(G,K,a)$ and $\mathrm{Ext}(\mathrm{Lie}\ G,\mathrm{Lie}\ K,\dot{a})$. If G is connected and simply connected, this can be done using the connection between Lie groups and Lie algebras.[7]

Remarks: (1) If G is connected but not simply connected, one proceeds as follows.[8] Let G^* be a connected and simply connected covering group for G, and $\epsilon: G^* \to G$ be the covering homomorphism. Let $c \in Z_a^2(G,K)$. Then we can define a^*: $G^* \to \mathrm{Aut}\,K$ by the formula $a^* = a \circ \epsilon$, and $c^* \in C^2(G^*,K)$ by the formula $c^*(g_1^*,g_2^*) = c(\epsilon(g_1^*),\epsilon(g_2^*))$. Then a^* is a homomorphism and $c^* \in Z_{a^*}^2(G^*,K)$. So we have a map from $Z_a^2(G,K)$ into $Z_{a^*}^2(G^*,K)$. It is easy to see that this map factorizes to a map from $H_a^2(G,K)$ into $H_{a^*}^2(G^*,K)$. We can analyze $H_{a^*}^2(G^*,K)$ with the help of the Hochschild theorem, and then we must study the inverse of this map.

A similar comment is also valid for $Z_a^1(G,K)$ and $H_a^1(G,K)$ if G is not simply connected.

(4) We see that in order to apply the main theorem from Sec. II we must study $H_a^2(\mathrm{Lie}\ G,\mathbb{R})$. This is another similarity with the problem studied in Ref. 3, at least at the formal level.

IV. EXAMPLES

We now apply the theory developed in the preceding sections on some examples also treated in Ref. 3 from the point of view of the Hamiltonian formalism, namely, the Galilei group (in three and in one dimension), the one-dimensional Poincaré group, and the one-dimensional Newton group. In all these cases, G acts transitively on variables $(\xi,\hat{p})$.

A. The Galilei group in three dimensions $\mathcal{G}^{\uparrow}_{+}$

As in Ref. 3, we identify this group with 5×5 real matrices of the form

$$(R,v,\eta,a) \equiv \begin{pmatrix} R & v & a \\ 0 & 1 & \eta \\ 0 & 0 & 1 \end{pmatrix},$$

where $R \in \mathrm{SO}(3)$, $v,a \in \mathbb{R}^3$, and $\eta \in \mathbb{R}$. The composition law is induced by matrix multiplication:

$$(R,v,\eta,a)(R',v',\eta',a') = (RR', Rv' + v, \eta + \eta', Ra' + a + \eta' v).$$

The group $\mathcal{G}^{\uparrow}_{+}$ acts in $\mathbb{R}^4$ by the following formula:

$$\varphi_{R,v,\eta,a}(t,x) = (t + \eta, Rx + tv + a) \quad (t \in \mathbb{R},\ x \in \mathbb{R}^3).$$

As in Ref. 1 we take as a reference point $(0,0) \in \mathbb{R}^4$. Then the stability subgroup H of $(0,0)$ is given by

$$H = \{(R,v,0,0) \mid R \in \mathrm{SO}(3),\ v \in \mathbb{R}^3\}.$$

Because $\mathcal{G}^{\uparrow}_{+}$ is a semidirect product of $\mathbb{R}^4$ and H, the bundle $(\mathcal{G}^{\uparrow}_{+}/H, \mathcal{G}^{\uparrow}_{+}, \pi)$ is trivial. In fact, a smooth cross section is [see Ref. 1, formula (49)]

$$\sigma(t,x) = (1,0,t,x).$$

Here $\mathcal{G}^{\uparrow}_{+}$ admits the following real one-dimensional representations[2]:

$$a_{R,v,\eta,a} = e^{\rho\eta} \quad (\rho \in \mathbb{R} \setminus \{0\}). \quad (4.1)$$

The Lie algebra of $\mathcal{G}^{\uparrow}_{+}$ is identified with the linear subspace of 5×5 matrices of the form

$$(A,u,t,x) \equiv \begin{pmatrix} A & u & x \\ 0 & 0 & t \\ 0 & 0 & 0 \end{pmatrix},$$

where $t \in \mathbb{R}$, $u,x \in \mathbb{R}^3$, and A is a real 3×3 antisymmetric matrix. Then the following representation of Lie $\mathcal{G}^{\uparrow}_{+}$ corresponds to (4.1):

$$\dot{a}(A,u,t,x) = \rho t.$$

(I) Because $a|H = 1$, $Z_{a|H}^2(H,\mathbb{R}) = Z^1(H,\mathbb{R})$. We analyze first $Z^1(H^*,\mathbb{R})$, where H^* is a covering of H given by

$$H^* = \{(U,v) \mid U \in \mathrm{SU}(2),\ v \in \mathbb{R}^3\},$$

with the composition law

$$(U,v)(U',v') = (UU',\epsilon(U)v' + v).$$

Here ϵ: SU(2)$\to$SO(3) is the well known covering homomorphism [see Ref. 8, Chap. XII, formula (48)]. The Lie algebra Lie H^* can be identified with the subalgebra of Lie $\mathscr{G}^{\uparrow}_{+}$ formed by elements of the form $(A,u,0,0)$.

We take the basis j_i, g_i $(i = 1,2,3)$ given by

$$j_i \equiv \begin{pmatrix} l_i & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad g_i = \begin{pmatrix} 0 & e_i & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

[*Here* $(l_i)_{jk} = -\epsilon_{ijk}$ and $(e_i)_j = \delta_{ij}$.] The cocycle identity is

$$c \in Z^1(\text{Lie } H^*,\mathbb{R}) \text{ iff } c([X,Y]) = 0, \quad \forall X,Y \in \text{Lie } H^*.$$

If we take $X = j_i$, $Y = j_k$, we get $c(j_l) = 0$. If we take $X = j_i$, $Y = g_j$, we get $c(g_e) = 0$. So $c = 0$. It follows that $Z^1(\text{Lie } H^*,\mathbb{R}) = \{0\}$. According to (A) from Sec. III, it follows that $Z^1(H^*,\mathbb{R}) = \{0\}$. Using the last remark in Sec. III we get $Z^1(H,\mathbb{R}) = \{0\}$.

(II) Let $(\mathscr{G}^{\uparrow}_{+})^*$ be the covering group of $\mathscr{G}^{\uparrow}_{+}$ defined as

$$(\mathscr{G}^{\uparrow}_{+})^* \equiv \{(U,v,\eta,a) \mid U \in \text{SU}(2),\ v,a \in \mathbb{R}^3,\ \eta \in \mathbb{R}\},$$

with the composition law

$$(U,v,\eta,a)(U',v',\eta',a') = (UU',\epsilon(U)v' + v,\eta + \eta',\epsilon(U)a' + a + \eta' v),$$

where ϵ: SU(2)$\to$SO(3) was used in I. We also denote by ϵ the covering homomorphism from $(\mathscr{G}^{\uparrow}_{+})^*$ into $\mathscr{G}^{\uparrow}_{+}$. One knows that Lie $(\mathscr{G}^{\uparrow}_{+})^* =$ Lie $\mathscr{G}^{\uparrow}_{+}$ and $\dot{a}^* = \dot{a}$. Also in Ref. 3 we have proved that $H^2_{\dot{a}}(\text{Lie } \mathscr{G}_{+},\mathbb{R}) = \{0\}$. Now let $c \in Z^2_{\dot{a}}(\mathscr{G}^{\uparrow}_{+},\mathbb{R})$. According to the last remark in Sec. III,

$$c^* \in Z^2_{\dot{a}^*}((\mathscr{G}^{\uparrow}_{+})^*,\mathbb{R}) = B^2_{\dot{a}^*}((\mathscr{G}^{\uparrow}_{+})^*,\mathbb{R}),$$

i.e., there exists b^*: $(\mathscr{G}^{\uparrow}_{+})^* \to \mathbb{R}$ such that

$$c(\epsilon(g_1^*),\epsilon(g_2^*)) = a_{\epsilon(g_1^*)} b^*(g_2^*) - b^*(g_1^* g_2^*) + b(g_1^*). \tag{4.2}$$

Taking in this equality $g_1^* = g_2^* = (\pm 1,0,0,0)$ we get easily that $b^*(\pm 1,0,0,0) = 0$. We now take in (4.2) $g_2^* = (-1,0,0,0)$ and $g_1^* = (U,v,\eta,a)$. We get

$$b^*(U,v,\eta,a) = b^*(-U,v,\eta,a).$$

So there is a map b: $\mathscr{G}^{\uparrow}_{+} \to \mathbb{R}$ such that $b^* = b \circ \epsilon$. Then from (4.2) and the surjectivity of ϵ it follows that $c = \delta b$, i.e., $c \in B^2_{\dot{a}}(\mathscr{G}^{\uparrow}_{+},\mathbb{R})$. We have proved that $H^2_{\dot{a}}(\mathscr{G}^{\uparrow}_{+},\mathbb{R}) = 0$.

(III) If we apply (8) from Sec. II we get $H^2_{\dot{a}}(\mathscr{G}^{\uparrow}_{+},H,\mathbb{R}) = 0$. From (7) in Sec. II we have $H_a = \{0\}$. We can take therefore $\Lambda = 0$, and (2.3) gives a Lagrangian of the form

$$L(t,x,v) = ce^{\rho t}, \text{ i.e., } \tilde{L} \sim 0, \quad \text{so } \mathscr{H} = 0.$$

We get nothing interesting from the physical point of view.

B. The Galilei group in one dimension[1]

As in Ref. 3 we identify this group with 3×3 real matrices of the form

$$(v,\eta,a) \equiv \begin{pmatrix} 1 & v & a \\ 0 & 1 & \eta \\ 0 & 0 & 0 \end{pmatrix},$$

where $v,\eta,a \in \mathbb{R}$. The composition law is induced by the matrix multiplication

$$(v,\eta,a)(v',\eta',a') = (v + v',\eta + \eta',a + a' + v\eta').$$

This group acts on $\mathbb{R}^2$ by the following law:

$$\varphi_{v,\eta,a}(t,x) = (t + \eta, x + vt + a), \quad t,x \in \mathbb{R}.$$

As in Ref. 1 we take as a reference point $(0,0) \in \mathbb{R}^2$. Then the stability subgroup H of $(0,0)$ is given by

$$H = \{(v,0,0) \mid v \in \mathbb{R}\}.$$

We have a smooth cross section given by

$$\sigma(t,x) = (0,t,x).$$

This group admits one-dimensional real representations of the form

$$a_{v,\eta,a} = e^{\zeta v + \rho\eta} \quad (\zeta,\rho \in \mathbb{R},\ \zeta^2 + \rho^2 \neq 0). \tag{4.3}$$

The Lie algebra of this group can be identified with the linear space of 3×3 matrices of the form

$$(u,t,x) \equiv \begin{pmatrix} 1 & u & x \\ 0 & 0 & t \\ 0 & 0 & 0 \end{pmatrix},$$

with $u,t,x \in \mathbb{R}$. Then the following representation of the Lie algebra corresponds to (4.3):

$$\dot{a}(u,t,v) = \zeta u + \rho t.$$

(I) From (4.3) we have $(a|H)_v = e^{\zeta v}$. Because H is connected and simply connected we can determine $Z^1_{a|H}(H,\mathbb{R})$ by studying first $Z^1_{\widehat{a|H}}(\text{Lie } H,\mathbb{R})$. The generic element of $C^1(\text{Lie } H,\mathbb{R})$ is of the form $c(u) = cu$, with $c \in \mathbb{R}$, and is evidently in $Z^1_{\widehat{a|H}}(\text{Lie } H,\mathbb{R})$. We find the corresponding element in $Z^1_{a|H}(H,\mathbb{R})$ following the method in part (A) of Sec. III:

$$b(v) = c(e^{\zeta v} - 1).$$

Let $b_1 \in C^1(G,\mathbb{R})$ be given by

$$b_1(v,\eta,a) = c(e^{\zeta v + \rho t} - 1).$$

Evidently we have $b = b_1|H$ and $b_1 \in Z^1_a(G,\mathbb{R})$ so it follows that $b \in r(Z^1_a(G,\mathbb{R}))$.

(II) We now study the group $H^2_a(G,\mathbb{R})$ using the method from part (B) of Sec. III because G is connected and simply connected. We first determine $H^2_{\dot{a}}(\text{Lie } G,\mathbb{R})$. We have determined in Ref. 3 the most general element from $Z^2_{\dot{a}}(\text{Lie } G,\mathbb{R})$ in the form

$$c((u,t,x),(u',t',x')) = \lambda\zeta(xu' - x'u) - \lambda\rho(tx' - t'x) + \sigma(tu' - t'u), \tag{4.4}$$

with $\lambda,\sigma \in \mathbb{R}$. On the other hand, the most general element in $C^1(\text{Lie } G,\mathbb{R})$ is of the form

$$b(u,t,x) = \alpha u + \beta t + \gamma x \quad (\alpha,\beta,\gamma \in \mathbb{R}).$$

A simple computation shows that by choosing α, β, and γ conveniently we can arrange that c given by (4.4) is of the

form δb. It follows that $H_a^2(\text{Lie } G,\mathbb{R}) = \{0\}$ so by part (B) of Sec. III, $H_a^2(G,\mathbb{R}) = \{0\}$.

(III) As in the preceding example, we have $H_a = \{[0]\}$. If we take $\Lambda = 0$, then (2.3) gives

$$L(t,x,v) = ce^{\rho t + \xi v}, \tag{4.5}$$

so $\hat{\mathscr{H}}$ contains only one element.

The case $\xi = 0$ is uninteresting. If $\xi \neq 0$, the Euler–Lagrange equation for (4.5) is

$$\ddot{x} = -\rho/\xi,$$

i.e., a uniform accelerated motion.

C. The Poincaré group in one dimension[1]

As in Ref. 3, we identify this group with the set of 3×3 real matrices of the form

$$(\chi,\eta,a) = \begin{pmatrix} \cosh\chi & \sinh\chi & \eta \\ \sinh\chi & \cosh\chi & a \\ 0 & 0 & 1 \end{pmatrix},$$

with $\chi,\eta,a\in\mathbb{R}$. The composition law is induced by the matrix multiplication

$$\begin{aligned}(\chi,\eta,a)(\chi',\eta',a') &= (\chi+\chi', \cosh(\chi)\eta' + \sinh(\chi)a' + \eta, \sinh(\chi)\eta' \\ &\quad + \cosh(\chi)a' + a).\end{aligned} \tag{4.6}$$

This group acts on $\mathbb{R}^2$ by the following formula:

$$\begin{aligned}\varphi_{\chi,\eta,a}(t,x) &= (\cosh(\chi)t + \sinh(\chi)x + \eta, \sinh(\chi)t \\ &\quad + \cosh(\chi)x + a).\end{aligned}$$

We take (0,0) as a reference point and the stability subgroup is then

$$H = \{(\chi,0,0)\,|\,\chi\in\mathbb{R}\}.$$

We have a smooth cross section

$$\sigma(t,x) = (0,t,x).$$

This group admits one-dimensional real representations of the form

$$a_{\chi,\eta,a} = e^{\rho\chi} \quad (\rho\in\mathbb{R}\setminus\{0\}). \tag{4.7}$$

The Lie algebra of this group can be identified with real 3×3 matrices of the form

$$(u,t,x) \equiv \begin{pmatrix} 0 & u & t \\ u & 0 & x \\ 0 & 0 & 0 \end{pmatrix}.$$

Then the following representation of the Lie algebra corresponds to (4.7):

$$\dot{a}(u,t,x) = \rho u.$$

(I) From (4.7) we have $(a|H)_\chi = e^{\rho\chi}$. Because H is connected and simply connected we determine $Z^2_{a|H}(H,\mathbb{R})$ by the method from part (A) of Sec. III. The most general element in $C^1(\text{Lie } H,\mathbb{R})$ is of the form

$$c(u) = cu, \quad \text{with } c\in\mathbb{R},$$

and is also in $Z^1_{a|H}(\text{Lie } H,\mathbb{R})$. As in (B), the corresponding element in $Z^1_{a|H}(H,\mathbb{R})$ is

$$b(\chi) = c(e^{\rho\chi} - 1).$$

Define $b_1\in C^1(G,\mathbb{R})$ by

$$b_1(\chi,\eta,a) = c(e^{\rho\chi} - 1).$$

Then $b_1\in Z^1_a(G,\mathbb{R})$ and evidently $b_1|H = b$, so $b\in r(Z^1_a(G,\mathbb{R}))$.

(II) We study $H_a^2(G,\mathbb{R})$ by the method of part (B) of Sec. III, because G is connected and simply connected. We first determine $H_a^2(\text{Lie } G,\mathbb{R})$. In Ref. 3 we have established that the most general element in $Z_a^2(\text{Lie } G,\mathbb{R})$ is of the form

$$c((u,t,x),(u',t',x')) = \lambda(xu' - x'u) + \mu(u't - ut'), \tag{4.8}$$

with $\lambda,\mu\in\mathbb{R}$. On the other hand, the most general element of $C^1(\text{Lie } G,\mathbb{R})$ is of the form

$$b(u,t,x) = \alpha u + \beta t + \gamma x \quad (\alpha,\beta,\gamma\in\mathbb{R}).$$

We have two distinct cases.

(a) $\rho \neq \pm 1$. Then by taking α,β,γ conveniently we can arrange such that c given by (4.8) is of the form δb. It follows that in this case $H_a^2(\text{Lie } G,\mathbb{R}) = \{0\}$, so by part (B) of Sec. III, $H_a^2(G,\mathbb{R}) = \{0\}$.

(b) $\rho = \pm 1$. In this case $H_a^2(\text{Lie } G,\mathbb{R})$ is one dimensional and every element in $Z_a^2(\text{Lie } G,\mathbb{R})$ is cohomologous with one of the form

$$c_f((u,t,x),(u',t',x')) = f(xu' - x'u). \tag{4.9}$$

By the result of part (B) of Sec. III, $H_a^2(G,\mathbb{R})$ is also one dimensional. One can determine (see the Appendix) the element in $Z_a^2(G,\mathbb{R})$ corresponding to (4.9):

$$\begin{aligned}\bar{c}_f((\chi,\eta,a),(\chi',\eta',a')) &= (f/2)[(\chi e^{\pm\chi} - \sinh\chi)\eta' \pm (\chi e^{\pm\chi} + \sinh\chi)a'] \end{aligned} \tag{4.10}$$

(the signs $\pm$ are in accordance with $\rho = \pm 1$). So every element in $Z_a^2(G,\mathbb{R})$ is cohomologous with one of the form (4.10).

(III) (a) In the case $\rho \neq \pm 1$, we have, as in the preceding examples, $H_a = \{0\}$. If we take $\Lambda = 0$, then (2.3) gives

$$L(t,x,v) = c\sqrt{1-v^2}\,((1+v)/(1-v))^{\rho/2}, \tag{4.11}$$

so $\hat{\mathscr{H}}$ is formed by a single element.

(b) In the case $\rho = \pm 1$, it follows from (8) of Sec. II that every admissible class in $H_a^2(G,\mathbb{R})$ contains one and only one class from $H_a^2(G,H,\mathbb{R})$. Because $\bar{c}_f$, given by (4.10), verifies $\bar{c}_f|H\times H = 0$, every class in $H_a^2(G,\mathbb{R})$ is admissible; thus $H_a^2(G,H,\mathbb{R})$ is in one to one correspondence with $H_a^2(G,\mathbb{R})$. We can now apply formula (2.12) with $c_0 = \bar{c}_f$ and $b = 0$. Taking (4.6) into account, we get the following gauge functions:

$$\begin{aligned}\Lambda((\chi,\eta,a),(t,x)) &= (f/2)[(\chi e^{\pm\chi} - \sinh\chi)t \pm (\chi e^{\pm\chi} + \sinh\chi)x]. \end{aligned} \tag{4.12}$$

If we now apply (2.3) we get

$$L(t,x,v) = (f)/4)(1\pm v)\ln[(1+v)/(1-v)]. \tag{4.13}$$

For $\Lambda = 0$, we get, from (2.3), formula (4.11) for $\rho = \pm 1$. These Lagrangians are equivalent to 0. Thus, $\hat{\mathscr{H}}$ is formed by a single element.

We remark that both (4.11) and (4.13) describe free motions.

D. The Newton group[1]

As in Ref. 3 we identify this group with 3×3 real matrices of the form

$$(\eta,a,v)\equiv\begin{pmatrix}\cosh\eta & \sinh\eta & 0\\ \sinh\eta & \cosh\eta & 0\\ a & v & 1\end{pmatrix},$$

with $\eta,a,v\in\mathbb{R}$. The composition law is induced by the matrix multiplication

$$(\eta,a,v)(\eta',a',v')$$
$$=((\eta+\eta',\cosh(\eta')a+\sinh(\eta')v+a',$$
$$\sinh(\eta')a+\cosh(\eta')v+v').$$

This group acts on $\mathbb{R}^2$ by the following formula:

$$\varphi_{\eta,a,v}(t,x)=(t+\eta,x+v\sinh(\eta)t+a\cosh(\eta)t).$$

We take, as before, (0,0) as the reference point. The stability subgroup is then

$$H=\{(0,0,v)\,|\,v\in\mathbb{R}\}.$$

We have a smooth cross section

$$\sigma(t,x)=(t,x,0). \tag{4.14}$$

This group admits one-dimensional real representations of the form

$$a_{\eta,a,v}=e^{\rho\eta}\quad(\rho\in\mathbb{R}\setminus\{0\}). \tag{4.15}$$

The Lie algebra of this group can be identified with real 3×3 matrices of the form

$$(t,x,u)\equiv\begin{pmatrix}0 & t & 0\\ t & 0 & 0\\ x & u & 0\end{pmatrix}.$$

Then the following representation of the Lie algebra corresponds to (4.15):

$$\dot{a}(t,x,u)=\rho t.$$

(I) From (4.15) we have $a|H=1$. We can easily determine that the elements of $Z^1(H,\mathbb{R})$ are of the form

$$b(v)=fv, \tag{4.16}$$

with $f\in\mathbb{R}$.

We turn now to $Z^1_a(G,\mathbb{R})$. First we determine $Z^1_a(\text{Lie } G,\mathbb{R})$. By simple calculations one establishes that there are two cases.

(a) $\rho\neq\pm 1$. In this case, $H^1_a(\text{Lie } G,\mathbb{R})=\{0\}$ so $H^1_a(G,\mathbb{R})=\{0\}$ by part (A) of Sec. III. It follows that the most general element in $Z^1_a(G,\mathbb{R})$ is of the form

$$c(\eta,a,v)=b(e^{\rho\eta}-1).$$

In particular, $r(Z^1_a(G,\mathbb{R}))=0$.

(b) $\rho=\pm 1$. In this case every cocycle in $Z^1_a(\text{Lie } G,\mathbb{R})$ is equivalent to one of the form

$$c(t,x,u)=\alpha(x\mp u)\quad(\alpha\in\mathbb{R}).$$

With the method of part (A) of Sec. III we get the corresponding element in $Z^1_a(G,\mathbb{R})$:

$$c(\eta,a,v)=\alpha e^{\pm\eta}(a\mp v).$$

In particular, $r(Z^1_a(G,\mathbb{R}))$ is formed by elements of the form (4.16).

(II) We now study $H^2_a(G,\mathbb{R})$. The most general element in $Z^2_a(\text{Lie } G,\mathbb{R})$ is of the form

$$c((t,x,u),(t',x',u'))=\lambda(xt'-x't)+\sigma(ut'-u't).$$

We again have two cases.

(a) $\rho\neq\pm 1$. In this case $H^2_a(\text{Lie } G,\mathbb{R})=\{0\}$, so $H^2_a(G,\mathbb{R})=\{0\}$.

(b) $\rho=\pm 1$. In this case $H^2_a(\text{Lie } G,\mathbb{R})$ is one dimensional, and every two-cocycle is cohomologous with one of the form

$$c_\alpha((t,x,u),(t',x',u'))=\alpha(t'x-tx').$$

It follows, by an analysis similar to the one for the one-dimensional Poincaré group, that $H^2_a(G,\mathbb{R})$ is one dimensional and every two-cocycle is cohomologous with one of the form

$$\bar{c}_\alpha((\eta,a,v),(\eta',a',v'))$$
$$=\tfrac{1}{2}\alpha e^{\pm\eta}[a(\eta'+e^{\pm\eta'}\sinh\eta')$$
$$+v(\eta'\mp e^{\pm\eta'}\sinh\eta')]. \tag{4.17}$$

(III) (a) $\rho\neq\pm 1$. From (8) in Sec. III it follows that $H^2_a(G,H,\mathbb{R})$ is in one to one correspondence with $Z^1(H,\mathbb{R})$. We can apply (2.12) with $c_0=0$, b given by (4.16), and σ given by (4.14). We get the following gauge functions:

$$\Lambda((\eta,a,v),(t,x))=fe^{\rho(\eta+t)}(a\sinh t+v\cosh t).$$

We now apply (2.3) and get

$$L(t,x,v)=fe^{\rho t}(x+\rho v+c),$$

i.e., $\tilde{L}\in\mathscr{L}_0$ so $\mathscr{H}=0$.

(b) $\rho=\pm 1$. From (7) and (8) of Sec. II it follows that H_a is in one to one correspondence with the admissible classes from $H^2_a(G,\mathbb{R})$. But $\bar{c}_\alpha|G\times H=0$, so every class in $H^2_a(G,\mathbb{R})$ is admissible. We now apply (2.12) with $c_0=\bar{c}_\alpha$ and $b=0$, and get that every gauge function is equivalent to one of the form

$$\Lambda((\eta,a,v),(t,x))$$
$$=(\alpha/2)e^{\pm\eta}[a(t+e^{\pm t}\sinh t)$$
$$+v(t\mp e^{\pm t}\sinh t)].$$

We now apply (2.2) and get

$$L(t,x,v)=\alpha e^{t}(x+c)$$

and, respectively,

$$L(t,x,v)=\alpha e^{-t}(v+c),$$

i.e., they are in $\mathscr{L}_0$. So $\mathscr{H}=0$. We get nothing interesting from the physical point of view.

E. Final remark

If we analyze in this spirit the cases studied in Ref. 1, i.e., we take the trivial one-dimensional representation of G ($a=1$, $\forall g\in G$), then the only minor modification is that one must identify, from the physical point of view, Lagrangians differing by a multiplicative constant. This is in accordance with a similar phenomenon from Ref. 3.

V. CONCLUSION

We have proved that the result in Ref. 1 extends to a more general case, related to a similar problem studied in Ref. 3. From the physical point of view the result is rather negative in the sense that it does not furnish new Lagrangians admitting a Lie group of symmetry (and thus interesting from the physical point of view), besides those in Ref. 1, at least for the four groups studied here. One exception is (4.5). Of course, it is possible in principle that for other groups appearing in physics this enlarged concept of symmetry gives new and interesting Lagrangians. This remains to be investigated.

Another problem that deserves clarification is the relation between the analysis in Ref. 3 and in this paper. We have remarked at the formal level some similarities in Secs. II and III. Another similarity is the following. Suppose that $X \in \text{Lie } G$ is such that $\dot{a}(X) = 0$. Then a conservation law corresponds to the X's by the Noether theorem. This has also been remarked in Ref. 3.

It would be interesting to formulate the connection between the two problems at the abstract level. This can probably be done using a more general concept of symmetry based on the Lagrange–Souriau form,[9] and it will be done elsewhere.

At last, we remark that this analysis can be extended to classical field theory following the lines in Ref. 10, and reformulated completely in cohomological terms.

ACKNOWLEDGMENT

The authors wishes to thank T. O. Popp for reading the manuscript and making many valuable comments.

APPENDIX: EXPLICIT COMPUTATIONS OF TWO-COCYCLES

From (3.5) we can find the extension of $\mathbb{R}$ by Lie G corresponding to the cocycle c_f. We denote a generic element of E_{c_f} by (k,u,t,x) and take the following basis:

$$\epsilon \equiv (1,0,0,0), \quad \bar{k} \equiv (0,1,0,0),$$
$$\bar{h} \equiv (0,0,1,0), \quad \bar{p} \equiv (0,0,0,1).$$

Then the commutation relations are

$$[p,\bar{k}] = -h - f\epsilon, \quad [\bar{k},\bar{h}] = \bar{p}, \quad [\bar{h},\bar{p}] = 0,$$
$$[\bar{p},\epsilon] = 0, \quad [\bar{k},\epsilon] = \pm\epsilon, \quad [\bar{h},\epsilon] = 0. \tag{A1}$$

To determine the corresponding group cocycle c_f, we must first determine the extension (H,i,j) of $\mathbb{R}$ by G related to the Lie algebra extension above, with H a connected and simply connected group.

One has, from (3.4),

$$(0,g_1)(0,g_2) = (\bar{c}_f(g_1,g_2),e)(0,g,g_2). \tag{A2}$$

From this relation one can now find $\bar{c}_f$.

The program goes as follows. First, we note that every element (χ,η,a) can be written as

$$(\chi,\eta,a) = e^{\eta h} e^{ap} e^{\chi k},$$

where h, a, and k are the 3×3 real matrices introduced in Ref. 2. Because we also have from (3.4)

$$(k,g) = (k,e)(0,g),$$

we can write any element in H in the form

$$(\exp k\epsilon)(\exp \eta\bar{h})(\exp a\bar{p})(\exp \chi\bar{k}),$$

where exp is the exponential map in H and we have identified the Lie algebra of $\mathbb{R}$ with the group itself, so the exponential map in the $\mathbb{R}$ subgroup of H is the identity: $\exp k\epsilon = k$. From (A2) and the composition law of the group it follows that $\bar{c}_f$ verifies the following relation:

$$\begin{aligned}(\exp \eta\bar{h})(\exp a\bar{p})(\exp \chi\bar{k})(\exp \eta'\bar{h})(\exp a'\bar{p})(\exp \chi'\bar{k}) \\ = (\exp \bar{c}_f((\chi,\eta,a),(\chi',\eta',a'))\epsilon)(\exp(\cosh(\chi)\eta' \\ + \sinh(\chi)a' + \eta)\bar{h})(\exp(\sinh(\chi)\eta' \\ + \cosh(\chi)a' + a)\bar{p})(\exp(\chi+\chi')\bar{k}).\end{aligned} \tag{A3}$$

From this relation we can now determine $\bar{c}_f$ with the help of the commutation relations (A1). We must commute $\exp \chi\bar{k}$ over $\exp \eta'\bar{h}$ and $\exp a'\bar{p}$.

This can easily be done as follows. First we use the well known identity

$$(\exp A)B\exp(-A) = \sum_{n=0}^{\infty} \frac{1}{n!}[A,B]_n, \tag{A4}$$

where $[A,B]_n$ are defined recursively by $[A,B]_0 \equiv B$, and $[A,B]_{n+1} = [A,[A,B]_n]$. With this identity one easily has, with the help of (A1),

$$\begin{aligned}&(\exp \chi\bar{k})\bar{h}\exp(-\chi\bar{k}) \\ &\quad = \cosh(\chi)\bar{h} + \sinh(\chi)\bar{p} + (f/2)(\chi e^{\pm\chi}\sinh\chi)\epsilon.\end{aligned}$$

Because $\bar{h}$, $\bar{p}$, and ϵ commute among themselves one now has

$$\begin{aligned}&(\exp \chi\bar{k})(\exp \eta'\bar{h})(\exp(-\chi\bar{k})) \\ &\quad = (\exp(f/2)\eta'(\chi e^{\pm\chi}\sinh\chi)\epsilon)(\exp \eta'\cosh(\chi)\bar{h}) \\ &\qquad \times(\exp \eta'\sinh(\chi)\bar{p}).\end{aligned}$$

Analogously one gets

$$\begin{aligned}&(\exp \chi\bar{k})(\exp a'p)(\exp(-\chi\bar{k})) \\ &\quad = (\exp \pm(f/2)a'(\chi e^{\pm\chi} + \sinh\chi)\epsilon)((\exp a'\sinh(\chi)\bar{h} \\ &\qquad \times(\exp a'\cosh(\chi)\bar{p}).\end{aligned}$$

Using the last two relations in (A3), we get (4.10).

Remarks: (1) One could also determine $\bar{c}_f$ by the procedure of Bargmann,[7,11] i.e., searching for a $\bar{c}_f$ as a polynomial in η,η',a,a', because these are changed linearly in the composition law. Starting with a polynomial of degree 1, with coefficients depending on χ and χ', we can get, after some computations, (4.10). The problem is that in this way we have no guarantee that $\bar{c}_f$ is not in $B_a^2(G,\mathbb{R})$. In the case studied by Bargmann, namely, the three-dimensional Galilei group (with $a = 1$), this is resolved by observing that if $c \in B^2(G,\mathbb{R})$ and g_1 and g_2 commute, then $c(g_1,g_2) = c(g_2,g_1)$. This property must be contradicted by some special choice of g_1 and g_2 (see Ref. 8, Chap. X).

In the case when a is nontrivial this trick does not work and there is no immediate generalization available, but there is an alternative method based on a generalization of Lemma

10.36 from Ref. 8. Namely, one can prove that if $\bar{c} \in Z_a^2(G,\mathbb{R}^n)$ then the corresponding element $c \in Z_a^2(\text{Lie } G,\mathbb{R}^n)$ is given by the following formula:

$$c(X,Y) \equiv \frac{\tilde{\partial}\bar{c}}{\partial t\, \partial u}(\exp tX, \exp uY)\,|_{t=u=0} - (X \leftrightarrow Y).$$

Using this relation, one can prove that $\bar{c}_f$ given by (4.10) is related to c_f given by (4.9).

(2) Of course, one can use the method developed here to recover the well known expression of a nontrivial cocycle in $\mathcal{G}^1_+$ (for $a = 1$) [formula (137) from Ref. 8, Chap. X], and also the cocycles used in Ref. 1.

[1] J. M. Levy-Lebond, Commun. Math. Phys. **12**, 64 (1969).
[2] T. O. Popp, "A theorem in cohomology for elementary classical systems," preprint FT-319-1987.
[3] D. R. Grigore, "A possible generalization of the concept of symmetry in analytical mechanics," preprint FT-316-1987, to appear in J. Math. Phys.
[4] A. Kirillov, *Elements de la theorie des représentations* (Mir, Moscow, 1974).
[5] S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[6] M. Goto and F. D. Grosshans, *Semisimple Lie Algebras* (Marcel Dekker, New York, 1978).
[7] G. Hochschild, Ann. Math. **34**, 96 (1951).
[8] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand Reinhold, New York, 1970), Vol. II.
[9] I. M. Souriau, *Structure de systemes dynamiques* (Dunod, Paris, 1970).
[10] T. O. Popp and D. R. Grigore, "Total divergence Lagrangians and Poincaré–Cartan form in classical field theory," preprint FT-314-1987.
[11] V. Bargmann, Ann. Math. **59**, 1 (1954).

Mastersymmetries, angle variables, and recursion operator of the relativistic Toda lattice

Walter Oevel, Benno Fuchssteiner, and Hongwei Zhang
Fachbereich Mathematik, Universität Paderborn, Paderborn, Federal Republic of Germany

Orlando Ragnisco
Dipartimento di Fisica, Università degli Studi "La Sapienza," Rome, Italy

(Received 5 July 1988; accepted for publication 10 May 1989)

Conserved quantities, bi-Hamiltonian formulation, and recursive structure of the relativistic Toda lattice (RT) are obtained in an algorithmic way without making use of the Lax representation. Furthermore, for the multisoliton solutions the gradients of the angle variables are described in terms of mastersymmetries. A new hierarchy of completely integrable systems is discovered, which turns out to correspond to the "negative" of the hierarchy of RT. Thus it is shown that the full algebra of time-dependent symmetry group generators for each member of the RT hierarchy is isomorphic to the algebra of first order differential operators with Laurent polynomials as coefficients. The surprising phenomenon is revealed that the members of the RT hierarchy are connected to their negative counterparts by explicit Bäcklund transformation.

I. INTRODUCTION

In the field of completely integrable Hamiltonian systems of finite dimension, outstanding results have recently been obtained by Ruijsenaars. Indeed, he first introduced and solved, together with Schneider, a relativistic version of the Calogero–Moser system[1–3] and then proposed an analogous relativistic extension of the Toda system: in a remarkable paper[4] (yet, to our knowledge, unpublished) he defined a Poincaré-invariant Hamiltonian system which, in the nonrelativistic limit, reduces to the Toda lattice; moreover, in the "free ends" case, he provided a Lax representation, proved the complete integrability, and solved the scattering problem, also establishing a close connection with soliton dynamics. New results on this relativistic Toda lattice, hereafter referred to as RT, have later on been obtained by Bruschi and one of the authors,[5,6] who, starting from the Lax representation, were able to produce a hereditary recursion operator, thus providing an alternative proof of complete integrability. The role played by the Lax representation was also crucial in order to prove the complete integrability and to solve the Cauchy problem in the periodic case.[7]

In this paper, we deal with the infinite lattice and investigate its integrability structure from a Lie-algebraic point of view. This then leads to the action-angle variables of the multisoliton solutions and, in addition, to a new hierarchy of integrable systems consisting of the "inverse members" of the RT hierarchy. In contrast to the previous work[5–7] our approach does not make any use of the Lax representation, but we will rather rely on the concept of "mastersymmetries."

We recall here the basic definitions for the RT system. In terms of the canonically conjugated variables $\{q_n,\Theta_n\}_{n\in\mathbb{Z}}$ the Hamiltonian for the infinite RT is given by

$$H(q,\Theta)=\sum_{n\in\mathbb{Z}}\{\exp(\Theta_n)[1+\exp(q_{n-1}-q_n)]^{1/2}\times[1+\exp(q_n-q_{n+1})]^{1/2}-2\},\quad(1.1)$$

so that the equations of motion read

$$\begin{bmatrix}q_n\\ \Theta_n\end{bmatrix}_t=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}\partial H/\partial q_n\\ \partial H/\partial\Theta_n\end{bmatrix}=\begin{bmatrix}b_n\\ 1/2a_{n-1}(b_n+b_{n-1})-1/2a_n(b_n+b_{n+1})\end{bmatrix}\quad(1.2)$$

with

$$b_n=\exp(\Theta_n)[1+\exp(q_{n-1}-q_n)]^{1/2}\times[1+\exp(q_n-q_{n+1})]^{1/2},\quad(1.3a)$$

$$a_n=\frac{\exp(q_n-q_{n+1})}{(1+\exp(q_n-q_{n+1}))},\quad(1.3b)$$

where we assume the natural boundary conditions

$$q_n-q_{n+1}\underset{|n|\to\infty}{\to}0;\quad\Theta_n\underset{|n|\to\infty}{\to}0.\quad(1.4)$$

Much simpler forms of the evolution equations, more convenient for our purposes, can be obtained by rewriting the dynamics in terms of the a_n and the b_n as defined above in (1.3). Then the boundary conditions at infinity are

$$b_n\underset{|n|\to\infty}{\to}2;\quad a_n\underset{|n|\to\infty}{\to}1/2,$$

the Hamiltonian reads

$$H(b,a)=\sum_{n\in\mathbb{Z}}(b_n-2),\quad(1.5)$$

and the evolution equations assume the form

$$\frac{d}{dt}\begin{bmatrix}b_n\\ a_n\end{bmatrix}=\begin{bmatrix}b_n(b_{n-1}a_{n-1}-a_nb_{n+1})\\ a_n(1-a_n)(b_n-b_{n+1})\end{bmatrix}=\hat{\Theta}\nabla H(b,a),\quad(1.6a)$$

with

 0022-2488/89/112664-07$02.50

$$\hat{\Theta} = \begin{bmatrix} bT^{-}ba - baT^{+}b; & b(T^{-}-1)a(1-a) \\ a(1-a)(1-T^{+})b; & 0 \end{bmatrix},$$

$$\nabla H(b,a) = \begin{bmatrix} \partial H/\partial b_n \\ \partial H/\partial a_n \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{1.6b}$$

Here 1 denotes the identity operator, the product ab of two sequences $a = (a_n)$ and $b = (b_n)$ is to be understood as the pointwise multiplication $(ab)_n := a_n b_n$, and T^{+} (resp. T^{-}) is to denote the raising (resp. lowering) operators on sequences: $T^{\pm} f_n = f_{n\pm 1}$. The implectic[8] operator $\hat{\Theta}$ arises from the canonical symplectic form using the transformation (1.3). A further useful form of the Hamiltonian and the evolution equations can be obtained by setting

$$d_n := b_n(1-a_n) = \exp(\Theta_n)\left[\frac{1+\exp(q_{n-1}-q_n)}{1+\exp(q_n-q_{n+1})}\right]^{1/2}, \tag{1.7a}$$

$$c_n := b_n a_n = \exp(\Theta_n)\exp(q_n - q_{n+1}) \times \left[\frac{1+\exp(q_{n-1}-q_n)}{1+\exp(q_n-q_{n+1})}\right]^{1/2}, \tag{1.7b}$$

so that now we have

$$H(d,c) = \sum_{n\in\mathbb{Z}} (c_n + d_n - 2), \tag{1.8}$$

and the evolution equations are

$$\begin{bmatrix} d_n \\ c_n \end{bmatrix}_t = \begin{bmatrix} d_n(c_{n-1}-c_n) \\ c_n(d_n - d_{n+1} + c_{n-1} - c_{n+1}) \end{bmatrix} = \Theta^{(0)}\nabla H(d,c), \tag{1.9a}$$

with

$$\Theta^{(0)} = \begin{bmatrix} 0; & d(T^{-}-1)c \\ c(1-T^{+})d; & c(T^{-}-T^{+})c \end{bmatrix},$$

$$\nabla H(d,c) = \begin{bmatrix} \partial H/\partial d_n \\ \partial H/\partial c_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{1.9b}$$

The operator $\Theta^{(0)}$ is obtained from $\hat{\Theta}$ in (1.6) by the usual transformation laws and hence is again an implectic operator.[8–10] The boundary conditions at infinity now are

$$d_n \underset{|n|\to\infty}{\to} 1; \quad c_n \underset{|n|\to\infty}{\to} 1.$$

For the form (1.9) of the relativistic Toda lattice, a hierarchy of mastersymmetries will be constructed in Sec. IV.

II. MASTERSYMMETRIES AND ANGLE VARIABLES

We briefly explain the background which leads to the results of this paper. We consider a Hamiltonian flow,

$$u_t = K(u) = \{H(u),u\} = \Theta(u)\nabla H, \tag{2.1}$$

where u is on a suitable manifold equipped with a Poisson bracket $\{.,.\}$ and $H(u)$ is to denote the Hamiltonian function. Then $K(u)$ denotes the vectorfield defining this flow. There is a homomorphism between the Lie algebra of scalar fields (with the Poisson bracket as Lie product) and the vectorfield Lie algebra; it is given by the operator $\Theta(u)$ arising from the Poisson bracket, and it maps the gradient of $H(u)$ onto the generator of the flow. We call Θ an implectic operator. Looking for conserved quantities and generators of symmetries amounts to finding the commutants of either $H(u)$ (in the Lie algebra given by the Poisson bracket) or $K(u)$ (in the vectorfield Lie algebra). There are many standard methods for recursive generation of commutants, for example by inverse scattering theory, by application of hereditary operators, or by the Hirota bilinear mechanism. Unfortunately, it turns out that in some cases either these standard mechanisms do not work or they are not suitable for algorithmic approaches. Therefore we choose to perform the construction of the commutant by way of mastersymmetries, which is a simple and straightforward method having the additional advantage of being rather accessible for computational methods. Mastersymmetries do exist for almost all known completely integrable systems. They were first discovered in the case of the BO (Benjamin–Ono equation) and the KP (Kadomtsev–Petviashvili equation); they exist for the nonlinear quantum mechanical XYZ-model and the XYh-model and many others. Systematic studies can be found in Ref. 11 and Refs. 9, 10, and 12.

Mastersymmetries are simple to explain: Let $\mathbf{L}$ be a Lie algebra. A *mastersymmetry* τ for $G\in\mathbf{L}$ is an element of $\mathbf{L}$ such that

$$[[\tau,G],G] = 0. \tag{2.2}$$

Thus

$$G^{(1)} := [\tau,G] \tag{2.3}$$

obviously is an element of the commutant of G since $[G^{(1)},G] = 0$. Because of the Jacobi identity,

$$G^{(2)} := [\tau,G^{(1)}] \tag{2.4}$$

also gives a symmetry. Let $G^{@}$ denote the commutant

$$G^{@} = \{\tilde{G}\in\mathbf{L} \mid [\tilde{G},G] = 0\}$$

of G. In the case of an Abelian $G^{@}$ one can continue this process of construction[11] to obtain a sequence

$$G^{(j+1)} := [\tau,G^{(j)}], \quad j\in\mathbb{N} \tag{2.5}$$

of Lie algebra elements commuting with G. Thus commutation with τ maps $G^{@}$ into $G^{@}$ and, under additional and reasonable assumptions, we can expect that $G^{@}$ is generated out of G by successive application of the commutator with τ. So our only task is to find a Lie algebra element τ in L–$G^{@}$ fulfilling (2.2).

If one carries out this task for the Lie algebra of vectorfields one discovers that basically there are two possible situations, depending on whether the mastersymmetries under consideration are Hamiltonian or not.

If the mastersymmetries are Hamiltonian, then applying the Lie algebra homomorphism going from the vectorfields onto the scalar fields, one finds conservation laws linear in time (i.e., angle variables). This happens for equations like BO, KP, and all the other completely integrable systems in $(2+1)$ dimensions. In these cases no recursion operators (in the usual sense) have been found, although recently recursion operators in an extended sense have been discovered.[13] The situation changes drastically when the mastersymmetries are not Hamiltonian, which means that the Lie derivative of the implectic operator Θ into the direction of the mastersymmetries do not vanish. According to an observation of Oevel[12,9] (see also Ref. 10), in most examples of

integrable hierarchies the Lie derivative of Θ into the direction of the mastersymmetry gives rise to a second implectic operator $\tilde{\Theta}$. Hence each nontrivial mastersymmetry yields a candidate for a bi-Hamiltonian formulation for the regarded system. This happens for equations like the Korteweg–de Vries, the modified Korteweg–de Vries, the sine-Gordon, the nonlinear Schrödinger, and so on. Now, taking

$$\Phi = \tilde{\Theta}\Theta^{-1}, \tag{2.6}$$

one finds a recursion operator for the system.

If the second operator $\tilde{\Theta}$ indeed turned out to a yield a second implectic operator, then the above Φ is hereditary automatically. So in any case a nontrivial mastersymmetry provides a heuristic tool for the recursive construction of the wanted hierarchy.

But in addition, the angle variables, at least for the multisoliton manifolds, can be easily obtained by the mastersymmetries even if those are not Hamiltonian. In the first case, where the mastersymmetries are Hamiltonian, this is obvious. In the other case this is far from being obvious and will be reported in detail elsewhere.[14]

Here we just review the results. Let $u_t = K(u)$ be a dynamical system admitting a hereditary recursion operator Φ and a mastersymmetry $\tau^{(1)}$. With suitable assumptions, that are fulfilled for most of the known examples of integrable systems,[9–12,15] we apply Φ to the first mastersymmetry $\tau^{(1)}$ and obtain a sequence of mastersymmetries

$$\tau^{(j)} = \Phi^{j-1}\tau^{(1)}, \quad j = 1,2,3,..., \tag{2.7}$$

satisfying

$$[\tau^{(j)},\tau^{(k)}] = (k-j)\tau^{(k+j)}. \tag{2.8}$$

As Φ is to be hereditary, the vector fields defined via

$$K^{(j+1)} := \Phi^j K \tag{2.9}$$

are the symmetry generators of the system (or the members of the hierarchy, if one likes). Now the following invariant submanifold,

$$\left\{u \middle| \sum_{i=1}^{N+1} \alpha_i K^{(i)} = 0, \quad \alpha_i \in \mathbb{R}\right\}, \tag{2.10}$$

turns out to be the N-soliton manifold[16,17] being of dimension $2N$. The parametrization of this manifold is given by time, the $N-1$ phases, and the N different asymptotic speeds (represented by the variables α_i). A detailed study reveals that, although the $\tau^{(n)}$ themselves are not Hamiltonian, the set of vectorfields consisting of the following linear combinations,

$$A_r = \beta_r \sum_{i=1}^{N} \alpha(r)_i \tau^{(i+r)}, \quad r = 1,...,N, \tag{2.11}$$

are Hamiltonian on the reduced manifold given by (2.10). Here the $\alpha(r)_i$ are the coefficients of the polynomials

$$\sum_{i=1}^{N} \alpha(r)_i \lambda^i = \frac{1}{\lambda - \lambda_r} \sum_{i=1}^{N+1} \alpha_i \lambda^i, \tag{2.12}$$

the β_r are suitable integrating factors, and the λ_r are the zeroes of the polynomial

$$P(\lambda) = \sum_{i=1}^{N+1} \alpha_i \lambda^i.$$

Since the corresponding scalar fields are linear in time [consequence of (2.9)] we thus have found the angle variables for these multisoliton manifolds.

Thus the program for finding the recursion operator and the angle variables for the multisoliton solutions is the following.

Let K be a Hamiltonian vector field with respect to the implectic Θ.

Step 1: Find the first nontrivial mastersymmetry $\tau^{(1)}$.

Step 2: Take $\tilde{\Theta}$ equal to the Lie derivative of Θ into the direction of $\tau^{(1)}$ in order to obtain the recursion operator $\Phi = \tilde{\Theta}\Theta^{-1}$.

Step 3: Check whether $\tilde{\Theta}$ is implectic (this implies that Φ is hereditary).

Step 4: If yes, then construct the angle variables using (2.11), where $\tau^{(j)} = \Phi^{j-1}\tau^{(1)}$.

The crucial step of course is Step 1. For this step a computer algebra algorithm has been implemented. Some of the principles of this computer implementation will be briefly described in the next section; a detailed report of the method will appear elsewhere.[15,18] In addition, computer programs for Steps 2–4 also have been developed in Paderborn.[19] It should be observed that for carrying out the procedure which has been described above, no advance knowledge of the special structure of the system or of its Lax representation is necessary.

III. COMPUTER ALGEBRA ASPECTS

About the algorithmic aspects and details of the implementation a detailed report will appear elsewhere. Here a few remarks must suffice (see also Refs. 20, 15, 18). A crucial role is played by a "highest-order projection" for vector fields. Recall that the vector fields under consideration are polynomial such that if one evaluates the vectorfield at the place n of the lattice also field variables at other places do enter because there is some interaction between neighboring points. Projecting a vectorfield G onto those terms where the interaction reaches farthest (highest distances with respect to lattice points) and then taking the highest polynomial degree of these projected terms constitutes the highest order projection. The result of this projection is denoted by $ho(G)$ and we are able to define a suitable degree function yielding these highest order terms. Now an important role in the program is played by an approximate solution of the division problem in the Lie algebra of vector fields. By approximate we mean that for given vector fields G and R we are able to find a vectorfield X such that

$$ho([ho(G),X]) = ho(R). \tag{3.1}$$

This routine is called CS(G,R) ("commutator solution") and is the heart of the whole matter. The reason why this is possible at all lies in the fact that restricting the attention only to terms less that a fixed degree more or less simulates the situation of a finite-dimensional Lie algebra.

Having this procedure, it is rather simple to find symmetries and mastersymmetries automatically. We start with a symmetry S for the vector field G, i.e., $[S,G] = 0$, where S and G are both assumed to be translation invariant, i.e., they are invariant w.r.t. the replacement $n \to n + N$, where n is the

lattice variable and N some fixed shift. We assume that the mastersymmetry is of the form

$$\tau = nS + Z, \quad \text{where } Z \text{ is translation invariant.} \tag{3.2}$$

Why such an ansatz for τ leads to success in most examples will be explained elsewhere. Since τ is assumed to be a mastersymmetry, commutation with G must produce another symmetry,

$$\mathrm{SYM}(G,S) := [\tau, G]. \tag{3.3}$$

Based on the observation that the highest order term of τ can be assumed to be given by the highest order term of nS, we know that the highest order term of $[\tau,G]$ is given by the highest order term of $[nS,G]$ and we can use the following algorithm of successive approximation to find a new symmetry $\mathrm{SYM}(G,S)$.

Procedure: $\mathrm{SYM}(G,S)$:

Step 1: Put $G^{(1)} := [nS,G]$.

Step 2: Put $R := [G^{(1)},G]$. If $R = 0$ then return $(G^{(1)})$. Otherwise go to Step 3.

Step 3: Determine $X = \mathrm{CS}(G,R)$, where CS ( ) is applied by restricting the considerations only to terms of degree less than the degree of $G^{(1)}$. If there is no solution then return. ("There is no symmetry of this form.") Otherwise go to Step 4.

Step 4: Put $G^{(1)} := G^{(1)} + X$ and go to Step 2.

Obviously, in Step 1 the quantity $G^{(1)}$ is computed correctly in its highest order and each run computes $G^{(1)}$ correctly up to one order less. Hence the algorithm has to stop either after a number of runs given by the degree of $G^{(1)}$ thus giving the correct $G^{(1)}$, or it stops before by telling us that for the given S there is no mastersymmetry of the form (3.2). Obviously this algorithm has to terminate since all descending chains (with respect to degree) are finite.

Of course, this algorithm is based on a symmetry S which has to be known already. But observe that one can always use $S = G$. In order to determine the mastersymmetry itself we need the following procedure.

Procedure: $\mathrm{GHO}(G,R,E)$: The procedure GHO ("given highest order") determines those X with

$$ho(X) = ho(E) \tag{3.4}$$

such that

$$[X,G] = R. \tag{3.5}$$

Step 1: Put $X := E$.

Step 2: If $[X,G] = R$ then return (X). Otherwise go to Step 3.

Step 3: Put $X := X + \mathrm{CS}(G,[X,G] - R)$, where CS ( ) is applied by restricting the considerations only to terms of degree less than the degree of X. If there is no solution then return ("There is no solution.") Otherwise go to Step 2.

Of course, once we have implemented this algorithm there is no need for SYM anymore because

$$\mathrm{SYM}(G,S) = \mathrm{GHO}(G,0,[nS,G]). \tag{3.6}$$

Now, the determination of the first nontrival mastersymmetry (if existent) is given by

$$\mathrm{MAS}(G,G) = \mathrm{GHO}(G, \mathrm{SYM}(G,G), nG). \tag{3.7}$$

This means that first we use SYM to determine one nontrivial symmetry and then we determine, via GHO, the mastersymmetry which commutes the G we started with into this symmetry.

Of course, if this is successful then by further commutation with τ we can compute as many higher order symmetries as we like, and these higher order symmetries then can be used to compute further mastersymmetries. Another, more tedious, way to compute symmetries would be successive use of SYM, i.e.,

$$\begin{aligned} G^{(1)} &= \mathrm{SYM}(G,G), \\ G^{(2)} &= \mathrm{SYM}(G,G^{(1)}), \\ G^{(n+1)} &= \mathrm{SYM}(G,G^{(n)}). \end{aligned} \tag{3.8}$$

The program package is implemented in MAPLE,[21] a formula manipulation system developed by the University of Waterloo. The choice for a formula manipulation system was mainly based on our desire for rapid prototyping and on the fact that for these systems many sophisticated algorithms are available.

In the next section we describe what the procedures yield in the case of the relativistic Toda lattice. The remarkable fact about this algorithm is that the only knowledge needed for the determination of the action-angle variables is the equation itself.

IV. RESULTS: MASTERSYMMETRIES AND MULTI-HAMILTONIAN STRUCTURE OF THE RUIJSENAARS–TODA SYSTEM

Throughout this section we will use the form of the RT system in the variables $\{u = (d_n, c_n)\}_{n\in\mathbb{Z}}$ as given by (1.7). We shall proceed in a heuristic way, making use of the principle of mastersymmetries.

The right-hand side of (1.9a) shall be denoted by $K^{(1)}u$. A first (trivial) mastersymmetry for the RT system is found easily from the observation that it quite obviously is invariant with respect to the scaling transformation $u(t) \to e^{\varepsilon} u(e^{\varepsilon} t)$ generated by the vector field $tK^{(1)}(u) + u$. As the time-depending symmetries are in one-to-one correspondence with their time-independent part[9,11] we only take into account this last part and denote it by $\tau^{(0)}$. One easily checks $[\tau^{(0)}, K^{(1)}] = K^{(1)}$. The Lie derivative of the Hamiltonian operator $\Theta^{(0)}$ given in (1.9) into the direction of $\tau^{(0)}$ vanishes, i.e., not only $K^{(1)}$ but also $\tau^{(0)}$ is Hamiltonian. Indeed, one finds

$$\tau^{(0)}(u) = u = \Theta^{(0)} \nabla \sum_{n\in\mathbb{Z}} n \ln\left(\frac{d_n}{c_n}\right). \tag{4.1}$$

This Hamiltonian generates a conserved quantity with linear time dependence and hence can be regarded as a first angle variable for the RT system.

The above scaling field is a trivial mastersymmetry, in the sense that it does not generate new invariants. In order to obtain higher invariants we have to look for a first nontrivial mastersymmetry $\tau^{(1)}$, say. Using the ansatz

$$\tau_n^{(1)} = -nK_n^{(1)} + \text{translation invariant terms},$$

we exploit the computer-algebra algorithms described in the previous section. The computer finds the vector field

$$\tau_n^{(1)} = -nK_n^{(1)} + \begin{bmatrix} d_n(d_n + 2c_{n-1}) \\ c_n(2d_n + d_{n+1} + c_n + c_{n+1} + 2c_{n-1}) \end{bmatrix} \tag{4.2}$$

having the property that $[\tau^{(1)},K^{(1)}]$ commutes with $K^{(1)}$. Hence the vector field $K^{(2)}:=[\tau^{(1)},K^{(1)}]$ is a higher symmetry of the RT system. We compute the Lie derivative of the Hamiltonian operator $\Theta^{(0)}$ and find

$$\Theta^{(1)}:=-L_{\tau^{(1)}}\Theta^{(0)}=\begin{bmatrix} \Theta_{11}^{(1)} & \Theta_{12}^{(1)} \\ \Theta_{21}^{(1)} & \Theta_{22}^{(1)} \end{bmatrix}, \tag{4.3}$$

where the matrix elements are

$$\Theta_{11}^{(1)} = d(T^- c - cT^+)d;$$
$$\Theta_{12}^{(1)} = d(-d - c - cT^+ + T^- c + dT^- + T^- cT^-)c;$$
$$\Theta_{21}^{(1)} = c(d + c + T^- c - cT^+ - T^+ d - T^+ cT^+)d;$$
$$\Theta_{22}^{(1)} = c(-T^+ c + cT^- - cT^+ + T^- c - 2T^+ d + 2dT^- - T^+ cT^+ + T^- cT^-)c.$$

Again using computer algebra it is checked that $\Theta^{(1)}$ is an implectic operator and gives rise to a further Hamiltonian of the RT system,

$$K^{(1)} = \Theta^{(1)}\nabla \sum_{n\in\mathbb{Z}} \ln(d_n). \tag{4.4}$$

As $\Theta^{(1)}$ is obtained from $\Theta^{(0)}$ via a Lie derivative it automatically is compatible with $\Theta^{(0)}$ in the sense that the operator $\Theta^{(1)}+\Theta^{(0)}$ is again implectic. Hence inverting $\Theta^{(0)}$,

$$(\Theta^{(0)})^{-1} = \begin{bmatrix} \frac{1}{d}\left(\frac{1}{T^+ - 1} + \frac{1}{1 - T^-}\right)\frac{1}{d}; & -\frac{1}{d}\frac{1}{T^+ - 1}\frac{1}{c} \\ -\frac{1}{c}\frac{1}{1 - T^-}\frac{1}{d}; & 0 \end{bmatrix}, \tag{4.5}$$

we find a hereditary recursion operator,

$$\Phi := \Theta^{(1)}(\Theta^{(0)})^{-1} = \begin{bmatrix} d; & d + d(1 - T^-)c\frac{1}{T^+ - 1}\frac{1}{c} \\ c(1 + T^+); & c(1 + T^+) + c(T^+ - 1)d\frac{1}{T^+ - 1}\frac{1}{c} + c(T^+ - T^-)c\frac{1}{T^+ - 1}\frac{1}{c} \end{bmatrix}, \tag{4.6}$$

for the RT system. Here the inverses of the difference operators

$$\frac{1}{1 - T^-}: (u_n) \to \left(\sum_{k=-\infty}^{n} u_k\right),$$
$$\frac{1}{T^+ - 1} = \frac{1}{1 - T^-} - 1: u_n \to \sum_{k=-\infty}^{n-1} u_k \tag{4.7}$$

are nonlocal operations.

It turns out (as expected) that (4.6) is just the recursion operator found in Refs. 5,6 by completely different methods (namely starting from the Lax representation).

Before further investigation of the hierarchy of vector fields generated by the recursion operator let us have a closer look at the two Hamiltonian operators $\Theta^{(0)}$ and $\Theta^{(1)}$. For both operators we found two Casimir functions, i.e., functions with gradients in the kernels of these operators. For $\Theta^{(0)}$ the second Hamiltonian $f_0:=\Sigma \ln(d_n)$ is such a Casimir function and the function $C:=\Sigma \ln(d_n^2/c_n)$ is a Casimir for both operators. In addition to this we found

$$f^{(-1)}:=-\sum_{k\in\mathbb{Z}} \frac{d_{n+1}+c_n}{d_n d_{n+1}} \tag{4.8}$$

as a further Casimir of $\Theta^{(1)}$. As such functions are conservation laws for all Hamiltonian equations with respect to the corresponding Hamiltonian operators, we thus have found two further conserved quantities C and $f^{(-1)}$ for the RT system. The implectic operators map gradients of conservation laws to generators of symmetries, hence we have a further symmetry,

$$K^{(-1)}:=\Theta^{(0)}\nabla f^{(-1)}, \quad K_n^{(-1)} = \begin{bmatrix} \frac{c_n}{d_{n+1}} - \frac{c_{n-1}}{d_{n-1}} \\ \frac{c_n}{d_n} - \frac{c_n}{d_{n+1}} \end{bmatrix}, \tag{4.9}$$

for the RT system. An additional mastersymmetry $\tau^{(-1)}$ corresponding to (4.9) is found as

$$\tau_n^{(-1)} = nK_n^{(-1)} + \begin{bmatrix} 1 + \frac{c_n}{d_{n+1}} + \frac{c_{n-1}}{d_{n-1}} \\ -\frac{c_n}{d_{n+1}} \end{bmatrix}. \tag{4.10}$$

Again, computing the Lie derivatives of the Hamiltonian operators into the direction of (4.10), one finds $L_{\tau^{(-1)}}\Theta^{(1)} = 2\Theta^{(0)}$ and

$$\Theta^{(-1)}:=L_{\tau^{(-1)}}\Theta^{(0)} = \begin{bmatrix} cT^+ - T^- c; & -(1 - T^-)c \\ -c(T^+ - 1); & 0 \end{bmatrix}, \tag{4.11}$$

which turns out to be Hamiltonian and hence is compatible with $\Theta^{(0)}$. The RT system is thus endowed with three *local* Hamiltonian structures, defined by $\Theta^{(-1)}$, $\Theta^{(0)}$, and $\Theta^{(1)}$. We recall that an analogous property holds for the classical Toda lattice, as was shown by Kupershmidt.[22]

One checks $\Phi\Theta^{(-1)} = \Theta^{(0)}$, i.e., the operator $\Theta^{(-1)}(\Theta^{(0)})^{-1}$ yields the explicit inverse

$$\Phi^{-1} = \Theta^{(-1)}(\Theta^{(0)})^{-1} = \begin{bmatrix} (cT^+ - T^-c)\frac{1}{d}\left(\frac{1}{T^+-1} + \frac{1}{1-T^-}\right)\frac{1}{d} + \frac{1}{d}; & (-cT^+ + T^-c)\frac{1}{d}\frac{1}{T^+-1}\frac{1}{c} \\ c(1-T^+)\frac{1}{d}\left(\frac{1}{T^+-1} + \frac{1}{1-T^-}\right)\frac{1}{d}; & c(T^+-1)\frac{1}{d}\frac{1}{T^+-1}\frac{1}{c} \end{bmatrix} \tag{4.12}$$

of the recursion operator (4.6). Of course, the inverses of the difference operators $T^+ - 1$ and $1 - T^-$ are defined only up to "integration constants." Via these nonlocal operations we have "inverted" the above operators although their kernels are not empty, e.g., for Φ and Φ^{-1} one finds $\Phi K^{(-1)} = 0 = \Phi^{-1}K^{(1)}$. By formally putting

$$\begin{aligned} &[1/(1-T^-)](1) = (n), \\ &[1/(T^+-1)](1) = (n) - (1), \end{aligned} \tag{4.13}$$

where (1) is the constant sequence (...,1,1,1,...) and (n) is (..., − 2, − 1,0,1,2,...), one verifies $\tau^{(1)} = \Phi\tau^{(0)}$ and $\tau^{(-1)} = \Phi^{-1}\tau^{(0)}$.

Let us now finish the derivation of the RT hierarchy and define the higher invariants in a systematic way. We start with the first symmetries $K^{(1)}$ (i.e., the RT system itself) and $K^{(-1)}$. Furthermore we consider the Hamiltonian operator $\Theta^{(0)}$, the recursion operator Φ, its inverse Φ^{-1}, and the scaling field $\tau^{(0)}$. Let these quantities be given as above. Let the first Hamiltonian function be given by $f_0 = \Sigma_{n\in\mathbb{Z}} \ln(d_n)$. Define

$$\begin{aligned} &K^{(i)} := \Phi^{i-1}K^{(1)}; \quad i = 1,2,3,..., \\ &K^{(i)} := \Phi^{i+1}K^{(-1)}; \quad i = -1,-2,-3,..., \\ &f^{(i)} := (1/i)\langle \nabla f^{(i-1)}, \tau^{(1)}\rangle; \quad i = 1,2,3,..., \\ &f^{(i)} := (1/i)\langle \nabla f^{(i+1)}, \tau^{(-1)}\rangle; \quad i = -1,-2,-3,..., \\ &\tau^{(i)} := \Phi^i\tau^{(0)}; \quad \Theta^{(i)} = \Phi^i\Theta^{(0)}; \quad i\in\mathbb{Z}. \end{aligned} \tag{4.14}$$

Furthermore we define the auxiliary vector field $K^{(0)} := 0$.

Using the hereditary property of Φ, i.e., the compatibility structure of the implectic operators $\Theta^{(0)}$ and $\Theta^{(1)}$ (or $\Theta^{(-1)}$ and $\Theta^{(0)}$) and the scaling properties of $K^{(1)}$, $K^{(-1)}$, $\Theta^{(0)}$, $\Theta^{(1)}$, and $f^{(0)}$, one obtains the following.

Results:

$$\begin{aligned} &K^{(i+j)} = \Theta^{(i)}\nabla f^{(j)}; \\ &[K^{(i)}, K^{(j)}] = 0; \\ &[\tau^{(i)}, K^{(j)}] = jK^{(i+j)}; \\ &[\tau^{(i)}, \tau^{(j)}] = (j-i)\tau^{(i+j)}; \\ &L_{\tau^{(i)}}\Theta^{(j)} = (j-i)\Theta^{(i+j)}; \\ &\langle \nabla f^{(j)}, \tau^{(i)}\rangle = (j+i)f^{(i+j)} \\ &\text{(modulo constant functions).} \end{aligned} \tag{4.15}$$

Here all $i, j\in\mathbb{Z}$, i.e., positive, negative, or mixed combinations of i and j are admitted. The way to prove these statements is by induction, using the derivation property of the Lie derivatives and the hereditary property of the recursion operator Φ (see Refs. 10 and 12). Obviously the above Lie algebra spanned by the vector fields $K^{(i)}$ and $\tau^{(i)}$ can also be realized in terms of the following operator algebra:

$$\begin{aligned} &\tau^{(i)} \leftrightarrow z^{i+1}(d/dz), \\ &K^{(i)} \leftrightarrow z^i. \end{aligned} \tag{4.16}$$

It is worthwhile to notice that in this example both the kernels of Φ and its inverse are given explicitly in terms of the field variables by the vector fields $K^{(-1)}$ and $K^{(1)}$, i.e., $\Phi K^{(-1)} = K^{(0)} = 0 = \Phi^{-1}K^{(1)}$. This situation is quite exceptional in the context of completely integrable bi-Hamiltonian systems.

It is interesting to ask how the equations of the "negative" hierarchy $K^{(-1)}, K^{(-2)}, ...$, can be interpreted in terms of the original "physical" variables (q_n, Θ_n). Using the transformation (1.7) one obtains the form of the dynamical system $u_t = K^{(-1)}(u)$ in terms of these canonical variables. It turns out that this new equation is related to the original RT (1.2) by the simple transformation

$$q_n \to q_n, \quad \Theta_n \to -\Theta_n. \tag{4.17}$$

Rewriting this transformation in terms of the (d,c)-coordinates, we obtain

$$d_n \to \frac{1}{c_n}\frac{1 + c_{n-1}/d_{n-1}}{1 + c_n/d_n}, \quad c_n \to \frac{1}{d_n}\frac{1 + c_{n-1}/d_{n-1}}{1 + c_n/d_n}, \tag{4.18}$$

i.e., a "Bäcklund" transformation in the sense that it does not map each lattice point $u(n)$ individually but involves interaction terms of shifted lattice points. The transformation obviously is an involution, i.e., it is its own inverse, mapping $K^{(-1)} \leftrightarrow K^{(1)}$. Checking the transformations $\tau^{(0)} \leftrightarrow -\tau^{(0)}$, $\Theta^{(0)} \leftrightarrow -\Theta^{(0)}$, and $\Theta^{(1)} \leftrightarrow -\Theta^{(-1)}$, one concludes that (4.18) maps

$$\begin{aligned} &K^{(i)} \leftrightarrow K^{(-i)}, \quad \tau^{(i)} \leftrightarrow -\tau^{(-i)}, \\ &\Theta^{(i)} \leftrightarrow -\Theta^{(-i)}, \quad f^{(i)} \leftrightarrow -f^{(-i)} \end{aligned} \tag{4.19}$$

for all indices $i\in\mathbb{Z}$. Hence the negative part of the RT hierarchy may be regarded as the transformed of the positive part. The above transformation is not an "auto-Bäcklund transformation" but a "Bäcklund" ("Miura") transformation, mapping the RT system and its "higher" symmetries to a different form. Nevertheless, and this is an unusual and remarkable phenomenon found for this example, the "Miura"

transformed equations commute with the original hierarcy and hence constitute a further (the negative) part of the whole picture. We remark that the commutation property $[K^{(-1)},K^{(1)}]=0$ is not surprising once its meaning in terms of the original variables has been elucidated; it is a simple consequence of the Poincaré invariance of the RT system (see Refs. 1, 4) that the discrete transformation $\Theta\leftrightarrow -\Theta$ yields a symmetry for the RT equation.

We finally remark that the Hamiltonian operator $\Theta^{(-1)}$ (linear in the fields) gives rise to the Lie–Poisson structure on the dual of some Lie algebra. Indeed, we have been able to identify this Lie algebra and establish an approach to the RT via so-called "R matrices" [see, e.g., Ref. 23 for the theoretical background and Ref. 23(a) for the application to the classical Toda lattice]. The application of this approach to the RT shall be published elsewhere.[24] Following the ideas in Ref. 23, the technique of translated Casimir functions thus leads to an extension of the RT system (in analogy to the classical Toda lattice).

[1]S. N. M. Ruijsenaars, "Relativistic Calogero Moser systems and solitons," in *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations*, edited by M. Ablowitz, B. Fuchssteiner, and M. Kruskal (World Scientific, Singapore, 1987), pp. 182–190.

[2]S. N. M. Ruijsenaars and H. Schneider, "A new class of integrable systems and its relation to solitons," Ann. Phys., **170**, 370 (1986).

[3]H. Schneider, "Integrable relativistic N-particle systems in an external potential," Physica D **26**, 203 (1987).

[4]S. N. M. Ruijsenaars, "Relativistic Toda systems," preprint, 1986.

[5]M. Bruschi and O. Ragnisco, "Recursion operator and Bäcklund transformations for the Ruijsenaars toda lattice," Phys. Lett. A, **129**, 21 (1988).

[6]M. Bruschi and O. Ragnisco, "Direct and Inverse Spectral problem for the infinite Ruijsenaars–Toda lattice," Inverse Problems **5**, 389 (1989).

[7]M. Bruschi and O. Ragnisco, "Lax representations and complete integrability for the periodic relativistic Toda lattice," Phys. Lett. A **134**, 365 (1989).

[8]B. Fuchssteiner and A. S. Fokas, "Symplectic structures, their Bäcklund transformations and hereditary symmetries," Physica D **4**, 47 (1981).

[9]W. Oevel, "A geometrical approach to integrable systems admitting time dependent invariants," in *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations*, edited by M. Ablowitz, B. Fuchssteiner, and M. Kruskal (World Scientific, Singapore, 1987), pp. 108–124.

[10]H. M. M. ten Eikelder, "Symmetries for dynamical and Hamiltonian systems," CWI Tract, 17, Centrum voor Wiskunde en Informatica, Amsterdam, 1985.

[11]B. Fuchssteiner, "Mastersymmetries, higher-order time-dependent symmetries and conserved densities of nonlinear evolution equations," Prog. Theor. Phys. **70**, 1508 (1983).

[12]W. Oevel, "Mastersymmetries: weak action angle structure for Hamiltonian and non-Hamiltonian dynamical systems," preprint, Paderborn, 1986.

[13]P. M. Santini and A. S. Fokas, "Recursion operators and bi-Hamiltonian structures of 2 + 1 dimensional systems," in *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations* edited by M. Ablowitz, B. Fuchssteiner, and M. Kruskal (World Scientific, Singapore, 1987), pp. 1–19.

[14]B. Fuchssteiner and G. Oevel, "Tangent space and angle variables for the multisoliton solutions," preprint, Paderborn, 1988.

[15]W. Oevel, H. Zhang, and B. Fuchssteiner, "Mastersymmetries and multi-Hamiltonian formulations for some integrable lattice systems," Prog. Theor. Phys. **81**, 294 (1989).

[16]B. Fuchssteiner, "Application of hereditary symmetries to non-linear evolution equations," Nonlinear Analysis TMA **3**, 849 (1979).

[17]B. Fuchssteiner and R. N. Aiyer, "Multisolitons, or the discrete eigenfunctions of the recursion operator of non-linear evolution equations. II. Background," J. Phys. A **20**, 375 (1987).

[18]H. Zhang and B. Fuchssteiner, "Computer algorithms for the treatment of completely integrable lattice systems," preprint, Paderborn, 1988.

[19]B. Fuchssteiner, W. Oevel, and W. Wiwianka, "Computer-algebra methods for investigation of hereditary operators of higher order soliton equations," Comput. Phys. Comm. **44**, 47 (1987).

[20]B. Fuchssteiner and U. Falck, "Computer algorithms for the detection of completely integrable quantum spin chains," in *Symmetries and Nonlinear Phenomena*, edited by D. Levi and P. Winternitz, Proceedings of the International School on Applied Mathematics, Columbia, 1988 (World Scientific, Singapore, 1988).

[21]K. O. Geddes, G. H. Gonnet, and B. W. Char, MAPLE User's Manual, Fifth ed. (WATCOM, 1988).

[22]B. A. Kupershmidt, "Discrete Lax Equation and Differential-Difference Calculus," Astérisque (Societé Mathematique de France, Paris, 1985), vol. 123.

[23]W. Oevel, (a) "R-matrices and related involution theorems," in *Proceedings of the Workshop on "Finite dimensional integrable nonlinear dynamical systems"* (Johannesburg, 1988), edited by P. G. L. Leach and W. H. Steeb (World Scientific, Singapore, 1988), pp. 60–73. (b) "R-structures, Yang–Baxter equations, and related involution theorems," J. Math. Phys., **30**, 1140 (1989).

[24]W. Oevel and O. Ragnisco, "R-matrices and higher poisson structures for integrable systems," to be published in Physica A.

The uniqueness of the energy momentum tensor in non-Abelian gauge field theories

R. J. Noriega
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

C. G. Schifini
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, and CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

(Received 28 June 1988; accepted for publication 24 May 1989)

The uniqueness of the energy momentum tensor in non-Abelian gauge field theories is established under minimal hypothesis.

I. INTRODUCTION

In the general theory of relativity, the interaction of the gravitational field (characterized by a metric tensor g_{ij}) and a source-free gauge field (characterized by a curvature form F^{α}_{ij}; see Ref. 1 for definitions and notations) is assumed to be governed by the Einstein–Yang–Mills field equations

$$R^{ij} - \tfrac{1}{2}g^{ij}R = B_{\alpha\beta}(F^{\alpha i}{}_{k}F^{\beta jk} - \tfrac{1}{4}g^{ij}F^{\alpha hk}F^{\beta}_{hk}), \tag{1}$$

$$F^{\alpha ij}{}_{\|j} = 0, \tag{2}$$

where $B_{\alpha\beta}$ are the coefficients of a bilinear symmetric form in LG, the Lie algebra of the Lie group G, which are Ad G invariant, i.e., $B_{\alpha\beta} = \mathrm{Ad}^{\gamma}_{\alpha}(a)\mathrm{Ad}^{\eta}_{\beta}(a)B_{\gamma\eta}$ for all a in G. Besides, the covariant gauge derivative of F^{α}_{ij} is defined as

$$F^{\alpha}_{ij\|h} = F^{\alpha}_{ij,h} - \Gamma^{s}_{ih}F^{\alpha}_{sj} - \Gamma^{s}_{jh}F^{\alpha}_{is} + C^{\alpha}_{\beta\gamma}A^{\beta}_{h}F^{\gamma}_{ij}, \tag{3}$$

where $C^{\alpha}_{\beta\gamma}$ are the structure constants of the Lie group and A^{β}_{h} are the gauge potentials (see Ref. 1 or Ref. 2) related to the curvature form by

$$F^{\alpha}_{ij} = A^{\alpha}_{j,i} - A^{\alpha}_{i,j} + C^{\alpha}_{\beta\gamma}A^{\beta}_{i}A^{\gamma}_{j}. \tag{4}$$

It is easy to see that with these definitions, the following identity holds:

$$F^{\alpha}_{ij\|h} + F^{\alpha}_{jh\|i} + F^{\alpha}_{hi\|j} = 0. \tag{5}$$

Since the Einstein tensor given by the left-hand side of (1) is divergence-free, the same must be true for T^{ij}_{0}, the right-hand side of (1). This is the case because of the identity

$$T^{ij}_{0\|j} = \beta_{\alpha\beta}F^{\alpha ih}F^{\beta}{}_{h}{}^{j}{}_{\|j}, \tag{6}$$

and Eq. (2). For any T^{ij} in the right-hand side it must be true that $T^{ij}{}_{\|j} = 0$, at least when (2) holds. In other words, it must be true that

$$F^{\alpha ij}{}_{\|j} = 0 \Rightarrow T^{ij}{}_{\|j} = 0. \tag{7}$$

The uniqueness of the energy momentum tensor was established recently[1] under the restrictive hypothesis $T^{ij}{}_{\|j} = C_{\alpha\beta}H^{\beta jr}F^{\alpha i}{}_{r\|i}$. Clearly (7) is weaker and it is mandatory because of (1) and (2). In this paper we will prove that $T_0{}^{ij}$ is essentially the only solution to the following problem: to find all gauge invariant symmetric tensors $T^{ij} = T^{ij}(g_{hk};F^{\alpha}_{hk})$ such that (7) holds. Our result generalizes Ref. 3.

We want to point out that, due to the condition (7), one cannot generate energy momentum tensors by adding terms to the action.

II. CONSEQUENCES OF THE IMPLICATION (7)

We will work in a coordinate system for which $(g_{ij}) = \mathrm{diag}(-1,1,1,1)$ and $g_{ij,h} = 0$ (which implies $\Gamma^{i}_{jk} = 0$). Then (2) reads

$$F^{\beta}_{1k\|1} = F^{\beta}_{2k\|2} + F^{\beta}_{3k\|3} + F^{\beta}_{4k\|4}. \tag{8}$$

It is easy to see that

$$T^{ij}{}_{\|j} = \frac{\partial T^{ij}}{\partial F^{\beta}_{hk}}F^{\beta}_{hk\|j} = T^{ijhk}_{\beta}F^{\beta}_{hk\|j}, \tag{9}$$

because of the gauge invariance of T^{ij} and its tensorial character. Then $T^{ij}{}_{\|j} = 0$ written out in full in the above mentioned coordinate system is

$$\begin{aligned}
&(T^{i223}_{\beta} + T^{il13}_{\beta})F^{\beta}_{23\|2} + (T^{i224}_{\beta} + T^{il14}_{\beta})F^{\beta}_{24\|2} + (T^{i323}_{\beta} - T^{il12}_{\beta})F^{\beta}_{23\|3} + (T^{i334}_{\beta} + T^{il14}_{\beta})F^{\beta}_{34\|3} + (T^{i424}_{\beta} - T^{il12}_{\beta})F^{\beta}_{24\|4}\\
&+ (T^{i434}_{\beta} - T^{il13}_{\beta})F^{\beta}_{34\|4} + (T^{il23}_{\beta} + T^{i213}_{\beta})F^{\beta}_{23\|1} + (T^{i312}_{\beta} + T^{i213}_{\beta})F^{\beta}_{12\|3} + (T^{i313}_{\beta} - T^{i212}_{\beta})F^{\beta}_{13\|3}\\
&+ (T^{il24}_{\beta} + T^{i214}_{\beta})F^{\beta}_{24\|1} + (T^{i412}_{\beta} + T^{i214}_{\beta})F^{\beta}_{12\|4} + (T^{i414}_{\beta} - T^{i212}_{\beta})F^{\beta}_{14\|4} + (T^{il34}_{\beta} + T^{i314}_{\beta})F^{\beta}_{34\|1}\\
&+ (T^{i413}_{\beta} + T^{i314}_{\beta})F^{\beta}_{13\|4} + (T^{i234}_{\beta} + T^{i324}_{\beta})F^{\beta}_{34\|2} + (T^{i423}_{\beta} + T^{i324}_{\beta})F^{\beta}_{23\|4} = 0.
\end{aligned} \tag{10}$$

 0022-2488/89/112671-03$02.50

Let us choose, for arbitrary but fixed g_{hk}, F^{α}_{hk}, the derivatives $F^{\beta}_{ij,h}$ such that (8) holds.

Taking account of (8) and (5), it is clear that the $F^{\beta}_{ij\|h}$ appearing in (10) are arbitrary and independent. Then we deduce

$$T^{i313}_{\beta} - T^{i212}_{\beta} = T^{i414}_{\beta} - T^{i212}_{\beta} = T^{i223}_{\beta} + T^{i113}_{\beta} = T^{i224}_{\beta} + T^{i114}_{\beta} = T^{i323}_{\beta} - T^{i112}_{\beta} = T^{i334}_{\beta} + T^{i114}_{\beta}$$
$$= T^{i424}_{\beta} - T^{i112}_{\beta} = T^{i434}_{\beta} - T^{i113}_{\beta} = T^{i123}_{\beta} + T^{i213}_{\beta} = T^{i312}_{\beta} + T^{i213}_{\beta} = T^{i124}_{\beta} + T^{i214}_{\beta} = T^{i412}_{\beta} + T^{i214}_{\beta}$$
$$= T^{i134}_{\beta} + T^{i314}_{\beta} = T^{i413}_{\beta} + T^{i314}_{\beta} = T^{i234}_{\beta} + T^{i324}_{\beta} = T^{i423}_{\beta} + T^{i324}_{\beta} = 0. \tag{11}$$

Taking $i = 1,2,3,4$ in (11), a tedious but straightforward computation proves that

$$1212 = 1313 = 1414 = 2323 = 2424 = 3434 = 1234 = 1324 = 1423 = 2314 = 2413 = 3412 = 0, \tag{12}$$
$$1323 = 1424 = -2212 = -2313 = -2414 = 3312 = 4412 = 1112, \tag{13}$$
$$-1223 = 1434 = 2213 = -2312 = -3313 = -3414 = 4413 = 1113, \tag{14}$$
$$-1224 = -1334 = 2214 = -2412 = 3314 = -3413 = -4414 = 1114, \tag{15}$$
$$-1213 = 1312 = 2223 = -2434 = 3323 = 3424 = -4423 = 1123, \tag{16}$$
$$-1214 = 1412 = 2224 = 2334 = -3324 = 3423 = 4424 = 1124, \tag{17}$$
$$-1314 = 1413 = -2234 = 2324 = -2423 = 3334 = 4434 = 1134, \tag{18}$$

where, for the sake of simplicity, we have used the notation

$$T^{ijhk}_{\beta} = ijhk$$

for a fixed β.

III. THE UNIQUENESS OF THE ENERGY MOMENTUM TENSOR

Let us denote, for fixed α, β, and γ,

$$T^{ijhkrslm} = \frac{\partial^3 T^{ij}}{\partial F^{\alpha}_{hk}\, \partial F^{\beta}_{rs}\, \partial F^{\gamma}_{lm}}.$$

We will prove that all these derivatives are zero. From (12)–(18) it is clear that it is enough to consider the cases $ijhk = 1213, 1214, 1223, 1224, 1323, 1314$.

(i) The case ijhk=1213: It is clear that $T^{1213rslm} = 0$ except perhaps for $(r,s) \neq (1,2) \neq (l,m)$ and $(r,s) \neq (3,4) \neq (l,m)$. Since $|T^{ijhk}_{\beta}| = |T^{hkij}_{\beta}|$ for $i \neq j$ and $h \neq k$ as a consequence of (12)–(18), then in this case all the pairs commute, and so it is enough to consider $(r,s) \neq (1,3) \neq (l,m) \neq (r,s)$, which leave us with the following cases:

(a) $(r,s,l,m) = (1,4,2,3)$,

(b) $(r,s,l,m) = (1,4,2,4)$,

(c) $(r,s,l,m) = (2,3,2,4)$.

In case (a) using (16), (15), and $|T^{ijhk}_{\beta}|$ for $i \neq j$ and $h \neq k$, we have

$$|T^{12131423}| = |T^{33231423}| = |T^{33142323}| = |T^{24122323}|$$
$$= |T^{23232412}| = 0. \tag{19}$$

In case (b) we have

$$|T^{12131424}| = |T^{13121424}| = |T^{13241214}| = 0. \tag{20}$$

Finally, in case (c) it is

$$|T^{12132324}| = |T^{13122324}| = |T^{13241223}| = 0, \tag{21}$$

where we have also used the equality of the cross derivatives. We conclude that

$$T^{1213rslm} = 0 \quad \text{for all } r,s,l,m. \tag{22}$$

(ii) The case ijhk=1214: It is easy to see that $T^{1214rslm} = 0$ except perhaps for $(r,s) \neq (1,2) \neq (l,m) \neq (r,s)$, $(r,s) \neq (1,4) \neq (l,m) \neq (r,s)$, $(r,s) \neq (3,4) \neq (l,m) \neq (r,s)$, and $(r,s) \neq (2,3) \neq (l,m) \neq (r,s)$, which leaves us only with the case $T^{12141324}$, and this is zero because of (20). Then

$$T^{1213rslm} = 0 \quad \text{for all } r,s,l,m. \tag{23}$$

(iii) The case ijhk=1223: As in case (ii) $T^{1223rslm} = 0$ except perhaps for $(r,s) \neq (1,2)$, $(2,3)$, $(3,4)$, $(1,4) \neq (l,m)$, and $(r,s) \neq (l,m)$. This leaves us only with the case $T^{12231324}$ which is zero by (21). Then

$$T^{1223rslm} = 0 \quad \text{for all } r,s,l,m. \tag{24}$$

(iv) The case ijhk=1224: As before, it is enough to consider the cases $(r,s) \neq (1,2)$, $(2,4)$, $(3,4)$, $(1,3) \neq (l,m)$, and $(r,s) \neq (l,m)$. Then there is only the case $|T^{12241423}| = |T^{14231224}| = 0$. Then

$$T^{1224rslm} = 0 \quad \text{for all } r,s,l,m. \tag{25}$$

(v) The case ijhk=1323: It is enough to consider $(r,s) \neq (1,3)$, $(2,3)$, $(2,4)$, $(1,4) \neq (l,m)$, and $(r,s) \neq (l,m)$. Then $|T^{13231234}| = |T^{12341323}| = 0$. Thus

$$T^{1323rslm} = 0 \quad \text{for all } r,s,l,m. \tag{26}$$

(vi) The case ijhk=1314: It is enough to consider $(r,s) \neq (1,3)$, $(1,4)$, $(2,4)$, $(2,3)$, $\neq (l,m)$, and $(r,s) \neq (l,m)$. But then $|T^{13141234}| = |T^{12341314}| = 0$, and so

$$T^{1314rslm} = 0 \quad \text{for all } r,s,l,m. \tag{27}$$

From (22)–(27) we conclude that

$$T^{ijhkrslm} = 0 \quad \text{for all } i,j,h,k,r,s,l,m,$$

and so T^{ij} is a polynomial in F^{α}_{ij} of degree not greater than two.

Consequently

$$T^{ij} = A^{ijhkrs}_{\epsilon\gamma}(g_{lm}) F^{\epsilon}_{hk} F^{\gamma}_{rs} + B^{ijhk}_{\epsilon}(g_{lm}) F^{\epsilon}_{hk} + C^{ij}(g_{lm}). \tag{28}$$

The tensorial concomitants of g_{lm} were recently found[4,5] for any valence of the tensor. Taking account of the fact that we are dealing with all coodinate systems, and not merely with those belonging to an oriented atlas, then it follows that

$$T^{ij} = (d_{\alpha\beta} F^{\alpha hk} F^{\beta}_{hk} + \lambda) g^{ij} + \tfrac{1}{2} a_{\alpha\beta} (F^{\alpha i}{}_{t} F^{\beta jt} + F^{\alpha j}{}_{t} F^{\beta it}), \tag{29}$$

where $d_{\alpha\beta}$, λ, and $a_{\alpha\beta}$ are real numbers and $a_{\alpha\beta} = a_{\beta\alpha}$, $d_{\alpha\beta} = d_{\beta\alpha}$. Then

$$T^{ij} = (d_{\alpha\beta} F^{\alpha hk} F^{\beta}_{hk} + \lambda) g^{ij} + a_{\alpha\beta} F^{\alpha i}{}_{t} F^{\beta jt}. \tag{30}$$

Assuming $F^{\alpha ij}{}_{\|j} = 0$, it follows that $T^{ij}{}_{\|j} = 0$, and so, using the identity (5) to change indices, we have

$$[2 d_{\alpha\beta} g^{ij} F^{\alpha hk} + \tfrac{1}{2} a_{\alpha\beta} g^{ij} F^{\alpha hk}] F^{\beta}_{hk\|j} = 0. \tag{31}$$

It is easy to see that if S^{ijhk}_{β} is the term within brackets in (31) then, because of (12)–(18), we have

$$S^{ijhk}_{\beta} g_{ij} = 0. \tag{32}$$

From (32) and the definition of S^{ijhk}_{β} it follows that

$$2 \cdot (4 d_{\alpha\beta} + a_{\alpha\beta}) F^{\alpha hk} = 0. \tag{33}$$

Differentiating (33) with respect to F^{α}_{hk} we obtain

$$(4 d_{\alpha\beta} + a_{\alpha\beta})(g^{hr} g^{ks} - g^{kr} g^{hs}) = 0. \tag{34}$$

Multiplying (34) by $g_{hr} \cdot g_{ks}$ we deduce

$$d_{\alpha\beta} = -\tfrac{1}{4} a_{\alpha\beta}, \tag{35}$$

and so, replacing (33) in (30), it follows that

$$T^{ij} = a_{\alpha\beta} (F^{\alpha i}{}_{t} F^{\beta jt} - \tfrac{1}{4} g^{ij} F^{\alpha hk} F^{\beta}_{hk}) + \lambda g^{ij}. \tag{36}$$

It follows easily from the gauge invariance of T^{ij} that the $a_{\alpha\beta}$ are Ad G invariant.

In summary, we have proved the following.

Theorem: If $T^{ij} = T^{ij}(g_{hk}: F^{\alpha}_{hk})$ is a gauge invariant tensor whose divergence vanishes when the divergence of $F^{\alpha ij}$ is zero, and if $T^{ij} = T^{ji}$, then

$$T^{ij} = T_0{}^{ij} + \lambda g^{ij},$$

where T^{ij}_{0} is the usual energy momentum tensor.

[1] R. J. Noriega and C. G. Schifini, Int. J. Theor. Phys. **24**, 1181 (1985).
[2] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley–Interscience, New York, 1963), Vol. I.
[3] B. Kerrighan, J. Math. Phys. **23**, 1979 (1982).
[4] R. J. Noriega and C. G. Schifini, Gen. Relativ. Gravit. **18**, 983 (1986).
[5] D. Prélat, "Tensorial concomitants of a metric and a convector," Utilitas Math. (in press).

Dyons in affine field theories

A. C. Cadavid and R. J. Finkelstein
Department of Physics, University of California, Los Angeles, California 90024-1547

(Received 28 February 1989; accepted for publication 17 May 1989)

By postulating that the gauge vector field and the Higgs field both lie in an affine algebra rather than a Lie algebra, an affine field theory with an expanded soliton sector is obtained. An infinite family of exact, special solutions of the new equations corresponding to the original 't Hooft–Polyakov solitonic solutions, is now found. A perturbation method for an arbitrary solution of the new equation is also proposed.

I. INTRODUCTION

Since the work of 't Hooft and Polyakov, it has been known that non-Abelian field theories may have a soliton sector. In particular, magnetic monopoles[1,2] as well as dyons[3] may arise in a spontaneously broken SU(2) gauge theory. Although there is no present evidence for these particles one can speculate that they are simply too heavy to have been observed or that they are bound preons.[4] If one is interested in gravitationally coupled theories, however, and therefore in extremely high energies, perhaps one should think more about these objects.

The theoretically known Yang–Mills solitons are nodeless, but in principle there could be noding solutions like the excited states of an atom or nucleus. In fact, one should in general expect higher modes of excitation in any spatially extended structure.[5] It may be, however, that the ensemble of higher modes is realized physically only if one passes from the finite Lie algebra to an infinite algebra. Moreover, such a formal extension is natural since there exists an infinite-dimensional Lie algebra corresponding to every semisimple Lie algebra. This formal extension is also suggested by the idea of regarding the infinite algebra and the corresponding internal space as a surrogate for particle extension in space-time.

If one does pass from Yang–Mills theories to their affine extensions,[6] one finds that the soliton sector is indeed correspondingly expanded. In this paper these new solitons will be studied. In taking this step our basic assumption is that the Higgs field, as well as the vector field, lies in a loop algebra rather than in a Lie or in a Kac–Moody algebra. In an earlier note[7] we discussed the Higgs splitting of a Kac–Moody field and found that the mass of the vector particle was linear in the loop index n. That result provided a simple interpretation of n as a label of the heavy vectors, but our assumptions here will be slightly different: for simplicity we limit the discussion to the loop rather than the full Kac–Moody algebra.

II. FORMULATION

We assume that the vector and Higgs field lie in the loop algebra

$$W_\mu = \sum (W_\mu)_n{}^a T_n{}^a, \tag{2.1}$$

$$\Phi = \sum (\Phi)_n{}^a T_n{}^a, \tag{2.2}$$

where

$$(T_n{}^a, T_m{}^b) = f^{abc} T^c_{n+m} \tag{2.3}$$

and the structure constants are real. Then

$$(T_n{}^a)^+ = -T^a_{-n}. \tag{2.4}$$

We also take both fields to be anti-Hermitian:

$$W_\mu{}^+ = -W_\mu, \tag{2.5}$$

$$\Phi^+ = -\Phi. \tag{2.6}$$

Then

$$(W_\mu)_n{}^a = (W_\mu)^a_{-n}, \tag{2.7}$$

$$\Phi_n{}^a = \Phi^a_{-n}, \tag{2.8}$$

if we take the component fields to be real. We consider the action

$$S = \int d^4x\, L, \tag{2.9}$$

$$L = \langle -\tfrac{1}{4} G_{\mu\lambda} (G^{\mu\lambda})^+ + \tfrac{1}{2} (D_\mu \Phi)(D^\mu \Phi)^+ \rangle + \lambda V(\langle \Phi^+ \Phi \rangle), \tag{2.10}$$

where $\langle\ \rangle$ is the invariant scalar product for the algebra, $G_{\mu\lambda}$ is the usual field strength, and $D_\mu \Phi$ is a covariant derivative:

$$G_{\mu\lambda} = (\nabla_\mu, \nabla_\lambda), \tag{2.11}$$

$$D_\mu \Phi = (\nabla_\mu, \Phi), \tag{2.12}$$

$$\nabla_\mu = \partial_\mu + W_\mu. \tag{2.13}$$

Here W_μ and Φ are the loop vector and scalar fields.

The equations of motion are formally the same as for a Lie algebra since $\langle\ \rangle$ shares with the trace the properties

$$\langle AB \rangle = \langle BA \rangle, \tag{2.14}$$

$$\langle (A,B)C \rangle = \langle (B,C)A \rangle. \tag{2.15}$$

Also

$$\langle T_n{}^a T_m{}^b \rangle = -\delta^{ab} \delta(n+m). \tag{2.16}$$

The equations of motion are

$$(\nabla_\lambda, G^{\mu\lambda}) = (\Phi, (\nabla^\mu, \Phi)), \tag{2.17}$$

$$(\nabla^\mu, (\nabla_\mu, \Phi)) = \lambda \frac{\partial V}{\partial \Phi}. \tag{2.18}$$

In addition there is the Bianchi identity

$$(\nabla_\mu, \tilde{G}^{\mu\lambda}) = 0, \tag{2.19}$$

where $\tilde{G}^{\mu\lambda}$ is the dual field. The symmetric energy momentum tensor is

 0022-2488/89/112674-09$02.50

$$T^{\mu\lambda} = \langle G^{\mu\sigma}(G_\sigma{}^\lambda)^+ + (D^\mu\Phi)(D^\lambda\Phi)^+ + g^{\mu\lambda}L\rangle \tag{2.20}$$

and the energy density is

$$T^{00} = \tfrac{1}{2}\langle \mathscr{E}^i(\mathscr{E}^i)^+ + \mathscr{B}^i(\mathscr{B}^i)^+ (D^0\Phi)(D^0\Phi)^+ + (D^i\Phi)(D^i\Phi)^+\rangle + \lambda V(\langle\Phi\Phi^+\rangle), \tag{2.21}$$

where

$$G^{0i} = \mathscr{E}^i, \tag{2.22}$$

$$G^{ij} = \epsilon^{ijk}\mathscr{B}^k, \tag{2.23}$$

$$\lambda V > 0. \tag{2.24}$$

A field configuration for which $T^{00} = 0$ everywhere will be called a vacuum configuration. Then by (2.21) a vacuum configuration implies

$$G^{\mu\lambda} = 0, \tag{2.25}$$

$$D^\mu\Phi = 0, \tag{2.26}$$

$$V = 0. \tag{2.27}$$

A region of space-time will be described as a Higgs vacuum if (2.26) and (2.27) but not necessarily (2.25) are satisfied.[8] The condition of finite energy for the entire soliton field will enforce the Higgs vacuum asymptotically at large distances.

One may ask for classical solitons that solve Eqs. (2.17)–(2.19) subject to the boundary conditions (2.26) and (2.27). Solutions are known where $G^{\mu\lambda}$ and Φ are both isotriplets lying in SO(3). For definiteness we limit our work to the loop generalization of SO(3). Therefore we shall investigate the generalization of the SO(3) solitons to the corresponding loop algebra.

III. THE HIGGS VACUUM AND THE ELECTROMAGNETIC FIELD[8]

In the Higgs vacuum the SO(3) symmetry is broken down to SO(2) or equivalently to U(1). Then the photon, remaining massless, is separated from the massive vectors which acquire mass $ae\hbar$, where $a = |\boldsymbol{\phi}|$ in the Higgs vacuum and e is the gauge coupling constant. The Higgs particle acquires the mass $a(2\lambda)^{1/2}\hbar$. In the same region of space-time where the Higgs field points in a fixed direction $\boldsymbol{\phi}$, the electromagnetic potential may be identified with $\boldsymbol{\phi}\mathbf{W}^\mu/a$ and the electric charge operator with $e\boldsymbol{\phi}\mathbf{T}\hbar/a$, where T^a $(a = 1,2,3)$ are the SO(3) generators and $\boldsymbol{\phi}\mathbf{T}/a$ is the equivalent U(1) generator. In this region the electromagnetic field is well defined. However, the expression for it is not unique since the electromagnetic field is not well defined in the interior of the soliton, although all such expressions must agree in the Higgs vacuum.

The electromagnetic field tensor proposed by 't Hooft is[1]

$$F_{\mu\lambda} = \hat{\phi}^a(G^a_{\mu\lambda} - e^{-1}\epsilon^{abc}D_\mu\hat{\phi}_b D_\lambda\hat{\phi}_c), \tag{3.1}$$

where

$$\hat{\phi}^a = \phi^a/|\boldsymbol{\phi}|. \tag{3.2}$$

For SO(3) this tensor may be rewritten as

$$F_{\mu\lambda} = M_{\mu\lambda} + H_{\mu\lambda}, \tag{3.3}$$

where

$$M_{\mu\lambda} = \partial_\mu A_\lambda - \partial_\lambda A_\mu, \tag{3.4}$$

$$A_\lambda = \hat{\phi}^a W_\lambda{}^a, \tag{3.5}$$

and

$$H_{\mu\lambda} = e^{-1}\epsilon^{abc}\hat{\phi}^a\partial_\mu\hat{\phi}^b\partial_\lambda\hat{\phi}^c. \tag{3.6}$$

The magnetic current is

$$k_\mu = \partial^\lambda\tilde{F}_{\mu\lambda}, \tag{3.7}$$

where $\tilde{F}_{\mu\lambda}$ is the dual field. Then

$$k_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}\partial^\nu M^{\rho\sigma} - (1/2e)\epsilon_{\mu\nu\rho\sigma}\epsilon^{abc}\partial_\nu\hat{\phi}^a\partial_\rho\hat{\phi}^b\partial_\sigma\hat{\phi}^c. \tag{3.8}$$

The magnetic current may arise from the first term if A_λ is singular on Dirac strings or entirely from the second term if A_λ has no string singularities. These equivalent descriptions of magnetic charge are connected by a singular gauge transformation. If there are no string singularities, the magnetic charge is

$$g = \frac{1}{4\pi}\int k_0\, d^3x \tag{3.9}$$

$$= \frac{1}{8\pi e}\int \epsilon_{ijk}\epsilon_{abc}\partial_i\hat{\phi}^a\partial_j\hat{\phi}^b g_k\hat{\phi}^c d^3x \tag{3.10}$$

or

$$eg = \lim_{R\to\infty}\frac{1}{8\pi}\int_{S^{(2)}_R}\epsilon_{ijk}\epsilon_{abc}\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c(d^2\sigma)_i. \tag{3.11}$$

The integral in the preceding section is a topological invariant that can assume only integral values. The total magnetic charge is therefore

$$g = n/e, \tag{3.12}$$

where n is the number of nodes in the Higgs function.[9]

Equation (3.1) for the electromagnetic field is gauge invariant and reduces by (2.26) to the required form in the Higgs vacuum. As already remarked, however, there is no unique definition of $F_{\mu\lambda}$ outside the Higgs vacuum. Another proposal that agrees with (3.1) in the Higgs vacuum is[10]

$$F_{\mu\lambda} = \boldsymbol{\phi}\cdot\mathbf{G}_{\mu\lambda}/a \tag{3.13a}$$

or

$$F_{\mu\lambda} = \mathrm{Tr}(\Phi G_{\mu\lambda}/a). \tag{3.13b}$$

We shall generalize (3.13b) for the loop algebra to

$$F_{\mu\lambda} = \langle\Phi^+ G_{\mu\lambda}\rangle/a, \tag{3.14}$$

where

$$a^2 = \lim_{r\to\infty}\langle\Phi^+\Phi\rangle \tag{3.14a}$$

and $\langle\ \rangle$ is the invariant scalar product for the loop algebra. Here $F_{\mu\lambda}$ is the generalized electromagnetic field, but only $n = 0$ is the Maxwell field. The corresponding generalized magnetic current is

$$k^\mu = \partial_\lambda\langle\Phi^+\tilde{G}^{\mu\lambda}\rangle/a \tag{3.15a}$$

or

$$= \langle D_\lambda \Phi^+ \tilde{G}^{\mu\lambda} \rangle / a, \tag{3.15b}$$

by the Bianchi identities (2.19).

The conservation of k^μ follows from its definition (3.15a) as the divergence of an antisymmetric tensor.

The generalized electric and magnetic charges following from (3.14) are

$$Q = \lim_{r\to\infty} \frac{r^2}{a} \int d\Omega \langle \Phi^+ \mathscr{E}_r \rangle \tag{3.16a}$$

$$= \frac{1}{a} \int d^3x \langle D_i \Phi^+ \mathscr{E}_i \rangle, \tag{3.16b}$$

$$M = \lim_{r\to\infty} \frac{r^2}{a} \int d\Omega \langle \Phi^+ \mathscr{B}_r \rangle \tag{3.17a}$$

$$= \frac{1}{a} \int d^3x \langle D_i \Phi^+ \mathscr{B}_i \rangle, \tag{3.17b}$$

where $\mathscr{E}$ and $\mathscr{B}$ are given by (2.22) and (2.23), respectively.

The actual electric and magnetic charges are, of course, given by the $n = 0$ components.

IV. DYONS[3,11]

We consider those special solutions obeying the conditions

$$(W^0, \Phi) = 0, \tag{4.1}$$

$$\partial_0 W^\mu = \partial_0 \Phi = 0. \tag{4.2}$$

Then

$$D_0 W^0 = D_0 \Phi = 0. \tag{4.3}$$

With these assumptions it is shown in the Appendix that Eqs. (2.17)–(2.19) are equivalent to the set

$$\mathscr{E}_i = \sin\alpha \, D_i \Phi, \tag{4.4}$$

$$\mathscr{B}_i = \cos\alpha \, D_i \Phi, \tag{4.5}$$

in the limit $\lambda \to 0^+$. This reduction holds for any algebra including loop algebras.

The mass of the dyon is

$$m = \int T_{00} \, d^3x. \tag{4.6}$$

By (2.21) and (4.3),

$$m = \frac{1}{2} \int \langle \mathscr{E}_i \mathscr{E}_i{}^+ + \mathscr{B}_i \mathscr{B}_i{}^+ + D_i \Phi D_i \Phi^+ \rangle d^3x. \tag{4.7}$$

By (4.4) and (4.5),

$$m = \int \langle D_i \Phi D_i \Phi^+ \rangle d^3x \tag{4.8a}$$

$$= \int \langle \mathscr{E}_i \mathscr{E}_i{}^+ + \mathscr{B}_i \mathscr{B}_i{}^+ \rangle d^3x. \tag{4.8b}$$

From (3.16), (3.17) and (4.4), (4.5), one finds

$$Q \sin\alpha = \frac{1}{a} \int \langle \mathscr{E}_i \mathscr{E}_i{}^+ \rangle d^3x, \tag{4.9}$$

$$M \cos\alpha = \frac{1}{a} \int \langle \mathscr{B}_i \mathscr{B}_i{}^+ \rangle d^3x, \tag{4.10}$$

and

$$m = a(Q \sin\alpha + M \cos\alpha). \tag{4.11}$$

Also

$$aQ = \sin\alpha \int (D_i \Phi)(D_i \Phi^+) d^3x, \tag{4.12}$$

$$aM = \cos\alpha \int (D_i \Phi)(D_i \Phi^+) d^3x, \tag{4.13}$$

$$Q/M = \tan\alpha. \tag{4.14}$$

By (4.11) and (4.14),

$$m = a(Q^2 + M^2)^{1/2}. \tag{4.15}$$

This relation between mass and generalized electric and magnetic charge holds for an exact solution of the affine field equations, and in relation to the finite algebra it is known as the Bogomolny bound.[8,12]

The solution found by Prasad and Sommerfield, a special solution of the general loop equations, is the following[8,11]:

$$eW_a{}^i = [K(r) - 1]\epsilon_{aij} \hat{r}^j / r, \tag{4.16}$$

$$e\phi_a = (H(r)/r) \hat{r}^a, \tag{4.17}$$

$$W_a{}^0 = (J(r)/r) \hat{r}^a, \tag{4.18}$$

where

$$K(r) = \lambda r / \sinh(\lambda r), \tag{4.19}$$

$$H(r)\cos\alpha = \lambda r \coth(\lambda r) - 1, \tag{4.20}$$

$$J(r) = H(r)\sin\alpha, \tag{4.21}$$

and

$$\lambda = ae\cos\alpha. \tag{4.22}$$

Then

$$\lim_{r\to\infty} \phi(r) = a\hat{r}. \tag{4.23}$$

The differential equations for the functions J, H, and K are scale invariant. The scale of the soliton is fixed by the value of a, the vacuum expectation value of the Higgs field.

V. GENERALIZATION TO LOOP ALGEBRA

As already noted, Eqs. (4.4) and (4.5) hold for the loop algebra. We again assume (4.1) and (4.2) as well. Then

$$\mathscr{E}_k = G_{0k} = -D_k W_0. \tag{5.1}$$

By (4.4)

$$D_k W_0 = -\sin\alpha \, D_k \Phi, \tag{5.2}$$

so that we may take W_0 proportional to the Higgs potential as shown in (4.21). Then it is only necessary to solve (4.5).

We make the simplest generalization of the Yang–Wu ansatz[13]:

$$eW_k{}^{an} = g_n(r)\epsilon^a{}_{kl} X^l, \tag{5.3}$$

$$eW_0{}^{an} = J_n(r) X^a, \tag{5.4}$$

$$e\Phi^{an} = h_n(r) X^a. \tag{5.5}$$

Then

$$\mathscr{B}_k{}^{an} = \delta_k{}^a \mathscr{B}_1{}^n + X^a X_k \mathscr{B}_2{}^n, \tag{5.6}$$

where

$$e\mathscr{B}_1{}^n = -\left(r\frac{d}{dr} + 2\right) g_n(r), \tag{5.6a}$$

$$e\mathscr{B}_2{}^n = \frac{1}{r}\frac{dg_n}{dr} + \sum_p g_{n-p}\, g_p, \tag{5.6b}$$

while

$$(\nabla_k, \Phi^{an}) = \delta_k{}^a \mathrm{I}^n + X^a X_k \mathrm{II}^n, \tag{5.7}$$

$$e\mathrm{I}^n = h_n - r^2 \sum_p g_{n-p}\, h_p, \tag{5.7a}$$

$$e\mathrm{II}^n = \frac{1}{r}\frac{dh_n}{dr} + \sum_p g_{n-p}\, h_p. \tag{5.7b}$$

Then (4.5) becomes

$$rg_n' + 2g_n = \cos\alpha\left[r^2 \sum_p g_{n-p}\, h_p - h_n\right], \tag{5.8}$$

$$\frac{1}{r} g_n' + \sum_p g_{n-p}\, g_p = \cos\alpha\left[\frac{1}{r} h_n' + \sum_p g_{n-p}\, h_p\right]. \tag{5.9}$$

In order to describe a soliton solution, these equations must be satisfied subject to the boundary conditions that all the components of Φ and W_μ be regular at $r = 0$ and that they approach the Higgs vacuum at $r = \infty$. We shall require finiteness and a flat tangent at the origin as well as

$$\lim_{r\to\infty} D_\mu \Phi = 0, \tag{5.10}$$

$$\lim_{r\to\infty} \Phi_m = a_m \hat{r}. \tag{5.11}$$

Then

$$\lim_{r\to\infty} \langle \Phi^+ \Phi \rangle = \sum a_m{}^2 = a^2. \tag{5.12}$$

VI. SPECIAL SOLUTIONS

Let us denote the nth soliton by the set (**g**,**h**), where $g_m = h_m = 0$ if $|m| > n$ and where

$$\mathbf{g} = (g_{-n} \cdots g_0 \cdots g_n), \tag{6.1}$$

$$\mathbf{h} = (h_{-n} \cdots h_0 \cdots h_n). \tag{6.2}$$

By (2.7) and (2.8), $g_m = g_{-m}$ and $h_m = h_{-m}$. According to (5.8) and (5.9) the components (g_0, h_0) satisfy

$$rg_0' = -2g_0 - \bar{h}_0 + r^2\left(g_0 \bar{h}_0 + 2\sum_1^n g_m \bar{h}_m\right), \tag{6.3}$$

$$r\bar{h}_0' = -2g_0 - \bar{h}_0 + r^2\left(g_0{}^2 + 2\sum_1^n g_m{}^2\right). \tag{6.4}$$

The remaining components satisfy

$$rg_m' = -2g_m - \bar{h}_m + r^2 \sum_0^m g_{m-p}\, \bar{h}_p, \tag{6.5}$$

$$r\bar{h}_m' = -2g_m - \bar{h}_m + r^2 \sum_0^m g_{m-p}\, g_p. \tag{6.6}$$

Here

$$\bar{h}_m = \cos\alpha\, h_m. \tag{6.7}$$

The complete set of components $(g_0 \cdots g_n; h_0 \cdots h_n)$ is codetermined by the nonlinear equations (6.3)–(6.6). Notice that (h_m, g_m) cannot be renormalized independently. They can be rescaled together, however, if lengths are also rescaled.

The Prasad–Sommerfield solution is recognized in this class of solutions as

$$\bar{h}_m = g_m = 0, \quad m \neq 0, \tag{6.8}$$

$$\bar{h}_0(r) = (1/r^2)[1 - \lambda_0 \coth(\lambda_0 r)], \tag{6.9}$$

$$g_0(r) = (1/r^2)[1 - \lambda_0 r/\sinh(\lambda_0 r)], \tag{6.10}$$

where

$$\lambda_0 = ae\cos\alpha. \tag{6.11}$$

This solution will be described as a "singlet."

Since all components are coupled nonlinearly, there is no obvious method to investigate the general solution of these equations. One may, however, obtain special solutions in the following way. With r corresponding to the time, the general solution will trace out an orbit in (**h**,**g**) space. We may find special solutions in this space by confining the motion to a subspace of the full space. In particular, we may ask for solutions lying in the following subspace:

$$\sum_1^n g_m\, h_m = K g_0 h_0, \tag{6.12}$$

$$\sum_1^n g_m{}^2 = K g_0{}^2. \tag{6.13}$$

Then (6.3) and (6.4) become

$$rg_0' = -2g_0 - \bar{h}_0 + (1 + 2K) r^2 g_0 \bar{h}_0, \tag{6.14}$$

$$r\bar{h}_0' = -2g_0 - \bar{h}_0 + (1 + 2K) r^2 g_0{}^2. \tag{6.15}$$

Now rescale the length by

$$x^2 = (1 + 2K) r^2. \tag{6.16}$$

Then

$$x\frac{dg_0}{dx} = -2g_0 - \bar{h}_0 + x^2 g_0 \bar{h}_0, \tag{6.17}$$

$$x\frac{d\bar{h}_0}{dx} = -2g_0 - \bar{h}_0 + x^2 g_0{}^2. \tag{6.18}$$

These rescaled equations have the explicit solutions for (g_0, h_0) given by (6.9) and (6.10) in which r is replaced by $(1 + 2K)^{1/2} r$. The rescaling, of course, depends on the other components $(g_1 \cdots g_n)$.

Since they are not integrals of the complete set of equations, (6.12) and (6.13) are nonlinear constraints on the set (6.5) and (6.6). These constraints may be satisfied for the special solutions

$$\begin{aligned} &\mathbf{g}_n = (g_{-n} \circ \cdots \circ g_0 \circ \cdots \circ g_n), \\ &g_m = 0, \quad |m| = 1, \ldots, n-1, \\ &\mathbf{h}_n = (h_{-n} \circ \cdots \circ g_0 \circ \cdots \circ h_n), \\ &h_m = 0, \quad |m| = 1, \ldots, n-1. \end{aligned} \tag{6.19}$$

These special solutions will be termed "triplets." Then (6.12) and (6.13) imply

$$g_n = kg_0,$$
$$\bar{h}_n = k\bar{h}_0, \tag{6.20}$$

where

$$k = \pm\sqrt{K} \tag{6.21}$$

and

$$\sum_0^n g_{n-p}\,\bar{h}_p = 2kg_0\,\bar{h}_0, \tag{6.22}$$

$$\sum_0^n g_{n-p}\,g_p = 2kg_0{}^2. \tag{6.23}$$

Equations (6.5) and (6.6) for the components (g_n, h_n) become

$$rg_0{}' = -2g_0 - \bar{h}_0 + 2r^2 g_0\,\bar{h}_0, \tag{6.24}$$

$$r\bar{h}_0{}' = -2g_0 - \bar{h}_0 + 2r^2 g_0{}^2. \tag{6.25}$$

These are identical with (6.14) and (6.15) for $(g_0,\bar{h}_0)$ with

$$1 + 2K = 2,$$
$$k = \pm\tfrac{1}{2}\sqrt{2}. \tag{6.26}$$

Therefore $(g_0,\bar{h}_0)$ and $(g_n,\bar{h}_n)$ are given by the explicit solutions (6.9) and (6.10) in which r is replaced by $\sqrt{2}r$.

By (6.20) and (6.26), however, g_n and h_n are normalized differently from g_0 and h_0:

$$g_n(x) = \tfrac{1}{2}\sqrt{2}g_0(x), \tag{6.27}$$

$$h_n(x) = \tfrac{1}{2}\sqrt{2}h_0(x). \tag{6.28}$$

We have now found an infinite set of exact solutions. Let us denote their h components by $h_m{}^{(n)}$. The simplest of these is the singlet for which

$$h_m{}^{(0)} = (r^2\cos\alpha)^{-1}(1 - \lambda_0 r\coth(\lambda_0 r))\delta_{m0}. \tag{6.29}$$

This is the previously obtained solution given by (6.8)–(6.11).

The new set contains the triplets for which

$$h_m{}^{(n)} = \delta_{m0}h_0{}^{(0)}(x) + \tfrac{1}{2}\sqrt{2}\,\delta_{n|m|}h_0{}^{(0)}(x), \quad n>0, \tag{6.30}$$

where

$$x = \sqrt{2}r. \tag{6.31}$$

Let us label the boundary values by $a_m{}^{(n)}$. Then, by (5.11),

$$\lim_{r\to\infty} r|h_m{}^{(n)}(r)| = ea_m{}^{(n)}, \tag{6.32}$$

and, by (6.30), one finds

$$a_0{}^{(n)} = \tfrac{1}{2}\sqrt{2}\,a_0{}^{(0)}, \tag{6.33}$$

$$a_n{}^{(n)} = a_{-n}{}^{(n)} = \tfrac{1}{2}a_0{}^{(0)}. \tag{6.34}$$

We are also interested in the sums

$$(a^{(n)})^2 \equiv \lim_{r\to\infty}\langle\Phi^{(n)}(\Phi^{(n)})^+\rangle \tag{6.35a}$$

$$= \sum_{-n}^{n}(a_m{}^{(n)})^2. \tag{6.35b}$$

For the triplets

$$(a^{(n)})^2 = (a_0{}^{(n)})^2 + 2(a_n{}^{(n)})^2. \tag{6.36}$$

Then by (6.33) and (6.34),

$$a^{(n)} = \pm a^{(0)}. \tag{6.37}$$

VII. ELECTRIC AND MAGNETIC CHARGES AND MASS

The formally generalized electric charge may be calculated by (3.17):

$$a^{(n)}Q^{(n)} = \lim_{r\to\infty} r^2\int d\Omega\langle\Phi^+\mathscr{E}_r\rangle \tag{7.1}$$

$$= 4\pi\lim_{r\to\infty} r^2\langle\Phi^+\mathscr{E}_r\rangle. \tag{7.2}$$

The scalar product is

$$\langle\Phi^+\mathscr{E}_r\rangle = \sin\alpha\langle\Phi^+D_r\Phi\rangle$$
$$= \sin\alpha\langle\Phi^+\,\partial_r\Phi + \Phi^+(W_r,\Phi)\rangle.$$

The commutator term depends on

$$\langle\Phi^+(W_r,\Phi)\rangle = -\langle W_r(\Phi,\Phi)\rangle = 0.$$

Then

$$\langle\Phi^+\mathscr{E}_r\rangle = \sin\alpha\langle\Phi^+\,\partial_r\Phi\rangle$$
$$= \tfrac{1}{2}\sin\alpha\sum_{an}\frac{d}{dr}(\Phi_n{}^a)^2. \tag{7.3}$$

The total charge of the nth solution is

$$Q^{(n)} = \sum_{-n}^{n} Q_m{}^{(n)}, \tag{7.4}$$

where

$$a^{(n)}Q_m{}^{(n)} = \frac{2\pi}{e^2}\sin\alpha\lim_{r\to\infty} r^2\frac{d}{dr}r^2[h_m{}^{(n)}(r)]^2. \tag{7.5}$$

Then for the singlet

$$a^{(0)}Q^{(0)} = \frac{2\pi}{e^2}\sin\alpha\lim_{r\to\infty} r^2\frac{d}{dr}r^2[h_0{}^{(0)}(r)]^2, \tag{7.6}$$

where $h_0{}^{(0)}(r)$ is given by (6.29) and

$$\lambda_0 = a^{(0)}e\cos\alpha. \tag{7.7}$$

Taking the limit one finds the known result

$$Q^{(0)} = (4\pi/e)\tan\alpha. \tag{7.8}$$

For the triplets, one has

$$Q^{(n)} = Q_0{}^{(n)} + 2Q_n{}^{(n)}, \tag{7.9}$$

where

$$a^{(n)}Q_0{}^{(n)} = \frac{2\pi}{e^2}\sin\alpha\lim_{r\to\infty} r^2\frac{d}{dr}r^2(h_0{}^{(n)}(r))^2. \tag{7.10}$$

By (6.30),

$$a^nQ_0{}^{(n)} = \frac{2\pi}{e^2}\sin\alpha\left(\frac{1}{\sqrt{2}}\right)^3\lim_{x\to\infty} x^2\frac{d}{dx}x^2(h_0{}^{(0)}(x))^2$$
$$= (1/\sqrt{2})^3a^{(0)}Q^{(0)} \tag{7.11}$$

or

$$Q_0{}^{(n)} = \tfrac{1}{4}\sqrt{2}Q^{(0)}, \tag{7.12}$$

by (6.30) and (6.37). The additional contributions to the total triplet charge are $Q_n^{(n)}$ and $Q_{-n}^{(n)}$. For these we have

$$a^{(n)}Q_{\pm n}^{(n)} = \frac{2\pi}{e^2}\sin\alpha \lim_{r\to\infty} r^2 \frac{d}{dr} r^2 (h_n^{(n)}(r))^2$$

$$= \left(\frac{1}{\sqrt{2}}\right)^3 \frac{1}{2}\frac{2\pi}{e^2}\sin\alpha$$

$$\times \lim_{x\to\infty} x^2 \frac{d}{dx} x^2 (h_0^{(0)}(x))^2$$

$$= (\sqrt{2}/8)a^{(0)}Q^{(0)}, \tag{7.13}$$

$$\therefore \quad Q_{\pm n}^{(n)} = (\sqrt{2}/8)Q^{(0)}. \tag{7.14}$$

Then the total triplet charge is

$$Q^{(n)} = (\sqrt{2}/4 + 2\sqrt{2}/8)Q^{(0)} = \tfrac{1}{2}\sqrt{2}Q^{(0)} \tag{7.15a}$$

or

$$Q^{(n)} = (\sqrt{2}/2)((4\pi/e)\tan\alpha). \tag{7.15b}$$

The corresponding magnetic charge is

$$M^{(n)} = (\sqrt{2}/2)(4\pi/e), \tag{7.16}$$

by (4.14).

The true electric and magnetic charges come from the Maxwell field only and for the singlet have the usual values:

$$q^{(0)} = Q_0^{(0)} = (4\pi/e)\tan\alpha, \tag{7.17}$$

$$g^{(0)} = 4\pi/e. \tag{7.18}$$

For the triplets these charges are reduced:

$$q^{(n)} = Q_0^{(n)} = (\sqrt{2}/4)q^{(0)}, \tag{7.19}$$

$$g^{(n)} = (\sqrt{2}/4)g^{(0)}. \tag{7.20}$$

The mass is determined by the total charge according to (4.14) and (4.15):

$$m = aQ/\sin\alpha. \tag{7.21}$$

Then

$$m^{(0)} = a^{(0)}Q^{(0)}/\sin\alpha = a^{(0)}q/\sin\alpha,$$

$$m^{(n)} = a^{(n)}Q^{(n)}/\sin\alpha. \tag{7.22}$$

By (7.15a),

$$m^{(n)} = (\sqrt{2}/2)a^{(0)}Q^{(0)}/\sin\alpha,$$

$$\therefore \quad m^{(n)} = (\sqrt{2}/2)m^{(0)}. \tag{7.23}$$

The triplet dyons have less charge and also less mass than the singlet. All triplets have the same charges and mass.

VIII. GENERAL SOLUTIONS

The special solutions just discussed are not only exact but are also the simplest. They also span an infinite-dimensional function space, but a general solution of the soliton equations (5.8) and (5.9) does not lie in this space since the soliton equations are nonlinear.

We have also studied a slightly different set of equations obtained by dropping the terms in (6.3) and (6.4) coupling (h_0,g_0) to the higher modes (h_m,g_m), namely, $\Sigma_1^n {g_m}^2$ and $\Sigma_1^n g_m h_m$. In this truncation the remaining equations are to be left unchanged. We propose the so-modified set of equations as the basis of a general perturbation method for investigating arbitrary solutions of the exact equations, provided that the higher amplitudes (h_m,g_m) are small compared to the zero set (h_0,g_0). The modified equations themselves may be solved exactly by the following procedure.

Let the general solution be

$$\vec{g} = (g_{-n}g_{-n+1}\cdots g_0\cdots g_{n-1}g_n),$$
$$\vec{h} = (h_{-n}h_{-n+1}\cdots h_0\cdots h_{n-1}h_n). \tag{8.1}$$

where (g_0,h_0) is the exact Prasad–Sommerfield solution since (g_0,h_0) is now decoupled from the other modes. The remaining components may be shown to satisfy

$$g_m'' + Pg_m' + Q_g g_m = S_m^{(g)}, \tag{8.2g}$$

$$h_m'' + Ph_m' + Q_h h_m = S_m^{(h)}, \tag{8.2h}$$

where

$$P = (2/r)F, \tag{8.3}$$

$$Q_g = [F^2 - 3G^2 - 1]/r^2, \tag{8.4}$$

$$S_m^{(g)} = G\sum_{p=1}^{m-1} g_{m-p}g_p, \tag{8.5}$$

and

$$Q_h = 2[F - G^2 - 1]/r^2, \tag{8.6}$$

$$S_m^{(h)}\cos\alpha = \left[\left(r\frac{d}{dr} + 2F\right)\sum_{p=1}^{m-1} g_{m-p}g_p + 2G\sum_{p=1}^{m-1} g_{m-p}h_p\right]. \tag{8.7}$$

These expressions all depend on only the two functions F and H, which are linear functionals of h_0 and g_0:

$$F = 2 - r^2\cos\alpha\, h_0, \tag{8.8}$$

$$G = r^2 g_0 - 1. \tag{8.9}$$

From (8.2h) and (8.2g) one sees that (h_m,g_m) satisfy linear equations where the coefficients depend on the pair (h_0,g_0), which in turn satisfy the nonlinear equations (6.14) and (6.15) with $K=0$. Since these coefficients depend only on (h_0,g_0), they are the same for all m. The different components are distinguished by only the source terms $(S_m^{(h)},S_m^{(g)})$, which do not depend on (h_m,g_m) but do depend on all the lower components (h_p,g_p), where $1\leqslant p\leqslant m-1$. As a consequence of this structure these coupled equations may be solved by first solving the nonlinear equations for (h_0,g_0) to obtain (6.9) and (6.10), and then proceeding along the sequence $(h_1,g_1)\cdots(h_m,g_m)\cdots(h_n,g_n)$ by solving at each step equations that are only linear in the new variables (h_m,g_m).

Since there is no source for the (h_1,g_1) equation, the required (h_1,g_1) functions satisfy homogeneous equations subject to the boundary conditions at the origin and infinity. To proceed to higher states we shall introduce the Green's function, which also satisfies the proper boundary conditions. The general solution of these equations is the sum of the solutions of the homogeneous equation and a special solution determined by the source:

$$g_m(r) = \mathring{g}_m(r) + \int_0^\infty W(r) G_g(r,r') S_m^{(g)}(r') dr', \tag{8.10g}$$

$$h_m(r) = \mathring{h}_m(r) + \int_0^\infty W(r) G_h(r,r') S_m^{(h)}(r') dr'. \tag{8.10h}$$

Since the differential equations are not Sturm–Liouville, it is necessary to introduce a weight function $W(r)$ independent of m:

$$W(r) = (r \sinh r)^2. \tag{8.11}$$

The Green's functions are also independent of m since they depend on only (h_1,g_1) as follows:

$$G_g(r,r') = \begin{cases} -g_{1<}(r') g_{1>}(r), & r > r', \\ -g_{1>}(r') g_{1<}(r), & r < r'; \end{cases} \tag{8.12g}$$

$$G_h(r,r') = \begin{cases} \frac{1}{2} h_{1<}(r') h_{1>}(r), & r > r', \\ \frac{1}{2} h_{1>}(r') h_{1<}(r), & r < r'; \end{cases} \tag{8.12h}$$

where a subscript $<$ refers to solutions of homogeneous equations satisfying boundary conditions at the origin and a subscript $>$ refers to an independent solution satisfying boundary conditions at infinity.

IX. THE LOWEST STATES OF THE MODIFIED EQUATIONS

The first excited state is characterized by $(\mathbf{h},\mathbf{g})$, where $\mathbf{h} = (h_0,h_1)$ and $\mathbf{g} = (g_0,g_1)$. The components (h_0,g_0) have already been given in (6.9) and (6.10). The new components satisfy the homogeneous linear equations

$$g_1'' + P g_1' + Q_g g_1 = 0, \tag{9.1g}$$

$$h_1'' + P h_1' + Q_h h_1 = 0, \tag{9.1h}$$

with solutions of the form

$$g_1(r) = \exp\left[-\frac{1}{2} \int^r P(s) ds \right] U_g(r), \tag{9.2g}$$

$$h_1(r) = \exp\left[-\frac{1}{2} \int^r P(s) ds \right] U_h(r), \tag{9.2h}$$

where

$$U_g'' + (Q_g - \tfrac{1}{2}P' - \tfrac{1}{4}P'') U_g = 0, \tag{9.3g}$$

$$U_h'' + (Q_h - \tfrac{1}{2}P' - \tfrac{1}{4}P'') U_h = 0. \tag{9.3h}$$

By (8.3), (8.4), and (8.6), one finds

$$P = (2/r)(1 + \lambda_0 r \coth \lambda_0 r), \tag{9.4}$$

$$Q_g - \tfrac{1}{2}P' - \tfrac{1}{4}P^2 = -2(\lambda_0/\sinh \lambda_0 r)^2, \tag{9.5}$$

$$Q_h - \tfrac{1}{2}P' - \tfrac{1}{4}P^2 = -\lambda_0^2 - 2(\lambda_0/\sinh \lambda_0 r)^2. \tag{9.6}$$

There are two independent solutions of both the h and g equations. These may be chosen to be

$$g_{1<} = \frac{1}{\sinh \lambda_0 r} \left(\coth \lambda_0 r - \frac{1}{\lambda_0 r} \right), \tag{9.7}$$

$$g_{1>} = \frac{1}{\lambda_0 r} \frac{\coth \lambda_0 r}{\sinh \lambda_0 r} \tag{9.8}$$

and

$$h_{1<} = \frac{1}{\lambda_0 r} \left[-\coth \lambda_0 r + \frac{\lambda_0 r}{\sinh^2 \lambda_0 r} \right], \tag{9.9}$$

$$h_{1>} = \frac{1}{\lambda_0 r} \frac{1}{\sinh^2 \lambda_0 r}. \tag{9.10}$$

Here $g_{1<}$ and $h_{1<}$ satisfy the boundary conditions of regularity at the origin; these functions also vanish properly at infinity. On the other hand, $g_{1>}$ and $h_{1>}$ vanish at infinity but blow up at the origin. The solutions $g_{1<}$ and $h_{1<}$ satisfy not only the second-order equations (9.1) but also the first-order equations (6.5) and (6.6), as required. The functions $g_{1<}$ and $h_{1<}$ therefore describe the approximate solitonic solution we seek.

Since both the first- and second-order equations are linear and homogeneous in g_1 and h_1, these equations do not fix the normalization of these functions. However, since the first-order equations (6.5) and (6.6) relate the normalization of g_1 and h_1, only one normalization condition needs to be supplied. We take this to be the boundary condition at infinity for $h_{1<}$. At the two limits we have

$$\lim_{r \to 0} g_{1<}(\lambda_0 r) = \tfrac{1}{3} A_1,$$

$$\lim_{r \to 0} h_{1<}(\lambda_0 r) = \tfrac{2}{3} A_1, \tag{9.11}$$

where A_1 normalizes $g_{1<}$ and $h_{1<}$ and

$$\lim_{r \to \infty} g_{1<}(\lambda_0 r) = A_1 e^{-\lambda_0 r},$$

$$\lim_{r \to \infty} |h_{1<}(\lambda_0 r)| = A_1/\lambda_0 r. \tag{9.12}$$

But we require, by (5.11),

$$\lim_{r \to \infty} r|h_n(r)| = e a_n \tag{9.13}$$

or

$$A_1 = \lambda_0 \lambda_1,$$

$$\lambda_1 = a_1 e \cos \alpha. \tag{9.14}$$

The approximate soliton $(h_0,g_0;h_1,g_1)$ is then entirely determined by the parameters appearing in the action and the constants $\{a_m\}$, the vacuum expectation values of the extended Higgs field.

The (h_2,g_2) solutions of the modified equations are

$$h_2 = \frac{1}{\cos \alpha} \frac{1}{y} \left[(A_2 + \lambda_1^2) \frac{y}{\sinh^2 y} - \lambda_1^2 \frac{y^2 \coth y}{\sinh^2 y} - A_2 \coth y \right], \tag{9.15h}$$

$$g_2 = (A_2 + \lambda_1^2) \frac{\coth y}{\sinh y} - A_2 \frac{1}{y \sinh y} - \frac{\lambda_1^2}{2} \left[\frac{y}{\sinh^3 y} + \frac{y \coth^2 y}{\sinh y} \right], \tag{9.15g}$$

where $y = \lambda_0 r$ and A_2 is the constant that multiplies the solution of the homogeneous equation. The value of A_2 is fixed by the limit of h_2 at large distances. By (9.13), $A_2 = \lambda_0 \lambda_2$, with $\lambda_2 = a_2 e \cos \alpha$.

There is no reason for this series of solutions to termi-

nate. The recursive procedure for generating new solutions is described in Sec. VIII. The explicit form of the Green's function appearing there is obtained by substituting the appropriate h and g solutions from (9.7)–(9.10) into (8.12). The general solution is the sum of solutions of the homogeneous equation and the special solution depending on the source and the Green's function.

At each stage of this recursive procedure one must determine the constant associated with the solution to the homogeneous equation. In the next step we encounter the constant $A_3 = \lambda_0\lambda_3$ with $\lambda_3 = a_3\, e \cos\alpha$.

Although we do not have a general formula for the solutions at any level n, only the contribution from the homogeneous solution is important in the calculation of (9.13).

For $n = 3$, we have

$$h_3 = \frac{1}{\cos\alpha}\left[\frac{1}{y}\left[(A_3 + 2\lambda_2\lambda_1)\frac{y}{\sinh^2 y} - A_3 \coth y\right] - \left(\frac{\lambda_1^3}{\lambda_0} + 2\lambda_1\lambda_2\right)\frac{y\coth y}{\sinh^2 y} + \frac{\lambda_1^3}{\lambda_0}\left[\frac{y^2\coth^2 y}{\sinh^2 y} - \frac{1}{3}\frac{y^2}{\sinh^2 y}\right]\right], \tag{9.16h}$$

$$g_3 = (A_3 + 2\lambda_2\lambda_1)\frac{\coth y}{\sinh y} - A_3\frac{1}{y\sinh y} - \left(\lambda_2\lambda_1 + \frac{1}{2}\frac{\lambda_1^3}{\lambda_0}\right)\left[\frac{2y}{\sinh^3 y} + \frac{y}{\sinh y}\right] + \frac{\lambda_1^3}{\lambda_0}\left(\frac{y^2\coth y}{\sinh^3 y} + \frac{1}{6}\frac{y^2\coth y}{\sinh y}\right). \tag{9.16g}$$

For $n = 1$ there is a test of this perturbative procedure since we have both the exact solution and the approximate solution as determined from the modified equations. For small r the approximate solution (9.9) becomes

$$h_1^{(1)} \sim (1/(\lambda_0 r)^2)\,[(\lambda_0 r)\coth(\lambda_0 r) - 1]. \tag{9.17}$$

This expression is of the same form as the exact solution (6.30). At large distances, on the other hand, the exact and approximate solutions do not have the same form; but at large distances one has a linear approximation to the unmodified equations.

In general, if the higher n components are small compared to (h_0, g_0), one would expect the preceding procedure based on the modified equations to be useful.

X. REMARKS

In order to investigate the dependence of the soliton sector on the affine index, we have studied a very special model that can be generalized in different ways. In particular, if the total field contained an additional component tranforming according to some irreducible and higher-dimensional representation of SU(2), this additional component carrying electric charge would interact with the magnetic charge so as to convert internal degrees of freedom into spin degrees of freedom with the result that the total angular momentum would no longer vanish; the total angular momentum would then be either integral or half integral, depending on the transformation properties of the additional field.[14,15] One would also expect that this additional structure might lift the infinite mass degeneracy of the triplet solitons. Then the members of the infinite family of dyons would differ not only in mass but also in angular momentum and statistics,[16] and in a way that can be calculated once the model is made definite.

The refinements just described have not yet been investigated. We have shown elsewhere, however, how the theory may be made globally[17] and also locally[18] supersymmetric.

The preceding comments refer to the soliton sector of the theory, which one may expect to be excited only at very high energies. At currently accessible energies, on the other hand, the vacuum or "meson" sector provides the appropriate formalism. Here the particle states have spin 1 and 0. In this sector, where one has the standard procedures, we have previously studied the Higgs splitting and found a linear mass spectrum for the vectors if the Higgs has a component lying in the L_0 direction of the associated Virasoro algebra. In the present paper our model is different and our object has been to concentrate on the affine solitons rather than to provide a unified consistent model for both sectors. One may, however, plausibly assume that the soliton and vacuum sectors exhibit different representations of the Kac–Moody algebra and, in particular, that the central operator k and the Virasoro operator L_0 vanish for the soliton sector. A more realistic model would also address this question.

APPENDIX: DERIVATION OF EQS. (4.4) AND (4.5)

Equations (4.4) and (4.5) may be obtained from the field equations in the following way. If we distinguish space from time coordinates, the set (2.17)–(2.19) becomes

$$(\nabla_k, G^{0k}) = 0, \tag{A1}$$

$$(\nabla_k, \tilde{G}^{0k}) = 0, \tag{A2}$$

$$(\nabla_k, (\nabla^k, \Phi)) = 0, \tag{A3}$$

$$(\nabla_\lambda, \tilde{G}^{k\lambda}) = 0, \tag{A4}$$

$$(\nabla_\lambda, G^{k\lambda}) = (\Phi^+, (\nabla^k, \Phi)). \tag{A5}$$

Let us try

$$G_{k0} = a(\nabla_k, \Phi), \tag{A6}$$

$$\tilde{G}_{k0} = b(\nabla_k, \Phi). \tag{A7}$$

By (A6) and (A7), Eqs. (A1)–(A3) become equivalent. Then it is enough to check (A2). But (A2) and (A4) together comprise the Bianchi identity. Since both may be checked in the same way, we write out (A4) only.

One equation of the set (A4) is

$$(\nabla_0, \tilde{G}^{10}) + (\nabla_2 \tilde{G}^{12}) + (\nabla_3 \tilde{G}^{13}) = 0$$

or

$$(\nabla_0, \tilde{G}_{01}) + (\nabla_2, G_{03}) + (\nabla_3, G_{20}) = 0.$$

By (A6) and (A7),

$$-b(\nabla_0, (\nabla_1, \Phi)) - a(\nabla_2, (\nabla_3, \Phi)) + a(\nabla_3, (\nabla_2, \Phi)) = 0.$$

By the Jacobi identity,

$$-b(\nabla_0, (\nabla_1, \Phi)) + a(\Phi, (\nabla_2, \nabla_3)) = 0.$$

By the Jacobi identity and (4.3),

$$b(\Phi,(\nabla_0,\nabla_1)) + a(\Phi,(\nabla_2,\nabla_3)) = 0.$$

By (2.11),

$$b(\Phi,G_{01}) + a(\Phi,G_{23}) = 0. \tag{A8}$$

Finally by (A6) and (A7) one sees that (A8) is satisfied identically, as it should.

By a similar reduction of (A5) one finds

$$a^2 + b^2 = 1. \tag{A9}$$

The preceding argument holds any for any algebra including the loop algebra.

[1]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[2]A. M. Polyakov, JETP Lett. **20**, 194 (1974).
[3]B. Julia and A. Zee, Phys. Rev. D **11**, 2227 (1975).
[4]J. Schwinger, Science **165**, 757 (1969).
[5]For a field theory example of noding solitons, see R. Finkelstein, R. LeLevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[6]A. C. Cadavid and R. Finkelstein, Lett. Math. Phys. **16**, 279 (1988).
[7]A. C. Cadavid and R. Finkelstein, Lett. Math. Phys. **17**, 157 (1989).
[8]P. Goddard and D. Olive, Rep. Prog. Phys. **41**, 1361 (1978).
[9]J. Arafune, P. Freund, and C. Goebel, J. Math. Phys. **16**, 433 (1975).
[10]E. B. Bogomolny, Sci. J. Nucl. Phys. **24**, 449 (1979); L. D. Faddeev, Lett. Math. Phys. **7**, 289 (1976).
[11]M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[12]S. Coleman, S. Parke, A. Neveu, and C. M. Sommerfield, Phys. Rev. D **15**, 544 (1977). Equations (4.4) and (4.5) in this reference follow the saturation of the Bogomolny bound. Here the argument is turned around. See the Appendix.
[13]T. T. Wu and C. N. Yang, *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1969).
[14]R. Jackiw and C. Rebbi, Phys. Rev. Lett. **36**, 1116 (1976).
[15]P. Hasenfratz and G. 't Hooft, Phys. Rev. Lett. **36**, 1119 (1976).
[16]A. S. Goldhaber, Phys. Rev. Lett. **36**, 1122 (1976).
[17]A. C. Cadavid and R. Finkelstein, Int. J. Mod. Phys. A **4**, 2967 (1989).
[18]A. C. Cadavid and R. Finkelstein, "Affine extension of supergravity," to be published in Int. J. Mod. Phys. A.

Quantum equations for simple interacting strings: Their exact solutions and physical implications

J. P. Hsu
Physics Department, Southeastern Massachusetts University, North Dartmouth, Massachusetts 02747

(Received 1 February 1989; accepted for publication 21 June 1989)

A string can have a simple potential interaction only when it has a constant radius. The Hamiltonian for such a string is characterized by a cyclic radial momentum p_r. Physical properties of this interacting string (or "ring") are illustrated by exact solutions of quantum ring equations (with a "flavor potential") in the four-dimensional space-time and in a particular reference frame. The energy eigenstates of the ring are ultrastable because flavor transitions between different states are forbidden. Furthermore, the ring can be permanently confined by a potential $\propto r^{-1}\ln(\lambda r)$, and the size of the ring can be detected by scattering processes. In the absence of the electroweak interaction, a simple composite quark model based on a particle–ring interaction is discussed. It is found that the relation $m_s/m_c \approx m_b/m_t \approx m_{b'}/m_{t'}$ holds independently of the parameters in the potential. Differences and similarities between quantum rings and Nambu strings are discussed.

I. INTRODUCTION

The idea of strings has a great influence on contemporary theoretical physics.[1] To explore its possible applications in particle physics, one should have at least a quantum-mechanical equation for a string moving in a potential field. We observed that *only when the string is closed and has a constant radius, can one have a simple string equation with a potential interaction*. Such a closed string may be termed a "ring." We first consider the classical motion of a string in a central force field based on a Hamiltonian with the cyclic radial momentum p_r (which implies constant radii). Classically, such a Hamiltonian is trivial, but it is important for passing to the quantum theory. Based on a classical analog, we obtain quantum equations for bosonic rings and fermionic rings with a mass $m \geqslant 0$ and a potential interaction.

To illustrate physical properties of an interacting ring in the four-dimensional space-time, we consider a simple model of a massless and spin-0 ring moving in a flavor potential $C_0(r)$ produced by a particle in a particular frame. The ring is described by the quantum equation ($c = \hbar = 1$)

$$\left[\left(i\frac{\partial}{\partial t} - fC_0(r)\right)^2 + \frac{\partial^2}{\partial\sigma^2}\right]\Phi(\sigma,t) = 0. \tag{1}$$

This closed *quantum string* has a constant radius r, so that $\sigma = r\phi$, $0 \leqslant \sigma \leqslant 2\pi r$, and it *differs* from the Nambu string.

A composite quark model based on the interacting ring has some interesting properties (in a strong interaction approximation).

(a) Its energy eigenstates are "ultrastable."

(b) Its extended structure in space cannot be detected by quark–lepton scatterings because leptons do not carry flavor charges.

(c) In the strong interaction approximation, the ultrastability of the ring states enables us to interpret, for example, the conservation of the quantum numbers $n = 0$ and $n = 1$ of the d-ring system as the conservation of the strangeness and the bottomness, respectively.

(d) The quantum ring can be permanently confined by the basic quark with a new flavor potential $C_0(r) = fr^{-1}\ln(\lambda r)$, which leads to a mass spectrum consistent with known quark masses.

Although a free ring described by (1) with $f = 0$ satisfies an equation similar to that of the Nambu string,

$$\left[\frac{\partial^2}{\partial\tau^2} - \frac{\partial^2}{\partial\sigma^2}\right]X^\mu(\sigma,\tau) = 0, \tag{2}$$

the physics derived from the quantum equation (1) is quite different from that derived from (2). We note that both (1) and (2) are *formally* analogous to the massless Klein–Gordon equation with cyclic radial momentum p_r and cyclic angular momentum p_θ:

$$\left[\left(i\frac{\partial}{\partial t} - fC_0(r)\right)^2 + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]\Phi(\sigma,t) = 0. \tag{3}$$

This analogy suggests that one can consider (1) as a quantum-mechanical equation of a ring and $\Phi(\sigma,t)$ as a probability amplitude. This viewpoint leads to a conceptual departure from the original Nambu string equation (2), which is regarded as a classical equation of a freely moving relativistic string whose coordinates in space-time are described by $X^\mu(\sigma,t)$. The interactions of general strings described by the coordinates $X^\mu(\sigma,t)$ are not simple.[2] On the other hand, the scalar function $\Phi(\sigma,t)$ in (1) is interpreted as the wave function of a spin-0 bosonic ring. According to this new viewpoint, we can also have a vector function $\Phi^\mu(\sigma,t)$ and a spinor function $\Psi(\sigma,t)$ for other types of rings. The interpretation of $\Phi(\sigma,t)$ is based on a well-established method for passing from the classical theory to the quantum theory, which is discussed in Secs. II and III.

II. CLOSED STRING AND A HAMILTONIAN WITH CYCLIC p_r AND p_θ

The analogy of Eqs. (3) and (2) [or (1)] suggests that a closed string with a constant radius (i.e., a ring) is described by a Hamiltonian with cyclic radial momentum p_r

 0022-2488/89/112683-06$02.50

and cyclic angular momentum p_θ. This property can be seen more clearly in classical mechanics and can be generalized to quantum mechanics. Let us consider the relativistic Lagrangian $L_r(r,\theta,\phi,\dot\phi)$ with cyclic $\dot r$ and $\dot\theta$ in the action $S_r = \int L_r\, dt$:

$$S_r = -\int m\, ds\Big|_{\dot r,\dot\theta\ \text{cyclic}} = -\int m(1 - r^2\dot\phi^2 \sin^2\theta)^{1/2}\, dt, \tag{4}$$

for a free ring. If this ring moves in a Coulomb-like potential field $V(r)$, the ring Lagrangian is assumed to have the usual form:

$$L_r(r,\theta,\phi,\dot\phi) = -m(1 - r^2\dot\phi^2 \sin^2\theta)^{1/2} - V(r). \tag{5}$$

We have the Lagrange equations

$$\frac{d}{dt}\frac{\partial L_r}{\partial \dot q_i} - \frac{\partial L_r}{\partial q_i} = 0, \quad i = 1,2,3, \tag{6}$$

where $q_1 = r$, $q_2 = \theta$, and $q_3 = \phi$. Since $\dot r$ and $\dot\theta$ are cyclic, the radius r and the angle θ are constants determined by

$$\frac{\partial L_r}{\partial r} = 0, \quad \frac{\partial L_r}{\partial \theta} = 0, \tag{7}$$

respectively. In order to have the corresponding quantum mechanical equation for such a ring, we follow Routh's procedure for treating cyclic variables[3] and define a new Hamiltonian function H_r:

$$\begin{aligned} H_r(r,\theta,\phi,p_\phi) &= p_\phi\dot\phi - L_r \\ &= (m^2 + p_\phi^2/(r^2\sin^2\theta))^{1/2} + V(r), \\ p_\phi &= \frac{\partial L_r}{\partial \dot\phi}. \end{aligned} \tag{8}$$

We have seen that the momenta p_r and p_θ are cyclic in H_r. The usual Hamiltonian equations for ϕ and p_ϕ can be obtained. We also have the following equations (in which H_r plays the role of the Lagranginan[3]) for the ring's motion:

$$\frac{\partial H_r}{\partial r} = \frac{d}{dt}\frac{\partial H_r}{\partial \dot r} = 0, \quad \frac{\partial H_r}{\partial \theta} = \frac{d}{dt}\frac{\partial H_r}{\partial \dot\theta} = 0. \tag{9}$$

They determine the constant values $r = \alpha_r$ and $\theta = \alpha_\theta$, respectively, and are equivalent to equations in (7).

For an arbitrary central potential $V(r)$ in H_r given by (8), the equation $\partial H_r/\partial\theta = 0$ in (9) leads to a general results:

$$\theta = \pi/2 = \alpha_\theta. \tag{10}$$

Thus the new Hamiltonian $H_r = H_r(\alpha_r,\alpha_\theta,\phi,p_\phi)$ involves only ϕ and p_ϕ as the dynamical variables. Classically, this Hamiltonian H_r can describe a particle with a mass m moving in a circular orbit. Since a rotating ring ($\dot\phi \neq 0$, $r = \text{const}$) can be pictured as a collection of N particles moving in the same circular orbit, we can also interpret H_r as the Hamiltonian for a rotating ring with a total rest mass m. In this paper, we consider H_r in (8), in which p_r and p_θ are cyclic, as a classical ring Hamiltonian. It is the basis for our discussion of quatnum rings and a composite quark model.

III. QUANTUM EQUATIONS FOR INTERACTING BOSONIC AND FERMIONIC STRINGS

Based on classical analogy, it is natural to postulate that the wave function $\Phi(\phi,t)$ for a bosonic string with a mass m and an arbitrary value for p_ϕ satisfies the quantum equation

$$\left(i\frac{\partial}{\partial t} - V\right)^2\Phi(\phi,t) = \left(\frac{1}{r^2\sin^2\theta}p_\phi^2 + m^2\right)\Phi(\phi,t),$$
$$p_\phi = -i\frac{\partial}{\partial\phi}. \tag{11}$$

Similarly, for an interacting fermionic ring we postulate the quantum equation

$$\left[i\frac{\partial}{\partial t} - V(r)\right]\Psi = \left[\frac{1}{r\sin\theta}\alpha p_\phi + \beta m\right]\Psi,$$
$$\alpha^2 = \beta^2 = 1, \quad \alpha\beta + \beta\alpha = 0, \tag{12}$$

where α and β are 2×2 matrices. From the quantum-mechanical viewpoint, it appears difficult to distinguish among a mass point, a section of a ring (an open string), and a ring, if they have the same mass m and move with $\dot\phi \neq 0$, $r = \text{const}$, and $\theta = \pi/2$ in a potential $V(r)$. The reason is that they are all described by the same quantum equation.

In the nonrelativistic case, i.e., $p_\phi/r \ll m$, the quantum ring equation (11) can be approximated by

$$\left[i\frac{\partial}{\partial t} - V(r)\right]\Phi = \left(m + \frac{1}{2mr^2}p_\phi^2\right)\Phi. \tag{13}$$

Since $\Phi(\phi,t)$ has only one component, Eq. (11) describes a scalar string (which may be open or closed) with spin 0. A vector string is postulated to be described by $\Phi^\mu(\phi,t)$, which satisfies

$$\left[i\frac{\partial}{\partial t} - V(r)\right]^2\Phi^\mu(\phi,t) = \left[\frac{1}{r^2\sin^2\theta}p_\phi^2 + m^2\right]\Phi^\mu(\phi,t). \tag{14}$$

The fermionic string described by (12) has only two components. It resembles the two-component neutrino rather than the four-component electron. In the special case $m = 0$ in (12), we have $\alpha = \pm 1$.

In this paper, we are interested in the physical properties of the quantum ring moving within a central potential field and in its application to the composite quark model. Thus we investigate the quantum equation (11) with $m = 0$ and a central potential $V(r)$ in a particular reference frame. We note that the four-dimensional symmetry of (11) and that of the Nambu equation (2) are nontrivial. Equation (11) involves cyclic p_r and p_θ, which are frame dependent. The Nambu equation (2) is related to the gauge conditions, which are also frame dependent. One may look at these properties from the viewpoint of a four-dimensional framework with a common scalar evolution variable.[4] At any rate, for our discussion, we choose a particular reference frame to solve the stationary states of the quantum equation (11).

IV. MASS SPECTRA OF COMPOSITE QUARKS

Let us consider a massless ring with $I_z = +\frac{1}{2}$ and spin 0 moving in a flavor potential produced by the basic d quark

with $I_z = -\frac{1}{2}$ and spin $\frac{1}{2}$. The ring carries only a flavor charge f and does not have any other charge. Suppose the d quark located at $\mathbf{r} = 0$ is the source of the flavor potential,

$$C_\mu = (C_0(r),0,0,0).$$

At this stage it becomes important to note that the only clue we have for the flavor potential $C_0(r)$ is that is must lead to the quark mass spectrum consistent with experiments. Let us consider the following two forms of potential:

$$C_0(r) = \begin{cases} -fr^{-1}\log(1/\lambda r) = V_1, \\ -fr^{-1}\log(e + 1/\lambda r) = V_2, \quad \log e = 1. \end{cases} \tag{15}$$

The potential V_1 is simple for solving the quantum ring equation. Nevertheless, the potential V_2 is perhaps more realistic because it approaches $-f/r$ as $r \gg \lambda^{-1}$. For small $r \ll \lambda^{-1}$, we have $V_1 = V_2$. Assuming that bound states can only be isospin singlet, we can identify them with s and b quarks with isospin 0, spin $\frac{1}{2}$, and charge quantum number $-\frac{1}{3}$.

Since $C_0(r)$ is time independent and ϕ is the only coordinate variable, we write $\Phi(\phi,t)$ in the form

$$\Phi(\phi,t) = e^{-iEt}F(\phi). \tag{16}$$

It follows from (11) with $m = 0$ and (16) that

$$\frac{d^2F(\phi)}{d\phi^2} + nF(\phi) = 0, \tag{17}$$

$$E - fC_0(r) = |n|/r, \quad V = fC_0(r), \quad n = \text{const}, \tag{18}$$

where we have used $\theta = \pi/2$ in (10). In analogy with Bohr's atomic model [using $(E - e^2/r)^2 = p^2 + m^2$], *we take the positive square root in (18).* The usual requirement that $F(\phi)$ and $dF(\phi)/d\phi$ be continuous throughout the domain 0 to 2π of ϕ demands that $n = 0, \pm 1, \pm 2, \ldots$ in (17). Thus we have

$$F_n(\phi) = e^{-in\phi}/(2\pi)^{1/2}. \tag{19}$$

The constant values of r in (18) for stationary states must satisfy $\partial E/\partial r = 0$, which corresponds to $\partial H_r/\partial r = 0$ in (9). We find

$$r_n = \lambda^{-1}\exp(1 - |n|/f^2), \quad C_0 = V_1, \quad |n| = 0,1,2,\ldots . \tag{20}$$

For $C_0 = V_2$, the constant radii are determined by the equation

$$-\frac{|n|}{r^2} + \frac{f^2}{r^2}\log\left(e + \frac{1}{\lambda r}\right) + \frac{f^2}{\lambda r^3(e + 1/\lambda r)} = 0. \tag{21}$$

When $1/(\lambda r) \gg e = 2.718$, the values for r in (21) can be approximated by (20) with $n \geqslant 1$. Note that if $n = 0$ there is no solution for r in (21). Thus n must be 1,2,3,... for r_n in (21), which cannot be solved exactly.

Let us now concentrate on the simple case $C_0 = V_1$, which can be solved exactly for the model of composite quarks. The eigenvalues of the energy E are

$$E_n = f^2\lambda\exp(|n|/f^2 - 1), \quad |n| = 0,1,2,\ldots . \tag{22}$$

Since the ring is massless and the d quark has a mass $m_d = 0.0089 \pm 0.000\,26$ GeV (see Ref. 5), we have the exponential mass formula for the d-type quarks:

$$M_n^d = m_d + f^2\lambda\exp(|n|/f^2 - 1), \quad |n| = 0,1,2,\ldots , \tag{23}$$

which holds approximately because the d quark's motion is ignored. Using $M_0^d = m_s = 0.175 \pm 0.055$ GeV and $M_1^d = m_b = 5.3 \pm 0.1$ GeV (see Ref. 5), we can determine the flavor charge f and the composite mass scale λ:

$$f = 0.54 \pm 0.03, \quad \lambda = 1.5 \pm 0.7 \text{ GeV}. \tag{24}$$

Based on (23) and (24), the model predicts the mass of the fourth-generation quark b' to be

$$m_{b'} = M_2^d \approx 170 \text{ GeV}. \tag{25}$$

Similarly, let us consider the bound states of the u quark and the ring with isospin $I_z = -\frac{1}{2}$. Based on symmetry considerations, their interaction should have the same flavor charge f. But they may involve a different mass scale, $\lambda' \neq \lambda$, because m_u differs from m_d. Thus their flavor potential is assumed to be

$$fC_0'(r) = -f^2r^{-1}\log(1/\lambda' r). \tag{26}$$

Following the steps (20)–(22), we obtain a similar mass formula for the u-type quarks:

$$M_n^u = m_u + f^2\lambda'\exp(|n|/f^2 - 1), \quad |n| = 0,1,2,\ldots . \tag{27}$$

Using $m_u = 0.005 \pm 0.0015$ GeV and $M_0^u = m_c = 1.35 \pm 0.05$ GeV (see Ref. 5), we obtain

$$\lambda' = 12.6 \pm 2.0 \text{ GeV}. \tag{28}$$

The exponential mass formula (27) predicts the masses of the top quark t and the fourth-generation quark t' to be

$$m_t = 45 \pm 23 \text{ GeV}, \quad m_{t'} \approx 1400 \text{ GeV}, \tag{29}$$

which are consistent with experiments.[5] Furthermore, the mass formulas (23) and (27) imply the approximate relation

$$m_s/m_c \approx m_b/m_t \approx m_{b'}/m_{t'}, \quad \text{for any } f, \lambda, \text{ and } \lambda'. \tag{30}$$

V. RING WAVE FUNCTIONS AND SIZES OF COMPOSITE QUARKS

According to (16), (19), and (20), the composite s and b quarks are described, respectively, by the wave function Φ_0 and Φ_1:

$$\Phi_0(\phi,t) = \begin{cases} e^{-iE_0t}/(2\pi)^{1/2}, & r = r_0, \quad \theta = \pi/2, \\ 0, & r \neq r_0 \quad \text{or} \quad \theta \neq \pi/2, \end{cases} \tag{31}$$

$$\Phi_1(\phi,t) = \begin{cases} e^{-iE_1t - i\phi}/(2\pi)^{1/2}, & r = r_1, \quad \theta = \pi/2, \\ 0, & r \neq r_1 \quad \text{or} \quad \theta \neq \pi/2. \end{cases} \tag{32}$$

Thus we may picture composite s and b quarks as rings with radii r_0 and r_1 and with the basic d quark at the center.

As far as the variable ϕ is concerned, $\Phi_0(\phi,t)$ and $\Phi_1(\phi,t)$ are orthogonal. Furthermore, these two wave functions have no overlap in space at all, as one can see from (31) and (32). Therefore, the transition between the $n = 0$ state (s quark) and the $n = 1$ state (b quark) through the flavor interaction is forbidden. This property could be changed by other interactions. In a class of processes, the quantum number is conserved. Consequently, it appears reasonable to identify the conservation of $n = 0$ ($n = 1$) as the conservation of the strangeness (bottomness) quantum number, in the absence of weak interactions.

Of course, it is possible to change the d quark and the ring with $I_z = +\frac{1}{2}$ to the u quark and the ring with $I_z = -\frac{1}{2}$ through other interactions. For example, suppose a d-type quark is initially in a state n and suppose the weak interaction causes such a change during a very short time interval (e.g., in a collision process). In this case, the flavor potential $C_0 = -fr^{-1}\log(1/\lambda r)$ is suddenly changed into another flavor potential $C_0' = -fr^{-1}\log(1/\lambda' r)$. To find out the change of the energy of the system, we use sudden approximation in quantum mechanics[6]:

$$\left[\left(i\frac{\partial}{\partial t} - fC_0\right)^2 + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2}\right]\Phi(\phi,t) = 0, \quad t<0, \tag{33}$$

$$\left[\left(i\frac{\partial}{\partial t} - fC_0'\right)^2 + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2}\right]\Phi'(\phi,t) = 0, \quad t>0. \tag{34}$$

After the interaction ($t>0$), we solve (34) and obtain the following results: (i) the ring remains in the same state n; (ii) the energy of the system increases from E_n, given by (22), to $E_n' = f^2\lambda'\exp(|n|/f^2 - 1)$; and (iii) the wave function of the system changes from

$$\Phi_n = \exp(-iE_n t - in\phi)/(2\pi)^{1/2},$$

at

$$r_n = \lambda^{-1}\exp(1 - |n|/f^2),$$

to

$$\Phi_n' = \exp(-iE_n' t - in\phi)/(2\pi)^{1/2},$$

at

$$r_n' = \lambda'^{-1}\exp(1 - |n|/f^2).$$

The detailed mechanism of such a physical process is probably not simple and should be treated by introducing annihilation and creation operators for the rings, etc.

Note that ϕ is the only coordinate variable. The normalization of Φ_n can be formally written as

$$1 = \int_0^{2\pi} |\Phi_n(\phi,t)|^2\, d\phi \equiv \int |\varphi_n(r,\theta,\phi,t)|^2\, d^3r,$$

$$|\varphi_n(r,\theta,\phi,t)|^2 = \delta(r - r_n)\delta(\cos\theta)|F_n(\phi)|^2/r^2. \tag{35}$$

The new function $\varphi_n(r,\theta,\phi,t)$ explicitly exhibits the important property that $\Phi_n \neq 0$ only at $r = r_n$ and $\theta = \pi/2$. We have seen that the ring's flavor charge is distributed in space according to $\rho(\mathbf{r}) = f^2|\varphi_n(r,\theta,\phi,t)|^2$. It is interesting to note that such a ringlike distribution of the flavor charge leads to an angle-dependent form factor:

$$F_f(\mathbf{k}) = \int f^{-2}\rho(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}\, d^3r = J_0(r_n k\sin\theta_k),$$

$$\sin\theta_k = k_\parallel/k, \quad k_\parallel^2 = k_x^2 + k_y^2, \tag{36}$$

where $\mathbf{k}_\parallel$ is parallel to the ring's plane and $J_0(z)$ denotes the Bessel function of the first kind: $J_0(z) = 1 - z^2/4 + \ldots$, for $|z| \ll 1$. For a spherically symmetric charge density, a form factor $F(\mathbf{k})$ can be expanded as $F(\mathbf{k}) = 1 - |\mathbf{k}|^2\langle r^2\rangle/6 + \ldots$. Comparing (36) to this form factor $F(\mathbf{k})$, we may say that, for a given $\theta_k \equiv \theta$, the mean square radius $\langle r^2\rangle_\theta$ of the ring's flavor-charge distribution is $\langle r^2\rangle_\theta = \frac{1}{2}r_n^2\sin^2\theta$, $\theta = \theta_k$. Thus when we take averages over θ, we have the mean square radius:

$$\langle r^2\rangle \equiv \frac{1}{\pi}\int_0^\pi \langle r^2\rangle_\theta\, d\theta = \frac{3r_n^2}{4} \tag{37}$$

$$\left[\text{or } \langle F_f\rangle \equiv \pi^{-1}\int_0^\pi F_f(\mathbf{k})d\theta_k = J_0^2\left(\frac{r_n k}{2}\right) = 1 - \frac{k^2(3r_n^2/4)}{6} + \ldots\right],$$

where

$$r_0 = \lambda^{-1}\exp(1) = 3.6\times10^{-14}\text{ cm, for an } s \text{ quark},$$

$$r_1 = \lambda^{-1}\exp(1 - 1/f^2) = 1.1\times10^{-15}\text{ cm, for a } b \text{ quark}. \tag{38}$$

The result (37) should be observable by the scattering of the ringlike quark and a pointlike object carrying the flavor charge. However, the ringlike structure of quarks probably can also be detected experimentally by the scattering of two ringlike quarks. In this case, the observed form factor will be more complicated than that in (36). We shall not discuss this case here.

VI. COMPARISONS OF QUANTUM RINGS AND CLOSED NAMBU STRINGS

Let us examine the heuristic connection between the classical equation (2) and the quantum equation (14) for a ring. In a particular frame, Eq. (2) has a solution $X^\mu(\delta,\tau)$ that describes a rotating ring:

$$X^0 = Rn\xi, \quad X^1 = R\cos(n\xi), \quad X^2 = R\sin(n\xi), \quad X^3 = 0, \tag{39}$$

where $n = 1,2,\ldots$ and $\xi = (\sigma+\tau)/R$ or $(\sigma-\tau)/R$. This solution also satisfies the constraints $(\partial X^\mu/\partial\tau)(\partial X_\mu/\partial\sigma) = 0$ and $(\partial X^\mu/\partial\tau)^2 + (\partial X^\mu/\partial\sigma)^2 = 0$. The spatial components X^1 and X^2 satisfy the periodic condition $X^i(\sigma + 2\pi R,\tau) = X^i(\sigma,\tau)$; but physically it is not necessary for the time component X^0 to satisfy the condition. The total energy p^0 of such a ring is

$$p^0 = \frac{1}{4\pi\alpha'}\int_0^{2\pi R} d\sigma\frac{\partial X^0}{\partial\tau} = \frac{Rn}{2\alpha'} \quad (p^k = 0, \quad k = 1,2,3), \tag{40}$$

where α' is the "Regge slope" that appears in the Nambu action:

$$-\frac{1}{4\pi\alpha'}\int(\cdots)d\sigma\, d\tau.$$

Intuitively, if one introduces a central potential $V(R)$ with a source located at the center of the ring, we expect that the total energy (40) of the ring will be just shifted by $V(R)$. Nevertheless, even in such a simple case, it does not seem possible to implement $V(R)$ in the string equation (2). In sharp contrast, we can easily implement a potential in the quantum ring equation (14). We note that Eq. (14) with $V(r) = m = 0$ can be obtained formally from Eq. (2) by the replacements

$$X^\mu(\sigma,\tau) \to \Phi^\mu(\phi,t), \quad \sigma \to (r\sin\theta)\phi, \quad \tau \to t. \tag{41}$$

It is interesting to observe that the resultant equation, which is formally obtained through (41), turns out to be the same as that obtained by passing from the classical theory with the Hamiltonian (8) to the quantum theory, as discussed in Sec. III. It should be stressed that we have followed the conventional and well-established method for passing to the quantum theory. Although the free massless ring equation (11) [or (14)] with $V=m=0$ is the same as the Nambu equation (2) with $\sigma=(r\sin\theta)\phi$, it has different solutions and different physical properties: The free closed Nambu string described by (2) has a quantized energy $\varepsilon_n = n$ (in a suitable unit). The discrete energy ε_n is kinematically determined by the boundary condition $X^\mu(0,\tau)=X^\mu(2\pi R,\tau)$, where the constancy of $R=r\sin\theta$ has been implicitly assumed in choosing a suitable unit of length for $\sigma=r(\sin\theta)\phi$.

On the other hand, if one considers a free quantum ring described by (11) [or (14)] with $m=V=0$ and

$$\Phi(\phi,t)=\exp(-iEt+i\beta\phi).$$

We obtain

$$E=0, \quad r=\infty, \quad \beta=n. \tag{42}$$

These results for a free ring are the same as those of an atomic system in the limit of vanishing coupling constant. Thus we have seen that discrete energy states of a quantum ring cannot exist if there is no potential. The reason for this is that, in our approach based on a Hamiltonian with cyclic radial momentum p_r, the radius and the energy eigenvalues of a ring are determined by $\partial E/\partial r=0$, as one can see in (20) and (22).

VII. PERMANENT CONFINEMENT OF THE QUANTUM RING

The ring in the composite quarks is characterized by its ultrastable states with

$$0<E_0<E_1<E_2<\cdots,$$
$$\lambda^{-1}\sim r_0>r_1>r_2>\cdots, \tag{43}$$

even though the flavor potential vanishes as $r\to\infty$. (The effective potential at $r=r_n$ has a maximum rather than a minimum). The positive and discrete energy spectrum (43) resembles that of the confining particle in the simple harmonic oscillator or in a linear potential field.[7] When one hits the composite quarks with a particle, however large the impact energy may be, the system will not collapse and break into pieces. (If the interaction of collision can change the state of a composite quark, the ring will jump to a higher energy state rather than become a free string.)

Furthermore, the bosonic ring moving in the flavor potential (26) has unusual behavior in the limit $f\to 0$. In this limit, r_n and E_n in (20) and (22) become

$$r_n\to\begin{cases}\lambda^{-1}\exp(1), \\ 0,\end{cases} \quad E_n\to\begin{cases}0, & n=0, \\ \infty, & n\neq 0,\end{cases} \tag{44}$$

which are completely different from those results (42) for a free ring. They indicate that once the ring exists in the stationary state, it cannot become a free ring even if the flavor charge f approaches zero.

All these properties indicate that the quantum ring is permanently confined by the basic quarks within a small radius $r\lesssim\lambda^{-1}\approx 10^{-14}$ cm.

VIII. DISCUSSIONS AND REMARKS

The usual free-string equation (2) with the closed string boundary conditions gives the well-known normal-mode expansion. It also leads to many particles with different spins. However, the free quantum ring, described by Eq. (3) with $C_0(r)=0$, does not have these results. According to our Hamiltonian approach with cyclic p_r and p_θ, there is simply no stationary state with discrete energy, as shown in (42).

The flavor potential $C_0(r)=V_1$ in (15) can be expressed in terms of an integration in momentum space:

$$C_0(r)=4\pi f\int\frac{d^3k}{(2\pi)^3}\frac{e^{ik\cdot r}}{k^2}\left(-\frac{\pi}{2}\right)\log\left(\frac{e^\gamma k}{\lambda}\right)$$
$$=fr^{-1}\log(\lambda r),$$

where $\gamma=0.577\,21$ is the Euler constant. The unusual logarithmic factor in $C_0(r)$ is probably related to higher-order corrections, which are complicated and puzzling beyond measure from the field-theoretic viewpoint.

If one takes the negative square root in (18), one gets

$$r_n=\lambda^{-1}\exp(1+|n|/f^2)$$

and

$$E_n=f^2\lambda\exp(-|n|/f^2-1)$$

instead of (20) and (22). These states correspond to systems with a negative "centrifugal energy" for a ring and do not seem to exist in nature.

The quark masses can be fitted by different potentials. For example, one can modify the flavor potential (26) for $r\gg 1/\lambda'$ in such a way that the energy levels are essentially unchanged. Usually, one does not expect a potential model to work in a high energy region as shown in (24) and (28). However, the present ring model is very much different from all previous potential models, as one can see from the properties (a)–(d) in Sec. I. We stress that these special properties for an interacting quantum ring are physically more important than a particular flavor potential and mass formula.

We may remark that the definite radius for a ring, $\Delta r=0$, as shown in (31) or (38), does not contradict the uncertainty relation.[8] The reason is that the ring's radial momentum p_r is completely undefined, i.e., $\Delta p_r=\infty$, in the formalism. This is the physical meaning of the cyclic p_r in quantum mechanics.

In this simple model of composite quarks, we assume that only isosinglet states can exist. This resembles the situation of color singlet states in quantum chromodynamics. It is hoped that this property can be understood dynamically in the future.

ACKNOWLEDGMENTS

The author would like to thank Antal Javicki and colleagues at SMU for their discussions.

This work was supported in part by the University Research Committee of SMU and by the Polaroid Foundation, Cambridge, Massachusetts.

[1]Y. Nambu, Lecture at the Copenhagen Summer Symposium, 1970 and *Proceedings of International Conference on Symmetries and Quark Model*, Wayne State University, 1969 (Gordon and Breach, New York, 1970), p. 269; T. Goto, Prog. Theor. Phys. **46**, 1560 (1971).
[2]A. Jevicki, in *Proceedings of the XXIII International Conference on High Energy Physics*, (World Scientific, Singapore, 1987), Vol. 1, p. 351; D. J. Gross and A. Jevicki, Nucl. Phys. B **283**, 279 (1981); **287**, 255 (1987).
[3]H. Goldstein, *Classical Mechanics*, (Addison–Wesley, Reading, MA, 1959), pp. 218–220.
[4]J. P. Hsu, Phys. Lett. A **97**, 137 (1983); Nature Editorial, Nature (London) **303**, 129 (1983). In such a framework, the four-dimensional transformation with a common time t for all observers is given by $b't = \gamma(ct - \beta x)$, $x' = \gamma(x - \beta ct)$, $y' = y$, $z' = z$, $\beta = V/c$, $\gamma = (1 - \beta^2)^{-1/2}$. It has advantages for treating statistical and composite systems. [See J. P. Hsu, Nuovo Cimento B **80**, 201 (1984).] In this four-dimensional framework, Eqs. (4), (12), and (14) are unchanged in the F frame in which the center of mass of the ring is at rest.
[5]J. Gasser and H. Leutwyler, Phys. Rep. **87**, 77 (1982); M. Althoff *et al.*, Phys. Lett. B **138**, 441 (1984).
[6]L. L. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), p. 213.
[7]J. P. Hsu, Phys. Rev. D. **24**, 802 (1981).
[8]It is possible to have a "fuzzy quantum ring" with Δr very small but not zero. For a discussion of fuzzy corrdinate variables within the framework of quantum mechanics, see J. P. Hsu and S. Y. Pei, Phys. Rev. A **37**, 1406 (1988); J. P. Hsu and C. Whan, *ibid.* **38**, 2248 (1988).

Signals in nonlinear electrodynamics invariant under duality rotations

Humberto Salazar Ibarguen
Escuela de Ciencias Físico Matemáticas, Universidad Autónoma de Puebla, Apartado Postal 1364, Puebla, Mexico

A. García and J. Plebanski
CINVESTAV, Apartado Postal 14-740, 07000 Mexico, D.F., Mexico

(Received 3 May 1989; accepted for publication 12 July 1989)

The signal propagation in nonlinear electrodynamics when an arbitrary Einstein–Born–Infeld theory is invariant under duality rotations is discussed. The quasimetric defining the characteristic surfaces that depend on the structural function is obtained. The propagation of nonlinear photons is also briefly discussed.

I. INTRODUCTION

Recently the ideas of nonlinear electrodynamics (NLE) have been found of interest, even in the strings and superstrings theories. Born–Infeld (BI) type actions were already discussed in connection with string theory in Refs. 1 and 2. Fradkin and Tseytein found that the BI Lagrangian is the exact solution of a constant external vector field problem in the open string theory.[3] Bergshoeff *et al.*[4] found that the bosonic part of the low-energy open superstring effective action is again the BI action. Supersymmetric BI action was established by Cerotti and Ferrara.[5]

These new approaches reinforce the importance of the study of classical BI-type theories coupled with gravitational fields. Recently, by restricting these theories to be invariant under duality rotations, it was possible to get a wide branch of new exact solutions.[6]

On the other hand, assuming Einstein–BI dynamics as a classical model of the vacuum polarization processes near strong gravitational fields,[7] it is interesting to study the nature of causal signals in this theory. The content of this paper is the study of the characteristic surfaces, along which discontinuities of the nonlinear electromagnetic field are propagated, when we impose the restriction to the theory to be invariant under duality rotations.

In Sec. II, we give a brief introduction to the Einstein–BI theory and the consequences of the duality rotations invariance. In Sec. III, we develop the corresponding simplifications on the theory of discontinuities in general relativistic nonlinear electrodynamics[8] when the duality rotation invariance is assumed. The special case corresponding to the original BI theory is discussed with some details. Concluding remarks are given in Sec. IV.

II. NL ELECTRODYNAMICS ENDOWED WITH THE FREEDOM OF DUALITY ROTATIONS

The nonlinear electrodynamic field is represented by two skew field tensors $f_{\mu\nu}$ and $P_{\mu\nu}$, which are interrelated through a single relation designated as the "structure" equation. The existence of a potential A_μ satisfying the Faraday field equation

$$f_{\mu\nu} = A_{\mu;\nu} - A_{\nu;\mu} \tag{2.1}$$

is assumed, yielding the electrodynamic Lagrangian

$$L_E = -(1/4\pi)\left[\tfrac{1}{2} f_{\mu\nu}P^{\mu\nu} - \mathscr{H}(P,\check{Q})\right]. \tag{2.2}$$

The "structure function" $\mathscr{H} = \mathscr{H}(P,Q)$, whose arguments

$$P := \tfrac{1}{4} P_{\mu\nu}P^{\mu\nu} \quad \text{and} \quad \check{Q} := \tfrac{1}{4} P_{\mu\nu}\check{P}^{\mu\nu}$$

are invariant (scalar) and pseudoinvariant, respectively, is a real Hamiltonian whose functional form is intentionally left unspecified.

The essential distinction between linear and nonlinear electrodynamics resides in the electromagnetic structure equations,

$$f_{\mu\nu} = 2\frac{\partial \mathscr{H}}{\partial P^{\mu\nu}} = \mathscr{H}_P P^{\mu\nu} + \mathscr{H}_{\check{Q}}\check{P}^{\mu\nu}. \tag{2.3}$$

When these equations can be inverted one may algebraically express $P_{\mu\nu}$ through $f_{\mu\nu}$, its dual, and the invariants

$$F := \tfrac{1}{4} f_{\mu\nu}f^{\mu\nu} \quad \text{and} \quad \check{G} := \tfrac{1}{4} f_{\mu\nu}\check{f}^{\mu\nu}$$

by the expression

$$P_{\mu\nu} = 2\frac{\partial L}{\partial f^{\mu\nu}} = L_F f_{\mu\nu} + L_{\check{G}}\check{f}_{\mu\nu}. \tag{2.4}$$

Hence $L = L(F,\check{G})$ is understood as a function of $f_{\mu\nu}$ depending on it through the arguments F and $\check{G}$.

The Hamiltonian function $\mathscr{H}$ and the Lagrangian function L are related by a Legendre transformation,

$$L = L(F,\check{G}) = 2P\mathscr{H}_P + 2\check{Q}\mathscr{H}_Q - \mathscr{H}. \tag{2.5}$$

We can introduce new independent parameters of the complex invariants and therefore invariants as such by the relations

$$P + Q = -\tfrac{1}{2}(D + i\check{H})^2 \neq 0,$$
$$F + \check{G} = -\tfrac{1}{2}(E + i\check{B})^2 \neq 0; \tag{2.6}$$

the interpretation of these parameters in an adequate null tetrad is discussed in Ref. 9.

From Eqs. (2.6) we can consider the Hamiltonian function $\mathscr{H}$ as a function of D and H and the Lagrangian function L as a function of E and B. Defining the function

$$M = ED + \mathscr{H}, \tag{2.7}$$

we have

$$dM = E\,dD + H\,dB; \tag{2.8}$$

then

 0022-2488/89/112689-03$02.50

$$M = M(D,B)\,. \tag{2.9}$$

The theory is equivalently uniquely defined with any of the structural functions $\mathscr{H}$,L of M.

With the freedom of the duality rotations defined by the condition that "given θ^a,$D + i\breve{B}$ and $E + i\breve{H}$, which satisfy the Einstein–Maxwell equations, then

$$\begin{aligned} &\theta^a, D' + i\breve{B}' := e^{i\phi_0}(D + i\breve{B})\,, \\ &E' + i\breve{H}' := e^{i\phi_0}(E + i\breve{H}) \end{aligned} \tag{2.10}$$

are also a solution of the Einstein–Maxwell equations for arbitrary $\phi_0 = \text{const}$," the function M is constrained to a function of one variable $\frac{1}{2}(D^2 + \breve{B}^2)$.

It is convenient to understand M as determined by an arbitrary dimensionless function of a dimensionless variable $f(x)$ in the form

$$M := b^2 f(x), \quad x := (1/2b^2)(D^2 + \breve{B}^2)\,, \tag{2.11}$$

where b is a constant of dimension of the electromagnetic field. An extended version of this section can be found in Ref. 6.

III. SIGNALS IN NONLINEAR ELECTRODYNAMICS ENDOWED WITH THE FREEDOM OF DUALITY ROTATIONS

According to the general theory developed in detail in Ref. 8, the discontinuities of the derivatives of the nonlinear electromagnetic field propagate along the two possible sets of the characteristic surfaces $S \equiv S(x) = 0$, these being submitted to the eikonal equation

$$\gamma^{\mu\nu} S_{,\mu} S_{,\nu} = 0\,, \tag{3.1}$$

where the quasimetrics $\gamma^{\mu\nu}$—determined modulo proportionality factor $\phi^{-2} \neq 0$—are, correspondingly,

$$\begin{aligned} \text{I:}\ \gamma^{\mu\nu} &= \phi^{-2}\{g^{\mu\nu} + [(\mathscr{H}_{\breve{H}\breve{H}} - \tau)/(\mathscr{H}_{\breve{H}\breve{H}} + \tau)] \\ &\quad \cdot \tau^{\mu\nu}/|Z|\}\,, \\ \text{II:}\ \gamma^{\mu\nu} &= \phi^{-2}\{g^{\mu\nu} + [(\mathscr{H}_{DD} + \tau)/(\tau - \mathscr{H}_{DD})] \\ &\quad \cdot \tau^{\mu\nu}/|Z|\}\,, \\ |Z| &:= \tfrac{1}{2}(D^2 + \breve{H}^2), \quad \tau_{\mu\nu} := -P_{\mu\rho}P_{\nu}{}^{\rho} + P g_{\mu\nu}\,. \end{aligned} \tag{3.2}$$

This description of the characteristic surfaces applies when one works with the structural function $\mathscr{H} = \mathscr{H}(D,\breve{H})$ considered as given as fundamental [recall that $P + Q = -\frac{1}{2}(D + i\breve{H})^2$].

The remaining structural element of (3.2)—apart from the Riemmanian metric $g_{\mu\nu}$—the invariant τ, is defined by

$$\begin{aligned} \delta &:= \{(\mathscr{H}_{D\breve{H}} + i\mathscr{H}_{Q})^2 + (\mathscr{H}_{P})^2 - \mathscr{H}_{DD}\mathscr{H}_{\breve{H}\breve{H}}\}\,, \\ k &:= \delta^2 + 4(\mathscr{H}_{P})^2 \mathscr{H}_{DD}\mathscr{H}_{\breve{H}\breve{H}}, \\ \tau &= (1/2k_P)(\sqrt{k} + \delta)\,. \end{aligned} \tag{3.3}$$

In the specific case of the theory endowed with the freedom of the duality rotations, deciding to work with the structural function M equivalent to $\mathscr{H}$, we find that the description of characteristic surfaces given above considerably simplifies.

In the first step, omitting manipulations related to the replacement of the independent electromagnetic variables $(D,\breve{H})$ by the equivalent $(D,\breve{B})$, we state that δ and 1 can be expressed in terms of f in the forms

$$\begin{aligned} \delta &= 1/M_{\breve{B}\breve{B}} \cdot [(D^2 + \breve{B}^2)/(D^2 + \breve{H}^2)] \\ &\quad \times \{2\times f^{\nabla\nabla} + f^{\nabla} + (f^{\nabla})^2\}\,, \\ k &= \nu^2\,, \\ \nu &:= 1/M_{\breve{B}\breve{B}} \cdot (D^2 + \breve{B}^2)/(D^2 + \breve{H}^2) \\ &\quad \cdot \{2\times f^{\nabla\nabla} + f^{\nabla} - (f^{\nabla})^2\}\,, \end{aligned} \tag{3.4}$$

while $\mathscr{H}_{HH} = 1/M_{\breve{B}\breve{B}}$. We must now distinguish two subcases,

$$\begin{aligned} &\text{(a)}\ \ \nu < 0 \to \tau = (f^{\nabla})^2/M_{\breve{B}\breve{B}}\,, \\ &\text{(b)}\ \ \nu > 0 \to \tau = 1/2M_{\breve{B}\breve{B}} \cdot 1/(f^{\nabla})^2 \\ &\qquad\qquad \cdot \{2\times f^{\nabla} f^{\nabla\nabla} + f^{\nabla 2}\}\,. \end{aligned} \tag{3.5}$$

By substituting these values of τ into (3.2) one obtains the final expressions for the quasimetrics $\gamma_{\mu\nu}$. It turns out that the case (a) evaluated in I overlaps with the case (b) evaluated in II, the final result amounting to

$$\begin{aligned} \text{I:}\ \gamma_{\mu\nu} &= \phi^{-2}\{g_{\mu\nu} + [1 - (f^{\nabla})^2]/[1 + (f^{\nabla})^2] \\ &\quad \cdot \tau_{\mu\nu}/|Z|\}\,, \\ \text{II:}\ \gamma_{\mu\nu} &= \phi^{-2}\{g_{\mu\nu} - x f^{\nabla\nabla}/(f^{\nabla} + x f^{\nabla\nabla}) \cdot \tau_{\mu\nu}/|Z|\}\,. \end{aligned} \tag{3.6}$$

These formulas are general: they apply also in the sense of a limiting transition when the electromagnetic field is null, i.e., when with $x \to 0$, $\tau_{\mu\nu}/|Z| \to k_\mu k_\nu$, $k^\mu k_\mu = 0$.

According to the results of Alarcon, the single structural function for which the quasimetrics (3.2) coincide, is that which is equivalent to the original Born–Infeld Lagrangian. This can be verified in the considered special case of the theory endowed with the freedom of the duality rotations by noticing that I–II quasimetrics coincide iff

$$(1 - f^{\nabla 2})/(1 + f^{\nabla 2}) + x f^{\nabla\nabla}/(f^{\nabla} + x f^{\nabla\nabla}) = 0\,. \tag{3.7}$$

This condition leads to f,

$$f = (2/\alpha)\{\sqrt{1 + \alpha x} - 1\}\,, \tag{3.8}$$

where a is an integration constant, which, because[6] of $f/x > f^{\nabla}$, must be positive. Thus

$$M = (2b^2/a)\{\sqrt{1 + (a/2b^2)(D^2 + B^2)} - 1\}\,, \tag{3.9}$$

so that renormalizing b according to $2b^2/a$ we end up precisely with M equivalent to the Borne–Infeld Lagrangian[6]

$$M = b^2\{\sqrt{1 + 2x} - 1\}\,.$$

Consider then the case of algebraically general electromagnetic field $P_{\mu\nu}$ described in terms of a null tetrad aligned along the eigenvectors of the field. The only nontrivial components of

$$\tau_{ab}/|Z| \quad \text{are} \quad \tau_{12}/|Z| = -1, \quad \tau_{34}/|Z| = 1\,. \tag{3.10}$$

Equation (3.1), spelled out in terms of the null tetrad, is of course $\gamma^{ab} S_{,a} S_{,b} = 0$, where a denotes the directional deriv-

ative. We then easily see that in the case I from (3.6) this equation amounts to

$$\text{I: } 2(f^{\nabla})^2 S_{,1}S_{,2} + 2S, \quad S_{,4} = 0 ,$$

while in the case II we have, assuming $d(xf^{\nabla}) \neq 0$,

$$\text{II: } 2(1 + 2x(f^{\nabla\nabla}/f^{\nabla}))S_{,1}S_{,2} + 2S_{,3}S_{,4} = 0 . \quad (3.11)$$

Now, the bicharacteristic lines $x^{\mu} = x^{\mu}(\tau)$, which are the integrals of

$$\gamma_{\mu\nu}{}^{(x)} \frac{dx^{\nu}}{d\tau} = S_{,\mu}(x) ,$$

in a sense "the trajectories of the nonlinear photons"—can be obtained by treating correspondingly (3.11) as the Hamilton–Jacobi equation. These lines are null geodesic in the sense of quasimetrics $\gamma_{\mu\nu}$, which, if taken as normalized as in (3.11), evidently correspond to the Riemannian line elements

$$\tilde{ds}^2 := \begin{cases} \text{I: } 2(f^{\nabla})^{-2}\theta^1 \underset{s}{\otimes} \theta^2 + 2\theta^3 \underset{s}{\otimes} \theta^4 ; \\ \text{II: } 2(1 + 2x(f^{\nabla\nabla}/f^{\nabla}))^{-1}\theta^1 \underset{s}{\otimes} \theta^2 + 2\theta^3 \underset{s}{\otimes} \theta^4 ; \end{cases} \quad (3.12)$$

it follows that the basic Riemannian line element

$$ds^2 = 2\theta^1 \underset{s}{\otimes} \theta^2 + 2\theta^3 \underset{s}{\otimes} \theta^4$$

is evaluated along these lines, because then $\tilde{ds}^2 = 0$ amounts to

$$ds^2 = \begin{cases} \text{I: } 2[1 - (f^{\nabla})^{-2}]\theta^1 \underset{s}{\otimes} \theta^2 , \\ \text{II: } 2[1 - (1 + 2x(f^{\nabla\nabla}/f^{\nabla}))^{-1}]\theta^1 \underset{s}{\otimes} \theta^2 . \end{cases} \quad (3.13)$$

With $\theta^1 \otimes \theta^2$ being positive, it is thus the sign of the coefficients in (3.13) that determines whether the corresponding bicharacteristic lines are spacelike $ds^2 > 0$ or timelike $ds^2 < 0$, i.e., they can be interpreted as corresponding to the "nonlinear photons" propagating with velocities faster or slower than light.

Notice that in the case of the original Born–Infeld structural function $f = \sqrt{1 + 2x} - 1$ and

$$\text{II} \to \text{I}, \quad 1 - (f^{\nabla})^2 = -2x < 0 \to ds^2 \leqslant 0 ,$$

and therefore the nonlinear light cone is contained within the Riemannian light cone bicharacteristic lines are timelike or null in the standard sense.

With a general structural function, with $f = f(x)$ constrained only by

$$f = x + 0[(x)^2], \quad f/x > f^{\nabla} > 0 ,$$

the spacelike or timelike nature of the bicharacteristic line depends on the specific form of this function.

We will close this section with a general comment concerned with the relationship between the causal horizons—if they occur—of the $g_{\mu\nu}$ metric and the quasimetrics $\gamma_{\mu\nu}$.

Working with a general structural function $\mathscr{H} = \mathscr{H}(D,H)$, i.e., not necessarily committed to the case of a theory endowed with the freedom of the duality rotations, in the case of the algebraically general electromagnetic field, and employing the null tetrad aligned along the eigenvectors of the field, one easily sees that the $\gamma^{\mu\nu}S_{,\mu}S_{,\nu} = 0$ equations with $\gamma^{\mu\nu}$ from (3.2) is equivalent to

$$\begin{aligned} &\text{I: } 2(\tau/\mathscr{H}_{\check{H}\check{H}})S_{,1}S_{,2} + 2S_{,3}S_{,4} = 0 , \\ &\text{II: } -2(\mathscr{H}_{DD}/\tau)S_{,1}S_{,2} + 2S_{,3}S_{,4} = 0 , \end{aligned} \quad (3.14)$$

where the invariant τ is defined by (3.3). The quasiline element $\tilde{ds}^2$ is then

$$\tilde{ds}^2 = \begin{cases} \text{I: } 2(\mathscr{H}_{HH}/\tau)\theta^1 \underset{s}{\otimes} \theta^2 + 2\theta^3 \underset{s}{\otimes} \theta^4 , \\ \text{II: } -2(\tau/\mathscr{H}_{DD})\theta^1 \underset{s}{\otimes} \theta^2 + 2\theta^3 \underset{s}{\otimes} \theta^4 . \end{cases} \quad (3.15)$$

Comparing this with the Riemannian metric $ds^2 = 2\theta^1 \otimes_s \theta^2 + 2\theta^3 \otimes_s \theta^4$, we see that the relevant segment of the metric, $2\theta^3 \otimes_s \theta^4$, i.e., that segment of signature $(+,-)$ essential for the reversal of signs $[(+,-) \to (-,+)]$ on both sides of a horizon, is the same for the metric ds^2 and quasimetrics $\tilde{ds}^2$. From this we now infer that the causal horizons—if they occur—of ds^2 and $\tilde{ds}^2$ do coincide. This generalizes a result of S. Alarcón[10] derived this for the case of the spherically symmetric solutions.

IV. CONCLUSIONS

The study of the characteristic surfaces, along which discontinuities of the nonlinear electromagnetic field are propagated in the specific case of the theory endowed with the freedom of the duality rotations is considerably simplified. The quasimetric defining the characteristic surfaces that depends on the structural function is given by Eq. (3.6). The trajectories of the nonlinear photons can be obtained by treating Eq. (3.11) as the Hamilton–Jacobi equation. Nonlinear photons can propagate with velocities faster or slower than light.

[1] H. C. Tze, Lett. Nuovo Cimento **11**, 401 (1973).
[2] H. C. Tze, Nuovo Cimento A **22**, 507 (1974).
[3] E. S. Fradkin and A. A. Tseytlin, Phys. Lett. B **163**, 123 (1985).
[4] E. Bergshoeff, E. Sezgin, C. N. Pope, and P. K. Townsend, Phys. Lett. B **188**, 70 (1987).
[5] S. 'Cecotti and S. Ferrara, Phys. Lett. B **187**, 335 (1987).
[6] H. Salazar I., A. Garcia, and J. F. Plebañski, J. Math. Phys. **28**, 2171 (1987).
[7] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[8] S. Alarcon G., A. L. Dudley, and J. F. Plebañski, J. Math. Phys. **22**, 2835 (1981).
[9] J. F. Plebañski, *Spinors, Tetrads and Forms*, A Monograph of CINVESTAV del IPN, 1975.
[10] S. Alarcon G., Ph.D. Thesis, CINVESTAV del IPN, 1979.

On some nonunitary representations of the Poincaré group and their use for the construction of free quantum fields

G Cassinelli, G. Olivieri, and P. Truini
Departimento di Fisica, Universitá di Genova, Istituto Nazionale di Fisica Nucleare, Sezione di Genova via Dodecaneso 33, 16146 Genova, Italy

V. S. Varadarajan
Department of Mathematics, University of California, Los Angeles, California 90024

(Received 11 January 1989; accepted for publication 7 June 1989)

In this paper a class of nonunitary infinite dimensional Hilbert space representations of a semidirect product is investigated. The equivalence of this category with the category of finite dimensional representations of the stability subgroups is shown. This theory is applied to the Poincaré group and to the construction of free quantum fields. In an appendix a method is introduced for building an infinite family of finite dimensional indecomposable representations of the noncompact Euclidean group in two dimensions. Such representations are used for carrying out the analysis of the massless fields.

I. INTRODUCTION

This paper represents the beginning of a detailed study of not necessarily unitary representations of the Poincaré group and their applications to problems of quantum field theory. In Sec. II we introduce the category $\mathscr{P}$ of Hilbert space representations of certain semidirect products $G = A \times' H$ of which the Poincaré group is a typical example. One of our fundamental results (Sec. III) is that this category is naturally equivalent to the category of finite dimensional representations of the appropriate stability subgroup H_0. This result permits us to reduce various problems concerning the infinite dimensional representations in $\mathscr{P}$ to corresponding problems involving the finite dimensional representations of H_0. While this principle is not new, going back as it does to Frobenius and Mackey in the category of *unitary* representations, the construction of a *Hilbertian functor* in the nonunitary context is believed to be new.

In order to apply these results to quantum field theory one should describe (cf. Sec. VI) all intertwining operators from the Schwartz space of classical fields to the representations in $\mathscr{P}$. We do this in Sec. VII. In the appendices we discuss various mathematical techniques necessary for applying the results of Secs. III and VII to physical problems.

The need for working with nonunitary categories is already clear from the Gupta–Bleuler theory of quantum electrodynamics, and the category $\mathscr{P}$ is a very conservative extension of the usual category of unitary representations of the Poincaré group with finite spin (or helicity). If one does not insist on working in Hilbert spaces, one can replace, at least when A, H, and G are Lie groups, the category $\mathscr{P}$ by the category of smooth vector bundles on H/H_0, while G is acting on the spaces of sections in the usual manner. In particular, the elements a of the group act as multiplication by $e^{i\langle a,\hat{a}\rangle}$. To obtain more extensive categories it is thus essential to give up this requirement for the action of A. Interesting examples of such more general representations of the Poincaré group have been constructed by Rideau[1] and studied further by Carey and Hurst,[2] Mintchev,[3] Pierotti[4] (see also Araki.[5]) We shall postpone to another paper a systematic study of this enlargement of $\mathscr{P}$.

The definition of the morphisms between objects of $\mathscr{P}$ also perhaps needs a comment. The point is that nonunitary representations allow, in general, *unbounded* operators intertwining them and so, due to the obvious complications involving the domains of these operators, care is needed in defining the notion of morphism. Our definition, inasmuch as it works with states of finite extent in momentum space, is very natural from the physical point of view.

In this paper we shall use freely the basic facts of the theory of representations of locally compact groups (in particular, of semidirect products) in Hilbert spaces. For this we refer to the book of Varadarajan.[6]

II. THE CATEGORY $\mathscr{P}$ OF PHYSICAL REPRESENTATIONS

We consider a semidirect product $G = A \times' H$ where A, H are lcsc groups, with A closed, Abelian, normal and H acting on A by $h, a \mapsto h[a]$. Here $\hat{A}$ is the dual group of A and H acts naturally $(h, \hat{a} \mapsto h[\hat{a}])$ on it. We fix $\hat{a}_0 \in \hat{A}$ and consider representations of G that are not necessarily irreducible but associated to the H orbit of $\hat{a}_0$. We put $X = H[\hat{a}_0]$, we denote by H_0 the stabilizer of $\hat{a}_0$ in H, and assume that X is locally closed in $\hat{A}$, so that $hH_0 \mapsto h[\hat{a}_0]$ is a homeomorphism of H/H_0 with X. We also assume that X has an invariant Borel measure α.

Definition 2.1: $\mathscr{P}$ is the class of pairs $(U, \mathscr{H})$ where

(a) $\mathscr{H}$ is a Hilbert space;

(b) U is a (strongly) continuous representation of G in $\mathscr{H}$;

(c) U_a is unitary for all $a \in A$;

(d) if P is the projection valued measure associated to $(U_a)_{a\in A}$ on $\hat{A}$, then $\mathrm{supp}(P) \subset X$;

(e) P has a uniform multiplicity $N < \infty$.

We call $(U, \mathscr{H})$ *unitary* if U is a unitary representation of G.

 0022-2488/89/112692-16$02.50

P is related to $(U_a)_{a\in A}$ by

$$U_a = \int_X e^{i\langle a,\hat{a}\rangle}\, dP(\hat{a}) \quad (a\in A).$$

The relation

$$U_h U_a U_h^{-1} = U_{h[a]} \quad (h\in H, a\in A)$$

leads to

$$U_h P_E U_h^{-1} = P_{h[E]} \quad (E\subset X). \tag{2.1}$$

Relation (2.1) shows that if $S(E)$ is the range of P_E, then U_h takes $S(E)$ to $S(h[E])$. Let

$$\mathscr{H}^0 = \cup\{S(C):\ C\subset X \text{ and relatively compact}\}.$$

Then $\mathscr{H}^0$ is a dense linear subspace of $\mathscr{H}$. Elements of $\mathscr{H}^0$ are called *finite*. It is obvious that

$$U_{a,h}[\mathscr{H}^0] = \mathscr{H}^0, \quad U_a[S(C)] = S(C).$$

Let $(U^i,\mathscr{H}^i)$ $(i=1,2)\in\mathscr{P}$. We write $\mathscr{L}((\mathscr{H}^1)^0;(\mathscr{H}^2)^0)$ for the space of linear maps

$$T:(\mathscr{H}^1)^0\to(\mathscr{H}^2)^0$$

such that for each relatively compact $C\subset X$, T maps $S^1(C)$ into $S^2(C)$ and defines a bounded operator on $S^1(C)$ [the norm of T on $S^1(C)$ may be unbounded as C varies]. It is easy to see that T has a closure and an adjoint T^* which is defined on $(\mathscr{H}^2)^0$ and defines a member of $\mathscr{L}((\mathscr{H}^2)^0;(\mathscr{H}^1)^0)$.

We make $\mathscr{P}$ into a category with the following definition.

Definition 2.2: If $(U^1,\mathscr{H}^1)$ and $(U^2,\mathscr{H}^2)$ are in $\mathscr{P}$, the morphisms $(U^1,\mathscr{H}^1)\to(U^2,\mathscr{H}^2)$ are precisely all the maps $T\in\mathscr{L}((\mathscr{H}^1)^0\to(\mathscr{H}^2)^0)$ such that (a) $T\in\mathscr{L}((\mathscr{H}^1)^0;(\mathscr{H}^2)^0)$; (b) T commutes with U_h^i, i.e., $TU_h^1 = U_h^2 T (h\in H)$. We say that $(U^1,\mathscr{H}^1)$ and $(U^2,\mathscr{H}^2)$ are isomorphic in $\mathscr{P}$ if there exists between them an isomorphism in the sense of the category $\mathscr{P}$.

We denote by $\mathrm{Morph}(U)$ the algebra of morphisms $(U,\mathscr{H})\to(U,\mathscr{H})$. A morphism is called *strict* if it is a bounded operator; in this case it extends to T $(\mathscr{H}^1\to\mathscr{H}^2)$ that intertwines U^1 and U^2. The set of strict elements of $\mathrm{Morph}(U)$ is the commuting algebra of the representation U in the usual sense.

It is easy to show that every morphism between unitary elements of $\mathscr{P}$ is strict. So our theory extends the standard one[6] of unitary representations of G. In particular, if $(U,\mathscr{H})$ is unitary, $\mathrm{Morph}(U)$ coincides with the commuting algebra of U in the usual sense.

III. THE INDUCED REPRESENTATION AS A FUNCTOR $(u,W)\mapsto(U,\mathscr{H})$

For any lcsc group L let $\mathscr{F}=\mathscr{F}(L)$ be the category of pairs (u,W), where W is a finite dimensional Hilbert space and u a continuous representation (not necessarily unitary) of L on W. We write $\mathscr{F}$ for $\mathscr{F}(H_0)$. Fix once and for all a Borel section $s(X\to H)$ such that $s(\hat{a}_0)=e_H$ and s is locally bounded on X, i.e., for any relatively compact $C\subset X$, $s(C)$ is relatively compact in H. Then if we put

$$\gamma(h:\hat{a}) = s(h[\hat{a}])^{-1}hs(\hat{a}),$$

it is clear that γ is a strict H_0-valued cocycle for (H,X). We define $C_u = u\circ\gamma$ by

$$C_u(h:\hat{a}) = u(s(h[\hat{a}])^{-1}hs(\hat{a}))$$

so that C_u is a strict $\mathrm{GL}(W)$ cocycle for (H,X).

Definition 3.1: A family of (positive definite) scalar products for W, $(\cdot,\cdot)_{\hat{a}}$ $(\hat{a}\in X)$, is said to be regular for u if (a) $(\cdot,\cdot)_{\hat{a}}$ is Borel, i.e., $(\theta_1,\theta_2)_{\hat{a}}$ is Borel in $\hat{a}$ $\forall\theta_1,\theta_2\in W$; (b) for any relatively compact $C\subset X$, there exist $k_1=k_1(C)>0$, $k_2=k_2(C)>0$, such that

$$k_1(\theta,\theta)_W \leqslant (\theta,\theta)_{\hat{a}} \leqslant k_2(\theta,\theta)_W \quad (\hat{a}\in C)$$

$[(\cdot,\cdot)_W$ is the given scalar product on $W]$. (c) Let $\mathscr{H}(W)$ be the Hilbert space of all (α-equivalence classes of) Borel functions f on X with values in W, such that

$$\int_X \|f(\hat{a})\|_{\hat{a}}^2\, d\alpha(\hat{a}) < \infty;$$

then the action

$$(a,h)\mapsto U_{ah}$$

$$(U_{ah}f)(\hat{a}) = e^{i\langle a,\hat{a}\rangle}C_u(h:h^{-1}[\hat{a}])\, f^h(\hat{a})$$

defines a continuous representation of G in $\mathscr{H}(W)$.

Note that $P_E(f\mapsto\chi_E f)$ are projections so that, writing $U(W)$ for the representation defined in (c), $(U(W),\mathscr{H}(W))\in\mathscr{P}$.

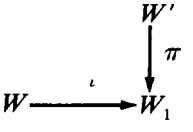
Condition (b) guarantees that $\mathscr{H}(W)^0 = L^2(X;W)^0$.

We shall now examine when a regular family of scalar products exists. It is clear that, in order to verify (c), it is enough to check that $h\mapsto U_{e_A h} = V_h$ is a continuous representation of H.

We introduce the following definition.

Definition 3.2: A pair $(u,W)\in\mathscr{F}$ is called *extendible* if there exist two pairs $(u_1,W_1)\in\mathscr{F}$, $(u',W')\in\mathscr{F}(H)$ such that (a) u is a subrepresentation of u_1, (b) u_1 is a quotient representation of $u_0' = u'|H_0$.

This means that we have a diagram



where ι is an injection and π a surjection and both ι and π are H_0-module maps, i.e., they commute with the action of H_0.

Theorem 3.1: If (U,W) is extendable, there exists a regular family of scalar products for u.

Proof: It is not a difficult exercise to see that, if $(\cdot,\cdot)_{W'}$ is a fixed scalar product for W', the family $(\cdot,\cdot)_{W',\hat{a}}$ defined by

$$(\theta_1,\theta_2)_{W',\hat{a}} = (u'(s(\hat{a}))\theta_1, u'(s(\hat{a}))\theta_2)_{W'}, \quad \theta_1,\theta_2\in W',$$

is a regular family for u_0'. From this the reduction to W is quite straightforward.

Extendability is an algebraic property. The following lemma is of wide applicability.

Lemma 3.1: Suppose H is a real algebraic matrix group and H_0 is an algebraic subgroup. Then all rational representations of H_0 are extendable (to rational representations of H).

Proof: Let R (resp. R_0) be the ring of regular functions on H (resp. H_0). Rationality of a representation is equivalent to the assertion that its matrix elements are in the ring of

regular functions. The group H (resp. H_0) acts on R (resp. R_0) by right translations and each element of R (resp. R_0) lies in a finite dimensional subspace stable under H (resp. H_0). Since the restriction map $R \to R_0$ is surjective and commutes with the action of H_0, it is immediate that any subrepresentation of R_0 (for H_0) is extendable. This property is then true for subrepresentations of $R_0^M = R_0 \oplus \cdots \oplus R_0$ (M copies). The proof is completed now with the well known observation that any rational representation of H_0 is a subrepresentation of R_0^M for some M. In fact, let r_0 be a rational representation of H_0 in a vector space W_0 of dimension M and let e_i^* ($1 \leqslant i \leqslant M$) be a basis for the dual of W_0. Let $\theta \in W_0$ and let $g_{\theta,i}(h) = \langle r_0(h)\theta, e_i^* \rangle$ ($h \in H$). Then $(g_{\theta,1},\ldots,g_{\theta,M}) \in R_0^M$ and

$$\theta \mapsto (g_{\theta,1},\ldots,g_{\theta,M})$$

is a map of W_0 into R_0^M that commutes with H_0. If $g_{\theta,i} = 0$, $\forall i$, $r_0(h)\theta = 0$, $\forall h$, so that $\theta = 0$. This is thus an embedding. The proof of the lemma is complete.

Let E be the group of all

$$\begin{pmatrix} z & a \\ 0 & z^{-1} \end{pmatrix} \quad (z,a \in \mathbf{C}, |z| = 1).$$

It is clear that SU(2) and E, the stabilizers of the orbits of the Poncaré group that we consider in the sequel, are real algebraic. For application to the Poincaré group we need the following Lemma.

Lemma 3.2: All continuous representations of SU(2) and E are rational.

Proof: For SU(2) this is by direct verification since the representations of SU(2) are explicitly realized in spaces of polynomials. We now take up E. We begin by recalling the well known fact that all irreducible representations of E are one dimensional and are given by

$$\begin{pmatrix} z & a \\ 0 & z^{-1} \end{pmatrix} \mapsto z^n \quad (n \in \mathbf{Z}).$$

Suppose now V carries an arbitrary representation and let $a = a_1 + ia_2$ act as $e^{a_1 L_1 + a_2 L_2}$ where L_1, L_2 elements of End V and commute with each other. Let

$$V = V_1 \supset V_2 \supset \cdots \supset V_r \supset 0$$

be a Jordan–Hölder series. Since V_i / V_{i+1} is irreducible, they are one dimensional and $L_1 = L_2 = 0$ on them. So $L_j(V_i) \subset V_{i+1}$ for all i and for $j = 1,2$, showing that the L_j are nilpotent. This means that $e^{a_1 L_1 + a_2 L_2}$ is a polynomial in a_1 and a_2, and the rationality of the representation is clear. The lemma is proved.

In view of the above discussion the assumption that for each $(u,W) \in \mathscr{F}$ there is a regular family of scalar products for u is seen to be a very reasonable one, satisfied in all cases of interest. We shall therefore suppose it to be true from now on. Then, for each $(u,W) \in \mathscr{F}$ we can associate a pair $(U(W), \mathscr{H}(W)) \in \mathscr{P}$; it is only necessary to select a regular family of scalar products. The following theorem is then our basic result; its proof is essentially similar to the corresponding result in the unitary context.

Theorem 3.2: The assignment

$$(u,W) \mapsto (U(W), \mathscr{H}(W))$$

is an exact covariant functor from the category $\mathscr{F}$ to the category $\mathscr{P}$ which is an equivalence of categories.

IV. COVARIANT SESQUILINEAR FORMS

We shall now discuss briefly the theory of covariant sesquilinear forms associated to $\mathscr{P}$.

Definition 4.1: Let $(U,\mathscr{H}) \in \mathscr{P}$; we say that $\langle \cdot,\cdot \rangle$ is a sesquilinear form for $(U,\mathscr{H})$ if (a) $\langle \cdot,\cdot \rangle$ is a sesquilinear form on $\mathscr{H}^0$; (b) $\langle \cdot,\cdot \rangle$ is invariant with respect to U_a for $a \in A$; (c) $\langle \cdot,\cdot \rangle$ is bounded on $S(C) \times S(C)$, for all $C \subset X$, relatively compact.

It is easy to see that these are precisely the forms defined by

$$\langle f,g \rangle = (f, \Gamma g), \quad f,g \in \mathscr{H}^0,$$

where Γ is a selfadjoint element of $\mathscr{L}(\mathscr{H}^0 : \mathscr{H}^0)$. The form $\langle \cdot,\cdot \rangle$ is called *covariant* if it is invariant with respect to U_h, for all $h \in H$; the condition for this is

$$U_h^* \Gamma U_h = \Gamma, \quad \forall h \in H.$$

We denote by $\Sigma(\mathscr{H})$ the space of sesquilinear forms covariant for $(U,\mathscr{H})$. We say that $\langle \cdot,\cdot \rangle$ is of *full rank* if Γ is invertible. A full rank covariant sesquilinear form is called an *indefinite metric*. We denote by $M(\mathscr{H})$ the space of indefinite metrics for $(U,\mathscr{H})$. Let now $(u,W) \in \mathscr{F}$; a sesquilinear form on W is defined by

$$\langle \theta_1, \theta_2 \rangle = (\theta_1, \gamma\theta_2)_W, \quad \theta_1, \theta_2 \in W,$$

where $\gamma \in \mathrm{Hom}_c(W,W)$. The form $\langle \cdot,\cdot \rangle$ is called *covariant* if it is invariant with respect to the action of H_0 on W (through u), i.e.,

$$u(k)^* \gamma u(k) = \gamma, \quad \forall k \in H_0.$$

We denote by $\sigma(W)$ the space of covariant sesquilinear forms for (u,W). In analogy with the previous case we can define the indefinite metrics for (u,W). We denote by $\mu(W)$ the space of the indefinite metrics for (u,W). The fundamental result for covariant sesquilinear forms is the following.

Theorem 4.1: Let $(u,W) \in \mathscr{F}$; then there is a canonical bijection between $\sigma(W)$ and $\Sigma(\mathscr{H}(W))$; this bijection also induces a bijection between the corresponding sets of indefinite metrics.

Proof: Let $\gamma \in \sigma(W)$, so that

$$u(k)^* \gamma u(k) = \gamma, \quad \forall k \in H_0.$$

Let us define on X a function $\gamma(\cdot)$ by $\gamma(\hat{a}_0) = \gamma$ and by defining $\gamma(h[\hat{a}_0])$ through the equation

$$C_u(h : \hat{a}_0)^* \gamma(h[\hat{a}_0]) C_u(h : \hat{a}_0) = \gamma, \quad h \in H.$$

Let us define Γ by

$$(\Gamma f)(\hat{a}) = \gamma(\hat{a}) f(\hat{a}), \quad f \in \mathscr{H}(W)^0.$$

It is easy to show that

$$\Gamma \in \mathscr{L}(\mathscr{H}(W)^0 : \mathscr{H}(W)^0) \quad \text{and} \quad U_h^* \Gamma U_h = \Gamma.$$

We thus have a well defined linear map $\gamma \to \Gamma$ from $\sigma(W)$ to $\Sigma(\mathscr{H}(W))$, which is obviously injective. It is in fact also surjective and it is the bijection referred to in the theorem. We omit the proof since it is essentially based on the same type of arguments used in the unitary case.

V. EXPLICITLY COVARIANT MODELS

Let us suppose that $(u_0, W_0) \in \mathcal{F}$ is such that there exists $(u, W) \in \mathcal{F}(H)$ with the properties: (a) $W_0 \subset W$; (b) if we denote by $u_0' = u|_{H_0}$, then $u_0'|_{W_0} = u_0$.

For all $\hat{a} \in X$ we define

$$W_0(\hat{a}) = u(h) W_0,$$

where $h \in H$ is any element such that

$$h[\hat{a}_0] = \hat{a}, \quad \hat{a} \in X.$$

The definition of $W_0(\hat{a})$ is independent of the choice of h. The map $\hat{a} \mapsto W_0(\hat{a})$ is a *covariant field of subspaces of* W, in the sense that

$$W(\hat{a}_0) = W_0; \; u(h) W_0(\hat{a}) = W_0(h[\hat{a}]).$$

Let us define B and $\pi(B \to X)$ by

$$B = \bigsqcup_{\hat{a} \in X} W_0(\hat{a}), \quad \pi(\theta(\hat{a})) = \hat{a}, \quad \forall \hat{a} \in X.$$

An action of H on B is defined by setting

$$h, \theta(\hat{a}) \mapsto u(h) \theta(\hat{a}) \in W_0(h[\hat{a}]).$$

With these definitions B is a *vector bundle on* X.

Let us denote by $\mathcal{H}^{\mathrm{cov}}(W_0)$ the Hilbert space of *square integrable sections* of B. Explicitly, $\mathcal{H}^{\mathrm{cov}}(W_0)$ is the Hilbert space of all (α-equivalence classes of) Borel maps $f(X \to W)$ such that (a) $f(\hat{a}) \in W_0(\hat{a})$ for all $\hat{a} \in X$; (b) $\int \|f(\hat{a})\|_W^2 \, d\alpha(\hat{a}) < \infty$.

Let us define $U^{\mathrm{cov}}(W_0)$ on $\mathcal{H}^{\mathrm{cov}}(W_0)$ by

$$(U^{\mathrm{cov}}(W_0)_{ah} f)(\hat{a}) = e^{i\langle a, \hat{a} \rangle} u(h) f^h(\hat{a}).$$

It can be shown that there exists a strict isometric isomorphism between $(U(W_0), \mathcal{H}(W_0))$ and $(U^{\mathrm{cov}}(W_0), \mathcal{H}^{\mathrm{cov}}(W_0))$.

The algebra $\mathrm{Morph}(U^{\mathrm{cov}}(W_0))$ is given by the operators

$$\Lambda_\lambda^{\mathrm{cov}}: f(\hat{a}) \mapsto L_\lambda(\hat{a}) f(\hat{a}),$$

where L_λ is the unique function from X in $\mathrm{End}(W_0(\hat{a}))$ satisfying

$$L_\lambda(h[\hat{a}_0]) = u(h) L_\lambda(\hat{a}) u(h)^{-1},$$

with $L_\lambda(\hat{a}_0) = \lambda$, $\lambda \in \mathcal{A}(u_0) =$ commuting algebra of u_0.

We remark that, in general, the functions L_λ will have *unbounded norms* as $\hat{a}$ varies in X; in fact, we are considering the commuting algebra of U^{cov} *in the sense of the category* $\mathcal{P}$.

A similar calculation allows us to transfer to the covariant model $(U^{\mathrm{cov}}(W_0), \mathcal{H}^{\mathrm{cov}}(W_0))$ the description of the bijection between $\sigma(W_0)$ and $\Sigma(\mathcal{H}(W_0))$. Given $\gamma = \bar{\gamma}(\hat{a}_0)$ satisfying

$$u(k)^* \gamma u(k) = \gamma \quad (k \in H_0),$$

the equation

$$u(h)^* \bar{\gamma}(h[\hat{a}_0]) u(h) = \gamma$$

defines $\bar{\gamma}(\hat{a})$ without ambiguities, for all $\hat{a} \in X$. The operator Γ becomes the operator Γ^{cov}, on $\mathcal{H}^{\mathrm{cov}}(W_0)$, defined by

$$(\Gamma^{\mathrm{cov}} f)(\hat{a}) = \bar{\gamma}(\hat{a}) f(\hat{a}).$$

We must stress again that the function $\hat{a} \mapsto \bar{\gamma}(\hat{a}) \in \mathrm{End}(W_0(\hat{a}))$ will have, in general, *unbounded norm* as $\hat{a}$ varies in X.

VI. APPLICATIONS TO THE POINCARÉ GROUP. CONSTRUCTION OF FREE QUANTUM FIELDS

In the following we shall consider some applications to the Poincaré group of the theory we have exposed so far. Such applications are chosen having in mind the problem of building free quantum fields in the sense of Wightman.[7]

We denote by $\mathrm{ISL}(2,\mathbf{C})$ the semidirect product $\mathbf{R}^4 \times' \mathrm{SL}(2,\mathbf{C})$ with the action of $\mathrm{SL}(2,\mathbf{C})$ on $\mathbf{R}^4$ defined by the covering homomorphism δ of the connected component to the identity of the Lorentz group

$$h: x \mapsto \delta(h) x, \quad h \in \mathrm{SL}(2,\mathbf{C}), \quad x \in \mathbf{R}^4.$$

$\mathrm{ISL}(2,\mathbf{C})$ is the universal covering group of the connected component to the identity of the Poincaré group. We denote by $x = (x_0, \mathbf{x}) = (x_0, x_1, x_2, x_3)$ the elements of $\mathbf{R}^4$; $\mathrm{ISL}(2,\mathbf{C})$ can be regarded as the Lie group of transformations on $\mathbf{R}^4$, whose action is

$$a, h: x \mapsto \delta(h) x + a, \quad a, h \in \mathrm{ISL}(2,\mathbf{C}), \quad x \in \mathbf{R}^4.$$

We denote by $\mathbf{P}^4$ ($\cong \mathbf{R}^4$) the dual group (character group) of $\mathbf{R}^4$; and by $p = (p_0, \mathbf{p}) = (p_0, p_1, p_2, p_3)$ the elements of $\mathbf{P}^4$; the dual action of $\mathrm{SL}(2,\mathbf{C})$ on $\mathbf{P}^4$ is given by

$$h: p \mapsto \delta(h) p, \quad h \in \mathrm{SL}(2,\mathbf{C}), \quad p \in \mathbf{P}^4.$$

The canonic duality between $\mathbf{R}^4$ and $\mathbf{P}^4$ is defined by the bilinear form

$$p \cdot x = p_0 x_0 - p_1 x_1 - p_2 x_2 - p_3 x_3,$$

which is invariant with respect to the action of $\mathrm{SL}(2,\mathbf{C})$.

We assume that the orbit structure of $\mathbf{P}^4$ is known, as well as the stabilizers of the various orbits and the properties of the invariant measures on the orbits. See Ref. 6 for details.

We denote by X one of the orbits X_m^+ ($m \geqslant 0$), and we denote by p^+ the point

$$p^+ = \begin{cases} (m,0,0,0) \in X_m^+, & \text{if } m > 0; \\ (1,0,0,1) \in X_0^+ & \text{if } m = 0. \end{cases}$$

We denote by H_0 the stabilizer of p^+, namely

$$H_0 = \mathrm{SU}(2), \text{ if } m > 0, \quad H_0 = E, \text{ if } m = 0.$$

The invariant measure on X (resp. X_m^+, X_0^+) is denoted by α (resp. α_m, α_0). We notice that α defines, for all $m \geqslant 0$, a Radon measure on $\mathbf{P}^4$.

We assume that the classification of the unitary representations (of physical interest) of $\mathrm{ISL}(2,\mathbf{C})$ and their realizations are known. See Refs. 6 and 8 for details.

In quantum field theory not only unitary representations of $\mathrm{ISL}(2,\mathbf{C})$ are considered but, more generally, also reducible nonunitary representations which leave a nonpositive definite Hermitian form invariant.[9–11]

The presence of the noncompact group E among the stability subgroups of $\mathrm{SL}(2,\mathbf{C})$ leads to consider indecomposable representations whose study may be based on the methods of cohomology.[1,5] The study of nonunitary repre-

sentations and of their indefinite metrics falls naturally in the framework of the theory we have developed in the first part of the paper. Moreover the Hilbert functor

$$(u,W) \mapsto (U(W),\mathscr{H}(W))$$

reduces the infinite dimensional cohomology problem of the indecomposable representations of ISL(2,C) to the much simpler study of finite dimensional representations of the stability subgroups [E and SU(2)]. In particular, in Appendix A, we give a general method for the construction of continuous finite dimensional indecomposable modules for the group E.

Let us now discuss how a quantum field can be associated to a representation of ISL(2,C). This is well known, and we consider it only for fixing the notations; for details see Ref. 12 in the unitary case, and Ref. 13 in the case of indefinite metric. We shall consider only the symmetric (boson) case. Let $\mathscr{H}$ be a Hilbert space and U a representation of ISL(2,C) such that $(U,\mathscr{H})\in\mathscr{P}$. Let us denote by $\mathscr{F}_S(\mathscr{H})$ the symmetrized Fock space defined on $\mathscr{H}$; let F_0 denote the dense linear subspace of $\mathscr{F}_S(\mathscr{H})$ consisting of the states with a *finite number of particles*. We notice that the structure of $\mathscr{F}_S(\mathscr{H})$ as a Hilbert space depends only on the scalar product of $\mathscr{H}$ which is not necessarily left invariant by U. We denote by $\langle\cdot,\cdot\rangle$ an indefinite metric for $(U,\mathscr{H})$ such that

$$\langle f,g\rangle = (f,\Gamma g), \quad f,g\in\mathscr{H},$$

where Γ is a bounded Hermitian operator with bounded inverse on $\mathscr{H}$. We can then define on $\mathscr{F}_S(\mathscr{H})$ the creation $a^*(f)$ and annihilation $a(f)$ operators with respect to the indefinite metric $\langle\cdot,\cdot\rangle$. In particular, we get the commutation rule on F_0,

$$[a(f),a^*(g)] = \langle f,g\rangle, \quad f,g\in\mathscr{H}.$$

The generalized *Segal quantization* operator $\Xi(f)$ is defined by

$$\Xi(f) = (1/\sqrt{2})(a(f) + a^*(f)), \quad f\in\mathscr{H}.$$

$\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ denotes the space of N-complex component Schwartz functions on $\mathbf{R}^4$; ISL(2,C) acts on it (see the next section). Let A be a continuous map from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ to $\mathscr{H}$ that commutes with the actions of ISL(2,C); we call it an *intertwining map* (cf. Definition 7.1, *infra*). Let A be the operator valued distribution on $\mathbf{R}^4$ defined by

$$\mathscr{A}(\varphi) = \Xi(A\varphi), \quad \varphi\in\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4).$$

Then the distribution $\mathscr{A}$ is a quantum field in the sense of Wightman. The field $\mathscr{A}$ is completely defined by $\mathscr{H}$, $\langle\cdot,\cdot\rangle$, and A. Having $\mathscr{H}$ and $\langle\cdot,\cdot\rangle$ fixed, and varying A within the possible intertwining maps between $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ and $\mathscr{H}$, we get a family of fields that differ from each other by *a gauge transformation*. On the other hand, suppose that $\mathscr{A}(\varphi\mapsto\mathscr{A}(\varphi))$ is a *free* covariant quantum field, acting in the Fock space $\mathscr{F}_S(\mathscr{H})$. If Ω denotes the vacuum state of $\mathscr{F}_S(\mathscr{H})$, then it is easily seen that $(\varphi\mapsto\mathscr{A}(\varphi)\Omega)$ is an intertwining map from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ to $\mathscr{H}$. Then we can conclude that on a fixed Fock space with a given indefinite metric *the free covariant quantum fields are in* 1-1 *correspondence with the intertwining maps* $\mathscr{A}(\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)\mapsto\mathscr{H})$.

The recipe for building a free quantum field may be summarized in the following three points: (1) we fix a Hilbert space $\mathscr{H}$ and a *unitary* representation U of ISL(2,C) acting on $\mathscr{H}$; this will fix the *physical content of the field*, (2) we classify all possible extensions of $(U,\mathscr{H})$, i.e., all possible objects $(U^1,\mathscr{H}^1)$ in $\mathscr{P}$ which admit $(U,\mathscr{H})$ as a subquotient, and all possible indefinite metrics for these extensions, (3) if $(U^1,\mathscr{H}^1)$ is the extension of $(U,\mathscr{H})$ that we want to consider, then we classify all intertwining maps

$$A(\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)\to\mathscr{H}^1).$$

For what concerns point (1) there is no problem since all unitary representations of ISL(2,C) are classified. For what concerns point (2) we have the following facts: $(U,\mathscr{H})\in\mathscr{P}$ since U is unitary; from Theorem 3.2 $(U,\mathscr{H})$ is isomorphic in $\mathscr{P}$, to a pair $(U(W),\mathscr{H}(W))$, where $(u,W)\in\mathscr{F}$. Moreover this theorem assures us that the extensions of $(U,\mathscr{H})$ that are elements of $\mathscr{P}$ are in 1-1 correspondence with the extensions of (u,W). Further, Theorem 4.1 completely solves the problem of the indefinite metrics for a pair $(U,\mathscr{H})\in\mathscr{P}$.

The problem raised in point (3) is discussed in the following section.

VII. THE INTERTWINING MAPS

In this section we denote by u an N-dimensional matrix representation of SL(2,C) so that $(u,\mathbf{C}^N)\in\mathscr{F}(\mathrm{SL}(2,\mathbf{C}))$; we denote by u_0 the restriction of u to H_0, then $(u_0,\mathbf{C}_N)\in\mathscr{F}$. We define an action of ISL(2,C) on $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ by

$$(a,h):\varphi(x)\to u(h)\varphi(\delta(h)^{-1}(x-a)), \quad \varphi\in\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4).$$

Let us suppose that $(U,\mathscr{H})$ be an element of $\mathscr{P}$.

Definition 7.1: A map

$$A:\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)\to\mathscr{H}$$

that commutes with the action of ISL(2,C) and is continuous [in the sense that if $\varphi_i\to 0$ in $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$, then $A\varphi_i\to 0$ in $\mathscr{H}$] is called an intertwining map from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ to $\mathscr{H}$.

As we show in the sequel, the problem of characterizing the intertwining maps from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ to $\mathscr{H}$ can be reduced, for the cases we are interested in, to the problem of characterizing the intertwining maps

$$A:\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)\to L^2(X:\mathbf{C}^N),$$

where the action of ISL(2,C) on $L^2(X:\mathbf{C}^N)$ is given in the explicitly covariant form. This is the problem we deal with in the present section.

Let us introduce the *covariant* Fourier transform

$$\varphi\mapsto\hat{\varphi}:\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)\to\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$$

as

$$\hat{\varphi}(p) = \frac{1}{(2\pi)^2}\int e^{ip\cdot a}\varphi(x)d^4x.$$

We define the action of ISL(2,C) on $\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$ by

$$(a,h):\hat{\varphi}(p)\to e^{ip\cdot a}u(h)\hat{\varphi}^h(p),$$

where we use the notation $\hat{\varphi}^h(p) = \hat{\varphi}(\delta(h)^{-1}p)$. The presence of the term $p\cdot a$ in its definition assures that the *covar-*

iant Fourier transform commutes with the action of ISL(2,C).

We introduce also the space $\mathscr{S}_{\mathbf{C}^N}(X)$ of the functions $f(X \to \mathbf{C}^N)$ that are the restrictions to X of functions in $\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$. We consider $\mathscr{S}_{\mathbf{C}^N}(X)$ as endowed with the quotient topology. The action of ISL(2,C) on $\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$ induces an action on $\mathscr{S}_{\mathbf{C}^N}(X)$; the restriction map

$$\text{Res}: \mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4) \to \mathscr{S}_{\mathbf{C}^N}(X)$$

commutes with the action of ISL(2,C) and is continuous.

A fundamental result that we shall use in the sequel is the following.

Lemma 7.1: (a) For each $\lambda \in \mathscr{A}(u_0)$ let L_λ be the unique map of X into End($\mathbf{C}^N$) such that

$$L_\lambda(p_0) = \lambda,$$

$$L_\lambda(\delta(h)p) = u(h) L_\lambda(p) u(h)^{-1}$$

for each $h \in \text{SL}(2,\mathbf{C})$, $p \in X$. Then there exists a polynomial map $\theta_\lambda(\mathbf{P}^4 \to \text{End}(\mathbf{C}^N))$ such that

$$L_\lambda = \theta_\lambda|_X.$$

(b) If Λ_λ denotes the operator on $C^\infty_{\mathbf{C}^N}(X)$ that is the multiplication by L_λ, then Λ_λ defines a continuous operator on $\mathscr{S}_{\mathbf{C}^N}(X)$ which commutes with ISL(2,C). (c) Λ_λ defines an operator on $L^2(X: \mathbf{C}^N)^0$ that is an element of Morph(U).

Proof: The proof of point (a) is given in Appendix B. Point (b) follows from point (a), since the multiplication by a matrix of polynomials defines a continuous operator on $\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$. For what concerns point (c), because of point (a), L_λ is a continuous (matrix valued) function on X, hence it defines an element of Morph(U). This proves the Lemma.

Remark: The multiplication operator L_λ also has the stronger property: if C is compact in $\mathbf{P}^4$, L_λ defines a bounded operator on $S(C \cap X)$. Notice that, since X is not in general a closed set in $\mathbf{P}^4$, then $C \cap X$ is not in general relatively compact in X.

Lemma 7.2: Let A_0 ($\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4) \to L^2(X: \mathbf{C}^N)$) be defined by

$$A_0\varphi = \text{Res} \circ \hat{\varphi}, \quad \varphi \in \mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4);$$

then A_0 is an intertwining map; moreover, for each $\lambda \in \mathscr{A}(u_0)$, $\Lambda_\lambda \circ A_0$ is an intertwining map.

Proof: Obvious from Lemma 7.1.

Remark: We stress that Λ_λ is *not* a continuous operator on $L^2(X: \mathbf{C}^N)$, nevertheless the fact that L_λ is a polynomial function guarantees that the operator $\Lambda_\lambda \circ A_0$ is a continuous operator from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4)$ to $L^2(X: \mathbf{C}^N)$.

Definition 7.2: We say that a map

$$A: \mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4) \to L^2(X: \mathbf{C}^N)$$

is continuous in the weak sense if

$$\varphi, f \mapsto (A\varphi, f)_{L^2}, \quad \varphi \in \mathscr{S}_{\mathbf{C}^N}(\mathbf{R}^4), \quad f \in \mathscr{S}_{\mathbf{C}^N}(X)$$

is continuous in each variable, in the Schwartz topology.

Lemma 7.3: Every intertwining map which is continuous in the weak sense

$$A: \mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4) \to L^2(X: \mathbf{C}^N)$$

is of the form

$$A = \Lambda_\lambda \circ A_0,$$

where $\lambda \in \mathscr{A}(u_0)$ uniquely determined.

Remark: The following proof makes use of the Schwartz's theorem on kernels, in order to express every continuous, in weak sense, intertwining map as a distribution having suitable invariance and support properties that force it to be in the form $\Lambda_\lambda \circ A_0$. Such a technique is due to Bruhat, who developed it in great generality and depth in his thesis.[14]

Proof of Lemma 7.3: Let A an intertwining map in the weak sense. Since the Fourier transform is a topological linear isomorphism, we can write $A\varphi = K\hat{\varphi}$, where K is an intertwining map in weak sense from $\mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$ to $L^2(X: \mathbf{C}^N)$. We have to prove that $K = \Lambda_\lambda \circ \text{Res}$ for a suitable Λ. Let

$$\{f,g\} = \sum_{i=1}^{N} \int_X f_i g_i \, d\alpha, \quad f,g \in \mathscr{S}_{\mathbf{C}^N}(X)$$

[this integral is well defined also in the case $X = X_0^+$ since α always defines (also in the case $m = 0$) a Radon measure on $\mathbf{P}^4$]. If we define

$$B(f,g) = \{Kf, \text{Res}\, g\} \quad [\, f,g \in \mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)\,],$$

then the hypothesis of weak continuity of K assures that B is a bilinear form continuous in each variable. Then, from the kernel theorem of Schwartz, there exists a unique matrix whose entries K_{ij} are tempered distributions on $\mathbf{P}^4 \times \mathbf{P}^4$ such that

$$\sum_{i,j=1}^{N} K_{ij}(f_j \otimes g_i) = B(f,g), \quad f,g \in \mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4).$$

Let us now suppose that K commutes with the action of the translation subgroup, namely

$$e^{ip\cdot a}(Kf) = K(e^{ip\cdot a} f).$$

From this equality we get

$$\sum_{i,j=1}^{N} (e^{ip\cdot a} - e^{iq\cdot a}) K_{ij}(f_j \otimes g_i) = 0.$$

If we choose f, g such that they have all vanishing components except $f_j = \zeta$ and $g_i = \xi$ with $\zeta, \xi \in \mathscr{S}(\mathbf{P}^4)$ then

$$(e^{ip\cdot a} - e^{iq\cdot a}) K_{ij}(\zeta \otimes \xi) = 0,$$

$$\forall \zeta, \xi \in \mathscr{S}(\mathbf{P}^4), \quad \forall a \in \mathbf{R}^4 \text{ and } \forall i,j = 1,\ldots,N.$$

By differentiating with respect to a and putting $a = 0$, we obtain

$$(p_\mu - q_\mu) K_{ij} = 0, \quad \forall \mu = 0,1,2,3.$$

Let us denote $p - q = v$, $p + q = w$, then $v_\mu K'_{ij} = 0$, where K'_{ij} is a distribution in the variables v, w. From a standard result of distribution theory we conclude that

$$K'_{ij} = \delta(v) \otimes \theta_{ij}(w).$$

The change of variables from p,q to v,w transforms the function $\zeta(p) \otimes \xi(q)$ into a function $\chi(v,w)$ such that

$$\chi(0,w) = \zeta(p)\xi(q)$$

so that

$$K_{ij}(\zeta \otimes \xi) = K'_{ij}(\chi) = \theta_{ij}(\zeta\xi),$$

where θ_{ij} is a uniquely determined tempered distribution on $\mathbf{P}^4$.

Let us consider now the intertwining condition with respect the action of $SL(2,\mathbf{C})$. Let us denote

$$[u(h)]_{ij} = u_{ij}(h), \quad [u^\vee(h)]_{ij} = u^{ji}(h), \quad 1 \leqslant i,j \leqslant N.$$

Since

$$\{u(h)f^h, u^\vee(h)g^h\} = \{f,g\}, \quad f,g \in \mathscr{S}_{\mathbf{C}^N}(X),$$

it follows that

$$\begin{aligned} B(u(h)f^h, u^\vee(h)g^h) &= \{K(u(h)f^h), \operatorname{Res} u^\vee(h)g^h\} \\ &= \{u(h)(Kf)^h, u^\vee(h)\operatorname{Res} g^h\} \\ &= \{Kf, \operatorname{Res} g\} = B(f,g). \end{aligned}$$

If we choose f, g such that they have all vanishing components except $f_j = \xi$ and $g_i = \xi$ with $\xi,\xi \in \mathscr{S}(\mathbf{P}^4)$ then we get

$$\theta_{ij}(\xi\xi) = \sum_{k,l=1}^{n} u^{ik}(h)\theta_{kl}((\xi\xi)^h)u_{lj}(h).$$

Since $\xi,\xi \mapsto \xi\xi$ has dense range in $\mathscr{S}(\mathbf{P}^4)$ we have

$$\theta_{ij} = \sum_{k,l=1}^{N} u^{ik}(h)\theta_{kl}^{h^{-1}}u_{lj}(h).$$

Hence the matrix Θ whose entries are the distributions $[\Theta]_{ij} = \theta_{ij}$ satisfies the relation

$$\Theta = u(h)\Theta^h u(h)^{-1}, \quad \forall h \in SL(2,\mathbf{C}). \tag{7.1}$$

Let us finally consider the condition on the support of the distribution Θ. We want to prove that Θ *lives* on the closure $\mathrm{Cl}(X)$ of X, in the following sense:

$$\varphi \in C_c^\infty(\mathbf{P}^4) \text{ and } \varphi = 0 \text{ on } \mathrm{Cl}(X) \Rightarrow \theta_{ij}(\varphi) = 0$$

(notice that X_0^+ is not closed in $\mathbf{P}^4$). In fact, given any φ such that $\varphi \in C_c^\infty(\mathbf{P}^4)$ and $\varphi = 0$ on $\mathrm{Cl}(X)$, one can always choose $\psi \in C_c^\infty(\mathbf{P}^4)$ such that $\psi = 1$ on $\mathrm{supp}(\varphi)$. Then $\varphi\psi = \varphi$ and so

$$\theta_{ij}(\varphi) = \theta_{ij}(\varphi\psi) = K_{ij}(\psi \otimes \varphi).$$

We choose f,g such that they have all vanishing components except $f_j = \psi$ and $g_i = \varphi$; then $f,g \in C_{c\mathbf{C}^N}^\infty(\mathbf{P}^4)$, and

$$\theta_{ij}(\varphi) = B(f,g) = 0,$$

since $\operatorname{Res} g = 0$.

Let us denote

$$\Omega = \begin{cases} \mathbf{P}^4, & \text{if } X = X_m^+, \\ \mathbf{P}^4 \setminus \{0\}, & \text{if } X = X_0^+ \end{cases}$$

In all cases Ω is open in $\mathbf{P}^4$ and X is a regular *closed* submanifold of Ω. Θ is a (matrix valued) distribution in Ω that lives, in the above sense, on X. Therefore we can write

$$\Theta(\varphi) = \Theta_0(\operatorname{Res}\varphi), \quad \varphi \in C_c^\infty(\Omega),$$

for a uniquely determined distribution Θ_0 on $X \simeq SL(2,\mathbf{C})/H_0$; moreover (7.1) implies that Θ_0 satisfies

$$\Theta_0 = u(h)\Theta_0^h u(h)^{-1}, \quad \forall h \in SL(2,\mathbf{C}). \tag{7.2}$$

We use at this point the result of the next lemma to conclude that Θ_0 is a C^∞ function on X; in other words, there exists a C^∞ matrix Φ on X such that

$$\Theta_0(\varphi) = \int_X \Phi(p)\varphi(p)d\alpha(p) \quad [\varphi \in C_c^\infty(X)].$$

The covariance condition (7.2) implies

$$\Phi(\delta(h)p) = u(h)\Phi(p)u(h)^{-1}, \quad \forall p \in X, \quad \forall h \in SL(2,\mathbf{C}).$$

By setting $p = p^+$ and choosing $h \in H_0$ we get

$$\Phi(p^+) = u_0(h)\Phi(p^+)u_0(h)^{-1}, \quad h \in H_0.$$

Thus $\lambda = \Phi(p^+) \in \mathscr{A}(u_0)$ and $\Phi(p) = L_\lambda(p)$; λ uniquely determined by K.

Hence, for all $g \in C_{c\mathbf{C}^N}^\infty(\Omega)$, $f \in \mathscr{S}_{\mathbf{C}^N}(\mathbf{P}^4)$,

$$B(f,g) = \sum_{ij}^{N} \theta_{ij}(f_j g_i) = \{L_\lambda \cdot \operatorname{Res} f, \operatorname{Res} g\}.$$

Since Λ_λ is the multiplication operator by L_λ, this equation can be written as

$$\{Kf, \operatorname{Res} g\} = \{\Lambda_\lambda \circ \operatorname{Res} f, \operatorname{Res} g\}.$$

Since the restrictions of the elements $g \in C_{c\mathbf{C}^N}^\infty(\Omega)$ to X are dense in $L^2(X;\mathbf{C}^N)$, we obtain $Kf = \Lambda_\lambda \circ \operatorname{Res} f$. This proves the lemma.

We still have to show the property, used in the proof of the preceding lemma, that Θ is a C^∞ function matrix. We notice that the vector space generated by the distributions θ_{ij}, $1 \leqslant i,j \leqslant N$, is stable with respect to the action of $SL(2,\mathbf{C})$, because of (7.2). Therefore, in order to show the property we are interested in, it is sufficient to prove the following lemma. We work with a connected Lie group L and a closed subgroup L_0 such that $Y = L/L_0$ admits an invariant measure.

Lemma 7.4: Let t be a distribution on $Y = L/L_0$ such that its transforms t^h, for $h \in L$, span a finite dimensional vector space. Then t is a C^∞ function.

Proof: Let us denote by β the canonical projection

$$\beta: L \to Y.$$

Then β induces a map, $f \mapsto f'$: $C_c^\infty(L) \to C_c^\infty(Y)$ defined by

$$f'(\beta(h)) = \int_{L_0} f(hh_0)d\mu_0(h_0), \quad \forall f \in C_c^\infty(L),$$

where μ_0 is a left Haar measure on L_0. It is possible to show (with a straightforward generalization of the methods used in the proof of Lemmas 5.13 and 5.14 of Ref. 6) that $f \mapsto f'$ has the following properties:

(1) it is surjective and continuous;

(2) $(f^h)' = (f')^h$.

If t is a distribution on X we can use the map $f \mapsto f'$ in order to lift t to a distribution on the manifold of the group L:

$$t'(f) = t(f'), \quad \forall f \in C_c^\infty(L).$$

The map $t \mapsto t'$ (pull back) is injective and t' is right L_0 invariant. So if we want to prove that t is a C^∞ function it is enough to prove that t' is a C^∞ function. Let us consider the action of L on itself by left translations; then L acts on the functions of $C_c^\infty(L)$, hence on the distributions on L. Let us denote such an action by

$$h: t' \to h \cdot t', \quad h \in L.$$

As h varies in L, $h \cdot t'$ span a finite dimensional vector space V. The action of L on V induces an action on V of the Lie

algebra of L, $t' \to \partial(X)t'$, and induces also an action of the universal enveloping algebra of this Lie algebra; if a denotes any element of the latter, then $\partial(a)t' \in V$. Let us choose a basis $\{X_1,...,X_n\}$ of Lie(L) and denote by $\Box = X_1^2 + ... + X_n^2$. $\Box$ is an element of the universal enveloping algebra and $\partial(\Box)$ is a differential operator on the L manifold such that $\partial(\Box)t' \in V$. Also $\Box^k$ (the k th power) is in the universal enveloping algebra and

$$\partial(\Box^k)t' \in V, \quad k = 0,1,... .$$

Since V is finite dimensional, then there exist constants $c_1,...,c_k$ such that

$$\partial(\Box^k + c_1\Box^{k-1} + ... + c_k)t' = 0.$$

The differential operator $\partial(\Box^k + c_1\Box^{k-1} + ... + c_k)$ is *elliptic* on the L manifold, as can be shown by calculating the symbol of the operator in a local coordinate system, ellipticity being assured by the fact that $\{X_1,...,X_n\}$ is a basis of Lie(L). Then by the elliptic regularity theorem, t' is a C^∞ function on L. The remarks made above then show that t is a C^∞ funciton on Y. This proves the lemma.

If A is an intertwining map in the sense of Definition 7.1, then A is continuous in the sense of Definition 7.2, hence A is of the form $A = \Lambda_\lambda \circ A_0$, by Lemma 7.3. Conversely, Lemma 7.2 assures that, if A is of the from $A = \Lambda_\lambda \circ A_0$ then it is an intertwining map. We can then conclude the whole discussion of this section by stating the following theorem.

Theorem 7.1: The intertwining maps

$$A: \mathscr{S}_{C^N}(\mathbf{R}^4) \to L^2(X: \mathbf{C}^N)$$

are all and only those of the form

$$A = \Lambda_\lambda \circ A_0$$

for λ varying in $\mathscr{A}(u_0)$.

VIII. THE GUPTA–BLEULER THEORY OF THE EM FIELD

In this final section we show how the theory of the representations (of the Poincaré group) outlined in this paper can be concretely used to build a free quantum field. We examine the Gupta–Bleuler theory of the EM field as it is discussed in Ref. 10. We choose this example because it is the prototype of a quantum field theory with indefinite metric.

There are other elementary examples (massive vector field,[9] Boulware–Gilbert model[15]) for which our machinery works. We do not review them here because of their triviality. We believe that our approach is useful also for treating less elementary models of covariant quantum fields with indefinite metrics.[11] In particular, we shall treat the free gravitational quantum field in another paper.

We denote by (n), for any relative integer n, the unitary irreducible representations of E defined by

$$(n): \begin{pmatrix} x & a \\ 0 & z^{-1} \end{pmatrix} \mapsto x^n.$$

We write $W^2 = (2) \oplus (-2)$; W^2 is a unitary representation of E. We also write W^1 for the trivial one-dimensional representation (0) of E.

We denote by $(U(W^2), \mathscr{H}(W^2))$ the *unitary* element of $\mathscr{P}$ that corresponds to W^2. $\mathscr{H}(W^2)$ is the space of the quantum states of the photon. We shall show in this section that it is not possible to define a covariant quantum four-vector field that acts on the Fock space defined on $\mathscr{H}(W^2)$ or on the carrier space of any other model of the representation $U(W^2)$ of ISL(2,C). This fact is at the root of the difficulties of the covariant quantization of the EM field and can be shown in a quite transparent way using the theorem on intertwining maps of the preceding section. For this reason, we have to look for pairs $(U,\mathscr{H})$, other than $(U(W^2),\mathscr{H}(W^2))$, that can be used to define a quantum EM field. They have to fulfill the following requirements: (1) U has not to be a fully reducible representation that contains $U(W^2)$ as direct summand; otherwise we are led again to the impossibility of defining a covariant EM field on $\mathscr{H}(W^2)$. (2) U has to be indecomposable [as consequence of (1)] and has to be an extension of $U(W^2)$ [in the sense that it contains $U(W^2)$ as a subquotient]. (3) $(U,\mathscr{H})$ has to admit an indefinite metric.

As explained in Sec. V the problem of finding extensions of $(U(W^2),\mathscr{H}(W^2))$ in $\mathscr{P}$ that admit an indefinite metric is equivalent to the problem of finding the extensions of W^2 as (finite dimensional) modules for E that have an indefinite metric. We shall restrict to extensions of W^2 up to dimension 4. This is for two reasons: the first is that it gives a *minimal solution*; the second is that we want to define a quantum field that transforms as a four-vector field, because it couples locally to the currents. We shall show that there is a unique solution to this problem (that leads to the Gupta–Bleuler theory of the EM field).

The classification of the indecomposable finite dimensional modules for E is fully discussed in Appendix A. The main result is that there is a unique indecomposable four-dimensional module for E that allows an indefinite metric and that extends W^2. As shown in Appendix A (the equivalence class of) such a representation is the (the equivalence class of the) restriction to E of the four-dimensional irreducible representation $D^{(1/2,1/2)}$ of SL(2,C). We stress the fact that there exist other (equivalence classes of) indecomposable three and four-dimensional extensions of W^2; the requirement of existence of an indefinite metric for the representation forces us to select only the (equivalence class of the) restriction of $D^{(1/2,1/2)}$.

We briefly describe the explicit realizations of the extensions of W^2 that we shall use.

We consider the realization of the representation $D^{(1/2,1/2)}$ of SL(2,C) defined by the covering homomorphism δ; in a natural way we consider δ as acting on $\mathbf{C}^4$

$$h \mapsto \delta(h): \mathrm{SL}(2,\mathbf{C}) \mapsto \mathrm{GL}(4,\mathbf{C}).$$

This realization preserves the *sesquilinear* form

$$v,w \to \bar{v}\cdot w = \bar{v}_0 w_0 - \bar{v}_1 w_1 - \bar{v}_2 w_2 - \bar{v}_3 w_3, \quad v,w \in \mathbf{C}^4.$$

In the sequel we shall denote by W^4 the restriction to E of this matrix realization of $D^{(1/2,1/2)}$.

The commuting algebra $\mathscr{A}(W^4)$ is easily seen to be

$$\mathscr{A}(W^4) = \mathbf{C}\mathbf{I} + \mathbf{C}\lambda_0,$$

where λ_0 is the operator on $\mathbf{C}^4$ defined by

$$\lambda_0 v = (p^+ \cdot v)p^+, \quad v \in \mathbf{C}^4$$

$[p^+ = (1,0,0,1)]$.

The sesquilinear covariant forms for W^4 are easily seen to be given by

$$v,w \mapsto a\bar{v}\cdot w + b(\bar{v}\cdot p^+)(p^+\cdot w)$$

as a, b vary in $\mathbf{R}$. We have an indefinite metric for W^4 if and only if $a \neq 0$.

W^4 has a flag of E-invariant submodules

$$W^4 \supset W^3 \supset W^1,$$

$$W^4/W^3 \simeq W^1; \quad W^3/W^1 \simeq W^2.$$

$W^3 = \text{span}\{(1,0,0,1), (0,1,0,0), (0,0,1,0)\}$ is an E-invariant subspace of $\mathbf{C}^4$ that is a realization of an indecomposable three-dimensional extension of W^2. No indefinite metric there exists for W^3. The restriction to W^3 of the indefinite metric $v,w \mapsto -\bar{v}\cdot w$ of W^4,

$$v,w \mapsto \bar{v}_2 w_2 + \bar{v}_3 w_3, \quad v,w \in W^3 \subset \mathbf{C}^4,$$

defines an E-invariant seminorm on W^3.

$W^1 = \text{span}\{(1,0,0,1)\}$ is an E-invariant subspaces of W^3 and coincides with the subspace of W^3 of null-seminorn elements. We have $W^2 \simeq W^3/W^1$. We notice that the scalar product of W^4 induces on W^3 an invariant (since W^2 is unitary for E) scalar product that coincides with the covariant positive definite sesquilinear form induced on W^3 by the indefinite metric $v,w \mapsto \bar{v}\cdot w$ of W^4.

The Hilbert functor allows to transfer at once all the properties of the extensions of W^3 to analogous properties of the extensions in $\mathscr{P}$ of $(U(W^2), \mathscr{H}(W^2))$.

W^4 is a nonunitary indecomposable representation of E, and so the representation of $\text{ISL}(2,\mathbf{C})$ that correspond to it in $\mathscr{P}$ is nonunitary and indecomposable. Since W^4 is the restriction to E of a representation of $\text{SL}(2,\mathbf{C})$, $(U(W^4), \mathscr{H}(W^4))$ has an explicitly covariant realization. We denote by $L^2(X: \mathbf{C}^4)$ the space of (α-equivalence classes of) functions $f(X \to \mathbf{C}^4)$ defined on $X = X_0^+$, such that

$$\int_X d\alpha(p)|f(p)|^2 < \infty$$

[where $|f(p)|$ denotes the Euclidean norm $|f(p)|^2 = \Sigma_{\mu=0}^3 |f_\mu(p)|^2$]. We define on $L^2(X: \mathbf{C}^4)$

$$(U_{ah}f)(p) = e^{ip\cdot a}\delta(h) f^h(p).$$

$(U, L^2(X: \mathbf{C}^4))$ is the explicitly covariant realization of $(U(W^4), \mathscr{H}(W^4))$. The algebra Morph$(U)$ is the algebra of the multiplication operators L_λ on $L^2(X: \mathbf{C}^4)$ of the form

$$L_\lambda(p) = \delta(h)\lambda\delta(h)^{-1},$$

where $\lambda \in \mathscr{A}(W^4)$ and h is any element of $\text{SL}(2,\mathbf{C})$ such that $\delta(h)p^+ = p$. For $f \in L^2(X: \mathbf{C}^4)$ we get

$$L_\lambda(p) f(p) = \alpha f(p) + \beta p(p\cdot f(p)), \quad \alpha,\beta \in \mathbf{C}.$$

The indefinite metrics for $(U, L^2(X: \mathbf{C}^4))$ are of the form

$$f,g \mapsto a\int_X d\alpha(p)\ \overline{f(p)}\cdot g(p) + b\int_X d\alpha(p)(\overline{f(p)}\cdot p)(p\cdot g(p)),$$

$a,b \in R$, $a \neq 0$. In particular,

$$f,g \mapsto -\int_X d\alpha(p)\ \overline{f(p)}\cdot g(p) \tag{8.1}$$

is defined by a bounded operator, with bounded inverse. This is the only (up to scalar multiplication) indefinite metric for $(U, L^2(X: \mathbf{C}^4))$ that has this property.

$\mathscr{H}(W^3)$ is a closed invariant subspace of $\mathscr{H}(W^4)$. In the sequel we shall denote

$$\mathscr{H}^\sim = \mathscr{H}^{\text{cov}}(W^3), \quad U^\sim = U^{\text{cov}}(W^3).$$

$\mathscr{H}^\sim$ is the space of the (α-equivalence classes of) functions $f(X \mapsto \mathbf{C}^4)$ such that

$$p\cdot f(p) = 0, \quad p \in X,$$

and

$$\int_X d\alpha(p)|f(p)|^2 < \infty.$$

$U^\sim$ is defined on $\mathscr{H}^\sim$ as

$$(U^\sim_{ah})(p) = e^{ip\cdot a}\delta(h) f^h(p).$$

$\mathscr{H}^\sim$ is an invariant closed subspace of $\mathscr{H}$. The restriction to $\mathscr{H}^\sim$ of the indefinite metric (8.1) is

$$f,g \mapsto \sum_{i=1}^{2}\int_X d\alpha(p)\ \overline{f_i(p)} g_i(p);$$

it defines an invariant seminorm on $\mathscr{H}^\sim$.

Since $W^4 \supset W^3 \supset W^1$, $\mathscr{H}(W^1)$ is an invariant closed subspace of (W^4) and $\mathscr{H}(W^3)$. We shall denote

$$\mathscr{H}^0 = \mathscr{H}^{\text{cov}}(W^1), \quad U^0 = U^{\text{cov}}(W^1),$$

$\mathscr{H}^0$ is the space of the (α-equivalence classes of) functions $f(X \mapsto \mathbf{C}^4)$ such that

$$f(p) = p\lambda_f(p), \quad p \in X,$$

where λ_f is an element of $L^2(X, d\alpha)$. The restriction to $\mathscr{H}^0$ of the indefinite metric (8.1) is the null form; in other words $\mathscr{H}^0$ coincides with the invariant closed subspace of $\mathscr{H}^\sim$ of all elements with invariant seminorm 0.

Corresponding to the flag

$$W^4 \supset W^3 \supset W^1,$$

we have the flag

$$L^2(X: \mathbf{C}^4) \supset \mathscr{H}^\sim \supset \mathscr{H}^0,$$

where

$$L^2(X: \mathbf{C}^4)/\mathscr{H}^\sim \simeq \mathscr{H}^0; \quad \mathscr{H}^\sim/\mathscr{H}^0 \simeq \mathscr{H}(W^2).$$

The representation $U^\sim/U^0$ that acts in the Hilbert space $\mathscr{H}^\sim/\mathscr{H}^0$ is the explicitly covariant realization of $U(W^2)$. We shall denote

$$\mathscr{H}^{\text{photon}} = \mathscr{H}^\sim/\mathscr{H}^0, \quad U^{\text{photon}} = U^\sim/U^0.$$

We notice that the invariant scalar product of $\mathscr{H}^{\text{photon}}$ coincides with the invariant positive definite sesquilinear form induces on it by (8.1).

We consider now the intertwining maps. Theorem 7.1 ensures that the intertwining maps $A\{\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4) \mapsto L^2(X: \mathbf{C}^4)\}$ are given by

$$(A\varphi)(p) = \lambda\ \text{Res}\,\hat{\varphi}(p), \quad \varphi \in \mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4),$$

where λ is an element of $\mathscr{A}(W^4)$. Therefore

$$(A\varphi)(p) = \alpha \operatorname{Res} \hat{\varphi}(p) + \beta p(p\cdot \operatorname{Res} \hat{\varphi}(p)),$$

where $\alpha,\beta\in\mathbf{C}$.

If an intertwining map $A(\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)\to L^2(X:\mathbf{C}^4))$ has its range contained in $\mathscr{H}^\sim$, then it defines an intertwining map $A^\sim(\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)\to\mathscr{H}^\sim)$ because $\mathscr{H}^\sim$ is an invariant subspace of $L^2(X:\mathbf{C}^4)$. Conversely, if $A^\sim(\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)\to\mathscr{H}^\sim)$ is an intertwining map, since $\mathscr{H}^\sim\subset L^2(X:\mathbf{C}^4)$ and the canonical imbedding $\iota(\mathscr{H}^\sim\to L^2(X:\mathbf{C}^4))$ commutes with the action of ISL(2,C), then $\iota\circ A^\sim(\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)\to L^2(X:\mathbf{C}^4))$ is intertwining. This means that *the maps $A^\sim$ are exactly the maps $A(\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)\to L^2(X:\mathbf{C}^4))$ whose range is in $\mathscr{H}^\sim$*. They are of the form

$$(A^\sim\varphi)(p) = \beta p(p\cdot \operatorname{Res} \hat{\varphi}(p)), \quad \beta\in\mathbf{C}.$$

We notice that these maps have range in $\mathscr{H}^0$, hence, in particular, all the intertwining maps from $\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)$ to $\mathscr{H}^0$ are of this form. Since $\mathscr{H}^{\text{photon}}=\mathscr{H}^\sim/\mathscr{H}^0$ we get immediately that *there are no intertwining maps from $\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)$ to $\mathscr{H}^{\text{photon}}$*, other than the trivial one.

Now we consider the problem of defining the free quantum EM, field. According to our discussion of Sec. V we define the *physical content* of the EM field by fixing the pair $(U^{\text{photon}},\mathscr{H}^{\text{photon}})$. We consider the Fock space $\mathscr{F}_S(\mathscr{H}^{\text{photon}})$ and the creation and annihilation operators defined with respect to the invariant scalar product of $\mathscr{H}^{\text{photon}}$. Then, a free quantum field that transforms covariantly as a vector field can be defined on $\mathscr{F}_S(\mathscr{H}^{\text{photon}})$ *if and only if* there exists an intertwining map from $\mathscr{S}_{\mathbf{C}^4}(\mathbf{R}^4)$ to $\mathscr{H}^{\text{photon}}$. We have seen that there exists no nontrivial map. This shows that *it is not possible to define a covariant quantum field acting in the Fock space $\mathscr{F}_S(\mathscr{H}^{\text{photon}})$*.

Now we discuss $(U^\sim,\mathscr{H}^\sim)$. By Theorem 4.1 the indefinite metrics for $(U^\sim,\mathscr{H}^\sim)$ are in 1-1 correspondence with the indefinite metrics for W^3. Since there exists no indefinite metric for W^3, there is no indefinite metric for $(U^\sim,\mathscr{H}^\sim)$. This shows that it is not possible to define unambiguously a pair of creation and annihilation operators on $\mathscr{F}_S(\mathscr{H}^\sim)$. Therefore *it is not possible to define a quantum field theory on $\mathscr{F}_S(\mathscr{H}^\sim)$*.

We are led to consider the extension $(U,L^2(X:\mathbf{C}^4))$ of $(U^{\text{photon}},\mathscr{H}^{\text{photon}})$. We fix $\mathscr{F}_S(L^2(X:\mathbf{C}^4))$ and the operators $a^*(f)$ and $a(f)$, for $f\in L^2(X:\mathbf{C}^4)$, using the indefinite metric (8.1). Then, following our discussion contained in Sec. V we define

$$A(\varphi) = \Xi(a \operatorname{Res} \hat{\varphi} + bp(p\cdot \operatorname{Res} \hat{\varphi})),$$

$$\varphi\in\mathscr{S}_{\mathbf{R}^4}(\mathbf{R}^4), \quad a,b\in\mathbf{R}, \quad a\neq 0.$$

Varying a,b we get all and only the covariant quantum fields that transform as four-vectors and acts on $\mathscr{F}_S(L^2(X:\mathbf{C}^4))$. They are the generalized Gupta–Bleuler gauges in the sense of Ref. 10.

ACKNOWLEDGMENTS

One of us (V.S.V.) is grateful for the hospitality he received from the INFN and the Department of Physics of the University of Genova in Genova, from the Mathematical Institute of the University of Utrecht in Utrecht, and from the Mathematical Institute of the University of Copenhagen in Copenhagen. We are also grateful to Professor Kolk and Professor de Meijer of the University of Utrecht, Professor Niels Vigand Pedersen and Professor Flemming Topsøe of the University of Copenhagen, and Doctor Saracco of the INFN of Genova for their help in the computer preparation of the manuscript of this paper.

APPENDIX A. *E-MODULES*

1. Introduction

In this appendix we shall make a few remarks on the continuous finite dimensional representations of E. The term representation will be used only for these and by a E module we shall mean one defined by a representation. In view of the Krull–Schmidt theorem[16] which asserts that any representation is a direct sum of indecomposable ones, the indecomposable constituents being essentially uniquely determined, it is enough to treat the indecomposable representations. We shall construct an infinite family of indecomposable E modules and give a classification of all indecomposable modules up to dimension 4.

2. Basic properties of $\mathscr{E}$ modules

Since

$$E = \left\{\begin{pmatrix} z & a \\ 0 & z^{-1}\end{pmatrix}: z,a\in\mathbf{C}, |z|=1\right\},$$

Lie $(E):=\mathscr{E}_0$ is spanned over $\mathbf{R}$ by the basis $\{H',X',Y'\}$ where

$$H' = \begin{pmatrix} i & 0 \\ 0 & -i\end{pmatrix}, \quad X' = \begin{pmatrix} 0 & 1 \\ 0 & 0\end{pmatrix}, \quad Y' = \begin{pmatrix} 0 & i \\ 0 & 0\end{pmatrix}$$

with commutation rules

$$[H',X'] = 2Y', \quad [H',Y'] = -2X', \quad [X',Y'] = 0.$$

We introduce the complexification $\mathscr{E}=\mathbf{C}\otimes_{\mathbf{R}}\mathscr{E}_0$ of $\mathscr{E}_0$ and put

$$H = -i\otimes H', \quad X = \tfrac{1}{2}(1\otimes X' - i\otimes Y'),$$

$$Y = \tfrac{1}{2}(1\otimes X' + i\otimes Y').$$

Then $\{H,X,Y\}$ is a $\mathbf{C}$ basis for $\mathscr{E}$ with commutation rules

$$[H,X] = 2X, \quad [H,Y] = -2Y, \quad [X,Y] = 0.$$

It follows from the compactness of the rotation subgroup of E that the category of E modules is identical with the category of finite dimensional $\mathscr{E}$ modules in which H acts semisimply with integer eigenvalues. The term $\mathscr{E}$ module will be used only for these. If V is any $\mathscr{E}$ module we shall write V_μ ($\mu\in\mathscr{Z}$) for the eigenspace in V of H for the eigenvalue μ. Also we put

$$V^0 = \{v\in V: X\cdot v = 0,\ Y\cdot v = 0\}.$$

Proposition A1: Let V be any $\mathscr{E}$ module. Then (a) X and Y are commuting nilpotents in $\operatorname{End}(V)$; (b) $X(V)\subset V_{\mu+2}$, $Y(V_\mu)\subset V_{\mu-2}$; (c) $V^0\neq 0$ if $V\neq 0$; (d) if $\dim(V^0)=1$, V will be indecomposable.

Proof: (b) is immediate from the commutation rules and implies (a). If U is any vector space and $N\in\operatorname{End}(U)$ is nilpo-

tent, $U^N := \{u \in U \mid Nu = 0\}$ is $\neq 0$ if $U \neq 0$. So $V^X \neq 0$ if $V \neq 0$, and as Y commutes with X, V^X is stable under Y. Since Y is also nilpotent, $(V^X)^Y = V^0 \neq 0$. If $V = U_1 \oplus U_2$ where U_i are submodules, $(U_i)^0 \neq 0$ so that $V^0 = (U_1)^0 \oplus (U_2)^0$ has dimension at least 2. This proves (d).

For any integer a let $l(a)$ be the one-dimensional module $\mathbf{C}$ on which X, Y are 0 and H acts as the scalar a. Since any module V can be tensored by $l(a)$ so that the H spectrum changes from S to $S + a$, we can normalize V so that the smallest eigenvalue of H in V^0 is 0; if $\dim(V^0) = 1$, this means $H = 0$ on V^0 when V is normalized.

Let

$$D = \left\{\delta(z) = \begin{pmatrix} z & 0 \\ 0 & z^{-1} \end{pmatrix} : |z| = 1\right\},$$

$$A = \left\{\alpha(a) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathbf{C}\right\}$$

so that $E = D \times' A$ and

$$\begin{pmatrix} z & a \\ 0 & z^{-1} \end{pmatrix} = \delta(z)\alpha(z^{-1}a).$$

For any integer μ let $R(\mu)$ be the space of rational functions f on E such that

$$f(\delta(z)g) = z^{\mu} f(g) \quad [\delta(z) \in D, g \in E].$$

The restriction map $f \mapsto f|_A$ allows us to identify $R(\mu)$ with the space $\mathscr{A}$ of functions on $\mathbf{C}$ that are polynomials in a_1 and a_2 ($a = a_1 + ia_2$ as usual). The action of E on $R(\mu)$ by right translation then gives an action on $\mathscr{A}$; we write $\mathscr{A}(\mu)$ for $\mathscr{A}$ equipped with this action

$$\alpha(b): \varphi(a) \mapsto \varphi(a + b),$$

$$\delta(z): \varphi(a) \mapsto \varphi(z^{-2}a)z^{\mu}.$$

The action of $\mathscr{E}_0$ on $\mathscr{A}$ is then computed to be:

$$X^1: \frac{\partial}{\partial a_1}, \quad Y^1: \frac{\partial}{\partial a_2},$$

$$H^1: 2\left(a_2 \frac{\partial}{\partial a_1} - a_1 \frac{\partial}{\partial a_2}\right) + i\mu.$$

As usual, let

$$\frac{\partial}{\partial a} = \frac{1}{2}\left(\frac{\partial}{\partial a_1} - i\frac{\partial}{\partial a_2}\right),$$

$$\frac{\partial}{\partial \bar{a}} = \frac{1}{2}\left(\frac{\partial}{\partial a_1} + i\frac{\partial}{\partial a_2}\right).$$

Then

$$H: -2\left(a\frac{\partial}{\partial a} - \bar{a}\frac{\partial}{\partial \bar{a}}\right) + \mu,$$

$$X: \frac{\partial}{\partial a}, \quad Y: \frac{\partial}{\partial \bar{a}}.$$

For $A(\mu)$ we can take the basis

$$(r,s) = \frac{a^r}{r!}\frac{\bar{a}^s}{s!}.$$

In this basis X, Y, H act as

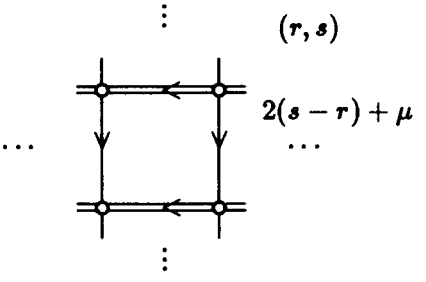
$$X: (r,s) \to (r-1,s), \quad (0,s) \to 0,$$

$$Y: (r,s) \to (r,s-1), \quad (r,0) \to 0,$$

$$H: (r,s) \to (2(s-r) + \mu)\cdot(r,s).$$

We represent (r,s) by the corresponding point in $\mathbf{Z}^2 \subset \mathbf{R}^2$ and write

for the actions of X and Y, respectively. We then get the following graphical representation:



where the label under a vertex gives the eigenvalue for the action of H.

3. Modules V with $\dim(V^0) = 1$

If $\dim(V^0) = 1$, V is indecomposable and we have an imbedding $V \hookrightarrow \mathscr{A}(\mu)$. We shall now determine completely all E-submodules of $\mathscr{A}(\mu)$. Let us consider the set $\mathbf{I}$ of all sequences

$$h = (h_r)_{r \geqslant 0},$$

$$h_0 \geqslant h_1 \geqslant \cdots \geqslant 0, \quad h_r \text{ integer and } = 0 \text{ for } r \gg 0.$$

Define

$$\mathscr{A}(\mu: h) = \bigoplus_{r \geqslant 0, 0 \leqslant s \leqslant h_r} \mathbf{C}\cdot(r,s).$$

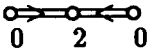
It is immediate that $\mathscr{A}(\mu: h)$ is a submodule and

$$\dim \mathscr{A}(\mu: h) = \sum_r h_r.$$

Proposition A2: We have (a) the $\mathscr{A}(\mu: h)$ (as h varies) are precisely all the submodules of $\mathscr{A}(\mu)$ (of finite dimension); $\mathscr{A}(\mu:h)^0 = \mathbf{C}\cdot(0,0)$; (b) $\mathscr{A}(\mu: h) \cong \mathscr{A}(\mu: k)$ if and only if $h = k$.

In particular, all the $\mathscr{A}(\mu:h)$ are indecomposable, and an E module V with $\dim(V^0) = 1$ is isomorphic to exactly one $\mathscr{A}(\mu:h)$.

Remark: Proposition A2 leads to graphs that describe the modules V with $\dim(V^0) = 1$. One must exercise some care in constructing modules out of more general graphs, especially when the eigenvalues of H have multiplicities. For instance the graph



defines an $\mathscr{E}$ module with basis e_α ($\alpha = 1,2,3$) such that

$$Ye_\alpha = 0, \quad Xe_1 = Xe_3 = e_2, \quad Xe_2 = 0,$$
$$He_1 = He_3 = 0, \quad He_2 = 2e_2.$$

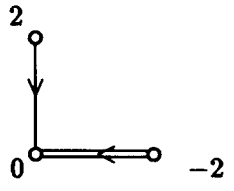
But although the graph is connected, the module is decomposable; indeed it is the direct sum of the submodules $\mathbf{C}\cdot(e_1 - e_3)$ and $\mathbf{C}\cdot e_2 \oplus \mathbf{C}\cdot e_3$. Of course if the graph is not connected, the module is decomposable.

4. The category S of $\mathscr{E}$ modules

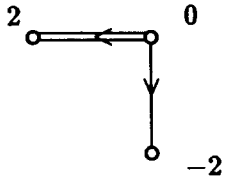
The class of $\mathscr{E}$ modules V with $\dim(V^0) = 1$ is not closed under passage to duals. For instance, let

$$V = \mathscr{A}(0{:}\ (2,1,0,\ldots))$$

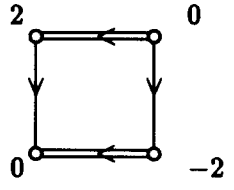
so that its graph is



A simple calculation shows that V^* corresponds to the graph



for V^* we have $\dim(V^*)^0 = 2$. It is clear that the graph of V^* is obtained from



by deleting the vertex with 0 eigenvalue on the bottom line, as well as the arrows going to it. This corresponds to the relation

$$V^* \cong \mathscr{A}(0{:}\ (2,2,0,\ldots))/\mathscr{A}(0{:}\ (1,0,\ldots)).$$

This suggests that one should try to enlarge the class of modules by adding to it all subquotients. We shall now proceed to give a more precise description of this extended family. Our analysis will show that it is closed under passage to duals.

Let $\mu \in Z$. Then $\mathscr{A}(\mu{:}k) \subset \mathscr{A}(\mu{:}h)$ $(h,k \in I)$ if and only if $k_r \leqslant h_r$ for all $r \geqslant 0$. We shall write

$$k \ll h$$

if the following conditions are satisfied: (a) $k_r \leqslant h_r$, for all $r \geqslant 0$; (b) there exist a,b, $0 \leqslant a < b$, such that

$$h_r = k_r, \quad r < a \text{ or } r > b,$$
$$h_{r+1} > k_r, \quad a \leqslant r < b.$$

It is immediate from the second condition in (b) that

$$k_r < h_r, \quad a \leqslant r \leqslant b.$$

The significance of this definition is clear from the following proposition.

Proposition A3: The condition $k \ll h$ is equivalent to the indecomposability of the module $\mathscr{A}(\mu{:}h)/\mathscr{A}(\mu{:}k)$.

Remark: We do not know whether the construction given here produces all indecomposable modules. We write $\mathscr{S}$ for the class of indecomposables of the form $\mathscr{A}(\mu{:}h)/\mathscr{A}(\mu{:}k)$. In low dimension this can be checked.

5. Classification in dimension $\leqslant 4$

By brutal calculation we can compute *all* indecomposables if the dimension is not high.

Proposition A4: All indecomposable modules for $\mathscr{E}$ in dimension $d \leqslant 4$ are of the class $\mathscr{S}$ and have the following graphs, the module $\mathscr{A}(\mu{:}h)/\mathscr{A}(\mu{:}k)$ being abbreviated as $(\mu{:}h{:}k)$ and $\mathscr{A}(\mu{:}h)$ as $(\mu{:}h)$.

$d = 1$

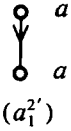
$(a{:}\ (1,0,\ldots))$

(a_1^1)

$d = 2$

$(a{:}\ (1,1,0,\ldots))$

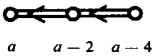
(a_1^2)



$(a{:}\ (2,0,\ldots))$

$(a_1^{2'})$

$d = 3$

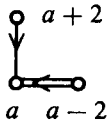


$(a{:}\ (1,1,1,0,\ldots))$

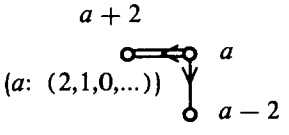
(a_1^3)

$(a{:}\ (3,0,\ldots))$

$(a_1^{3'})$



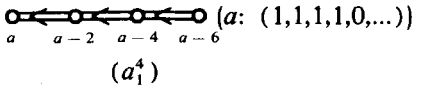
(a_2^3)



$(a_1^{3'})$

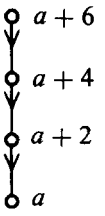
$(a{:}\ (2,2,0,\ldots){:}(1,0,\ldots))$

$d = 4$



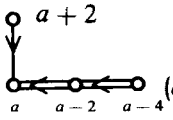
$(a: (1,1,1,1,0,\ldots))$

(a_1^4)



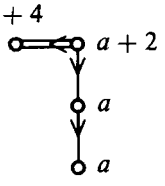
$(a: (4,0,\ldots))$

$(a_1^{4'})$



$(a: (2,1,1,0,\ldots))$

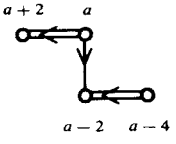
(a_2^4)



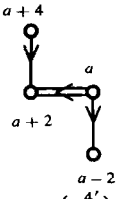
$(a: (3,3,0,\ldots): (2,0,$

$(a_2^{4'})$

$(a: (2,2,1,0,\ldots): (1,0,\ldots))$



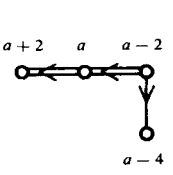
(a_3^4)



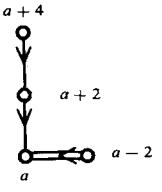
$(a: (3,2,0,\ldots): (1,0,\ldots))$

$(a_3^{4'})$

$(a: (2,2,2,0,\ldots): (1,1,0,\ldots))$

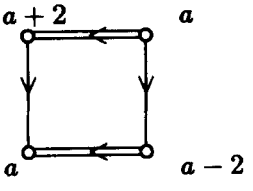


(a_4^4)



$(a: (3,1,0,\ldots))$

$(a_4^{4'})$



$(a: (2,2,0,\ldots))$

(a_5^4)

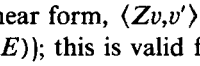
6. Modules admitting an indefinite metric in dimension $\leqslant 4$

It is easy to decide which of the above modules admit W^2 (photon representation) as subquotient and admit an indefinite metric. If V is an E module and $\langle \cdot, \cdot \rangle$ is an invariant sesquilinear form, $\langle Zv,v' \rangle + \langle v,Zv' \rangle = 0$ for $v,v' \in V$, $Z \in \mathscr{E}_0$ $(= \mathrm{Lie}(E))$; this is valid for $Z = H,X',X''$, and hence the condition in terms of H,X,Y becomes

$$\langle Hv,v' \rangle = \langle v,Hv' \rangle,$$

$$\langle Xv,v' \rangle + \langle v,Yv' \rangle = 0.$$

The first condition implies that the eigenspace for H for distinct eigenvalue are mutually orthogonal. In particular, if $\langle \cdot, \cdot \rangle$ is nondegenerate, it remains nondegenerate when restricted to any of the eigenspaces for H. Now W^2 has the graph



and the following modules admit W^2 as a subquotient (we use the notation $U \twoheadrightarrow V$ to indicate that V is a quotient of U):

$d=1,2$ None

$d=3$ $W^3=0_2^3$, $W^1=0_1^1$ (trivial module),
$W^1\subset W^3$, $W^3/W^1\simeq W^2$,
$(W^3)^*=0_2^{3'}$, $W^2\subset(W^3)^*$,

$d=4$ $0_2^4\supset W^3$, $2_2'\twoheadrightarrow(W^3)^*$,
$0_3^4\supset(W^3)^*$, $2_3^{4'}\twoheadrightarrow W^3$,
$2_3^4\twoheadrightarrow W^3$, $0_3^{4'}\supset(W^3)^*$,
$2_4^4\twoheadrightarrow(w^3)^*$, $0_4^{4'}(\supset(W^3$,
$0_5^4\supset W^3$,
$0_5^4\twoheadrightarrow(W^3)^*$.

However only 0_5^4 admits an indefinite metric. To exclude the others we use the following lemma.

Lemma A5: Suppose V is an indecomposable $\mathscr{E}$ module with two eigenvalues b, $b+2$ being simple. Then V does not admit an indefinite metric.

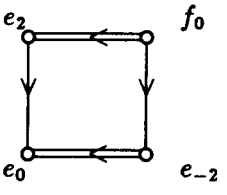
Proof: If both $X\cdot V_b$ and $Y\cdot V_{b+2}$ are zero, then V splits as the direct sum of the *submodules* $\oplus_{c<b}V_c$ and $\oplus_{c>b+2}V_c$ which are both different from zero. Let $0\neq u\in V_b$ and $0\neq v\in V_{b+2}$. The two possibilities are essentially

(a) $Xu=v$, $Yv=0$, (b) $Xu=0$, $Yv=u$

for otherwise either XY or YX will have a nonzero eigenvalue, contradicting its nilpotency. For an indefinite metric, $\langle Xw,w'\rangle+\langle w,Yw'\rangle=0$ which is impossible under both (a) and (b), in fact, assuming both (a) and (b), we get

$\langle Xu,v\rangle=\langle v,v\rangle\neq 0$, $\langle u,Yv\rangle=0$,
$\langle Xu,v\rangle=0$, $\langle u,Yv\rangle=\langle u,u\rangle\neq 0$.

To construct all indefinite metrics for 0_5^4 we write it as spanned by basis vectors e_2, e_0, e_{-2}, f_0 with



Then the indefinite metrics are given by

$\langle e_0,e_0\rangle=0$, $\langle e_2,e_2\rangle=\langle e_{-2},e_{-2}\rangle=\alpha$,
$\langle e_0,f_0\rangle=-\alpha$, $\langle f_0,f_0\rangle=\beta$,

where $\alpha\neq 0$ and β are arbitrary constants and all other scalar products are zero. It is trivial to verify that these are indefinite metrics. To see that there are no others, let $(\cdot,\cdot)$ be an indefinite metric. Then all scalar products other than the above ones are zero. Write $(e_2,e_2)=\alpha$, $(e_{-2},e_{-2})=\gamma$, $(f_0,f_0)=\beta$, $(e_0,f_0)=\gamma'$, $(e_0,e_0)=\delta$. Then

$$\delta=(e_0,e_0)=(e_0,Ye_2)=-(Xe_0,e_2)=0,$$
$$\gamma'=(e_0,f_0)=(Ye_2,f_0)$$
$$=-(e_2,Xf_0)=-(e_2,e_2)=-\alpha,$$
$$\gamma'=(e_0,f_0)=(Xe_{-2},f_0)$$
$$=-(e_{-2},Yf_0)=-(e_{-2},e_{-2})=-\beta.$$

Let $\langle\cdot,\cdot\rangle_{\alpha,\beta}$ denote the indefinite metric corresponding to fixed values of $\alpha\neq 0$ and β. Let us now take the representation of $SL(2,\mathbf{C})$ in $\mathbf{R}^4$ acting through δ and write $\mathbf{C}^4$ for the complexification of this representation. It then turns out that if we denote by W^4 the space $\mathbf{C}^4$ viewed as an E module, then

$0_5^4\cong W^4$.

Indeed, this isomorphism is given by

$$e_0\to\begin{pmatrix}1\\0\\0\\1\end{pmatrix},\quad e_2\to\begin{pmatrix}0\\1\\-i\\0\end{pmatrix},$$
$$e_{-2}\to\begin{pmatrix}0\\1\\i\\0\end{pmatrix},\quad f_0\to\begin{pmatrix}1\\0\\0\\-1\end{pmatrix},$$

as an easy calculation shows. Furthermore,

$\langle\cdot,\cdot\rangle_{-2,0}\cong$ standard Minkowski metric.

Summarizing we have the following proposition.

Proposition A6: There is, up to isomorphism, a unique indecomposable $\mathscr{E}$-module admitting an indefinite metric and extending W^2, in dimension $\leqslant 4$. This is the module W^4 of dimension 4, obtained by restricting the standard four-dimensional representation of $SL(2,\mathbf{C})$ in $\mathbf{C}^4$ to the subgroup E; $W^4\cong 0_5^4$. The indefinite metrics form a two-parameter family $\langle\cdot,\cdot\rangle_{\alpha,\beta}$ with $\alpha\neq 0$, and the standard Minkowski metric is obtained when $\alpha=-2$, $\beta=0$. In the basis $\{e_0,e_2,e_{-2},f_0\}$, the matrix of $\langle\cdot,\cdot\rangle_{\alpha,\beta}$ is given by

$$\langle\cdot,\cdot\rangle_{\alpha,\beta}\sim\begin{pmatrix}0&0&0&-\alpha\\0&\alpha&0&0\\0&0&\alpha&0\\-\alpha&0&0&\beta\end{pmatrix}.$$

APPENDIX B. SL(2, C)

We discuss some aspects of the finite dimensional representation theory of $SL(2,\mathbf{C})$. The holomorphic irreducible representation are the $\mathscr{D}^j$, $j\geqslant 0$ integer, with $\dim(\mathscr{D}^j)=j+1$; $\mathscr{D}^j$ is the representation obtained from the natural action of $SL(2,\mathbf{C})$ on the space of polynomial functions on $\mathbf{C}^2$ that are homogeneous of degree j. If

$$\mathsf{H}=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad \mathsf{X}=\begin{pmatrix}0&1\\0&0\end{pmatrix},\quad \mathsf{Y}=\begin{pmatrix}0&0\\1&0\end{pmatrix}$$

then $\{\mathsf{H},\mathsf{X},\mathsf{Y}\}$ is a $\mathbf{C}$ basis for the Lie algebra of $SL(2,\mathbf{C})$; and the infinitesimal representation, also denoted by $\mathscr{D}^j$, is $\mathbf{C}$ linear. If we write

$$u_k^{(j)}:=u_k:=\frac{(-1)^k}{\sqrt{k!}\sqrt{(j-k)!}}z_1^k z_2^{j-k},$$

$k=0,1,\ldots,j$ ($u_{-1}=0$ and $u_{j+1}=0$ by convention), the $\{u_k\}$ is a basis for the space of $\mathscr{D}^j$ which can be identified with $\mathbf{C}^j$. We have

$$\mathscr{D}^j(\mathsf{H})u_k = (j-2k)u_k,$$
$$\mathscr{D}^j(\mathsf{X})u_k = \sqrt{k(j-k+1)}\,u_{k-1},$$
$$\mathscr{D}^j(\mathsf{Y})u_k = \sqrt{(j-k)(k+1)}u_{k+1},$$

from which it follows that on the Lie algebra of SU(2), the matrices of $\mathscr{D}^j$ are skew Hermitian. So, in the usual scalar product for $\mathbf{C}^j$, $\mathscr{D}^j(h_0)$ is *unitary* for $h_0 \in \mathrm{SU}(2)$. The continuous (i.e., real analytic) irreducible representations of SL(2,C) are the $\mathscr{D}^{j,k}$ ($j,k \geqslant 0$, integers)

$$\mathscr{D}^{j,k}(h) = \mathscr{D}^j(h) \otimes \mathscr{D}^k(\bar{h})$$

($\bar{h}$ = complex conjugate of h).

The $\mathscr{D}^{j,k}$ are all mutually inequivalent and $\mathscr{D}^{j,k}$ is of dimension $(j+1)(k+1)$. As we shall see presently we are interested in the $\mathscr{D}^{j,j}$.

Restriction to SU(2): $\mathscr{D}^j$ restricts to the representation d_j with spin $\frac{1}{2}j$. Thus, by the Clebsch–Gordan formula,

$$\mathscr{D}^{j,k}|_{\mathrm{SU}(2)} \sim d^{|j-k|} \oplus d^{|j-k|+2} \oplus \cdots \oplus d^{j+k}.$$

In particular, (I) $\mathscr{D}^{j,k}|_{\mathrm{SU}(2)}$ contains the trivial representation of SU(2) if and only if $j=k$; and then it contains the trivial representation exactly once.

Restriction to E: We have the following: (II) $\mathscr{D}^{j,k}|_E$ contains the trivial representation of E if and only if $j=k$; and then it contains the trivial representation exactly once.

To prove this we note that $\{i\mathsf{H},\mathsf{X},i\mathsf{X}\}$ is an $\mathbf{R}$ basis of the Lie algebra of E, and we want the condition for the existence of a nonzero vector θ such that $\mathscr{D}^{j,k}(a)\theta = 0$, $\forall \alpha \in \mathrm{Lie}(E)$,

$$\mathscr{D}^{j,k}(\alpha) = \mathscr{D}^j(\alpha) \otimes \mathbf{I} + \mathbf{I} \otimes \mathscr{D}^k(\bar{a}),$$
$$a \in \mathrm{Lie}(\mathrm{SL}(2,\mathbf{C})),$$

the condition reduces to

$$(\mathscr{D}^j(\mathsf{H}) \otimes \mathbf{I} - \mathbf{I} \otimes \mathscr{D}^k(H))\theta = 0,$$
$$(\mathscr{D}^j(\mathsf{X}) \otimes \mathbf{I})\theta = 0, \quad (\mathbf{I} \otimes \mathscr{D}^k(\mathsf{X}))\theta = 0;$$

and so, from the first we get

$$\theta = \sum_{j-2\alpha = k-2\beta} c_{\alpha\beta} u^j_\alpha \otimes u^k_\beta.$$

A simple calculation shows that $\theta = 0$ unless $j=k$, and when $j=k$;

$$\theta = \mathrm{const}\; u^j_0 \otimes u^j_0 \qquad \text{(B1)}$$

(using the remaining relations).

Explicit realizations: We shall confine ourselves only to the $\mathscr{D}^{j,j}$.

(1) As before $\mathscr{D}^j$ acts on C^j. $\mathscr{D}^{j,j}$ acts on the space of endomorphisms of C^j, i.e., on the space of $j \times j$ complex matrices M. The action is

$$\mathscr{D}^{j,j}(h): M \to \mathscr{D}^j(h) M \mathscr{D}^j(h)^*.$$

The vectors fixed by SU(2) are the scalar multiples of the identity. The vectors fixed by E are the multiples of

$$\begin{pmatrix} 1 & \\ & \end{pmatrix}.$$

(2) We consider polynomials on $\mathbf{P}^4$, i.e., polynomials f in p_μ ($\mu = 0,1,2,3$) which satisfy (a) f is homogeneous of degree j; (b) f is harmonic, i.e., $\Box f = 0$.

We have SL(2,C) acting via δ and the above space is carried into itself by this action. The representation of SL(2,C) thus defined is known to be precisely $\mathscr{D}^{j,j}$. We need only the following.

Proposition 1: Let W be a module for SL(2,C) (with representation u) and suppose that $f(X_0^+ \to W)$ is a function such that

$$f(\delta(h)p) = u(h)f(p), \quad h \in \mathrm{SL}(2,\mathbf{C}), \quad p \in X_0^+.$$

Then there exists a polynomial map F: $\mathbf{P}^4 \to W$ such that

$$F|_{X_0^+} = f.$$

Proof: We identify (p_0,p_1,p_2,p_3) with the matrix

$$\begin{pmatrix} p_0+p_3 & p_1+ip_2 \\ p_1-ip_2 & p_0-p_3 \end{pmatrix}$$

as usual. Let $p^+ = (1,0,0,1)$ so that E is the stabilizer of p^+ and $X_0^+ \simeq H/E$. The condition on f shows that $f(p^+)$ is fixed by $u(h_0)$, for $h_0 \in E$. Splitting u as a direct sum of $\mathscr{D}^{j,k}$ and noting that $\mathscr{D}^{j,k}$ with $j \neq k$ do not contribute to $f(p^+)$ by (II), we see that it is enough to consider the case $u = \mathscr{D}^{j,j}$ for some $j \geqslant 0$. By (B1) we may assume that $f(p^+) = u^j_0 \oplus u^j_0$. Hence

$$f(\delta(h)p^+) = \mathscr{D}^j(h)u^j_0 \otimes \mathscr{D}^j(\bar{h})u^j_0.$$

Let

$$h = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

and recall hat $\mathscr{D}^j(h)$ acts on the homogeneous polynomial $P(z)$ by $P(g^{-1}z)$, where

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \mathbf{C}^2.$$

Hence $\mathscr{D}^j(h) \cdot z_2^j = (-\gamma z_1 + \alpha z_2)^j$ showing that

$$\mathscr{D}^j(h)u^j_0 = \sum_{0 \leqslant k \leqslant j} c_k \gamma^k \alpha^{j-k} u^j_k,$$

where c_k are constants independent of h. Hence $\mathscr{D}^j(h)u^j_0 \otimes \mathscr{D}^j(\bar{h})u^j_0$ is a linear combination of

$$\gamma^k \bar{\gamma}^l \alpha^{j-k} \bar{\alpha}^{j-l} \cdot (u^j_k \otimes u^j_l), \quad 0 \leqslant k,l \leqslant j,$$

with coefficients which are independent of h. It is thus a question of showing that the functions

$$f_{kl}\left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\right) = \gamma^k \bar{\gamma}^l \alpha^{j-k} \bar{\alpha}^{j-l}$$

actually are defined on $H/E \simeq X_0^+$ and are restrictions to X_0^+ of polynomials on $\mathbf{P}^4$. Changing h to hh_0, $h_0 \in E$ changes α to $\alpha\varepsilon$ and γ to $\gamma\varepsilon$, with $|\varepsilon| = 1$; and it is clear that $f_{kl}(h)$ is the same as $f_{kl}(hh_0)$. On the other hand,

$$\delta(h)p^+ \sim 2h \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} h^*,$$

so that $\delta(h)p^+ = (p_0,p_1,p_2,p_3)$ with

$$p_0 + p_3 = 2\alpha\bar{\alpha}, \quad p_0 - p_3 = 2\gamma\bar{\gamma}, \quad p_1 + ip_2 = 2\alpha\bar{\gamma}.$$

It is thus a question of showing that f_{kl} is a polynomial in these quantities. We may suppose that $k \leqslant l$ by symmetry. Then

$$\gamma^k\bar{\gamma}^l\alpha^{j-k}\bar{\alpha}^{j-l} = (\gamma\bar{\gamma})^k(\alpha\bar{\alpha})^{j-1}(\alpha\bar{\gamma})^{l-k}$$

$$= \left(\frac{p_0 - p_3}{2}\right)^k \left(\frac{p_0 + p_3}{2}\right)^{j-l} \left(\frac{p_1 + ip_2}{2}\right)^{l-k}.$$

Remark: It can be shown that F can be chosen to be harmonic, i.e., $\Box F = 0$. It is then even uniquely determined.

A similar result is also true for X_m^+ $(m > 0)$.

Proposition 2: Let W be a module for SL(2,C) (with representation u) and suppose that $f(X_m^+ \to W)$ $(m > 0)$, is a function such that

$$f(\delta(h)p) = u(h)f(p), \quad h \in \mathrm{SL}(2,\mathbf{C}), \quad p \in X_m^+.$$

Then there exists a polynomial map F: $\mathbf{P}^4 \to W$ such that

$$F|_{X_m^+} = f.$$

Proof: The base point is now $p^+ = (m,0,0,0) \sim m\mathbf{I}$. As before we assume $u \simeq \mathscr{D}^{jj}$ and use the matrix model to conclude that $f(p^+) = 1$. Then f is a function $h \mapsto \mathscr{D}^j(h)\mathscr{D}^j(h)^*$. We saw that $\mathscr{D}^j(a)$ is skew Hermitian for $a \in \mathrm{Lie(SU(2))}$ so that $\mathscr{D}^j(b)^* = \mathscr{D}^j(b^*)$ for $b \in \mathrm{Lie(SL(2,C))}$. Thus $\mathscr{D}^j(h)^* = \mathscr{D}^j(h^*)$ for $h \in \mathrm{SL}(2,\mathbf{C})$, so that $f(h) = \mathscr{D}^j(hh^*)$. But the matrix entries of $\mathscr{D}^j(h)$ are polynomials in the matrix entries of h so that the entries of $f(h)$ are polynomials in the entries of hh^*. Since $\delta(h)p^+ = mhh^*$, the proposition is proven.

Remark: Here again F can be chosen to be harmonic, and then it is unique.

[1] G. Rideau, J. Math. Phys. **19**, 1627 (1978).
[2] A. L. Carey and C. A. Hurst, Lett. Math. Phys. **2**, 227 (1978).
[3] M. Mintchev, Rep. Math. Phys. **17**, 257 (1980).
[4] D. Pierotti, J. Math. Phys. **26**, 143 (1985).
[5] H. Araki, Commun. Math. Phys. **97**, 149 (1985).
[6] V. S. Varadarajan, *Geometry of Quantum Theory*, 2nd ed. (Springer, New York, 1985).
[7] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[8] U. H. Niederer and L. O'Raifeartaigh, Fortschr. Phys. **22**, 111, 131 (1974).
[9] N. Nakanishi, Prog. Theor. Phys. Suppl. **51**, 1 (1972).
[10] F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[11] T. Kugo and I. Ojima, Prog. Theor. Phys. Suppl. **66**, 1 (1979).
[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. 2.
[13] M. Mintchev, J. Phys. **13**, 1841 (1980).
[14] F. Bruhat, Bull. Soc. Math. France **84**, 97 (1956).
[15] N. Nakanishi, Prog. Theor. Phys. **49**, 640 (1973).
[16] N. Jacobson, *Lectures in Abstract Algebra* (Van Nostrand, Princeton, 1953), Vol. 2.

Linear realizations of the superrotation and super-Lorentz symmetries. I

Smain Kouadik and Pierre Minnaert
Laboratoire de Physique Théorique, Bordeaux, France

(Received 7 March 1989; accepted for publication 17 May 1989)

This paper is the first of a series in which various aspects of the superrotation and super-Lorentz symmetries will be studied. In this paper the superrotation algebra and the super-Lorentz algebra are defined. Their linear, finite-dimensional representations are classified. A realization of the low-dimensional representations is provided under the form of irreducible tensor operator multiplets.

I. INTRODUCTION

Supersymmetry has been introduced in theoretical physics by Gol'fand and Likhtman[1] when they showed that it is possible to add to the ten generators $(L_{\mu\nu},P_\rho)$ of the Poincaré algebra a set of four spinorial operators Q_α, whose anticommutation relations close on the translation operators P_ρ. Interest in this super-Poincaré symmetry was raised by the fact that each irreducible linear multiplet contains two states with the same mass but with spins differing by $\frac{1}{2}$. The two states are mixed by the action of the spinorial operators Q_α.

Following Wess and Zumino[2] and Salam and Strathdee,[3] the study of the scalar multiplet (with spins 0 and $\frac{1}{2}$) and of the vector multiplet (with spins $\frac{1}{2}$ and 1) has lead to many developments in supersymmetric field theory and to the concept of superspace and superfields.[4] The study of the tensor multiplet (with spins $\frac{3}{2}$ and 2) by Freedman, Van Nieuwenhuizen and Ferrara,[5] and by Deser and Zumino[6] has provided an extension of Einstein gravitation theory known as supergravity.[7] This theory has also been built by MacDowell and Mansouri[8] as a gauge theory based on the super-Poincaré group.

Although the physical manifestations of supersymmetry and supergravity are inexistent for the moment, the general idea of extending a bosonic symmetry to a supersymmetry, by addition of operators that satisfy anticommutation relations, remains very attractive. It may therefore be interesting and pedagogical to have at hand simple examples of supersymmetric extensions of some well known bosonic symmetries.

This paper (I) is the first of a series in which we study various aspects of the supersymmetric extension of the rotation and Lorentz symmetries. It is devoted to the definition of the algebras and to the classification and realizations of their linear finite-dimensional representations.

In the second paper (II) we define the corresponding superrotation and super-Lorentz groups, we study their actions in their vectorial representation spaces, and we define the corresponding inhomogeneous algebras.

The third paper (III) will be devoted to the nonlinear realizations of the superrotation and super-Lorentz groups.

Finally, in a fourth paper (IV) we shall show that the nonlinear action of the super-Lorentz group into Minkowski space and on the fields leads to a nonconventional theory of supergravity.

II. THE SUPERROTATION ALGEBRA (sr)

A. Definition of the superrotation algebra

Since the rotation algebra is so(3) with covering su(2), it would seem that the natural supersymmetric extension of this algebra is the eight-dimensional superunitary algebra su(2/1), which contains four bosonic and four fermionic generators. However, a smaller algebra exists that contains su(2) as a subalgebra. It is the superalgebra defined by Pais and Rittenberg,[9] which is a complex version of the orthosymplectic algebra osp(1/2). It contains only three bosonic generators J_a $(a=1,2,3)$, those of the rotation subalgebra, and two fermionic generators J_α $(\alpha=\pm\frac{1}{2})$. Its supercommutation relations are

$$\begin{aligned}&[J_a,J_b]=i\varepsilon_{abc}J_c,\\&[J_a,J_\alpha]=\tfrac{1}{2}J_\beta(\sigma_a)^\beta{}_\alpha,\\&\{J_\alpha,J_\beta\}=\tfrac{1}{2}J_a(\Gamma\sigma_a)_{\alpha\beta}\end{aligned}\tag{1}$$

where the σ_a are the Pauli matrices and $\Gamma=i\sigma_2$. We call this algebra the superrotation algebra (sr), its generators are denoted by J_A $(A=a,\alpha)$. Let us recall the main results obtained by Pais and Rittenberg.[9]

The irreducible linear representations are labeled by a spin J $(J=0,\frac{1}{2},1,\frac{3}{2},\ldots)$, and they will be denoted by J^{SR}. For $J\neq 0$, the decomposition of J^{SR} into irreducible representations of the rotation subalgebra (r) contains the representation J^{R} and $(J-\frac{1}{2})^{\mathrm{R}}$, i.e.,

$$J^{\mathrm{SR}}=J^{\mathrm{R}}\oplus(J-\tfrac{1}{2})^{\mathrm{R}}.\tag{2}$$

For $J=0$ one has $0^{\mathrm{SR}}=0^{\mathrm{R}}$.

The standard basis of the J^{SR} representation is formed by $(4J+1)$ vectors $|JIm\rangle$ with $I=J,J-\frac{1}{2}$ and $|m|<I$. The label J is related to the Casimir operator

$$K_2=J_aJ_a+J_\alpha\Gamma^{\alpha\beta}J_\beta,\tag{3}$$

such that

$$K_2|JIm\rangle=J(J+\tfrac{1}{2})|JIm\rangle.\tag{4}$$

B. The fundamental representation $(1/2)^{\mathrm{SR}}$

Pais and Rittenberg[9] have given explicitly the matrix elements of J_a and J_α in the standard basis of the rotation group for any representation J^{SR}. In this paper we shall not be concerned with such matrix representations; however, it is interesting to recall the structure of the fundamental representation $(\frac{1}{2})^{\mathrm{SR}}$.

 0022-2488/89/112708-06$02.50

In the standard basis the three-dimensional representation can be written $J_A = \frac{1}{2}\Sigma_A$, where the matrices $\Sigma_A = (\Sigma_a, \Sigma_\alpha)$ ar 3×3 matrices we call Pauli supermatrices. They are

$$\Sigma_a = \begin{bmatrix} 0 & & \\ & \sigma_a & \\ & & \end{bmatrix}, \quad \Sigma_{1/2} = \begin{bmatrix} 0 & 0 & -1 \\ -1 & & 0 \\ 0 & & \end{bmatrix},$$
$$\Sigma_{-1/2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & & 0 \\ -1 & & \end{bmatrix}. \tag{5}$$

They satisfy the properties

$$\Sigma_A^{\mathrm{ST}} = -G\Sigma_A G^{-1}, \ \Sigma_A^{\neq} = G_{AB}\Sigma_B, \ \mathrm{Str}(\Sigma_A\Sigma_B) = 2G_{AB}, \tag{6}$$

where G is the 3×3 matrix

$$G = \begin{bmatrix} 1 & & \\ & 0 & 1 \\ & -1 & 0 \end{bmatrix}. \tag{7}$$

We use the notation, $^{\mathrm{ST}}$ = supertransposition, $^{\neq}$ = super-Hermitian conjugation, and Str = supertrace.

C. Realization of the first representations

For each representation J^{SR} we can build a realization as a multiplet of $(4J+1)$ tensorial operators whose supercommutation relations with the generators J_A satisfy the super-Jacobi identities.

According to Eq. (2), the multiplet J^{SR} is decomposed into irreducible rotation multiplets $J^{\mathrm{R}} \oplus (J-\frac{1}{2})^{\mathrm{R}}$. So the commutation relation of these multiplets with the generators J_a are the usual relations satisfied by irreducible tensorial operators of the rotation algebra. Therefore, what characterizes the superrotation algebra are the commutation relations of the multiplet with the fermionic generators J_α. We shall see that the action of J_α mixes the two irreducible rotation multiplets.

(i) The trivial representation $J^{\mathrm{SR}} = 0$ contains only one generator, a *superscalar* T, which commutes with all generators J_A:

$$[J_A, T] = 0, \quad A = (a,\alpha). \tag{8}$$

(ii) The three-dimensional multiplet $\Psi = (S,\psi_\alpha)$ of the fundamental representation $J^{\mathrm{SR}} = \frac{1}{2}$ is called a *superspinor*. It contains a scalar S and a two-component spinor ψ_α. Their supercommutation relations with J_A read

$$\begin{aligned} &[J_a, S] = 0, \\ &[J_a, \psi_\alpha] = \tfrac{1}{2}\psi_\beta(\sigma_a)^\beta{}_\alpha, \\ &[J_\alpha, S] = \tfrac{1}{2}\psi_\alpha, \\ &\{J_\alpha, \psi_\beta\} = \tfrac{1}{2}\Gamma_{\alpha\beta}S. \end{aligned} \tag{9}$$

The last two relations show that the action of the fermionic generators J_α mixes the bosonic and fermionic components of the supermultiplet. Besides, it is easy to verify that the quadratic operator

$$X_{1/2} = S^2 + \psi_\alpha\Gamma^{\alpha\beta}\psi_\beta \tag{10}$$

built with the components of Ψ is a superscalar, that is, it commutes with all the generators, $[J_A, X_{1/2}] = 0$.

(iii) The five-dimensional multiplet $V = (\chi_\alpha, V_a)$ of the adjoint representation $J^{\mathrm{SR}} = 1$ is called a *supervector*. It contains a two-component spinor χ_α and a three-dimensional vector V_a, with supercommutation relations

$$\begin{aligned} &[J_a, V_b] = i\varepsilon_{abc}V_c, \\ &[J_a, \chi_\alpha] = \tfrac{1}{2}\chi_\beta(\sigma_a)^\beta{}_\alpha, \\ &[J_\alpha, V_a] = -\tfrac{1}{2}\chi_\beta(\sigma_a)^\beta{}_\alpha, \\ &\{J_\alpha, \chi_\beta\} = \tfrac{1}{2}V_a(\Gamma\sigma_a)_{\alpha\beta}. \end{aligned} \tag{11}$$

Again we notice that J_α mixes the bosonic and fermionic components of V, and that one can build a superscalar

$$X_1 = \chi_\alpha\Gamma^{\alpha\beta}\chi_\beta + V_aV_a, \tag{12}$$

such that $[J_A, X_1] = 0$.

(iv) The spin $\frac{3}{2}$ is described by a spinor–vector $R_{a\alpha}$ satisfying the condition

$$R_{a\alpha} = i\,\varepsilon_{abc}R_{b\beta}(\sigma_c)^\beta{}_\alpha, \tag{13}$$

which leaves four independent components. The seven-dimensional multiplet $\Phi = (R_{a\alpha}, W_b)$ is called *superspinor–vector*. It contains a spinor–vector $R_{a\alpha}$ and a vector W_b. Their supercommutation relations with J_A are

$$\begin{aligned} &[J_a, W_b] = i\varepsilon_{abc}W_c, \\ &[J_a, R_{b\alpha}] = i\varepsilon_{abc}R_{c\alpha} + \tfrac{1}{2}R_{b\beta}(\sigma_a)^\beta{}_\alpha, \\ &[J_\alpha, W_a] = \tfrac{1}{2}R_{a\alpha}, \\ &\{J_\alpha, R_{a\beta}\} = \Gamma_{\alpha\beta}W_a + (i/2)\varepsilon_{abc}(\Gamma\sigma_c)_{\alpha\beta}W_b. \end{aligned} \tag{14}$$

Note that the verification of the super-Jacobi identities requires the identity

$$(\Gamma\sigma_a)_{\alpha\beta}(\sigma_b)^\delta{}_\gamma - (\Gamma\sigma_b)_{\alpha\beta}(\sigma_a)^\delta{}_\gamma = i\varepsilon_{abc}(\Gamma\sigma_c)_{\alpha\beta}\delta^\delta\gamma. \tag{15}$$

D. Tensor product of two representations

The reduction formula for the tensor product of the representations J^{SR} and J'^{SR} is

$$J^{\mathrm{SR}} \otimes J'^{\mathrm{SR}} = \bigoplus_{p=0}^{4\,\mathrm{inf}(J,J')} (J + J' - p/2)^{\mathrm{SR}}, \tag{16}$$

which means that every finite-dimensional representation can be constructed from the fundamental three-dimensional representation $(\frac{1}{2})^{\mathrm{SR}}$. Let us consider two examples of this reduction formula.

(i) The tensor product of two *fundamental* representation reduces as

$$(\tfrac{1}{2})^{\mathrm{SR}} \otimes (\tfrac{1}{2})^{\mathrm{SR}} = 0^{\mathrm{SR}} \oplus (\tfrac{1}{2})^{\mathrm{SR}} \oplus (1)^{\mathrm{SR}}. \tag{17}$$

Therefore, given two superspinors $\Psi_1 = (S_1,\psi_1)$ and $\Psi_2 = (S_2,\psi_2)$ one can build a superscalar T, a superspinor $\Psi = (S,\psi)$, and a supervector $V = (\chi, V)$. Explicitly they are given by

$$\begin{aligned} &T = S_1S_2 + \psi_{1\alpha}\Gamma^{\alpha\beta}\psi_{2\beta}; \\ &\Psi = \begin{cases} S = S_1S_2 + \frac{1}{2}\psi_{1\gamma}\Gamma^{\gamma\delta}\psi_{2\delta}, \\ \psi_\alpha = \frac{1}{2}(S_1\psi_{2\alpha} + \psi_{1\alpha}S_2); \end{cases} \end{aligned} \tag{18}$$

$$V=\begin{cases}\chi_\alpha = S_1\psi_{2\alpha} - \psi_{1\alpha}S_2,\\ V_a = \psi_{1\gamma}(\sigma_a\Gamma)^{\gamma\delta}\psi_{2\delta}.\end{cases}$$

(ii) The reduction formula for the tensor product of the representations $(1)^{SR}$ and $(\frac{1}{2})^{SR}$ is

$$(1)^{SR}\otimes(\tfrac{1}{2})^{SR} = (\tfrac{3}{2})^{SR}\oplus(1)^{SR}\oplus(\tfrac{1}{2})^{SR}.$$

Then, given a supervector $V=(\chi_\alpha,V_a)$ and a superspinor $\psi=(S,\psi_\alpha)$ one can build a superspinor $\Lambda=(L,\lambda_\alpha)$, a supervector $U=(\varepsilon_\alpha,U_a)$, and a superspinor–vector multiplet $\Phi=(R_{a\alpha},W_b)$:

$$\Lambda=\begin{cases}L=\chi_\delta\Gamma^{\delta\gamma}\psi_\gamma,\\ \lambda_\alpha = V_a\psi_\gamma(\sigma_a)\gamma_\alpha - \chi_\alpha S;\end{cases}$$
$$U=\begin{cases}\varepsilon_\alpha = 3\chi_\alpha S - V_a\psi_\gamma(\sigma_a)\gamma_\alpha,\\ U_a = 2V_aS+\chi_\delta(\sigma_a\Gamma)^{\delta\gamma}\psi_\gamma;\end{cases}\qquad(19)$$
$$\Phi=\begin{cases}R_{a\alpha}=\frac{2}{3}V_a\psi_\alpha - (i/3)\varepsilon_{abc}(\sigma_b)^\beta{}_\alpha V_c\psi_\beta,\\ W_a=\frac{1}{3}(V_aS-\chi_\delta(\sigma_a\Gamma)^{\delta\gamma}\psi_\gamma).\end{cases}$$

It can be verified that the multiplets built in Eqs. (18) and (19) satisfy their respective commutation relations (9), (11), and (14).

III. THE SUPER LORENTZ ALGEBRA (sl)

Lorentz symmetry is a fundamental symmetry of physics. However, since Lorentz algebra is the homogeneous part of Poincaré algebra, the supersymmetric extension of which is known, it has not been considered an interesting problem for finding a supersymmetric extension of Lorentz algebra. However, Lukierski and Novicki[10] have defined the graded Lorentz algebras in three, four, and five dimensions and, in his book *Supermanifolds*, De Witt[11] has devoted some pages to the study of the super-Lorentz group.

A. Definition of the super-Lorentz algebra

It is well known that sl(2,c) is the complexification of the rotation algebra su(2). This means that it can be built by adding to the three generators J_a of su(2), three generators K_a defined formally by $K_a = iJ_a$. In much the same way we shall define the super-Lorentz algebra sl by complexification of the sr algebra.[12] To the five generators J_A $(A=a,\alpha)$, we add five generators K_A defined formally by $K_A = iJ_A$. If the supercommutation relations (1) of sr are written in the concise form

$$[J_A,J_B\} = J_C C^C_{AB},\qquad(20)$$

the supercommutation relations of J_A with K_B and of K_B with K_C read

$$[J_A,K_B\} = K_C C^C_{AB},$$
$$[K_A,K_B\} = -J_C C^C_{AB}.\qquad(21)$$

Equations (20) and (21) define the super-Lorentz algebra.

The structure of these relations suggests the definition of complex combinations of J_A and K_A:

$$M_A=\tfrac{1}{2}(J_A - iK_A),\quad N_A=\tfrac{1}{2}(J_A+iK_A).\qquad(22)$$

These satisfy the supercommutation relations

$$[M_A,M_B\} = M_C C^C_{AB},$$
$$[N_A,N_B\} = N_C C^C_{AB},\qquad(23)$$
$$[M_A,N_B\} = 0.$$

This means that the generators M_A and N_A form the superalgebra sr ⊗ sr. This remark allows us to build sl representations from tensor products of sr representations.

B. Representations of the super-Lorentz algebra

The super-Lorentz algebra has two quadratic Casimir operators. One of them is denoted by K_2^{SL} for the M_A superalgebra, and the other is denoted by $K_2'^{SL}$ for the N_A superalgebra.

A finite-dimensional irreducible representation will be characterized by two spins J and J', and denoted by $(J,J')^{SL}$. Its dimension is $(4J+1)\times(4J'+1)$. The lowest-dimensional representations are listed in Table I.

In this paper we shall not be interested in matrix realizations of super-Lorentz algebra representations. We only give the matrix realization of the two fundamental representations $(\frac{1}{2},0)^{SL}$ and $(0,\frac{1}{2})^{SL}$. They can be written with the help of the Pauli supermatrices defined in Eq. (5):

$$(\tfrac{1}{2},0)^{SL}:\ J_A=\tfrac{1}{2}\Sigma_A,\quad K_A=(i/2)\Sigma_A,$$
$$(0,\tfrac{1}{2})^{SL}:\ J_A=\tfrac{1}{2}\Sigma_A,\quad K_A=-(i/2)\Sigma_A.\qquad(24)$$

The super-Lorentz algebra contains several subalgebras; thus any representation may be decomposed into irreducible representations of these subalgebras.

(a) For the sr subalgebra, which is generated by J_A, from (16) one has the decomposition

$$(J,J')^{SL} = \bigoplus_{p=0}^{4\inf(J,J')}(J+J'-p/2)^{SR}.\qquad(25)$$

(b) For the Lorentz subalgebra, generated by (J_a,K_a), from (2) one has the decomposition into at most four representations:

$$(J,J')^{SL} = (J,J')^{L}\oplus(J,J'-\tfrac{1}{2})^{L}$$
$$\oplus(J-\tfrac{1}{2},J')^{L}\oplus(J-\tfrac{1}{2},J'-\tfrac{1}{2})^{L},\qquad(26)$$

where $(J,J')^{L}$ are the usual representations of the Lorentz algebra.

(c) For the rotation subalgebra, generated by J_a, one has the decomposition

TABLE I. Notation and dimension for the first representations of the super-Lorentz algebra.

Representation	Notation	Dimension
Superscalar	$(0,0)^{SL}$	1
Fundamental or spinorial	$(\frac{1}{2},0)^{SL}$ $(0,\frac{1}{2})^{SL}$	3
Super-Dirac	$(\frac{1}{2},0)^{SL}\oplus(0,\frac{1}{2})^{SL}$	6
Supervectorial	$(\frac{1}{2},\frac{1}{2})^{SL}$	9
Adjoint	$(1,0)^{SL}\oplus(0,1)^{SL}$	10

$$(J,J')^{SL} = (J+J')^{R} \oplus 2 \overset{2\inf(J,J')}{\underset{p=0}{\oplus}} (J+J'-p/2)^{R} \oplus (|J-J'|-\tfrac{1}{2})^{R}. \quad (27)$$

If $J = J'$, the last representation $(|J-J'|-\frac{1}{2})^{R}$ does not appear in Eq. (27).

C. Covariant form of the super-Lorentz algebra

The generators (J_A, K_A) belong to the adjoint representation, whose reduction onto the Lorentz subalgebra contains a skew-symmetric tensor and a four-component spinor. Indeed, the six Lorentz algebra generators (J_a, K_a) can be gathered into the antisymmetric tensor $L_{\mu\nu} = -L_{\nu\mu}$ $(\mu,\nu = 0,1,2,3)$ defined by

$$L_{ab} = \varepsilon_{abc} J_c, \quad L_{a0} = K_a, \quad (28)$$

and the four fermionic generators $(J_{\pm 1/2}, K_{\pm 1/2})$ or $(M_{\pm 1/2}, N_{\pm 1/2})$ can be gathered into the spinor l_α $(\alpha = 1,2,3,4)$, defined by

$$l_\alpha = \sqrt{2}(M_{1/2}, M_{-1/2}, N_{1/2}, N_{-1/2}). \quad (29)$$

Then, using for Dirac matrices the Weyl representation in which one has

$$\gamma_0 = \sigma_1 \otimes 1, \quad \gamma_a = \Gamma \otimes \sigma_a, \quad \gamma^5 = \sigma_3 \otimes 1, \quad c = \sigma_3 \otimes \Gamma,$$
$$\sigma_{ab} = \tfrac{1}{2}\varepsilon_{abc}(1 \otimes \sigma_c), \quad \sigma_{0a} = -(i/2)(\sigma_3 \otimes \sigma_a), \quad (30)$$

the supercommutation relations of the super-Lorentz algebra can be written as

$$[L_{\mu\nu}, L_{\rho\sigma}] = i(\eta_{\nu\rho} L_{\mu\sigma} + \eta_{\mu\sigma} L_{\nu\rho} - \eta_{\mu\rho} L_{\nu\sigma} - \eta_{\nu\sigma} L_{\mu\rho}),$$
$$[L_{\mu\nu}, l_\alpha] = l_\beta (\sigma_{\mu\nu})^{\beta}{}_{\alpha}, \quad (31)$$
$$\{l_\alpha, l_\beta\} = \tfrac{1}{2}(c\gamma^5 \sigma^{\mu\nu})_{\alpha\beta} L_{\mu\nu}.$$

The matrix c is the Dirac charge conjugation matrix such that $c\gamma^\mu c^{-1} = -\gamma^{\mu T}$, and $\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$ is the metric tensor of Minkowski space.

The two Casimir operators K_2^{SL} and $K_2'^{SL}$ can also be written in a Lorentz covariant form. If we define

$$\kappa_2 = 2(K_2^{SL} + K_2'^{SL}),$$
$$\kappa_2' = 2(K_2^{SL} - K_2'^{SL}),$$

then we have

$$\kappa_2 = \tfrac{1}{2} L_{\mu\nu} L^{\mu\nu} + l_\alpha (c\gamma^5)^{\alpha\beta} l_\beta,$$
$$\kappa_2' = i(\tfrac{1}{2} {}^*L_{\mu\nu} L^{\mu\nu} + l_\alpha c^{\alpha\beta} l_\beta), \quad (32)$$

where ${}^*L_{\mu\nu} = -(i/2)\varepsilon_{\mu\nu\rho\sigma} L^{\rho\sigma}$ is the dual tensor of $L_{\mu\nu}$.

D. Realization of the first representations

As for the superrotation algebra we shall give realizations of the low-dimensional representations $(J,J')^{SL}$ as multiplets of $(4J+1)\times(4J'+1)$ tensor operators whose supercommutation relations with the generators $(L_{\mu\nu}, l_\alpha)$ satisfy the super-Jacobi identities.

1. The superspinorial representations $(\frac{1}{2},0)^{SL}$ and $(0,\frac{1}{2})^{SL}$

The superspinorial representations $(\frac{1}{2},0)^{SL}$ and $(0,\frac{1}{2})^{SL}$ of the super-Lorentz algebra can be realized by superspinors with three independent components, denoted, respectively, by Φ^+ and Φ^-:

$$\Phi^+ = (S^+, \psi_\alpha^+), \quad \Phi^- = (S^-, \psi_\alpha^-). \quad (33)$$

Here the $\psi_\alpha^\pm$ are four component Dirac spinors with given chirality, that is, they satisfy the condition

$$\gamma^5 \psi^\pm = \pm \psi^\pm, \quad (34)$$

which leaves only two independent components.

The supercommutation relations of these irreducible tensor operators with the generators of the super-Lorentz algebra read

$$[L_{\mu\nu}, S^\pm] = 0,$$
$$[L_{\mu\nu}, \psi_\alpha^\pm] = \psi_\beta^\pm (\sigma_{\mu\nu})^{\beta}{}_{\alpha},$$
$$[l_\alpha, S^\pm] = -(1/\sqrt{2})\psi_\alpha^\pm, \quad (35)$$
$$\{l_\alpha, \psi_\beta^\pm\} = \mp (1/2\sqrt{2})(c(1 \pm \gamma^5))_{\alpha\beta} S^\pm.$$

We notice that the generators l_α mix the bosonic and fermionic components of the multiplet.

2. The super-Dirac representation $(\frac{1}{2},0)^{SL} \oplus (0,\frac{1}{2})^{SL}$

Such a supermultiplet can be constructed with two scalars S and P and a four-component spinor ψ_α, defined from the multiplets (33) by

$$S = (1/\sqrt{2})(S^+ + S^-),$$
$$P = (1/\sqrt{2})(S^+ - S^-), \quad (36)$$
$$\psi_\alpha = \psi_\alpha^+ \oplus \psi_\alpha^-.$$

Their supercommutation relations can be derived in a straightforward manner from (35):

$$[L_{\mu\nu}, S] = 0, \quad [L_{\mu\nu}, P] = 0,$$
$$[L_{\mu\nu}, \psi_\alpha] = \psi_\beta (\sigma_{\mu\nu})^{\beta}{}_{\alpha},$$
$$[l_\alpha, S] = -\tfrac{1}{2}\psi_\alpha, \quad (37)$$
$$[l_\alpha, P] = -\tfrac{1}{2}\psi_\beta (\gamma^5)^{\beta}{}_{\alpha},$$
$$\{l_\alpha, \psi_\beta\} = -\tfrac{1}{2} c_{\alpha\beta} P - \tfrac{1}{2}(c\gamma^5)_{\alpha\beta} S.$$

3. The supervectorial representation $(\frac{1}{2},\frac{1}{2})^{SL}$

The supervectorial representation $(\frac{1}{2},\frac{1}{2})^{SL}$ is realized by a nine-dimensional multiplet that contains a scalar V_5, a four-component spinor V_α, and a quadrivector V_μ with supercommutation relations

$$[L_{\mu\nu}, V_\rho] = i(\eta_{\nu\rho} V_\mu - \eta_{\mu\rho} V_\nu),$$
$$[L_{\mu\nu}, V_\alpha] = V_\beta (\sigma_{\mu\nu})^{\beta}{}_{\alpha},$$
$$[L_{\mu\nu}, V_5] = 0,$$
$$[l_\alpha, V_\mu] = -\tfrac{1}{2} V_\beta (\gamma^5 \gamma_\mu)^{\beta}{}_{\alpha}, \quad (38)$$
$$[l_\alpha, V_5] = (1/\sqrt{2}) V_\beta (\gamma^5)^{\beta}{}_{\alpha},$$
$$\{l_\alpha, V_\beta\} = -\tfrac{1}{2}(c\gamma^\mu)_{\alpha\beta} V_\mu + (1/\sqrt{2}) c_{\alpha\beta} V_5.$$

4. The representations $(1,0)^{SL}$ and $(0,1)^{SL}$

The $(1,0)^{SL}$ representation can be realized by a five-dimensional multiplet that contains a self-dual skew-symmetric tensor $F_{\mu\nu}^+$ and a chiral spinor F_α^+:

$$-(i/2)\varepsilon_{\mu\nu}{}^{\rho\sigma} F_{\rho\sigma}^+ = F_{\mu\nu}^+, \quad F_\alpha^+ (\gamma^5)^{\alpha}{}_{\beta} = F_\beta^+. \quad (39)$$

In the same way, the $(0,1)^{SL}$ representation can be realized by an anti-self-dual skew-symmetric tensor $F^-_{\mu\nu}$ and an antichiral spinor F^-_α:

$$-(i/2)\varepsilon_{\mu\nu}{}^{\rho\sigma}F^-_{\rho\sigma} = -F^-_{\mu\nu}, \quad F^-_\alpha(\gamma^5)^\alpha{}_\beta = -F^-_\beta. \tag{40}$$

The supercommutation relations are

$$\begin{aligned}
&[L_{\mu\nu},F^\pm_{\rho\sigma}] = i(\eta_{\nu\rho}F^\pm_{\mu\sigma} + \eta_{\mu\sigma}F^\pm_{\nu\rho} - \eta_{\mu\rho}F^\pm_{\nu\sigma} - \eta_{\nu\sigma}F^\pm_{\mu\rho}),\\
&[L_{\mu\nu},F^\pm_\alpha] = F^\pm_\beta(\sigma_{\mu\nu})^\beta{}_\alpha,\\
&[l_\alpha,F^\pm_{\mu\nu}] = -F^\pm_\beta(\sigma_{\mu\nu})^\beta{}_\alpha,\\
&\{l_\alpha,F^\pm_\beta\} = \tfrac{1}{2}(c\gamma^5\sigma^{\mu\nu})_{\alpha\beta}F^\pm_{\mu\nu}.
\end{aligned} \tag{41}$$

5. The supertensorial representation $(1,0)^{SL}\oplus(0,1)^{SL}$

It is the adjoint representation $(1,0)^{SL}\oplus(0,1)^{SL}$. It contains a six-component tensor $F_{\mu\nu}$ and a four-component spinor F_α with supercommutation relations

$$\begin{aligned}
&[L_{\mu\nu},F_{\rho\sigma}] = i(\eta_{\nu\rho}F_{\mu\sigma} + \eta_{\mu\sigma}F_{\nu\rho} - \eta_{\mu\rho}F_{\nu\sigma} - \eta_{\nu\sigma}F_{\mu\rho}),\\
&[L_{\mu\nu},F_\alpha] = F_\beta(\sigma_{\mu\nu})^\beta{}_\alpha,\\
&[l_\alpha,F_{\mu\nu}] = -F_\beta(\sigma_{\mu\nu})^\beta{}_\alpha,\\
&\{l_\alpha,F_\beta\} = \tfrac{1}{2}(c\gamma^5\sigma^{\mu\nu})_{\alpha\beta}F_{\mu\nu}.
\end{aligned} \tag{42}$$

Of course, these anticommutation relations are identical to the relations (31) satisfied by the algebra itself.

E. Tensor product of two representations

The reduction formula for the tensor product of two representations $(J,J')^{SL}$ and $(K,K')^{SL}$ is deduced from the formula (16):

$$(J,J')^{SL}\otimes(K,K')^{SL} = \bigoplus_{p=0}^{4\inf(J,K)}\ \bigoplus_{q=0}^{4\inf(J',K')}(J+K-p/2,J'+K'-q/2)^{SL}. \tag{43}$$

This shows that all the representations can be built by tensor products from the two three-dimensional fundamental representations $(\frac{1}{2},0)^{SL}$ and $(0,\frac{1}{2})^{SL}$. Let us consider some examples of tensor products.

(i) The reduction formula for the tensor product of the representations $(\frac{1}{2},0)^{SL}$ and $(0,\frac{1}{2})^{SL}$ is

$$(\tfrac{1}{2},0)^{SL}\otimes(0,\tfrac{1}{2})^{SL} = (\tfrac{1}{2},\tfrac{1}{2})^{SL}.$$

Thus the supervector $V=(V^5,V_\alpha,V_\mu)$ constructed as the product of two superspinors $\Phi^+=(S^+,\psi^+)$ and $\Phi^-=(S^-,\psi^-)$, which are chiral and antichiral, respectively, is given by

$$V=\begin{cases} V^5 = S^+S^-,\\ V_\alpha = S^+\psi^-_\alpha - \psi^+_\alpha S^-,\\ V_\mu = -(1/\sqrt{2})\psi^+_\alpha(\gamma_\mu c)^{\alpha\beta}\psi^-_\beta. \end{cases} \tag{44}$$

(ii) The tensor product of two representations $(\frac{1}{2},0)^{SL}$ reduces as

$$(\tfrac{1}{2},0)^{SL}\otimes(\tfrac{1}{2},0)^{SL} = (1,0)^{SL}\oplus(\tfrac{1}{2},0)^{SL}\oplus(0,0)^{SL}.$$

Therefore, given two chiral superspinors $\Phi_1^+=(S_1^+,\psi_1^+)$ and $\Phi_2^+=(S_2^+,\psi_2^+)$ one can build a self-dual supertensor $F^+=(F^+_{\mu\nu},F^+_\alpha)$, a chiral superspinor $\Phi^+=(S^+,\psi^+_\alpha)$ and a superscalar T^+, namely,

$$\begin{aligned}
&F^+ = \begin{cases} F^+_{\mu\nu} = \psi^+_{1\alpha}(\sigma_{\mu\nu})^{\alpha\beta}\psi^+_{2\beta},\\ F^+_\alpha = (1/\sqrt{2})(S_1^+\psi^+_{2\alpha} - \psi^+_{1\alpha}S_2^+);\end{cases}\\
&\Phi^+\begin{cases} S^+ = 2S_1^+S_2^+ + \psi^+_{1\alpha}c^{\alpha\beta}\psi^+_{2\beta},\\ \psi^+_\alpha = S_1^+\psi^+_{2\alpha} + \psi^+_{1\alpha}S_2^+;\end{cases}\\
&T^+ = S_1^+S_2^+ + \psi^+_{1\alpha}c^{\alpha\beta}\psi^+_{2\beta}.
\end{aligned} \tag{45}$$

(iii) The reduction formula for the tensor product of two $(0,\frac{1}{2})^{SL}$ representations is

$$(0,\tfrac{1}{2})^{SL}\otimes(0,\tfrac{1}{2})^{SL} = (0,1)^{SL}\oplus(0,\tfrac{1}{2})^{SL}\oplus(0,0)^{SL}.$$

Then from two antichiral superspinors $\Phi_1^-=(S_1^-,\psi^-_{1\alpha})$ and $\Phi_2^-=(S_2^-,\psi^-_{2\alpha})$ one can construct an anti-self-dual supertensor $F^-=(F^-_{\mu\nu},F^-_\alpha)$, an antichiral superspinor $\Phi^-=(S^-,\psi^-)$, and a superscalar T^-. Explicitly, they are given by

$$\begin{aligned}
&F^- = \begin{cases} F^-_{\mu\nu} = \psi^-_{1\alpha}(\sigma_{\mu\nu})^{\alpha\beta}\psi^-_{2\beta},\\ F^-_\alpha = -(1/\sqrt{2})(S_1^-\psi^-_{2\alpha} - \psi^-_{1\alpha}S_2^-);\end{cases}\\
&\Phi^-\begin{cases} S^- = 2S_1^-S_2^- - \psi^-_{1\alpha}c^{\alpha\beta}\psi^-_{2\beta};\\ \psi^-_\alpha = S_1^-\psi^-_{2\alpha} + \psi^-_{1\alpha}S_2^-;\end{cases}\\
&T^- = S_1^-S_2^- - \psi^-_{1\alpha}c^{\alpha\beta}\psi^-_{2\beta}.
\end{aligned} \tag{46}$$

F. Quadratic invariants of the representations

According to Eq. (43), the reduction formula for the tensor product of a representation with itself reads

$$(J,J')^{SL}\otimes(J,J')^{SL} = \bigoplus_{p=0}^{4J}\ \bigoplus_{q=0}^{4J'}(2J-p/2,2J'-q/2)^{SL}; \tag{47}$$

that is, the reduction always contains the trivial representation $(0,0)^{SL}$. Thus for every realization $(J,J')^{SL}$ one can build a quadratic invariant and for the representations that are the direct sum of the form $(J,J')^{SL}\oplus(J',J)^{SL}$ one can construct two quadratic invariants.

In Table II we give the quadratic invariants of the various realizations described above.

G. Bosonic and fermionic dimensions of the representations

The decomposition (27) allows us to separate the representation space $(J,J')^{SL}$ into two subspaces: a bosonic sub-

TABLE II. The quadratic invariants for the first representations of the super-Lorentz algebra.

Representation	Quadratic invariants
$(\frac{1}{2},0)^{SL}$	$(S^+)^2 + \psi^+_\alpha c^{\alpha\beta}\psi^+_\beta$
$(0,\frac{1}{2})^{SL}$	$(S^-)^2 - \psi^-_\alpha c^{\alpha\beta}\psi^-_\beta$
	$S^2 + P^2 + \psi_\alpha(c\gamma^5)^{\alpha\beta}\psi_\beta$
$(\frac{1}{2},0)^{SL}$	$SP + \psi_\alpha c^{\alpha\beta}\psi_\beta$
$(1,0)^{SL}$	$\frac{1}{2}F^+_{\mu\nu}F^{+\mu\nu} + F^+_\alpha c^{\alpha\beta}F^+_\beta$
$(0,1)^{SL}$	$\frac{1}{2}F^-_{\mu\nu}F^{-\mu\nu} - F^-_\alpha c^{\alpha\beta}F^-_\beta$
	$\frac{1}{2}F_{\mu\nu}F^{\mu\nu} + F_\alpha(c\gamma^5)^{\alpha\beta}F_\beta$
$(1,0)^{SL}\oplus(0,1)^{SL}$	$\frac{1}{2}{}^*F_{\mu\nu}F^{\mu\nu} + F_\alpha c^{\alpha\beta}F_\beta$

TABLE III. Bosonic and fermionic dimensions of the first representations of the super-Lorentz algebra.

$(J,J')^{SL}$	$[d_B,d_F]$	Lorentz decomposition
$(0,0)^{SL}$	[1,0]	$(0,0)^{L}$
$(\frac{1}{2},0)^{SL}$	[1,2]	$(\frac{1}{2},0)^{L} \oplus (0,0)^{L}$
$(0,\frac{1}{2})^{SL}$	[1,2]′	$(0,\frac{1}{2})^{L} \oplus (0,0)^{L}$
$(\frac{1}{2},0)^{SL} \oplus (0,\frac{1}{2})^{SL}$	[2,4]	$(\frac{1}{2},0)^{L} \oplus (0,\frac{1}{2})^{L} \oplus (0,0)^{L} \oplus (0,0)^{L}$
$(\frac{1}{2},\frac{1}{2})^{SL}$	[5,4]	$(\frac{1}{2},\frac{1}{2})^{L} \oplus (\frac{1}{2},0)^{L} \oplus (0,\frac{1}{2})^{L} \oplus (0,0)^{L}$
$(1,0)^{SL}$	[3,2]	$(1,0)^{L} \oplus (\frac{1}{2},0)^{L}$
$(0,1)^{SL}$	[3,2]′	$(0,1)^{L} \oplus (0,\frac{1}{2})^{L}$
$(1,0)^{SL} \oplus (0,1)^{SL}$	[6,4]	$(1,0)^{L} \oplus (0,1)^{L} \oplus (\frac{1}{2},0)^{L} \oplus (0,\frac{1}{2})^{L}$
$(1,\frac{1}{2})^{SL}$	[7,8]	$(1,\frac{1}{2})^{L} \oplus (1,0)^{L} \oplus (\frac{1}{2},\frac{1}{2})^{L} \oplus (\frac{1}{2},0)^{L}$
$(\frac{1}{2},1)^{SL}$	[7,8]′	$(\frac{1}{2},1)^{L} \oplus (0,1)^{L} \oplus (\frac{1}{2},\frac{1}{2})^{L} \oplus (0,\frac{1}{2})^{L}$
$(1,\frac{1}{2})^{SL} \oplus (\frac{1}{2},1)^{SL}$	[14,16]	[7,8] ⊕ [7,8]′
$(1,1)^{SL}$	[13,12]	$(1,1)^{L} \oplus (1,\frac{1}{2})^{L} \oplus (\frac{1}{2},1)^{L} \oplus (\frac{1}{2},\frac{1}{2})^{L}$

space with dimension d_B, the direct sum of all the representation subspaces J^R with even spin and a fermionic subspace with dimension d_F, the direct sum of all the representation subspaces with odd spin.

The representations can be characterized by the dimensions d_B and d_F. We note, however, that the representations $(K,L)^{SL}$ and $(L,K)^{SL}$ have the same dimensions. By convention, we will distinguish between these two representations by denoting $[d_B,d_F]$ the representation with $J<J'$ and $[d_B,d_F]'$ the representation with $J>J'$.

We note that, in Ref. 11 some representations are mentioned but the fundamental representations [1,2] and [1,2]′ do not appear. Furthermore, in the reduction formula of the tensor product [2,4] ⊗ [5,4] [Eq. (4,462)] a representation [8,12] appears that does not exist. The correct reduction formula reads

$$[2,4] \otimes [5,4] = [14,16] \oplus [5,4] \oplus [5,4] \oplus [2,4].$$

In Table III we give the dimensions of the representations with spins J and $J' < 1$, and their decomposition (26) on the Lorentz subalgebra.

[1]Y. A. Gol'fand and E. P. Likhtman, JETP Lett. **13**, 323 (1971).
[2]J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[3]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[4]A. Salam and J. Strathdee, Fortschr. Phys. **26**, 5 (1978).
[5]D. Z. Freedman, P. Van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976).
[6]S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).
[7]P. Van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).
[8]S. W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[9]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[10]J. Lukierski and A. Novicki, Fortschr. Phys. **30**, 75 (1982).
[11]B. De Witt, *Supermanifolds* (Cambridge U.P., Cambridge, 1984).
[12]P. Minnaert, C. R. Acad. Sci. Paris **301** (II), 1395 (1985).

Positive discrete series representations of the noncompact superalgebra Osp(4/2,*R*)

H. A. Schmitt, P. Halse, and B. R. Barrett
Department of Physics, University of Arizona, Tucson, Arizona 85721

A. B. Balantekin
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706

(Received 23 February 1988; accepted for publication 24 May 1989)

Some physically realizable positive discrete series representations of the noncompact orthosymplectic superalgebra Osp(4/2,*R*) are considered. The decomposition of these Osp(4/2,*R*) representations on reduction to Sp(2,*R*)×SO(4) is studied in detail, and the corresponding state vectors are explicitly constructed by acting with the generators on a general lowest weight state. Some examples are given to illustrate these results for particular single-particle spaces.

I. INTRODUCTION

Algebraic techniques have found widespread use in physics.[1] In nuclear physics, the observed symmetries have found expression in the language of group theory and their associated algebras.[2,3] In particular, the interacting boson model[3] (IBM) has been shown to be able to correlate many properties of the spectra of even–even nuclei. The interacting boson fermion model[4] (IBFM) is an extension of the IBM to handle odd–odd nuclei by the addition of a fermionic degree of freedom. In the IBM and the geometrical model[5] (GM), even–even nuclei are described completely by bosons (with $l=2$ for the GM and $l=0,2$ for the IBM), odd–A nuclei by a system of both bosons and fermions. Of particular interest to us are the dynamical combined Bose–Fermi symmetries[6] of the IBFM, and the possibility of physically relevant supersymmetries.[7] These Bose–Fermi symmetries are associated with algebras whose generators are the sums of generators of the separate Bose and Fermi groups.[8,9] It has been found that spectra for even–even and odd–even nuclei can be described reasonably well using these generators.

In a previous paper,[10] we have shown that, for the single-particle space of $l=2$ bosons and $j=\frac{3}{2},\frac{5}{2}$ $((l=2)\times(s=\frac{1}{2}))$ fermions, a realization of the noncompact superalgebra Osp(4/2,*R*) is complementary to the combined Bose–Fermi algebra $SO^{BF}(5)$. Similarly, for $l=0,2$ and $j=\frac{1}{2},\frac{3}{2},\frac{5}{2}$, this superalgebra is complementary to the combined $SO^{BF}(6)$ algebra. Use of the noncompact superalgebra Osp(4/2,*R*), which then provides a simultaneous classification to that of the combined Bose–Fermi algebras, might, in some cases, result in some calculational simplification. It is therefore useful to study the representations of Osp(4/2,*R*). The approach is similar in principle to earlier work[11,12] involving the quasispin-seniority classification for bosons and fermions.

In Sec. II, we briefly review the structure of Osp(4/2,*R*) and its complementarity to combined Bose–Fermi algebras. We study the decomposition of the representations of this superalgebra in Sec. III, and the representation complementarity is deduced in Sec. IV. Several examples are considered in Sec. V.

II. ORTHOSYMPLECTIC SUPERALGEBRAS IN BOSE–FERMI PAIRING

In this section, we present for completeness the important aspects of Ref. 10, concerning algebras corresponding to a partial pairing of bosons and fermions. Consider a boson space of dimension n, and a fermion space (of dimension m) which can be factored into two parts having dimensions n (with a symmetric metric) and p $(n\times p=m)$,

$$a^{\dagger}{}_{ik},a^{ik}=(a^{\dagger}{}_{ik})^{\dagger}:\quad \{a^{ik},a^{\dagger}{}_{jl}\}=\delta^{ik}{}_{jl},\qquad i,j=1,\ldots,n;\quad k,l=1,\ldots,p, \tag{2.1a}$$

$$b^{\dagger}{}_{i},b^{i}=(b^{\dagger}{}_{i})^{\dagger}:\quad [b^{i},b^{\dagger}{}_{j}]=\delta^{i}_{j},\quad i,j=1,\ldots,n. \tag{2.1b}$$

The sets of operators which are combined as scalars in the n space [a summation over i is implied in Eqs. (2.2)–(2.4)] close to give[13,14]

$$\text{Fermions: SO}(2p):\quad a^{\dagger i}{}_{k}a^{\dagger}{}_{il},\quad a^{ik}a_{i}{}^{l},\quad a^{\dagger i}{}_{k}a_{i}{}^{l}-(n/2)\delta_{k}{}^{l}, \tag{2.2a}$$

$$\text{Bosons: Sp}(2,R):\quad b^{\dagger i}b^{\dagger}{}_{i},\quad b^{i}b_{i},\quad b^{\dagger i}b_{i}+n/2. \tag{2.2b}$$

Augmenting these sets with the mixed Bose–Fermi operators constructed using the same prescription (i.e., scalar in the n space), the $(2p^2+3p+3)$-element graded algebra closes on that of Osp(2*p*/2,*R*),[15]

$$\text{Osp}(2p/2,R):\quad b^{\dagger i}a^{\dagger}{}_{ik},\quad b^{\dagger}{}_{i}a^{ik},\quad b^{i}a^{\dagger}{}_{ik},\quad b^{i}a_{i}{}^{k}\quad [\text{SO}(2p),\text{Sp}(2,R)]. \tag{2.3}$$

These operators commute with the generators of $SO^{BF}(n)$,

$$SO^{BF}(n):\quad (b^{\dagger}{}_{i}b_{j}-b^{\dagger}{}_{j}b_{i})+(a^{\dagger}{}_{ik}a_{j}{}^{k}-a^{\dagger}{}_{jk}a_{i}{}^{k}). \tag{2.4}$$

Thus there exists a direct product structure

$$\text{Osp}(2p/2,R)\times SO^{BF}(n). \tag{2.5}$$

Two subgroups of Osp(2*p*/2,*R*) with potential physical relevance are

$$\text{Osp}(2p/2,R)\supset[\text{SO}(2p)\supset\text{U}(p)]\times[\text{Sp}(2,R)\supset\text{U}(1)], \tag{2.6a}$$

 0022-2488/89/112714-07$02.50

$$\mathrm{Osp}(2p/2,R)\supset \mathrm{U}(1/p)\supset[\mathrm{U}(p)\times \mathrm{U}(1)], \tag{2.6b}$$

where $\mathrm{U}(1/p)$ corresponds to a restriction to number-conserving operators.

In the case that the fermion levels have j values corresponding to coupling a pseudospin s to the l values of the boson states, the orthosymplectic supersymmetry corresponding to l pairing, which can be described by $\mathrm{SO}^{\mathrm{BF}}(\Sigma_i(2l_i+1))$ [Eq. (2.4)], is $\mathrm{Osp}(4s+2/2,R)$, becoming $\mathrm{Osp}(4/2,R)$ for the commonly encountered value $s=\frac{1}{2}$.

The superalgebra $\mathrm{Osp}(4/2,R)$ is generated by 17 elements of which the eight even elements close under commutation to generate the subalgebra [Eq. (2.6a)]

$$\mathrm{Osp}(4/2,R)\supset \mathrm{Sp}(2,R)\times \mathrm{SO}(4)$$

$$\cong \mathrm{Sp}(2,R)\times \mathrm{SO}(3)\times \mathrm{SO}(3). \tag{2.7}$$

Here we concentrate exclusively on this subalgebra chain, deferring a study of the $\mathrm{U}(1/2)$ chain to a future paper. In Table I, we present the specific realization for the generators of $\mathrm{Osp}(4/2,R)$ and $\mathrm{SO}^{\mathrm{BF}}(n)$ appropriate to this Bose–Fermi pairing scheme, while the commutation and anticommutation relations and Hermiticity properties for $\mathrm{Osp}(4/2,R)$ are given in Appendix A.

In the present work, we consider only the representations appropriate to this realization in terms of boson and fermion creation and annihilation operators.

III. DECOMPOSITION OF OSP(4/2,R) REPRESENTATIONS ON REDUCTION TO SP(2,R)×SO(4)

We construct the irreducible representations of $\mathrm{Osp}(4/2,R)$ by acting with the raising operators on lowest weight states (LWS), those states which are annihilated by all the lowering operators. The diagonal operators K_0, R_0, and S_0 generate the Cartan subalgebra of $\mathrm{Osp}(4/2,R)$, the corresponding eigenvalues giving the weight of each state. We choose the first, second, and third components of the weight to be the eigenvalues of K_0, R_0, and S_0, respectively. From the (anti-)commutation relations, it is then clear that the lowering operators are W_μ, Y_μ, K_-, R_-, and S_-, whereas the raising operators are V_μ, X_μ, K_+, R_+, and S_+. The $\mathrm{Osp}(4/2,R)$ representations can be labeled by the $\mathrm{Sp}(2,R)\times \mathrm{SO}(3)\times \mathrm{SO}(3)$ quantum numbers of the LWS, where we have employed the homomorphism [Eq. (2.7)] between SO(4), with labels p_1, p_2, and $\mathrm{SO}(3)\times \mathrm{SO}(3)$, with labels R, S; these quantum numbers being related by

$$R=\tfrac{1}{2}(p_1+p_2), \tag{3.1a}$$

$$S=\tfrac{1}{2}(p_1-p_2). \tag{3.1b}$$

The state vectors are further specified by including the labels K, M_K, R, M_R, S, and M_S.

To find the representations of the subalgebra it is only necessary to consider the raising operators which generate the coset $\mathrm{Osp}(4/2,R)/[\mathrm{Sp}(2,R)\times \mathrm{SO}(4)]$, i.e., X_μ and V_μ, since K_+, R_+, and S_+ only ladder within these representations. By using the anticommutation relations of Appendix A, it can be shown that there are 16 independent combinations of these operators. These are listed in Table II, together with their transformation properties under $\mathrm{Sp}(2,R)\times \mathrm{SO}(3)\times \mathrm{SO}(3)$, and the allowed $\mathrm{Sp}(2,R)\times \mathrm{SO}(3)\times \mathrm{SO}(3)$ quantum numbers obtained by acting on a LWS with labels K, R, and S. Thus each $\mathrm{Osp}(4/2,R)$ representation decomposes into at most 16 representations of the subgroup. As can be seen from this table, there can be two independent states with $\mathrm{Sp}(2,R)\times \mathrm{SO}(3)\times \mathrm{SO}(3)$ quantum numbers $K+1, R, S$.

There is one more problem to be addressed; the labels in the far right-hand column of Table II refer to orthogonal states, whereas the states formed by the action of the coset raising operators upon the LWS are in general not orthogonal. In the general case, there is some freedom in choosing a suitable inner product; however, as stated above, in this paper we are interested in the representations appropriate for the particular realization discussed, associated with the positive definite inner product defined in Fock space, as is conventional in many-body quantum mechanics.[16]

TABLE I. Generators of the groups involved in the $\mathrm{Osp}(4/2,R)\times \mathrm{SO}^{\mathrm{BF}}(n)$ scheme. We have assumed a single pseudo-orbital angular momentum l ($n=2l+1$), the generalization to multiple values being straightforward. The tensor components of the annihilation operators are defined as $a^{(l\,1)}_{m\,\mu}=(-)^{l-m+1/2-\mu}(a^{\dagger(l\,1/2)}_{-m\,-\mu})^{\dagger}$.

$\mathrm{Osp}(4/2,R)$:	$R_+=\sqrt{(n/2)}\,[a^\dagger a^\dagger]^{(0,0)}_{0\,0}$, $R_-=(R_+)^\dagger$, $R_0=\frac{1}{2}(N_{\mathrm{F}}-n)$
	$S_\mu=-\sqrt{(n/2)}\,[a^\dagger a]^{(0,1)}_{0\,\mu}$
	$K_+=(\sqrt{n}/2)[b^\dagger b^\dagger]^{(0)}_0$, $K_-=(K_+)^\dagger$, $K_0=\frac{1}{2}(N_{\mathrm{B}}+n/2)$
	$V_\mu=(1/\sqrt{2})b^\dagger\cdot a^{(1/2)}_\mu\equiv\sqrt{(n/2)}\,[b^\dagger a]^{(0,1)}_{0\,\mu}$
	$=\dfrac{1}{\sqrt{2}}\sum_m(-)^m b^\dagger_m a^{(1)}_{-m\,\mu}$
	$W_\mu=(1/\sqrt{2})b\cdot a^{\dagger(1/2)}_\mu$, $X_\mu=(1/\sqrt{2})b^\dagger\cdot a^{\dagger(1/2)}_\mu$, $Y_\mu=(1/\sqrt{2})b\cdot a^{(1/2)}_\mu$
$\mathrm{SO}(4)$:	R_+, R_-, R_0, S_μ
$\mathrm{Sp}(2,R)$:	K_+, K_-, K_0
$\mathrm{U}(2/1)$:	R_0, K_0, S_μ, V_μ, W_μ
$\mathrm{SU}^S(2)$:	S_μ
$\mathrm{U}^{\mathrm{F}}(1)$:	R_0
$\mathrm{U}^{\mathrm{B}}(1)$:	K_0
$\mathrm{SO}^{\mathrm{F}}(n)$:	$\mathrm{U}^{\mathrm{F}(k)}_m=-\sqrt{2}[a^\dagger a]^{(k,0)}_{m\,0}$, $k=1,3,\ldots,n-2$
$\mathrm{SO}^{\mathrm{B}}(n)$:	$\mathrm{U}^{\mathrm{B}(k)}_m=[b^\dagger b]^{(k)}_m$, $k=1,3,\ldots,n-2$
$\mathrm{SO}^{\mathrm{BF}}(n)$:	$\mathrm{U}^{\mathrm{B}(k)}_m+\mathrm{U}^{\mathrm{F}(k)}_m$, $k=1,3,\ldots,n-2$

It might be added that there is some freedom in the way that one constructs the states, however, the number of independent states in a given Osp(4/2,R) representation does not depend on this choice. Upon evaluating the norms of these states, we find that all 16 independent multiplets are not always realized; the conditions for the existence of any multiplet can be expressed in terms of the Osp(4/2,R) quantum numbers K, R, and S. This is a generalization of the results we have obtained[17] for Osp(2/2,R) and that Scheunert *et al.*[18] have found for the compact superalgebra Spl(2,1).

The Osp(4/2,R) representations so generated are irreducible, as follows from the properties of our chosen realization. The decomposition is presented in Eq. (3.2), the conditions for the existence of each multiplet being given below the corresponding labels:

$$
\begin{aligned}
(KRS) \rightarrow\ & (K)(R,S) \\
& \oplus \underset{2K-R-S\neq 0}{(K+\tfrac{1}{2})(R+\tfrac{1}{2},S+\tfrac{1}{2})} \oplus \underset{S\neq 0}{(K+\tfrac{1}{2})(R+\tfrac{1}{2},S-\tfrac{1}{2})} \oplus \underset{R\neq 0}{(K+\tfrac{1}{2})(R-\tfrac{1}{2},S+\tfrac{1}{2})} \oplus \underset{S\neq 0,R\neq 0}{(K+\tfrac{1}{2})(R-\tfrac{1}{2},S-\tfrac{1}{2})} \\
& \oplus \underset{2K-R-S\neq 0}{(K+1)(R+1,S)} \oplus \underset{R\neq 0,\,2K-R-S\neq 0}{(K+1)(R,S)} \oplus \underset{R\neq 0,\,\frac{1}{2}}{(K+1)(R-1,S)} \\
& \oplus \underset{2K-R-S\neq 0}{(K+1)(R,S+1)} \oplus \underset{S\neq 0,\,2K+R-S\neq 0}{(K+1)(R,S)} \oplus \underset{S\neq 0,\,\frac{1}{2}}{(K+1)(R,S-1)} \\
& \oplus \underset{2K-R-S\neq 0}{(K+\tfrac{3}{2})(R+\tfrac{1}{2},S+\tfrac{1}{2})} \oplus \underset{S\neq 0,\,2K-R-S\neq 0}{(K+\tfrac{3}{2})(R+\tfrac{1}{2},S-\tfrac{1}{2})} \oplus \underset{R\neq 0,\,2K-R-S\neq 0}{(K+\tfrac{3}{2})(R-\tfrac{1}{2},S+\tfrac{1}{2})} \oplus \underset{R\neq 0,S\neq 0}{(K+\tfrac{3}{2})(R-\tfrac{1}{2},S-\tfrac{1}{2})} \\
& \oplus \underset{2K-R-S\neq 0}{(K+2)(R,S)}. \qquad (3.2)
\end{aligned}
$$

With our realization of the algebra, the allowed multiplet labels $(K)(R,S)$ satisfy the inequality $R+S\leqslant 2K$. The decomposition [Eq. (3.2)] can be verified using the Kronecker product for the complementary $\mathrm{SO}^{\mathrm{BF}}(\Sigma_i 2l_i+1)$ algebra and the complementarity relations as discussed in Sec. IV.

As noted above, there can be two independent multiplets with the $\mathrm{Sp}(2,R)\times\mathrm{SO}(3)\times\mathrm{SO}(3)$ quantum numbers $K+1$, R, S. This is in contrast to the case of Osp(2/2,R),[17] where the superalgebra is not sufficiently complex for such a situation to occur. We should note that there is some ambiguity in how one chooses to construct these states; we chose to first orthogonalize the vector formed by the action of $A^{(01)}$ (Table II), then that obtained using $A^{(10)}$. This introduces an otherwise absent asymmetry in R and S; however, the multiplicity of these representations can be expressed in the symmetrical form $M_{K+1,R,S}=2-\delta_{RS,0}-\delta_{R+S,2K}$. The orthogonalized states are presented in Appendix B as functions of the generators acting on the LWS. The R-S asymmetry for $K+1$, R, S is again evident; it is possible that there are linear combinations of these two states which have a simpler form.

TABLE II. Transformation properties of the independent vectors with respect to $\mathrm{SO}(4)\times\mathrm{Sp}(2,R)$. For convenience, we have defined the following combinations:
$A^{(10)}=X_{1/2}V_{-1/2}+V_{1/2}X_{-1/2}=X_{1/2}V_{-1/2}-X_{-1/2}V_{1/2}+K_+$,
$A^{(01)}=X_{1/2}V_{-1/2}+X_{-1/2}V_{1/2}$.

Operator	SO(4)	Sp(2,R)	$(K)(R,S)$
1	(0,0)	(0)	$(K)(R,S)$
V_μ, X_μ	$(\frac{1}{2},\frac{1}{2})$	$(\frac{1}{2})$	$(K+\frac{1}{2})(R\pm\frac{1}{2},S\pm\frac{1}{2})$
$X_{1/2}X_{-1/2}, A^{(10)}, V_{1/2}V_{-1/2}$	(1,0)	(1)	$(K+1)\left(R\,{}^{+1}_{-1},S\right)$
$X_{1/2}V_{1/2}, A^{(01)}, X_{-1/2}V_{-1/2}$	(0,1)	(1)	$(K+1)\left(R,S\,{}^{+1}_{-1}\right)$
$X_\mu V_{1/2}V_{-1/2}$, $V_\mu X_{1/2}X_{-1/2}$	$(\frac{1}{2},\frac{1}{2})$	$(\frac{3}{2})$	$(K+\frac{3}{2})(R\pm\frac{1}{2},S\pm\frac{1}{2})$
$V_{1/2}V_{-1/2}X_{1/2}X_{-1/2}$	(0,0)	(2)	$(K+2)(R,S)$

IV. RELATION BETWEEN THE REPRESENTATIONS OF Osp(4/2,R) AND THE COMPLEMENTARY ORTHOGONAL GROUP

With the realization of Osp(4/2,R) in terms of the boson–fermion number conserving and pair creation operators scalar in a n-dimensional space (Table I), a larger algebra exists as discussed in Sec. II. The number conserving operators which are scalar in the p space and of odd rank in the n space close under commutation to give the algebra $\mathrm{SO}^{\mathrm{BF}}(n)$, and moreover commute with the generators of Osp(4/2,R) giving the direct product structure

$$\mathrm{Osp}(4/2,R)\times\mathrm{SO}^{\mathrm{BF}}(n). \qquad (4.1)$$

This is completely analogous to the complementarity between the Sp(2,R) and SO(4) subgroups [Eqs. (2.7)] and the individual orthogonal groups giving the quasispin-seniority classification for bosons[12] and fermions[11] separately

$$\mathrm{Sp}(2,R)\times\mathrm{SO}^{\mathrm{B}}(n), \qquad (4.2a)$$

$$\mathrm{SO}(4)\times\mathrm{SO}^{\mathrm{F}}(n). \qquad (4.2b)$$

For these cases, the corresponding relations between the Weyl representation labels[19] of Sp(2,R) and $\mathrm{SO}^{\mathrm{B}}(n)$ [K and τ, respectively], and SO(4) and $\mathrm{SO}^{\mathrm{F}}(n)$ [$(p_1 p_2)$ and $(\rho)=(\rho_1,\rho_2,\ldots)$, respectively] have been deduced

$$K=\tfrac{1}{4}n+\tfrac{1}{2}\tau, \qquad (4.3a)$$

$$p_i=\tfrac{1}{2}n-(\tilde{\rho})_{3-i},\quad i=1,2, \qquad (4.3b)$$

where $\tilde{\rho}$ denotes the transpose of the diagram implied by the unmodified[20] $\mathrm{SO}^{\mathrm{F}}(n)$ label ρ. Equations (3.1) and (4.3b) give

$$R=\tfrac{1}{2}n-\tfrac{1}{2}((\tilde{\rho})_1+(\tilde{\rho})_2), \qquad (4.4a)$$

$$S=\tfrac{1}{2}((\tilde{\rho})_1-(\tilde{\rho})_2). \qquad (4.4b)$$

The relations between corresponding Osp(4/2,R) and

$SO^{BF}(n)$ representation labels $(K\ R\ S)$ and (σ) can be found from similar considerations. The former are just the eigenvalues in the LWS of the operators

$$K: \quad M_K = \tfrac{1}{4}n + \tfrac{1}{2}N_b, \tag{4.5a}$$

$$R: \quad -M_R = \tfrac{1}{2}n - \tfrac{1}{2}(N_{1/2} + N_{-1/2}), \tag{4.5b}$$

$$S: \quad -M_S = -\tfrac{1}{2}(N_{1/2} - N_{-1/2}), \tag{4.5c}$$

where $N_{\pm 1/2}$ are the numbers of fermions with $M_S = \pm\frac{1}{2}$. Since the boson and fermion $SO(n)$ labels are constrained as in Eqs. (4.6),

$$SO^B(n): \quad (\tau), \tag{4.6a}$$

$$SO^F(n): \quad (2^{\mu}\ 1^{\nu}), \tag{4.6b}$$

those of $SO^{BF}(n)$, obtained by taking the inner tensor product, can only be of the form

$$(\sigma_1\ 2^{\mu}\ 1^{\nu}). \tag{4.7}$$

It will be convenient to represent this by the labels

$$[\sigma_1'\ \mu'\ \nu'], \tag{4.8}$$

where

$$\sigma_1' = \sigma_1 - 2, \quad \mu' = \mu + 1, \quad \nu' = \nu \quad (\sigma_1 \geqslant 2), \tag{4.9a}$$

$$\sigma_1' = 0, \quad \mu' = \mu \quad (=0), \quad \nu' = \nu + \sigma_1 \quad (\sigma_1 < 2). \tag{4.9b}$$

The corresponding LWS then has

$$N_b = \sigma', \tag{4.10a}$$

$$N_{1/2} = \mu', \tag{4.10b}$$

$$N_{-1/2} = \mu' + \nu', \tag{4.10c}$$

giving

$$K = \tfrac{1}{4}n + \tfrac{1}{2}\sigma', \tag{4.11a}$$

$$R = \tfrac{1}{2}n - \tfrac{1}{2}(2\mu' + \nu'), \tag{4.11b}$$

$$S = \tfrac{1}{2}\nu'. \tag{4.11c}$$

V. EXAMPLES

In this section, we illustrate the abstract notions of the previous sections with some specific examples.

A. Single-particle space $l=0, j=\frac{1}{2}$

A system of scalar (s) bosons and $j = \frac{1}{2}$ fermions is the simplest one for which the $Osp(4/2,R)$ superalgebra can be realized. We present an explicit treatment for this reason alone, the likelihood that such a simple model will have no application to physics notwithstanding. Indeed, this case is particularly simplified because $Osp(4/2,R)$ is the "large" superalgebra $Osp(2m/2n,R)$[10]; correspondingly, the boson–fermion orthogonal group $SO^{BF}(n)$ does not exist for $n = 1$. Thus there are only two allowed representations, containing all the states with even and odd values, respectively, for the total number of particles.[10]

For the representation containing the even states, the LWS (obtained by successively minimizing the eigenvalues of K_0, R_0, and S_0) is clearly the boson–fermion vacuum with $Osp(4/2,R)$ labels

$$K_e = 0 + \tfrac{1}{4} = \tfrac{1}{4}, \tag{5.1a}$$

$$R_e = -(0 - \tfrac{1}{2}) = \tfrac{1}{2}, \tag{5.1b}$$

$$S_e = 0. \tag{5.1c}$$

From Table I it is indeed clear that this state is annihilated by the lowering operators (K_-, R_-, S_-, W_μ, and Y_μ). Following the same procedure with the constraint that the number of particles is odd, one obtains

$$K_0 = 0 + \tfrac{1}{4} = \tfrac{1}{4}, \tag{5.2a}$$

$$R_0 = -(\tfrac{1}{2} - \tfrac{1}{2}) = 0, \tag{5.2b}$$

$$S_0 = \tfrac{1}{2}. \tag{5.2c}$$

On reduction to $Sp(2,R)\times SO(3)\times SO(3)$, these representations decompose to [Eq. (3.2)]

$$(\tfrac{1}{4}\,\tfrac{1}{2}\,0) \to (\tfrac{1}{4})(\tfrac{1}{2},0) \oplus (\tfrac{3}{4})(0,\tfrac{1}{2}), \tag{5.3a}$$

$$(\tfrac{1}{4}\,0\,\tfrac{1}{2}) \to (\tfrac{1}{4})(0,\tfrac{1}{2}) \oplus (\tfrac{3}{4})(\tfrac{1}{2},0). \tag{5.3b}$$

Since

$$Sp(2,R): \ (\tfrac{1}{4}) \to N_b \text{ even}, \tag{5.4a}$$

$$(\tfrac{3}{4}) \to N_b \text{ odd}, \tag{5.4b}$$

$$SO(3)\times SO(3): \ (\tfrac{1}{2}\,0) \to N_f = 0,2; \quad S = 0, \tag{5.5a}$$

$$(0\,\tfrac{1}{2}) \to N_f = 1; \quad S = \tfrac{1}{2}, \tag{5.5b}$$

it is seen that Eqs. (5.3) are consistent with the conditions that the total number of particles be even and odd for the two representations, respectively.

B. Single-particle space $l=2, j=\frac{3}{2}, \frac{5}{2}$

This model space allows a description in the framework of the GM for nuclei for which the fermions can be treated as occupying $p_{3/2}\,f_{5/2}$ or $d_{3/2}\,d_{5/2}$ orbits. Examples of such a region may be provided by the Nickel isotopes.[21]

The large algebra is here $Osp(20/10,R)$,[10] so that more than two representations of $Osp(4/2,R)\times SO^{BF}$ are permitted; indeed, the number of representations is infinite. Since $SO(5)$ is a rank 2 algebra, the standard $SO^{BF}(5)$ representations are of the forms

$$\begin{matrix} (\sigma\,0) & (\sigma\,1) & (\sigma\,2) \\ & \sigma \geqslant 1 & \sigma \geqslant 2 \end{matrix}, \tag{5.6}$$

with corresponding [Eqs. (4.11)] $Osp(4/2)$ labels

$$\begin{aligned} &(0\,0): \ (\tfrac{5}{4}\,\tfrac{5}{2}\,0) \\ &(1\,0): \ (\tfrac{5}{4}\,2\,\tfrac{1}{2}) \quad (1\,1): \ (\tfrac{5}{4}\,\tfrac{3}{2}\,1) \\ &(\sigma\,0): \ (\tfrac{1}{2}\sigma + \tfrac{1}{4}\,\tfrac{3}{2}\,0) \quad (\sigma\,1): \ (\tfrac{1}{2}\sigma + \tfrac{1}{4}\,1\,\tfrac{1}{2}) \\ &(\sigma\,2): (\tfrac{1}{2}\sigma + \tfrac{1}{4}\,\tfrac{1}{2}\,0) \quad \sigma \geqslant 2. \end{aligned} \tag{5.7}$$

From the R-S symmetry of the model, for each representation $(K\ R\ S)$ corresponding to a standard $SO^{BF}(5)$ label (σ), there is one labeled (KSR), these being associated with the nonstandard labels which give (σ) on modification.[20]

For the reductions, Eq. (3.2) gives for example

$$(\tfrac{5}{4}\,\tfrac{5}{2}\,0)\to(\tfrac{5}{4})(\tfrac{5}{2},0)\oplus(\tfrac{7}{4})(2,\tfrac{1}{2})\oplus(\tfrac{9}{4})(\tfrac{3}{2},0), \tag{5.8a}$$

$$\begin{aligned}(K\,1\,\tfrac{1}{2})\to{}&(K)(1,\tfrac{1}{2})\oplus(K+\tfrac{1}{2})(\tfrac{3}{2},1)\oplus(K+\tfrac{1}{2})(\tfrac{3}{2},0)\oplus(K+\tfrac{1}{2})(\tfrac{1}{2},1)\oplus(K+\tfrac{1}{2})(\tfrac{1}{2},0)\\&\oplus(K+1)(2,\tfrac{1}{2})\oplus(K+1)(1,\tfrac{3}{2})\oplus 2(K+1)(1,\tfrac{1}{2})\oplus(K+1)(0,\tfrac{1}{2})\\&\oplus(K+\tfrac{3}{2})(\tfrac{3}{2},1)\oplus(K+\tfrac{3}{2})(\tfrac{3}{2},0)\oplus(K+\tfrac{3}{2})(\tfrac{1}{2},1)\oplus(K+\tfrac{3}{2})(\tfrac{1}{2},0)\\&\oplus(K+2)(1,\tfrac{1}{2}).\end{aligned} \tag{5.8b}$$

The complementary $SO^B(5)\times SO^F(5)$ products for Eq. (5.8a) are [Eqs. (4.3),(4.4)]

$$(0\,0)\times(0\,0)\oplus(1\,0)\times(1\,0)\oplus(2\,0)\times(2\,0), \tag{5.9}$$

which indeed are the only ways to construct the corresponding $SO^{BF}(5)$ representation (0 0) [Eq. (5.7)]. The decompositions of the other representations in Eq. (5.7) may also be verified in this way. We note that the multiplicity of 2 for the $Sp(2,R)\times SO(3)\times SO(3)$ representation $(K+1)(R,S)$ only occurs for $Osp(4/2,R)$ label $(K\,1\,\frac{1}{2})$ [Eq. (3.2)], corresponding to (2 1) of $SO^F(5)$; it is indeed the case that $(2\,1)\times(\sigma\,0)\to 2(\sigma\,1)\oplus\cdots$ for $\sigma\geqslant 2$ [cf. Eq. (5.7)], while products for all the other $SO^F(5)$ representations involved in this case are simply reducible.

VI. CONCLUSION

We have studied the positive discrete series representations of the noncompact superalgebra $Osp(4/2,R)$ appropriate to the realization discussed in Sec. II. For this superalgebra, there are at most 16 $Sp(2,R)\times SO(3)\times SO(3)$ multiplets for each representation; however, these multiplets need not all occur. We have given specific criteria, in terms of conditions on the $Osp(4/2,R)$ representation labels, for when each multiplet occurs.

We have also given several examples of our technique. These examples illustrate the calculational advantage of the superalgebra over the Bose–Fermi algebras; the decomposition given by Eq. (3.2) holds for all values of n, while a different Bose–Fermi algebra would have to be introduced for each new single-particle space. The $Osp(4/2,R)$ superalgebra has been applied to a physical example.[21]

ACKNOWLEDGMENTS

The authors would like to thank J. Cresson, G. Rosensteel, D. Pickrell, and W. Klink for useful and enlightening discussions. H. A. S. and P. H. would also like to thank the members of the Theoretical Nuclear Physics group at the University of Wisconsin for their hospitality during their stay.

The research of H. A. S., P. H., and B. R. B. was supported by U. S. National Science Foundation Grant No. PHY-8603883. The research of A. B. B. was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by U. S. National Science Foundation Grant No. PHY-8610851, and in part by a Presidential Young Investigator Award administered by the U. S. National Science Foundation.

APPENDIX A: Osp(4/2,R) SUPERALGEBRA

In this Appendix, we present the defining commutation and anticommutation relations for $Osp(4/2,R)$, where we have introduced, when it results in a simplification, the shorthand notation $Z^1_\mu=V_\mu$, $Z^2_\mu=Y_\mu$, $Z^3_\mu=X_\mu$, $Z^4_\mu=W_\mu$, and where $S_{\pm1}=(\mp 1/\sqrt{2})\,S_\pm$.

$$[K_0,K_\pm]=\pm K_\pm,\quad [R_0,R_\pm]=\pm R_\pm,\quad [S_0,S_\pm]=\pm S_\pm, \tag{A1a}$$

$$[K_+,K_-]=-2K_0,\quad [R_+,R_-]=2R_0,\quad [S_+,S_-]=2S_0, \tag{A1b}$$

$$[R_0,K_\pm]=[R_\pm,K_0]=[R_\pm,K_+]=[R_\pm,K_-]=0, \tag{A1c}$$

$$[S_0,R_\pm]=[S_\pm,R_0]=[S_\pm,R_+]=[S_\pm,R_-]=0, \tag{A1d}$$

$$[K_0,S_\pm]=[K_\pm,S_0]=[K_\pm,S_+]=[K_\pm,S_-]=0, \tag{A1e}$$

$$[R_0,K_0]=[S_0,R_0]=[K_0,S_0]=0, \tag{A1f}$$

$$[K_+,Z^i_\mu]=\tfrac{1}{2}((-)^i-1)Z^{i-1}_\mu,\quad [K_-,Z^i_\mu]=\tfrac{1}{2}(1-(-)^i)Z^{i+1}_\mu,$$
$$[K_0,Z^i_\mu]=\tfrac{1}{2}(-)^iZ^i_\mu, \tag{A1g}$$

$$[R_+,W_\mu]=[R_+,X_\mu]=0,\quad [R_+,V_\mu]=X_\mu,\quad [R_+,Y_\mu]=W_\mu, \tag{A1h}$$

$$[R_-,V_\mu]=[R_-,Y_\mu]=0,\quad [R_-,W_\mu]=Y_\mu,\quad [R_-,X_\mu]=V_\mu, \tag{A1i}$$

$$[R_0,V_\mu]=-\tfrac{1}{2}V_\mu,\quad [R_0,W_\mu]=\tfrac{1}{2}W_\mu,\quad [R_0,X_\mu]=\tfrac{1}{2}X_\mu,\quad [R_0,Y_\mu]=-\tfrac{1}{2}Y_\mu, \tag{A1j}$$

$$[S_\pm,Z^i_\mu]=\delta_{-\mu,\pm1/2}Z^i_{\mu\pm1},\quad [S_0,Z^i_\mu]=\mu Z^i_\mu,\quad \{Z^i_\mu,Z^i_{\mu'}\}=0, \tag{A1k}$$

$$\{V_\nu,X_\mu\}=(-)^{1/2-\nu}\delta_{\nu,-\mu}K_+,\quad \{W_\nu,Y_\mu\}=(-)^{1/2-\mu}\delta_{\nu,-\mu}K_-, \tag{A1l}$$

$$\{W_\nu,X_\mu\}=\tfrac{1}{2}(-)^{1/2-\nu}\delta_{\nu,-\mu}R_+,\quad \{V_\nu,Y_\mu\}=\tfrac{1}{2}(-)^{1/2-\nu}\delta_{\nu,-\mu}R_-, \tag{A1m}$$

$$\{V_\nu,W_\mu\}=-\sqrt{\tfrac{1}{2}}\,\delta_{\mu,\nu}S_{\mu+\nu}+\tfrac{1}{2}\delta_{\mu,-\nu}(-)^{1/2+\mu}\{2\mu S_0+2K_0+R_0\}, \tag{A1n}$$

$$\{X_\nu, Y_\mu\} = \sqrt{\tfrac{1}{2}}\,\delta_{\mu,\nu} S_{\mu+\nu} + \tfrac{1}{2}\delta_{\mu,-\nu}(-)^{1/2-\mu}\{2\mu S_0 + 2K_0 - R_0\}. \tag{A1o}$$

These relations [(A1)] remain invariant under Hermitian conjugation as defined by the following:

$$(K_0)^\dagger = K_0, \quad (R_0)^\dagger = R_0, \quad (S_0)^\dagger = S_0, \tag{A2a}$$

$$(K_+)^\dagger = K_-, \quad (R_+)^\dagger = R_-, \quad (S_+)^\dagger = S_-, \tag{A2b}$$

$$(V_\mu)^\dagger = (-)^{1/2-\mu} W_{-\mu}, \quad (X_\mu)^\dagger = (-)^{1/2+\mu} Y_{-\mu}. \tag{A2c}$$

APPENDIX B: ORTHOGONALIZED BASIS STATES

It is convenient to work with the combinations of the two $(M_K\ M_R\ M_S) = (1\,0\,0)$ raising operators ($X_{1/2}\ V_{-1/2}$ and $X_{-1/2}\ V_{1/2}$) which are symmetrized with respect to SO(3)×SO(3):

$$A^{(01)} = X_{1/2}V_{-1/2} + X_{-1/2}V_{1/2}, \tag{B1a}$$

$$A^{(10)} = X_{1/2}V_{-1/2} + V_{1/2}X_{-1/2} = X_{1/2}V_{-1/2} - X_{-1/2}V_{1/2} + K_+. \tag{B1b}$$

where $A^{(01)}(A^{(10)})$ is a scalar (vector) in R and a vector (scalar) in S. We denote the Sp(2,R)×SO(3)×SO(3) LWS by

$$|K\,R\,S;\ K\,K\,R-R\,S-S\rangle \equiv |K\,R\,S;\ K\,R\,S\rangle, \tag{B2}$$

where the first three quantum numbers in the ket on the right-hand side of Eq. (B2) label the Osp(4/2,R) representation, and the second set specifies that of Sp(2,R)×SO(3)×SO(3). The LWS for the Osp(4/2,R) representation $|\phi_0\rangle$ has Sp(2,R)×SO(3)×SO(3) quantum numbers K, R, S.

It is then possible to express the 16 normalized LWS of the Sp(2,R)×SO(3)×SO(3) representations in the right-hand column of Table II in terms of the 16 operators in the left-hand column, and where necessary to give orthogonality, combinations including K_+, R_+, and S_+,

$$|KRS;\ KRS\rangle = |\phi_0\rangle, \tag{B3a}$$

$$|KRS;\ K+\tfrac{1}{2}\,R+\tfrac{1}{2}\,S+\tfrac{1}{2}\rangle = \left(\frac{2}{(2K-R-S)}\right)^{1/2} V_{-1/2}|\phi_0\rangle, \tag{B3b}$$

$$|KRS;\ K+\tfrac{1}{2}\,R+\tfrac{1}{2}\,S-\tfrac{1}{2}\rangle = \left(\frac{1}{S(2S+1)(2K-R+S+1)}\right)^{1/2}\{(2S+1)V_{1/2} - S_+V_{-1/2}\}|\phi_0\rangle, \tag{B3c}$$

$$|KRS;\ K+\tfrac{1}{2}\,R-\tfrac{1}{2}\,S+\tfrac{1}{2}\rangle = \left(\frac{1}{R(2R+1)(2K+R-S+1)}\right)^{1/2}\{(2R+1)X_{-1/2} - R_+V_{-1/2}\}|\phi_0\rangle, \tag{B3d}$$

$$|KRS;\ K+\tfrac{1}{2}\,R-\tfrac{1}{2}\,S-\tfrac{1}{2}\rangle = \left(\frac{1}{2R(2R+1)S(2S+1)(2K+R+S+2)}\right)^{1/2} \times\{(2R+1)(2S+1)X_{1/2} - (2S+1)R_+V_{1/2} - (2R+1)S_+X_{-1/2} + S_+R_+V_{-1/2}\}, \tag{B3e}$$

$$|KRS;\ K+1\,R+1\,S\rangle = \left(\frac{4}{(2K-R-S)(2K-R+S+1)}\right)^{1/2} V_{1/2}V_{-1/2}|\phi_0\rangle, \tag{B3f}$$

$$|KRS;\ K+1\,R\,S+1\rangle = \left(\frac{4}{(2K-R-S)(2K+R-S+1)}\right)^{1/2} X_{-1/2}V_{-1/2}|\phi_0\rangle, \tag{B3g}$$

$$|KRS;\ K+1\,R\,S\rangle_{(01)} = \left(\frac{1}{2KS(S+1)(KZ-S(S+1))}\right)^{1/2}\{2K(S+1)A^{(01)} - S(S+1)K_+ - 2KS_+X_{-1/2}V_{-1/2}\}|\phi_0\rangle, \tag{B3h}$$

$$|KRS;\ K+1\,R\,S\rangle_{(10)} = \{[2(K+1)R(R+1)(2K-R-S)(2K+R-S+1)(2K-R+S+1)(2K+R+S+2) \times(KZ-S(S+1))]^{-1}\}^{1/2}\{2(R+1)(KZ-S(S+1))A^{(10)} - R(R+1)(Z+2S(S+1))K_+ - (2KZ-S(S+1))R_+V_{1/2}V_{-1/2} + 2(2K+1)R(R+1)[(S+1)A^{(01)} - S_+X_{-1/2}V_{-1/2}]\}|\phi_0\rangle, \tag{B3i}$$

$$|KRS;\ K+1\,R-1\,S\rangle = \{[(2R-1)R^2(2R+1)(2K+R-S+1)(2K+S+R+2)]^{-1}\}^{1/2}\{2R(2R+1)X_{1/2}X_{-1/2} - (2R+1)R_+A^{(10)} + R_+^2V_{1/2}V_{-1/2}\}|\phi_0\rangle, \tag{B3j}$$

$$|KRS;\ K+1\,R\,S-1\rangle = \{[(2S-1)S^2(2S+1)(2K-R+S+1)(2K+R+S+2)]^{-1}\}^{1/2}\{2S(2S+1)X_{1/2}V_{1/2} - (2S+1)S_+A^{(01)} + S_+^2X_{-1/2}V_{-1/2}\}|\phi_0\rangle, \tag{B3k}$$

$$|KRS;\ K+\tfrac{3}{2}R+\tfrac{1}{2}S+\tfrac{1}{2}\rangle = \{[(2K+1)(K+1)(2K-R-S)(2K+R-S+1)(2K-R+S+1)]^{-1}\}^{1/2}$$
$$\times\{2(2K+1)X_{-1/2}V_{1/2}V_{-1/2} - (2K-R+S+1)K_+V_{-1/2}\}, \tag{B3l}$$

$$|KRS;\ K+\tfrac{3}{2}R+\tfrac{1}{2}S-\tfrac{1}{2}\rangle$$
$$= [2(2K+1)(K+1)S(2S+1)(2K-R-S)(2K-R+S+1)(2K+S+R+2)]^{-1}\}^{1/2}$$
$$\cdot\{2(2K+1)(2S+1)X_{1/2}V_{1/2}V_{-1/2} - (2S+1)(2K-R-S)K_+V_{1/2}$$
$$+ (2K-R-S)K_+S_+V_{-1/2} - 2(2K+1)S_+X_{-1/2}V_{1/2}V_{-1/2}\}|\phi_0\rangle, \tag{B3m}$$

$$|KRS;\ K+\tfrac{3}{2}R-\tfrac{1}{2}S+\tfrac{1}{2}\rangle$$
$$= \{[2(2K+1)(K+1)R(2R+1)(2K-R-S)(2K+R-S+1)(2K+S+R+2)]^{-1}\}^{1/2}$$
$$\cdot\{2(2K+1)(2R+1)X_{1/2}X_{-1/2}V_{-1/2} - (2R+1)(2K-R-S)K_+X_{-1/2}$$
$$+ (2K-R-S)R_+K_+V_{-1/2} - 2(2K+1)R_+V_{1/2}X_{-1/2}V_{-1/2}\}|\phi_0\rangle, \tag{B3n}$$

$$|KRS;\ K+\tfrac{3}{2}R-\tfrac{1}{2}S-\tfrac{1}{2}\rangle$$
$$\{[(2K+1)(K+1)(2R+1)R(2S+1)S(2K+R-S+1)(2K-R+S+1)(2K+R+S+2)]^{-1}\}^{1/2}$$
$$\cdot\{2(2K+1)(2R+1)(2S+1)V_{1/2}X_{1/2}X_{-1/2} + (2K+R-S+1)K_+R_+S_+V_{-1/2}$$
$$- (2S+1)(2K+R-S+1)K_+R_+V_{1/2} - (2R+1)(2K+R-S+1)K_+S_+X_{-1/2}$$
$$+ 2(2K+1)(2S+1)R_+X_{1/2}V_{1/2}V_{-1/2} - 2(2K+1)(2R+1)S_+V_{-1/2}X_{+1/2}X_{-1/2}$$
$$- 2(2K+1)R_+S_+X_{-1/2}V_{+1/2}V_{-1/2} + (2R+1)(2S+1)(2K+R-S+1)K_+X_{1/2}\}|\phi_0\rangle, \tag{B3o}$$

$$|KRS;\ K+2\ R\ S\rangle$$
$$= \{[4(2K+1)(K+1)^2(2K+3)(2K-R-S)(2K+R-S+1)(2K-R+S+1)(2K+R+S+2)]^{-1}\}^{1/2}$$
$$\cdot\{8(2K+1)(K+1)X_{1/2}X_{-1/2}V_{1/2}V_{-1/2} - 2(2K+1)K_+R_+V_{1/2}V_{-1/2} + 2(2K+1)K_+S_+X_{-1/2}V_{-1/2}$$
$$- 2(2K+1)(S+1)K_+A^{(01)} - 2(2K+1)(2K-R+1)K_+A^{(10)} + (Z+2S(S+1))K^2_+\}|\phi_0\rangle, \tag{B3p}$$

where

$$Z = \{(2K+1)(2K+2) - R(R+1) - S(S+1)\}. \tag{B4}$$

[1]E. P. Wigner, Bull. Am. Math. Soc. **74**, 793 (1968).
[2]B. F. Bayman and Silverberg, Nucl. Phys. **16**, 625 (1960); J. P. Elliott, Proc. R. Soc. London A **245**, 128 (1958); J. N. Ginocchio, Ann. Phys. **126**, 234 (1980).
[3]A. Arima and F. Iachello, Phys. Rev. Lett. **35**, 1069 (1975).
[4]F. Iachello and O. Scholten, Phys. Rev. Lett. **43**, 679 (1979).
[5]A. Bohr and B. R. Mottelson, *Nuclear Structure* (Benjamin, New York, 1975), Vol. II.
[6]F. Iachello and S. Kuyucak, Ann. Phys. **136**, 19 (1981); A. B. Balantekin, I. Bars, R. Bijker, and F. Iachello, Phys. Rev. C **27**, 1761 (1983).
[7]A. B. Balantekin, I. Bars, and F. Iachello, Phys. Rev. Lett. **47**, 19 (1981).
[8]A. B. Balantekin, I. Bars, and F. Iachello, Nucl. Phys. A **370**, 284 (1981).
[9]J. Jolie, K. Heyde, P. van Isacker, and A. Frank, Nucl. Phys. A **466**, 1 (1987).
[10]H. A. Schmitt, P. Halse, A. B. Balantekin, and B. R. Barrett, Phys. Lett. B **210**, 1 (1988).
[11]K. Helmers, Nucl. Phys. A **69**, 593 (1965).
[12]H. Ui, Ann. Phys. (NY) **49**, 69 (1968).
[13]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), p. 349.
[14]See. Ref. 13, p. 290.
[15]V. Rittenberg, *VI International Colloquium on Group Theoretical Methods in Physics*, edited by P. Kramer and A. Rieckers (Springer, Berlin, 1978); M. Scheunert, *The Theory of Lie Superalgebras, An Introduction* (Springer, Berlin, 1979).
[16]J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison–Wesley, New York, 1988), pp. 4–20.
[17]A. B. Balantekin, H. A. Schmitt, and P. Halse, J. Math. Phys. **30**, 274 (1989).
[18]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[19]H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[20]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970), p. 42.
[21]H. A. Schmitt, P. Halse, A. B. Balantekin, and B. R. Barrett, Phys. Rev. C **39**, 2419 (1989).

The universal current distribution near the end of a tubular antenna

Hao-Ming Shen and Tai T. Wu
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts 02138

(Received 3 July 1989; accepted for publication 12 July 1989)

An asymptotic solution for the charge density and current near the end of a linear tubular antenna is obtained by means of the Wiener–Hopf technique. It is found that, due to the charge-repulsion effect, the distributions of charge per unit length and of current within a distance of the order of the radius from the end are significantly different from the sinusoidal distributions that are the solutions from approximate theories. Furthermore, the relative distributions are independent of the length of the antenna and of the frequency of operation.

I. INTRODUCTION

Since the dipole antenna is one of the most useful forms of antenna, it has been analyzed extensively for more than half a century. Most of these analyses have dealt with the case where the radius of the dipole is small. In such cases, since the vector potential on the surface of the dipole is roughly proportional to the current,[1] its distribution obtained theoretically is approximately sinusoidal or, more generally, a sum of several sinusoidal terms. This traveling-wave-like solution is mainly caused by the aspect of electromagnetic induction.

There is, however, another aspect of this problem of the current distributions on the dipole antenna. Coulomb repulsion between charges of the same sign is always present. The specific case of the Coulomb repulsion on a thin rod was analyzed by Maxwell himself,[2] who obtained interesting results more than a century ago. This repulsion leads to a larger charge density near the end of the rod than at the center. For the dipole antenna, this Coulomb repulsion leads to the result that the charge density and, hence, also the current cannot be described naturally and accurately as a sum of sinusoidal terms. Similar to the electrostatic case of Maxwell, this effect is most pronounced near the ends of the antenna, and is the subject of study for this paper. Indeed, this effect of Coulomb repulsion has been obtained in many experiments[3] but, to the best of our knowledge, the present paper is the first systematic theoretical analysis.

The basic idea of the present paper is to use the Wiener–Hopf procedure of solving integral equations.[4] This is intimately related to the observation that, except possibly for an overall constant, the current distribution near one end of a dipole antenna is approximately independent of the details near the other end. This idea has previously been used to obtain the current distribution and the input admittance of a long dipole antenna,[5] but the present problem is quite different and in many ways more difficult.

The procedure to be followed is as follows. First, the problem of the current distribution near the end of the center-driven dipole antenna where Coulomb repulsion is most pronounced—as already mentioned above—is formulated in terms of a Wiener–Hopf integral equation. Second, this Wiener–Hopf integral equation is solved exactly. Finally, this exact solution is evaluated approximately using the fact that the radius of the dipole antenna is small. This last step is technically complicated. Much effort has been directed toward changing the contours of the integration to improve the convergence. It is found that the expression for the current from the inverse Fourier transform is a universal formula in which no parameter other than the radius is involved. This means that within a range of about a quarter-wavelength, the repulsion effect dominates the distributions of the charge density and the current.

These universal current and charge-density distributions are shown in Fig. 1. With this result, the current distribution from other approximate theories, for which the accuracy near the end is poor, can be improved significantly. This is presented in a separate paper.[6]

II. WIENER–HOPF SOLUTION

A. Wiener–Hopf equations

The integral equation for the current is

$$\int_{-h}^{h} I(z')K(z-z')dz' = \frac{i}{\zeta_0}(C\cos kz + \tfrac{1}{2}V\sin k|z|), \quad (1)$$

where $\zeta_0 = 120\pi\,\Omega$, V is the driving voltage, k is the wave number, and the kernel is

$$K(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{ikR}}{4\pi R}d\theta, \quad (2)$$

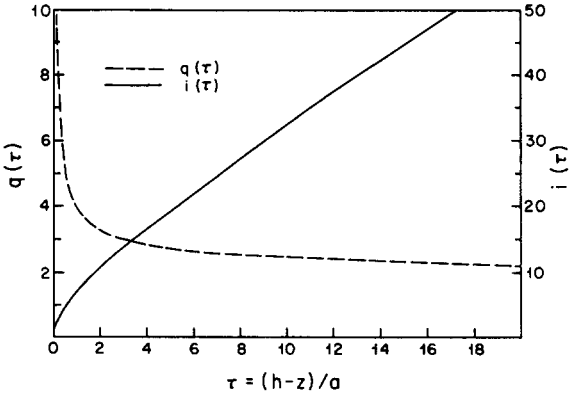


FIG. 1. Distributions of relative charge density $q(\tau)$ and relative current $i(\tau)$.

 0022-2488/89/112721-09$02.50

with

$$R = [z^2 + (2a \sin(\theta/2))^2]^{1/2}. \tag{3}$$

The requirement that the antenna be electrically thin is usually imposed in the form

$$ka \ll 1, \quad a/h \ll 1. \tag{4}$$

It is desired to determine the current $I(z)$ near the end of the tube from the integral equation (1) under the conditions (4). The origin is located at one end of the tube and the following notation is introduced:

$$x = h - z, \quad x' = h - z', \quad F(x) = I(h - x). \tag{5}$$

Then Eq. (1) becomes

$$\int_0^\infty dx' \, K(x - x')F(x') = G(x), \tag{6}$$

where

$$F(x) = 0, \quad \text{for } x > 2h,$$

$$G(x) = \begin{cases} (iV/\zeta_0)[\frac{1}{2}\sin k|h - x| + C\cos k(h - x)], \\ \quad \text{for } 0 < x < 2h, \\ \text{unknown}, \quad \text{for } x < 0, \ x > 2h. \end{cases} \tag{7}$$

The Fourier transform of the integral equation (6) is

$$\overline{K}(\zeta)\overline{F}(\zeta) = \overline{G}_+(\zeta) + \overline{G}_-(\zeta), \tag{8}$$

where

$$\overline{K}(\zeta) = \int_{-\infty}^{\infty} K(x)e^{-i\zeta x}\,dx$$

$$= \frac{i}{4} J_0(a\sqrt{k^2 - \zeta^2}) H_0^{(1)}(a\sqrt{k^2 - \zeta^2})$$

$$= \frac{1}{2\pi} I_0(a\sqrt{\zeta^2 - k^2}) K_0(a\sqrt{\zeta^2 - k^2}), \tag{9a}$$

$$\overline{F}(\zeta) = \int_0^\infty F(x)e^{-i\zeta x}\,dx = \int_0^{2h} F(x)e^{-i\zeta x}\,dx, \tag{9b}$$

$$\overline{G}_+(\zeta) = \int_0^\infty G(x)e^{-i\zeta x}\,dx, \tag{9c}$$

$$\overline{G}_-(\zeta) = \int_{-\infty}^0 G(x)e^{-i\zeta x}\,dx, \tag{9d}$$

and where $I_0(z)$ and $K_0(z)$ are modified Bessel functions.

B. Formal solution for $\overline{F}(\zeta)$

It is clear that $\overline{F}(\zeta)$ and $\overline{G}_+(\zeta)$ are analytic in the upper ζ plane, and $\overline{G}_-(\zeta)$ is analytic in the lower ζ plane. After factorization, (8) becomes

$$\overline{L}_+(\zeta)\overline{F}(\zeta) = \overline{L}_-(\zeta)\overline{G}_+(\zeta) + \overline{L}_-(\zeta)\overline{G}_-(\zeta), \tag{10}$$

where

$$\overline{L}_\pm(\zeta) = \exp\{[\ln \overline{K}(\zeta)]_\pm\}$$

$$= \exp\left[\frac{1}{2\pi i}\int_{-\infty \mp i\epsilon}^{\infty \mp i\epsilon} \frac{d\zeta'}{\zeta' - \zeta} \ln \overline{K}(\zeta')\right]. \tag{11}$$

After further decomposition of $\overline{L}_-(\zeta)\overline{G}_+(\zeta)$, Eq. (10) becomes

$$\overline{L}_+(\zeta)\overline{F}(\zeta) - [\overline{L}_-(\zeta)\overline{G}_+(\zeta)]_+ = [\overline{L}_-(\zeta)\overline{G}_+(\zeta)]_- + \overline{L}_-(\zeta)\overline{G}_-(\zeta). \tag{12}$$

The left and right sides of Eq. (12) are analytic in the upper and lower ζ planes, respectively, and have an overlapping strip. Thus, the left-hand side must be constant at zero, and the solution for $\overline{F}(\zeta)$ is

$$\overline{F}(\zeta) = \frac{1}{\overline{L}_+(\zeta)} \frac{1}{2\pi i} \int_{-\infty - i\epsilon}^{\infty - i\epsilon} \frac{d\zeta'}{\zeta' - \zeta} \overline{L}_-(\zeta')\overline{G}_+(\zeta'). \tag{13}$$

C. Alternative form of the solution

Since from (7) $G(x)$ is unknown in the region $(2h, \infty)$ and since from (9c) the same is true of $\overline{G}_+(\zeta)$, the expression (13) is only a formal solution. In other words, for a finite tubular antenna, it seems to be impossible to obtain a solution for the current $F(x)$ in the whole region $(0,2h)$ with the Wiener–Hopf technique alone. However, since $F(x)$ for x small corresponds to $\overline{F}(\zeta)$ for ζ large, it follows that, if an asymptotic solution for $\overline{F}(\zeta)$ can be obtained explicitly, it should be possible to derive a solution for the current within a distance of the order of the radius a from the end of the antenna.

In order to derive the asymptotic form of $\overline{F}(\zeta)$, the expression (13) must first be changed. From the definitions (9a)–(9c) and (11), it is not difficult to find the following asymptotic properties:

$$\overline{F}(\zeta) \sim \zeta^{-3/2}, \quad \text{for large } \zeta, \tag{14}$$

$$\overline{L}_+(\zeta) \sim \zeta^{-1/2}, \quad \text{for large } \zeta, \tag{15}$$

where (14) follows from $F(x) \sim \sqrt{x}$ for $x \ll a$. Thus, from (13),

$$\int_{-\infty - i\epsilon}^{\infty - i\epsilon} \frac{d\zeta'}{\zeta' - \zeta} \overline{L}_-(\zeta')\overline{G}_+(\zeta') = O(\zeta^{-2}), \quad \text{for large } \zeta. \tag{16}$$

Since, for large ζ,

$$\overline{G}_+(\zeta) = \int_0^\infty G(x)e^{-i\zeta x}\,dx \sim \frac{i}{\zeta} G(0), \tag{17}$$

it is convenient to introduce

$$\tilde{G}_+(\zeta) = \overline{G}_+(\zeta) - [i/(\zeta - k)]G(0), \tag{18}$$

which has the following asymptotic property:

$$\tilde{G}_+(\zeta) = O(\zeta^{-2}) \quad \text{for large } \zeta, \tag{19}$$

and the following identity:

$$\int_{-\infty - i\epsilon}^{\infty - i\epsilon} \frac{d\zeta'}{\zeta' - \zeta} \overline{L}_-(\zeta')\tilde{G}_+(\zeta') = \int_{-\infty - i\epsilon}^{\infty - i\epsilon} \frac{d\zeta'}{\zeta' - \zeta} \overline{L}_-(\zeta')\overline{G}_+(\zeta'). \tag{20}$$

When $\zeta \to \infty$ and with the asymptotic properties (16) and (19), it follows that

$$\int_{-\infty}^{\infty} d\zeta' \, \overline{L}_-(\zeta')\tilde{G}_+(\zeta') = 0. \tag{21}$$

Let the following quantity be defined:

$$\overline{M}(\zeta) = \int_{-\infty}^{\zeta} d\zeta' \, \overline{L}_-(\zeta')\tilde{G}_+(\zeta'), \tag{22}$$

so that

$$\overline{M}(-\infty)=\overline{M}(\infty)=0, \tag{23}$$

and

$$\frac{d\overline{M}(\zeta)}{d\zeta}=\overline{L}_-(\zeta)\widetilde{G}_+(\zeta). \tag{24}$$

The substitution of (20) and (22) into (13) leads to

$$\begin{aligned}\overline{F}(\zeta)&=\frac{1}{\overline{L}_+(\zeta)}\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\overline{L}_-(\zeta')\widetilde{G}_+(\zeta')\\&=\frac{1}{\overline{L}_+(\zeta)}\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\frac{d\overline{M}(\zeta')}{d\zeta'}.\end{aligned} \tag{25}$$

After integration by parts and with the condition (23), it follows that

$$\overline{F}(\zeta)=\frac{1}{\overline{L}_+(\zeta)}\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}d\zeta'\frac{\overline{M}(\zeta')}{(\zeta'-\zeta)^2}. \tag{26}$$

This is the alternative form for $\overline{F}(\zeta)$, in place of (13).

D. Asymptotic solution

Thus far, the alternative form (26) is still a precise solution and since $\overline{G}_+(\zeta)$ and $\overline{M}(\zeta)$ are unknown, it cannot be used to evaluate $F(x)$ in the whole region.

When ζ is large or of the order of $1/a$, an asymptotic approximation for $\overline{F}(\zeta)$ is readily obtained from (26), i.e., the integral in (26) may be approximated as follows:

$$\begin{aligned}&\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}d\zeta'\frac{\overline{M}(\zeta')}{(\zeta'-\zeta)^2}\\&\quad\sim\frac{1}{(\zeta+i\epsilon)^2}\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}d\zeta'\overline{M}(\zeta')\\&\quad=\frac{\mathscr{C}}{(\zeta+i\epsilon)^2},\end{aligned} \tag{27}$$

where $\mathscr{C}$ is an unknown constant. The substitution of (27) into (26) leads to an explicit solution, viz.,

$$\overline{F}(\zeta)\sim\frac{\mathscr{C}}{\overline{L}_+(\zeta)(\zeta+i\epsilon)^2},\quad\text{for }\zeta=O(a^{-1}), \tag{28}$$

where $\overline{L}_+(\zeta)$ is given in (11). The current near the end is given by the inverse Fourier transform of (28).

III. EVALUATION OF THE CURRENT

A. Modified form for $\overline{L}_+(\zeta)$

In order to evaluate the current near the end of the antenna, the formula for $\overline{L}_+(\zeta)$ needs to be modified. When $\zeta\to\infty$, it follows from (9a) that

$$\overline{K}(\zeta)\to\frac{1}{4\pi}\frac{1}{a\sqrt{\zeta^2-k^2}}.$$

This implies that the logarithmic integrand in (11) converges very slowly. Let

$$\widetilde{K}(\zeta)=4\pi a\sqrt{\zeta^2-k^2}\,\overline{K}(\zeta), \tag{29}$$

so that

$$\widetilde{K}(\zeta)\to1\quad\text{as }\zeta\to\pm\infty. \tag{30}$$

Now let

$$\widetilde{K}(\zeta)=\widetilde{L}_+(\zeta)/\widetilde{L}_-(\zeta). \tag{31}$$

From (11) it then follows that

$$\begin{aligned}\overline{L}_+(\zeta)&=\frac{1}{\sqrt{4\pi a(\zeta+k)}}\widetilde{L}_+(\zeta)\\&=\frac{1}{\sqrt{4\pi a(\zeta+k)}}\\&\quad\times\exp\left[\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\ln\widetilde{K}(\zeta')\right],\end{aligned} \tag{32}$$

$$\begin{aligned}\overline{L}_-(\zeta)&=\sqrt{4\pi a(\zeta-k)}\,\widetilde{L}_-(\zeta)\\&=\sqrt{4\pi a(\zeta-k)}\\&\quad\times\exp\left[\frac{1}{2\pi i}\int_{-\infty+i\epsilon}^{\infty+i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\ln\widetilde{K}(\zeta')\right].\end{aligned} \tag{33}$$

Due to (30), the integrand in (32) or (33) converges much faster than that in (11).

B. Charge density $q(\tau)$ near the end

The substitution of (32) into (28), and with the conditions (4), leads to

$$\begin{aligned}-i\zeta\overline{F}(\zeta)\sim\text{const}&\frac{1}{\sqrt{\zeta+i\epsilon}}\\&\times\exp\left[-\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\ln\widetilde{K}(\zeta')\right].\end{aligned} \tag{34}$$

The charge density near the end is the inverse Fourier transform of (34),

$$\begin{aligned}q(x/a)\sim\text{const}&\int_{-\infty}^{\infty}d\zeta\,e^{i\zeta x}\frac{1}{\sqrt{\zeta+i\epsilon}}\\&\times\exp\left[-\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}\frac{d\zeta'}{\zeta'-\zeta}\ln\widetilde{K}(\zeta')\right].\end{aligned} \tag{35}$$

Now if we let $\xi=a\zeta$, $\xi'=a\zeta'$, and $\tau=x/a$, and with (4) and (9a), it follows that

$$q(\tau)\sim\text{const}\int_{-\infty}^{\infty}d\xi\,e^{i\xi\tau}\frac{1}{\sqrt{\xi+i\epsilon}}\exp[A(\xi)], \tag{36}$$

where

$$A(\xi)=-\frac{\xi}{\pi i}\int_0^{\infty}\frac{d\xi'}{\xi'^2-\xi^2}\ln[2\xi' I_0(\xi')K_0(\xi')]. \tag{37}$$

This is the solution for the charge density near the end, except for an unknown factor.

The expression (36) is a universal formula, in which no parameter other than the radius a is involved. In other words, when the frequency in k or the length h is changed, the relative distributions of the charge per unit length (or current) near the end of the linear antenna remain the same, provided the thin-antenna conditions in (4) are satisfied.

The integral in (36) cannot be carried out analytically. Since there is no parameter involved, the numerical evaluation is universal. The integral (36) converges rapidly only when τ is small. The numerical evaluation is more convenient when the paths of integration are changed. These changes are described below.

C. Change in the path of integration for $A(\xi)$

From (37), $A(-\xi) = -A(\xi)$ and is analytic for $\bar{I}(\xi) > 0$. In the following derivation, ξ is initially confined to the region $\mathrm{Re}(\xi) > 0$. From (37),

$$A(\xi) = -\frac{1}{2\pi i}\int_0^\infty d\xi'\left(\frac{1}{\xi'-\xi} - \frac{1}{\xi'+\xi}\right) \times \ln[2\xi' I_0(\xi')K_0(\xi')]$$

$$= -\frac{1}{2\pi i}\lim_{T\to\infty}\left\{\int_0^T \frac{d\xi'}{\xi'-\xi}\ln[2\xi' I_0(\xi')K_0(\xi')] - \int_0^T \frac{d\xi'}{\xi'+\xi}\ln[2\xi' I_0(\xi')K_0(\xi')]\right\}. \quad (38)$$

After changing the paths of integration for the two integrals, respectively, from the real axis to the $\mp$ imaginary axes, (38) becomes

$$A(\xi) = -\frac{1}{2\pi i}\lim_{T\to\infty}\left\{\int_\epsilon^{-iT+\epsilon} \frac{d\xi'}{\xi'-\xi}\ln[2\xi' I_0(\xi')K_0(\xi')] - \int_\epsilon^{iT+\epsilon} \frac{d\xi'}{\xi'+\xi}\ln[2\xi' I_0(\xi')K_0(\xi')]\right\}. \quad (39)$$

We now let $\xi' = \mp it + \epsilon$, respectively, in the two integrals so that (39) becomes

$$A(\xi) = -\frac{1}{2\pi i}\lim_{T\to\infty}\left\{\int_0^T \frac{-i\,dt}{-it+\epsilon-\xi}\ln\left[2(-it+\epsilon)J_0(t+i\epsilon)\left(\frac{i\pi}{2}\right)H_0^{(1)}(t+i\epsilon)\right] - \int_0^T \frac{-i\,dt}{it+\epsilon+\xi}\ln\left[2(it+\epsilon)J_0(t-i\epsilon)\left(\frac{-i\pi}{2}\right)H_0^{(2)}(t-i\epsilon)\right]\right\} = -\frac{1}{\pi}\int_0^\infty \frac{dt}{t-i\xi}\Phi(t), \quad (40)$$

where

$$\Phi(t) = \frac{1}{2i}\ln\left[\frac{J_0(t+i\epsilon)H_0^{(1)}(t+i\epsilon)}{J_0(t-i\epsilon)H_0^{(2)}(t-i\epsilon)}\right] = \tan^{-1}\left[\frac{Y_0(t)}{J_0(t)}\right], \quad (41)$$

$\tan^{-1}(x)$ takes the principal value,[7] which is given in Fig. 2, and $J_0(z)$, $Y_0(z)$, $H_0^{(1)}(z)$, and $H_0^{(2)}(z)$ are Bessel functions.

D. Change in the path of integration for $q(\tau)$

$A(\xi)$ in the form (40) is analytic in the whole ξ plane except on the negative imaginary axis. With (40), it is convenient to change the path of ξ in the integral (36). Let

$$\left.\begin{matrix}\mathscr{E}(u)\\ \mathscr{F}(u)\end{matrix}\right\} = A(-iu) = -\frac{1}{\pi}\int_0^\infty \frac{dt}{t-u}\tan^{-1}\left[\frac{Y_0(t)}{J_0(t)}\right], \quad \begin{cases}\mathrm{Im}(u) > 0,\\ \mathrm{Im}(u) < 0.\end{cases} \quad (42)$$

From (36), let $\xi = -iu$,

$$q(\tau) = \int_0^{i\infty} \frac{-i\,du}{\sqrt{-iu+i\epsilon}}\exp[\mathscr{E}(u)] + \int_{-i\infty}^0 \frac{-i\,du}{\sqrt{-iu+i\epsilon}}\exp[\mathscr{F}(u)]. \quad (43)$$

After rotating the two paths along the negative and positive imaginary axes to the paths C_1 and C_2 (shown in Fig. 3) along the positive real axis, (43) becomes

$$q(\tau) = \int_{C_1} \frac{-i\,du\,e^{-u\tau}}{e^{-i\pi/4}\sqrt{u}}\exp[\mathscr{E}(u)] - \int_{C_2} \frac{-i\,du\,e^{-u\tau}}{e^{i3\pi/4}\sqrt{u}}\exp[\mathscr{F}(u)]. \quad (44)$$

In Fig. 3, μ_n, $n = 1,2,\ldots$, are the roots of $J_0(u)$, except $\mu_0 = 0$.

Contributions to the integral along the paths C_1 and C_2 come from the poles at μ_n and the straight path. After separating the path into two parts, we obtain

$$q(\tau) = e^{-i\pi/4}[q_p(\tau) + q_r(\tau)], \quad (45)$$

where

$$q_p(\tau) = \int_{C_0} \frac{du\,e^{-u\tau}}{\sqrt{u}}\{\exp[\mathscr{E}(u)] + \exp[\mathscr{F}(u)]\}, \quad (46a)$$

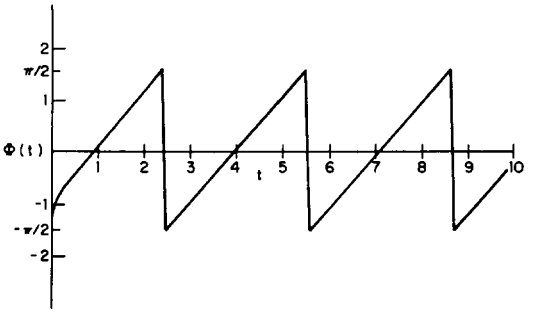


FIG. 2. Plot of $\Phi(t) = \tan^{-1}[Y_0(t)/J_0(t)]$ versus t.

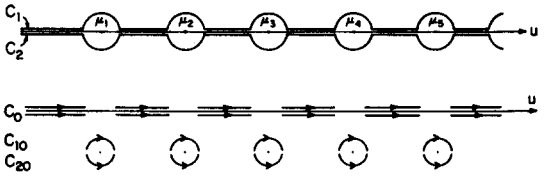


FIG. 3. Paths of integration for $q(\tau)$.

$$q_r(\tau)=\int_{C_{10}}\frac{du\,e^{-u\tau}}{\sqrt{u}}\exp[\mathscr{E}(u)]+\int_{C_{20}}\frac{du\,e^{-u\tau}}{\sqrt{u}}\exp[\mathscr{F}(u)]. \tag{46b}$$

The paths C_0, C_{10}, and C_{20} are shown in Fig. 3.

E. Contribution from the semicircles

Since, from the definition (42),

$$\mathscr{F}(u)=\mathscr{E}(u^*), \tag{47}$$

it is only necessary to evaluate $\mathscr{E}(u)$ near μ_n, $n=1,2,\ldots$. We define

$$\Phi_n(t)=\begin{cases}\tan^{-1}[Y_0(t)/J_0(t)]-\pi/2, & \mu_{n-1}<t<\mu_n,\\ \tan^{-1}[Y_0(t)/J_0(t)]+\pi/2, & \mu_n<t<\mu_{n+1},\\ \tan^{-1}[Y_0(t)/J_0(t)], & \text{otherwise}\end{cases} \tag{48}$$

which is shown in Fig. 4. Thus (42) becomes

$$\begin{aligned}\mathscr{E}(u)&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi_n(t)-\frac{1}{\pi}\int_{\mu_{n-1}}^{\mu_n}\frac{dt}{t-u}\frac{\pi}{2}\\&\quad+\frac{1}{\pi}\int_{\mu_n}^{\mu_{n+1}}\frac{dt}{t-u}\frac{\pi}{2}\\&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi_n(t)-\frac{1}{2}\left[\ln\left(\frac{-\mu_n+u}{-\mu_{n-1}+u}\right)\right.\\&\quad\left.-\ln\left(\frac{\mu_{n+1}-u}{\mu_n-u}\right)\right].\end{aligned} \tag{49}$$

When u is close to μ_n,

$$\mathscr{E}(u)=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi_n(t)+\frac{1}{2}\ln(\mu_n-\mu_{n-1})\times(\mu_{n+1}-\mu_n)-\ln(u-\mu_n)+\frac{\pi i}{2}. \tag{50a}$$

Also, from (47),

$$\mathscr{F}(u)=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi_n(t)+\frac{1}{2}\ln(\mu_n-\mu_{n-1})\times(\mu_{n+1}-\mu_n)-\ln(u-\mu_n)-\frac{\pi i}{2}. \tag{50b}$$

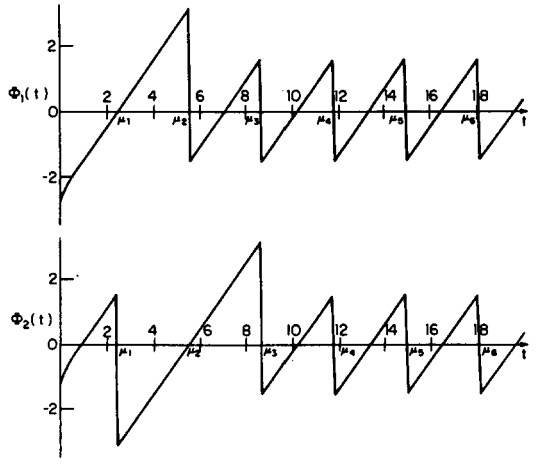


FIG. 4. Plot of $\Phi_n(t)$, $n=1,2$, versus t.

The substitution of (50a) and (50b) into (46b) leads to

$$q_r(\tau)=\sum_{n=1}e^{-\mu_n\tau}\frac{1}{\sqrt{\mu_n}}M_n\left(\int_{\frown}\frac{i\,d\mu}{u-\mu_n}+\int_{\smile}\frac{-i\,du}{u-\mu_n}\right), \tag{51}$$

where the symbols $\int_{\frown}$ and $\int_{\smile}$ denote the upper and lower semicircles around μ_n, respectively, and

$$M_n=\exp\left[-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-\mu_n}\Phi_n(t)+\frac{1}{2}\ln(\mu_{n+1}-\mu_n)(\mu_n-\mu_{n-1})\right]. \tag{52}$$

After taking the residue, (51) becomes

$$q_r(\tau)=\sum_{n=1}^\infty 2\pi e^{-\mu_n\tau}\frac{M_n}{\sqrt{\mu_n}}. \tag{53}$$

F. Contribution from the straight path

The contribution from the straight path is given by (46a) where, according to (42),

$$\begin{aligned}\mathscr{E}(u)=A(-iu+\epsilon)&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u-i\epsilon}\Phi(t)\\&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi(t)-i\Phi(u),\end{aligned} \tag{54a}$$

and

$$\begin{aligned}\mathscr{F}(u)=A(-iu-\epsilon)&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u+i\epsilon}\Phi(t)\\&=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi(t)+i\Phi(u).\end{aligned} \tag{54b}$$

Note that $\int_0^\infty$ denotes the principal value of the integral, i.e.,

$$\int_0^\infty=\lim_{\delta\to0}\left(\int_0^{u-\delta}+\int_{u+\delta}^\infty\right).$$

The substitution of (54a) and (54b) into (46a) leads to

$$q_p(\tau)=\int_0^\infty\frac{du\,e^{-u\tau}}{\sqrt{u}}\exp[B(u)], \tag{55}$$

where

$$B(u)=-\frac{1}{\pi}\int_0^\infty\frac{dt}{t-u}\Phi(t)+\ln[2\cos\Phi(u)]. \tag{56}$$

Formula (56) is not good for numerical evaluation. Since $\Phi(x)$ jumps at the zeros of the Bessel function J_0, two terms in (56) have opposite peaks at $x=\mu_n$. An alternative formula for numerical evaluation is given in the Appendix.

G. Results

The current at the end is, from (45),[8]

$$i(\tau)=ika\int_0^\tau q(\tau')d\tau'=e^{i\pi/4}ka[i_p(\tau)+i_r(\tau)], \tag{57}$$

where

$$i_p(\tau)=\int_0^\tau q_p(\tau')d\tau'=\int_0^\infty\frac{1-e^{-u\tau}}{u^{3/2}}\exp[B(u)], \tag{58a}$$

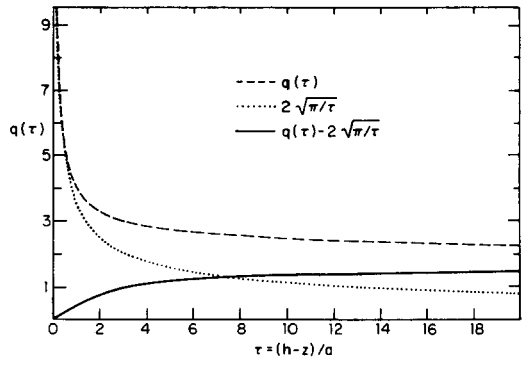


FIG. 5. Distribution of relative charge density $q(\tau)$ and square-root distribution as functions of $\tau = (h - z)/a$.

$$i_r(\tau) = \int_0^\tau q_r(\tau')d\tau' = \sum 2\pi \frac{1 - e^{-\mu_n \tau}}{\mu_n^{3/2}} M_n. \quad (58b)$$

The relative charge per unit length and current are shown, respectively, in Figs. 5 and 6 and listed in Table I. The figures also show the square-root distribution for purposes of comparison. It is well known that the square-root distribution of the current is valid only within the region x $(= h - z) \ll a$. When x is larger than a, the departure from the square-root distribution is significant.

Figure 7 gives the comparison with the experiment. The currents measured at the ends of tubular antennas of different lengths are taken from a report by Mack.[9] The figure shows that the asymptotic solution agrees very well with the measurements in the region $h - z < 10a$, in which other approximations are poor. The figure also indicates that the relative distribution near the end is universal, i.e., when the length increases from $h/\lambda = 0.25$ to 0.75, the distribution remains the same.

ACKNOWLEDGMENTS

The authors wish to thank Ms. Margaret Owens for her help in preparing this manuscript for publication.

This research was supported in part by the Strategic Defense Initiative Organization, Office of Innovative Science and Technology, and managed by the Harry Dia-

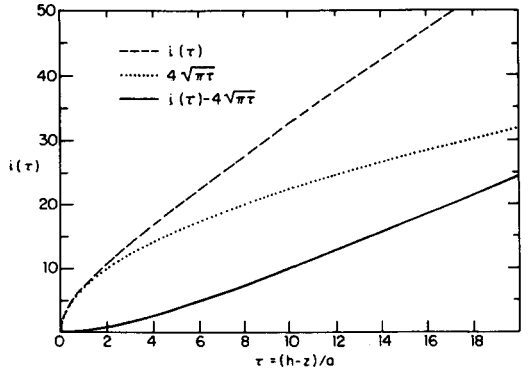


FIG. 6. Distribution of relative current $i(\tau)$ and square-root distribution as functions of $\tau = (h - z)/a$.

TABLE I. Asymptotic solution for relative charge density $q(\tau)$ and current $i(\tau)$ near end of tubular antenna; $\tau = (h - z)/a$.

τ	$q(\tau)$	$i(\tau)$
0.00	∞	0.000 000
0.05	15.888 23	1.586 858
0.10	11.275 75	2.133 377
0.15	9.244 049	2.641 178
0.20	8.041 176	3.071 435
0.25	7.226 934	3.451 876
0.30	6.631 376	3.797 680
0.35	6.173 275	4.118 281
0.40	5.808 206	4.417 569
0.45	5.509 525	4.700 063
0.50	5.260 304	4.969 698
0.55	5.048 866	5.228 570
0.60	4.867 229	5.476 535
0.65	4.709 419	5.716 188
0.70	4.571 166	5.948 951
0.75	4.448 999	6.175 151
0.80	4.340 320	6.394 034
0.85	4.243 127	6.609 552
0.90	4.155 676	6.820 101
0.95	4.076 613	7.025 929
1.00	4.004 820	7.228 565
1.05	3.939 366	7.428 108
1.10	3.879 466	7.624 542
1.15	3.824 455	7.817 417
1.20	3.773 767	8.007 630
1.25	3.726 915	8.195 458
1.30	3.683 478	8.381 082
1.35	3.643 094	8.565 387
1.40	3.605 446	8.747 086
1.45	3.570 259	8.926 984
1.50	3.537 290	9.105 198
1.55	3.506 329	9.281 848
1.60	3.477 187	9.457 012
1.65	3.449 702	9.630 760
1.70	3.423 727	9.803 162
1.75	3.399 133	9.974 284
1.80	3.375 806	10.144 19
1.85	3.353 595	10.312 58
1.90	3.332 450	10.480 37
1.95	3.312 289	10.647 20
2.00	3.293 041	10.812 81
2.05	3.274 640	10.977 05
2.10	3.256 967	11.141 79
2.15	3.240 013	11.304 78
2.20	3.223 742	11.466 82
2.25	3.208 110	11.628 28
2.30	3.193 021	11.789 18
2.35	3.178 467	11.948 97
2.40	3.164 441	12.108 21
2.45	3.150 916	12.267 29
2.50	3.137 770	12.425 03
2.55	3.125 070	12.581 98
2.60	3.112 794	12.738 14
2.65	3.100 857	12.893 91
2.70	3.089 269	13.049 86
2.75	3.078 045	13.204 99
2.80	3.067 118	13.358 53
2.85	3.056 475	13.512 40
2.90	3.046 151	13.665 55
2.95	3.036 077	13.817 58
3.00	3.026 250	13.969 90
3.05	3.016 705	14.121 71
3.10	3.007 358	14.272 90
3.15	2.998 243	14.423 57
3.20	2.989 381	14.573 54
3.25	2.980 660	14.723 54
3.30	2.972 175	14.873 02
3.35	2.963 878	15.022 36
3.40	2.955 740	15.171 47

TABLE I. (*Continued*).

τ	$q(\tau)$	$i(\tau)$
3.45	2.947 815	15.319 85
3.50	2.940 014	15.467 76
3.55	2.932 397	15.615 12
3.60	2.924 939	15.762 07
3.65	2.917 609	15.908 50
3.70	2.910 463	16.054 16
3.75	2.903 406	16.200 32
3.80	2.896 518	16.345 85
3.85	2.889 744	16.491 16
3.90	2.883 098	16.636 14
3.95	2.876 588	16.780 66
4.00	2.870 169	16.924 76
4.05	2.863 903	17.068 11
4.10	2.857 697	17.212 04
4.15	2.851 638	17.355 29
4.20	2.845 654	17.498 51
4.25	2.839 791	17.641 25
4.30	2.834 014	17.783 81
4.35	2.828 336	17.926 06
4.40	2.822 753	18.067 98
4.45	2.817 250	18.209 49
4.50	2.811 849	18.350 41
4.55	2.806 512	18.491 60
4.60	2.801 283	18.632 35
4.65	1.796 105	18.773 16
4.70	2.791 035	18.913 37
4.75	2.786 008	19.053 36
4.80	2.781 089	19.192 66
4.85	2.776 206	19.332 53
4.90	2.771 431	19.471 74
4.95	2.766 686	19.611 17
5.00	2.762 045	19.749 90
5.05	2.757 432	19.888 50
5.10	2.752 918	20.026 40
5.15	2.748 431	20.164 56
5.20	2.744 038	20.302 05
5.25	2.739 672	20.440 11
5.30	2.735 394	20.577 57
5.35	2.731 144	20.714 54
5.40	2.726 975	20.850 88
5.45	2.722 837	20.987 95
5.50	2.718 771	21.124 65
5.55	2.714 740	21.261 41
5.60	2.710 772	21.397 76
5.65	2.706 844	21.533 60
5.70	2.702 971	21.668 91
5.75	2.699 142	21.804 43
5.80	2.695 358	21.940 13
5.85	2.691 625	22.075 43
5.90	2.687 927	22.210 47
5.95	2.684 286	22.345 10
6.00	2.680 670	22.479 86
6.05	2.677 118	22.614 05
6.10	2.673 579	22.748 19
6.15	2.670 103	22.881 78
6.20	2.666 650	23.016 02
6.25	2.663 246	23.150 04
6.30	2.659 875	23.283 80
6.35	2.656 541	23.417 24
6.40	2.653 249	23.550 41
6.45	2.649 982	23.683 62
6.50	2.646 767	23.816 35
6.55	2.643 563	23.949 35
6.60	2.640 412	24.081 83
6.65	2.637 281	24.214 42
6.70	2.634 189	24.346 70
6.75	2.631 130	24.478 70

TABLE I. (*Continued*).

τ	$q(\tau)$	$i(\tau)$
6.80	2.628 096	24.610 47
6.85	2.625 105	24.741 92
6.90	2.622 126	24.874 16
6.95	2.619 194	25.005 93
7.00	2.616 278	25.137 07
7.05	2.613 397	25.267 71
7.10	2.610 545	25.398 43
7.15	2.607 714	25.529 55
7.20	2.604 925	25.660 22
7.25	2.602 142	25.791 50
7.30	2.599 601	25.922 34
7.35	2.596 678	26.052 80
7.40	2.593 981	26.182 88
7.45	2.591 318	26.312 69
7.50	2.588 664	26.442 95
7.55	2.586 049	26.572 77
7.60	2.583 447	26.702 69
7.65	2.580 873	26.832 41
7.70	2.578 326	26.961 80
7.75	2.575 791	27.091 10
7.80	2.573 291	27.219 98
7.85	2.570 803	27.349 00
7.90	2.568 340	27.477 79
7.95	2.565 904	27.606 34
8.00	2.563 478	27.735 05
8.05	2.561 083	27.863 35
8.10	2.558 701	27.991 97
8.15	2.556 341	28.120 49
8.20	2.554 007	28.248 73
8.25	2.551 681	28.377 17
8.30	2.549 384	28.505 22
8.35	2.547 101	28.632 96
8.40	2.544 836	28.760 37
8.45	2.542 599	28.887 40
8.50	2.540 365	29.015 32
8.55	2.538 158	29.142 88
8.60	2.535 968	29.270 34
8.65	2.533 790	29.397 84
8.70	2.531 639	29.524 98
8.75	2.529 495	29.651 84
8.80	2.527 372	29.778 37
8.85	2.525 271	29.904 70
8.90	2.523 174	30.032 04
8.95	2.521 102	30.159 02
9.00	2.519 044	30.285 57
9.05	2.516 997	30.411 68
9.10	2.514 974	30.537 44
9.15	2.512 958	30.663 56
9.20	2.510 960	30.789 58
9.25	2.508 983	30.915 28
9.30	2.507 008	31.041 29
9.35	2.505 056	31.166 95
9.40	2.503 118	31.292 49
9.45	2.501 188	31.418 14
9.50	2.499 279	31.543 45
9.55	2.497 380	31.668 81
9.60	2.495 492	31.794 13
9.65	2.493 626	31.919 12
9.70	2.491 764	32.044 30
9.75	2.489 918	32.169 29
9.80	2.488 092	32.293 97
9.85	2.486 265	32.418 69
9.90	2.484 458	32.543 09
9.95	2.482 667	32.667 35
10.00	2.480 879	32.792 04

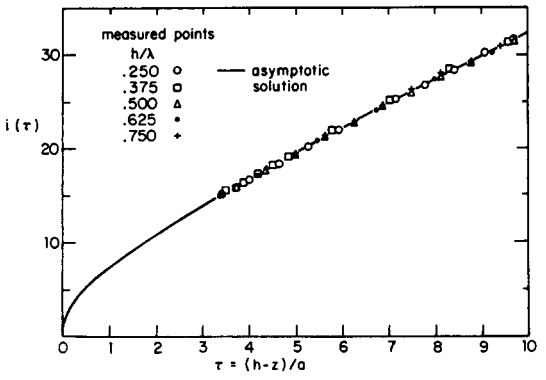


FIG. 7. Comparison of asymptotic solution for current $i(\tau)$ with experimental data of Mack (Ref. 9) for several values of h/λ.

mond Laboratories. The project also received partial support from the Joint Services Electronics Program under Grant No. N00014-89-J-1023 and from the Air Force Electronics System Division under Contract No. F19628-88-K-0024.

APPENDIX: ALTERNATIVE FORMULA FOR $B(u)$

We begin by introducing the saw-toothed function $\Phi_0(u)$,

$$\Phi_0(u) = u - \frac{\pi}{4} - \pi \cdot \mathrm{Int}\left(\frac{u}{\pi} + \frac{1}{4}\right). \tag{A1}$$

From (56), let

$$B(u) = -\frac{1}{\pi} \oint_0^\infty \frac{dt}{t-u}[\Phi(t) - \Phi_0(t)] - \frac{1}{\pi} \oint_0^\infty \frac{dt}{t-u} \Phi_0(t) + \ln[2\cos\Phi(u)]. \tag{A2}$$

Using integration by parts, the first integral in (A2) becomes

$$\begin{aligned} B_1(u) &= -\frac{1}{\pi} \oint_0^\infty \frac{dt}{t-u}[\Phi(t) - \Phi_0(t)] \\ &= -\frac{1}{\pi} \ln|t-u|[\Phi(t) - \Phi_0(t)]\Big|_0^\infty \\ &\quad + \frac{1}{\pi} \int_0^\infty dt \ln|t-u|[\Psi'(t) - 1] \\ &\quad + \sum_{n=1}^\infty \int_{\nu_n}^{\nu_{n+1}} dt \ln|t-u| \\ &\quad \times [\delta(t-\mu_{0n}) - \delta(t-\mu_n)], \end{aligned} \tag{A3}$$

where[10]

$$\Psi'(t) = \frac{2}{\pi t[J_0^2(t) + Y_0^2(t)]}, \tag{A4a}$$

$$\nu_n = (n - \tfrac{3}{4})\pi, \quad \mu_{0n} = (n - \tfrac{1}{4})\pi, \quad J_0(\mu_n) = 0, \quad n = 1,2,..., \tag{A4b}$$

so that

$$\nu_n < \mu_n \text{ (or } \mu_{0n}) < \nu_{n+1}. \tag{A5}$$

Since

$$\lim_{t\to\infty}[\Phi(t) - \Phi_0(t)] = 0, \tag{A6}$$

it follows that

$$\begin{aligned} B_1(u) &= -\frac{1}{4}\ln u + \frac{1}{\pi}\ln u \int_0^\infty dt\,[\Psi'(t) - 1] \\ &\quad + \frac{1}{\pi}\int_0^\infty dt \ln\left|1 - \frac{t}{u}\right|[\Psi'(t) - 1] \\ &\quad + \sum_{n=1}^\infty \frac{|\mu_{0n} - u|}{|\mu_n - u|}. \end{aligned} \tag{A7}$$

The first integral in (A7) is constant, viz.,

$$\begin{aligned} \int_0^\infty dt\,[\Psi'(t) - 1] &= \lim_{n\to\infty} \int_0^{\nu_n} dt\,[\Psi'(t) - 1] \\ &= \lim_{n\to\infty}[\Psi(\nu_n) - \Psi(0) - \nu_n] \\ &= \lim_{n\to\infty}\left[(n-1)\pi + \frac{\pi}{2} - \left(n - \frac{3}{4}\right)\pi\right] \\ &= \frac{\pi}{4}. \end{aligned} \tag{A8}$$

The substitution of (A8) into (A7) leads to

$$\begin{aligned} B_1(u) &= \frac{1}{\pi}\int_0^\infty dt \ln\left|1 - \frac{1}{u}\right|[\Psi'(t) - 1] \\ &\quad + \ln \prod_{n=0}^\infty \frac{|t - \mu_{0n}|}{|t - \mu_n|}. \end{aligned} \tag{A9}$$

The integral in (A9) is smooth, converges quickly, and is convenient for numerical analysis.

The second integral in (A2) can be carried out analytically as follows.

$$\begin{aligned} B_2(u) &= -\frac{1}{\pi} \oint_0^\infty \frac{dt}{t-u} \Phi_0(t) \\ &= -\frac{1}{\pi} \oint_0^{\pi/4} \frac{dt}{t-u}\left(t - \frac{\pi}{4}\right) \\ &\quad - \frac{1}{\pi} \sum_{n=1}^\infty \oint_{\nu_n}^{\nu_{n+1}} \frac{dt}{t-u} \Phi_0(t), \end{aligned} \tag{A10}$$

where ν_n is given in (A4b). After integration by parts and with $\Phi_0(\nu_n) = 0$,

$$B_2(u) = -\frac{1}{4} + \left(\frac{1}{4} - \frac{u}{\pi}\right)\ln\frac{|u-\pi/4|}{u} + \frac{1}{\pi}\sum_{n=1}\int_{\nu_n}^{\nu_{n+1}} dt \ln|t-u|[1-\pi\delta(t-\mu_n)]$$

$$= -\frac{1}{4} + \left(\frac{1}{4} - \frac{u}{\pi}\right)\ln\frac{|u-\pi/4|}{u} + \sum_{n=1}\left[\left(n+\frac{1}{4}-\frac{u}{\pi}\right)\ln\left|n+\frac{1}{4}-\frac{u}{\pi}\right| - \left(n-\frac{3}{4}-\frac{u}{\pi}\right)\ln\left|n-\frac{3}{4}-\frac{u}{\pi}\right|\right]$$

$$-1-\ln\left|n-\frac{1}{4}-\frac{u}{\pi}\right|$$

$$= -\frac{1}{4} + \left(\frac{1}{4} - \frac{u}{\pi}\right)\ln\frac{|u-\pi/4|}{u} + \lim_{N\to\infty}(A+B+C), \tag{A11}$$

where

$$A = \sum_{n=1}^{N}(n+u_0)\ln|n+u_0| - (n-1+u_0)\ln|n-1+u_0|$$

$$= \sum_{n=1}^{N}(n+u_0)\ln|n+u_0| - \sum_{n=0}^{N-1}(n+u_0)\ln|n+u_0|$$

$$= -u_0\ln|u_0| + (N+u_0)\ln|N+u_0|, \tag{A12}$$

$$B = \sum_{n=1}^{N}(-1) = -N, \tag{A13}$$

$$C = \sum_{n=1}^{N} -\ln\left|n-\frac{1}{2}+u_0\right| = -\ln\frac{\Gamma(N+\frac{1}{2}+u_0)}{|\Gamma(\frac{1}{2}+u_0)|}, \tag{A14}$$

and

$$u_0 = \tfrac{1}{4} - u/\pi. \tag{A15}$$

The substitution of (A12)–(A14) into (A11) leads to

$$B_2(u) = -\tfrac{1}{4} - u_0\ln(u/\pi) + \ln|\Gamma(\tfrac{1}{2}+u_0)|$$

$$+ \lim_{N\to\infty}[(N+u_0)\ln(N+u_0)$$

$$- N - \ln\Gamma(N+\tfrac{1}{2}+u_0)]. \tag{A16}$$

With Stirling's formula (when $z \gg 1$), viz.,

$$\ln\Gamma(z) = (z-\tfrac{1}{2})\ln z - z + \ln\sqrt{2\pi}, \tag{A17}$$

(A16) becomes

$$B_2(u) = -\frac{1}{4} - u_0\ln\frac{u}{\pi} + \ln\left|\Gamma\left(\frac{1}{2}+u_0\right)\right| - \ln\sqrt{2\pi} + \frac{1}{2} + u_0 - \lim_{N\to\infty}(N+u_0)\ln\left(1+\frac{\frac{1}{2}}{N+u_0}\right)$$

$$= -\frac{1}{4} - \ln\sqrt{2\pi} + \left(\frac{u}{\pi}-\frac{1}{4}\right)\left(\ln\frac{u}{\pi}-1\right) + \ln\left|\Gamma\left(\frac{3}{4}-\frac{u}{\pi}\right)\right|. \tag{A18}$$

The substitution of (A9) and (A18) into (A2) leads to the final formula for numerical evaluation. It is

$$B(u) = \frac{1}{\pi}\int_0^\infty dt \ln\left|1-\frac{t}{u}\right|[\Psi'(t)-1] + \left(\frac{u}{\pi}-\frac{1}{4}\right)\left(\ln\frac{u}{\pi}-1\right) + \ln\left[\Gamma\left(\frac{4}{3}-\frac{u}{\pi}\right)\cdot 2\cos\Phi_0(u)\right] - \frac{1}{4} - \frac{1}{2}\ln 2\pi$$

$$+ \ln\left[\frac{2\cos\Phi(u)}{2\cos\Phi_0(u)}\prod_{n=1}^{\infty}\frac{|u-\mu_{0n}|}{|u-\mu_n|}\right]. \tag{A19}$$

[1]T. T. Wu, in *Antenna Theory, Part I*, edited by R. E. Collin and F. J. Zucker (McGraw-Hill, New York, 1969), pp. 306–351.
[2]J. C. Maxwell, *The Scientific Papers of James Clerk Maxwell*, edited by W. D. Niven (Dover, New York, 1952), Vol. II, pp. 672–680.
[3]R. W. P. King and T. T. Wu, J. Res. Natl. Bur. Stand. D **69**, 429 (1965).
[4]B. Noble, *Wiener–Hopf Technique and Methods Based on Wiener–Hopf Technique* (Pergamon, New York, 1965).
[5]T. T. Wu, J. Math. Phys. **2**, 550 (1961).
[6]H.-M. Shen, R. W. P. King, and T. T. Wu, to be published in J. Electro. Waves Appl.

[7]From the argument of the logarithm, it is not difficult to find that when t passes the zeros of J_0, $\Phi(t)$ jumps from $\pi/2$ to $-\pi/2$.
[8]Here $i(\tau)$ is proportional to $F(\tau a)$ in (5).

[9]R. B. Mack, Cruft Laboratory Technical Report, No. 383, Harvard University, Cambridge, MA, 1963.

[10]$\Phi'(t) = \Psi'(t) - \Sigma_{n=1}^{\infty}\delta(t-\mu_n)$.
[11]$\nu_n = (n-\frac{3}{4})\pi$ are the zeros of $Y_0(t)$ when n is large and, thus, $\Psi(0) = -\pi/2$, $\Psi(\nu_n) = (n-1)\pi$.