Affine fields and the operator formulation of augmented scalar fields

John R. Klauder

Bell Laboratories, Murray Hill, New Jersey 07974

Affine fields, which can be used to replace the usual canonical fields, and which induce strictly homogeneous transformations of the underlying configuration space, are shown to be relevant in the operator formulation of augmented scalar field models. A characterization of the Hamiltonian and other basic generators by means of the expectation functional of the square of the field replaces the standard one based on the expectation functional of the field. Connection with previous work on augmented models is established through the form of the equation of motion for the field.

I. INTRODUCTION

Efforts to understand the singular nature of nonrenormalizable quantum field theories continue at a steady pace. In one point of view, which can be dubbed the "hard-core picture" and the essentials of which were outlined several years ago,[1] it was argued that from the standpoint of functional integration nonrenormalizable interactions act partially as relative hard cores projecting out certain of the field histories otherwise allowed by the free theory. This viewpoint manifestly explains the difficulties inherent in a perturbation expansion about the conventional free theory, and simultaneously suggests the search for a "pseudofree" theory, an alternative to the free theory that includes the hard-core interaction effects and about which any meaningful perturbation must necessarily take place. But such a general picture, however good it may prove to be, does not specify just how to go about constructing a theory incorporating the hard-core effects, either pseudofree or fully interacting. More recently, guided by highly specialized soluble models, a fairly definitive proposal was made as to just how to formulate a theory in such a way as to take into account the hard-core effects and thus to constitute a potential theory of nonrenormalizable interactions. At the same time this specific kind of proposal could be applied to models conventionally regarded as (super) renormalizable so as to construct new, noncanonical solutions.

The initial formulation of these alternative models was in the context of functional integration in which the conventional action was "augmented" by an additional term, a term which in no way altered the classical theory, but which effectively changed the basic measure in the functional integration.[2] Lattice-space techniques were used to give some precision to this formulation. Subsequently, functional differential equations were derived for the generating functional of the time-ordered Green's functions,[3] but this analysis, while providing a technically convenient formulation, is nevertheless a fairly straightforward consequence of the functional integration approach initially taken.

In this paper we reformulate these models once again, this time on the basis of an "operator approach," i.e., field operators, Hamiltonian, Hilbert space, and all that. This is not as straightforward as it might appear at first glance since the models in question are definitely *non*canonical, and so the ultimate formulation cannot be canonical. In particular, instead of the usual canonical Heisenberg group we find an alternative kinematical group, the affine group, to be relevant. In its simplest form the affine group is the two-parameter group of transformations $y \to p^{-1}y - q$, $p > 0$, of the real line into itself. Representations of this group have been studied previously,[4] and in addition we shall be able to draw on our own earlier studies of its kinematical applications in quantum theory.[5,6]

In our quantum field applications here, we are ultimately led to consider affine fields, $\phi(\mathbf{x})$ and $\kappa(\mathbf{x})$, $\mathbf{x}$ a spatial variable, that obey the (equal time) affine commutation relations (ACR), $[\phi(\mathbf{x}), \kappa(\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y})\phi(\mathbf{x})$, rather than the conventional canonical fields, $\phi(\mathbf{x})$ and $\pi(\mathbf{x})$, that obey the (equal time) canonical commutation relations (CCR), $[\phi(\mathbf{x}), \pi(\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y})$.

The reasons for choosing such an alternative, noncanonical operator formulation are not immediately evident, and our efforts in Sec. II are directed toward some heuristic motivational arguments. Section III is devoted to developing the affine field formulation and a number of results are derived that may be compared with those from the canonical approach. Especially noteworthy are the relations we derive that are analogs of the well-known characterization of a scalar field theory by the expectation functional of the sharp-time field as developed by Araki.[7] Our basic philosophy and primary results are briefly summarized in Sec. IV.

II. SIGNIFICANCE OF SCALE TRANSFORMATIONS IN AUGMENTED FIELD THEORIES

Let us here briefly recall the formal functional integral defining the augmented model for a quartic self-interacting scalar field.[2] In an n-dimensional Euclidean space-time the generating functional of interest takes the form ($dx \equiv d^n x$)

$$S'(h) = \mathcal{N}' \int \exp\{i \int h\Phi\, dx - \int \{\tfrac{1}{2}[(\nabla\Phi)^2 + (m^2 + X^2)\Phi^2] + \lambda\Phi^4\}\, dx\} \mathcal{D}\Phi \mathcal{D}X, \tag{1}$$

where X denotes an auxiliary field variable and the conventional action has been augmented by the term $\frac{1}{2}\int X^2\Phi^2\, dx$. With the formal convention that

$$\mathcal{D}\Phi \equiv \prod_x d\Phi(x), \quad \mathcal{D}X \equiv \prod_x dX(x), \tag{2}$$

it follows upon integrating out the X field that

$$S'(h) = \mathcal{N}' \int \exp[i \int h\Phi\, dx - W(\Phi)]\mathcal{D}'\Phi, \tag{3}$$

where W denotes the conventional Euclidean action,

$$W(\Phi) \equiv \int \{\tfrac{1}{2}[(\nabla\Phi)^2 + m^2\Phi^2] + \lambda\Phi^4\}\, dx, \tag{4}$$

and

$$\mathcal{D}'\Phi \equiv \prod_x d\Phi(x)/|\Phi(x)|. \tag{5}$$

In Ref. 2 meaning was given to such expressions [including (5)] by appealing to a lattice-space formulation (regularization).

The conventional approach to quartic self-interacting scalar field models is characterized instead by the generating functional

$$S(h) \equiv \mathcal{N} \int \exp[i \int h\Phi\, dx - W(\Phi)]\mathcal{D}\Phi, \tag{6}$$

where $\mathcal{D}\Phi$ is given in (2).

Formally speaking, the principal distinction between the conventional and augmented formulations is just the choice of the basic formal measure, $\mathcal{D}\Phi$ or $\mathcal{D}'\Phi$. These two measures have different formal invariances: $\mathcal{D}\Phi$ is *translation invariant*,

$$\mathcal{D}(\Phi + \Lambda) \equiv \mathcal{D}\Phi, \tag{7}$$

where $(\Phi + \Lambda)(x) \equiv \Phi(x) + \Lambda(x)$, for arbitrary $\Lambda(x)$; $\mathcal{D}'\Phi$ is *scale invariant*,

$$\mathcal{D}'(S\Phi) \equiv \mathcal{D}'\Phi, \tag{8}$$

where $(S\Phi)(x) \equiv S(x)\Phi(x)$, for arbitrary $S(x) > 0$. Such different properties lead to rather distinct consequences.

Recall that it is the goal of regularization and renormalization to convert a formal expression such as (6) into one like

$$S(h) \equiv \int \exp[i(h, \Phi)]\, d\mu(\Phi), \tag{9}$$

where μ denotes a probability measure (e.g., on the space of tempered distributions, $\mathcal{S}'$), and (h, Φ) is simply a more proper way of writing $\int h\Phi\, dx$ when Φ is not necessarily a function. Now suppose that for some real vector space $\mathcal{V}$ (e.g., C_0^∞ functions), it follows that

$$\mu(\Phi + \Lambda) \sim \mu(\Phi) \tag{10}$$

for all $\Lambda \in \mathcal{V}$; in words, the translated measure is equivalent to itself (has the same sets of zero measure). If (10) holds true, then (Euclidean space–time!) canonical fields exist; and in fact, for the $(\phi^4)_2$ and $(\phi^4)_3$ models of constructive quantum field theory[5] that is undoubtedly the case. For given the validity of (10) one can introduce the functional

$$S(h, k) \equiv \int \exp[i(h, \Phi)]\left(\frac{d\mu(\Phi - k)}{d\mu(\Phi)}\right)^{1/2} d\mu(\Phi), \tag{11}$$

for suitably many $h = h(x)$ and $k = k(x)$, which *defines* a representation of the canonical commutation relations (in integrated form), and which may be given concrete realization through a GNS (Gel'fand, Naimark, Segal) construction as unitary operators acting in a separable Hilbert space.[9]

It is intuitively clear that the translation invariant form $\mathcal{D}\Phi$ is needed in order for (10) to hold true. Indeed, if (6) is given meaning by regularizing $W(\Phi)$ (e.g., through a momentum-space cutoff and a finite space–time volume), then the so regularized form implicitly defines a measure that *automatically* satisfies (10). So much for the conventional approach.

As for the augmented formulation it can be expected that regularization and renormalization convert (3) into the expression

$$S'(h) \equiv \int \exp[i(h, \Phi)]\, d\mu'(\Phi), \tag{12}$$

where $\mu'(\Phi)$ is another probability measure on fields (say on $\mathcal{S}'$, again). Now suppose, instead of (10), that there exists a real vector space $\mathcal{V}'$ (e.g., C_0^∞ functions, again) such that

$$d\mu'(S\Phi) \sim d\mu'(\Phi) \tag{13}$$

for all $S(x) \equiv \exp[\hat{s}(x)]$ where $\hat{s} \in \mathcal{V}'$; in words, the scale-transformed measure is equivalent to itself. If (13) holds true, then the functional

$$S'(h, r) \equiv \int \exp[i(h, \Phi)]\left(\frac{d\mu'(r^{-1}\Phi)}{d\mu'(\Phi)}\right)^{1/2} d\mu'(\Phi) \tag{14}$$

is well defined for suitably many $h = h(x)$ and $r = r(x) > 0$. This functional *defines* a representation of a group, *the affine group*, and the GNS construction again provides a concrete realization of that group as unitary operators acting in a separable Hilbert space. Further discussion of this group and its representations are reserved to the next section.

In order for (13) to hold, it is intuitively clear that the scale invariant form $\mathcal{D}'\Phi$ is needed. Moreover, if (3) is given meaning by regularizing $W(\Phi)$ as above, then the so regularized form implicitly defines a measure that *automatically* satisfies (13).

In fact, even more is true. When a field problem is reduced by regularization to a problem of essentially a finite number of degrees of freedom, the associated probability measure is generally equivalent to itself under *both* translation and scaling. For instance, the probability measure $\mu^*(y)$ on the real line characterized by $d\mu^*(y) \equiv \exp(-\pi y^2)\, dy$ is equivalent to itself under both translation $(y \to y + a)$ and scaling $(y \to by,\ b > 0)$. Consequently, there is, in the regularized formulation of the model, no argument to favor the canonical approach over the affine one. The relevant choice is the one that survives the limit when the regularization is removed; and for the augmented formulation as in (3) it seems reasonably clear that the affine approach has the best chance of survival.

Another argument for scale transformations

Another argument in favor of scale transformations and thus for the affine group can be readily advanced. To make this point most clearly let us first recall the anharmonic oscillator Hamiltonian given by $(\lambda > 0)$

$$H=-\frac{1}{2}\frac{\partial^2}{\partial x^2}+\frac{1}{2}x^2+\frac{\lambda}{x^4}. \tag{15}$$

This example has been treated earlier by path-integral techniques,[10] and the hard-core effects alluded to at the outset of this paper arise in this case as well in that all paths that would otherwise meet or cross $x=0$ are projected out due to the singular potential. This projection of certain paths persists even for $\lambda\to 0^+$, and the limiting Hamiltonian differs from the harmonic oscillator. The analog for the differential equation of this path projection is represented by boundary conditions. In particular, the singular nature of the term x^{-4} permits one to freely rewrite (15) as

$$H=-\frac{1}{2}\frac{\partial^2}{\partial x^2}\bigg|_{\text{B.C.}}+\frac{1}{2}x^2+\frac{\lambda}{x^4}, \tag{16}$$

where B.C. denotes a Dirichlet boundary condition at $x=0$, i.e., the wavefunction $\psi(0)=0$. As $\lambda\to 0^+$ the effects of the potential disappear, but the boundary condition remains.

In this problem the Dirichlet boundary conditions effectively split the real line into two, dynamically disconnected parts. And what dynamics keeps separate, kinematics has no right to mix! Being guided by this dictum the relevant kinematic group for this singular problem is not the Heisenberg group—which, along with $\exp(ipx)$, includes translations $\exp[iq(-i\partial/\partial x)]$ that *change* the location of the singularity—but the affine group—which "replaces" translations by $\exp\{is[(-i\partial/\partial x)x+x(-i\partial/\partial x)]\}$, i.e., scale changes or dilations, and thus does *not* change the location of the singularity.[11] The moral of this example asserts that in the presence of differential operators burdened with Dirichlet boundary conditions *translation of the underlying configuration space should be avoided.*

Consider now, by analogy, the formal expression representing the Hamiltonian for a quartic self-coupled scalar field. Here we drop our Euclidean space formulation, choose sharp-time field operators (say at $t=0$) and employ a formal Schrödinger-like representation for the field and its conjugate momentum. The relevant expression for the Hamiltonian is

$$H=\int\left\{\frac{1}{2}\left[-\left(\frac{\delta}{\delta\phi(\mathbf{x})}\right)^2+[\nabla\phi(\mathbf{x})]^2+m^2\phi^2(\mathbf{x})\right]+\lambda\phi^4(\mathbf{x})\right\}d\mathbf{x} \tag{17}$$

(modulo normal ordering, etc.), where the spatial variable $\mathbf{x}\in R^{n-1}$ and we use ϕ (rather than Φ) for the Minkowski-space field variable. This is, of course, a standard prescription for H.[12]

Now, one can imagine (e.g., in high enough space dimensions) that the potential term $\int\phi^4(\mathbf{x})\,d\mathbf{x}$ diverges for field configurations otherwise allowed by the free Hamiltonian, and just as one finds for Schrödinger wave mechanics, it becomes appropriate to introduce Dirichlet boundary conditions into (17) so that the wave functionals vanish on the boundaries of the regions of infinite potential. Consequently, (17) is more properly interpreted as

$$H=\int\left\{\frac{1}{2}\left[-\left(\frac{\delta}{\delta\phi(\mathbf{x})}\right)^2\bigg|_{\text{B.C.}}+[\nabla\phi(\mathbf{x})]^2+m^2\phi^2(\mathbf{x})\right]+\lambda\phi^4(\mathbf{x})\right\}d\mathbf{x}. \tag{18}$$

Here B.C., roughly speaking, means that the wave functional vanishes for field configurations $\phi(\mathbf{x})$ which are allowed by the quadratic terms but for which the quartic term diverges. Correctly stated, B.C. means that the configuration space itself is restricted to regions of finite potential and the wave functional vanishes on the boundaries.[13] Never mind that this prescription falsely implies that nonrenormalizability arises for quartic models when $n>5$ (rather than $n>4$); the Hamiltonian, after all, does *not* provide the ultimate criterion in cases of *relative* divergence, and it is even a false indicator of the need for Dirichlet boundary conditions for problems similar to (15) when the potential has been generalized to $|x|^{-\alpha}$, $\alpha>0$.[10] What is ultimately relevant are the space—time configurations, i.e., the field histories. But we are here arguing for motivation and not for specifics, and the general idea behind (18) is perfectly valid—and this is so even though the boundaries involved are really pretty pathological.

Granting, then, the basic premise behind (18), one sees immediately that any kinematic field variable which leads to translation of the underlying field configuration space is highly undesirable. More acceptable is a transformation that preserves the singularity location such as the *scaling* transformation; for whenever $0<a\le S(\mathbf{x})\le b<\infty$, it follows that

$$a^4\int\phi^4(\mathbf{x})\,d\mathbf{x}\le\int S^4(\mathbf{x})\phi^4(\mathbf{x})\,d\mathbf{x}\le b^4\int\phi^4(\mathbf{x})\,d\mathbf{x}, \tag{19}$$

and so the location of the configuration space boundaries, cryptically symbolized in (18) by B.C., is preserved. It is not difficult to see that this argument is not limited to quartic interactions, nor for that matter to monomials.

Other arguments for scale transformations and affine fields can be advanced, but we limit ourselves to those presented above.

III. AFFINE FIELD FORMULATION OF SCALAR FIELDS

A. Canonical fields

The time-honored canonical approach to scalar quantum field theory begins with the introduction of a Hilbert space $\mathfrak{H}$ and a representation of the field $\phi(\mathbf{x})$ and momentum $\pi(\mathbf{x})$, local operators subject to the familiar CCR

$$[\phi(\mathbf{x}),\pi(\mathbf{y})]=i\delta(\mathbf{x}-\mathbf{y}). \tag{20}$$

Hamiltonians and other important generators are constructed out of π and ϕ. Of course, to make genuine operators out of π and ϕ themselves, the fields need to be smeared with test functions, and the domain on which the smeared CCR holds needs to be specified.

In a more rigorous approach to canonical quantization it is postulated that the smeared fields are self-adjoint operators and the Weyl (integrated) form of the commutation relations in terms of unitary operators is taken

as a replacement for (20). But this is just the step we must avoid, for when Dirichlet boundary conditions are basic to the Hamiltonian in a Schrödinger-like representation, the smeared momentum operator is at best only a *symmetric* operator and not self-adjoint. This is just a reflection of the nonexistence of unitary operators for general translations of the request configuration space. As stressed in the last section translations are to be avoided in such situations. Consequently, we must choose an alternative approach.

B. Affine fields: General case

Let us introduce the (bi)local field $\kappa(\mathbf{x}, \mathbf{y})$ formally defined as

$$\kappa(\mathbf{x}, \mathbf{y}) \equiv \tfrac{1}{2}[\pi(\mathbf{x})\phi(\mathbf{y}) + \phi(\mathbf{y})\pi(\mathbf{x})] \tag{21}$$

which is Hermitian (symmetric as an operator) but *not* symmetric in the variables $\mathbf{x}$ and $\mathbf{y}$. This local operator obeys the relation

$$[\kappa(\mathbf{x}, \mathbf{y}), \kappa(\mathbf{u}, \mathbf{v})]$$

$$= i\delta(\mathbf{y} - \mathbf{u})\kappa(\mathbf{x}, \mathbf{v}) - i\delta(\mathbf{x} - \mathbf{v})\kappa(\mathbf{u}, \mathbf{y}), \tag{22}$$

and thus the local operators $\kappa(\mathbf{x}, \mathbf{y})$ form the elements (formally) of a Lie algebra. We note also that

$$[\kappa(\mathbf{x}, \mathbf{y}), \phi(\mathbf{u})] = -i\delta(\mathbf{x} - \mathbf{u})\phi(\mathbf{y}) \tag{23}$$

so that the local operators $\kappa(\mathbf{x}, \mathbf{y})$ and $\phi(\mathbf{u})$ taken together also form a Lie algebra. Lastly we note as a consequence of (23) that

$$[\kappa(\mathbf{x}, \mathbf{y}), \phi(\mathbf{u})\phi(\mathbf{v})]$$

$$= -i\delta(\mathbf{x} - \mathbf{u})\phi(\mathbf{y})\phi(\mathbf{v}) - i\delta(\mathbf{x} - \mathbf{v})\phi(\mathbf{u})\phi(\mathbf{y}) \tag{24}$$

so that the local operators $\kappa(\mathbf{x}, \mathbf{y})$ and $\phi(\mathbf{u})\phi(\mathbf{v})$ form yet another Lie algebra. The fields $\kappa(\mathbf{x}, \mathbf{y})$ and $\phi(\mathbf{u})$ are the (general) affine fields.

Let us next consider the operators

$$\bar{U}[f, s] \equiv \exp[i\int f(\mathbf{u})\phi(\mathbf{u})\, d\mathbf{u}]$$

$$\times \exp[-i\int s(\mathbf{x}, \mathbf{y})\kappa(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y}], \tag{25}$$

and

$$U[w, s] \equiv \exp[i\int w(\mathbf{u}, \mathbf{v})\phi(\mathbf{u})\phi(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v}]$$

$$\times \exp[-i\int s(\mathbf{x}, \mathbf{y})\kappa(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y}]. \tag{26}$$

Here w is a symmetric function and s is in general a *non*symmetric function of their respective arguments, and for convenience we assume (say) that f, w, $s \in C_0^\infty$. We suppose the operators $\bar{U}$ and U are unitary and individually satisfy group laws that follow from the formal relations (22), (23) and (22), (24). [Of course, a rigorous formulation presumes unitary group representations and weak continuity, and deduces (25), (26), (22), (23), and (24).]

We shall require only a few basic properties of such groups of unitary operators and their generators. Note first that

$$\bar{U}^\dagger[f, s]\phi(\mathbf{z})\bar{U}[f, s] = U^\dagger[w, s]\phi(\mathbf{z})U[w, s]$$

$$= \int K(\mathbf{z}, \mathbf{t})\phi(\mathbf{t})\, d\mathbf{t}, \tag{27}$$

where K is a functional of s given by

$$K(\mathbf{z}, \mathbf{t}) \equiv \delta(\mathbf{z} - \mathbf{t}) + s(\mathbf{z}, \mathbf{t}) + \frac{1}{2!}\int s(\mathbf{z}, \mathbf{x})s(\mathbf{x}, \mathbf{t})\, d\mathbf{x}$$

$$+ \frac{1}{3!}\int s(\mathbf{z}, \mathbf{x})s(\mathbf{x}, \mathbf{y})s(\mathbf{y}, \mathbf{t})\, d\mathbf{x}\, d\mathbf{y} + \cdots. \tag{28}$$

It is to be noted that K is generally *non*symmetric in its arguments, and apart from an identity component [i.e., $\delta(\mathbf{z} - \mathbf{t})$], is also a C_0^∞ function whenever s is. Of course, if more general s are allowed, then the K are correspondingly generalized.

It should be remarked that (27) involves the field *homogeneously*. No field translation takes place, and so the location of the singularity and integrity of the boundary conditions in an expression such as (18) is preserved under action of the group U (or $\bar{U}$).

Next we note that

$$U^\dagger[w, s] = U[-w^{**}, -s], \tag{29}$$

where

$$w^{**}(\mathbf{u}, \mathbf{v}) \equiv \int w(\mathbf{x}, \mathbf{y})K(\mathbf{x}, \mathbf{u})K(\mathbf{y}, \mathbf{v})\, d\mathbf{x}\, d\mathbf{y}. \tag{30}$$

C. Distinguished states

We shall make particular use of the states

$$|w\rangle \equiv U[w, 0]\,|0\rangle$$

$$= \exp[i\int w(\mathbf{u}, \mathbf{v})\phi(\mathbf{u})\phi(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v}]\,|0\rangle. \tag{31}$$

Here $|0\rangle$ denotes a normalized fiducial vector that will be identified with the ground state of the Hamiltonian, and which is also a state invariant under spatial rotations and translations.

One should also consider the states defined by

$$|f\rangle \equiv \bar{U}[f, 0]\,|0\rangle \equiv \exp[i\int f(\mathbf{x})\phi(\mathbf{x})\, d\mathbf{x}]\,|0\rangle \tag{32}$$

as well. As usual the functional

$$\bar{E}(f) \equiv \langle 0|f\rangle \tag{33}$$

based on (32) characterizes a cyclic representation of the field. In like manner the functional

$$E(w) \equiv \langle 0|w\rangle \tag{34}$$

based on (31) characterizes the *even* subspace of that very representation.

It is evident that the functional $\bar{E}(f)$ uniquely determines the functional $E(w)$, and under certain circumstances, which we shall assume to hold, the converse is true. Specifically, suppose that $\bar{E}(-f) = \bar{E}(f)$; namely that the state $|0\rangle$ is itself *even*, as is the case when the Hamiltonian is field-reversal invariant, i.e., unchanged for $\phi \to -\phi$. This invariance holds for the models under consideration, and with $\bar{E}(f)$ an even functional, then the functional $E(w)$ uniquely determines the functional $\bar{E}(f)$.

D. Time-reversal invariance

When $|0\rangle$ is a time-reversal invariant state—a condition we shall assume satisfied—$\bar{E}(f)$ uniquely deter-

mines the state $|0\rangle$ (up to one—irrelevant—overall phase). And, moreover, since the state $|0\rangle$ is itself *even*, then the functional $E(w)$ also uniquely determines the time-reversal invariant state $|0\rangle$ just as well as $\bar{E}(f)$ does.

The time-reversal invariance of $|0\rangle$ has the following standard consequence [cf., Ref. 7]:

$$\langle 0|\, U[w, s]\,|0\rangle^* = \langle 0|\, U[-w, s]\,|0\rangle, \tag{35}$$

which when combined with (29) [and (30)] leads to the relation

$$\langle 0|\, U[w^{**}, -s]\,|0\rangle = \langle 0|\, U[w, s]\,|0\rangle. \tag{36}$$

On expanding both sides of this expression to lowest order in s, one finds that

$$\langle w|\, \kappa(\mathbf{x}, \mathbf{y})\,|0\rangle = \int w(\mathbf{x}, \mathbf{u})\, d\mathbf{u}\, \langle w\,|\, \phi(\mathbf{u})\phi(\mathbf{y})\,|0\rangle, \tag{37}$$

and after a further manipulation it follows that

$$\langle w|\, \kappa(\mathbf{x}, \mathbf{y})\,|w'\rangle = \int [w(\mathbf{x}, \mathbf{u}) + w'(\mathbf{x}, \mathbf{u})]\, d\mathbf{u} \langle w|\, \phi(\mathbf{u})\phi(\mathbf{y})\,|w'\rangle. \tag{38}$$

This expression determines the matrix element of $\kappa(\mathbf{x}, \mathbf{y})$ in the even subspace. Note that

$$\langle w|\, \phi(\mathbf{u})\phi(\mathbf{y})\,|w'\rangle \tag{39}$$

is just a functional of the difference variable $w'(\mathbf{x}, \mathbf{y}) - w(\mathbf{x}, \mathbf{y})$. In fact, this functional is nothing but

$$\frac{\delta}{i\delta v(\mathbf{u}, \mathbf{y})} E(v), \tag{40}$$

evaluated for $v(\mathbf{x}, \mathbf{y}) = w'(\mathbf{x}, \mathbf{y}) - w(\mathbf{x}, \mathbf{y})$.

E. Dynamical considerations

As standard we assume the Hamiltonian formally appears like

$$H = \tfrac{1}{2} \int \pi^2(\mathbf{x})\, d\mathbf{x} + W(\phi). \tag{41}$$

Traditionally one exploits the relation

$$i[H, \phi(\mathbf{x})] = \pi(\mathbf{x}) \tag{42}$$

based on the CCR, but in view of anticipated domain difficulties we prefer to avoid the use of operator π in isolation. Instead we appeal to the formal relation

$$\begin{aligned} i[H, \phi(\mathbf{x})\phi(\mathbf{y})] &= \pi(\mathbf{x})\phi(\mathbf{y}) + \phi(\mathbf{x})\pi(\mathbf{y}) \\ &= \tfrac{1}{2}\{\pi(\mathbf{x})\phi(\mathbf{y}) + \phi(\mathbf{y})\pi(\mathbf{x})\} \\ &\quad + \tfrac{1}{2}\{\pi(\mathbf{y})\phi(\mathbf{x}) + \phi(\mathbf{x})\pi(\mathbf{y})\}, \end{aligned} \tag{43}$$

and we take the conclusion of this exercise, namely

$$i[H, \phi(\mathbf{x})\phi(\mathbf{y})] = \kappa(\mathbf{x}, \mathbf{y}) + \kappa(\mathbf{y}, \mathbf{x}), \tag{44}$$

as our dynamical replacement for (42). Here is a relation among self-adjoint fields that respects the imposition of Dirichlet boundary conditions.

Because of the dependence of (39) only on $w' - w$, a relation for the matrix elements of H is readily obtained. As usual we assume $H|0\rangle = 0$ and $H \geq 0$. The stated conditions uniquely imply that

$$\langle w|H|w'\rangle = 2\int w(\mathbf{u}, \mathbf{t}) w'(\mathbf{t}, \mathbf{v}) \langle w|\, \phi(\mathbf{u})\phi(\mathbf{v})\,|w'\rangle\, d\mathbf{t}\, d\mathbf{u}\, d\mathbf{v}, \tag{45}$$

which, besides filling the desired spectral conditions, satisfies (44) as may be seen on examining

$$\left(\frac{\delta}{\delta w(\mathbf{x}, \mathbf{y})} + \frac{\delta}{\delta w'(\mathbf{x}, \mathbf{y})}\right) \langle w|H|w'\rangle = \langle w|\, i[H, \phi(\mathbf{x})\phi(\mathbf{y})]\,|w'\rangle \tag{46}$$

and using (38) [always remembering that $w(\mathbf{x}, \mathbf{y})$ is symmetric in its arguments]. Thus (45) fixes the matrix elements of the Hamiltonian in the even subspace.

F. The even subspace has got it all

It is not difficult to see that in fact the matrix elements of the Hamiltonian in the even subspace actually determine the general matrix elements. For, in a Schrödinger-like representation, there is a dense set of smooth elements $\Psi_e(\phi) = \Psi_e(-\phi)$ in the even subspace with support away from $(f_0, \phi) \equiv 0$ to which there corresponds a dense set of smooth elements $\Psi_0(\phi) \equiv \mathrm{sgn}(f_0, \phi)\Psi_e(\phi) = -\Psi_0(-\phi)$ in the odd subspace, where $f_0 = f_0(\mathbf{x})$ is an arbitrary but fixed, nonzero test function. Since the Hamiltonian involves only a finite number of derivatives (thus making no finite shift) and maps the even (odd) subspace into itself, it follows that

$$\langle \Psi_0|H|\Psi_0'\rangle \equiv \langle \Psi_e|H|\Psi_e'\rangle \tag{47}$$

and

$$\langle \Psi_0|H|\Psi_e'\rangle \equiv 0 \tag{48}$$

for arbitrary elements Ψ_e (or Ψ_0) and Ψ_e' (or Ψ_0') of the dense set. Closure does the rest.[14] A completely parallel argument can be used to determine the general matrix elements of $\kappa(\mathbf{x}, \mathbf{y})$.

The relations derived above for the matrix elements of the Hamiltonian, etc., for an affine formulation are analogous to those derived by Araki[7] for a canonical formulation. In our discussion we have only used the momentum operator $\pi(\mathbf{x})$ as a heuristic guide, and we have not relied on it in our basic results in any way.

G. Conjugate momentum—No

In fact, one can extract from the foregoing narrative a scenario in which the canonical momentum $\pi(\mathbf{x})$ *never* appears. In this case one supposes that the field $\phi(\mathbf{x})$ and the affine field $\kappa(\mathbf{x}, \mathbf{y})$, which satisfies (22), are basic, and that they are related to the Hamiltonian H through (44). The rest of the scenario is basically unchanged: The functional $\bar{E}(f)$ in (33) uniquely determines the functional $E(w)$ in (34); A total set of even matrix elements of $\kappa(\mathbf{x}, \mathbf{y})$ and H can be expressed in terms of the functional $E(w)$, granting only that the ground state of H is time-reversal invariant; and for an even time-reversal invariant ground state, $E(w)$ uniquely fixes $\bar{E}(f)$, and general matrix elements of $\kappa(\mathbf{x}, \mathbf{y})$ and H are determined.

The elimination of the conjugate momentum $\pi(\mathbf{x})$ in

the description is of great value when the nature of $\pi(\mathbf{x})$ is uncertain. Moreover, there are models for which the conjugate momentum exists as no more than a form.

H. Conjugate momentum—Yes

This next paragraph is in the nature of a remark. While we have been doing our best to convince the reader that the affine formulation can do without $\pi(\mathbf{x})$, the preceding development is fully consistent with the existence of a self-adjoint local field $\pi(\mathbf{x})$. A case in point is the free relativistic scalar field of mass m for which the expectation functional (33) reads

$$\bar{E}_0(f) = \exp\{-\tfrac{1}{4}\int (k^2+m^2)^{-1/2}|\tilde{f}(\mathbf{k})|^2\,d\mathbf{k}\}$$
$$\equiv \int \exp[i(f,\phi)]\,d\mu_0(\phi). \tag{49}$$

It is known that the measure $\mu_0(\phi)$ is equivalent to itself under translation, $\mu_0(\phi+\Lambda)\sim\mu_0(\phi)$, for Λ such that $\int(k^2+m^2)^{1/2}|\tilde{\Lambda}(\mathbf{k})|^2\,d\mathbf{k}<\infty$. This equivalence guarantees the existence of the canonical field $\pi(\mathbf{x})$ and the CCR. But this particular measure is also known to be equivalent to itself under certain homogeneous transformations, $\mu_0(\phi+T\phi)\sim\mu_0(\phi)$, provided $(1+T)^{-1}$ is bounded and T is sufficiently smooth, namely if

$$\int|\tilde{T}(\mathbf{k},\mathbf{q})|^2\,d\mathbf{k}\,d\mathbf{q}<\infty, \tag{50}$$

where $\tilde{T}(\mathbf{k},\mathbf{q})$ is the kernel of T in momentum space.[15] This equivalence guarantees the existence of the affine field $\kappa(\mathbf{x},\mathbf{y})$ and the ACR in the general sense we have discussed in this section up to the present. [The existence of $\kappa(\mathbf{x},\mathbf{y})$ can be traced from (27), (28), and related relations.]

On the other hand, strictly local field scaling, $\phi(\mathbf{x})\to S(\mathbf{x})\phi(\mathbf{x})$, $S(\mathbf{x})>0$, falls outside the class of homogeneous transformations covered by (50) [provided $S(\mathbf{x})\neq 1$], and it is known that $\mu_0(\phi)$ is *not* equivalent to itself after such a transformation.

I. Affine fields: Special case

Based on the analysis up to this point, we now begin to discuss affine fields that lead to local field scaling. We first observe that we can restrict (22), (23), and (24) to $\kappa(\mathbf{x},\mathbf{x})$, i.e., to equal arguments, without any formal difficulty. Specifically, if we set

$$\kappa(\mathbf{x})\equiv\kappa(\mathbf{x},\mathbf{x}), \tag{51}$$

then (22) implies

$$[\kappa(\mathbf{x}),\kappa(\mathbf{y})]=0 \tag{52}$$

while (23) and (24) become

$$[\kappa(\mathbf{x}),\phi(\mathbf{u})]=-i\delta(\mathbf{x}-\mathbf{u})\phi(\mathbf{u}) \tag{53}$$

and

$$[\kappa(\mathbf{x}),\phi(\mathbf{u})\phi(\mathbf{v})]=-i[\delta(\mathbf{x}-\mathbf{u})+\delta(\mathbf{x}-\mathbf{v})]\phi(\mathbf{u})\phi(\mathbf{v}), \tag{54}$$

respectively. A Lie algebra structure of these relations is apparent; and indeed, one can say that the new Lie algebras (52)–(54) are subalgebras of the previous Lie algebras (22)–(24).

To see how such a subalgebra can be extracted consider the generator

$$\int s(\mathbf{x},\mathbf{y})\kappa(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} \tag{55}$$

that enters (25) and (26). Let attention be restricted to only special s, namely those of the form

$$s(\mathbf{x},\mathbf{y})=\hat{s}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}), \tag{56}$$

in which case (55) becomes

$$\int\hat{s}(\mathbf{x})\kappa(\mathbf{x})\,d\mathbf{x}. \tag{57}$$

More properly, the transition to (57) arises through consideration of a *sequence* of functions $s_j(\mathbf{x},\mathbf{y})$ that converge to $\hat{s}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})$ as $j\to\infty$.

An immediate and important property of the so derived subgroup of operators pertains to homogeneous field transformations. For it follows from (27) and (28) that

$$\exp[i\int\hat{s}(\mathbf{y})\kappa(\mathbf{y})\,d\mathbf{y}]\phi(\mathbf{x})\exp[-i\int\hat{s}(\mathbf{y})\kappa(\mathbf{y})\,d\mathbf{y}]$$
$$=S(\mathbf{x})\phi(\mathbf{x}), \tag{58}$$

where $S(\mathbf{x})\equiv\exp[\hat{s}(\mathbf{x})]$. Consequently, when this type of singular subgroup extraction works properly, it leads to the local field scaling of interest.

The usefulness of the kind of subgroup extraction we have in mind here is highly representation dependent. In particular, it is necessary that the representation of the affine field $\kappa(\mathbf{x},\mathbf{y})$ *admit* a sequence of smearing functions $s_j(\mathbf{x},\mathbf{y})$ that weakly converges to $\hat{s}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})$ and in addition yields a meaningful operator. This is not always the case, and fails to be true for the representation built on the free field as characterized by (49).

To examine this convergence question in detail one can study the functional

$$E(w,s)\equiv\langle 0|\,U[w,s]\,|0\rangle \tag{59}$$

that determines the representation of U uniquely up to unitary equivalence. Consider a sequence $s_j(\mathbf{x},\mathbf{y})$ that converges to $\hat{s}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})$, and evaluate the limiting form of E; namely introduce

$$E(w,\hat{s})\equiv\lim_{j\to\infty}E(w,s_j). \tag{60}$$

For the general case this limit will vanish, and thus be of no use to us; however, if $E(w,\hat{s})$ is nonvanishing for sufficiently many functions $w(\mathbf{x},\mathbf{y})$ and $\hat{s}(\mathbf{x})$, and has reasonable continuity properties, then we have obtained a useful subgroup extraction. The expression $E(w,\hat{s})$ uniquely characterizes the relevant affine subgroup up to unitary equivalence.

An exercise similar to that in (60) can be carried through for $\bar{E}(f,s)\equiv\langle 0|\,\bar{U}[f,s]\,|0\rangle$ to determine $\bar{E}(f,\hat{s})$, and the resultant expression can be given a comparatively simple integral representation. Namely,

$$\bar{E}(f,\hat{s}) = \langle 0|\exp[i\phi(f)]\exp[-i\kappa(\hat{s})]|0\rangle = \int \exp[i(f,\phi)]\left(\frac{d\mu(S^{-1}\phi)}{d\mu(\phi)}\right)^{1/2} d\mu(\phi), \tag{61}$$

where $S(\mathbf{x}) = \exp[\hat{s}(\mathbf{x})]$, and where we have used a standard notation for smeared fields. This relation is evidently similar to (14), except that the present case deals with the sharp-time local field operators $\kappa(\mathbf{x})$ and $\phi(\mathbf{x})$. It is reasonable to assume that every representation of the special affine fields $\kappa(\mathbf{x})$ and $\phi(\mathbf{x})$ arises as the subalgebra (for $\mathbf{x}=\mathbf{y}$) of some representation of the general affine fields $\kappa(\mathbf{x},\mathbf{y})$ and $\phi(\mathbf{x})$.

J. Dynamics reconsidered

The kinematical operator $\kappa(\mathbf{x})$ singled out above leads to important consequences for dynamics. In showing this we proceed in a formal and heuristic fashion. Observe that

$$\kappa(\mathbf{x}) = \tfrac{1}{2}[\pi(\mathbf{x})\phi(\mathbf{x}) + \phi(\mathbf{x})\pi(\mathbf{x})] = \phi(\mathbf{x})\pi(\mathbf{x}) + (c\text{-number}), \tag{62}$$

where in addition $\pi(\mathbf{x}) = \dot{\phi}(\mathbf{x})$ according to (42). Therefore,

$$\dot{\kappa}(\mathbf{x}) = \phi(\mathbf{x})\ddot{\phi}(\mathbf{x}) + \dot{\phi}(\mathbf{x})\dot{\phi}(\mathbf{x}); \tag{63}$$

on the other hand, with the presumed form for H [cf., (17)], it follows that

$$\dot{\kappa}(\mathbf{x}) \equiv i[H,\kappa(\mathbf{x})] = \dot{\phi}(\mathbf{x})\dot{\phi}(\mathbf{x}) + \phi(\mathbf{x})\nabla^2\phi(\mathbf{x}) - m^2\phi(\mathbf{x})\phi(\mathbf{x}) - 4\lambda\phi(\mathbf{x})\phi^3(\mathbf{x}). \tag{64}$$

Here we have made use of the formal relation

$$[\kappa(\mathbf{x}),\pi(\mathbf{y})] = i\delta(\mathbf{x}-\mathbf{y})\pi(\mathbf{y}) \tag{65}$$

in addition to (53). On equating these two expressions for $\dot{\kappa}(\mathbf{x})$, it follows that

$$\phi(\mathbf{x})[\Box\phi(\mathbf{x}) + m^2\phi(\mathbf{x}) + 4\lambda\phi^3(\mathbf{x})] = 0, \tag{66}$$

where $\Box \equiv \partial_t^2 - \nabla^2$.

At this point, we introduce $\phi(x) \equiv \phi(\mathbf{x},t)$ and rewrite (66) as an equation valid for all time, namely

$$\phi(x)[\Box\phi(x) + m^2\phi(x) + 4\lambda\phi^3(x)] = 0; \tag{67}$$

this relation is the form taken by the equation of motion in the present context.

K. Connection to earlier work

Equation (67) is important for several reasons. First of all, this relation bears a close resemblance to the conventional equation of motion (the brackets alone), and since $\phi(x)$ almost never vanishes (67) could even be accepted, at least formally, as a substitute for the conventional equation of motion. Having said this, it must also be realized that (67) is not as simple as it looks. If, as we expect,[3] nontrivial multiplicative renormalizations are entailed in defining the local products, then one simply cannot "chip off" the factor $\phi(x)$ to recover the conventional field equation. This would be true if the first factor read $\phi(y)$, i.e., was evaluated at another, independent point, but that is just not the case. Second of all, *and this is a very important point*, (67) represents the operator form of the equation of motion implicit in the functional differential equation for the generating functional of time-ordered Green's functions for a quartic self-interacting augmented scalar field as presented in Eq. (14) of Ref. 3.[16] Equation (67) thus provides the all-important connection between the operator formalism discussed in this paper and the functional approaches treated in Refs. 2 and 3.

L. Other Poincaré group generators

Besides (45), which expresses certain matrix elements of the Hamiltonian in terms of the functional $E(w)$ and known operations, analogous expressions can be given for the other Poincaré-group generators. In particular, the space-translation generator P is determined from the relation

$$\langle w|\exp(i\mathbf{a}\cdot P)|w'\rangle = \langle w|w'_{\mathbf{a}}\rangle = E(w'_{\mathbf{a}} - w) \tag{68}$$

where

$$w'_{\mathbf{a}}(\mathbf{x},\mathbf{y}) \equiv w'(\mathbf{x}+\mathbf{a},\mathbf{y}+\mathbf{a}). \tag{69}$$

Thus,

$$\langle w|P|w'\rangle = \frac{1}{i}\frac{\partial}{\partial \mathbf{a}} E(w'_{\mathbf{a}} - w)\Big|_{\mathbf{a}\equiv 0}. \tag{70}$$

In similar fashion matrix elements for the spatial rotation generators $\vec{\mathcal{J}}$ are determined from the relation

$$\langle w|\exp(i\vec{b}\cdot\vec{\mathcal{J}})|w'\rangle = \langle w|w'_{\vec{b}}\rangle = E(w'_{\vec{b}} - w), \tag{71}$$

where

$$w'_{\vec{b}}(\mathbf{x},\mathbf{y}) \equiv w'(R\mathbf{x}, R\mathbf{y}) \tag{72}$$

and

$$R \equiv \exp(i\vec{b}\cdot\vec{L}) \tag{73}$$

with $\vec{L}$ the $\frac{1}{2}(n-1)(n-2)$ conventional (defining representation) generators of $SO(n-1)$ on R^{n-1}. Thus,

$$\langle w|\vec{\mathcal{J}}|w'\rangle = \frac{1}{i}\frac{\partial}{\partial \vec{b}} E(w'_{\vec{b}} - w)\Big|_{\vec{b}\equiv 0}. \tag{74}$$

In deriving these relations use has been of the invariance of $|0\rangle$, namely $P|0\rangle = \vec{\mathcal{J}}|0\rangle = 0$.

Both the spatial translation and rotation operators are even operators, and consequently their general matrix elements may be determined in the manner outlined above in the case of the Hamiltonian.

The case of the boost operators K is only a little harder. From the conventional formal relation (at $t=0$) that

$$K = \int \mathbf{x}H(\mathbf{x})\,d\mathbf{x}, \tag{75}$$

where $H(\mathbf{x})$ is the Hamiltonian density,[17] it follows, in analogy to (44), that we should adopt

$$i[K,\phi(\mathbf{x})\phi(\mathbf{y})] = \mathbf{x}\kappa(\mathbf{x},\mathbf{y}) + \mathbf{y}\kappa(\mathbf{y},\mathbf{x}) \tag{76}$$

as the suitably basic commutator. And, using (38) just as we did to derive (45) for the Hamiltonian, it follows

that

$$\langle w|K\,|\,w'\rangle = 2\int w(\mathbf{u},\mathbf{t})\mathbf{t}w'(\mathbf{t},\mathbf{v}) \times\langle w|\,\phi(\mathbf{u})\phi(\mathbf{v})\,|\,w'\rangle\,d\mathbf{t}\,d\mathbf{u}\,d\mathbf{v}. \tag{77}$$

In deriving this relation we have also assumed that $K|0\rangle = 0$.

As with all the other generators, K is an even operator and thus there is sufficient information to obtain general matrix elements of the boost operators.

IV. SUMMARY

The program to study augmented model field theories had its inception in the unsatisfactory handling of singular, nonrenormalizable interactions in the conventional formalism. This paper complements our previous functional approaches to these models through the use here of operator techniques. Motivation for our noncanonical analysis comes directly from the nature of the singular interactions themselves and a desire to preserve any boundary conditions they may impose, such as in (18). In consequence, our formulation for augmented models avoids the use of the canonical momentum $\pi(\mathbf{x})$, which cannot generate configuration-space translations when none are allowed. Instead we employ the affine field $\kappa(\mathbf{x},\mathbf{y})$ which induces homogeneous transformations on the configuration space thus preserving any boundary conditions that may be present. Of course, the affine fields $\kappa(\mathbf{x},\mathbf{y})$ and $\phi(\mathbf{x})$, which obey the ACR given in (22) and (23), are not necessarily inconsistent with the existence of a self-adjoint local field $\pi(\mathbf{x})$, particularly for less singular models where no special boundary conditions exist. But the important point is that the affine fields do not require the canonical momentum $\pi(\mathbf{x})$. This seems especially significant when the general affine fields $\kappa(\mathbf{x},\mathbf{y})$ and $\phi(\mathbf{x})$ admit a bona fide representation of the special affine fields $\kappa(\mathbf{x})$ $[\equiv \kappa(\mathbf{x},\mathbf{x})]$ and $\phi(\mathbf{x})$, which satisfy (53), for then the existence of a canonical momentum $\pi(\mathbf{x})$ in such cases is seriously cast in doubt. And it is just such affine field representations that are fundamental, for then local field scaling $\phi(\mathbf{x}) \to S(\mathbf{x})\phi(\mathbf{x})$, $S(\mathbf{x}) > 0$, which is seemingly so basic to augmented models, is in fact unitarily implementable.

The difficulties with $\pi(\mathbf{x})$ pose potential problems for conventional canonical dynamical formulations as well. The meaning of the conventional relation (42) becomes dubious, and as an alternative basic dynamical statement we take (44) which does not compromise our concern regarding boundary conditions. An analogously save formula (76) applies for the boost generators of the Poincaré group.

Once the proper kinematical variables and basic dynamical statements have been identified, there are several ways to proceed. None of the foregoing was particularly special to a quartic interaction, and for the most part of our discussion in Sec. III we chose to characterize the model by means of the expectation functional $E(w)$ for the field product $\phi(\mathbf{x})\phi(\mathbf{y})$, as specified in (34). With the aid of the basic relations identified earlier, a total set of matrix elements in the even field subspace for the operators $\kappa(\mathbf{x},\mathbf{y})$, H, and the other Poincaré-group generators could be expressed in terms of $E(w)$. The matrix elements for $\kappa(\mathbf{x})$ follow from (38) for $\mathbf{y} \equiv \mathbf{x}$, but such a relation may only define a form unless the general affine field representation admits a special affine field representation. It was pointed out that for even Hamiltonians, such as a quartic self-coupling, general matrix elements of all the basic operators are determined, at least in principle. In essence, then, the specific dynamics is fully contained in the expectation functional $E(w)$, which plays the role in some sense of "model selector." The analysis in this form bears a close relationship to that of Araki,[7] which was based on the canonical formalism.

Another approach to the dynamical equations is as follows. We pointed out that for a quartic self-coupled field the equation of motion for the special affine field $\kappa(\mathbf{x})$ led to (67). The meaning of this equation can be seen in another way. Let us adopt the not unfamiliar expression

$$I \equiv \int\{\tfrac{1}{2}[(\partial_\mu\phi)^2 - m^2\phi^2] - \lambda\phi^4\}\,dx \tag{78}$$

as the quantum action, and derive the equation of motion by a stationary principle under field variations of the form

$$\delta\phi(x) = \delta S(x)\cdot\phi(x), \tag{79}$$

where $\delta S(x)$ is arbitrary. Such a class of variations, suggested by local scaling, respects the homogeneity of field transformations and any boundary conditions that may exist. The result of such a variational principle is just

$$\phi(x)\Box\phi(x) + m^2\phi^2(x) + 4\lambda\phi^4(x) = 0, \tag{80}$$

which is recognized as (67). Moreover, this is just the operator equation of motion implicit in the functional studies of Ref. 3, and it is our present view that a study of this approach to dynamics is best carried out by a study of the Green's functional or coupled Green's function equations as spelled out in Ref. 3.

[1] J.R. Klauder, in *Recent Developments in Mathematical Physics*, proceedings of the XII Schladming conference on nuclear physics, edited by P. Urban (Springer, Berlin, 1973) [Acta Phys. Austriaca Suppl. **11** (1973)], p. 341; Phys. Lett. B **47**, 523 (1973); in *Lecture Notes in Physics*, edited by H. Araki (Springer, New York, 1975), Vol. 39, p. 160.
[2] J.R. Klauder, Phys. Rev. D **14**, 1952 (1976).
[3] J.R. Klauder, Phys. Rev. D **15**, 2830 (1977).
[4] I.M. Gel'fand and M.A. Naimark, Dokl. Akad. Nauk SSSR **55**, 570 (1947).
[5] E.W. Aslaksen and J.R. Klauder, J. Math. Phys. **9**, 206 (1968); **10**, 2267 (1969); J.R. Klauder and E.W. Aslaksen, Phys. Rev. D **2**, 272 (1970).
[6] J.R. Klauder, J. Math. Phys. **11**, 609 (1970), Sec. 4; in *Proceedings of the Relativity Conference in the Midwest*, edited by M. Carmeli *et al.* (Plenum, New York, 1970), p. 1.
[7] H. Araki, J. Math. Phys. **1**, 492 (1960).
[8] J. Glimm, A. Jaffe, and T. Spencer, Ann. Math. **100**, 585 (1974); B. Simon and R. Griffiths, Commun. Math. Phys. **33**, 145 (1973); J.S. Feldman and K. Osterwalder, Ann. Phys. (N.Y.) **97**, 80 (1976); J. Magnen and R. Seneor, Ann. Inst. Henri Poincaré **24**, 95 (1976).
[9] I.M. Gel'fand and N.Ya. Vilenkin, *Generalized Functions*, *Vol.* 4: *Applications of Harmonic Analysis*, translated by

A. Feinstein (Academic, New York, 1964); G.C. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Interscience, New York, 1972); H. Araki, Commun. Math. Phys. **20**, 9 (1971); G.C. Hegerfeldt, J. Math. Phys. **13**, 45 (1972).

[10]H. Ezawa, J.R. Klauder, and L.A. Shepp, J. Math. Phys. **16**, 783 (1975).

[11]In this regard compare also J.R. Klauder and E.W. Aslaksen, Phys. Rev. D **2**, 272 (1970).

[12]See, e.g., S.S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1962).

[13]This general behavior is entirely analogous to the treatment of hard-core interactions in nuclear physics.

[14]This argument is no different in principle than one for an even quantum mechanical problem in a finite number of dimensions.

[15]F.A. Berezin, *The Method of Second Quantization*, translated by N. Mugibayashi and A. Jeffrey (Academic, New York, 1966); A.V. Skorohod, *Integration in Hilbert Space*, translated by K. Wickwire (Springer, New York, 1974); W.J. Eachus and L. Streit, Rep. Math. Phys. **4**, 161 (1973); see also L. Rosen, J. Math. Phys. **13**, 918 (1972), which contains references to earlier works.

[16]In terms of equations in this paper, the alternative equation of motion (67) can be said to arise from the *scale* invariance of the formal measure in (3) as opposed to the *translation* invariance of the formal measure in (6). For further details consult Ref. 3.

[17]See, e.g., J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Reading, Massachusetts, 1955).

On generating forms of *K*-generalized Lagrangian and Hamiltonian systems

Paulo R. Rodrigues*

Département de Mathématiques, Université de Dijon, 214, rue de Mirande, 21000-Dijon, France
(Received 20 October 1976)

From a $(n+1)$-form Ω on the manifold J^kM of k-jets of local sections of the vector bundle (M,π,N) we study the conditions to obtain the Lagrangian and Hamiltonian formalisms for a theory which involves higher order derivatives. The results generalize those of Gallissot and others for $k=1$.

1. INTRODUCTION

The utilization of differential forms in mechanics, see for example the books of Abraham,[1] Godbillon,[2] Souriau,[3] and Hermann,[4] has allowed not only a more rigorous formulation from the mathematical point of view, but also a better understanding of its physical content. Further, the use of the jets theory introduced by Ehresmann[5] seems to be an adequated instrument for the study of theories which involve higher order derivatives, see, for example, Goldschmidt[6] and Sniatycki.[7]

Recently, Goldschmidt and Sternberg,[6] working on intrinsic geometrical properties of variational calculus (as for as we know, first studied by Dedecker[8] in 1953), have obtained a sophisticated generalization of this theory and in particular have generalized Lagrangian and Hamiltonian equations of motion (these equations had already been established by Gallissot[9] for a generalized continuous systems). Others authors,[7,10,11] and particularly Krupka[12] had already worked on the subject, with different ideas. These generalizations will be called *generalizations of order* 1, because they have developped this theory on the manifold of 1-jets of local sections of a vector bundle or more generally of a fibered manifold.

The aim of this paper, following Mimura,[10] Goldschmidt and Sternberg, and others, is to study the conditions to obtain the equations of motions of the k-generalized Lagrangian and Hamiltonian formalism. For this, we consider a $(n+1)$ form on the manifold J^kM of k-jets of local sections of the vector bundle (M,π,N) (M,N are C^∞-differentiable manifolds) from which it is possible to establish the formalisms for a theory which involves higher order derivatives. Concerning the Hamiltonian formalism, we remark that the equations obtained are those of De Donder,[13] Eq. (622). If the Lagrangian and Hamiltonian functions are canonically defined as integrals of the respective densities, we can obtain equations formally similar to those obtained in classical mechanics, where the usual derivatives will be replaced by generalized formal functional derivatives [see Definition 5 and Eqs. (8) and (9) in Coelho de Souza and Rodrigues[14]].

2. NOTATIONS

Throughout this paper it is assumed that all differential structures are of class C^∞. Let N, M be differentiable manifolds of dimensions n, m, $(m \geq n)$, respectively, and (M,π,N) a fibered manifold.[15] We denote by (J^kM,ψ^k,N), for sake of brevity J^kM, the fibered manifold of k-jets of local sections of M, $\tilde{s}^k_x$ the k-jet of a section s of M in $x\in N$, $\tilde{s}^k: N\to J^kM$ the k-jet extension of s defined by $\tilde{s}^k(x)=\tilde{s}^k_x$ and $\pi^k_t: J^kM\to J^tM$, $(t\leq k)$, $\pi^k: J^kM\to M$, the projections defined by

$$\psi^k(\tilde{s}^k_x)=x,\quad \pi^k_t(\tilde{s}^k_x)=\tilde{s}^t_x,\quad \pi^k(\tilde{s}^k_x)=s(x).$$

Let $I(j)=(i_1,\ldots,i_j)$ denote an ordered j-tuple of integers $1,\ldots,n$. If $r\in J^kM$, then we put locally $r=(x_i,y^\alpha,y^\alpha_{I(j)})$, where $x_i=(x_1,\ldots,x_n)$, $y^\alpha=(y^1,\ldots,y^m)$ are local coordinates for M on a local chart such that $(x_1,\ldots,x_n)$ is the local coordinate on N induced by the projection π, and

$$y^\alpha_{I(j)}(\tilde{s}^k(x))=\frac{\partial^j s^\alpha}{\partial x_{i_1}\cdots\partial x_{i_j}}(x)=s_{I(j)}(x),$$

$$s^\alpha=y^\alpha\circ s,\quad 1\leq j\leq k,\quad 1\leq\cdots\leq i_j\leq\cdots\leq n,$$

for any local section s of M. Let f be a function defined on an open set in J^kM such that its ψ^k projection is in the domain of the local coordinates (x_i). The *formal derivative of* f by x_j, d_jf is by definition

$$d_jf=D_{1j}f+D_{2\alpha}f\cdot y^\alpha_j+\cdots+\sum_{i_1\leq\cdots\leq i_k}D_{k+2,I(k)\alpha}f\cdot y^\alpha_{I(k)},$$

where, for the coordinate system $(x_i,y^\alpha,y^\alpha_{I(j)})$ on J^kM, the symbols $D_{1j},\ldots,D_{k+2,I(k)\alpha}$ are used for the partial derivative operators with respect to the variables $x_i,y^\alpha,\ldots,y^\alpha_{i_1\ldots i_k}$. Now, consider the 1-form

$$d_jf\cdot dx_j=D_{1j}f\,dx_j+\cdots+\sum D_{k+2,I(k)\alpha}f\cdot y^\alpha_{I(k)}dx_j, \tag{1}$$

and, for brevity take $k=1$. Then, developing (1), we have

$$d_jf\cdot dx_j=\frac{\partial f}{\partial x_j}dx_j+\frac{\partial f}{\partial y^\alpha}y^\alpha_j\,dx_j+\sum_i\frac{\partial f}{\partial y^\alpha_i}y^\alpha_{ij}dx_j,$$

where $y^\alpha_i=\partial y^\alpha/\partial x_j$, $y^\alpha_{ij}=\partial y^\alpha_i/\partial x_j$. Taking the sum over the j's and α's, we get

$$\sum_j d_jf\cdot dx_j=df, \tag{2}$$

where df denotes the differential of f.

Let V, W be two manifolds. By $C^\infty(V,W)$ we denote the set of all C^∞ mapping from V to W. If $f\in C^\infty(V,W)$, then by $Tf: TV\to TW$ we denote the tangent mapping (where TV, TW are the tangent bundles of V and W, respectively).

Let (M, π, N) be a vector bundle. By $\Gamma(N)$ we represent the $C^\infty(N)$-module of local sections on M [here $C^\infty(N)$ is the R-algebra of C^∞-functions on N]. Let $C^\infty_{\psi 1}(J^1M, N)$ be the $C^\infty(J^1M)$-module of C^∞-mappings $f: J^1M \to M$ such that $\pi \circ f = \psi^1$. If $(s^1, \ldots, s^m)$ is a basis of $\Gamma(N)_V$, where V is a open subset on a local chart (U, x_i). Then there is a induced basis on $C^\infty_{\psi 1}(J^1M, N)_{(\psi^1)^{-1}(V)}$ defined by $F^\alpha = s^\alpha \cdot \psi^1$.

Now, we may consider (locally) a basis of $\Gamma(J^1M)_{(\psi^1)^{-1}(V)}$ as being formed by the m-local sections (s^α) of M and its 1-jets. If $(x_i, y^\alpha, y^\alpha_i)$ are local coordinates on J^1M, we consider $F^\alpha = s^\alpha \cdot \psi^1$, $F^\alpha_i = s^\alpha_i \cdot \psi^2$ as being the elements of the induced basis on $C^\infty_{\psi 2}(J^2M, J^1M)_{(\psi^2)^{-1}(V)}$, and so on. For the general situation, we will say that F^α, $F^\alpha_{I(1)}, \ldots, F^\alpha_{I(k-1)}$ is the basis for $C^\infty_{\psi k}(J^kM, J^{k-1}M)$ induced by the local coordinates $(x_i, y^\alpha, y^\alpha_{I(1)}, \ldots, y^\alpha_{I(k)})$ on J^kM. Let[16] $F = f^{I(j)}_\alpha (F^\alpha_{I(j)})^*$, [resp. $\theta = \sum_{I(j)} \theta^\alpha_{I(j)} F^\alpha_{I(j)}$], where $f^{I(j)}_\alpha \in C^\infty(J^kM)$, $(F^\alpha_{I(j)})^*$ is the dual basis of $(F^\alpha_{I(j)})$, [resp. $(\theta^\alpha_{I(j)})$ are 1-forms on J^kM], with $0 \leq j \leq k-1$ and $F^\alpha_{I(0)} = F^\alpha$ [resp. $\theta^\alpha_{I(0)} = \theta^\alpha$]. Then we have $F\theta = f^{I(j)}_\alpha \theta^\alpha_{I(j)}$. We remark that sometimes we put ∂x for the symbol $\partial/\partial x$.

3. LAGRANGIAN EQUATIONS

Let (M, π, N) be a vector bundle and z a point of J^kM (here and in the following we suppose that N is oriented). If s is a section of M over a neighborhood of $x = \psi^k(z)$, consider the mapping

$$T\tilde{s}^{k-1} : T_xN \to T_r(J^{k-1}M),$$

where $r = \pi^k_{k-1}(z)$, $\tilde{s}^{k-1}(x) = z$. Consider also the following difference:

$$T\pi^k_{k-1} - T\tilde{s}^{k-1} \circ T\psi^k.$$

If $X \in T_z(J^kM)$, then[7]

$$(T\pi^k_{k-1} - T\tilde{s}^{k-1} \circ T\psi^k)_r(X) \in \text{Ver } T_r(J^{k-1}M), \tag{3}$$

where Ver $T_r(J^{k-1}M)$ denotes the vertical subspace of $T_r(J^{k-1}M)$. Consider the mapping

$$\theta_{[k]} : z \in J^kM \to (T\pi^k_{k-1} - T\tilde{s}^{k-1} \circ T\psi^k)_r(X) \in \text{Ver } T_r(J^{k-1}M).$$

Then we have a unique 1-form, which is also represented by $\theta_{[k]}$, on J^kM. This $T(J^{k-1}M)$-valued 1-form on J^kM is such that $\langle X, \theta_{[k]} \rangle$ is equal to the sum of the components of (3), where X is a vector field on J^kM (for a proof of the unicity, see Ref. 6). We call $\theta_{[k]}$ the *fundamental form* on J^kM. Let $(x_i, y^\alpha, y^\alpha_{I(1)}, \ldots, y^\alpha_{I(k)})$ be a local coordinate system on J^kM and

$$G_{[k+1]} = f_\alpha (F^\alpha)^* + \cdots + f^{I(k)}_\alpha (F^\alpha_{I(k)})^* \in C^\infty_{\psi k+1}(J^{k+1}M, J^kM)^*.$$

Consider the mapping

$$G_{[k+1]} \in C^\infty_{\psi k+1}(J^{k+1}M, J^kM)^* \to F_{[k]} \in C^\infty_{\psi k}(J^kM, J^{k-1}M)^*,$$

where

$$F_{[k]} \overset{\text{def}}{=} \sum_{I(1)} f^{I(1)}_\alpha (F^\alpha)^* + \cdots + \sum_{I(k)} f^{I(k)}_\alpha (F^\alpha_{I(k-1)})^*,$$

then

$$F_{[k]}\theta_{[k]} = \sum_{I(1)} f^{I(1)}_\alpha \theta^\alpha + \cdots + \sum_{I(k)} f^{I(k)}_\alpha \theta^\alpha_{I(k-1)},$$

where $\theta_{[k]}$ is the fundamental form on J^k. Let ω be a volume form on N and $\omega_j = i(\partial x_j)\omega$, where $i(\ )\omega$ denotes the interior product. The $(n+1)$-form on J^kM

$$\Omega_{[k]} = d\sigma_{[k]} + \rho \wedge \omega,$$

where

$$\begin{aligned}\sigma_{[k]} &= f^{i_1}_\alpha \theta^\alpha \wedge \omega_{i_1} + f^{i_1 i_2}_\alpha \theta^\alpha_{i_1} \wedge \omega_{i_2} + \cdots \\ &\quad + f^{i_1 \cdots i_k}_\alpha \theta^\alpha_{i_1 \cdots i_{k-1}} \wedge \omega_{i_k} \\ &= f^{I(j)}_\alpha \theta^\alpha_{I(j-1)} \wedge \omega_{i_j}\end{aligned} \tag{4}$$

and ρ is a 1-form on J^kM, is called *generating form*.

Proposition 1: Let $u: N \to J^2M$ be a section of J^2M such that $u = \tilde{s}^2$, for some section s of (M, π, N). For all vector fields X on J^2M, $(u)^*(i(X)\Omega_{[2]}) = 0$ iff

$$\sum_j d_j(u^* f^{I(2)}_\alpha) + u^* f^{I(1)}_\alpha - u^* U^{I(1)}_\alpha = 0, \tag{5a}$$

$$\sum_j d_j(u^* f^{I(1)}_\alpha) - u^* U_\alpha = 0, \tag{5b}$$

$$u^* f^{I(2)}_\alpha - u^* U^{I(2)}_\alpha = 0, \tag{5c}$$

where u^* is the pullback mapping and $U^{I(j)}_\alpha$ are the component functions of ρ.

Proof: Let $(x_i, y^\alpha, y^\alpha_{I(1)}, y^\alpha_{I(2)})$ be the local coordinates on J^2M

$$X = X_i \partial x_i + X_\alpha \partial y^\alpha + X^{I(1)}_\alpha \partial y^\alpha_{I(1)} + X^{I(2)} \partial y^\alpha_{I(2)},$$

a vector field on J^2M (to simplify the calculus, we will take $X_i = X_\alpha = \cdots = X^{I(2)}_\alpha = 1$ throughout the proof),

$$\omega = dx^1 \wedge \cdots \wedge dx^n, \quad [\text{resp. } i(\partial x_i)\omega = \omega_i],$$

a volume form on N [resp. the contraction of ω by ∂x_i] and

$$\rho = U_i\, dx^i + U_\alpha\, dy^\alpha + U^{I(1)}_\alpha\, dy^\alpha_I + U^{I(2)}_\alpha\, dy^\alpha_{I(2)},$$

an arbitrary 1-form on J^2M.

Then (3) is

$$(T\pi^2_1 - T\tilde{s}^1 \circ T\psi^2)(X) = (1 - y^\alpha_{I(1)})\, \partial y^\alpha + (1 - y^\alpha_{I(2)}) \partial y^\alpha_{I(1)},$$

which gives rise to the 1-forms

$$\theta^\alpha_{I(j-)} = dy^\alpha_{I(j-1)} - y^\alpha_{I(j)}\, dx_i, \quad 1 \leq j \leq 2$$

$$\theta_{[2]} = \sum_{\alpha, j} \theta^\alpha_{I(j-1)}$$

so, we have

$$\begin{aligned}d\sigma_{[2]} &= df^{I(j)}_\alpha \wedge \theta^\alpha_{I(j-1)} \wedge \omega_{i_j} + f^{I(j)}_\alpha d\theta^\alpha_{I(j-1)} \wedge \omega_{i_j} \\ &= df^{I(j)} \wedge dy^\alpha_{I(j-1)} \wedge \omega_{i_j} - f^{I(j)}_\alpha dy^\alpha_{I(j)} \wedge \omega,\end{aligned}$$

since, by (2)

$$y^\alpha_{I(j)}\, df^{I(j)}_\alpha \wedge \omega = y^\alpha_{I(j)} \Big(\sum_k d_k f^{I(j)}_\alpha\, dx_k\Big) \wedge \omega = 0.$$

After some calculations, we have

$$u^*i(X)\Omega_{[2]} = u^*(d\sigma_{[2]} + \rho \wedge \omega)(X)$$

$$= \sum_j \left\{ [d_j(u^* f_\alpha^{I(2)}) + u^* f_\alpha^{I(1)} - u^* U_\alpha] \left(\frac{\partial y_{I(1)}^\alpha}{\partial x_i} - 1 \right) \right.$$

$$+ [d_j(u^* f_\alpha^{I(1)}) - u^* U_\alpha] \left(\frac{\partial y^\alpha}{\partial x_i} - 1 \right)$$

$$\left. + [u^* f_\alpha^{I(2)} - u^* U_\alpha^{I(2)}] \left(\frac{\partial y_{I(2)}^\alpha}{\partial x_i} - 1 \right) \right\} \omega$$

which gives the desired result.

From Eq. (5b) we have the Eqs. (5a) and (5c) in the following forms:

$$\sum_{i_1, i_2} d_{i_1} d_{i_2} (u^* f_\alpha^{I(2)}) - \sum_{i_1} d_{i_1} (u^* U_\alpha^{I(1)}) + u^* U_\alpha = 0$$

$$u^* f_\alpha^{I(2)} - u^* U_\alpha^{I(2)} = 0. \tag{5'}$$

Suppose now that $\rho = d\mathcal{L}$, for $\mathcal{L} \in C^\infty(J^2 M)$ (which is true locally); then the equations (5′) takes the classical form

$$\sum_{i_1, i_1} \frac{d}{dx_{i_1}} \frac{d}{dx_{i_1}} \left(\frac{\partial \mathcal{L}}{\partial y_{i_1 i_2}^\alpha} \right) - \sum_{i_1} \frac{d}{dx_{i_1}} \left(\frac{\partial \mathcal{L}}{\partial y_{i_1}} \right) + \frac{\partial \mathcal{L}}{\partial y^\alpha} = 0, \tag{5''}$$

or, since Proposition 1 is true for $k > 2$, for the general situation we have

$$\sum_{i_1, \ldots, i_k} (-1)^k \frac{d^k}{dx_{i_1} \cdots dx_{i_k}} \left(\frac{\partial \mathcal{L}}{\partial y_{i_1 \ldots i_k}^\alpha} \right) + \cdots$$

$$+ (-1) \sum_{i_1} \frac{d}{dx_{i_1}} \left(\frac{\partial \mathcal{L}}{\partial y_{i_1}^\alpha} \right) + \frac{\partial \mathcal{L}}{\partial y^\alpha} = 0,$$

which are the *Euler—Lagrange equations* for a field theory with Lagrangian density $\mathcal{L}$ depending on field derivatives of higher order.[14]

4. THE VARIATIONAL FORMALISM

Let N be an oriented manifold, $A \subset N$ a compact, s a section of M over a neighborhood of A, and $u = \tilde{s}^k$. Consider the n-form $\sigma_{\mathcal{L}}$ over J^k defined by

$$\sigma_{\mathcal{L}} = \sigma_{[k]} + \mathcal{L}\omega,$$

where $\mathcal{L} \in C^\infty(J^k M)$. If $(s_{(t)})$ an arbitrary smooth one-parameter family of sections on $J^k M$ such that $s_0 = s$ and $s_{(t)} = s$ on ∂A, for all t; then s is a extremal of I_A if

$$\frac{d}{dt} I_A(s) \Big|_{t=0} = 0, \tag{6}$$

where

$$I_A(s) = \int_A u^* \sigma_{\mathcal{L}}. \tag{7}$$

Now,[17]

$$\frac{d}{dt} u_t^* \sigma_{\mathcal{L}} \Big|_{t=0} = u^*(i(X)\, d\sigma_{\mathcal{L}}) + du^*(i(X)\sigma_{\mathcal{L}}),$$

where X is an arbitrary[18] vector field on N with supp $X \subset A^0$. From (6), (7), and Stoke's theorem,

$$\int_A u^*(i(X)\, d\sigma_{\mathcal{L}}) = 0.$$

Since this last equality holds for all X with supp $X \subset A^0$, we conclude that on A^0

$$u^*(i(X)\, d\sigma_{\mathcal{L}}) = u^*(i(X)\Omega_{[k]}) = 0,$$

which gives the following.

Proposition 2: With the notations above, u is an extremal of $I_A(s)$ if and only if $u^*(i(X)\Omega_{[k]}) = 0$, for all vector fields X on N with supp $X \subset A^0$.

5. THE HAMILTONIAN EQUATIONS

Let (Y, π, N) be a fibered manifold over N, where $\dim Y = (n + m + nm)$, $\dim N = n$. Let $\mathcal{L} \in C^\infty(J^1 Y)$ and consider the mapping

$$\Lambda_{\mathcal{L}} : J^1 Y \to TN \otimes_Y V^*(Y)$$

defined by $\Lambda_{\mathcal{L}}(p) = d\mathcal{L}_p$ (d means the fiber derivative). $\Lambda_{\mathcal{L}}$ is called the *Legendre transformation*, and is said to be an *Hamilton—Lagrange duality* (H—LD) if it is a diffeomorphism. In the following $\Lambda_{\mathcal{L}}$ is considered H-LD locally.

Let $\theta_{[k]}$ be the fundamental form on $J^k M$ and consider the following linear combination, (cf. Sec. 3)

$$P_{[k]} \overset{\text{def}}{=} \sum_{I(1)} p_\alpha^{I(1)} (F^\alpha)^* + \cdots + \sum_{I(k)} p_\alpha^{I(k)} (F_{i(k-1)}^\alpha)^*.$$

Then, locally, the generating form may be put in the form

$$\Omega_{[k]} = \sum_j d[(P_{[k]}\theta_{[k]}) \wedge \omega_{i_j}] + \rho \wedge \omega$$

$$= dp_\alpha^{I(j)} \wedge dy_{I(j-1)}^\alpha \wedge \omega_{I_j} - \Theta \wedge \omega,$$

where

$$\Theta = d(p_\alpha^{I(j)} y_{I(j)}^\alpha) - \rho.$$

Put

$$\rho = d\mathcal{L},$$

$$H = p_\alpha^{I(j)} y_{I(j)}^\alpha - \mathcal{L}, \tag{8}$$

$$\sigma_H^{[k]} = \lambda - H\omega,$$

where

$$\lambda = f_\alpha^{I(j)}\, dy_{I(j-1)}^\alpha \wedge \omega_{i_j}.$$

Then

$$\Theta = dH, \quad \Omega_{[k]} = d\sigma_H^k.$$

Now, consider $k = 2$ and define

$$p_\alpha^{I(1)} = \frac{d\mathcal{L}}{dy_{I(1)}^\alpha} - \sum_{i_2} \frac{d}{dx_{i_2}} \frac{\partial \mathcal{L}}{\partial y_{I(2)}^\alpha},$$

$$p_\alpha^{I(2)} = \frac{d\mathcal{L}}{dy_{I(2)}^\alpha}, \tag{9}$$

and let $(x^i, y^\alpha, z_i^\alpha)$ be a local coordinate system on an open subset U of Y, $(x^i, y^\alpha, z_{I(1)}^\alpha, z_{I(2)}^\alpha)$ the induced coordinates on $(\pi^1)^{-1}(U)$. If we take $Y = J^1 M$, $\Lambda_{\mathcal{L}}$ H—LD locally, then we may consider $(x^i, y^\alpha, y_{I(1)}^\alpha, p_\alpha^{I(1)}, p_\alpha^{I(2)})$ as coordinates on an open subset of $(\pi^1)^{-1}(U)$. It is clear

that if X is such that $u^*(i(X)\,d\sigma_H^{[2]}) = 0$, then, by (9) and (5″), the integral curves of X are solutions of the *generalized Hamilton's equations*

$$\frac{\partial H}{\partial y^\alpha} = -\frac{\partial L}{\partial y^\alpha} = -\sum_{i_1} \frac{d}{dx_{i_1}}(p_\alpha^{I(1)}),$$

$$\frac{\partial H}{\partial p_\alpha^{I(1)}} = y_{I(1)}^\alpha = \frac{d}{dx_{i_1}}(y^\alpha),$$

$$\frac{\partial H}{\partial y_{I(1)}^\alpha} = p_\alpha^{i(1)} - \frac{dL}{dy_{I(1)}^\alpha} = -\sum_{i_2} \frac{d}{dx_{i_2}}(p_\alpha^{I(2)}),$$

$$\frac{\partial H}{\partial p_\alpha^{I(2)}} = y_{I(2)}^\alpha = \frac{d}{dx_{i_2}}(y_{I(1)}^\alpha).$$

To generalize this situation (i.e., $k > 2$), we remark that there is a mapping which permits the identification of $J^k M$ with a manifold regularly immersed in $J^1(J^{k-1}M)$, and so we may consider Λ_L defined on $J^k M$. Then we will get the equations

$$\frac{\partial H}{\partial y_{I(j-1)}^\alpha} = -\sum_{i_1} \frac{d}{dx_{i_1}}(p_\alpha^{I(j)}),$$

$$\frac{\partial H}{\partial p_\alpha^{I(j)}} = \frac{d}{dx_{i_j}}(y_{I(j-1)}^\alpha), \tag{10}$$

for $1 \leq j \leq k$, $k > 2$.

6. THE GENERALIZED CLASSICAL MECHANICS

Following the results in symplectic mechanics systematized in the literature,[1-3] it is found that the Lagrangian, respectively Hamiltonian, formalism can be characterized by geometric structures canonically associated to the tangent (velocities space), respectively cotangent (phase space), bundle of a differentiable manifold (configuration space). Concerning the Lagrangian point of view, Klein[19] showed that special structures are necessary like the *almost symplectic forms* and a special exterior differential calculus characterized by an endomorphism over the double tangent bundle, called *almost tangent structure* (in the Hamiltonian formalism, the symplectic structure over the cotangent bundle and the usual exterior calculus are sufficient). Recently[20] the author has studied some aspects of the theory for a generalized velocities space and showed that in this case the structures are not canonically induced by the usual ones. It was proved that for this type of mechanics, which we call *generalized mechanics*, a new structure, called almost horizontal symplectic subordinated to some hypothesis, is necessary for the development of the theory.

To develop a generalized classical mechanics from the present geometrical formalism, it is sufficient to consider the trivial bundle $(R\times M, \pi, R)$. Then we may identify $J^k(R\times M)$ with the manifold of k-jets of mappings of R into M. Let $\tau_k M$ be the k-jets of nonconstant mappings with source $O \in R$. A natural k-jet chart on $(\tau_k M, \psi^k, R)$ is then $(t, q_\alpha, \dot{q}_\alpha, \ldots, q_\alpha^{(k)})$, where $q_\alpha^{(j)} = (d^j/dt^j)(q_\alpha)$. It is clear now that it is possible to describe a system of mechanics including higher time derivatives of the coordinates $q_\alpha(t)$. So, the introduction of higher derivatives into Lagrangian and Hamiltonian formalism seems to be not only an increase of a great number of coordinates.

*On leave of absence from Universidade Federal Fluminense, Rio do Janeiro, Brazil.

[1]R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
[2]C. Godbillon, *Geometrie Differentielle et Mécanique Analytique* (Hermann, Paris, 1968).
[3]J.M. Souriau, *Structure des systèmes dynamiques* (Dunod, Paris, 1970).
[4]P. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968).
[5]Ch. Ehresmann, C.R. Acad. Sci. Paris 233, 598—600 (1951); 233, 777—779 (1951); IV Congresso Unione Mat. Italiana, Taormina (1951).
[6]H. Goldschmidt and S. Sternberg, Ann. Inst. Fourier (Grenoble) 23 (1); 203—267 (1973).
[7]J. Sniatycki, Proc. Cambridge Phil. Soc. 68, 475—484 (1970).
[8]P. Dedecker, "Calcul de variations, formes différentielles et champs géodésiques," Coll. Int. de géom. Diff. Strasbourg (1953).
[9]G. Gallissot, Ann. Inst. Fourier (Grenoble) 8, 291—335 (1958).
[10]F. Mimura, Bull. Kyushu Inst. Techn. 21, 1.20 (1974).
[11]P.L. Garcia, thesis (Barcelona) (1968), pp. 1—66.
[12]D. Krupka, J. Math. Anal. Appl. 49, 180—206 (1975); 49, 469—475 (1975).
[13]Th. De Donder, *Théorie Invariantive du Calcul des Variations* (Nouvelles Edit. Gauthier-Villars, Paris, 1935).
[14]L.M.C. Coelho de Souza and P.R. Rodrigues, J. Phys. A 2, 304—310 (1969).
[15]The definitions may be found in Ref. 6, for example.
[16]Here and inthe following, we consider the usual convention on summation.
[17]See Ref. 6 for a proof.
[18]In general we take X as a vertical vector field, but this restriction is superfluous, see Ref. 6, p. 221.
[19]J. Klein, Ann. Inst. Fourier 12, 1—124 (1962).
[20]P.R. Rodrigues, C.R. Acad. Sci. Paris, 281, 643—646 (1975); 282, 1307—1309 (1976).

On dual series equations involving Konhauser biorthogonal polynomials

K. R. Patil

Department of Mathematics, Science College, Satara (Maharashtra), India

N. K. Thakare

Department of Mathematics, Shivaji University, Kolhapur (Maharashtra), India—416 004
(Received 9 August 1976)

By using Abel's integral equations, we solve dual series equations involving Konhauser's biorthogonal polynomial set of the first kind.

1. INTRODUCTION

Spencer and Fano[1] introduced a certain pair of biorthogonal polynomial sets in carrying out calculations involving the penetration of gamma rays through matter. They did not establish any general properties of biorthogonal polynomial sets but essentially utilized biorthogonality of polynomials in x and polynomials in x^2 with respect to the weight function $x^\alpha \exp(-x)$, $(\alpha > 0)$ over the interval $(0, \infty)$. A general foundation for the theory of biorthogonal polynomials was established by Konhauser.[2] As an application of his general theory he introduced two polynomial sets (related to Laguerre polynomials) which are biorthogonal with respect to $x^\alpha \exp(-x)$ over the interval $(0, \infty)$. Incidently his polynomials give as particular cases the biorthogonal sets of Spencer and Fano[1] and also the biorthogonal sets considered by Preiser.[3]

The technique of dual series equations is frequently used to solve mixed boundary value problems of potential theory. For instance, suppose we wish to find the axisymmetric solution $V(\rho, z)$ of Laplace's equation in the semi-infinite cylinder $0 \leq \rho \leq a$, $z \geq 0$ satisfying the boundary conditions:

$$V(\rho, z) \to 0 \quad \text{as } z \to \infty,$$

$$V(a, z) = 0, \quad z \geq 0,$$

$$V(\rho, 0) = f(\rho), \quad 0 \leq \rho < 1,$$

and

$$\{\partial V/\partial z\}_{z=0} = 0, \quad 1 < \rho \leq a.$$

Konhauser[4] studied the biorthogonal polynomial sets $\{Z_n^\alpha(x; k)\}$ and $\{Y_n^\alpha(x; k)\}$ which are biorthogonal over the interval $(0, \infty)$ with respect to the weight function $x^\alpha \exp(-x)$, $(\alpha > -1)$. The polynomial $Z_n^\alpha(x; k)$ is of degree n in x^k and the polynomial $Y_n^\alpha(x; k)$ is of degree n in x for $n = 0, 1, 2, \cdots$. In fact Konhauser stated that

$$Z_n^\alpha(x; k) = \frac{\Gamma(kn + \alpha + 1)}{n!} \times \sum_{j=0}^{n} (-1)^n \binom{n}{j} \frac{x^{kj}}{\Gamma(kj + \alpha + 1)}, \tag{1.1}$$

and

$$Y_n^\alpha(x; k) = \frac{k}{n!} \frac{\partial^n}{\partial t^n} \times \left\{ \frac{\exp(-xt)(t+1)^{\alpha+kn}}{(t^{k-1} + kt^{k-2} + \cdots + k)^{n+1}} \right\} \Big/_{t=0}. \tag{1.2}$$

For $k = 1$, both these polynomials reduce to the generalized Laguerre polynomial $L_n^\alpha(x)$. We shall hence afterwards call $\{Z_n^\alpha(x; k)\}$ the Konhauser biorthogonal set of the first kind and $\{Y_n^\alpha(x; k)\}$ the Konhauser biorthogonal set of the second kind.

In this paper we solve dual series equations involving the biorthogonal polynomial set $\{Z_n^\alpha(x; k)\}$. The method which we use is a simple and direct method in which Abel's integral equations are involved. The solution of the problem follows by using biorthogonal polynomial sets of first and second kind and the known solutions of Abel's integral equations. Our approach is formal. We have also obtained a new property of $Z_n^\alpha(x; k)$ which was required in the course of our investigation.

Karande and Thakare[5] have solved similar dual series equations involving the Konhauser biorthogonal polynomial set of the first kind by using the multiplying factor technique due to Noble.[6] The dual series equations considered by Saxena, Sethi, and Banerji[7] are as follows:

$$\sum_{n=0}^{\infty} \frac{A_n}{\Gamma\{(n-q) + (\alpha + q + 1)/k\}} \times Y_n^\alpha(x; k) = f(x), \quad 0 \leq x \leq b,$$

and

$$\sum_{n=0}^{\infty} \frac{A_n}{\Gamma\{(n + \beta - q) + (\alpha + q)/k\}} \times Z_n^\mu(x; k) = g(x), \quad b < x < \infty,$$

where

$$\frac{(q-1)(1-k) + \alpha}{k} + \beta + 1 > \beta > 1 - m,$$

$$\mu + 1 > \frac{q(k-1)}{1} + \alpha + \beta > 0,$$

k and m are positive integers and $f(x)$ and $g(x)$ are prescribed functions.

In this connection we also recall similar work of Thakare[8,9], Patil,[10] H.M. Srivastava,[11,12] K.N. Srivastava,[13] R.P. Srivastava,[14–16] Lowndes,[17] and Askey.[18] Note that for $k = 1$ we get as particular cases the results of[11–13,17] after adjusting the constants appropriately.

In the present paper we solve dual series equations

of the form:

$$\sum_{n=0}^{\infty} \frac{A_n}{\Gamma(\delta+2\beta+kn)} Z_n^{\delta+2\beta-1}(x;k)=f(x), \quad 0\le x<y, \tag{1.3}$$

and

$$\sum_{n=0}^{\infty} \frac{A_n}{\Gamma(\delta+\beta+1+kn)} Z_n^{\delta}(x;k)=g(x), \quad y<x<\infty,$$

where $\beta>0$, $\delta>-1$, $f(x)$, and $g(x)$ are known functions and the A_n are unknown constants which are to be determined.

2. RESULTS REQUIRED IN THE COURSE OF THE ANALYSIS

We require the biorthogonal property of the Konhauser biorthogonal polynomials (Konhauser[4])

$$\int_0^{\infty} \exp(-x)x^{\delta} Z_n^{\delta}(x;k)\cdot Y_m^{\delta}(x;k)\,dx = 0, \quad \text{if } m\neq n,$$
$$= \Gamma(1+\delta+kn)/n!, \quad \text{if } m=n, \tag{2.1}$$

where $\delta>-1$ and $\{Y_m^{\delta}(x;k)\}$ is the Konhauser biorthogonal set of second kind.

The second formula required is the Weyl integral stated by Karande and Thakare[5]

$$\int_{\xi}^{\infty} \exp(-x)(x-\xi)^{\beta-1} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \Gamma(\beta)\exp(-\xi) Z_n^{\delta}(\xi;k), \tag{2.2}$$

where $\beta>0$, $\delta+1>\beta$.

The third result that we require is

$$\frac{d}{d\xi}\int_0^{\xi} (\xi-x)^{\beta-1}x^{\delta+\beta} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \frac{\Gamma(\beta)\Gamma(\delta+\beta+1+kn)\xi^{\delta+2\beta-1}}{\Gamma(\delta+2\beta+kn)} Z_n^{\delta+2\beta-1}(\xi;k), \tag{2.3}$$

where $\delta+2\beta>0$, $\beta>0$, $\delta>-1$.

We have the Riemann–Liouville fractional integral (see Erdelyi[19]) given by Prabhakar[20]

$$\int_0^{\xi} x^{\delta+\beta}(\xi-x)^{\beta-1} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \frac{\Gamma(\delta+\beta+1+kn)\Gamma(\beta)\xi^{\delta+2\beta} Z_n^{\delta+2\beta}(\xi;k)}{\Gamma(1+\delta+2\beta+kn)}, \tag{2.4}$$

where $\beta>0$, $\delta+1>0$.

Incidentally we shall give the proof of the result (2.3). Differentiating both sides of (2.4) w.r.t. ξ and using the result [Konhauser,[4] (8), p. 306]

$$\frac{d}{d\xi}\{Z_n^{\alpha}(\xi;k)\} = -k\xi^{k-1} Z_{n-1}^{\alpha+k}(\xi;k),$$

we get

$$\frac{d}{d\xi}\int_0^{\xi} (\xi-x)^{\beta-1}x^{\delta+\beta} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \frac{\Gamma(\delta+\beta+1+kn)\Gamma(\beta)}{\Gamma(1+\delta+2\beta+kn)}\xi^{\delta+2\beta-1}$$
$$\times\{(\delta+2\beta)Z_n^{\delta+2\beta}(\xi;k) - k\xi^{k} Z_{n-1}^{\delta+2\beta+k}(\xi;k)\}.$$

Now using the pure recurrence relation for $Z_n^{\alpha}(x;k)$ obtained by Karande and Thakare[5]

$$-kx^{k} Z_n^{\alpha+k}(x;k) = (kn+\alpha)Z_n^{\alpha-1}(x;k) - \alpha Z_n^{\alpha}(x;k),$$

the proof of (2.3) is complete.

If $f(\xi)$ and $f'(\xi)$ are continuous in $0\le x<\infty$ and if $0<\beta<1$, then the solutions of the Abel integral equations

$$f_1(\xi) = \int_0^{\xi} \frac{F_1(x)}{(\xi-x)^{\beta}}\,dx, \tag{2.5}$$

and

$$f_2(\xi) = \int_{\xi}^{\infty} \frac{F_2(x)}{(x-\xi)^{\beta}}\,dx, \tag{2.6}$$

are given by

$$F_1(x) = \frac{\sin(\beta\pi)}{\pi}\frac{d}{dx}\int_0^{x} (x-\xi)^{\beta-1} f_1(\xi)\,d\xi, \tag{2.7}$$

and

$$F_2(x) = -\frac{\sin(\beta\pi)}{\pi}\frac{d}{dx}\int_x^{\infty} f_2(\xi)(\xi-x)^{\beta-1}\,d\xi, \tag{2.8}$$

respectively.

3. SOLUTION OF THE EQUATIONS

From (2.3) and (1.3) we get

$$\frac{d}{d\xi}\int_0^{\xi} (\xi-x)^{\beta-1}x^{\delta+\beta}\sum_{n=0}^{\infty}\frac{A_n}{\Gamma(\delta+\beta+1+kn)} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \sum_{n=0}^{\infty}\frac{\Gamma(\beta)A_n}{\Gamma(\delta+2\beta+kn)}\xi^{\delta+2\beta-1} Z_n^{\delta+2\beta-1}(\xi;k)$$
$$= \Gamma(\beta)\xi^{\delta+2\beta-1} f(\xi).$$

Hence

$$\sum_{n=0}^{\infty}\frac{A_n}{\Gamma(\delta+\beta+1+kn)}\frac{d}{d\xi}\int_0^{\xi} (\xi-x)^{\beta-1}x^{\delta+\beta} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \Gamma(\beta)\xi^{\delta+2\beta-1} f(\xi). \tag{3.1}$$

Similarly from (2.2) and (1.4), we get

$$\sum_{n=0}^{\infty}\frac{A_n}{\Gamma(\delta+\beta+1+kn)}\int_{\xi}^{\infty} \exp(-x)(x-\xi)^{\beta-1} Z_n^{\delta+\beta}(x;k)\,dx$$
$$= \Gamma(\beta)\exp(-\xi) g(\xi). \tag{3.2}$$

Let

$$f_1(x) = x^{\delta+\beta} p(x), \tag{3.3}$$

where

$$p(x) = \sum_{n=0}^{\infty}\frac{A_n}{\Gamma(\delta+\beta+1+kn)} Z_n^{\delta+\beta}(x;k). \tag{3.4}$$

Multiplying both sides of (3.3) by $(\xi-x)^{\beta-1}$ and integrating w.r.t. x over $(0,\xi)$ and then differentiating w.r.t. ξ

$$\frac{d}{d\xi}\int_0^{\xi} (\xi-x)^{\beta-1} f_1(x)\,dx = \frac{d}{d\xi}\int_0^{\xi} x^{\delta+\beta}(\xi-x)^{\beta-1} p(x)\,dx.$$

Now using (2.7) and (3.1) we get

$$F_1(\xi) = \frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)\xi^{\delta+2\beta-1} f(\xi). \tag{3.5}$$

Again dividing both sides of (3.5) by $(x-\xi)^{\beta}$, integrating w.r.t. ξ over $(0,x)$ and then using (2.5)

$$f_1(x)=x^{\delta+\beta}p(x)$$
$$=\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)\int_0^x \frac{\xi^{\delta+2\beta-1}f(\xi)}{(x-\xi)^{\beta}}\,d\xi. \tag{3.6}$$

Let

$$f_2(x)=\exp(-x)p(x), \tag{3.7}$$

where $p(x)$ is given by (3.4).

Similarly, multiplying both sides of (3.7) by $(x-\xi)^{\beta-1}$, and integrating w.r.t. x over (ξ,∞) and differentiating w.r.t. ξ, we get

$$F_2(\xi)=-\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)\frac{d}{d\xi}(\exp(-\xi)g(\xi)). \tag{3.8}$$

Dividing both sides of (3.8) by $(\xi-x)^{\beta}$, integrating w.r.t. ξ over (x,∞) and then using (2.6) we get

$$f_2(x)=\exp(-x)p(x)$$
$$=-\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)\int_x^{\infty}\frac{(d/d\xi)(\exp(-\xi)g(\xi))\,d\xi}{(\xi-x)^{\beta}}. \tag{3.9}$$

From (3.6) and (3.9) we write, respectively,

$$p(x)=x^{-\delta-\beta}\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)$$
$$\times\int_0^x \frac{\xi^{\delta+2\beta-1}f(\xi)}{(x-\xi)^{\beta}}\,d\xi,\quad 0<x<y, \tag{3.10}$$

and

$$p(x)=-\exp(+x)\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)$$
$$\times\int_x^{\infty}\frac{(d/d\xi)(\exp(-\xi)g(\xi))}{(\xi-x)^{\beta}}\,d\xi,\quad y<x<\infty. \tag{3.11}$$

The left-hand sides of (3.10) and (3.11) are identical hence multiplying both by $x^{\delta+\beta}\exp(-x)\,Y_m^{\delta+\beta}(x;k)$, integrating (3.10) w.r.t. x over $(0,y)$, integrating (3.11) w.r.t. x over (y,∞), adding and using the orthogonality relation (2.1), we get the solution of the dual series equations (1.3) and (1.4) in the form:

$$A_n=\frac{n!\,\sin(\beta\pi)}{\pi}\Gamma(\beta)\int_0^y \exp(-x)Y_n^{\delta+\beta}(x;k)$$
$$\times\left\{\int_0^x \frac{\xi^{\delta+2\beta-1}f(\xi)}{(x-\xi)^{\beta}}\,d\xi\right\}dx-n!\,\frac{\sin(\beta\pi)}{\pi}\Gamma(\beta)$$
$$\times\int_y^{\infty}x^{\delta+\beta}Y_n^{\delta+\beta}(x;k)$$
$$\times\left\{\int_x^{\infty}\frac{(d/d\xi)(\exp(-\xi)g(\xi))}{(\xi-x)^{\beta}}\,d\xi\right\}dx. \tag{3.12}$$

$$A_n=\frac{n!\,\sin(\beta\pi)}{\pi}\Gamma(\beta)\left\{\int_0^y \exp(-x)\,Y_n^{\delta+\beta}(x;k)\right.$$
$$\left.\times f^*(x)\,dx-\int_y^{\infty}x^{\delta+\beta}\,Y_n^{\delta+\beta}(x;k)\,g^*(x)\,dx\right\}, \tag{3.13}$$

with $\delta+1>0$, $\beta>0$, where

$$f^*(x)=\int_0^x \frac{\xi^{\delta+2\beta-1}f(\xi)}{(x-\xi)^{\beta}}\,d\xi,$$

$$g^*(x)=\int_x^{\infty}\frac{(d/d\xi)(\exp(-\xi)g(\xi))}{(\xi-x)^{\beta}}\,d\xi.$$

In particular, we note that for $k=1$, $\beta=\frac{1}{2}$ we get the dual series equations for $\delta=\alpha$ considered by Chaturvedi[21] a paper which contains several mistakes.

Also if we put $k=1$ and $A_n=\Gamma(\delta+n+1)\Gamma(\delta+\beta+1+n)C_n$ we get the dual series equations considered by Lowndes.[17]

ACKNOWLEDGMENT

The authors are grateful to the referee for suggesting a number of improvements in the original version of the paper.

[1]L. Spencer and U. Fano, J. Res. Natl. Bur. Std. **46**, 446—51 (1951).
[2]J. D. E. Konhauser, J. Math. Anal. Appl. **11**, 242—60 (1965).
[3]S. Preiser, J. Math. Anal. Appl. **4**, 38—64 (1962).
[4]J. D. E. Konhauser, Pacific J. Math. **21**, 303—14 (1967).
[5]B. K. Karande and N. K. Thakare, Indian J. Pure Appl. Math. (accepted for publication).
[6]B. Noble, Proc. Cambridge Philos. Soc. **59**, 363—72 (1963).
[7]R. K. Saxena, P. L. Sethi, P. K. Banerji, paper presented at the 38th Annual Conference of the Indian Mathematical Society (1972).
[8]N. K. Thakare, Shivaji Univ. J. **5**, 67—72 (1972).
[9]N. K. Thakare, Z. Ang. Math. Mech. **54**, 283—84 (1974).
[10]K. R. Patil, "A note on dual series equations involving Konhauser Bi-orthogonal polynomials" (communicated).
[11]H. M. Srivastava, J. Math. Anal. Appl. **31**, 287—94 (1970).
[12]H. M. Srivastava, Pacific J. Math. **30**, 525—27 (1969).
[13]K. N. Srivastava, Pacific J. Math. **19**, 529—33 (1966).
[14]R. P. Srivastava, Proc. R. Soc. Edinburgh, Sect. A **66**, 161—72 (1964).
[15]R. P. Srivastava, Proc. R. Soc. Edinburgh, Sect. A **66**, 173—84 (1964).
[16]R. P. Srivastava, Proc. R. Soc. Edinburgh, Sect. A **66**, 184—91 (1964).
[17]J. Lowndes, Pacific J. Math. **25**, 123—27 (1968).
[18]R. Askey, J. Math. Anal. Appl. **24**, 677—85 (1968).
[19]A. Erdelyi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1953), Vol. 2.
[20]T. R. Prabhakar, Pacific J. Math. **35**, 213—19 (1970).
[21]H. C. Chaturvedi, An elementary solution of dual series equations involving Laguerre Polynomials, Math. Student 40A, 208—10 (1972).

Equatorial circular geodesics in the Kerr–Newman geometry

Naresh Dadhich and P. P. Kale

Department of Mathematics and Statistics, University of Poona, Poona 411 007, India
(Received 2 August 1976)

The conditions for existence, boundedness, and stability are obtained for equatorial circular geodesics in the Kerr–Newman geometry. These conditions are extensions of the Bardeen *et al.* conditions in the Kerr geometry.

I. INTRODUCTION

The circular geodesics in the vicinity of a black hole are gaining significance owing to their interesting astrophysical applications, in particular, the phenomenon of gravitational synchrotron radiation.[1–4] In general relativity, circular geodesics of arbitrary radii are not possible; there occurs a minimum radius (the existence threshold) below which circular geodesics cannot exist. The conditions for existence, boundedness, and stability of circular geodesics have been studied in the Schwarzschild,[5] the Nordström[6,7] and the Kerr[8,9] geometries. As far as we know, this has not been considered in the Kerr–Newman geometry.[10] We shall, in this paper, obtain these conditions for equatorial circular geodesics in the Kerr–Newman geometry which will be extensions of the Bardeen *et al.*[8,9] conditions in the Kerr geometry.

II. THE KERR-NEWMAN GEOMETRY

We shall take the Kerr–Newman metric as a generic metric for a charged and rotating black hole. The metric in the Boyer–Lindquist[11] form is given by

$$ds^2=\frac{\rho^2}{\Delta}\,dr^2+\rho^2\,d\theta^2+\frac{\sin^2\theta}{\rho^2}\,[a\,dt-(r^2+a^2)\,d\phi]^2 - \frac{\Delta}{\rho^2}\,(dt-a\sin^2\theta\,d\phi)^2, \tag{1}$$

where

$$\rho^2=r^2+a^2\cos^2\theta,\quad \Delta=r^2-2mr+a^2+e^2. \tag{2}$$

Here the relativistic units are employed with $G=c=1$ and m, e, and a respectively stand for the mass, the charge, and the specific rotation (angular momentum per unit mass) of the black hole.

We shall consider motion in the equatorial plane $\theta=\pi/2$. The first integrals of motion are well known and are as follows[11]:

$$r^2\frac{dr}{d\lambda}=\sqrt{R}, \tag{3}$$

$$r^2\frac{d\phi}{d\lambda}=-(aE-L)+\frac{a}{\Delta}\,P, \tag{4}$$

$$r^2\frac{dt}{d\lambda}=-a(aE-L)+\frac{r^2+a^2}{\Delta}\,P, \tag{5}$$

with

$$P=E(r^2+a^2)-aL, \tag{6}$$

$$R=P^2-\Delta[\mu^2r^2+(aE-L)^2], \tag{7}$$

E is the energy, L is the angular momentum (axial component), μ is the proper mass of the particle, and λ is the particle's proper time per unit mass, τ/μ, and is an affine parameter for photons.

In the geometry described by Eq. (1), the orthonormal frames of locally nonrotating observers have their basis 1-forms as[11]:

$$\omega^{\hat{t}}=\left|g_{tt}-\frac{g_{t\phi}^2}{g_{\phi\phi}}\right|^{1/2}dt,\quad \omega^{\hat{\phi}}=(g_{\phi\phi})^{1/2}\left(d\phi+\frac{g_{t\phi}}{g_{\phi\phi}}\,dt\right),$$

$$\omega^{\hat{r}}=\frac{\rho}{(\Delta)^{1/2}}\,dr,\quad \omega^{\hat{\theta}}=\rho\,d\theta. \tag{8}$$

Equation (3) would read in the explicit form as

$$\left(\frac{dr}{d\tau}\right)^2=\frac{R}{r^4}=r^{-2}\{[(r^2+a^2)^2-a^2\Delta]E^2 - 2a(r^2+a^2-\Delta)EL+(a^2-\Delta)L^2-\mu^2\Delta r^2\}. \tag{9}$$

The physically accessible and acceptable values of E lie above $E_{\min}$ for a given L at a given r with the requirement $R(r)>0$. We obtain from Eq. (9)

$$E_{\min}=\{aL(2mr-e^2)+r\Delta^{1/2}(r^2L^2 + [r^2(r^2+a^2)+a^2(2mr-e^2)]\mu^2)^{1/2}\} \times[r^2(r^2+a^2)+a^2(2mr-e^2)]^{-1}. \tag{10}$$

Here as $r\to\infty$, $E_{\min}/\mu\to 1$. For stable bound circular orbits $E_{\min}(r)$ is minimum and for unstable unbound circular orbits (unbound in the sense of Wilkins,[12] $E/\mu>1$) $E_{\min}$ is maximum. For some particular value of E, orbit of the particle is such that $E>E_{\min}$ with the turning points at $E=E_{\min}$. $E_{\min}$ is essentially the threshold energy for a given r.

For retrograde orbits $aL<0$, $E_{\min}$ could become negative when r is close to the event horizon. In this region, particles having positive energy in the local observer's frame can have negative energy relative to infinity; their gravitational binding energy exceeds their rest mass energy. Such a region lying between the event horizon $\Delta=0$ and $g_{tt}>0$ is called the ergosphere and existence of the ergosphere is the prime requirement for the possibility of energy extraction from a black hole by the Penrose process.[13]

For a photon, $\mu=0$, Eq. (10) simplifies to

$$E_{\min}=\pm L\left(\frac{r^2\Delta^{1/2}\pm a(2mr-e^2)}{r^2(r^2+a^2)+a^2(2mr-e^2)}\right). \tag{11}$$

The upper sign is for the direct orbit $L > 0$ and the lower sign is for the retrograde orbit $L < 0$.

III. CIRCULAR GEODESICS

For circular orbits we must have $R(r) = 0$ and $R'(r) = 0$ which yield after considerable algebraic manipulation,

$$\frac{E}{\mu} = \frac{r^2 - 2mr + e^2 \pm a(mr - e^2)^{1/2}}{r[r^2 - 3mr + 2e^2 \pm 2a(mr - e^2)^{1/2}]^{1/2}} \tag{12}$$

and

$$\frac{L}{\mu} = \pm \frac{(mr - e^2)^{1/2}[r^2 + a^2 \mp 2a(mr - e^2)^{1/2}] \mp ae^2}{r[r^2 - 3mr + 2e^2 \pm 2a(mr - e^2)^{1/2}]^{1/2}} . \tag{13}$$

Here again the upper sign is for the direct orbit ($L > 0$) and the lower one is for the retrograde orbit ($L < 0$).

The angular velocity of the particle relative to infinity is given by

$$\dot{\phi} = \pm \frac{(mr - e^2)^{1/2}}{r^2 \pm a(mr - e^2)^{1/2}} . \tag{14}$$

The physical velocity of rotation relative to the locally nonrotating observer's frame specified by (8) would read as

$$\dot{\phi}_{\text{phys}} = \pm \frac{(mr - e^2)^{1/2}[r^2 + a^2 \mp 2a(mr - e^2)^{1/2}] \mp ae^2}{(\Delta)^{1/2}[r^2 \pm a(mr - e^2)^{1/2}]} . \tag{15}$$

(i) Existence threshold

The existence threshold for circular orbits is given by

$$r^2 - 3mr + 2e^2 \pm 2a(mr - e^2)^{1/2} \geq 0 \tag{16}$$

so as to have E and L real. The limiting case of equality corresponds to a photon orbit. Let r_{ph} be the smallest root of Eq. (16). The existence condition for timelike circular goedesics is $r > r_{ph}$. The photon orbit with radius $r = r_{ph}$ is the closest possible circular orbit to the black hole.

(ii) Bound threshold

Not all circular orbits with $r > r_{ph}$ are bound; they are bound geometrically but may be unbound in energetics with $E/\mu > 1$. That is, when a particle having $E/\mu > 1$, is given an infinitesimal outward perturbation then it escapes to infinity along an asymptotically hyperbolic trajectory.[12] Such unbound circular orbits would obviously be unstable.

For bound orbits we must have $1 - E/\mu \geq 0$ which reads as follows:

$$\left(r^{3/2} + 2\sqrt{m}r - \frac{e^2}{\sqrt{m}} \mp \frac{a}{\sqrt{m}}(mr - e^2)^{1/2}\right) \times \left(r^{3/2} + 2\sqrt{m}r + \frac{e^2}{\sqrt{m}} \pm \frac{a}{\sqrt{m}}(mr - e^2)^{1/2}\right) \geq 0. \tag{17}$$

Let r_{mb} be the smallest root of the above equation. r_{mb} is the radius of the marginally bound (parabolic) circular orbit which would be the closest bound orbit to the black hole.

(iii) Stability threshold

The stability for orbits (not all bound orbits are stable) further requires $R''(r) \leq 0$. So we obtain

$$mr\Delta - 4(mr - e^2)[(mr - e^2)^{1/2} \mp a]^2 \geq 0 \tag{18}$$

which could equivalently be written as

$$1 - \frac{E^2}{\mu^2} \geq \frac{2}{3}\frac{m}{r} - \frac{e^2[a \mp (mr - e^2)^{1/2}]^{1/2}}{3r^2[\Delta - (a \mp (mr - e^2)^{1/2})^2]} . \tag{19}$$

Let r_{ms} (the smallest root) be the radius of marginally stable circular orbit, that would be the closest stable orbit to the black hole.

One can easily see that all the conditions of Bardeen *et al.*[8,9] could be obtained by putting $e = 0$ in the above relations. There one could fortunately manage to get the conditions in the explicit terms which is not possible here. Putting $a = 0$, we get the conditions for the Nordström geometry[6,7] and putting both $a = 0$, $e = 0$, the well-known conditions $r \geq 3m$, $r \geq 4m$, and $r \geq 6m$ for existence, bound and stable orbits in the Schwarzschild geometry[5] result.

ACKNOWLEDGMENTS

We thank Professor J. V. Narlikar, T.I.F.R. Bombay, for reading the first write up of the paper.

[1]C.W. Misner, Phys. Rev. Lett. 28, 994 (1972).
[2]C.W. Misner, R.A.Breuer, D.R.Brill, P.L. Chrzanowski, H.G. Hughes III, and C.M. Pereira, Phys. Rev. Lett. 28, 998 (1972).
[3]H.G. Hughes, Ann. Phys. 80, 463 (1973).
[4]S.M. Chitre, J.V. Narlikar, and R.C. Kapoor, Nature 252, 460 (1974).
[5]C. Darwin, Proc. Roy. Soc. London 249, 180 (1959).
[6]A. Jr. Armenti, Nuovo Cimento B 25, 442 (1975).
[7]N. Dadhich and P.P. Kale, "Timelike and null geodesics in the Nordström field," Pramana (to be published).
[8]J.M. Bardeen, W.H. Press, and S.A. Teukolsky, Astrophys. J. 178, 347 (1972).
[9]J.M. Bardeen, in *Black Holes*, edited by C. DeWitt and B.S DeWitt (Gordon and Breach, New York 1972).
[10]B. Carter, Phys. Rev. 141, 1242 (1966).
[11]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), pp. 877, 896, 899.
[12]D.C. Wilkins, Phys. Rev. D 5, 814 (1972).
[13]R. Penrose, Riv. Nuovo Cimento 1, 252 (1969).

The Kirkwood–Salsburg equations for a bounded stable Kac potential. I. General theory and asymptotic solutions

N. Grewe and W. Klein*

Institut für Theoretische Physik der Universität zu Köln, 5 Köln 41, Zölpicher Str. 77, Germany
(Received 25 August 1976)

We derive for Kac potentials of the form $\gamma^s\psi(\lambda x)$ an expansion, for all the distribution functions, in powers of γ^s (s = dimension) and prove for $z < \bar{z}_{cr}$ that the expansion is at least asymptotic. The coefficients in the expansion are shown to be solutions of linear operator equations similar to the Kirkwood–Salsburg equation. We also explicitly obtain a rather simple expression for the coefficients of γ^s and show that they are given by solving the Ornstein–Zernicke integral equation with the choice of $-\beta\psi(y)$ for the direct correlation function.

I. INTRODUCTION

In a previous paper[1] (subsequently referred to as GK), we established that the classical m-particle distribution functions $\rho_m(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ for $z < z_1$ reduce to

$$\prod_{j=1}^{m} \rho_1(\mathbf{x}_j)$$

in the limit $\gamma \to 0$, if a Kac potential of the form

$$\phi(x_{12}) = \gamma^s \psi(\gamma x_{12}) \tag{1.1}$$

is chosen with the hard core excluded. $\psi(\gamma x_{ij})$ is bounded and integrable, and $\sum_{i<j} \psi(\gamma x_{ij}) > -mA$. s is the dimension, and $\rho_1(\mathbf{x})$ is given by

$$\rho_1(\mathbf{x}_1) = z \exp[-\beta \int_{\mathbb{R}^s} d^s\mathbf{x}_2 \rho_1(\mathbf{x}_2)\, \phi(x_{12})], \tag{1.2}$$

where z is the activity. Equation (1.2) is known as the Kirkwood—Monroe equation.[2]

This result, derived by Gates with a different method,[3] establishes the mean field nature of the Kirkwood—Monroe theory of melting. Mean field theories are well known in statistical physics and many authors[4] have attempted to systematically improve them. This is generally done by obtaining a formal expansion of some quantity (i.e., free energy or pair distribution function) in powers of γ and establishing that the zeroth power coefficient is the mean field result. Higher powers would then be corrections. It is however generally thought[4a] that these expansions in γ are at best asymptotic and that to obtain useful information about $\gamma \neq 0$ would require resummation of the series. Unfortunately, these series are usually so complicated that it is difficult to establish simple forms for the coefficients and it is almost impossible to analyze convergence properties much less resum the series.

In Sec. II of this paper we derive, in a rather simple way, a formal expansion in powers of γ^s for all the distribution functions for systems with potentials of the form (1.1) with $\psi(\gamma x) \geq 0$. The coefficients of $(\gamma^s)^\nu$ are seen to be solutions of sets of linear equations similar in structure to the Kirkwood—Salsburg[5] equation.

We solve the equation for the coefficient of γ^s and thereby obtain an explicit form for the first order contribution to all the distribution functions. The coefficient for $\rho_2(\mathbf{x}_1, \mathbf{x}_2)$ to order γ^s is seen to obey an Ornstein—Zernicke[6] equation. (The details are presented in Appendix A.) We then show that the γ^s expansion is asymptotic to order γ^s. These results are anticipated by the work of Lebowitz, Stell, and Baer[4d] who obtained, nonrigorously, quite similar expressions for the more complicated Kac potential plus hard core.

In Secs. III and IV we generalize the above results to higher powers of γ^s and to a class of nonpositive definite potentials, respectively. The significance of these results and possible paths of future research are discussed in Sec. V.

II. ASYMPTOTIC SOLUTION OF THE KIRKWOOD-SALSBURG EQUATIONS TO FIRST ORDER IN γ^s FOR NONNEGATIVE POTENTIALS

In this and the following section, we want to consider a system of particles in s dimensions interacting via a Kac potential $\phi(x_{12}) = \gamma^s\psi(\gamma x_{12})$ with $\gamma \geq 0$ and a nonnegative bounded and integrable function ψ,

$$0 \leq \psi(y) \leq A < \infty, \quad C_1 = \int_{\mathbb{R}^s} d^s\mathbf{y}\psi(y) < \infty. \tag{2.1}$$

This pair-potential is regular, since the integral of the absolute value of the Mayer function $f(x_{12}) = \exp[-\beta\phi(x_{12})] - 1$, $\beta = 1/k_B T$ is bounded,

$$C(\beta) = \int_{\mathbb{R}^s} d^s\mathbf{x} |f(x)| \leq \beta C_1 < \infty. \tag{2.2}$$

All information about the system in equilibrium is contained in the set of m-particle distribution functions $\rho_m(\mathbf{x}_1, \ldots, \mathbf{x}_m)$. The vector $\boldsymbol{\rho}(\rho_1, \rho_2, \cdots)$ of these functions obeys the Kirkwood—Salsburg (K—S) equations,[5] which may be written in compact form as[7]

$$(\hat{I} - z\hat{K})\boldsymbol{\rho} = z\boldsymbol{\alpha}, \tag{2.3}$$

where $\hat{I}$ is the unit operator, z is the activity, $\boldsymbol{\alpha} = (1, 0, 0, \cdots)$ and $\hat{K}$ is the Kirkwood—Salsburg operator, defined by

$$\begin{aligned}(\hat{K}\varphi)_m(\mathbf{x}_1, \ldots, \mathbf{x}_m) &= \Big[\prod_{j=2}^{m} (1 + f(x_{1j}))\Big]\Big((1 - \delta_{m,1})\varphi_{m-1}(\mathbf{x}_2, \ldots, \mathbf{x}_m) \\ &+ \sum_{l=1}^{\infty} \frac{1}{l!} \int_{\mathbb{R}^s} \varphi_{m+l-1}(\mathbf{x}_2, \ldots, \mathbf{x}_{m+l}) \\ &\times \prod_{j=m+1}^{m+l} [f(x_{1j})\, d^s\mathbf{x}_j]\Big).\end{aligned} \tag{2.4}$$

$\hat{K}$ operates on the Banach space E_ξ of vectors of functions $\varphi = (\varphi_1, \varphi_2, \cdots)$ where $\varphi_m : \mathbb{R}^{ms} \to \mathbb{R}$ is Lebesque measurable and bounded, the norm in E_ξ being

$$\|\varphi\|_\xi = \sup_{m\in\mathbb{N}}\left(\operatorname*{ess\,sup}_{(\mathbf{x}_1,\ldots,\mathbf{x}_m)\in\mathbb{R}^{ms}} \frac{|\varphi_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)|}{\xi^m}\right),$$
$$\xi > 0 \text{ fixed.} \quad (2.5)$$

The Kirkwood–Salsburg operator is bounded in the operator norm which corresponds to this vector norm,

$$\|\hat{K}\| \le (1/\xi)\exp[\xi C(\beta)] = z_0(\beta,\psi,\xi,\gamma)^{-1}. \quad (2.6)$$

The bound can be made independent of γ, since with (2.2) we have $z_0^{-1} \le (1/\xi)\exp[\xi\beta C_1]$.

A formal expansion of the distribution functions in γ^s can be obtained in the following way. It was shown in GK that for z small enough (2.3) could be solved by iteration and that the solution ρ did not depend on the positions of the particle but on γ times the distance.

If we assume that the solution of (2.3) has this form for all z then we can expand all the Mayer functions $f(x_{ij})$ in (2.4) in series in the variable $\beta\phi(x_{ij}) = \gamma^s\beta\psi(\gamma x_{ij})$, redefine our variables $\gamma x_{ij} \to y_{ij}$, and equate powers of γ^s. We obtain, for the coefficient of each power of γ^s, a linear operator equation to solve. Before discussing the form of these equations it is interesting to note that the simple form for this expansion in powers of γ^s instead of γ comes from the fact that we have no hard core or lattice constant to impose another length scale on the problem. The only length we must concern ourselves with is γ^{-1}. Therefore, we can scale it away by a simple variable transformation $\gamma x \to y$.

An expansion of the Kirkwood–Salsburg operator $\hat{K}$ in terms of γ^s by expanding the Mayer functions in (2.4) meets with serious difficulties because of the product of $[1+f(x_{1j})]$ factors. If we split $\hat{K}$ by setting

$$\hat{K} = \hat{P}\circ\hat{R},$$
$$(\hat{P}\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \left[\prod_{j=2}^{m}(1+f(x_{1j}))\right]\cdot\varphi_m(\mathbf{x}_1,\ldots,\mathbf{x}_m), \quad (2.7)$$

where $\hat{R}$ contains the rest of $\hat{K}$, we see that the expansion coefficients of $\hat{P}$ are unbounded operators from the first order on

$$\hat{P} \to \hat{P}_1 = \hat{I} + \gamma^s\hat{S}_1,$$
$$(\hat{S}_1\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = -\beta\left[\sum_{j=2}^{m}\psi(\gamma x_{1j})\right]\varphi_m(\mathbf{x}_1,\ldots,\mathbf{x}_m). \quad (2.8)$$

Nevertheless, we will see that in this way one can get an asymptotic expansion for the vector ρ of correlation functions. An expansion of the rest operator $\hat{R}$ up to the first order gives

$$\hat{R} \to \hat{R}_1 = Q_0 + \gamma^s\hat{Q}_1,$$
$$(\hat{Q}_0\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = (1-\delta_{m,1})\varphi_{m-1}(\mathbf{x}_2,\ldots,\mathbf{x}_m) + \sum_{l=1}^{\infty}\frac{(-\beta)^l}{l!}\int_{\mathbb{R}^{ls}}\varphi_{m+l-n}(\mathbf{x}_2,\ldots,\mathbf{x}_{m+l}) \times\prod_{j=m+n}^{m+l}[\psi(\gamma x_{1j})\,d^s(\gamma\mathbf{x}_j)], \quad (2.9)$$

$$(\hat{Q}_1\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = -\tfrac{1}{2}\beta\sum_{l=1}^{\infty}\frac{(-\beta)^l}{l!}\sum_{i=m+1}^{m+l}\int_{\mathbb{R}^{ls}}\varphi_{m+l-1}(\mathbf{x}_2,\ldots,\mathbf{x}_{m+l}) \times\psi(\gamma x_{1i})\prod_{j=m+1}^{m+l}(\psi(\gamma x_{1j})\,d^s(\gamma\mathbf{x}_j)). \quad (2.10)$$

The first order approximation to the Kirkwood–Salsburg operator is therefore

$$\hat{K} \to \hat{K}_1 = \hat{Q}_0 + \gamma^s(\hat{S}_1\hat{Q}_0 + \hat{Q}_1). \quad (2.11)$$

If we formally expand the solution ρ of (2.3) up to order γ^s,

$$\rho \to \rho^{(1)} = \chi^{(0)} + \gamma^s\chi^{(1)}, \quad (2.12)$$

and collect terms of the same order in (2.3), we get two hierarchies which are the zeroth order approximation and its first-order correction to the Kirkwood–Salsburg equations,

$$(\hat{I} - z\hat{Q}_0)\chi^{(0)} = z\alpha, \quad (2.13)$$
$$(\hat{I} - z\hat{Q}_0)\chi^{(1)} = z(\hat{S}_1\hat{Q}_0 + \hat{Q}_1)\chi^{(0)}. \quad (2.14)$$

These two equations can be solved uniquely as long as 1 is not an eigenvalue of the (bounded) operator $z\hat{Q}_0$. If we extend E_ξ to complex-valued functions and z to the complex plane we have, for example, by Neumann's theorem that $(\hat{I} - z\hat{Q}_0)$ has a bounded inverse as long as $|z|\|\hat{Q}_0\| < 1$ and the right-hand side of (2.14) is bounded which is the case for $|z| < z_1$. However, it shall be pointed out that this restriction may in general be too strong. In fact, with our approach we obtain a criterion which, although satisfied for all z smaller than the bound found by Ruelle,[7] may be satisfied for much larger z.

The zeroth order equation (2.13) has been studied in GK. It is equivalent—for small enough density—to the Kirkwood–Monroe equation for the zeroth order one-particle distribution function (density) $\chi_1^{(0)} \equiv n$,

$$n(\mathbf{x}_1) = z\exp[-\beta\textstyle\int_{\mathbb{R}^s} d^s\mathbf{x}_2 n(\mathbf{x}_2)\phi(x_{12})]. \quad (2.15)$$

This can be shown easily by an ansatz of the form

$$(\chi^{(0)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \prod_{j=1}^{m} n(\mathbf{x}_j) \quad (2.16)$$

for the hierarchy (2.13). In fact, for all $z \ge 0$ there is a constant solution n of (2.15) ($n = z\exp[-\beta C_1 n]$) which only depends on β, z and not on γ, since $\int_{\mathbb{R}^s} d^s\mathbf{x}_2\phi(x_{12}) = C_1$. This n also determines, via (2.16), the unique solution of (2.13) for $0 \le z \le z_{cr}^{(MF)}$ where $z_{cr}^{(MF)}$ is defined as the infimum of all positive z with $1/z$ belonging to the spectrum of $\hat{Q}_0$. The result, that this is also an asymptotic solution to zeroth order in γ^s of the full Kirkwood–Salsburg equations (2.3) for $0 \le z \le z_1$, has already been proven in GK. An extension to perhaps a larger range of z, namely

$$\|\rho - \chi^{(0)}\|_\xi \le \gamma^s\circ z\,\|(\hat{I} - z\hat{K})^{-1}\|\circ B_0(\beta,\psi;\xi,z), \quad (2.17)$$
$$0 \le z < \min\{z_{cr}, z_{cr}^{(MF)}\} \equiv \bar{z}_{cr},$$

can be achieved by using the direct method, which is presented in Appendix B for the first order case. Here z_{cr} is the infimum of all positive z with $1/z$ belonging to the spectrum of $\hat{K}$.

In order to solve the first order hierarchy (2.14) we make an ansatz

$$(\chi^{(1)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \left[\sum_{j=1}^{m} n(\mathbf{x}_j)\right]\cdot\sum_{i=1}^{m}\left[r(\mathbf{x}_1)+\sum_{i<k\le m} h(\mathbf{x}_1,\mathbf{x}_k)\right], \tag{2.18}$$

where r is the relative first order correction to the density and h is the first order correction to the two-particle correlation function. It is shown in Appendix A that the hierarchy (2.13), (2.14) reduces to the following two integral equations for the functions r and h:

$$h(\mathbf{x}_1,\mathbf{x}_2) = -\beta\psi(\gamma x_{12}) - \beta\int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0) \times n(\mathbf{x}_0)\psi(\gamma x_{10})h(\mathbf{x}_2,\mathbf{x}_0), \tag{2.19}$$

$$r(\mathbf{x}_1) = \tfrac{1}{2}[h(\mathbf{x}_1,\mathbf{x}_1)+\beta\psi(0)] - \beta\int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0)\, n(\mathbf{x}_0)\,\psi(\gamma x_{10})\, r(\mathbf{x}_0). \tag{2.20}$$

In the liquid phase r will be a constant and h will only depend on the absolute distance between the particles, $h(\mathbf{x}_1,\mathbf{x}_2)=h(x_{12})$. In this case the two equations can be simplified,

$$h(x_{12}) = -\beta\psi(\gamma x_{12}) - \beta n\int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0)\psi(\gamma x_{10})h(x_{20}), \tag{2.19'}$$

$$r = (1+n\beta C_1)^{-1}\cdot\tfrac{1}{2}[h(0)+\beta\psi(0)]. \tag{2.20'}$$

By (2.20'), r is directly given in terms of h. From (2.19') it is clear that h scales with γ. Therefore, by defining new variables $\mathbf{y}_i=\gamma\mathbf{x}_i$ and a scaled correlation function $h_{sc}(\gamma x_{ij})=h(y_{ij})$ we have—instead of (2.19)—to consider an equation which does not depend on γ, [8]

$$h_{sc}(y_{12}) = -\beta\psi(y_{12}) - \beta n\int_{\mathbb{R}^s} d^s\mathbf{y}_0\psi(y_{10})h_{sc}(y_{20}). \tag{2.21}$$

The solution of (2.21) determines via (2.20') and (2.18) the unique first-order correction $\chi^{(1)}$ to the vector of correlation functions in the range $0\le z<z_{cr}^{(MF)}$. The full first order approximation $\rho^{(1)}=\chi^{(0)}+\gamma^s\chi^{(1)}$ to the solution ρ of the Kirkwood–Salsburg equations (2.3) is asymptotic to first order in γ^s,

$$\|\rho-\rho^{(1)}\|_\xi \le \gamma^{2s}\cdot z\,\|(\hat{I}-z\hat{K})^{-1}\|\cdot B_1(\beta,\psi;\xi,z), \qquad 0\le z<\bar{z}_{cr}. \tag{2.22}$$

This result is proven in Appendix B. In order to derive this inequality for the maximal range of z, $0\le z<\bar{z}_{cr}$ one has to choose

$$\xi \ge n_{cr} \quad \text{with } n_{cr}=\bar{z}_{cr}\exp(-\beta C_1 n_{cr}) \tag{2.22a}$$

in (2.22) [and also for the corresponding result (2.17)]. Furthermore it is assumed that the functions r and h are bounded in each interval $[0,z]$ with $z<\bar{z}_{cr}$. [9]

III. ASYMPTOTIC SOLUTIONS OF THE KIRKWOOD-SALSBURG EQUATIONS TO HIGHER ORDERS IN γ^s FOR NONNEGATIVE POTENTIALS

The procedure of solving the Kirkwood–Salsburg equations to first order in γ^s which has been developed in the last section can be generalized to higher orders in γ^s. Expansion of the Mayer functions in the product operator $\hat{P}$ and the rest operator $\hat{R}$, defined in (2.7), gives the approximation of these operators to the νth order in γ^s,

$$\hat{P}\to\hat{P}_\nu=\sum_{\mu=0}^{\nu}\gamma^{\mu s}\hat{S}_\mu, \quad \hat{R}\to\hat{R}_\nu=\sum_{\mu=0}^{\nu}\gamma^{\mu s}\hat{Q}_\mu. \tag{3.1}$$

It is easy to give the $\hat{S}_\mu$ explicitly,

$$(\hat{S}_\mu\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \frac{1}{\mu!}\left[-\beta\sum_{j=2}^{m}\psi(\gamma x_{1j})\right]^\mu\cdot\varphi_m(\mathbf{x}_1,\ldots,\mathbf{x}_m), \tag{3.2}$$

or $\hat{S}_\mu=\hat{S}_1^\mu/\mu!$.

They are unbounded operators for $\mu\ge 1$. To write down the $\hat{Q}_\mu$ in a short form, we use the identity $f(x) = (-\phi(x))\cdot\int_0^\beta d\tau\exp[-\tau\phi(x)]$ in the definition (2.4), (2.7) of the rest operator $\hat{R}$,

$$(\hat{R}\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \delta_{\mu,0}(1-\delta_{m,1})\varphi_{m-1}(\mathbf{x}_2,\ldots,\mathbf{x}_m) + \sum_{l=1}^{\infty}\frac{(-1)^l}{l!}\int_{\mathbb{R}^{ls}\times[0,\beta]^l}\varphi_{m+l-1}(\mathbf{x}_2,\ldots,\mathbf{x}_{m+l}) \times\exp\left[-\gamma^s\sum_{j=m+1}^{m+l}\tau_j\psi(\gamma x_{1j})\right]\prod_{j=m+1}^{m+l}(\psi(\gamma x_{1j})\,d^s(\gamma\mathbf{x}_j)\,d\tau_j). \tag{3.3}$$

Now the expansion in terms of γ^s is trivial and we obtain

$$(\hat{Q}_\mu\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_m) = \delta_{\mu,0}(1-\delta_{m,1})\varphi_{m-1}(\mathbf{x}_2,\ldots,\mathbf{x}_m) + \frac{(-1)^\mu}{\mu!}\int_{\mathbb{R}^{ls}\times[0,\beta]^l}\varphi_{m+l-1}(\mathbf{x}_2,\ldots,\mathbf{x}_{m+l}) \times\left[\sum_{i=m+1}^{m+l}\tau_i\psi(\gamma x_{1i})\right]^\mu\sum_{j=m+1}^{m+l}[\psi(\gamma x_{1j})\,d^s(\gamma\mathbf{x}_j)\,d\tau]. \tag{3.4}$$

From this explicit expression for the $\hat{Q}_\mu$ one can deduce that

$$\hat{R}=\sum_{\mu=0}^{\infty}\gamma^{\mu s}\hat{Q}_\mu, \quad 0\le\gamma<\infty \tag{3.5}$$

is an operator-norm convergent series expansion of the rest operator, and in particular that $\|\hat{R}-\hat{R}_\nu\|\to 0$ $(\nu\to\infty)$. The proof is concluded in the following way: With use of the identity

$$\int_0^\infty d\tau_1\cdots\int_0^\beta d\tau_n\left[\sum_{j=1}^{l}\tau_j\right]^\mu=\beta^{\mu+l}\frac{d^\mu}{dx^\mu}\left(\frac{l^x-1}{x}\right)^l\Big|_{x=0}$$

it is straightforward to show that

$$\gamma^{\mu s}\|\hat{Q}_\mu\| \le \frac{1}{\xi}\frac{(\beta A\gamma^s)^\mu}{\mu!}\frac{d^\mu}{dx^\mu}\exp\left(\xi\beta C_1\cdot\frac{l^x-1}{x}\right)\Big|_{x=0}. \tag{3.6}$$

Apart from the factor $1/\xi$, the right-hand side of (3.6) is the $(\mu+1)$th term in the Taylor series of $\exp[\xi\beta C_1(l^x-1)/x]|_{x=\beta A\gamma^s}$ around the point $x=0$. Clearly, $\exp[\xi\beta C_n(l^x-1)/x]$ is an analytic function at $x=0$, so that the Taylor series converges. We have found in this way a convergent majorant to $\sum_{\mu=0}^{\infty}\gamma^{\mu s}\|\hat{Q}_\mu\|$ which proves the proposition.

Expanding the solution ρ of the Kirkwood–Salsburg equations in powers of γ^s up to the νth order

$$\rho\to\rho^{(\nu)}=\sum_{\mu=0}^{\nu}\gamma^{\mu s}\chi^{(\mu)}, \tag{3.7}$$

one gets, by collecting terms of the same order in (2.3), $\nu+1$ hierarchies which determine successively

the zeroth-order solution and its 1st to νth order corrections,

$$(\hat{I}-z\hat{Q}_0)\chi^{(0)}=z\alpha,$$
$$(\hat{I}-z\hat{Q}_0)\chi^{(\mu)}=z\sum_{\lambda=1}^{m}\left(\sum_{\sigma+\tau=\lambda}\hat{S}_\sigma\hat{Q}_\tau\right)\chi^{(\mu-\lambda)}\quad(1\le\mu\le\nu). \tag{3.8}$$

For all orders the operator on the left, $\hat{I}-z\hat{Q}_0$, remains the same and only the inhomogeneous term on the right-hand side changes. This means that the solution of the hierarchies is always unique up to the same value $z_{cr}^{(MF)}$ of the cativity, which is only determined by the zeroth-order rest operator $\hat{Q}_0$.

If one looks at the particular form of the solutions (2.16) and (2.18) of the zeroth and the first-order hierarchies, one is led to the idea that only correlations involving not more than $\nu+1$ particles should enter the solution of the νth order hierarchy. This is physically clear, since a (nonseparable) correlation between $\nu+1$ particles involves at least ν interactions, each of them giving a factor γ^s to the correlation function. One can check this assumption formally, for $z<z_{cr}^{(MF)}$ by inspecting the hierarchies (3.8). Since the operator $\hat{Q}_0$ according to its definition (2.9) [or (3.4)] only involves correlations between particle 1 and others, whose positions are integrated out, no new correlations are created on the left-hand sides of the hierarchies. Therefore, the degree of correlation involved in a vector $\chi^{(\nu)}$ is determined by the action of the operators on the right-hand sides of the hierarchies on the vectors $\chi^{(\mu)}$ $(\mu=0,\ldots,\nu-1)$, which degree of correlation has already been established in the ν lower hierarchies. It is especially clear that the zeroth order hierarchy has an uncorrelated solution, since in the circle $|z|<z_1$ we have by Neumann's theorem $\chi^{(0)}=\sum_{l=0}^{\infty}(z\hat{Q}_0)^l z\alpha$, which remains an uncorrelated solution after analytic continuation in z. In the higher hierarchies correlations are now introduced successively by the factors $\psi(\gamma x_{1j})$ appearing on the right-hand sides of (3.8) in the operators $\hat{S}_\sigma$. The number σ of these factors and the order $\mu-\lambda$ of the vector $\chi^{(\mu-\lambda)}$ have a sum not greater than the order μ of the hierarchy. Therefore, the assumption follows by induction.

It is clear now that for $\nu\ge 1$ we can make an ansatz of the form

$$(\chi^{(\nu)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)=\left[\prod_{j=1}^{m}n(\mathbf{x}_j)\right]\cdot\sum_{k_1=1}^{m}\left[\sigma_1^{(\nu)}(\mathbf{x}_{k_1})+\sum_{k_1<k_2\le m}\left[\sigma_2^{(\nu-1)}(\mathbf{x}_{k_1},\mathbf{x}_{k_2})+\ldots+\sum_{k_1<\cdots<k_{\nu+1}\le m}\sigma_{\nu+1}^{(0)}(\mathbf{x}_{k_1},\ldots,\mathbf{x}_{k_{\nu+1}})\right]\cdots\right]\quad(\nu\ge 1), \tag{3.9}$$

for the solution of the νth order hierarchy, where $\sigma_1^{(\nu)}$ is the νth relative correction to the density, $\sigma_2^{(\nu-1)}$ is the $(\nu-1)$th correction to the two-particle correlation function—whose first (nonseparable) part is of first order in γ^s—and so on, and finally $\sigma_{\nu+1}^{(0)}$ is the first (nonseparable) part of the $(\nu+1)$-particle correlation function.

Inserting (3.9) into the νth-order hierarchy (3.8), with the use of the lower order hierarchies, leads to a system of $\nu+1$ linear integral equations for the functions $\sigma_1^{(\nu)},\ldots,\sigma_{\nu+1}^{(0)}$. As coefficient the lower order correlation functions $\sigma_\tau^{(\mu-\tau+1)}$ $(\tau=1,\ldots,\mu+1;\mu=0,\ldots,\nu)$ will appear. All the $\sigma_\tau^{(\mu-\tau+1)}$ $(\tau=1,\ldots,\mu+1;\mu=0,\ldots,\nu)$ together determine via (3.9) and (3.7) the νth-order approximation $\rho^{(\nu)}$ to the solution ρ of the Kirkwood–Salsburg equations. In analogy to the first-order situation of Sec. II and Appendix B, one can prove the asymptoticity of this approximation (in the limit $\gamma\to 0$),

$$\|\rho-\rho^{(\nu)}\|_\xi\le\gamma^{(\nu+1)s}\cdot z\,\|(\hat{I}-z\hat{K})^{-1}\|\times B_\nu(\beta,\psi;\xi,z),\quad 0\le z<\bar{z}_{cr}. \tag{3.10}$$

Again, this is valid for $\xi\ge n_{cr}$ and under the additional assumption that all the correlation functions, which are involved in $\rho^{(\nu)}$, are bounded in each interval $[0,z]$ with $z<\bar{z}_{cr}$. The technique of proving (3.10) is the same as in Appendix B. Typically the constant B_ν contains terms of the form

$$\max_{m\in\mathbb{N}}\left[m^\lambda\left(\frac{n}{\xi}\right)^m\right]\approx\left(\frac{\lambda}{\ln\xi/n}\right)^\lambda\left(\frac{n}{\xi}\right)^{\lambda/(\ln\xi/n)}, \tag{3.11}$$

where λ increases with the order ν, so that the B_ν diverges for $\nu\to\infty$. This indicates that the asymptotic expansion of ρ is not a convergent one.

IV. EXTENSION TO NEGATIVE-VALUED STABLE POTENTIALS

In the preceding two sections we considered only nonnegative potentials $\psi\ge 0$. The reason for this was that for potentials with negative values the product operator $\hat{P}$ in general is unbounded. Then it is not possible to derive statements about the quality of approximative solutions. However, also in this case one can formally expand operators and vectors in powers of γ^s and the possibility and the method of solving a hierarchy, which gives a contribution of a certain order to the vector of distribution functions, apparently does not depend on the sign of the potential. By thoroughly checking the proofs which led to the asymptoticity results (2.17), (2.22), and (3.10), one notices the possibility of a generalization to potentials with negative values, if one could bound the operator $\hat{P}$.

This, effectively, can be done with the help of a trick introduced by Ruelle.[7] Pair potentials, taking negative values, lead to a nonthermodynamic behavior, if they are not stable. Therefore, one has to impose the stability condition on the potential

$$\sum_{1\le i<j\le m}\phi(x_{ij})\ge -m(B/2)\quad(m\in\mathbb{N};\mathbf{x}_1,\ldots,\mathbf{x}_m\in\mathbb{R}^s). \tag{4.1}$$

B is a positive constant. If (4.1) is fulfilled, one can always find an index $i_0=i_0(\mathbf{x}_1,\ldots,\mathbf{x}_m)$, so that $\sum_{j\ne i_0}\phi(x_{i_0j})\ge -B$. One now defines a permutation operator $\hat{\Pi}_\phi$ by

$$(\hat{\Pi}_\phi\varphi)_m(\mathbf{x}_1,\ldots,\mathbf{x}_{i_0},\ldots,\mathbf{x}_m)=\varphi_m(\mathbf{x}_{i_0},\ldots,\mathbf{x}_1,\ldots,\mathbf{x}_m). \tag{4.2}$$

Then one easily sees that

$$\|\hat{\Pi}_\phi\hat{P}\|\le\exp(\gamma^s\beta B). \tag{4.3}$$

Since the distribution functions ρ_m are invariant with

respect to a permutation of particles, one can consider, instead of (2.3), the modified Kirkwood–Salsburg equations

$$(\hat{I} - z\hat{\Pi}_{\phi}\hat{K})\rho = z\alpha, \tag{4.4}$$

where the modified Kirkwood–Salsburg operator can be factorized into bounded operators $\hat{\Pi}\hat{K} = (\hat{\Pi}\hat{P})\hat{R}$. For technical reasons we also redefine $C_1 = \int_{\mathbb{R}^s} d^s\mathbf{y}\,|\psi(y)| < \infty$. In this new situation all of the proofs go through in the same way as before. For example, in Appendix B one only has to introduce some additional factors $\exp(\gamma^s\beta B)$ and change (B8)–(B10) to inequalities with the factor $\exp[nC(\beta)]$ instead of $\exp[-nC(\beta)]$ on the right-hand side. Also the systems of integral equations, to which the hierarchies of all orders reduce, remain the same because of the invariance property of the distribution functions. Therefore, all results of Secs. II and III are also valid in the more general case of stable, bounded, integrable Kac potentials.

V. RESULTS AND CONCLUSIONS

Our main result is that for potentials of the form (1.1) with the restrictions (2.1) and $z < \bar{z}_{cr}$ a series expansion for all of the distribution functions in powers of γ^s can be found which is at least asymptotic in the $\gamma \to 0$ limit. We also derive explicit expressions for the coefficient of γ^s in the expansion for $\rho_m(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ and indicate how to obtain explicit expressions for coefficients of higher powers.

The coefficient of γ^s is of particular importance as it is necessary to obtain, for example, the density–density correlation function

$$\langle \rho_k \rho_{-k} \rangle \equiv \rho\left[1 + \frac{1}{\rho}\int_{\mathbb{R}^s} d^s\mathbf{x}\exp(i\mathbf{kx})(\rho_2(x) - \rho^2)\right] \tag{5.1}$$

and the isothermal compressibility

$$\chi_T \equiv \frac{\beta}{\rho^2}\lim_{k\to 0}\langle \rho_k \rho_{-k}\rangle. \tag{5.2}$$

These thermodynamic quantities exhibit rather interesting behavior which is investigated in a subsequent paper.

It is clear from the foregoing work that there exist several unsolved problems. We have proven that the expansion is asymptotic only up to $\bar{z}_{cr}$ which is the minimum of z_{cr} and $z_{cr}^{(MF)}$. It is clearly important to extend this range. It has not yet been proven that the expansion is not in fact convergent in some neighborhood of $\gamma = 0$ although it is suspected that it is not. It is also quite important to locate the phase transition and determine its order. These questions and others such as the form of the coefficients for higher powers of γ^s are being investigated.

There are at least two reasons for studying potentials of this kind. First, investigations of the structure of the expansion may indicate possible paths to proving similar theorems about the more complicated hard core plus Kac potential systems. Second, and perhaps more interesting, is the possibility of resummation of the γ expansion and obtaining information about systems with potentials which are finite in range.

In conclusion we would stress that the advantage to investigating systems with potentials of interaction given by (1.1) is that there exists only one length in the problem, that is γ^{-1}. Therefore, one can absorb this length into a scale factor, as we have done, and investigate the behavior of the system on the scale γ^{-1}. This is clearly impossible for lattice systems or potentials with hard cores.

APPENDIX A: REDUCTION OF THE FIRST-ORDER HIERARCHY

We want to show here how the hierarchy for the first-order correction $\chi^{(1)}$ to the solution ρ of the Kirkwood–Salsburg equations (2.3), namely

$$(\hat{I} - z\hat{Q}_0)\chi^{(1)} = z(\hat{S}_1\hat{Q}_0 + Q_1)\chi^{(0)}, \tag{A1}$$

where $\hat{Q}_0$, $\hat{Q}_1$, $\hat{S}_1$, and $\chi^{(0)}$ are defined by (2.9), (2.10), (2.8), and (2.16) is reduced by the ansatz

$$(\chi^{(1)})_m(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \left[\prod_{j=1}^{m} n(\mathbf{x}_j)\right]\cdot \sum_{i=1}^{m}\left[r(\mathbf{x}_i) + \sum_{i<k\le m} h(\mathbf{x}_i, \mathbf{x}_k)\right] \tag{A2}$$

to the integral equations (2.19) and (2.20) for the functions r and h. Applying the operators, which appear in (A1), to the vectors $\chi^{(0)}$ [defined in (2.16)] and $\chi^{(1)}$ one gets infinite sums containing integrals of the functions n, r, h. These sums can be substituted with the aid of the integral equation (2.15) for n by simple expressions. In this way we get

$$([\hat{S}_1\hat{Q}_0 + \hat{Q}_1]\chi^{(0)})_m(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \frac{1}{z}\left[\prod_{j=1}^{m} n(\mathbf{x}_j)\quad \tfrac{1}{2}\beta^2 I_1(1) - \beta\sum_{i=2}^{m}\psi(\gamma x_{1i})\right], \tag{A3}$$

$$([\hat{I} - z\hat{Q}_0]\chi^{(1)})_m(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \left[\prod_{j=1}^{m} n(\mathbf{x}_j)\right]\cdot\left[r(\mathbf{x}_1) + \sum_{i=2}^{m} h(\mathbf{x}_1, \mathbf{x}_i) + \beta I_2(1) + \beta\sum_{i=2}^{m} I_3(1,i) - \tfrac{1}{2}\beta^2 I_4(1)\right], \tag{A4}$$

with

$$I_1(1) = \int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0) n(\mathbf{x}_0)\psi(\gamma x_{10})^2,$$
$$I_2(1) = \int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0) n(\mathbf{x}_0) r(\mathbf{x}_0)\psi(\gamma x_{10}),$$
$$I_3(1,i) = \int_{\mathbb{R}^s} d^s(\gamma\mathbf{x}_0) n(\mathbf{x}_0) h(\mathbf{x}_i, \mathbf{x}_0)\psi(\gamma x_{10}),$$
$$I_4(1) = \int_{\mathbb{R}^{2s}} d^s(\gamma\mathbf{x}_0)\, d^s(\gamma\mathbf{x}_{0'}) n(\mathbf{x}_0) n(\mathbf{x}_{0'}) \times h(\mathbf{x}_0, \mathbf{x}_{0'})\psi(\gamma x_{10})\psi(\gamma x_{10'}).$$

In this way, Eq. (A1) is satisfied, if

$$r(\mathbf{x}_1) + \beta I_2(1) - \tfrac{1}{2}\beta^2[I_1(1) + I_4(1)] + \sum_{j=2}^{m}[h(\mathbf{x}_1, \mathbf{x}_j) + \beta I_3(1,j) + \beta\psi(\gamma x_{1j})] = 0. \tag{A5}$$

Especially for $m = 1$ the sum gives zero and it follows that

$$r(\mathbf{x}_1) + \beta I_2(1) - \tfrac{1}{2}\beta^2[I_1(1) + I_4(1)] = 0. \tag{A6}$$

Therefore, the sum must also be zero for all m and since $\mathbf{x}_2, \ldots, \mathbf{x}_m$ are arbitrary, we have

$$h(\mathbf{x}_1, \mathbf{x}_2) + \beta I_3(1,2) + \beta\psi(\gamma x_{12}) = 0. \tag{A7}$$

If one uses (A7) one easily derives that $I_1(1)+I_4(1) = -(1/\beta)I_3(1,1)$ and with this and (A7) one gets instead of (A6),

$$r(\mathbf{x}_1)+\beta I_2(1)-\tfrac{1}{2}[h(\mathbf{x}_1,\mathbf{x}_1)+\beta\psi(0)]=0. \tag{A8}$$

(A7) and (A8) are identical to Eqs. (2.19) and (2.20), respectively.

APPENDIX B: ASYMPTOTICITY OF THE FIRST-ORDER SOLUTION

In this appendix we want to derive the asymptoticity property (2.22) for the first-order approximation $\rho^{(1)} = \chi^{(0)}+\gamma^s\chi^{(1)}$ to the solution ρ of the Kirkwood–Salsburg equations (2.3). The vectors $\chi^{(0)}$ and $\chi^{(1)}$, respectively are solutions of the zeroth and first-order hierarchies

$$\begin{aligned}(\hat{I}-z\hat{Q}_0)\chi^{(0)}&=z\alpha,\\ (\hat{I}-z\hat{Q}_0)\chi^{(1)}&=z(\hat{S}_1\hat{Q}_0+\hat{Q}_1)\chi^{(0)},\end{aligned} \tag{B1}$$

where $\hat{Q}_0$, $\hat{Q}_1$, and $\hat{S}_1$ are defined by (2.9), (2.10), and (2.8), and $\chi^{(0)}$ and $\chi^{(1)}$ are explicitly given by

$$\begin{aligned}(\chi^{(0)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)&=n^m,\\ (\chi^{(1)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)&=n^m\sum_{i=1}^{m} r+\Big[\sum_{i<k\le m}h(x_{ik})\Big],\end{aligned} \tag{B2}$$

n, r, and h being determined by the integral equations (2.15), (2.19), and (2.20).

By combining the two equations (B1) we get

$$(\hat{I}-z\hat{K}_1)\rho^{(1)}=z\alpha-\gamma^{2s}z(\hat{S}_1\hat{Q}_0+\hat{Q}_1)\chi^{(1)} \tag{B3}$$

with $\hat{K}_1=\hat{Q}_0+\gamma^s(\hat{S}_1\hat{Q}_0+\hat{Q}_1)$. Subtracting (B3) from the Kirkwood–Salsburg equation $(\hat{I}-z\hat{K})\rho=z\alpha$ and adding and subtracting a term $z\hat{K}\rho^{(1)}$, one obtains

$$\begin{aligned}&(\hat{I}-z\hat{K})(\rho-\rho^{(1)})\\ &\quad=z(\hat{K}-\hat{K}_1)\rho^{(1)}+\gamma^{2s}z(\hat{S}_1\hat{Q}_0+\hat{Q}_1)\chi^{(1)},\end{aligned} \tag{B4}$$

from which we get the following bound:

$$\begin{aligned}\|\rho-\rho^{(1)}\|_\xi \le z\,\|(\hat{I}-z\hat{K})^{-1}\|\cdot[\,&\|(\hat{K}-\hat{K}_1)\rho^{(1)}\|_\xi\\ &+\gamma^{2s}(\|\hat{S}_1\hat{Q}_0\chi^{(1)}\|_\xi+\|\hat{Q}_1\chi^{(1)}\|_\xi)].\end{aligned} \tag{B5}$$

To simplify the calculation we split off the first vector norm on the right-hand side,

$$\begin{aligned}\|(\hat{K}-\hat{K}_1)\rho^{(1)}\|_\xi \le\ &\|(\hat{P}-\hat{P}_1)\hat{R}\rho^{(1)}\|_\xi\\ &+\|\hat{P}_1(\hat{R}-\hat{R}_1)\rho^{(1)}\|_\xi+\gamma^{2s}\|\hat{S}_1\hat{Q}_1\rho^{(1)}\|_\xi.\end{aligned} \tag{B6}$$

It remains to be shown that $\|(\hat{P}-\hat{P}_1)\hat{R}\rho^{(1)}\|_\xi$, $\|\hat{P}_1(\hat{R}-\hat{R}_1)\rho^{(1)}\|_\xi$ are bounded to second-order and $\|\hat{S}_1\hat{Q}_0\rho^{(1)}\|_\xi$, $\|\hat{Q}_1\rho^{(1)}\|_\xi$, $\|\hat{S}_1\hat{Q}_1\rho^{(1)}\|_\xi$ to zeroth-order in γ^s. We will demonstrate this explicitly in the first case: By use of $|e^{-x}-1+x|\le\frac{1}{2}x^2$ for $x\ge 0$ we have

$$\begin{aligned}&|([\hat{P}-\hat{P}_1]\hat{R}\rho^{(1)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)|\\ &\quad\le\tfrac{1}{2}\beta^2\gamma^{2s}A^2(m-1)^2[\,|(\hat{R}\chi^{(0)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)|\\ &\qquad+\gamma^s|(\hat{R}\chi^{(1)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)|\,].\end{aligned} \tag{B7}$$

The two terms on the right-hand side can be calculated directly,

$$(\hat{R}\chi^{(0)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)=n^{m-1}\exp[-nC(\beta), \tag{B8}$$

$$\begin{aligned}&(\hat{R}\chi^{(1)})_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)\\ &\quad=n^{m-1}\exp[-nC(\beta)]\cdot\Big[(m-1)r-nC(\beta)r\\ &\qquad+\sum_{2\le i<j\le m}h(x_{ij})+n\sum_{i=2}^{m}\mathcal{J}_1(1,i)+\tfrac{1}{2}n^2\mathcal{J}_2(1)\Big],\end{aligned} \tag{B9}$$

with

$$\begin{aligned}\mathcal{J}_1(1,i)&=\int_{\mathbb{R}^s}d^s\mathbf{x}_0 f(x_{10})h(x_{i0}),\\ \mathcal{J}_2(1)&=\int_{\mathbb{R}^s}d^s\mathbf{x}_0\int_{\mathbb{R}^s}d^s\mathbf{x}_{0'}f(x_{10})f(x_{10'})h(x_{00'}).\end{aligned}$$

If we assume that the absolute values of the functions r and h are bounded in the interval $[0,\bar{z}_{cr}-\epsilon]$ by $\bar{r}$ and $\bar{h}$ respectively, (B7)–(B9) give

$$\begin{aligned}&\|(\hat{P}-\hat{P}_1)\hat{R}\rho^{(1)}\|_\xi\\ &\quad\le\gamma^{2s}\frac{1}{2\xi}\beta^2A^2\exp[-nC(\beta)]\cdot\max_{m\in\mathbb{N}}\{m^2(n/\xi)^m\\ &\qquad\times[1+\gamma^s[(m+nC(\beta))\bar{r}+(\tfrac{1}{2}m(m-1)+mnC(\beta)\\ &\qquad+\tfrac{1}{2}n^2C(\beta)^2)\bar{h}]]\}.\end{aligned} \tag{B10}$$

With $\xi\ge n_{cr}$ we have $0\le x\equiv n/\xi<1$ for all $0\le z<\bar{z}_{cr}$. Therefore, the function $f(m)=m^\lambda x^m$ has a finite maximum for $m=-\lambda/\ln x$. With this and the fact that the bounds $\bar{r}$ and $\bar{h}$ are clearly also functions of β, ψ, ξ, and z, one finally gets

$$\|(\hat{P}-\hat{P}_1)\hat{R}\rho^{(1)}\|_\xi\le\gamma^{2s}B(\beta,\psi;\xi,z). \tag{B11}$$

Bounds for the other four expressions can be obtained in a similar way. Inserting all bounds into (B6) and (B5) gives the desired result,

$$\|\rho-\rho^{(1)}\|_\xi\le\gamma^{2s}z\,\|(\hat{I}-z\hat{K})^{-1}\|\,B(\beta,\psi;\xi,z),\quad 0\le z<\bar{z}_{cr}. \tag{B12}$$

Here, $\xi\ge n_{cr}$, and we remark that B_1 may be chosen independent of z if r and h are bounded in the closed interval $[0,\bar{z}_{cr}]$.

*Present address: Physics Dept., Boston University, Boston, Mass. 02215.

[1]N. Grewe and W. Klein, J. Math. Phys. 17, 699 (1976).
[2]J.G. Kirkwood and E. Monroe, J. Chem. Phys. 9, 514 (1941).
[3]D.J. Gates, Ann. Phys. (N.Y.) 71, 395 (1972).
[4]G.E. Uhlenbeck, P.C. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963); (b) P.C. Hemmer, J. Math. Phys. 5, 75 (1964); (c) J.L. Lebowitz and O. Penrose, J. Math. Phys. 7, 98 (1966); (d) J.L. Lebowitz, G. Stell, and S. Baer, J. Math. Phys. 6, 1282 (1965); (e) J.B. Jalikee, A.J.F. Siegert, and D.J. Vezetti, J. Math. Phys. 10, 1442 (1969).
[5]J.G. Kirkwood and Z.W. Salsburg, Discuss. Faraday Soc. 15, 23 (1953).
[6]See, for example, A. Münster, *Statistical Thermodynamics* (Academic, New York, 1969), Vol. I.
[7]D. Ruelle, *Statistical Mechanics Rigorous Results* (Benjamin, New York, 1969), Chap. 4.
[8]During the preparation of the manuscript a paper by Gates [D.J. Gates, Physica A 81, 47 (1975)] was brought to our attention in which he obtains the result that the direct correlation function in the limit $\gamma\to 0$ is $\gamma^sC(y/\gamma,\gamma)=-\beta\psi(y)$.
[9]It is clear that r and h are bounded from the assumption that we are considering those values of z such that $(\hat{I}-z\hat{K})^{-1}$ exists. We would like to point out however that h and r are possibly bounded for much higher values of z than we are restricted to here.

The Kirkwood–Salsburg equations for a bounded stable Kac potential. II. Instability and phase transitions

N. Grewe and W. Klein*

Institut für Theoretische Physik Universität zu Köln, 5 Köln 41, Germany
(Received 20 September 1976)

We prove that systems interacting via potentials of the form
$\phi(\mathbf{x}_1,\mathbf{x}_2)=\gamma^s\psi(\gamma x_{12})$
where ψ is bounded stable and defined on bounded support are unstable to fluctuations of wavenumber $k'_{\min}\neq 0$ at a particular value v_0 of $v\equiv n\beta$, where n is the density and $\beta=1/k_BT$ in the limit $\gamma\to 0$ (VdW1). We also prove (in the VdW1) that the solution to the equation for the single particle distribution function bifurcates at this same value v_0, that the nonconstant solution is periodic and has a reciprocal lattice vector with a magnitude $k'_{\min}$, and that there exists a type of long range order at v_0. These results are interpreted to indicate the existence of a spinodal point on the liquid isotherm, and similarities between this system and the known properties of the hard sphere fluid are discussed. A theorem is also proven about the range of activity where one has a unique fluid phase, and it is shown that this system has no coexistence region in the usual sense.

I. INTRODUCTION

In the preceding paper[1] (referred to as I) we derived an expansion in powers of γ^s for all the distribution functions for systems interacting via a two-particle potential of the form

$$\phi(x_{12})=\gamma^s\psi(\gamma x_{12}),\quad x_{12}\equiv|\mathbf{x}_2-\mathbf{x}_1|, \tag{1.1}$$

where γ is an arbitrary parameter, its inverse being proportional to the range of the potential, ψ is a bounded, stable,[2] and absolutely integrable function, and s is the dimension. For $0\le z<\bar{z}_{cr}$ the expansion was proven to be at least asymptotic and explicit expressions for the coefficients of the zeroth and first powers of γ^s were found to be

$$\begin{aligned}\rho_m(\mathbf{x}_1,\ldots,\mathbf{x}_m)&=\chi_m^{(0)}(\mathbf{x}_1,\ldots,\mathbf{x}_m)\\&\quad+\gamma^s\chi_m^{(1)}(\mathbf{x}_1,\ldots,\mathbf{x}_m)+O(\gamma^{2s}),\\ \chi_m^{(0)}(\mathbf{x}_1,\ldots,\mathbf{x}_m)&=\prod_{j=1}^{m}n_{sc}(\mathbf{y}_j),\\ \chi_m^{(1)}(\mathbf{x}_1,\ldots,\mathbf{x}_m)&=\Big[\prod_{j=1}^{m}n_{sc}(\mathbf{y}_j)\Big]\cdot\sum_{i=1}^{m}\Big[r_{sc}(\mathbf{y}_i)\\&\quad+\sum_{i<k\le m}h_{sc}(\mathbf{y}_i,\mathbf{y}_k)\Big],\end{aligned} \tag{1.2}$$

where $\mathbf{y}=\gamma\mathbf{x}$ is the scaled position vector and the functions n_{sc}, r_{sc}, and h_{sc} are given by

$$n_{sc}(\mathbf{y}_1)=z\cdot\exp[-\beta\textstyle\int_{\mathbb{R}^s}n_{sc}(\mathbf{y}_2)\psi(y_{12})\,d^s\mathbf{y}_2], \tag{1.3a}$$

$$h_{sc}(\mathbf{y}_1,\mathbf{y}_2)=-\beta\psi(y_{12})-\beta\textstyle\int_{\mathbb{R}^s}n_{sc}(\mathbf{y}_0)\times\psi(y_{10})h_{sc}(\mathbf{y}_2,\mathbf{y}_0)\,d^s\mathbf{y}_0, \tag{1.3b}$$

$$r_{sc}(\mathbf{y}_1)=\tfrac{1}{2}[h_{sc}(\mathbf{y}_1,\mathbf{y}_1)+\beta\psi(0)]-\beta\textstyle\int_{\mathbb{R}^s}n_{sc}(\mathbf{y}_0)\psi(y_{10})r_{sc}(\mathbf{y}_0)\,d^s\mathbf{y}_0. \tag{1.3c}$$

In this paper we prove the existence of an instability in this system in the Van der Waals limit $\gamma\to 0$ (VdW1) for the restricted class of potentials (1.1) which have bounded support as functions of $y_{12}\equiv\gamma x_{12}$,

$$\psi(y_{12})=0\quad\text{for } y_{12}>R_0. \tag{1.1'}$$

This instability is shown to be due to fluctuations of wave vectors $|\mathbf{k}'_{\min}|\neq 0$. In order to gain greater insight into the nature of this instability we assume, as argued for by Gates[3] for all densities ρ and proven by the authors[4] for small ρ, that the solution to the Kirkwood–Monroe (KM) equation (1.3a) correctly gives $\rho_1(\mathbf{x}_1)$ in the VdW1. Assuming the validity of this equation we show that the instability we have found is connected with a bifurcation point of the nonlinear integral equation (1.3a). Following an argument of Gates at this point, the free energy corresponding to the constant solution ceases to be a minimum.

The structure of this paper is as follows. In Sec. II we derive expressions for the energy density, pressure, isothermal compressibility, and structure factor in the VdW1. The structure factor is shown to diverge for particular wave vectors $|\mathbf{k}'_{\min}|\neq 0$ at a certain value v_0 of $v\equiv n\beta$.

In Sec. III we examine the above quantities and illustrate the behavior of the two-particle correlation function g_2 to order γ^s, given by $g_2(x_{12})=1+\gamma^s h_{sc}(y_{12})$, in general and for a particular choice of $\psi(y_{12})$ to gain some additional insight into the meaning of the divergence of the structure factor.

In Sec. IV we show that v_0 is a bifurcation point of the KM equation and that for $\beta n\gtrsim v_0$ there exists a periodic solution of this equation with a period characterized by a reciprocal lattice vector $\mathbf{k}'_{\min}$ with the same magnitude as a wave vector of the critical fluctuations.

Section V contains the proof of a theorem about uniqueness of the solutions of the Kirkwood–Monroe equation and it is also shown that there does not exist for any density a solution which would describe a coexistence region in the usual way.

In the final section, VI, we discuss the meaning of these results and compare them to the work of other authors.

II. THERMODYNAMICS AND STRUCTURE FUNCTION

In I it was shown that the pair correlation function g_2 for $z<\bar{z}_{cr}$ could be written as

$$g_2(x;\gamma^s) = 1 + \gamma^s h_{sc}(y) + \gamma^{2s}\tilde{B}(y,z,\beta;\gamma^s), \tag{2.1}$$

where $\tilde{B}$ is a uniformly bounded function for $y \in \mathbb{R}$ and $z < \bar{z}_{cr}$, B, γ^s restricted to compact sets of $\mathbb{R}$. Using (2.1) we can calculate the energy per particle, pressure, isothermal compressibility, and the structure factor in the limit $\gamma \to 0$. Expressions for these quantities are obtained by inserting $g_2(x;\gamma^s)$ in the standard formulas[5] given below and then taking the VdWl. We have for the energy per particle

$$e(T,n) = (s/2)k_B T + \tfrac{1}{2}n\int_{\mathbb{R}^s}\phi(x)g_2(x)\,d^s\mathbf{x}, \tag{2.2}$$

the pressure

$$p(T,n) = nk_B T - \frac{1}{2s}n^2\int_{\mathbb{R}^s} x\frac{d\phi}{dx}(x)g_2(x)\,d^s\mathbf{x}, \tag{2.3}$$

and the structure factor[6]

$$S(k) = \frac{1}{nk_B T}\left[1 + n\int_{\mathbb{R}^s} g_2(x)\exp(i\mathbf{kx})\,d^s\mathbf{x}\right], \tag{2.4}$$

the $k \to 0$ limit of which is the isothermal compressibility. We have anticipated in the above equations that $n(\mathbf{x})$ is a constant and that in taking the limit $\gamma \to 0$ we will have no contributions from coefficients of powers of $(\gamma^s)^m$ with $m \geq 1$ in the expansion for $\rho_1(\mathbf{x};\gamma^s)$. This will become clear in the course of the calculations.

Inserting (1.1) and (2.1) into (2.2), defining $\mathbf{y} = \gamma\mathbf{x}$ and realizing from (1.3a) and (1.3b) that $h_{sc}(y)$ like $\tilde{B}(y)$ is uniformly bounded for $z < \bar{z}_{cr}$, one has in the limit $\gamma \to 0$

$$e(T,n) = (s/2)k_B T + \tfrac{1}{2}n\hat{\psi}(0), \tag{2.5}$$

where $\hat{\psi}(0) = \int_{\mathbb{R}^s}\psi(y)\,d^s\mathbf{y} < \infty$. Clearly, higher powers of γ^s in the expansion for $\rho_1(\mathbf{x}_1;\gamma^s)$ would give no contribution to the zeroth order result (2.5).

In calculating the pressure according to (2.3) we restrict ourselves to the class of at least piecewise differentiable functions $\psi(y)$ with bounded support, so that

$$\frac{d\psi}{dy}(y) = \tilde{\psi}(y) + \sum_{j=1}^{J} d_j\delta(y - a_j). \tag{2.6}$$

If we then again scale the integration variable to $\mathbf{y} = \gamma\mathbf{x}$ it is obvious that in the VdWl we only have a contribution from the zeroth-order term in (2.1) and one gets by partial integration for $\gamma \to 0$,

$$p(T,n) = nk_B T + \tfrac{1}{2}n^2\hat{\psi}(0). \tag{2.7}$$

The structure function presents a much more difficult problem and a much more interesting result. We can rewrite (2.4) for $k \neq 0$ as

$$S(k) = \frac{1}{nk_B T}\left[1 + n\int_{\mathbb{R}^s}(g_2(x) - 1)\exp(i\mathbf{kx})\,d^s\mathbf{x}\right]. \tag{2.8}$$

Inserting (2.1) and using the explicit form of the Fourier transform $\hat{h}_{sc}(k')$ of $h_{sc}(y)$, where $\gamma k' = k$,

$$\hat{h}_{sc}(k') = -\frac{\beta\hat{\psi}(k')}{1 + \beta n\hat{\psi}(k')}, \tag{2.9}$$

which can simply be derived from the linear integral equation (1.3b), we have

$$S_{sc}(k') \equiv S(k) = \frac{1}{nk_B T}\left(1 - \frac{n\beta\hat{\psi}(k')}{1 + n\beta\hat{\psi}(k')}\right) + \gamma^s F(z,\beta;\gamma). \tag{2.10}$$

The problem is to show that

$$\lim_{\gamma\to 0}\gamma^s F(z,\beta;\gamma) = 0. \tag{2.11}$$

The difficulty lies in the fact that the integral in (2.8) is over an infinite range and is not cut off (or damped for more general potentials) by the bounded support of $\psi(y)$ as in (2.2) and (2.3). Formulating condition (2.11) in a different way,

$$\lim_{\gamma\to 0}\int_{\mathbb{R}^s}(g_2(x) - 1)\exp(i\mathbf{kx})\,d^s\mathbf{x} = \int_{\mathbb{R}^s} h_{sc}(y)\exp(i\mathbf{k'y})\,d^s\mathbf{y}, \tag{2.11'}$$

and scaling the integration variable on the left, we have to prove that the limit and the integral in (2.11') can be commuted, since we know from I that

$$\lim_{\gamma\to 0}\frac{1}{\gamma^s}\left(g_2\left(\frac{y}{\gamma}\right) - 1\right) = h_{sc}(y)$$

for $z < \bar{z}_{cr}$. For this it is sufficient that the $\mathbf{y}$-integral over the absolute value $|g_2(y/\gamma) - 1|$ is uniformly convergent with respect to γ. The idea now is to express this integrand by a cluster function of a system with the scaled distances $\mathbf{y}_{ij} = \gamma\mathbf{x}_{ij}$ between the particles which may have some other scaled inverse temperature β' and activity z', and then to apply certain bounds on the integral over this cluster function using results of Penrose[7] and Ruelle.[8]

For this reason we express $g_2(y/\gamma) - 1$ by the original distribution functions

$$g_2\left(\frac{y}{\gamma}\right) - 1 = \frac{\rho_2(y/\gamma;\gamma^s,z,\beta) - [\rho_1(\gamma^s,z,\beta)]^2}{[\rho_1(\gamma^s,z,\beta)]^2}, \tag{2.12}$$

where we explicitly point out the dependence on γ^s, z, and β on the right-hand side. The distribution functions are defined by,[8]

$$\begin{aligned}\rho_m(\mathbf{x}_1,\ldots,\mathbf{x}_m;\gamma^s,z,\beta) &= \lim_{\Lambda\to\infty}\Xi(\Lambda,\gamma^s;z,\beta)^{-1}\sum_{l=0}^{\infty}\frac{z^{m+l}}{l!}\\ &\quad\times\int_{\Lambda^l} d^s\mathbf{x}_{m+1}\cdots d^s\mathbf{x}_{m+l}\prod_{1\leq i<j\leq m+l}(1 + f_{ij}),\\ \Xi(\Lambda,\gamma^s;z,\beta) &= \sum_{l=0}^{\infty}\frac{z^l}{l!}\int_{\Lambda^l} d^s\mathbf{x}_1\cdots d^s\mathbf{x}_l\prod_{1\leq i<j\leq l}(1 + f_{ij}).\end{aligned} \tag{2.13}$$

For potentials of the form (1.1) we have

$$f_{ij} = \exp[-\beta\gamma^s\psi(\gamma x_{ij})] - 1. \tag{2.14}$$

Convergence of the limit in (2.13) with respect to the Volume Λ is guaranteed inside the Ruelle circle $|z| < \exp(-\beta B - 1)C(\beta)^{-1}$, $C(\beta) \equiv \int_{\mathbb{R}^s}|f_{ij}|\,d^s\mathbf{x}_{ij}$, and B is given by the stability condition for the potential ϕ,

$$\sum_{1\leq i<j\leq m}\phi(x_{ij}) \geq -\frac{m}{2}B \quad (m \in \mathbb{N};\mathbf{x}_1,\ldots,\mathbf{x}_m \in \mathbb{R}^s). \tag{2.15}$$

If now we scale the integration variables in (2.13), $\mathbf{y}_i = \gamma\mathbf{x}_i$, and define $\gamma^s z' = z$, $\beta' = \gamma^s\beta$ we get

$$\rho_m\left(\frac{\mathbf{y}_1}{\gamma},\ldots,\frac{\mathbf{y}_m}{\gamma};\gamma^s,z,\beta\right) = \gamma^{ms}\rho'_m(\mathbf{y}_1,\ldots,\mathbf{y}_m;z',\beta'), \tag{2.16}$$

where ρ'_m is the distribution function of particles at points $\mathbf{y}_i$, interacting via the potential $\psi(y_{ij})$, for activity z' and inverse temperature β'. The representation of ρ'_m as a Λ limit in the sense of (2.13) is valid,

since z lying in the Ruelle circle implies that $|z'| < \exp(-\beta' B' - 1) C'(\beta')^{-1}$, where $C'(\beta') \equiv \int_{\mathbb{R}^s} |f'_{ij}|\, d^s\mathbf{y}_{ij} = \gamma^s C(\beta)$ with $f'_{ij} = \exp[-\beta'\psi(y_{ij})] - 1$, and $B' = B/\gamma^s$ gives the stability condition for the potential ψ like B in (2.15) does for ϕ.

Equation (2.12) can now be written as

$$g_2\left(\frac{y}{\gamma}\right) - 1 = \frac{\omega'(y; z', \beta')}{\rho'_1(z', \beta')^2}\,, \tag{2.17}$$

with the two-particle cluster function ω', generally defined as:

$$\omega'(\mathbf{y}_1, \mathbf{y}_2; z', \beta') = \rho'_2(\mathbf{y}_1, \mathbf{y}_2; z', \beta') - \rho'_1(\mathbf{y}_1; z', \beta')\rho'_1(\mathbf{y}_2; z', \beta'). \tag{2.18}$$

The uniform convergence of the integral

$$\int_{\mathbb{R}^s} \frac{1}{\gamma^s} \left| g_2\left(\frac{y}{\gamma}\right) - 1 \right| d^s\mathbf{y}$$

with respect to γ, which is left to be shown in order (2.11) to be right, therefore is equivalent to the possibility of choosing for any fixed $\epsilon \in 0$ a y_0 independent of γ, so that for all γ,

$$\frac{1}{\gamma^s \rho'_1(z', \beta')} \int_{|\mathbf{y}_{12}| \geq y_0} |\omega'(y_{12}; z', \beta')|\, d^s\mathbf{y}_{12} < \epsilon. \tag{2.19}$$

To show this we use the following results of Ruelle[8]:

$$\omega'(\mathbf{y}_1, \mathbf{y}_2) = \sum_{l=0}^{\infty} \frac{z'^{l+2}}{l!} \int_{\mathbb{R}^{ls}} \varphi'_{l+2}(\mathbf{y}_1, \ldots, \mathbf{y}_{l+2}) \times d^s\mathbf{y}_3 \cdots d^s\mathbf{y}_{l+2}, \tag{2.20}$$

$$\varphi'_m(\mathbf{y}_1, \ldots, \mathbf{y}_m) = \sum_{\Gamma} \prod_{1 \leq i < j \leq m}^{\Gamma} f'_{ij},$$

where the sum extends over all connected graphs Γ with vertices $1, \ldots, m$ and the product is over all pairs $1 \leq i < j \leq m$ such that i and j are joined by a line in Γ. The above expansion for ω' converges uniformly with respect to $\mathbf{y}_1$ and $\mathbf{y}_2$ inside the Ruelle circle and may therefore be integrated term by term. Integrals over the φ'_m functions can be bounded in the following way[7,8]:

$$\int_{\mathbb{R}^{(m-1)s}} |\varphi'_m(\mathbf{y}_1, \ldots, \mathbf{y}_m)|\, d^s\mathbf{y}_2 \cdots d^s\mathbf{y}_m \leq m^{m-2} \exp[(m-2)\beta' B'] C'(\beta')^{m-1}. \tag{2.21}$$

Since the potential $\psi(y)$ has finite support and all the graphs in the representation (2.20) of φ'_m are connected there exists for each $y_0 > 0$ a number $l(y_0)$, so that for each $|\mathbf{y}_1 - \mathbf{y}_2| > y_0$ the first $l(y_0) - 1$ terms in the sum (2.20) for $\omega'(\mathbf{y}_1, \mathbf{y}_2)$ are zero. Clearly, $l(y_0)$ can be chosen to go to infinity if and only if y_0 goes to infinity. Using (2.20) and (2.21) we then have

$$\int_{|\mathbf{y}| \geq y_0} |\omega'(y)|\, d^s\mathbf{y} \geq z'^2 C'(\beta') \sum_{l=l(y_0)}^{\infty} \frac{[z'(l+2)\exp(\beta' B') C'(\beta')]^l}{l!}$$

and with[9] $l! > (l/e)^l$, $z'C'(\beta') = zC(\beta)$, assuming $l(y_0) \geq 2$ and $z < \frac{1}{2}\exp(-\beta B - 1)C(\beta)^{-1}$,

$$\int_{|\mathbf{y}| \geq y_0} |\omega'(y)|\, d^s\mathbf{y} \leq \frac{1}{\gamma^s} z^2 C(\beta) \frac{[2zC(\beta)\exp(\beta B + 1)]^{l(y_0)}}{1 - 2zC(\beta)\exp(\beta B + 1)}\,. \tag{2.22}$$

Choosing a suitable y_0, we can make $l(y_0)$ so large that the right-hand side is smaller than $1/\gamma^s \in \rho_L(z\beta)^2$, where $\rho_L(z, \beta) > 0$ is a lower bound to $\rho_1(\gamma^s, z, \beta)$ in a region around $\gamma = 0$, which exists according to I. (The trivial case $z = 0$ or $\beta = 0$ is to be excluded.) With (2.16) this concludes the proof of the criterion (2.19) for uniform convergence of the scaled integral on the left of (2.11'), so that the limit and the integral can be commuted and (2.10) together with (2.11) is valid.

We therefore have obtained for $z < \frac{1}{2}\exp(-\beta B - 1) \times C(\beta)^{-1}$—which apparently can be improved to z inside the Ruelle circle—the result that in the VdWl

$$S_{sc}(k') = \frac{\beta}{n}[1 + n\beta\hat{\psi}(k')]^{-1}, \tag{2.23}$$

with the corresponding expression for the isothermal compressibility,

$$\mathcal{H}_T = \lim_{k' \to 0} S_{sc}(k') = \frac{\beta}{n}[1 + n\beta\hat{\psi}(0)]^{-1}. \tag{2.24}$$

We examine the meaning of these results in the next section.

III. INSTABILITY AND LONG RANGE ORDER

In the previous section the following results were derived in the VdWl:

$$e(T, n) = (s/2)k_B T + \tfrac{1}{2} n\hat{\psi}(0), \tag{3.1}$$

$$p(T, n) = nk_B T + \tfrac{1}{2} n^2\hat{\psi}(0), \tag{3.2}$$

$$S_{sc}(k'; T, n) = \frac{1}{nk_B T}[1 + n\beta\hat{\psi}(k')]^{-1}, \tag{3.3}$$

$$\mathcal{H}_T(T, n) = \frac{1}{nk_B T}[1 + n\beta\hat{\psi}(0)]^{-1}. \tag{3.4}$$

Here, the zeroth-order density n is given by the KM equation,

$$n = z \cdot \exp[-\beta n\hat{\psi}(0)]. \tag{3.5}$$

The above equations have been shown to be valid for $z < \bar{z}_{cr}$ or z inside the Ruelle circle respectively. However, in the VdWl they are exact and therefore their range of validity includes all parts of the thermodynamic surface which can be reached by analytic continuation. Clearly, this only has meaning in the intersection of the range of analytic continuation of all the above equations.

It is clear that e and p, given by (3.1) and (3.2), are entire functions of n for a fixed T. Since $\psi(y)$ is bounded and stable we must have $\hat{\psi}(0) \geq 0$ [10] and therefore $\kappa_T(T, n)$ is an analytic function of n for all $n \neq 0$. This last result is somewhat surprising in that the VdWl for a hard core plus attractive Kac potential produces a divergence in the analytic continuation of the isothermal compressibility[11] which signifies an instability at the spinodal point of the Van der Waals equation. On the other hand, we can see that the structure factor (3.3) does show an instability for the class of potentials such that $\hat{\psi}(k') < 0$ for some values of k'. However, this instability is at some $k'_{\min} \neq 0$. It is immediately clear that Eqs. (3.1), (3.2), and (3.4) do not have any physical significance as analytic continuations in n beyond the point where (3.3) diverges. This is not to say that they have no physical significance beyond this point, but only that we cannot infer it from analytic continuation.

To gain some further insight into the nature of this instability we will study the first-order contribution $h_{sc}(y)$ to the pair correlation function g_2. Its Fourier transform, which has been obtained in (2.9) as a solution of the integral equation (1.3b), can be written as

$$\hat{h}_{sc}(k') = -\frac{\beta\hat{\psi}(k')}{D(k')}, \quad D(k') \equiv 1 + v\hat{\psi}(k'), \quad v \equiv n\beta. \tag{3.6}$$

$\hat{\psi}$ is a real-valued, integrable, continuous function bounded from below and takes its minimum value at some argument $k'_{\min}$. (There may be several wavenumbers $k'_{\min}$ in which case the following considerations can easily be generalized.) If $\hat{\psi}$ is nonnegative for all k', $\hat{h}_{sc}$ is defined and integrable over the whole range of k' for all physical β and n. In the case $\hat{\psi}(k'_{\min}) < 0$, however, there exists some minimum value v_0 such that

$$D_{cr}(k'_{\min}) \equiv 1 + v_0\hat{\psi}(k'_{\min}) = 0. \tag{3.7}$$

For $v = v_0$ the function $\hat{h}_{sc}(k')$ will have a singularity at $k' = k'_{\min}$ which signals an instability under a periodic perturbation with wavenumber $k'_{\min} \neq 0$. To be specific, we consider the case $s = 3$, where

$$h_{sc}(y) = -\frac{\beta}{4\pi^2 y}\int_{-\infty}^{\infty} k' \sin yk' \frac{\hat{\psi}(k')}{D(k')}\, dk'. \tag{3.8}$$

Since $\psi(y)$ has finite support, $\hat{\psi}(k')$ is an entire function of the complex variable k' and the zeroes of the function D are discrete.[12] Let us assume additionally that the integrand in (3.8) has only simple poles in the complex k' plane for $v \leq v_0$, given by $D(k') = 0$. As an example of a potential fulfilling all assumptions one can use

$$\psi(y) = A\theta(|y| - a), \quad A, a > 0. \tag{3.9}$$

In this example the condition $D(k') = 0$ gives

$$1 - vA\frac{4\pi a}{k'^2}\left[\cos ak' - \frac{\sin ak'}{ak'}\right] = 0, \tag{3.10}$$

which can be solved numerically for complex k'.

From $\hat{\psi}(k') = \hat{\psi}(-k') = \hat{\psi}^*(k'^*)$ we know that the zeroes of D lie symmetrically to the real and to the complex axis and can therefore be labeled as

$$k'^{(4)}_{2\mu+\nu} = \pm\delta \pm i\sigma_\mu, \quad \delta_\mu, \sigma_\mu > 0 \quad (\mu = 0, 1, 2, \cdots, \quad \nu = 1, 2, 3, 4)$$
$$k'^{(2)}_{2\bar{\mu}+\bar{\nu}} = \pm i\theta_{\bar{\mu}}, \quad \theta_{\bar{\mu}} > 0 \quad (\bar{\mu} = 0, 1, 2, \cdots, \quad \nu = 1, 2, 3, 4). \tag{3.11}$$

If all the following sums converge, the integral in (3.8) can be performed by applying the theory of residues. Defining reduced residues by

$$R^{(4)}_\mu = \frac{8i\delta_\mu\sigma_\mu(\delta_\mu + i\sigma_\mu)}{D'(\delta_\mu + i\sigma_\mu)}\,\hat{\psi}(\delta_\mu + i\sigma_\mu),$$

$$R^{(2)}_{\bar{\mu}} = \frac{2i\theta_{\bar{\mu}}}{D'(i\theta_{\bar{\mu}})}\,\hat{\psi}(i\theta_{\bar{\mu}}),$$

we can write down $h_{sc}(y)$ in the following closed form:

$$h_{sc}(y) = -\frac{\beta}{4\pi y}\left[\sum_\mu \frac{e^{-\sigma_\mu y}}{2\delta_\mu\sigma_\mu}(\cos\delta_\mu y\, \mathrm{Im}R^{(4)}_\mu + \sin\delta_\mu y\, \mathrm{Re}R^{(4)}_\mu) + \sum_{\bar{\mu}} \exp(-\theta_{\bar{\mu}} y)\, \mathrm{Re}R^{(2)}_{\bar{\mu}}\right]. \tag{3.12}$$

With varying v the poles $k'^{(2,4)}_\lambda$ of D move in the complex k' plane and a part of them will approach the real axis at wavenumber $k'_{\min}$ for $v \to v_0$. If, for simplicity, there is only a single pair of poles $\delta_0 \pm i\sigma_0$ reaching $k'_{\min}$ and the above sums converge uniformly in y,[13] then we have in the neighborhood of v_0,

$$h_{sc}(y) \approx -\frac{\beta R^{(4)}_0}{8\pi y k'_{\min}} \sin k'_{\min} y \frac{\exp(-\sigma_0 y)}{\sigma_0} \quad (v \lesssim v_0). \tag{3.13}$$

Since $\sigma_0 \to 0$, we can see from (3.13) that v approaching its critical value v_0 implies long range order in h_{sc}, that is the damping term $\exp(-\sigma_0 y)$ in the first-order pair correlation function goes to one. This long range order has a characteristic periodicity which has the same magnitude as the wavenumber of the instability in (3.3).

We note that as one would expect from other work with γ expansions,[14] the coefficient of order γ^s diverges coincident with the appearance of long range order. This implies that the γ expansion derived in I cannot be analytically continued past the critical value of $\beta\rho$, i.e., there is a breakdown at v_0. To restate the argument, it is the long range order in the pair correlation function to order γ^s which produces the singularity in the structure factor.

Before interpreting the results of this section, we present in the next section a proof that the Kirkwood–Monroe equation has a bifurcation point at v_0 and that a periodic solution exists, in which the basis reciprocal lattice vectors have magnitudes $k'_{\min}$.

IV. BIFURCATION OF THE KIRKWOOD-MONROE EQUATION AND MINIMUM OF THE FREE ENERGY

Before discussing the bifurcation of the solutions of Eq. (1.3a) a work should be said about rigor. The results obtained in Secs. II and III are rigorous. We wish in this section to examine the properties of the KM equation for values of β and n for which no rigorous derivation exists. Nevertheless, an argument does exist for the validity of (1.3a) with one rather plausible, however unproven, assumption. This is a result of Gates[3] based on a variational principle derived by Gates and Penrose.[15]

We will in this section assume that Gates is correct and that the solution to the Kirkwood–Monroe equation gives the single particle distribution function with the corresponding lowest free energy in the VdWl. We will show using a theorem proven by Krasnoselskii[16] and used previously by Weeks, Rice, and Kozak[17] and Raveché and Stuart[18] that the KM equation (1.3a) has a bifurcation point at $n\beta = v_0$ and that the constant solution of it does not have the lowest free energy at this point. We also prove that the solution is periodic with basis reciprocal lattice vectors with magnitude $k_{\min}$ given by (3.7).

We essentially follow the argument presented by Raveché and Stuart. With a particular lattice in mind we define a Hilbert space $\mathcal{H}$ to be those functions which are invariant under the space group of the particular lattice chosen. We additionally demand reflection symmetry about the origin

$$\varphi(-\mathbf{y}) = \varphi(\mathbf{y}), \quad \varphi \in \mathcal{H}. \tag{4.1}$$

An inner product can be defined rather simply and it can be shown that the operator, defined by the exponent of the KM equation

$$\rho_1(\mathbf{y}_1) = z \cdot \exp[-\beta \int_{\mathbb{R}^s} \rho_1(\mathbf{y}_2)\psi(y_{12})\, d^s\mathbf{y}_3], \tag{4.2}$$

is compact.[18] To employ the bifurcation theorem of Krasnoselskii, which is quoted in the Appendix, we must rewrite the integral equation (4.2). To do this we define a function p, given by

$$\rho_1(\mathbf{y}) = n[1 + p(\mathbf{y})], \tag{4.3}$$

where n is the constant solution of (4.2). Inserting (4.2) in (4.1) gives the following equation for p:

$$p(\mathbf{y}_1) = \exp[-n\beta \int_{\mathbb{R}^s} p(\mathbf{y}_2)\psi(y_{12})\, d^s\mathbf{y}_2] - 1. \tag{4.4}$$

This equation now satisfies the conditions of the theorem stated in the Appendix. We see immediately that (4.4) can only have a bifurcation point in $\mathcal{H}$ when the linear equation

$$\tilde{p}(\mathbf{y}_1) + \beta n \int_{\mathbb{R}^s} \tilde{p}(\mathbf{y}_2)\psi(y_{12})\, d^s\mathbf{y}_2 = 0 \tag{4.5}$$

has a nonzero solution in this Hilbert space. It can be seen from the arguments in Ref. 18 that (4.4) will have a bifurcation point in one, two, and three dimensions for $\rho\beta = v_0$, if v_0 is defined as the first positive value of $\rho\beta \equiv v$ such that

$$1 + v\hat{\psi}(k') = 0 \tag{4.6}$$

can be satisfied for $k' = |\mathbf{k}'_{\min}|$, where $\mathbf{k}'_{\min}$ is a basis reciprocal lattice vector of the lattice on which the Hilbert space $\mathcal{H}$ has been defined.

We note immediately that v_0 and $k'_{\min}$, defined by (4.6), are the same values of these variables at which the instability, obtained in Sec. II, occurs. By employing the above mentioned variational method of Gates[3] one can show that the constant solution n of (4.2) is no longer the one with the lowest free energy.

To sum up the results of this section: We have shown that at $\rho\beta = v_0$ the KM equation has a bifurcation point. The nonconstant part of the solution that appears at v_0 is periodic with a reciprocal lattice vector $\mathbf{k}'_{\min}$. These values of v_0 and $|\mathbf{k}'_{\min}|$ are the same as those calculated for the instability in Sec. II.

V. UNIQUENESS OF SOLUTIONS OF THE KIRKWOOD–MONROE EQUATION AND NONEXISTENCE OF A COEXISTENCE REGION

In this section only potentials are considered which in addition to all previous conditions are nonnegative. We prove a theorem which determines a region of uniqueness of the solution of the KM equation. We also prove that at no value of the activity z can we have solutions which are nontrivial linear combinations of a constant solution n and a nonconstant solution $m(\mathbf{y})$, both at this same value z. This indicates that the KM equation does not have a coexistence region in the usual sense.

First we show that the solution of the Kirkwood–Monroe equation (1.3a) is unique in the space of real functions if $0 \leq n\beta\hat{\psi}(0) < 1$ [19] and is therefore identical to the constant solution which exists everywhere. An examination of (1.3a) immediately reveals that $n_{sc}(\mathbf{y}_1) \geq 0$ for all $\mathbf{y}_1$. That in turn implies that $n_{sc}(\mathbf{y}_1) \leq z$, since $\psi(y_{12})$ here is a positive definite function. The process of inserting lower and upper bounds in (1.3a) to obtain new upper and lower bounds for $n_{sc}(\mathbf{y}_1)$ can be continued to generate two sequences, one for the upper bound and one for the lower. If the iteration procedure approaches a fixed point then the two sequences must converge to the same limit and the solution of (1.3a) is a unique constant.

In order to show that for $0 \leq n\beta\hat{\psi}(0) < 1$ there is a unique fixed point we introduce dimensionless variables $\tilde{n} \equiv n\beta\hat{\psi}(0)$ and $\tilde{z} \equiv z\beta\hat{\psi}(0)$ and define functions f and g by

$$f(\lambda) = \tilde{z}e^{-\lambda}, \quad g(\lambda) = f(f(\lambda)), \quad \lambda \in \mathbb{R}. \tag{5.1}$$

We will prove now that any sequence, recursively defined by

$$\tilde{z} \geq \lambda_0 \geq 0, \quad \lambda_\nu = g(\lambda_{\nu-1}) \quad (\nu = 1, 2, 3, \cdots), \tag{5.2}$$

approaches a unique fixed point $\lambda^* \in [0, \tilde{z}]$ of g for $\tilde{z} < e$. Then it is clear that the sequence $\lambda'_0 = 0$, $\lambda'_\mu = f(\lambda'_{\mu-1})$ $(\mu = 1, 2, 3, \cdots)$ will also approach λ^*, since this is true for both of the subsequences $\{\lambda'_{2\nu}\}_{\nu=0}^{\infty}$ and $\{\lambda'_{2\nu+1}\}_{\nu=0}^{\infty}$ and that $f(\lambda^*) = \lambda^*$. The above condition $\tilde{z} < e$ is easily seen from the KM equation $\tilde{z} = \tilde{n}e^{\tilde{n}}$ to be fulfilled if $\tilde{n} < 1$. This completes the proof.

A sufficient condition for any sequence (5.2) approaching a unique fixed point λ^* of g is that $\tilde{z} \geq g(\lambda) \geq 0$ and $|g'(\lambda)| \leq A < 1$ for all $\tilde{z} \geq \lambda \geq 0$. The first inequality is trivial, since clearly $\tilde{z} \geq f(\lambda) \geq 0$ for $\tilde{z} \geq \lambda \geq 0$. With $f'(\lambda) = -f(\lambda)$ we have

$$g'(\lambda) = f(\lambda) \cdot g(\lambda), \quad g''(\lambda) = f(\lambda)g(\lambda)[f(\lambda) - 1]. \tag{5.3}$$

Since for $\tilde{z} \geq \lambda \geq 0$, clearly $f(\lambda)$ and $g(\lambda)$ are greater than zero, $|g'|$ takes its maximum value either at $\lambda = 0$ or at $\lambda = \tilde{z}$ or at $\bar{\lambda}$, where $f(\bar{\lambda}) = 1$. But $|g'(0)|$, $|g'(\tilde{z})| \leq \tilde{z}^2 e^{-\tilde{z}} \leq 4e^{-2} < 1$ and $g'(\lambda) = \tilde{z}e^{-1}$, which finally gives the above stated condition $\tilde{z} < e$.

The last thing we will show in this section is that the KM equation cannot have a solution of the form

$$n_{sc}(\mathbf{y}) = an + bm(\mathbf{y}), \quad a + b = 1, \quad b \neq 0, 1, \tag{5.4}$$

where n is a constant solution, m a nonconstant solution, both for the same value of z, at which one might expect a coexistence region. Inserting (5.4) into (1.3a) and using that n and m themselves are solutions of that equation, gives

$$an + bm(\mathbf{y}) = z\left(\frac{n}{z}\right)^a \left(\frac{m(\mathbf{y})}{z}\right)^b$$

and since $a + b = 1$ we have

$$an + bm(\mathbf{y}) = n^a m(\mathbf{y})^b. \tag{5.5}$$

Because the potential is nonnegative, any solution of the KM equation is bounded and hence continuous. Therefore, we can find a $\mathbf{y}_0$ with $m(\mathbf{y}_0) > 0$, so that $m(\mathbf{y}) \neq m(\mathbf{y}_0)$ in the neighborhood of $\mathbf{y}_0$. Then we get, by subtracting equation (5.5) for $\mathbf{y} = \mathbf{y}_0$ and $\mathbf{y} = \mathbf{y}_0 + \Delta\mathbf{y}$ from one another,

$$\frac{b}{n^a} = m(\mathbf{y}_0)^{b-1} \cdot \left[1 - \frac{m(\mathbf{y}_0 + \Delta\mathbf{y})}{m(\mathbf{y}_0)}^b\right]\left(1 - \frac{m(\mathbf{y}_0 + \Delta\mathbf{y})}{m(\mathbf{y}_0)}\right)^{-1}. \tag{5.6}$$

Since in the limit $\Delta\mathbf{y} \to 0$ the right-hand side goes to

$bm(\mathbf{y}_0)^{b-1}$ it follows that $n^{-a} = m(\mathbf{y}_0)^{b-1}$ for $b \neq 0$. Repeating the argument for some different $\mathbf{y}_0'$ in the neighborhood of $\mathbf{y}_0$, giving a different value of m, we have a contradiction for $0 \neq b \neq 1$. In the next section we discuss the significance of this result.

VI. RESULTS AND CONCLUSIONS

Our main results are that for the potentials we have stipulated, there exists an instability in the system in the VdWl at $n\beta = v_0$ which is due to fluctuations of wave vectors with a magnitude $k'_{\min} \neq 0$. This instability is shown to be connected with the onset of a type of long range order. Furthermore, we showed that the solutions of the KM equation bifurcated in a specially chosen Hilbert space at $\rho\beta = v_0$ and that the nonconstant solution that appeared at this point had a reciprocal lattice vector with a magnitude $k'_{\min}$. We also showed that the KM equation does not admit a solution of the usual coexistence form. All of these above results were obtained rigorously.

These results seem to indicate that the systems dealt with here have a phase transition (perhaps first order) with a fluid isotherm that ends at a point of instability. The instability is to fluctuations of a wavenumber $\mathbf{k}'_{\min} \neq 0$ which is also the magnitude of a reciprocal lattice vector of the solid which is presumably stable at this activity and temperature. The results also seem to indicate that this system will have no coexistence region.

The above results are not only interesting as properties of a mean field model arrived at rigorously but they also bear a striking similarity to results obtained for hard sphere fluids. Of particular interest in this context is the result of Raveché and Stuart[18] who showed that an equation quite similar in form to (4.4) has a bifurcation at the end of the fluid branch. We show in a subsequent paper that the hard sphere system also has an instability at this point in analogy to the VdWl case. We also note that the mean field model for these potentials also predicts the absence of divergences in the thermodynamic quantities such as κ_T. This is also found for hard spheres.[20] The nonexistence of a coexistence region in the VdWl is of some interest in light of the above similarity between the hard sphere and mean field repulsive potential cases. We merely note that a coexistence region, i. e., Maxwell construction, is found in the hard core plus attractive Kac potential case.[11]

APPENDIX

For the sake of completeness we state a theorem of Krasnoselskii[16]:

Let A be a completely continuous operator having a Fréchet derivative B at the point O and satisfying $AO = 0$. Then each characteristic value of odd multiplicity of the linear operator B is a bifurcation point of the operator A.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge useful conversations with Professor J. Zittartz and Dr. H. Moraal.

*Present address: Physics Dept., Boston University, Boston, Mass., 02215.

[1]N. Grewe and W. Klein, J. Math. Phys. **18**, 1729 (1977).
[2]D. Ruelle, *Statistical Mechanics Rigorous Results* (Benjamin, New York, 1969).
[3]D.J. Gates, Ann. Phys. (N.Y.) **71**, 395 (1972).
[4]N. Grewe and W. Klein, J.Math. Phys. **17**, 699 (1976).
[5]P. Schofield, Proc. Phys. Soc. **88**, 149 (1966).
[6]We have defined $S(k)$ so that $\lim_{k\to 0} S(k)$ is the isothermal compressibility.
[7]O. Penrose, J. Math. Phys. **4**, 1312 (1963).
[8]Ref. 2, Chap. 4.
[9]W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968).
[10]Ref. 2, Chap. 4.
[11]M. Kac, G.E. Uhlenbeck, and P. Hemmer, J. Math. Phys. **4**, 216 (1963); J.L. Lebowitz and O. Penrose, J. Math. Phys Phys. **7**, 98 (1965).
[12]J. Pierpont, *Functions of a Complex Variable* (Ginn, Boston, 1914), p. 105.
[13]W. Klein, J. Math. Phys. **14**, 1049 (1973).
[14]F.S. Høye, Phys. Rev. B **9**, 2390 (1974); J.B. Jalickee, A.J.F. Siegert, and D.J. Vezzetti, J. Math. Phys. **10**, 1442 (1969).
[15]D.J. Gates and O. Penrose, Commun. Math. Phys. **15**, 255 (1969).
[16]M.A. Krasnoselskii, *Topological Methods in the Theory of Nonlinear Integral Equations* (MacMillan, New York, 1964). Also see the Appendix.
[17]J.D. Weeks, S.A. Rice, and J.J. Kozak, J. Chem. Phys. **52**, 2416 (1970).
[18]H.J. Raveché and C.A. Stuart, J. Chem. Phys. **63**, 1094 (1975).
[19]The idea for this proof arose in a conversation with H. Moraal.
[20]W.G. Hoover and F.H. Ree, J. Chem. Phys. **49**, 3609 (1968).

Lie theory and the wave equation in space-time. 5. *R*-separable solutions of the wave equation $\psi_{tt}-\Delta_3\psi=0$

E. G. Kalnins

Department of Mathematics, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 9 November 1976)

A detailed classification is made of orthogonal coordinate systems for which the wave equation $\psi_{tt}-\Delta_3\psi=0$ admits an R-separable solution. Only those coordinate systems are given which are not conformally equivalent to coordinate systems that have been found in previous articles. We find 106 new coordinates to give a total of 367 conformally inequivalent orthogonal coordinates for which the wave equation admits an R-separation of variables.

INTRODUCTION

In this article we continue our investigation of the orthogonal R-separable coordinate systems for which the wave equation in space–time,

$$\psi_{tt}-\Delta_3\psi=0 \qquad (*)$$

admits an R-separation of variables.[1–4] In a previous article[4] we have studied coordinate systems for which the Klein–Gordon equation

$$\psi_{tt}-\Delta_3\psi=\lambda\psi \qquad (**)$$

admits a separation of variables. Such coordinate systems also admit a separation of variables for the wave equation (*). In Paper 4 of this series we found 261 conformally inequivalent coordinate systems of this type. It is the purpose of this article to find coordinate systems for which (*) admits a strictly R-separable solution. By this we mean those coordinate systems for which (*) admits an R-separable solution and for which there is no conformally equivalent coordinate system such that (*) is simply separable. As with the treatment of the wave equation in two space dimensions,[5] we classify the different types of orthogonal coordinate systems whose coordinate curves are cyclides or their degenerate forms.

The content of the article is arranged as follows. In Sec. I we discuss the relevant details concerning coordinate systems whose coordinate curves are cyclides of most general type. This is a development of the methods in the fundamental book by Bôcher.[6] Also in this section we give the various differential forms corresponding to the coordinate systems of interest. In Sec. II we present the coordinate systems together with the corresponding separation equations and triplet of mutually commuting operators $\{L_1, L_2, L_3\}$ which describe each such system.

I. *R*-SEPARABLE DIFFERENTIAL FORMS FOR THE WAVE EQUATION

Here we classify orthogonal differential forms for which the wave equation (*) admits a strictly "R-separable" separation of variables. We recall that if ψ is a solution of (*) which is R-separable in terms of some new coordinates x_i $(i=1, 2, 3, 4)$, then ψ can be written in the form

$$\psi=\exp[Q(x_1, x_2, x_3, x_4)]\phi \qquad (1.1)$$

where the equation for the function ϕ is such that it admits a separation of variables. The factor $\exp Q$ is called the modulation function and has a definite form for each R-separable coordinate system. In addition no part of the function Q should contain the sum of functions f_i of only one of the variables x_i. For a strict R-separable system the modulation function Q should not be zero. In a previous article[5] where we treated the wave equation in two space variables, it was shown that only cyclidic coordinate systems whose coordinate surfaces were degenerate forms of confocal cyclides of the most general type were strictly R-separable. All remaining cylidic R-separable coordinate systems could be transformed into coordinate systems for which the Klein Gordon equation $(\partial_{tt}-\Delta_2)\psi=\lambda\psi$ also admits a separation of variables. This was done by a suitable transformation of the $O(3, 2)$ conformal symmetry group of $(\partial_{tt}-\Delta_2)\psi=0$. The same situation holds in the case of three spatial dimensions, and it is accordingly the purpose of this section to discuss confocal families of cyclides of general type and their degenerate forms. We now briefly outline the properties of cyclides of this type and refer the reader for details to our previous article[5] and the book by Bôcher.[6] Families of confocal cyclides have their natural setting in a six-dimensional projective space. Elements of this space are specified by six homogeneous coordinates $y_1 : y_2 : y_3 : y_4 : y_5 : y_6$, which are not all simultaneously zero and which are connected by the relation

$$y_1^2+y_2^2+y_3^2+y_4^2+y_5^2+y_6^2=0. \qquad (1.2)$$

The space–time coordinates are related to the homogeneous coordinates via the relations

$$\begin{aligned} y_1&=i(p^2-q^2-r^2-s^2+w^2),\\ y_2&=p^2-q^2-r^2-s^2-w^2,\\ y_3&=2pw,\quad y_4=2iqw,\\ y_5&=2irw,\quad y_6=2isw, \end{aligned} \qquad (1.3)$$

where $t=p/w$, $x=q/w$, $y=r/w$, $z=s/w$. A cyclide is then defined as the locus of points lying on the quadric surface

$$\Phi=\sum_{i,j=1}^{6} a_{ij}y_iy_j=0$$

with $a_{ij}=a_{ji}$ and $\det(a_{ij})\neq 0$. The classification of cyclides under the group of orthogonal transformations which preserves the form

$$\sum_{i=1}^{6} {y_i}^2$$

is then the problem of classifying the intersections of two quadratic forms in six-dimensional projective space. This is performed by the method of elementary divisors applied to the two quadratic forms. (For the details of this classification see Refs. 5, 6).

The equation describing the most general family of confocal cyclides in this six-dimensional projective space is

$$\sum_{i=1}^{6}\frac{{y_i}^2}{\lambda-e_i}=0,\quad \sum_{i=1}^{6}{y_i}^2=0. \tag{1.4}$$

Here λ is one of the new curvilinear coordinates and $e_i\neq e_j$, if $i\neq j$ $(i,j=1,\ldots,6)$. If we choose an orthogonal coordinate system in space–time whose coordinate surfaces have equations of the type (1.4), then the line element in terms of these new coordinates becomes

$$ds^2=\frac{1}{4\sigma w^4}\left[\sum_{i=1}^{4}\frac{(x_i-x_j)(x_i-x_k)(x_i-x_l)}{f(x_i)}{dx_i}^2\right] \tag{1.5}$$

where

$$f(x_i)=\prod_{j=1}^{6}(x_i-e_j)\quad\text{and}\quad\frac{-1}{\sigma}=\sum_{i=1}^{6}e_i{y_i}^2.$$

The coordinates y_i are related to the curvilinear coordinates x_i via the equations

$$y_i=\phi(e_i)/f'(e_i),\quad i=1,\ldots,6, \tag{1.6}$$

where $\phi(\lambda)=\prod_{j=1}^{4}(\lambda-x_j)$. If we write the solution ψ of the wave equation as

$$\psi=(\sigma^{1/2}w^2)\Phi, \tag{1.7}$$

then Φ satisfies the differential equation

$$\sum_{j=1}^{4}\left[\left(\frac{1}{\phi'(x_j)}\frac{\partial^2\Phi}{\partial {v_j}^2}\right)+3x_j\Phi\right]-2\left(\sum_{i=1}^{6}e_i\right)\Phi=0, \tag{1.8}$$

where $2dv_j=dx_j/\sqrt{f(x_j)}$. This equation admits separable solutions for the function Φ, i.e.,

$$\Phi=\prod_{j=1}^{4}E_j(x_j).$$

Each of the functions E_j satisfies the differential equation

$$\frac{d^2E_j}{{dv_j}^2}+\left[3{x_j}^4-2\left(\sum_{i=1}^{6}e_i\right){x_j}^3+A{x_j}^2+Bx_j+C\right]E_j=0. \tag{1.9}$$

We now proceed to classify coordinate systems of this type by considering the expression inside the square brackets in (1.5) and finding out what ranges of the coordinates x_i permit this differential form to have overall negative signature. We must also consider degenerate forms of these general coordinate systems which result when some of the e_i become equal. In addition we should mention that two confocal families of cyclides of type (1.4) are equivalent under the action of real linear transformations of the coordinates y_i which preserve the quantity $\sum_{i=1}^{6}{y_i}^2$ if their parameters e_i, e_i' and coordinates x_i, x_i' are related by the equations

$$e_i=\frac{\alpha e_i'+\beta}{\gamma e_i'+\delta},\quad x_i=\frac{\alpha x_i'+\beta}{\gamma x_i'+\delta}, \tag{1.10}$$

where $\alpha,\beta,\gamma,\delta\in\mathbb{R}$ and $\alpha\delta-\beta\gamma\neq 0$.

We now give the classification of the strictly R-separable coordinate systems, in particular the differential forms.

[1] The first type of differential form corresponds to R-separable coordinate systems of the type (1.6) for which all the e_i are real. In addition the relations (1.10) can be used to standardize these quantities so that $e_1=\infty$, $e_2=a$, $e_3=b$, $e_4=c$, $e_5=1$, $e_6=0$ with $a>b>c>1$. The differential form then becomes

$$ds^2=\left(\frac{-{y_1}^2}{4w^4}\right)\left[\sum_{i=1}^{6}\frac{(x_i-x_j)(x_i-x_k)(x_i-x_l)}{h(x_i)}{dx_i}^2\right] \tag{1.11}$$

where $h(x)=(x-a)(x-b)(x-c)(x-1)x$. The ranges of variation of the variables x_i are

$$\begin{aligned}
&x_1,x_2,x_3>a>x_4>b;\quad x_1,x_2>a>b>x_3>c>x_4>1;\\
&x_1,x_2,x_3>a>b>x_4>c;\\
&x_1>a>x_2,x_3>b>c>x_4>1;\\
&x_1,x_2>a>x_3>b>x_4>c;\\
&x_1>a>b>x_2>c>x_3>x_4>0.
\end{aligned} \tag{1.12}$$

[2] The differential forms of this type are as in (1.11) but with $b=a^*=\alpha-i\beta$, $\alpha,\beta\in\mathbb{R}$. The ranges of variation of the variables x_i are

$$\begin{aligned}
&x_1,x_2,x_3>c>x_4>1>0;\\
&x_1,x_2>c>x_3>1>x_4>0.
\end{aligned} \tag{1.13}$$

[3] In this case the quantities e_i can be taken to be

$$e_1=\infty,\quad e_2^*=e_3=\gamma+i\delta,$$

$$e_4=e_5=\alpha+i\beta,\quad e_6=0,\quad \alpha,\beta,\gamma,\delta\in\mathbb{R}.$$

The differential form is given as in (1.11) with

$$h(x)=[(x-\gamma)^2+\delta^2][(x-\alpha)^2+\beta^2]x.$$

The ranges of variation of the variables x_i are then

$$x_1,x_2,x_3>0>x_4;\quad x_1>0>x_2,x_3,x_4. \tag{1.14}$$

The simplest types of degenerate differential forms corresponding to cyclides of general type (1.4) are obtained by allowing pairs of the quantities e_i to become equal. This is achieved by the prescription given by Bôcher,[6] e.g., if e_1 and e_2 become equal then they do so according to the prescription

$$e_1=e_2+\epsilon,\quad x_1=e_2+\epsilon x_1' \tag{1.15}$$

where ϵ is a first order quantity. With this substitution and the subsequent use of the relations (1.10) to take

$e_2=\infty$, the differential form becomes

$$ds^2=\left(-\frac{(y_1^2+y_2^2)}{4w^4}\right)\times\left[\frac{dx_1'^2}{x_1'(x_1'-1)}-\sum_{i=2}^{4}\frac{(x_i-x_j)(x_i-x_k)}{h(x_i)}dx_i^2\right], \quad (1.16)$$

where $h(x)=(x-a)(x-b)(x-c)(x-d)$. If we make the same substitution in (1.6) relating the coordinates y_i^2, we obtain

$$\begin{aligned}
y_1^2&=1-x_1', \quad y_2^2=x_1'\\
y_3^2&=\frac{(x_2-e_3)(x_3-e_3)(x_4-e_3)}{(e_3-e_4)(e_3-e_5)(e_3-e_6)},\\
y_4^2&=\frac{(x_2-e_4)(x_3-e_4)(x_4-e_4)}{(e_4-e_3)(e_4-e_5)(e_4-e_6)},\\
y_5^2&=\frac{(x_2-e_5)(x_3-e_5)(x_4-e_5)}{(e_5-e_3)(e_5-e_4)(e_5-e_6)},\\
y_6^2&=\frac{(x_2-e_6)(x_3-e_6)(x_4-e_6)}{(e_6-e_3)(e_6-e_4)(e_6-e_5)},
\end{aligned} \quad (1.17)$$

In addition we note that the coordinate surface for the coordinate x_1' has the equation

$$y_1^2/(x_1'-1)+y_2^2/x_1'=0. \quad (1.18)$$

From the form of the coordinates in (1.6) we see that the real linear transformations which preserve the quantity $\sum_{i=1}^{6}y_i^2$ form a group isomorphic to O(4, 2). In fact the representation of a point in space–time by the six coordinates is such that the generators $L_{ij}=y_i\partial_{y_j}-y_j\partial_{y_i}$ are directly related to the canonical generators of the conformal symmetry group of the wave equation.[3] More specifically we have the relations

$$\begin{aligned}
&L_{12}=\tfrac{1}{2}(K_0-P_0), \quad L_{13}=(i/2)(K_1-P_1), \quad L_{14}=(i/2)(K_2-P_2),\\
&L_{15}=(i/2)(K_3-P_3), \quad L_{16}=iD, \quad L_{23}=iN_1, \quad L_{24}=iN_2,\\
&L_{25}=iN_3, \quad L_{26}=(i/2)(P_0+K_0), \quad L_{34}=M_3, \quad L_{35}=M_2,\\
&L_{36}=-\tfrac{1}{2}(P_1+K_1), \quad L_{45}=M_1, \quad L_{46}=-\tfrac{1}{2}(P_2+K_2),\\
&L_{56}=-\tfrac{1}{2}(P_3+K_3).
\end{aligned} \quad (1.19)$$

Here we have used the notation of Refs. 3 and 4 for the generators of the conformal symmetry group.

Taking note of these relations, we see that coordinate systems of the type given by (1.17) correspond to the diagonalization of the generator $L_{12}=y_1\partial_{y_2}-y_2\partial_{y_1}$. This generator may correspond to a rotation or a hyperbolic rotation in pentaspherical space. If a hyperbolic rotation, we may always use an O(4, 2) group motion to ensure that $L_{12}=D$. The resulting coordinate system in space–time is then equivalent to one of the radial coordinate systems discussed in reference 4. Accordingly in classifying differential forms of type (1.16) we need only consider those for which $0<x_1'<1$.

[4] If we choose $a>b>c=1>d=0$ then we have the possibilities

$$\begin{aligned}
&a>x_2>b>x_3>1>x_4>0,\\
&x_2>a>x_3, x_4>b; \quad x_2>a, \quad 1>x_3, x_4>0;\\
&x_2>a>x_3>b>1>x_4>0; \quad b>x_2>1>x_3, x_4>0;\\
&x_2, x_3, x_4>a; \quad b>x_2, x_3, x_4>1; \quad 0>x_2, x_3, x_4,\\
&x_2, x_3>a, \quad b>x_4>1; \quad x_2>a;\\
&b>x_3, x_4>1, \quad 0>x_3, x_4, \quad b>x_3>1>0>x_4;\\
&b>x_2, x_3>1>0>x_4; \quad b>x_2>1>0>x_3, x_4;\\
&a>x_2, x_3>b; \quad b>x_4>1, \quad 0>x_4;\\
&a>x_2>b>1>x_3>0>x_4.
\end{aligned} \quad (1.20)$$

[5] If $a=b^*=\alpha+i\beta$, $\alpha, \beta\in\mathbf{R}$ and $c=1$, $d=0$, then we have the possibilities

$$\begin{aligned}
&x_2, x_3, x_4>1; \quad x_2, x_3>1>0>x_4;\\
&x_2>1>x_3, x_4>0.
\end{aligned} \quad (1.21)$$

[6] If we have $a=b^*$ as above and $c=d^*=\gamma+i\delta$, $\gamma, \delta\in\mathbf{R}$ then the variables x_2, x_3, x_4 can be any real numbers.

If in addition we allow e_3 and e_4 to become equal according to the prescription of Bôcher,[6]

$$e_3=e_4+\epsilon, \quad x_2=e_4+\epsilon x_2'. \quad (1.22)$$

The differential form is then

$$ds^2=\left(\frac{-(y_1^2+y_2^2)}{4w^4}\right)\left[\frac{dx_1'^2}{x_1'(x_1'-1)}+\frac{(e_4-x_3)(e_4-x_4)}{(e_4-e_5)(e_4-e_6)}\frac{dx_2'^2}{x_2'(1-x_2')}+(x_4-x_3)\left(\frac{dx_3^2}{P(x_3)}-\frac{dx_4^2}{P(x_4)}\right)\right], \quad (1.23)$$

where $P(x)=(x-e_4)(x-e_5)(x-e_6)$. For all such differential forms $0<x_2'<1$. Differential forms of this type fall into classes in which the quantities e_4, e_5, e_6 can be chosen as 0, 1, or a.

[7] $e_4=0$, $e_5=1$, $e_6=a$; $a>1$.

The variables x_3, x_4 vary in the ranges:

$$0<x_3<1<x_4<a; \quad 1<x_3<a<x_4; \quad x_3<0<1<x_4<a. \quad (1.24)$$

[8] $e_4=1$, $e_5=0$, $e_6=a$; $a>1$.

$$1<x_3<a<x_4; \quad x_3<0<x_4<1; \quad x_3<0<1<x_4<a;$$
$$0<x_3<1<a<x_4. \quad (1.25)$$

Now by the usual prescription, the differential form becomes, with $e_4=1$ and $e_6=0$,

$$ds=\left(\frac{-(y_1^2+y_2^2)}{4w^4}\right)\left[\frac{dx_1'^2}{x_1'(x_1'-1)}+(1-x_4)\frac{dx_2'^2}{x_2'(1-x_2')}+x_4\frac{dx_3'^2}{x_3'(1-x_3')}+\frac{dx_4^2}{x_4(1-x_4)}\right]. \quad (1.26)$$

There is only one differential form of this type.

[9] For this case all the variables x_i' $(i=1, 2, 3)$, x_4 lie in the interval $[0, 1]$.

A further class of differential forms can be obtained by taking

$$e_4=e_6+a\epsilon, \quad e_5=e_6+\epsilon, \quad x_i=e_6+\epsilon x_i', \quad i=3, 4. \quad (1.27)$$

If we also put $e_6=\infty$ in the resulting differential form we obtain

$$ds=\left(\frac{-(y_4^2+y_5^2+y_6^2)}{4w^2}\right)\left[(x_2-x_1)\left(\frac{dx_1^2}{P(x_1)}-\frac{dx_2^2}{P(x_2)}\right)+(x_3'-x_4')\left(\frac{dx_3'^2}{Q(x_3')}-\frac{dx_4'^2}{Q(x_4')}\right)\right], \quad (1.28)$$

where $P(x)=(x-e_1)(x-e_2)(x-e_3)$ and $Q(x)=(x-a)\times(x-1)x$. This differential form corresponds to the reductions $O(4,2)\supset O(3)\otimes O(2,1)$ and $O(4,2)\supset O(2,1)\otimes O(2,1)$ when expressed in elliptic coordinates in the case of the two reductions $O(3)\supset L$ and $O(2,1)\supset L'$. Here L and L' are second order symmetric operators in the enveloping algebras of O(3) and O(2, 1), respectively.

With the exception of the reduction $O(2,1)\supset O(1,1)$, which can be conformally transformed into a radial system, we can in principle write down all the differential forms corresponding to the reductions of type $O(4,2)\supset O(3)\otimes O(2,1)$ and $O(4,2)\supset O(2,1)\otimes O(2,1)$ by considering degenerate forms of the differential form (1.28), but we do not do this here.

The remaining distinct type of differential form of interest in this section is obtained by taking $x_2=e_6+\epsilon' x_2'$ and $e_3=e_6+\epsilon'$ subsequent to the substitutions (1.27) and then allowing $e_6\to\infty$. We then obtain the differential form

$$ds^2=\left(\frac{(y_3^2+y_4^2+y_5^2+y_6^2)}{4w^4}\right)\left[\frac{dx_1^2}{x_1(1-x_1)}+\frac{dx_2'^2}{x_2'(x_2'-1)}+x_2'(x_3'-x_4')\left(\frac{dx_3'^2}{Q(x_3')}-\frac{dx_4'^2}{Q(x_4')}\right)\right]. \tag{1.29}$$

[10] In each class we have $0<x_1<1$, $0<x_2'<1$. The remaining variables vary in the ranges

$$0<x_3'<1<x_4'<a;\quad 1<x_3'<a<x_4';$$
$$x_3'<0<1<a<x_4';\quad x_3'<0<x_4'<1.$$

[11] A further dfiferential form can be obtained from taking the limits $a=1+\epsilon''$, $x_3'=1+\epsilon'' x_3''$. This gives one new differential form

$$ds^2=\left(\frac{-(y_1^2+y_2^2+y_3^2+y_4^2)}{4w^4}\right)\left[\frac{dx_1^2}{x_1(1-x_1)}+\frac{dx_2'^2}{x_2'(x_2'-1)}+x_2'\left((1-x_4')\frac{dx_3''^2}{x_3''(x_3''-1)}+\frac{dx_4'^2}{x_4'(x_4'-1)}\right)\right] \tag{1.30}$$

where all the variables lie between 0 and 1.

We have shown in this section how to get orthogonal coordinate systems by various limiting procedures applied to coordinate systems of the most general cyclidic type. We have as yet not fully understood in what sense these procedures are complete.

II. *R*-SEPARABLE COORDINATES FOR THE WAVE EQUATION

In this section we give the coordinate systems corresponding to the differential forms in section I together with the separation equations. We also present the triplet L_1, L_2, L_3 of mutually commuting second order symmetric operators in the enveloping algebra of O(4, 2) whose eigenvalues are the separation constants for each coordinate system presented. We tabulate the coordinate systems of interest starting with the most general real cyclidic type.

Coordinate systems of Class I

(1)—(5): (a) A suitable choice of coordinates is

$$t=\frac{1}{R}\left[-\frac{(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-1)a}\right]^{1/2},$$
$$x=\frac{1}{R}\left[\frac{(x_1-b)(x_2-b)(x_3-b)(x_4-b)}{(b-a)(b-c)(b-1)b}\right]^{1/2},$$
$$y=\frac{1}{R}\left[\frac{(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-1)c}\right]^{1/2}, \tag{2.1}$$
$$z=\frac{1}{R}\left[\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(1-a)(1-b)(1-c)}\right]^{1/2},$$
$$\text{where } R=\left(1+\left[\frac{x_1x_2x_3x_4}{abc}\right]^{1/2}\right).$$

The solution of the wave equation then assumes the form $\psi=R\Phi$, where $\Phi=\prod_{i=1}^4 E_i(x_i)$ typically. The separation equations for the functions E_i are

$$\frac{d^2E_j}{dx_j^2}+\frac{1}{2}\left(\frac{1}{x_j-a}+\frac{1}{x_j-b}+\frac{1}{x_j-c}+\frac{1}{x_j-1}+\frac{1}{x_j}\right)\frac{dE_j}{dx_j}$$
$$+\frac{(-2x_j^3+l_1x_j^2+l_2x_j+l_3)}{4(x_j-a)(x_j-b)(x_j-c)(x_j-1)x_j}E_j=0. \tag{2.2}$$

The operators L_i whose eigenvalues l_i are the separation constants are

$$\begin{aligned}L_1&=\tfrac{1}{4}(a+b+c)(P_3+K_3)^2+\tfrac{1}{4}(a+b+1)(P_2+K_2)^2\\&\quad+\tfrac{1}{4}(a+c+1)(P_1+K_1)^2-\tfrac{1}{4}(b+c+1)(P_0+K_0)^2\\&\quad+(a+b)M_1^2+(a+c)M_2^2-(b+c)N_3^2\\&\quad-(c+1)N_1^2-(b+1)N_2^2+(a+1)M_3^2,\\L_2&=\tfrac{1}{4}(ac+bc+ab)(P_3+K_3)^2\\&\quad+(ab+a+b)(P_2+K_2)^2+\tfrac{1}{4}(ac+a+c)(P_1+K_1)^2\\&\quad-\tfrac{1}{4}(bc+b+c)(P_0+K_0)^2+abM_1^2\\&\quad+acM_2^2-bcN_3^2-cN_1^2-bN_2^2+aM_3^2,\\L_3&=-\tfrac{1}{4}abc(P_3+K_3)^2-\tfrac{1}{4}ab(P_2+K_2)^2\\&\quad-\tfrac{1}{4}ac(P_1+K_1)^2+\tfrac{1}{4}bc(P_0+K_0)^2.\end{aligned} \tag{2.3}$$

The coordinates x_i vary in the ranges

$$x_1>a>b>x_2>c>x_3>1>x_4>0.$$

There are four more coordinate systems of this type. We list below the complex transformation of the space time coordinates which relates the coordinates of type (a) to the new system, together with the new ranges of variation of the coordinates x_i. The separation equations for the $E_j(x_j)$ are the same in each case and the basis defining operators can be obtained by the substitution given. We now list the possibilities:

(b) $(t,x,y,z)\to(iz,x,y,it)$,

$x_1, x_2>a>x_3>b>x_4>c$,

(c) $(t,x,y,z)\to(x,t,iy,iz)$,

$x_1>a>x_2, x_3>b>c>x_4>1$; $x_1, x_2, x_3>a>b>c>x_4>1$.

(d) $(t,x,y,z)\to(it,ix,iy,iz)$,

$x_1, x_2>a>b>x_3>c>x_4>1$.

(e) $(t,x,y,z)\to(t,ix,y,iz)$,

$x_1, x_2, x_3>a>x_4>b$.

(6)—(7) A suitable choice of coordinates is

$$x+it=\frac{1}{R}\left[\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-1)a}\right]^{1/2},$$
$$y=\frac{1}{R}\left[\frac{(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-1)c}\right]^{1/2},$$
$$z=\frac{1}{R}\left[\frac{(x_1-1)(x_2-1)(x_3-1)(x_4-1)}{(1-a)(1-b)(1-c)}\right]^{1/2}, \tag{2.4}$$

where $R=[1+(x_1x_2x_3x_4/abc)^{1/2}]$ and $a=b^*=\alpha+i\beta$, $\alpha,\beta\in\mathbf{R}$.

The solution of the wave equation has the form $\psi=R\Phi$, where each of the E_j satisfy Eq. (2.2). The operators whose eigenvalues are the separation constants are

$$\begin{aligned}L_1=&\tfrac{1}{4}(2\alpha+c)(P_3+K_3)^2+\tfrac{1}{4}(2\alpha+1)(P_2+K_2)^2\\&+2\alpha M_1^2+\tfrac{1}{4}(\alpha+c+1)[(P_1+K_1)^2-(P_0+K_0)^2]\\&-(\beta/4)\{(P_0+K_0,P_1+K_1\}+(\alpha+c)(M_2^2-N_3^2)\\&+\beta\{N_3,M_3\}+(\alpha+1)(M_3^2-N_2^2)\\&+\beta\{N_2,M_3\}-(c+1)N_2^2,\\L_2=&-\tfrac{1}{4}(2\alpha c+\alpha^2+\beta^2)(P_3+K_3)^2\\&-\tfrac{1}{4}(2\alpha+\alpha^2+\beta^2)(P_2+K_2)^2\\&-(\alpha^2+\beta^2)M_1^2+\tfrac{1}{4}(\alpha c+\alpha+c)[(P_0+K_0)^2\\&-(P_1+K_1)^2]+\tfrac{1}{4}\beta(c+1)\{P_1+K_1,P_0+K_0\}\\&+\alpha c(N_3^2-M_2^2)-c\beta\{M_2,N_3\}+cN_1^2\\&+\alpha(N_2^2-M_3^2)-\beta\{M_3,N_2\},\end{aligned} \tag{2.5}$$
$$\begin{aligned}L_3=&-\tfrac{1}{4}(\alpha^2+\beta^2)[c(P_3+K_3)^2+(P_2+K_2)^2]\\&+(\alpha c/4)[(P_0+K_0)^2-(P_1+K_1)^2]\\&-(c\beta/4)\{P_1+K_1,P_0+K_0\},\end{aligned}$$

where $\{A,B\}=AB+BA$.

The coordinates x_i can vary in the ranges

(a) $x_1,x_2>c>x_3>1>x_4>0$.

(b) $(t,x,y,z)\to(it,ix,iy,iz)$,

where $x_1,x_2,x_3>c>x_4>1>0$.

(8) A suitable choice of coordinates is

$$t+iy=\left[\frac{2(x_1-c)(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-d)c}\right]^{1/2}\Big/R,$$
$$x=\mathrm{Im}\left[\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)a}\right]^{1/2}\Big/R, \tag{2.6}$$
$$z=[-x_1x_2x_3x_4/abcd]^{1/2}/R,$$

where

$$R=1+\mathrm{Re}\left[-\frac{2(x_1-a)(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)a}\right]^{1/2}$$

and $a=b^*=\alpha+i\beta$, $c=d^*=\gamma+i\delta$, $\alpha,\beta,\gamma,\delta\in\mathbf{R}$.

The solution of the wave equation has the form $\psi=R\Phi$, where each of the E_j satisfy the equation

$$\frac{d^2E_j}{dx_j^2}+\frac{1}{2}\left(\frac{1}{x_j-a}+\frac{1}{x_j-b}+\frac{1}{x_j-c}+\frac{1}{x_j-d}+\frac{1}{x_j}\right)\frac{dE_j}{dx_j}+\frac{(-2x_j^3+l_1x_j^2+l_2x_j+l_3)}{4(x_j-a)(x_j-b)(x_j-c)(x_j-d)x_j}E_j=0. \tag{2.7}$$

The operators whose eigenvalues are the separation constants are

$$\begin{aligned}L_1=&(2\alpha+\gamma)(M_1^2-N_3^2)+\delta\{M_1,N_3\}\\&+(2\gamma+\alpha)[M_2^2-\tfrac{1}{4}(P_3-K_3)^2]\\&+\tfrac{1}{2}\beta\{M_2,P_3-K_3\}+\tfrac{1}{2}\gamma(P_0-K_0)^2-2\alpha N_2^2\\&+(\alpha+\beta)[\tfrac{1}{4}(P_0-K_0)^2-\tfrac{1}{4}(P_2-K_2)^2+M_3^2-N_1^2]\\&+\tfrac{1}{2}\beta\{N_1,P_0-K_0\}+\tfrac{1}{4}\delta\{P_0-K_0,P_2-K_2\}\\&-\delta\{N_1,M_3\}-(\beta/2)\{M_3,P_2-K_2\},\\L_2=&(\alpha^2+\beta^2+2\alpha\gamma)(N_3^2-M_1^2)-2\alpha\delta\{M_1N_3\}\\&+(\gamma^2+\delta^2+2\alpha\gamma)[\tfrac{1}{4}(P_3-K_3)^2-M_2^2]-\gamma\beta\{M_2,P_3-K_3\}\\&+(\alpha^2+\beta^2)N_2^2+\tfrac{1}{4}(\gamma^2+\delta^2)(P_1-K_1)^2\\&+\alpha\gamma[\tfrac{1}{4}(P_2-K_2)^2-\tfrac{1}{4}(P_0-K_0)^2+N_1^2-M_3^2]\\&-(\gamma\beta/2)\{P_0-K_0,N_1\}-(\alpha\delta/4)\{P_0-K_0,P_2-K_2\}\\&-(\beta\delta/2)\{M_3,P_0-K_0\}-(\beta\delta/2)\{N_1,P_2-K_2\}\\&+\alpha\delta\{N_1,M_3\}+(\gamma\beta/2)\{P_2-K_2,M_3\},\end{aligned} \tag{2.8}$$
$$\begin{aligned}L_3=&(\alpha^2+\beta^2)[\gamma(N_3^2-M_1^2)-\delta\{N_3,M_1\}]+(\gamma^2+\delta^2)\\&\times[\alpha(\tfrac{1}{4}(P_3-K_3)^2-M_2^2)-(\beta/2)\{P_3-K_3,M_2\}].\end{aligned}$$

The variables x_i can vary in the ranges $x_1>0>x_2,x_3,x_4$ and $x_1,x_2,x_3>0>x_4$.

Coordinate systems of Class II

These are the coordinate systems in which the operator $\frac{1}{2}(P_0-K_0)$ is diagonal.

As has been discussed in Ref. 3, the R-separable solutions of (*) then have the form $\psi=(Y_0-\cos\psi)\times\exp[i(2F+1)\psi]\Phi(Y_0,Y_1,Y_2,Y_3)$, where $Y_0^2+Y_1^2+Y_2^2+Y_3^2=1$ and the space–time coordinates are given by

$$t=\frac{\sin\psi}{Y_0-\cos\psi},\quad x=\frac{Y_1}{Y_0-\cos\psi},$$
$$y=\frac{Y_2}{Y_0-\cos\psi},\quad z=\frac{Y_3}{Y_0-\cos\psi}. \tag{2.9}$$

$i(2F+1)$ is the eigenvalue of the operator $\frac{1}{2}(P_0-K_0)$, and F is a positive integer or half-integer. The function Φ satisfies the equation

$$(\Gamma_{12}^2+\Gamma_{13}^2+\Gamma_{14}^2+\Gamma_{23}^2+\Gamma_{24}^2+\Gamma_{34}^2)\Phi=-4F(F+1)\Phi, \tag{2.10}$$

where $\Gamma_{12}=-\frac{1}{2}(P_1+K_1)$, $\Gamma_{13}=-\frac{1}{2}(P_2+K_2)$, $\Gamma_{14}=-\frac{1}{2}(P_3+K_3)$, $\Gamma_{23}=M_3$, $\Gamma_{24}=-M_2$, and $\Gamma_{34}=M_1$. Here we are using the notation of Ref. 3. The problem of separation of variables for coordinate systems in which $\frac{1}{2}(P_0-K_0)$ is diagonal reduces to the problem of separation of variables on the three-dimensional sphere S_3 in 4-space. Acting on the functions Φ, the operators given above have the form

$$\begin{aligned}&\Gamma_{12}=Y_0\partial_1-Y_1\partial_0,\quad \Gamma_{13}=Y_0\partial_2-Y_2\partial_0,\\&\Gamma_{14}=Y_0\partial_3-Y_3\partial_0,\quad \Gamma_{23}=Y_1\partial_2-Y_2\partial_1,\\&\Gamma_{24}=Y_1\partial_3-Y_3\partial_1,\quad \Gamma_{34}=Y_2\partial_3-Y_3\partial_2.\end{aligned} \tag{2.11}$$

This problem has been solved by Olevski[7] and the six coordinate systems on S_3 for which (2.10) admits separation of variables have recently been investigated.[8]

In the interest of completeness we give here the six coordinate systems, the separation equations, the operators describing the separation, and some comment on the actual solutions.

(9) *Ellipsoidal coordinates*: A suitable choice of coordinates is

$$Y_0^2=-\frac{(x_1-a)(x_2-a)(x_3-a)}{(b-a)(1-a)a},$$
$$Y_1^2=-\frac{(x_1-b)(x_2-b)(x_3-b)}{(a-b)(1-b)b},$$
$$Y_2^2=-\frac{(x_1-1)(x_2-1)(x_3-1)}{(a-1)(b-1)}, \tag{2.12}$$
$$Y_3^2=\frac{x_1x_2x_3}{ab},$$

where $0<x_3<1<x_2<b<x_1<a$. The separation equations for $\Phi=E_1(x_1)E_2(x_2)E_3(x_3)$ have the form

$$\frac{dE_i}{dx_i}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{1}{x_i-b}+\frac{1}{x_i-1}+\frac{1}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{4(x_i-a)(x_i-b)(x_i-1)x_i}E_i=0. \tag{2.13}$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}b(P_2+K_2)^2+\tfrac{1}{4}(b+1)(P_3+K_3)^2$$
$$+aM_3^2+(a+1)M_2^2-(a+b)M_1^2,$$
$$L_2=\tfrac{1}{4}b(P_3+K_3)^2-aM_2^2-abM_1^2. \tag{2.14}$$

(10) *Elliptic cylindrical coordinates of Type* I: A suitable choice of coordinates is

$$Y_0=\sqrt{x_1x_2/a}\cos\phi,\quad Y_1=\sqrt{x_1x_2/a}\sin\phi,$$
$$Y_2=\sqrt{(x_1-a)(x_2-a)/a(a-1)}, \tag{2.15}$$
$$Y_3=\sqrt{(x_1-1)(x_2-1)/(1-a)},$$

where $0<x_1<1<x_2<a$.

The separation equations have the form for $\Phi=E_1(x_1)\times E_2(x_2)A(\phi)$:

$$\frac{d^2E_i}{dx_i^2}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{1}{x_i-1}+\frac{2}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{(x_i-a)(x_i-1)x_i^2}E_i=0 \tag{2.16}$$

where $i=1,2$,

$$a\frac{d^2A}{d\phi^2}+l_2A=0.$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$L_1=M_3^2+\tfrac{1}{4}(P_2+K_2)^2+a[M_2^2+\tfrac{1}{4}(P_3+K_3)^2]$$
$$+\tfrac{1}{4}(a+1)(P_1+K_1)^2, \tag{2.17}$$
$$L_2=-\tfrac{1}{4}a(P_1+K_1)^2;$$

an alternative choice of coordinates is obtained by taking $x_1=\mathrm{sn}^2(\rho_1,k)$ and $x_2=(1/k^2)\,\mathrm{dn}^2(\rho_2,k')$ where $a=1/k^2$.

We then have that

$$y_0=\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2\cos\phi,$$
$$y_1=\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2\sin\phi,$$
$$y_2=\mathrm{dn}\rho_1\,\mathrm{sn}\rho_2, \tag{2.18}$$
$$y_3=\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2,$$

where $0\leq\rho_1<2K$ and $-K'<\rho_2<K'$. [Note: $\mathrm{sn}(z,k)$ is a Jacobi elliptic function.] In terms of these coordinates the solution for Φ has the form

$$\Phi=(\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2)^m K_{Fn}^{Ps}(\mathrm{dn}\rho_2)$$
$$\times K_{Fn}^{Ps}(k\,\mathrm{sn}\rho_1)\begin{bmatrix}\cos m\phi\\ \sin m\phi\end{bmatrix} \tag{2.19}$$

Here $K_{Fn}^{Ps}(z)$ is an associated Lamé polynomial as defined in Ref. 8.

(11) *Elliptic cylindrical coordinates of Type II*: A suitable choice of coordinates is

$$Y_0=\sqrt{(x_1-1)(x_2-1)/(1-a)}\cos\phi,$$
$$Y_1=\sqrt{(x_1-1)(x_2-1)/(1-a)}\sin\phi,$$
$$Y_2=\sqrt{x_1x_2/a}, \tag{2.20}$$
$$Y_3=\sqrt{(x_1-a)(x_2-a)/a(a-1)},$$

where $0<x_1<1<x_2<a$.

The separation equations have the form $\Phi=E_1(x_1)\times E_2(x_2)A(\phi)$:

$$\frac{d^2E_i}{dx_i^2}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{2}{x_i-1}+\frac{1}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{4(x_i-a)(x_i-1)^2x_i}E_i=0 \tag{2.21}$$

where $i=1,2$,

$$(a-1)\frac{d^2A}{d\phi^2}+l_2A=0.$$

The operators whose eigenvalues are the separation constants l_1 and l_2 are

$$L_1=M_1^2+\tfrac{1}{4}(a-1)(P_1+K_1)^2+a(M_3^2+\tfrac{1}{4}(P_2+K_2)^2),$$
$$L_2=\tfrac{1}{4}(1-a)(P_1+K_1)^2. \tag{2.22}$$

These coordinates can also be written in terms of Jacobi elliptic functions by the same substitution as used for system 10. We then obtain

$$Y_0=\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2\cos\phi,$$
$$Y_1=\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2\sin\phi,$$
$$Y_3=\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2, \tag{2.23}$$
$$Y_4=\mathrm{dn}\rho_1\,\mathrm{sn}\rho_2.$$

In terms of these coordinates the solution for Φ has the form

$$\Phi=(\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2)^m K_{Fn}^{Ps}(-(ik'/k)\,\mathrm{cn}\rho_2)$$
$$\times K_{Fn}^{Ps}(\mathrm{cn}\rho_1)\begin{bmatrix}\cos m\phi\\ \sin m\phi\end{bmatrix} \tag{2.24}$$

(12) *Spheroelliptic coordinates*: A suitable choice of

coordinates is

$$Y_0=\sin\alpha\sqrt{x_1x_2}/a,\quad Y_1=\sin\alpha\sqrt{(x_1-1)(x_2-1)/(1-a)},$$
$$Y_2=\sin\alpha\sqrt{(x_1-a)(x_2-a)/a(a-1)},\quad Y_3=\cos\alpha, \tag{2.25}$$

where $0<x_1<1<x_2<\alpha$, $0<\alpha<\pi$.

The coordinate system can also be written in terms of elliptic functions as with coordinate systems 10 and 11. This gives the parametrization:

$$Y_0=\sin\alpha\,\mathrm{sn}\rho_1\,\mathrm{dn}\rho_2,\quad Y_1=\sin\alpha\,\mathrm{cn}\rho_1\,\mathrm{cn}\rho_2,$$
$$Y_2=\sin\alpha\,\mathrm{dn}\rho_1\,\mathrm{sn}\rho_2,\quad Y_3=\cos\alpha. \tag{2.26}$$

A typical solution for Φ is of the form $A(\alpha)E_1(\rho_1)E_2(\rho_2)$, where

$$E_1(\rho_1)E_2(\rho_2)=F_{ln}^{pq}(-i\rho_1+iK+K',\rho_2) \tag{2.27}$$

a product of Lamé polynomials defined in Ref. 7 and

$$A(\alpha)=(\sin\alpha)^lC_{2F-l}^{l+1}(\cos\alpha).$$

[Here $C_m^v(z)$ is a Gegenbauer polynomial.] The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}(P_2+K_2)^2+M_3^2,$$
$$L_2=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}a(P_2+K_2)^2 \tag{2.28}$$

with eigenvalues $-l(l+1)$ and λ_n^{pq}, respectively.

(13) *Spherical coordinates*: A suitable choice of coordinates is

$$Y_0=\sin\alpha\sin\beta\cos\phi,\quad Y_1=\sin\alpha\sin\beta\sin\phi,$$
$$Y_2=\sin\alpha\cos\beta,\qquad Y_3=\cos\alpha, \tag{2.29}$$

where $0\leq\alpha,\ \beta\leq\pi,\ 0\leq\phi<2\pi$.

A typical solution of the form $A(\alpha)B(\beta)C(\phi)$ is

$$\Phi=(\sin\alpha)^l\,C_{2F-l}^{l+1}(\cos\alpha)P_l^m(\cos\beta)\exp(im\phi). \tag{2.30}$$

The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2+\tfrac{1}{4}(P_2+K_2)^2+M_3^2,$$
$$L_2=\tfrac{1}{4}(P_2+K_1)^2 \tag{2.31}$$

with eigenvalues $-l(l+1)$ and $-m^2$, respectively.

(14) *Cylindrical coordinates*: A suitable choice of coordinates is

$$Y_0=\sin\alpha\cos\beta,\quad Y_2=\sin\alpha\sin\beta,$$
$$Y_3=\cos\alpha\cos\phi,\quad Y_3=\cos\alpha\sin\phi, \tag{2.32}$$

where $0<\alpha<\pi$ and $0<\beta,\ \phi<2\pi$.

A typical solution $A(\alpha)B(\beta)C(\phi)$ is

$$\Phi=\exp[im\phi+ip\beta](\sin\alpha)^{a+b}(\cos\alpha)^{2F-a-b}$$
$$\times{}_2F_1(b-F,a-F,a+b+1;-\tan^2\alpha), \tag{2.33}$$

where $m=a+b$, $p=a-b$. The two operators characterizing this system are

$$L_1=\tfrac{1}{4}(P_1+K_1)^2\ \text{ and }\ L_2=M_1^2 \tag{2.34}$$

with eigenvalues $-p^2$ and $-m^2$, respectively.

Coordinate systems of Class III

These are the analogs of the elliptical coordinates of type 9. The difference is that coordinate systems of this type correspond to the diagonalization of M_3^2 rather than $\frac{1}{4}(P_0-K_0)^2$. We now list the possibilities.

(15) (a) A suitable choice of coordinates is

$$t=(1/R)\sqrt{(x_2-a)(x_3-a)(x_4-a)/(b-a)(a-1)a},$$
$$x=(1/R)\cos\phi,\quad y=(1/R)\sin\phi, \tag{2.35}$$
$$z=(1/R)\sqrt{(x_2-b)(x_3-b)(x_4-b)/(b-a)(b-1)b}$$

where

$$R=[\sqrt{(x_2-1)(x_3-1)(x_4-1)/(a-1)(b-1)}+\sqrt{x_2x_3x_4/ab}].$$

The typical solution of the wave equation is $\psi=R\Phi$, where $\theta=E_2(x_2)E_3(x_4)A(\phi)$. The separation equations are the same as for system 9 with $A(\phi)=\exp[i(2F+1)\phi]$. The variables x_2, x_3, x_4 vary in the ranges

$$x_2,x_3>a>b>x_4>1;$$
$$b>x_2>1>x_3,x_4>0;\quad b>x_2,x_3,x_4>1;$$
$$b>x_2>1>0>x_3,x_4;\quad a>x_2,x_3>b>x_4>1.$$

The operators whose eigenvalues are the separation constants are

$$L_1=(a+b)D^2-\tfrac{1}{4}(a+1)(P_3-K_3)^2$$
$$+\tfrac{1}{4}(b+1)(P_0-K_0)^2+\tfrac{1}{4}a(P_3+K_3)^2$$
$$-\tfrac{1}{4}b(P_0+K_0)^2-N_3^2, \tag{2.36}$$
$$L_2=abD^2+\tfrac{1}{4}a(P_3-K_3)^2+\tfrac{1}{4}b(P_0-K_0)^2,$$

and, of course, $L_3=M_3^2$.

There are five further coordinate systems of this type. In each case we choose the x and y coordinates to be of the form

$x=(1/R)\cos\phi$, $y=(1/R)\sin\phi$, and the operator $L_3=M_3^2$.

The separation equations are the same as in system 9. For each of these five systems we give the choice of R and the coordinates t and z together with the form of the operators L_1 and L_2.

(16) (b) The modulation function R is

$$R=\Big[\sqrt{(x_2-1)(x_3-1)(x_4-1)/(a-1)(b-1)}$$
$$+\sqrt{(x_2-b)(x_3-b)(x_4-b)/(a-b)(b-1)b}\Big] \tag{2.37}$$

and the coordinates t and z are given by

$$t=(1/R)\sqrt{x_2x_3x_4/ab},$$
$$z=(1/R)\sqrt{(x_2-a)(x_3-a)(x_4-a)/(a-b)(a-1)a}. \tag{2.38}$$

The operators L_1 and L_2 are

$$L_1=\tfrac{1}{4}(a+b)(P_0+K_0)^2-\tfrac{1}{4}(a+1)(P_0-K_0)^2$$
$$+(b+1)N_3^2+aD^2-\tfrac{1}{4}b(P_3+K_3)^2+\tfrac{1}{4}(P_3-K_3)^2$$
$$L_2=-\tfrac{1}{4}ab(P_0+K_0)^2+\tfrac{1}{4}a(P_0-K_0)^2-bN_3^2. \tag{2.39}$$

The ranges of variation of the coordinates x_2, x_3, and

x_4 are

$$x_2 > a > x_3, x_4 > b; \quad x_2 > a > b > x_3, x_4 > 1;$$
$$x_2, x_3, x_4 > a; \quad b > x_2, x_3, x_4 > 1;$$
$$a > x_2, x_3 > b > x_4 > 1; \quad \text{and } x_2 > a > b > 1 > 0 > x_3, x_4.$$

(17) (c) This coordinate system is related to 16(b) via the transformation $(t, x, y, z) \to (it, ix, iy, iz)$ of the space–time coordinates. The variables x_2, x_3, x_4 vary in the ranges

$$x_2 > a > x_3 > b > 1 > x_4 > 0 \text{ and } x_2 > a > b > 1 > x_3, x_4 > 0.$$

(18) (d) This coordinate system is related to 16(b) via the transformation $(t, x, y, z) \to (z, it, iy, t)$ of the space–time coordinates. The variables x_2, x_3, and x_4 vary in the ranges $x_2, x_3 > a > b > 1 > 0 > x_4$; $b > x_2, x_3 > 1 > 0 > x_4$, and $a > x_2, x_3 > b > 1 > 0 > x_4$.

(19) (e) This coordinate system is related to 15(a) via the transformation $(t, x, y, z) \to (z, ix, iy, t)$ of the space–time coordinates. The variables x_2, x_3, and x_4 vary in the ranges $x_2 > a > b > x_3, x_4 > 1$.

(20) (f) This coordinate system is related to 16(b) via the transformation $(t, x, y, z) \to (iz, x, y, it)$ of the space–time coordinates. The variables x_2, x_3 and x_4 vary in the ranges $a > x_2 > b > 1 > x_3 > 0 > x_4$.

In addition to the six types of coordinate systems we have discussed in Class III we will also include coordinate systems corresponding to the differential form of type (1.16).

(21) A suitable choice of coordinates is

$$(z+it) = \frac{1}{R}\left[\frac{2(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-1)a}\right]^{1/2}, \qquad x = \frac{1}{R}\cos\phi, \quad y = \frac{1}{R}\sin\phi, \tag{2.40}$$

where

$$R = [\sqrt{(x_2-1)(x_3-1)(x_4-1)/(a-1)(b-1)} + \sqrt{x_2x_3x_4/ab}]. \tag{2.41}$$

The separation equations are given by (2.13). The operators whose eigenvalues are l_1 and l_2 are

$$L_1 = 2\alpha D^2 + \tfrac{1}{4}(\alpha+1)[(P_3-K_3)^2 - (P_0-K_0)^2] - \tfrac{1}{2}\beta(P_0P_3 + K_0K_3) + \tfrac{1}{4}\alpha[(P_3+K_3)^2 - (P_0+K_0)^2] - N_3^2, \tag{2.42}$$

$$L_2 = (\alpha^2+\beta^2)D^2 + \tfrac{1}{4}\alpha[(P_3-K_3)^2 - (P_0-K_0)^2] + \tfrac{1}{4}\beta\{P_3-K_3, P_0-K_0\}.$$

The variables x_2, x_3 and x_4 vary in the ranges

$$x_2, x_3, x_4 > 1; \quad x_2 > 1 > x_3, x_4 > 0;$$
$$x_2 > 1 > 0 > x_3, x_4.$$

(22) Coordinate systems of this type can be obtained from those of type 21 via the transformation $(t, x, y, z) \to (it, ix, iy, iz)$. The variables x_2, x_3, and x_4 lie in the ranges $x_2, x_3 > 1 > 0 > x_4$; $0 > x_2, x_3, x_4$; and $1 > x_2, x_3 > 0 > x_4$.

(23) A suitable choice of coordinates is

$$(z+it) = \frac{1}{R}\left[\frac{2(x_2-a)(x_3-a)(x_4-a)}{(a-b)(a-c)(a-d)}\right]^{1/2}, \qquad x = \frac{1}{R}\cos\phi, \quad y = \frac{1}{R}\sin\phi, \tag{2.43}$$

where

$$R = \mathrm{Re}\omega - \mathrm{Im}\omega,$$
$$\omega = \left[\frac{2(x_2-c)(x_3-c)(x_4-c)}{(c-a)(c-b)(c-d)}\right]^{1/2}.$$

The separation equations in the variables x_2, x_3, and x_4 are

$$\frac{d^2E_i}{dx_i^2} + \frac{1}{2}\left[\frac{1}{x_i-a} + \frac{1}{x_i-b} + \frac{1}{x_i-c} + \frac{1}{x_i-d}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{4(x_i-a)(x_i-b)(x_i-c)(x_i-d)}E_i = 0. \tag{2.44}$$

The operators whose eigenvalues are l_1 and l_2 are

$$L_1 = -2\alpha D^2 - 2\gamma N_3^2 + \tfrac{1}{2}(\alpha+\gamma)[\{P_3, K_3\} - \{P_0, K_0\}] + \tfrac{1}{2}\delta[P_0^2 - P_3^2 + K_3^2 - K_0^2] - \tfrac{1}{2}\beta\{P_0, K_3\} + \{P_3, K_0\}, \tag{2.45}$$

$$L_2 = (\alpha^2+\beta^2)D^2 + (\gamma^2+\delta^2)N_3^2 + \tfrac{1}{2}\alpha\gamma[\{P_3, K_3\} - \{P_0, K_0\}] + \tfrac{1}{2}\alpha\delta[P_0^2 - P_3^2 + K_3^2 - K_0^2] + \beta\delta(P_0P_3 - K_0K_3) - \tfrac{1}{2}\beta\gamma[\{P_3, K_0\} + \{P_0, K_3\}];$$

the variables x_2, x_3, and x_4 can assume any real values.

(24) A suitable choice of coordinates is

$$t+z = \frac{2}{R}\,\mathrm{Im}\left[\frac{(x_1-a)(x_2-a)(x_3-a)}{(a-b)^2}\right]^{1/2},$$
$$t-z = \frac{1}{R}\,\mathrm{Im}\left[\frac{1}{(a-b)} - \frac{1}{2}\left\{\frac{1}{x_1-a} + \frac{1}{x_2-a} + \frac{1}{x_3-a}\right\}\right],$$
$$x = \frac{1}{R}\cos\phi, \quad y = \frac{1}{R}\sin\phi,$$
$$\text{where } R = 2\,\mathrm{Re}\left[\frac{(x_1-a)(x_2-a)(x_3-a)}{(a-b)^2}\right]^{1/2}. \tag{2.46}$$

The separation equations in the variables x_2, x_3, and x_4 are

$$\frac{d^2E_i}{dx_i} + \left[\frac{1}{x_i-a} + \frac{1}{x_i-b}\right]\frac{dE_i}{dx_i} + \frac{[4F(F+1)x_i^2 + l_1x_i + l_2]}{4(x_i-a)^2(x_i-b)^2}E_i = 0. \tag{2.47}$$

The operators whose eigenvalues are l_1 and l_2 are

$$L_1 = \alpha[\tfrac{1}{4}(P_3 - P_0 - K_3 - K_0)^2 - (D+N_1)^2] + \tfrac{1}{2}\beta\{P_3 - P_0 - K_3 - K_0, D+N_1\} + \alpha[\tfrac{1}{4}(P_0+P_3+K_3-K_0)^2 - \tfrac{1}{4}(P_0-P_3+K_0+K_3)^2 - (N_1-D)^2 + \tfrac{1}{4}(P_0+P_3+K_0-K_3)^2] - \tfrac{1}{4}\{P_0+K_3, P_3+P_0+K_0-K_3\},$$

$$L_2 = -\tfrac{1}{4}(P_0+K_3)^2 + \tfrac{1}{2}(\alpha^2+\beta^2)[\tfrac{1}{4}(P_0+P_3+K_3-K_0)^2 - \tfrac{1}{4}(P_0-P_3+K_0+K_3)^2 - (N_1-D)^2 - \tfrac{1}{4}(P_0+P_3+K_0-K_3)^2] + \tfrac{1}{2}(\alpha^2-\beta^2) \times [\tfrac{1}{4}(P_3-P_0-K_0-K_3)^2 - (D+N_1)^2] - \tfrac{1}{4}\alpha\beta\{P_3 - P_0 - K_3 - K_0, D+N_1\} + \tfrac{1}{4}\{(P_0+K_3), \beta(D-N_1) - \tfrac{1}{2}\alpha(P_0+P_3+K_0-K_3)\}. \tag{2.48}$$

(25) This coordinate system is of similar type to coordinate systems 10 and 11 appearing in Class II. A suitable choice of coordinates is

$$t=(1/R)\sqrt{(x_1-a)(x_2-a)/a(a-1)},$$
$$x=(1/R)\cos\psi\sqrt{(x_1-1)(x_2-1)/(a-1)},$$
$$y=(1/R)\cos\phi,\quad z=(1/R)\sin\phi,$$

where (2.49)

$$R=\sqrt{(x_1-1)(x_2-1)/(a-1)}\sin\psi+\sqrt{x_1x_2/a},$$

and $x_1, x_2<0$, or $0<x_1, x_2<1$.

The solution ψ of the wave equation has the form $\psi=R\Phi$. The separation equations for $\Phi=E_1(x_1)E_2(x_2)\times A(\phi)B(\psi)$ are

$$\frac{d^2E_i}{dx_i^2}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{2}{x_i-1}+\frac{1}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)(x_i-1)^2+l_1(x_i-1)+l_2]}{4(x_i-a)(x_i-1)^2x_i}E_i=0, \qquad (2.50)$$

where $i=1, 2$:

$$\frac{d^2A}{d\phi^2}=-(2F+1)^2A,\quad (a-1)\frac{d^2B}{d\psi^2}=l_2B.$$

The operators whose eigenvalues are the separation constants are

$$L_1=(a-1)[D^2+\tfrac{1}{4}(P_1-K_1)^2] \qquad (2.51)$$
$$-[N_1^2+\tfrac{1}{4}(P_0+K_0)^2]+\tfrac{1}{4}(a-2)(P_1+K_1)^2$$
$$L_2=\tfrac{1}{4}(a-1)(P_1+K_1)^2,\quad L_3=M_1^2.$$

(26) A suitable choice of coordinates is

$$t=(1/R)\sqrt{(x_1-a)(x_2-a)/(a-1)},$$
$$x=1/R\cos\psi\sqrt{-x_1x_2/a},$$
$$y=(1/R)\cos\phi,\quad z=(1/R)\sin\phi, \qquad (2.52)$$

where

$$R=[\sqrt{-x_1x_2/a}\sin\psi+\sqrt{(x_1-1)(x_2-1)/(1-a)}$$

and $x_1<0<1<x_2<a$.

The solution ψ of the wave equation has the form $\psi=R\Phi$. The separation equation for $\Phi=E_1(x_1)E_2(x_2)\times A(\phi)B(\psi)$ are

$$\frac{d^2E_i}{dx_i^2}+\frac{1}{2}\left[\frac{1}{x_i-a}+\frac{1}{x_i-1}+\frac{2}{x_i}\right]\frac{dE_i}{dx_i}$$
$$+\frac{[4F(F+1)x_i^2+l_1x_i+l_2]}{4(x_i-a)(x_i-1)x_i^2}E_i=0 \qquad (2.53)$$

where $i=1, 2$:

$$\frac{d^2A}{d\phi^2}=-(2F+1)^2A,\quad a\frac{d^2B}{d\psi^2}=l_2B.$$

The operators whose eigenvalues are the separation constants are

$$L_1=-a[D^2+\tfrac{1}{4}(P_1-K_1)^2]-N_1^2+\tfrac{1}{4}(P_2+K_2)^2$$
$$+\tfrac{1}{4}(a+1)(P_1+K_1)^2,$$
$$L_2=-\tfrac{1}{4}a(P_1+K_1)^2,\quad L_3=M_1^2. \qquad (2.54)$$

This completes the list of coordinate systems of Class III.

Coordinate systems of Class IV

Coordinate systems of this type correspond to the two direct product reductions $SO(4, 2)\supset SO(2, 1)\otimes SO(2, 1)$ and $SO(4, 2)\supset SO(3)\otimes SO(1, 2)$. In each of these cases coordinates can be chosen from the nine separable classes of orthogonal coordinates on the two-sheeted and one-sheeted two-dimensional hyperboloids and the two separable classes of orthogonal coordinate systems on the two-dimensional sphere. The coordinate systems on these manifolds are given in the Appendix. In classifying coordinates of this type we give the general form of space–time coordinates in terms of the above mentioned two-dimensional manifolds.

(1) Coordinate systems corresponding to the reduction

$SO(4, 2)\supset SO(3)\otimes SO(1, 2)$.

A suitable choice of space–time coordinates is

$$t=\xi_2/(\xi_1+\xi_3),\quad x=\zeta_1/(\xi_1+\xi_3),$$
$$y=\zeta_2/(\xi_1+\xi_3),\quad z=\zeta_3/(\xi_1+\xi_3), \qquad (2.55)$$

where $\xi_1^2-\xi_2^2-\xi_3^2=-1$ and $\zeta_1^2+\zeta_2^2+\zeta_3^2=1$.

With the exception of coordinate systems of type 8 (which can always be chosen such that D is diagonal) there are 16 coordinate systems of this type on the single and double sheeted hyperboloids.

In each case the solution of the wave equation has the form

$$\psi=(\xi_1+\xi_3)\phi(\zeta_1, \zeta_2, \zeta_3)\theta(\xi_1, \xi_2, \xi_3),$$

where the functions ϕ and θ satisfy the equations

$$(M_1^2+M_2^2+M_3^2)\phi=-l(l+1)\phi,$$
$$[\{P_0, K_0\}+D^2]\theta=l(l+1)\theta, \qquad (2.56)$$

and l is a positive integer. The operators corresponding to each of the 16 possible coordinate systems can then be read off from the Appendix, if we make the identifications $N_1=\frac{1}{2}(P_0+K_0)$, $N_2=D$, and $M_3=\frac{1}{2}(P_0-K_0)$ in the case of the SO(1, 2) coordinates.

(2) Coordinate systems corresponding to the reduction

$SO(4, 2)\supset SO(2, 1)\otimes SO(2, 1)$.

A suitable choice of space–time coordinates is

$$t=\xi_1/(\xi_1+\xi_3),\quad x=\xi_2/(\xi_1+\xi_3),$$
$$y=\zeta_2/(\xi_1+\xi_3),\quad z=\zeta_3/(\xi_1+\xi_3), \qquad (2.57)$$

where $\xi_1^2-\xi_2^2-\xi_3^2=\epsilon$, $\xi_1^2-\xi_2^2=-\epsilon$, $\epsilon=\pm1$.

Again with the exception of coordinate systems of type 8 there are 64 coordinate systems. In each case the solution of the wave equation has the form $\psi=(\xi_1+\xi_3)\phi(\xi_1, \xi_2, \xi_3)\theta(\zeta_1, \zeta_2, \zeta_3)$, where the functions ϕ and θ satisfy the equations

$$(N_2^2+N_3^2-M_1^2)\theta=j(j+1)\theta,$$
$$[-\{P_1, K_1\}+D^2]\phi=j(j+1)\phi, \qquad (2.58)$$

and $j=-\frac{1}{2}+iq$, $0<q<\infty$, for globally defined solutions.

The operator corresponding to the SO(2, 1) algebra associated with the vector (ξ_1, ξ_2, ξ_3) can be read off from the Appendix with the identification $N_2=\frac{1}{2}(P_1-K_1)$, $N_2=D$, and $M_3=\frac{1}{2}(P_1+K_1)$.

We have looked at four classes of coordinate systems for which the wave equation (*) is strictly R-separable and found 106 distinct such coordinate systems. This added to the results of Ref. 4, gives a total of 367 inequivalent R-separable coordinate systems for the wave equation (*).

APPENDIX

In this appendix we list the orthogonal separable coordinate systems for the two-dimensional sphere, single-sheeted and double-sheeted hyperboloids. In each case we list the symmetric second order operator in the enveloping algebras of the symmetry groups of these manifolds which describes the coordinate system. The coordinates (with the exception of the single-sheet hyperboloid) can be found in the article by Olevski[7] and the operator characterization is due to Winternitz *et al.*[9]

I. Coordinate systems separable on the two-dimensional sphere

$$\zeta_1^2+\zeta_2^2+\zeta_3^2=1$$

If we write the generators $M_1=\zeta_2\partial_{\zeta_3}-\zeta_3\partial_{\zeta_2}$, $M_2=\zeta_1\partial_{\zeta_3}-\zeta_3\partial_{\zeta_1}$, and $M_3=\zeta_1\partial_{\zeta_2}-\zeta_2\partial_{\zeta_1}$ the coordinate systems and operators are:

$$(\xi_1^{(1)})^2=x_1x_2/a,\quad (\xi_2^{(1)})^2=(x_1-1)(1-x_2)/(a-1),$$
$$(\xi_3^{(1)})^2=(x_1-a)(x_2-a)/a(a-1),\quad 0<x_1<1<x_2<a. \tag{A1}$$

The operator is $L=aM_2^2+M_3^2$.

$$\xi^{(2)}=(\cos x_1, \sin x_1\cos x_2, \sin x_1\sin x_1\sin x_2). \tag{A2}$$

The operator is $L=M_1^2$.

II. Coordinate systems on the one- and two-sheeted two-dimensional hyperboloids

$$\xi_1^2-\xi_2^2-\xi_3^2=\pm 1.$$

We adopt the notation $N_1=\xi_1\partial_{\xi_2}+\xi_2\partial_{\xi_1}$, $N_2=\xi_1\partial_{\xi_3}+\xi_3\partial_{\xi_1}$, and $M_3=\xi_2\partial_{\xi_3}-\xi_3\partial_{\xi_2}$.

$$(\xi_1^{(1)})^2=x_1x_2/a,\quad (\xi_2^{(1)})^2=(x_1-1)(x_2-1)/(a-1),$$
$$(\xi_3^{(1)})^2=(x_1-a)(a-x_2)/a(a-1),\quad 1<x_1<a<x_2,$$
$$\xi^{(1)}\cdot\xi^{(1)}=(\xi_1^{(1)})^2-(\xi_2^{(1)})^2-(\xi_3^{(1)})^2=1. \tag{A3}$$

The coordinates on $\xi\cdot\xi=-1$ are obtained by the substitution $\xi^{(1)}\to i\xi^{(1)}$ and $x_1<0<1<x_2<a$. The operator is $L=N_1^2+aN_2^2$.

$$(\xi_1^{(2)})^2=(x_1-1)(1-x_2)/(a-1),$$
$$(\xi_2^{(2)})^2=-x_1x_2/a,$$
$$(\xi_3^{(2)})^2=(x_1-a)(a-x_2)/a(a-1),$$
$$x_1<0<1<a<x_2,\quad \xi^{(2)}\cdot\xi^{(2)}=1. \tag{A4}$$

The coordinates on the single-sheeted hyperboloid $\xi\cdot\xi=-1$ are obtained via the substitution $\xi\to i\xi$ and $1<x_1, x_2<a$; $x_1, x_2>a$. The operator is $L=N_1^2-aM_3^2$.

$$(\xi_1^{(3)}+i\xi_2^{(3)})^2=2(x_1-a)(x_2-a)/(a-b),$$
$$a=b^*=\alpha+i\beta,\quad (\xi_3^{(3)})^2=-x_1x_2/ab,\quad x_1<0<x_2, \tag{A5}$$

$\xi^{(3)}\cdot\xi^{(3)}=1$. For $\xi^{(3)}\cdot\xi^{(3)}=-1$ we have $\xi\to i\xi$ and $x_1, x_2>0$. The operator is $L=\alpha(M_3^2-N_2^2)+\beta\{M_3, N_2\}$.

$$\xi_1^{(4)}+\xi_2^{(4)}=\sqrt{-x_1x_2},$$
$$\xi_1^{(4)}-\xi_2^{(4)}=\sqrt{-x_1/x_2}+\sqrt{-x_2/x_1}+\sqrt{-x_1/x_2},$$
$$\xi_3^{(4)}=\sqrt{(1-x_1)(x_2-1)},\quad x_1<0<1<x_2. \tag{A6}$$

$\xi^{(4)}\cdot\xi^{(4)}=1$. The coordinates on the single-sheeted hyperboloid are obtained via the substitution $\xi\to i\xi$ with $x_1, x_2>1$; $0<x_1, x_2<1$; $x_1, x_2<0$. The operator is $L=N_1^2-(N_2+M_3)^2$.

$$\xi_1^{(5)}+\xi_2^{(5)}=\sqrt{x_1x_2},$$
$$\xi_1^{(5)}-\xi_2^{(5)}=-(\sqrt{x_1/x_2}+\sqrt{x_2/x_1}+\sqrt{x_1x_2}),$$
$$\xi_3^{(5)}=\sqrt{(x_1-1)(x_2-1)},\quad 0<x_1<1<x_2, \tag{A7}$$

$\xi^{(5)}\cdot\xi^{(5)}=1$. The coordinates on the single-sheeted hyperboloid are obtained via the substitution $\xi\to i\xi$ with $x_1<0<x_2<1$. The operator is $L=N_1^2+(N_2+M_3)^2$.

$$\xi_1^{(6)}+\xi_2^{(6)}=\sqrt{-x_1x_2},$$
$$\xi_1^{(6)}-\xi_2^{(6)}=(x_1-x_2)/[4(-x_1x_2)^{3/2}],$$
$$\xi_3^{(6)}=\tfrac{1}{2}[\sqrt{-x_2/x_1}-\sqrt{-x_1/x_2}],\quad x_1<0<x_2. \tag{A8}$$

$\xi^{(6)}\cdot\xi^{(6)}=1$. The coordinates on the single-sheeted hyperboloid are obtained via the substitution $\xi\to i\xi$ with $x_1, x_2>0$. The operator is $L=\{N_1, N_2-M_3\}$.

$$\xi_1^{(7)}+\xi_2^{(7)}=\sqrt{x_1},\quad \xi_1^{(7)}-\xi_2^{(7)}=1/\sqrt{x_1}+\sqrt{x_1}x_2^2,$$
$$\xi_3^{(7)}=x_2\sqrt{x_1},\quad x_1, x_2>0. \tag{A9}$$

$\xi^{(7)}\cdot\xi^{(7)}=1$. The coordinates on the single-sheeted hyperboloid are obtained via the substitution $\xi\to i\xi$ with $x_1<0<x_2$. The operator is $L=(N_2+M_3)^2$.

$$\xi^{(8)}=(\cosh x_1\cosh x_2, \cosh x_1\sinh x_2, \sinh x_1),$$
$$\xi^{(8)}\cdot\xi^{(8)}=1,$$
$$\hat{\xi}^{(8)}=(\sinh x_1\cosh x_2, \sinh x_1\sinh x_2, \cosh x_1),$$
$$\hat{\xi}^{(8)}=(\sin x_1\sinh x_2, \sin x_1\cosh x_2, \cos x_1),$$
$$\hat{\xi}^{(8)}\cdot\hat{\xi}^{(8)}=-1. \tag{A10}$$

The operator is $L=N_1^2$.

$$\xi^{(9)}=(\cosh x_1, \sinh x_1\cos x_2, \sinh x_1\sin x_2),$$
$$\xi^{(9)}\cdot\xi^{(9)}=1,$$
$$\hat{\xi}^{(9)}=(\sinh x_1, \cosh x_1\cos x_2, \cosh x_1\sin x_2),$$
$$\hat{\xi}^{(9)}\hat{\xi}^{(9)}=-1. \tag{A11}$$

The operator is $L=M_3^2$.

[1]E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space—time. I. The Lorentz Group," J. Math. Phys. **18**, 1 (1977).
[2]E. G. Kalnins and W. Miller, Jr., "Lie theory and the wave equation in space time 2. The group $SO(4, C)$," SIAM J. Math. Anal. (to appear).
[3]E. G. Kalnins and W. Miller Jr., "Lie theory and the wave equation in space—time. 3. Semisubgroup coordinates," J. Math. Phys. **18**, 271 (1977).
[4]E. G. Kalnins and W. Miller Jr., "Lie theory and the wave equation in space time 4. The Klein Gordon equation and the Poincaré group," (in preparation).
[5]E. G. Kalnins and W. Miller Jr., J. Math. Phys. **17**, 331 (1976).
[6]M. Bocher, *Ueber die Reihenentwickelungen der Potentialtheorie* (Teubner, Leipzig, 1894).
[7]M. P. Olevski, Mat. Sb. **27**, 379 (1950).
[8]E. G. Kalnins, W. Miller, and P. Winternitz, "The group O(4), separation of variables and the hydrogen atom," SIAM J. Appl. Math. (to appear).
[9]P. Winternitz, I. Lukac, and Ya. A. Smorodinski, Sov. J. Nucl. Phys. **7**, 139 (1968).

Continuum calculus. II. The heterogeneous continuous functional differentiation applied to the Feynman path integral

L. L. Lee

School C.E.M.S., University of Oklahoma, Norman, Oklahoma 73019
(Received 7 October 1976)

The continuum calculus proposed previously [L. L. Lee, J. Math. Phys. **17**, 1988 (1976)] is here extended to the study of continuous differentiation of a functional. The result, called HCFD, is shown to be the inverse operation of the functional integration for Feynman path integrals, in analogy to the case in ordinary differential calculus. The class of "separable" functionals is defined, which are useful in the derivation of the theory, playing a role similar to that of the characteristic functions in the Lebesgue theory of integration. A Radon–Nikodym type derivative is introduced in the definition of the continuous derivative for a general Banach algebra. This development constitutes a functional calculus of the continuum type. Comparisons with other types of functional derivatives are also made.

I. INTRODUCTION

In a previous paper[1] (hereafter referred to as I), we developed an operational calculus, called the continuum calculus, consisting of two operations, the r differentiation and the p integrations on functions. We were able to characterize the Feynman[2] path integrals of certain type of functionals found in quantum mechanics as the outcome of the interaction between the proposed p integral and the ordinary integral (Riemann or Lebesgue). The resulting formula for the functional integral was given in closed form of known mathematical operations, in contrast to the previous P projection method of cylinder functionals[3] and subsequent n-fold integration with respect to a weak distribution of measures.[4] This formula is general in that it does not depend on the Gaussian measure which underlies most past calculations and admits finite limits of integration. Applications to the quantum harmonic oscillator[2] and the electron–phonon interaction[5] reproduced the well-known results. It was also applied to the probability theory of infinitely divisible distributions[6] and other branches of mathematics. In this paper, we report another application of the continuum calculus, the interaction of the p integration with the ordinary differentiation in functional calculus. The outcome will be called the heterogeneous continuous functional differentiation (or HCFD). The HCFD is found to be the "inverse" operation of the functional integration proposed previously,[1] in the same sense that the ordinary differentiation is the inverse operation of ordinary integration. The HCFD and the functional integration constitute the basis of a functional calculus, upon which our future developments will depend.

The utility of the Feynman (–Wiener) integral in many branches of physics has generated recent interests in its study. Notably the applications of the Feynman integral method to the studies of fluid turbulence,[7] laser beam propagation in random media,[8] multichannel scattering,[9] etc., aside from its usual applications in quantum mechanics and field theory. Past studies of this integral met the difficulty of establishing a satisfactory measure in an infinite-dimensional space, as pointed out by Skorohod,[4] and touched upon by Xia.[10] This difficulty is partially removed by the continuum calculus approach.[1] We demonstrated in I that for certain types of functionals the integral exhibits a two-tier structure involving two measures, one on the function space, the other on the base space. Here we present some further developments of the continuum calculus, the functional differentiation.

In Sec. II, we give a heuristic demonstration of the HCFD method on a simple "separable" functional of the exponential type. With insight gained therefrom, we formulate the definition of HCFD in Sec. III in a general complex Banach algebra. We prove the important "inverse" theorem, showing that the HCFD is the inverse operation of the functional integration introduced earlier. In Sec. IV, the method is applied to a known functional as an illustration. A comparison is made with the functional derivative introduced by Friedrichs.[3] We cite future developments in the concluding section.

To make this paper relatively self-contained, we recapitulate some of the important results from I. Let B be a complex Banach space, and A_B the Banach algebra constructed from the collection, C^B, of functions from B to C, the complex numbers. Let (B, S_B, μ) be a measure space[11] on B, and (A_B, S_A, m) a measure space on A_B. (S_B, S_A being the σ algebras, and μ, m the measures on B and A_B, respectively). The r differentiation, R/Rt, of a function $f(t)\in A_B$ has been shown to be

$$\frac{R}{Rt}f(t)\equiv\lim_{\|b\|\to 0}\left|\frac{f(t+b)}{f(t)}\right|^{\|b\|^{-1}}=\exp\left[\frac{d}{dt}\ln f(t)\right], \tag{1.1}$$

for $t\in B-\bar{N}$, $\bar{N}$ being the closure of the kernel, N, of f, whenever the limit exists. A more rigorous limiting ϵ-δ type definition can be given as follows: for every $\epsilon>0$, there exists $\delta>0$ such that

$$\left|f(t+b)-f(t)[f^*(t)^{\|b\|}]\right|<\epsilon \tag{1.2}$$

whenever $\|b\|<\delta$. If such an f^* exists at $t\in B-\bar{N}$, we call it the r deviative of $f(t)$ at t, i.e., $Rf(t)/Rt=f^*(t)$. We can recognize (1.1) as a measure of the instantaneous *ratio* of a function at the neighborhood of a point t, in

the same sense that $df(t)/dt$ is the measure of the instantaneous *difference* of $f(t)$, which is of concern in ordinary differential calculus.

We point out some prominent algebraic features of this new operation. The Banach algebra, A_B, can be considered as a commutative ring over the complex field with some scalar multiplication properties. The ordinary differentiation, d/dt, is linear with respect to the additive part of the ring operations, i.e.,

$$\frac{d}{dt}[\alpha f(t)+\beta g(t)]=\alpha\frac{d}{dt}f(t)+\beta\frac{d}{dt}g(t),\quad f,g\in A_B;\ \alpha,\beta\in C. \tag{1.3}$$

Whileas the r differentiation, R/Rt, is "linear" with respect to the multiplicative part of the ring operations, i.e., it is "additive" (or rather "multiplicative") with respect to products,

$$\frac{R}{Rt}(f(t)g(t))=\left(\frac{R}{Rt}f(t)\right)\left(\frac{R}{Rt}g(t)\right),\quad f,g\in A_B \tag{1.4}$$

and "homogeneous" with respect to exponentiation by scalars,

$$\frac{R}{Rt}[f(t)^\alpha]=\left[\frac{R}{Rt}f(t)\right]^\alpha,\quad f\in A_B,\ \alpha\in C \tag{1.5}$$

[for proofs, see (I.2.4) and (I.2.5)]. In a certain sense, the situation is similar (but not identical) to that for Fourier transforms. The Fourier transformation, $\mathcal{J}$, is "additive" ("multiplicative", *sic*) with respect to convolution, $*$,[12]

$$\mathcal{J}(f(t)*g(t))=(\mathcal{J}f(t))(\mathcal{J}g(t)). \tag{1.6}$$

However, $\mathcal{J}$ is homogeneous with respect to multiplication by scalars,

$$\mathcal{J}(\alpha f(t))=\alpha(\mathcal{J}f(t)). \tag{1.7}$$

The p integral was then obtained through a search for the "primitive" of the r differentiation, and is related to the ordinary integral by

$$P_E[f(t)]^{\mu(dt)}=\exp[\textstyle\int_E\mu(dt)\ln f(t)],\quad E\subset B. \tag{1.8}$$

The functional integral, I_f, for the exponential type functionals,

$$\phi[y]\equiv\exp\textstyle\int_E\mu(dt)f(y(t)),\quad f:C\to C \tag{1.9}$$

was found by interacting the p integration with ordinary integration,

$$I_f(\phi)=\exp\textstyle\int_E\mu(dt)\ln[\int_F m(dy(t))\exp f(y(t))],\quad E\subset B,\ F\subset A_B, \tag{1.10}$$

whenever the successive integrals exist. Details can be found in I.

II. AN EXAMPLE OF FUNCTIONAL DIFFERENTIATION

We demonstrate in this section the method of continuous functional differentiation by investigating a simple functional of the type,

$$\phi[y]=\exp\textstyle\int_0^1 dt\,b(t)\,y(t)^2, \tag{2.1}$$

where y is a complex function defined on real numbers, $y\in C^R$, ϕ is defined on C^R with complex values. In conventional theory of functionals (see, e.g., Volterra[13]), the one-point functional derivative of a given functional, $\psi[y]$, at point t, $0<t<1$, is defined as,

$$\Delta^1\psi[y]\equiv\textstyle\int_0^1 dt\frac{\delta\psi}{\delta y(t)}\,\delta y(t), \tag{2.2}$$

where $\Delta^1\psi$ denotes the first order variation of the functional $\psi[y]$ with respect to the variation $\delta y(t)$ in $y(t)$. The nth order functional derivative at n distinct points, $0<t_1<t_2<\ldots<t_n<1$, can be defined recursively and given the notation,

$$\frac{\delta^n}{\delta y(t_1)\cdots\delta y(t_n)}\psi[y]=\left[\prod_{i=1}^n\frac{\delta}{\delta y(t_i)}\right]\psi[y]. \tag{2.3}$$

We intend to find the limit of the nth order derivative (2.3) as $n\to\infty$ and t becomes the continuous index, $0\le t\le 1$. We first observe that the discrete (P projection) representation of (2.1) is,

$$\phi[P_n y]=\exp\sum_{i=1}^n b(t_i)y(t_i)^2. \tag{2.4}$$

We need a formal definition of a functional that is "separable".

2.1 Defintion: Separable functionals: A complex valued functional, $\phi[y]$ on C^R is called separable iff its P projection under the projector, P_n, $\phi[P_n y]$, can be decomposed into an n product, $\forall\, n$,

$$\phi[P_n y]=\prod_{i=1}^n f_i(y_i) \tag{2.5}$$

for some functions $f_i:C\to C$, and y_i being the projection of $y(t)$ by P_n into the ith interval, (cylinder functionals).[3]

This definition can also be extended to the complex Banach space. Examination of (2.4) shows that (2.1) is separable, with $f_i(y_i)=\exp[b(t_i)y(t_i)^2]$,

$$\phi[P_n y]=\prod_{i=1}^n\exp[b(t_i)y(t_i)^2]. \tag{2.6}$$

Now we apply the nth functional derivative (2.3) to (2.6),

$$D_n\equiv\prod_{i=1}^n\left(\frac{\delta}{\delta y(t_i)}\right)\phi[P_n y]=\prod_{i=1}^n\left[\frac{\delta}{\delta y(t_i)}\exp b(t_i)y(t_i)^2\right]. \tag{2.7}$$

Carrying out the differentiation,

$$D_n=\prod_{i=1}^n[2b(t_i)y(t_i)\exp b(t_i)y(t_i)^2]. \tag{2.8}$$

As $n\to\infty$ and $\sup\Delta t_i\to 0$, the n product approaches the p integral,[1]

$$\lim_{\substack{n\to\infty\\ \sup\Delta t_i\to 0}}\prod_{i=1}^n[2b(t_i)y(t_i)\exp b(t_i)y(t_i)^2]^{\Delta t_i}=P\int_0^1 dt\gg[2b(t)y(t)\exp b(t)y(t)^2]$$
$$=\exp\textstyle\int_0^1 dt\ln[2b(t)y(t)\exp b(t)y(t)^2]=2[\exp\int_0^1 dt\,b(t)y(t)^2][\exp\int_0^1 dt\ln b(t)y(t)]. \tag{2.9}$$

If we apply the p integral formula earlier, expressing the nth order derivation in the operator form, we have

$$\lim_{\substack{n\to\infty\\ \sup\Delta t_i\to 0}}\prod_{i=1}^n\left(\frac{\delta}{\delta y(t_i)}\right)^{\Delta t_i}\cdot\phi[y]=\exp\textstyle\int_0^1 dt\ln\frac{\delta}{\delta y(t)}\circ\phi[y]$$
$$=\exp\textstyle\int_0^1 dt\ln\frac{\delta}{\delta y(t)}\exp b(t)y(t)^2=\exp\int_0^1 dt\ln 2b(t)y(t)\exp b(t)y(t)^2=2\phi[y]\exp\int_0^1 dt\ln b(t)y(t). \tag{2.10}$$

The same result is obtained. Therefore the operator approach is applicable to "separable" functionals in carrying out the continuous functional differentiation. We shall call the derivative thus obtained the heterogeneous continuous functional derivative of $\phi[y]$, to stress the fact that the continuous index, t, runs from $0 \leq t \leq 1$. We extend the definition to a general Banach space in the following.

III. THE CONTINUOUS FUNCTIONAL DIFFERENTIATION

Let B be a complex Banach space. Let A_B be the Banach algebra formed from the usual construction on the collection of complex valued functions from B to C, i.e., C^B, with C complex numbers. Let (B, S_B, μ) be a measure space on B, and (A_B, S_A, m) a measure space on A_B (S_B and S_A being the σ algebras, and μ, m the corresponding measures for B and A_B, respectively) The functionals are defined from A_B to C, $\phi: A_B \to C$. We consider the special type,

$$\phi[y] = \exp \int_E \mu(dt)\, f(y(t)), \tag{3.1}$$

where f is a complex vauled function, $f: C \to C$, and $E \subseteq B$.

Under suitable conditions, which for the moment we do not wish to elaborate, a Radon–Nikodym[14] type derivative, $g(y(t))$, $g: C \to C$, can be found such that

$$\begin{aligned}\int_E \mu(dt) f(y(t)) &= \int_E \mu(dt) \ln \exp[f(y(t))] \\ &= \int_E \mu(dt) \ln \int_F m(dy(t)) g(y(t)),\end{aligned} \tag{3.2}$$

for some $F \subseteq A_B$. In case such derivative exists for the given f, we can define the continuous functional derivative of (3.1) as follows.

3.1 *Definition*: *The heterogeneous continuous functional derivative in a Banach space*: The continuous functional derivative, $D_f(\phi)$, of the exponential functional (3.1) is given to be

$$D_f[\Phi] = \exp \int_E \mu(dt) \ln g(y(t)) \tag{3.3}$$

if such integral exists. g being the (*pseudo*) Radon–Nikodym derivative of $\exp(f)$ as defined by (3.2) for some $F \subseteq A_B$.

We called g the *pseudo*-Radon–Nikodym derivative because it is not clear whether $\exp(f)$ is a normal measure or can be transformed into an equivalent measure on F such that m is absolutely continuous with respect to it.[13] We adhere here to the definition (3.2). More research is required.

We can also write (3.3) in the operator form,

$$\begin{aligned}D_f[\phi] &= P_E \left[\frac{\partial}{\partial m(y(t))}\right]^{\mu(dt)} \cdot \phi[y] \\ &= \exp \int_E \mu(dt) \ln \frac{\partial}{\partial m(y(t))} \cdot \phi[y].\end{aligned} \tag{3.4}$$

It is then an easy matter to show that HCFD is the inverse of functional integration.

3.2 *Theorem*. *The inverse operation of functional integration*: For a functional of the type (3.1), the heterogeneous continuous functional differentiation defined in (3.3) is the inverse operation of functional integration on $\phi[y]$. Let the functional integral of ϕ be I_f, then

$$D_f[I_f[\phi]] = \phi[y]. \tag{3.5}$$

Proof: If the functional integral exists for $\phi[y]$ on $F \subseteq A_B$, (1.10) gives

$$I_f[\phi] = \exp \int_E \mu(dt) \ln \int_F m(dy(t)) \exp f(y(t)). \tag{3.6}$$

To differentiate (3.6) according to (3.3), the *pseudo*-Radon–Nikodym derivative can be easily identified as $\exp(f)$ in (3.6), and

$$\begin{aligned}D_f[I_f[\phi]] &= \exp \int_E \mu(dt) \ln \exp f(y(t)) \\ &= \exp \int_E \mu(dt) f(y(t)) \quad \text{Q.E.D.}\end{aligned} \tag{3.7}$$

We see that $D_f(I_f[\phi]) = \phi$; therefore D_f is the left inverse of I_f, similar to the situation in differential calculus.

IV. APPLICATION OF HCFD AND THE DERIVATIVE OF FRIEDRICHS

In this section, we shall integrate a known functional with a finite limit of integration, then differentiate it according to HCFD to recover the original functional. Consider the functional,

$$\phi[y] = \exp \int_0^1 dt\, b(t) y(t) \tag{4.1}$$

whose functional integral with a Gaussian measure has been investigated in I. When a Gaussian measure is used, the limits of integration are usually improper, and the final integral is independent of the limits, i.e., it is no longer a functional after integration. In the continuum calculus, it is possible to integrate with finite limits without the Gaussian measure. Now we choose the limits to be, $-\infty \leq y(t) \leq x(t)$, $x(t)$ a finite function. The functional integral of (4.1) is then

$$\begin{aligned}I_f^x[\Phi] &= \exp \int_0^1 dt \ln \int_{-\infty}^{x(t)} dy(t) \exp b(t) y(t) \\ &= \exp \int_0^1 dt \ln \frac{\exp b(t) x(t)}{b(t)} \\ &= \phi[x] \exp - \int_0^1 dt \ln b(t).\end{aligned} \tag{4.2}$$

The resulting integral is again a functional of $x(t)$. Now we seek the continuous functional derivative of (4.2) with respect to $x(t)$. Noting that (4.2) can be written as

$$\begin{aligned}I_f^x[\phi] &= \exp \int_0^1 dt \ln \frac{\exp b(t) x(t)}{b(t)} \\ &= \exp \int_0^1 dt \ln \int_{-\infty}^{x(t)} dy(t) \exp b(t) y(t).\end{aligned} \tag{4.3}$$

We identify $\exp[f(x(t))]$ of (3.2) as $[\exp b(t)x(t)]/b(t)$ and $g(x(t))$ to be its ordinary derivative, $\exp[b(t)x(t)]$. The derivative, D_f, is then, according to (3.3),

$$D_f[I_f^x[\phi]] = \exp \int_0^1 dt \ln \exp b(t) x(t) = \phi[x]. \tag{4.4}$$

This implies again $D_f(I_f[\phi]) = \phi$. That D_f is inverse to I_f has been shown for this example.

We note that at times the proper derivate $\delta/\delta y(t)$ to be used with our formulas differs from the conventional functional derivative by a factor of the Dirac delta function. The application of our derivative is also akin to that of the ordinary partial derivative, $\partial/\partial y_i$, again here without the Kronecker delta.

Also a remark on the functional derivative proposed by Friedrichs[3] is in order. Friedrichs defined the

derivative, $\delta/\delta y(t)dt$ for a cylinder functional $\phi[y] = \phi[P_n y]$ under the projection operator P_n by

$$\frac{\delta}{\delta y(t)dt}\,\phi[y] \equiv \frac{1}{\Delta_i}\frac{\partial}{\partial y_i}\phi[P_n y] = \frac{1}{\Delta_i}\frac{\partial}{\partial y_i}\phi(y_1, \cdots, y_n), \qquad (4.5)$$

for $t \in C_i$, C_i being a measurable interval, and Δ_i is the measure of C_i, also $P_n y(t) = (y_1, \cdots, y_i, \cdots, y_n)$. In the discussions ensued for the case of quadratic functionals,

$$\phi_2[y] = \int dr \int ds\, y(r)\, b(r,s)\, y(s), \quad b(r,s) \text{ symmetric.} \qquad (4.6)$$

The definition (4.5) does not differ from the conventional functional derivative[13] (2.2) appreciably, at least for the particular case considered by Friedrichs,

$$\int w(t)\,\frac{\delta}{\delta y(t)dt}\,dt\,\phi_2[y] = 2\int ds \int dt\, w(t)\, b(t,s)\, y(s) \qquad (4.7)$$

for some integrable function on $w(t)$. While for the ordinary functional derivative,

$$\frac{\delta\phi_2}{\delta y(t)} = \int dr \int ds\, y(r)\, b(r,s)\, \delta(s,t) + \int dr \int ds\, \delta(t,r)\, b(r,s) y(s)$$
$$= 2\int ds\, b(t,s) y(s). \qquad (4.8)$$

Therefore,

$$\int dt\, w(t)\frac{\delta\phi_2}{\delta y(t)} = 2\int ds \int dt\, w(t)\, b(t,s)\, y(s), \qquad (4.9)$$

which is exactly the same result.

V. CONCLUDING REMARKS

We presented in this work a formula for the continuous functional differentiation of a functional, which was subsequently shown to be the "left inverse" of functional integration given in paper I. In the development, the operation of *potentiation*, or p integration, again played an important role. The continuous derivative is essentially the result of the interaction of the p integration with ordinary functional differentiation. We hope to expound the detailed theoretical implications by employing measure theoretical, topological, and algebraic methods. Some ground has been broken by the efforts of Morette-DeWitt,[15] Skorohod,[4] and Xia.[10] We shall have occasion to comment on this. New avenues of development remain to be explored, such as the homogeneous continuous functional differentiation, integral formulas for new types of functionals, and also differential equations for functionals. The physical applications of Feynman integrals often times involve complicated functionals and their various derivatives, such as in the theories of turbulence,[7] and laser beam propagation.[8] Careful analyses are required.

Of immediate interest is the characterization of the functional integrals of functionals of the type,

$$\phi[y] = \int_0^1 dt f(y(t)), \quad f: C \to C. \qquad (5.1)$$

A differentiable homotopy approach was proposed in paper I. However, a uniqueness theorem is needed to define its applicability in general. The solution might be offered by a generalized distribution theory[16] approach. This is under investigation and will be reported subsequently.

[1] L. L. Lee, J. Math. Phys. **17**, 1988 (1976).
[2] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[3] K. O. Friedrichs *et al.*, *Integration of Functionals* Seminar (New York University, New York, 1957).
[4] A. V. Skorohod, *Integration in Hilbert Space* (Springer, New York, 1974).
[5] R. Abé, Busseiron Kenkyu **79**, 101 (1954).
[6] B. V. Gnedenko, and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, Reading, Mass., 1968).
[7] See review by J. Lee, J. Math. Phys. **15**, 1571 (1974); also A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, Mass., 1975).
[8] P.-L. Chow, J. Stat. Phys. **12**, 93 (1975).
[9] J. R. Laing and K. F. Freed, Phys. Rev. Lett. **34**, 849 (1975).
[10] D. X. Xia, *Measure and Integration Theory on Infinite-Dimensional Spaces* (Academic, New York, 1972).
[11] N. Dinculeanu, *Vector Measures* (Pergamon, New York, 1967).
[12] C. E. Rickart, *General Theory of Banach Algebra* (Van Nostrand, New York, 1960).
[13] V. Volterra, *Theory of Functionals and of Integral and Integro-differential Equations* (Dover, New York, 1959).
[14] A. C. Zaanen, *Integration* (North-Holland, Amsterdam, 1967).
[15] C. Morette DeWitt, Commun. Math. Phys. **28**, 47 (1972); **37**, 63 (1974).
[16] L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1966); I. M. Guelfand and G. E. Chilov, *Les Distributions* (Dunod, Paris, 1962—70), 5 volumes.

The topology of Euclidean Higgs fields

G. Woo[a), b)]

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 18 April 1977)

It is proved that in a class of Euclidean Higgs theories, a single index serves to describe both the gauge field and the Higgs field topologies.

I. INTRODUCTION

The study of nonperturbative aspects of field theories has led to considerable interest in their global topological features.[1] In spontaneously broken gauge theories, this was centered at the beginning[2,3] just on the behavior of the Higgs fields at spatial infinity. In particular, Arafune, Freund, and Geobel[2] showed that, for the SU(2) Georgi–Glashow model, the covering of the 2-sphere at infinity by the isotriplet of Higgs fields has a mathematical description in terms of the Kronecker index.

Recently, however,[4] it has been realized that in four-dimensional Euclidean space–time, the Yang–Mills field itself has a topological characteristic of its own, independently of the scalar fields. This is the Pontryagin index which has an integral representation in terms of the product of the curvature and its dual.

In Euclidean space–time, of course, a new analysis of Higgs topology has to be undertaken since the relevant domain is now a 3-sphere. This is the subject of the present paper. Whereas, in Minkowski space, it is the requirement of energy finiteness that restricts the range of the Higgs fields, in Euclidean space this is replaced by the requirement of action finiteness. The range of the fields at Euclidean infinity thus defined will be called the Higgs asymptotic space.

There are two cases where the asymptotic space is such as to allow an integral formula for the Higgs topological invariant to be written. They occur when the asymptotic space is either a 3-sphere or a 2-sphere. Since $\pi_3(S^3) = \pi_3(S^2) = Z$, the invariant both times is an integer. Higher dimension spheres are not relevant as $\pi_3(S^n) = 0$ for $n > 3$.

The former applies to an SU(2) theory with a complex doublet χ_a of scalars with a Higgs potential of the form $U(\chi) = (\chi_a^\dagger \chi_a - 1)^2$. For simplicity here and elsewhere, all parameters have been set to unity. The latter would apply to a theory of the type devised by Georgi and Glashow rather than Weinberg, which involves an isotriplet ϕ_k with the potential $U(\phi) = (\phi_k \phi_k - 1)^2$.

It will be proved that in each theory the associated topological integer is identically equal to the Pontryagin index. This means that there is a one-to-one correspondence between sectors in the configuration manifolds of the gauge and Higgs fields. In particular, in the single pseudoparticle sector, the index of the Higgs fields must be unity. Account of this fact should be taken in any analysis of tunneling in the vacuum of a spontaneously broken gauge theory.

II. AN EXAMPLE WITH S^3 HIGGS ASYMPTOTIC SPACE

In this section, the following SU(2) Euclidean Lagrangian theory is studied:

$$\mathcal{L} = -\tfrac{1}{4} F_{k\mu\nu} F_{k\mu\nu} + \tfrac{1}{2} (D_\mu \chi)_a^\dagger (D_\mu \chi)_a - (\chi_a^\dagger \chi_a - 1)^2. \tag{1}$$

$F_{k\mu\nu}$ is the Yang-Mills curvature tensor for an isovector gauge field, and $(D_\mu \chi)_a$ is the gauge covariant derivative acting on a complex doublet of Higgs scalars.

In terms of real components, let $\chi_1 = \phi_1 + i\phi_2$ and $\chi_2 = \phi_3 + i\phi_4$. The topological invariant categorizing the homotopy class of $(\phi_1, \phi_2, \phi_3, \phi_4)$ can be written as an integral over the 3-sphere S^3_∞ at Euclidean infinity:

$$I_H = (1/12\pi^2) \int_{S^3_\infty} \epsilon_{ijkl} \phi_i d\phi_j \wedge d\phi_k \wedge d\phi_l. \tag{2}$$

The indices on ϕ, of course, run from 1 to 4.

Now the finite action condition implies from Eq. (1) that

$$d\chi_a - i\omega_k (\sigma_k/2)_{ab} \chi_b \to 0 \quad \text{asymptotically.} \tag{3}$$

Here the usual notation has been used for the Pauli matrices, and ω_k is the Yang-Mills connection 1-form.

The substitution equation (3), together with the boundary condition that $\phi_i \phi_i \to 1$, implies that the invariant integral can be rewritten as

$$I_H = (1/16\pi^2) \int_{S^3_\infty} \omega^1 \wedge \omega^2 \wedge \omega^3. \tag{4}$$

The final use of the finite action requirement is in guaranteeing the asymptotic vanishing of the curvature, so that

$$\omega_k \sigma_k / 2i \to g^{-1} dg, \tag{5}$$

where g is a matrix of the adjoint representation of SU(2). This allows a transcription of Eq. (4) in terms of the Haar group measure:

$$I_H = \pm (1/24\pi^2) \int_{S^3_\infty} d\mu(g). \tag{6}$$

The uncertainty in sign reflects an arbitrariness in sign in the original definition of I_H given in Eq. (2). If the plus sign is taken, one obtains

$$I_H = I_P, \tag{7}$$

where I_P is the Pontryagin index of the gauge field configuration. This is the result which was to be proved.

a)Research supported in part by the National Science Foundation under Grant No. MPS75-20427.

b)Junior Fellow, Harvard University Society of Fellows.

III. THE HOPF INVARIANT

Although it was an easy matter to write formula (2) for the Kronecker index of a mapping from $S^3 \to S^3$, and although one can do the same with equal facility for the general case $S^n \to S^n$, it is not trivial to construct an integral expression for the Hopf invariant which labels homotopy classes of mappings from $S^3 \to S^2$.

That this is so was pointed out by Patani, Schlindwein, and Shafi.[5] The difficulty is that the obvious conserved tensor density one would write for, say, the O(3) nonlinear σ-model in three space dimensions, is the following:

$$J^{\mu\nu} = \epsilon^{\mu\nu\lambda\sigma}\epsilon_{abc}\partial_\lambda \pi_a \partial_\sigma \pi_b \pi_c. \tag{8}$$

The triplet of fields π_a are assumed to be nonsingular and normalized to satisfy $\pi_a\pi_a = 1$. The regrettable feature of this density is that all charges defined as either surface or volume integrals of it are identically zero. Thus sadly it is homotopically insensitive. The resolution of this problem was found somewhat earlier by Whitehead,[6] whose construction procedure will now be outlined.

This procedure is not in fact applicable to all continuous mappings from $S^3 \to S^2$, but only those which are twice differentiable. This smoothness condition, however, is certainly acceptable from a physical viewpoint.

The first step is to define a 2-form area element σ on S^2 normalized according to $\int_{S^2}\sigma = 1$. Given a twice differentiable map ϕ: $S^3 \to S^2$, the next step is to use its dual ϕ^* to define the 2-form $\phi^*\sigma$ on S^3. σ is, of course, not exact on S^2 since S^2 has no boundary, but $\phi^*\sigma$ is exact on S^3 since this space has a trivial second cohomology group. It is thus possible to find a 1-form Σ on S^3 which satisfies $d\Sigma = \phi^*\sigma$. Furthermore, this 1-form is unique up to the differential of a function.

The Whitehead integral representation for the Hopf invariant of ϕ is then

$$I_H = \int_{S^3}\Sigma\wedge d\Sigma. \tag{9}$$

An application of Stokes' theorem shows that it is well defined. In a coordinate frame it is not difficult to show that I_H is invariant under deformations. The fact that I_H is integer-valued can be established first for fiber mappings by integrating Σ along a fiber and then using Stokes' theorem to equate the result with a surface integral which is essentially a winding number. For the algebraic theorems required to complete the proof, and indeed for the proof in full, the reader is advised to consult the original reference.[6]

It is worth remarking that I_H as given above is even under a change in the orientation of S^2, but odd under a change in the orientation of S^3. This is in accord with general algebraic results.[7] The physical significance of this can be seen in the context of discrete space–time symmetries.

IV. AN EXAMPLE WITH S^2 HIGGS ASYMPTOTIC SPACE

In this section, the integral expression just obtained will be applied to a coordinate-free study of the Higgs topology in the Euclidean version of the Georgi–Glashow model:

$$\mathcal{L} = -\tfrac{1}{4}F_{i\mu\nu}F_{i\mu\nu} + \tfrac{1}{2}(D_\mu\phi)_i(D_\mu\phi)_i - (\phi_i\phi_i - 1)^2. \tag{10}$$

$F_{i\mu\nu}$ is the Yang–Mills curvature for an isovector gauge field, and $(D_\mu\phi)_i$ is the covariant derivative of an isotriplet of Higgs scalars. From Eq. (10), it is clear that the following are three necessary asymptotic conditions for the Euclidean action to be finite:

$$\phi_i\phi_i \to 1, \tag{11a}$$

$$d\phi_i + \epsilon_{ijk}\phi_k\omega_j \to 0, \tag{11b}$$

$$\Omega_i \equiv d\omega_i + \tfrac{1}{2}\epsilon_{ijk}\omega_j\wedge\omega_k \to 0. \tag{11c}$$

The notation ω_i is used here for the Yang–Mills connection 1-form, and Ω_i for the curvature 2-form. Equation (11c) implies that

$$\omega_i\sigma_i/2i \to g^{-1}dg, \tag{12}$$

where g is a matrix of the adjoint representation of SU(2).

To construct the Whitehead integral, an area element σ on S^2 must first be defined. This is done by imbedding S^2 in $\mathbf{R}^3 = \{(x_1, x_2, x_3)\,|\,x_i \in \mathbf{R} \text{ for } i = 1, 2, 3\}$. σ is then defined by the Hodge star operation:

$$\sigma = (1/4\pi)^*(x_i dx_i) = (1/8\pi)\epsilon_{ijk}x_i dx_j\wedge dx_k. \tag{13}$$

Then

$$\phi^*\sigma = (1/8\pi)\epsilon_{ijk}\phi_i d\phi_j\wedge d\phi_k. \tag{14a}$$

Use of Eq. (11b) now leads to the expression:

$$\phi^*\sigma = (1/8\pi)\epsilon_{ijk}\phi_i\omega_j\wedge\omega_k. \tag{14b}$$

By de Rham's theorem, there exists a 1-form Σ on S^3 with the property that $d\Sigma = \phi^*\sigma$. Noting Eq. (11c), one obtains

$$\Sigma = (1/4\pi)\phi_i\omega_i. \tag{15}$$

The Hopf invariant is then given as:

$$I_H = \int_{S^3_\infty}\Sigma\wedge d\Sigma = (1/16\pi^2)\int_{S^3_\infty}\omega^1\wedge\omega^2\wedge\omega^3. \tag{16}$$

Just as in the earlier study of $S^3 \to S^3$, there is an uncertainty of sign in the definition of I_H. Using Eq. (12), one writes

$$I_H = \pm(1/24\pi^2)\int_{S^3_\infty} d\mu(g). \tag{17}$$

Again, taking the plus sign, one arrives at the identity

$$I_H = I_P, \tag{18}$$

where I_P is the Pontryagin index.

V. CONCLUSION

It has been shown that, in two typical SU(2) theories, a single index is able to describe both the gauge field and the Higgs field topologies in Euclidean space–time. The noteworthy fact is that the Higgs asymptotic spaces have different dimensions in the two cases. Thus one can speculate that the phenomenon discovered might generalize considerably. In this respect, straightforward imbedding of SU(2) into higher-rank groups should present no difficulty.

The equivalence between the topological sectors associated with the vector and scalar fields highlights the underlying geometrical structure of spontaneously broken gauge theories. Physical applications of this result to vacuum processes[8] are under investigation.

ACKNOWLEDGMENT

The author is indebted to R. Shankar for introducing him to Hopf mappings.

APPENDIX

The notation of differential forms has allowed the mathematical arguments in this paper to be presented in an elegant and compact manner. For the benefit of those readers who may be unfamiliar with the terminology, the rudiments of the theory of the differentiation and integration of forms will be described.

Given a completely antisymmetric covariant tensor $\Omega_{\mu_1\cdots\mu_p}$ of rank p, one defines the corresponding p-form Ω according to

$$\Omega \equiv (1/p!)\Omega_{\mu_1\cdots\mu_p} dx^{\mu_1}\wedge dx^{\mu_2}\wedge\cdots\wedge dx^{\mu_p}. \tag{A1}$$

The wedge product satisfies $dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu$ and obeys the obvious rules of addition and scalar multiplication.

From this p-form, one can define a $(p+1)$-form, $d\Omega$, called the exterior derivative of Ω, which is written as

$$d\Omega = \frac{1}{p!}\frac{\partial \Omega_{\mu_1\cdots\mu_p}}{\partial x_\nu} dx^\nu \wedge dx^{\mu_1}\wedge\cdots\wedge dx^{\mu_p}. \tag{A2}$$

The integral $\int_S \Omega$ of Ω over some p-dimensional surface S parametrized by the p coordinates $\lambda^1, \ldots, \lambda^p$, is defined by

$$\int_S \Omega = \int_S \Omega_{1\cdots p}(\lambda^1, \ldots, \lambda^p) d\lambda^1 \cdots d\lambda^p. \tag{A3}$$

For numerous physical applications of these formulas, Ref. 9 is recommended.

[1] For a review, see R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).
[2] J. Arafune, P. G. O. Freund, and C. J. Goebel, J. Math. Phys. **16**, 433 (1975).
[3] Yu. S. Tyupkin, V. A. Fateev, and A. S. Schwarts, Zh. Eksp. Teor. Fiz. Pis. Red. **21**, 91 (1975) [JETP Lett. **21**, 42 (1975)]; M. I. Monastyrskiy and A. M. Perelemov, *ibid.* **21**, 94 (1975) [*ibid.* **21**, 43 (1975)]; S. Coleman, Erice Lecture Notes (1975).
[4] A. A. Belavin, A. M. Polyakov, A. S. Schwarts, and Y. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[5] A. Patani, M. Schlindwein, and Q. Shafi, J. Phys. A **9**, 1513 (1976).
[6] J. H. C. Whitehead, Proc. Natl. Acad. Sci. USA **33**, 117 (1947). The formula was extended to describe Hopf invariants for $S^{2n-1} \to S^n$ for arbitrary n by M. Kervaire, C. R. Acad. Sci. Paris **237**, 1489 (1953).
[7] P. J. Hilton and S. Wylie, *Homology Theory* (Cambridge U. P., Cambridge, 1960).
[8] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976); C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. B **63**, 334 (1976).
[9] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Franscisco, 1973).

Topology of Euclidean Yang–Mills fields: Instantons and monopoles[a)]

Tamiaki Yoneya[b)]

Department of Physics, The City College of the City University of New York, New York, New York 10031
(Received 28 March 1977)

We discuss the topological properties of the Yang–Mills fields from a unified point of view of the Wu–Yang global formulation. In four-dimensional Euclidean space, the gauge type is characterized by the second Chern class. The multi-instanton solution recently found by 't Hooft and generalized by Jackiw, Nohl, and Rebbi is discussed from this point of view. We also discuss the generalization of the topology of instantons and monopoles in higher dimensional spaces and the relation among them.

1. INTRODUCTION

Based on their consideration of nonintegrable phase factors, Wu and Yang[1] proposed to generalize the notion of gauge in such a way that it is independent of specific gauge field. A *gauge* is defined as follows: First one chooses regions $R_1, R_2, \cdots$ in space–time such that the union of $R_1, R_2, \cdots$ covers the whole space–time, and then specifies gauge transformations g_{ij} for each overlapping region $R_i \cap R_j$. The g_{ij} is called the transition function. In each region R_i, the gauge field $A_\mu^{(i)}$ is defined such that $A_\mu^{(i)}$ and $A_\mu^{(j)}$ are related by

$$A_\mu^{(i)}(x) = g_{ij}^{-1}(x) A_\mu^{(j)}(x) g_{ij}(x) + g_{ij}^{-1}(x) \partial_\mu g_{ij}(x) \qquad (1.1)$$

for $x \in R_i \cap R_j$. *A global gauge transformation* consists of deformation of the regions R_i and of usual local gauge transformation for each $A_\mu^{(i)}$. The collection of gauges that can be globally transformed into each other is said to belong to the same *gauge type*.

In the Abelian case, an advantage of this generalization is that it enables us to accommodate the physics of Dirac monopole without string singularities.[2] The gauge types are characterized by the Dirac quantization condition.

As emphasized by Wu and Yang,[1,2] this generalization corresponds, mathematically, to the description of the Yang–Mills theory on the general notion of fibre bundle.[3,4] The U(1) magnetic charges are proportional to the first Chern class number of a complex line bundle.

The purpose of this paper is to extend Wu and Yang's bundle picture to non-Abelian Yang–Mills fields in higher-dimensional Euclidean spaces. As we shall see, the bundle picture is indeed suitable to analyze the topological properties of the multi-instanton solution. We then discuss the generalization of the topology of instantons and monopoles to arbitrary dimensional spaces and clarify their relationship.

In the next section, the gauge types of the (singularity free) Yang–Mills field in four-dimensional Euclidean space is discussed. It is shown that the second Chern class[5] characterizes the gauge types. The multi-instanton solution is discussed in Sec. 3. In the final section, we discuss the generalization of the topology of instantons and monopoles to arbitrary dimensions. In this paper, we only consider the gauge fields; in bundle language, only principal fibre bundle. The gauge group G is assumed to be semisimple.

2. THE GAUGE TYPES AND THE CHERN CLASS

We consider the Yang–Mills fields in compactified four-dimensional Euclidean space E^4, which is conformally equivalent to the four-dimensional sphere S^4 embedded in five-dimensional Euclidean space.

First consider a subspace X of $E^4 + \infty = S^4$.

Theorem 1: Let X be a subspace which has a countable covering and is contractible on itself to a point. Then there exists a global gauge transformation such that all the transition functions defined on the overlapping regions in X become the identity transformation.

It is known[3] that any fibre bundle with a base space which has a countable covering and is contractible on itself is a trivial bundle. The trivial bundle (or product bundle) is a bundle which has only one coordinate neighborhood. Therefore, there must be a global gauge transformation which satisfies the above property.

Thus, locally the gauge type is always trivial. The compactified Euclidean space (S^4) is covered by two hemispheres R_1 and R_2 of S^4, whose overlapping region is essentially $S^3 = E^3 + \infty$. Since the hemisphere satisfies the assumptions of Theorem 1, we have the following result:

Theorem 2: In the compactified Euclidean space, a gauge is characterized by choosing two regions R_1 and R_2 whose union covers the whole space and specifying the transition function g_{12} on the overlapping region $R_1 \cap R_2$ ($\simeq S^3 = E^3 + \infty$).

A simple example of these theorems is the Möbius band which is a fibre bundle with the base S^1, the fibre a line segment, and the group a cyclic group of order two. If we only look at a portion of the Möbius band, we cannot discriminate the topological structure of the Möbius band from that of the nontwisted band. The "gauge type" of the Möbius band is characterized by dividing the band into two open bands and specifying the group elements for the overlapping region (i.e., the

[a)] Work supported by the National Science Foundation under Grant No. PHY75-07376 A01 and in part by the Research Foundation of the City University of New York under Grant No. 11118.

[b)] On leave from Department of Physics, Hokkaido University, Sapporo, Japan.

identity transformation to the one end of an open band and the reflection transformation of the line segment to the other end).

Next, we explain the following very important theorem,[3] which is a basis of our work.

Theorem 3: The gauge type in the compactified Euclidean space S^4 is uniquely characterized by the homotopy class $\pi_3(G)$ of the transition function g_{12} as a mapping from S^3 to the group manifold if G is arcwise connected.

Suppose that two transition functions $g_{12}^{(a)}$ and $g_{12}^{(b)}$ are of the same gauge type. Then there exist transformations $g_1(x)$ and $g_2(x)$ which are well defined on R_1 and R_2, respectively, such that

$$g_{12}^{(a)}(x)=g_2^{-1}(x)g_{12}^{(b)}(x)g_1(x), \quad x\in R_1\cap R_2.$$

Since R_1 and R_2 are contractible on themselves to one point and G is arcwise connected, $g_1(x)$ and $g_2(x)$ are homotopically equivalent to the identity transformation. Thus, $g_{12}^{(a)}(x)$ and $g_{12}^{(b)}(x)$ belong to the same homotopy class. Conversely, if $g_{12}^{(a)}$ and $g_{12}^{(b)}$ belong to the same homotopy class, $g_{12}^{(a)-1}g_{12}^{(b)}$ belongs to the trivial homotopy class. Then $g_{12}^{(a)-1}g_{12}^{(b)}$ can be extended to a transformation well defined on R_1. Denote the extended transformation by g on R_1. Consider the gauge specified by $g_{12}^{(a)}$. Perform the gauge transformation $g(x)$ in R_1. Then after the transformation, the transition function becomes $g_{12}^{(b)}(x)$. Thus the gauge specified by $g_{12}^{(a)}$ and $g_{12}^{(b)}$ are of the same gauge type.

If the group is not arcwise connected, the gauge type is uniquely characterized by the equivalence classes of $\pi_3(G)$ under the translation $[\simeq\pi_0(G)]$ of the fixed point.[3] Again a simple example is provided by the Möbius band, where the "gauge group" is the cyclic group of order two consisting of the identity transformation and the reflection of the line segment. The gauge type is characterized by $\pi_0(G)=Z_2$. Thus there are two gauge types. The nontrivial one is the Möbius band, and the trivial one is the ordinary nontwisted band.

As we shall discuss in the final section, from the point of view of fibre bundle, the topological characterization of both monopoles and instantons can be based on Theorem 3.

We now introduce the gauge covariant field $F_{\mu\nu}$:

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu+[A_\mu,A_\nu]. \tag{2.1}$$

Theorem 4: Under the same assumptions as Theorem 3, the gauge types are characterized by the second Chern class[5] integrated over the whole space:

$$q=\int c_2, \tag{2.2}$$

where

$$c_2=(1/8\pi^2)\,\mathrm{Tr}(\Omega\wedge\Omega) \tag{2.3}$$

$$\Omega=\tfrac{1}{2}F_{\mu\nu}dx_\mu\wedge dx_\nu. \tag{2.4}$$

To verify this, we need the well-known identity[7]

$$c_2=\frac{1}{32\pi^2}\epsilon_{\mu\nu\alpha\beta}\,\mathrm{Tr}(F_{\mu\nu}F_{\alpha\beta})d^4x=(1/8\pi^2)\partial_\mu J_\mu{}^A,$$

$$J_\mu{}^A=\epsilon_{\mu\nu\alpha\beta}\,\mathrm{Tr}(A_\nu\partial_\alpha A_\beta+\tfrac{2}{3}A_\nu A_\alpha A_\beta).$$

Then we have

$$\begin{aligned}\int c_2&=\int_{R_1}c_2+\int_{R_2}c_2-\int_{R_1\cap R_2}c_2\\&=\frac{1}{8\pi^2}\int_{S^3,\ R_1\cap R_2}(J_\mu{}^{A^{(1)}}-J_\mu{}^{A^{(2)}})d\sigma^\mu\\&=-\frac{1}{24\pi^2}\int_{S^3}\epsilon_{\mu\nu\alpha\beta}\,\mathrm{Tr}(g_{12}^{-1}\partial_\nu g_{12}g_{12}^{-1}\partial_\alpha g_{12}g_{12}^{-1}\partial_\beta g_{12})d\sigma^\mu\end{aligned} \tag{2.5}$$

$$=\text{homotopy class number }\pi_3(g_{12})\text{ of the transition function }g_{12}\text{ in the overlapping region }R_1\cap R_2. \tag{2.6}$$

This proves Theorem 4 because of Theorem 3. If $\pi_3(G)=Z$, the second Chern class number uniquely characterizes the gauge types. Compare this theorem with Theorem 8 of Wu and Yang.[1]

Theorem 2 assures us that we can always arrange the gauge field on *compactified* Euclidean space to be singularity free by choosing, at most, two regions. However, for the gauge fields with a large Chern class number, it is practically convenient to provide regions more than two. (For an explicit example, see the next section.) Then the gauge type is characterized by the sum of the (suitably oriented) homotopy class numbers for each transition function.

$$q=\sum_{R_i\cap R_j}(\text{homotopy class number of the transition functions }g_{ij}) \tag{2.7}$$

3. TOPOLOGY OF THE MULTI-INSTANTON SOLUTION

First we summarize the solution of the Euclidean[7] SU(2) Yang–Mills field found by 't Hooft[8] and generalized by Jackiw, Nohl, and Rebbi[9]:

$$A_\mu=\tfrac{1}{2}i\sigma_a A_\mu{}^a, \tag{3.1a}$$

$$A_\mu{}^a=\eta_{a\mu\nu}\partial_\nu\ln\rho, \tag{3.1b}$$

$$\rho=\sum_{i=1}^{n}\frac{\lambda_i^2}{(x-x_i)^2}, \tag{3.1c}$$

where

$$\eta_{a\mu\nu}=\epsilon_{4a\mu\nu}-\tfrac{1}{2}\epsilon_{abc}\epsilon_{bc\mu\nu}. \tag{3.2}$$

(3.2) satisfies

$$(\partial_\mu\omega)\omega^{-1}=\tfrac{1}{2}i\sigma_a\eta_{a\mu\nu}2x_\nu/x^2, \tag{3.3a}$$

$$\omega=(x_4+ix_a\cdot\sigma_a)/\sqrt{x^2}. \tag{3.3b}$$

Our convention is $\epsilon_{4123}=1$. (3.1c) is a general solution of the equation

$$(\Box\rho)/\rho=0, \tag{3.4}$$

which is equivalent to the self-duality condition[7] on $F_{\mu\nu}$ under the ansatz (3.1b).

By using (3.3), (3.1) can be rewritten as

$$A_\mu=-\frac{1}{\rho}\sum_{i=1}^{n}\frac{\lambda_i^2}{(x-x_i)^2}(\partial_\mu\omega_i)\omega_i^{-1}, \tag{3.5a}$$

where

$$\omega_i=[(x-x_i)_4+i(x-x_i)_a\sigma_a]/[(x-x_i)^2]^{1/2}. \tag{3.5b}$$

The gauge covariant tensor $F_{\mu\nu}$ is given by

$$F_{\mu\nu} = i\frac{2}{\rho}\left(\sum_{i=1}^{n}\frac{\lambda_i^2}{(x-x_i)^4} - \frac{1}{\rho}\sum_{i,j}^{n}\frac{\lambda_i^2\lambda_j^2(x-x_i\cdot x-x_j)}{(x-x_i)^4(x-x_j)^4}\right)\eta_{a\mu\nu}\sigma_a. \tag{3.6}$$

It is easy to see that $F_{\mu\nu}$ is *not* singular at $x=x_i$, although the gauge field (3.5) is singular at $x=x_i$. This is because the singularities contained in (3.5) are only pure gauge. Therefore, each singularity can be eliminated by an appropriate gauge transformation. For the study of global property of this solution, the formulation of the last section provides an ideal framework.

Divide space into n regions $R_1,\ldots,R_n$ such that each region R_i contains only one singular point x_i and the overlapping regions meet only at infinity. We arrange the ordering of the regions so that the only adjacent regions R_i and R_{i+1} overlap. In each region R_i, we define the gauge field

$$A_\mu^{(i)} = \frac{1}{\rho}\sum_{l\neq i}\left(\frac{\lambda_l^2}{(x-x_l)^2}\omega_i^{-1}\partial_\mu\omega_i - \frac{\lambda_l^2}{(x-x_l)^2}\omega_i^{-1}(\partial_\mu\omega_l)\omega_l^{-1}\omega_i\right), \tag{3.7}$$

which has no singularity in R_i although w_i is singular at $x=x_i$. For each overlapping region $R_i\cap R_{i+1}(\simeq S^3)$ we specify the transition function

$$g_{i,i+1} = \omega_{i+1}^{-1}\omega_i. \tag{3.8}$$

Then we have

$$A_\mu^{(i)} = g_{i,i+1}^{-1}A_\mu^{(i+1)}g_{i,i+1} + g_{i,i+1}^{-1}\partial_\mu g_{i,i+1}. \tag{3.9}$$

We see that $g_{i,i+1}\to 1$ and $A_\mu^{(i)}\to O(1/|x|^2)$ as $|x|\to\infty$. This property conforms our S^4 picture. The field $A_\mu^{(i)}$ give the field strength which is equivalent to (3.6) because it is related to the original singular form (3.5) by a singular gauge transformation

$$A_\mu^{(i)} = \omega_i^{-1}A_\mu^{TJNR}\omega_i + \omega_i^{-1}\partial_\mu\omega_i. \tag{3.10}$$

The global gauge type of this solution is characterized by the second Chern class number. By (2.7), we have

$$q = \sum_{i=1}^{n-1}(\text{homotopy class number of } g_{i,i+1}) = n-1 \tag{3.11}$$

since the homotopy class number (or winding number) of $g_{i,i+1}$, as we shall see below, is one for each i. Choose a coordinate system such that the singular points x_i and x_{i+1} lie on the $x_4(\equiv t)$ axis with $t_i>0$ and $t_{i+1}<0$. For simplicity, first suppose that the overlapping region $R_i\cap R_{i+1}$ is a thin hyperplane (or rather hypersphere S^3 because of compactification) between t_{i+1} and t_i. Then we have

$$g_{i,i+1} = \{[r^2+(t-t_{i+1})^2][r^2+(t-t_i)^2]\}^{-1} \times[(t-t_{i+1})(t-t_i)+r^2-ix_a\cdot\sigma_a(t_{i+1}-t_i)] \quad (r^2=x_1^2+x_2^2+x_3^2). \tag{3.12}$$

For any hyperplane between t_{i+1} and t_i, we have

$$g_{i,i+1}\to 1 \quad \text{as } r\to\infty$$

and

$$g_{i,i+1}\to -1 \quad \text{as } r\to 0.$$

Thus, every point on the SU(2) group manifold is covered by $g_{i,i+1(x)}$ once and only once when x sweeps the hyperplane. The homotopy class number is one. Since any continuous deformation of the overlapping region induces a continuous change in $g_{i,i+1}$ as long as it does not cross the singularities of w_i and w_{i+1}; the homotopy class number for any overlapping region defined above is 1.

The result $q=n-1$ was obtained by Jackiw, Nohl, and Rebbi by direct calculation of (2.2). Our derivation clarifies the intrinsic topological property of the solution.

From this point of view, the occurrence of tunneling[10] from one classical vacuum to a gauge rotated one corresponds to the occurrence of homotopically nontrivial transition functions in some region of space–time. If we adopt this picture for tunneling, it is quite easy to modify the field configuration (3.5) which represents n instantons (or n anti-instantons) to a possible configuration with any mixture of instantons and anti-instantons by taking the inverse of the transition functions corresponding to "antitunneling." Of course, after doing this, the field configuration is not the solution of the field equations, while the action remains finite. We only give the simplest example of an instanton–anti-instanton configuration,

$$A_\mu = -\frac{1}{\rho}\left(\frac{\lambda_1^2}{(x-x_1)^2}(\partial_\mu\omega_1)\omega_1^{-1} + \frac{\lambda_2^2}{(x-x_2)^2}(\partial_\mu\omega_2)\omega_2^{-1} + \frac{\lambda_3^2}{(x-x_3)^2}(\partial_\mu\omega_{32})\omega_{32}^{-1}\right), \tag{3.13}$$

where $\omega_{32}=\omega_2\omega_3^{-1}\omega_2$. (3.13) corresponds to two-transition functions g_{12} and g_{23}^{-1}. It is of some interest to study the interaction between instanton and anti-instanton, which, however, is not the subject of this paper.

Our picture that the vacuum tunneling corresponds to the homotopically nontrivial transition functions is equivalent to the description[10] given by Jackiw and Rebbi and by Callan, Dashen, and Gross, if the overlapping regions are moved to infinity (or, in the S^4 picture, to the "north pole"). The original BPST[7] solution takes this form. In general Theorem 1 ensures that is always possible. Practically, however, performing such transformations is not simple for the multi-instanton solution and may not be necessary.

4. INSTANTONS AND MONOPOLES IN d-DIMENSIONAL SPACE

The discussions of Sec. 2 can be extended to arbitrary dimensional space. Suppose that the d-dimensional Euclidean space is conformally mapped onto S^d. Then, Theorem 3 generalized to d dimensions shows that the gauge types of the singularity free gauge fields are characterized by the homotopy class, $\pi_{d-1}(G)$, of the transition functions defined on the overlapping region S^{d-1}. If the dimension is even, the homotopy class can be described by the $(d/2)$th Chern class[5] defined by

$$[(-1)^{d/2}/(2\pi i)^{d/2}(d/2)!]\sum\delta_{i_1\cdots i_{d/2}}^{j_1\cdots j_{d/2}}\Omega_{j_1}^{i_1}\wedge\cdots\wedge\Omega_{j_{d/2}}^{i_{d/2}}, \tag{4.1}$$

where the summation is taken over all ordered subsets $(i_1,\ldots,i_{d/2})$ of $d/2$ elements from $(1,\ldots,r)$ and all permutations $(j_1,\ldots,j_{d/2})$ of $(i,\ldots,i_{d/2})$, the symbol δ denotes the sign of the permutation, r is the dimension of the matrix of the gauge group. Ω is the curvature form in matrix representation defined by (2.4). For a semisimple group, the proof is a straightforward extension of that for Theorem 4. If the dimension is odd, we have no characterization of the homotopy class $\pi_{d-1}(G)$ *in terms of the curvature form.* This is because any characteristic class expressible in terms of $F_{\mu\nu}$ must be an even form.

Next, consider a $(d-1)$-dimension sphere in d-dimensional Euclidean space and consider the restriction of the Yang–Mills field on this subspace. Thus, we are considering a relative bundle[3] of the original bundle. By applying Theorem 3 to this subspace, we conclude that the gauge of the relative bundle is determined by choosing two hemispheres covering S^{d-1} and by specifying the transition function for the overlapping region $(\simeq S^{d-2})$. Hence, the gauge type of the relative bundle is characterized by $\pi_{d-2}(G)$. If the dimension d is even, there is no characterization of $\pi_{d-2}(G)$ in terms of the curvature form because $d-1$ is odd. If the dimension is odd, on the other hand, it is characterized by the $[(d-1)/2]$th Chern class which is defined by (4.1) with $d/2$ being replaced by $(d-1)/2$. It is a generalization of magnetic flux in three dimensions, which is the first Chern class for the U(1) bundle, to higher dimensions. It is easy to see that the integral of the $[(d-1)/2]$th Chern class over the sphere S^{d-1} does not depend on the choice of the sphere as long as it does not cross a singularity. Thus, if the dimension is odd, the $[(d-1)/2]$th Chern class can be used as a characterization of the gauge type. If the gauge field has nontrivial $(d-1)/2$-th Chern class, its behavior at infinity is not pure gauge.[11] Hence, we cannot compactify the space and the topology of the base space is different from S^d.

We define an instanton as a classical field configuration whose nontrivial topological characterization is $\pi_{d-1}(G)$ in d-dimensional Euclidean space. We define a monopole[12] as a (possibly singular) classical field configuration whose nontrivial topological characterization is $\pi_{d-2}(G)$ in d-dimensional Euclidean space.

Thus, in our definition the appropriate pseudoparticle is the instanton if the space–time dimension is even and is the monopole if it is odd, as long as we require that the pseudoparticle be characterized by some topological charge in terms of $F_{\mu\nu}$.

The magnetic monopole in three dimensions is characterized by $\pi_1(G)$[13] which, in Wu and Yang's terminology, is the class circuit.[1] It is well known that $\pi_1(\mathrm{SU}(n))=0$ and $\pi_1(\mathrm{SO}(n))=Z_2$ $(n\geq 3)$. Therefore, there is no monopole in the SU(n) Yang–Mills theory in three dimensions according to our definition. In fact, the first Chern class identically vanishes for SU(n). For SO(n), in general, there is no characterization of $\pi_1(\mathrm{SO}(n))$ in terms of the curvature tensor, while there are only two types of class circuits which cannot be globally transformed into each other. For SO(3), this was explicitly demonstrated by Wu and Yang.[1] If we consider the generalization of magnetic monopole to five-dimensional space, $\pi_3(G)$, and hence the second Chern class is the topological characterization of the monopole. It is known[3] that $\pi_3(\mathrm{SU}(n))=\pi_3(\mathrm{Sp}(n))=Z$ for $n\geq 2$ and $\pi_3(\mathrm{SO}(n))=Z$ for $n\leq 5$. [We note that for SO(n) the first Pontjagin class[5] is the appropriate characterization of the monopole in terms of the curvature tensor in five dimensions.] Then, the five-dimensional monopole may have infinite number of gauge types. This is also true for the four-dimensional instanton.

After the completion of this work, the author came to know, at the "Five Decades in Weak Interaction" Symposium (City College, Jan. 21–22, 1977), that Professor C. N. Yang generalized the monopole to five-dimensional space and obtained spherically symmetric classical solutions.

ACKNOWLEDGMENTS

The author would like to thank Professor B. Sakita for his kind interest in this work and valuable comments on the manuscript. The author also thanks S. Wadia for his help in correcting the manuscript.

[1] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[2] T. T. Wu and C. N. Yang, Phys. Rev. D **14**, 437 (1976); Nucl. Phys. B **107**, 365 (1976).
[3] See, e.g., N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, New Jersey, 1951); D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[4] Besides Refs. 1 and 2, the following are the examples of recent physics literatures which use the fibre bundle picture: L. N. Chang, K. J. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976); Y. M. Cho, J. Math. Phys. **16**, 2029 (1975), Z. F. Ezawa and H. C. Tze, J. Math. Phys. **17**, 2228 (1976).
[5] See, e.g., S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry, Vol. II* (Interscience, New York, 1969) J. W. Milnor and J. D. Stasheff, *Characteristic Classes* (Princeton U. P., Princeton, New Jersey, 1974). The second Chern class number is the same as the quantity which is referred to as the Pontjagin index in Ref. 7. Since we are generally discussing complex bundles, the Chern class is an appropriate term.
[6] See the table given in Wu and Yang[1] for the translation of terminology.
[7] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[8] G. 't Hooft, unpublished.
[9] R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[10] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976) and Harvard preprint; R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976); C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. B **63**, 334 (1976); A. M. Polyakov, Copenhagen preprint (1976).
[11] This is easy to see by a dimensional consideration.
[12] The term "instanton"[10] means the object which is localized in space and time. Hence, a "monopole" in our definition can be an instanton if we continue the $[(d-1)+1]$-dimensional space–time to (odd) d-dimensional Euclidean space. My intention here is to emphasize an important difference of the topology of the Yang–Mills fields between even and odd dimensions.
[13] For the cases where Higgs fields are included, see J. Arafune, P. G. O. Freund, and C. J. Goebel, J. Math. Phys. **16**, 433 (1975); Yu S. Tyupkin, J. A. Fateev, and A. S. Schwartz, Zh. Eksp. Teor. Fiz. Pis. Red. **21**, 91 (1975) [JETP Lett. **21**, 42 (1975)]; M. I. Monastysskii and A. M. Porelormov, *ibid.*, 94 (1974) [*ibid.*, 43 (1975)].

Notes on the symmetries of systems of differential equations

F. Gonzalez-Gascon

Instituto de Estructura de la Materia del CSIC y GIFT, Serrano 119, Madrid 6, Spain
(Received 29 November 1976)

The concept of symmetry of the solutions of a system of differential equations is clarified. The functional character of the symmetry transformations is stressed in contrast with the pointlike character of the ordinary transformations considered by Lie. It is shown that any differential equation of arbitrary order possesses infinitely many symmetries, in strong contrast with a general theorem denying the existence of pointlike transformations of symmetry for an arbitrary differential equation of order greater than one. The relevance of local differential symmetries in theoretical mechanics is discussed, and some unsolved questions are raised.

I. INTRODUCTION

In a recent paper[1] Anderson *et al.* briefly discussed the great usefulness of introducing general transformations of the kind:

$$\begin{aligned} &\bar{x}^i = f^i(x, u, \partial_I u, \partial_I \partial_p u, \cdots), \\ &\bar{u}^j = g^j(x, u, \partial_I u, \partial_I \partial_p u, \cdots), \\ &i, l, p = 1, \ldots, m, \quad j = 1, \ldots, n, \end{aligned} \tag{1}$$

in the framework of the partial differential equations

$$P_k(x, u, \partial_i u, \partial_i \partial_I u, \cdots) = 0, \quad k = 1, \ldots, n, \tag{2}$$

$x_1, \ldots, x_m$ being the independent variables and $u_1, \ldots, u_n$ being the unknown functions.

The transformations (1) are important in the context of nonlinear wave phenomena, solitons, etc.[2]

We give additional information concerning the relevance of the transformations (1) for the particular case of the theory of ordinary differential systems (Sec. IV) and classical mechanical systems (Sec. V). We shall see that transformations similar to (1) arise in a natural way when studying the symmetries of a system of ordinary differential equations (Secs. II and III). These symmetries were first considered by Lie[3] in a more restricted framework than the one presented here.

II. THE SYMMETRIES OF A SYSTEM OF DIFFERENTIAL EQUATIONS

Assume that we are given a certain system of ordinary differential equations,

$$D_i(x; y_1, \ldots, y_1^{(s_1)}; \ldots; y_n, \ldots, y_n^{(s_n)}) = 0, \quad i = 1, \ldots, n, \tag{3}$$

x being the independent variable, $y_1, \ldots, y_n$ the unknown functions, and $y_1^{(s_1)}, \ldots, y_n^{(s_n)}$ the derivatives of maximum order appearing in (3).

We shall call symmetry of (3) any prescription or rule s permitting us to transform *any* generic solution $y_1(x), \ldots, y_n(x)$ of Eq. (3) into another solution $s \circ y_i(x)$, $i = 1, \ldots, n$, of Eq. (3). That is, s is an operator whose domain of definition is the set S consisting of the formal solutions of (3).

By defining the composition of symmetries s_1 and s_2 in the standard way, it is obvious that the set S of all the symmetries of (3) is an abstract monoid (semigroup with identity).[4]

The symmetries considered by Lie in relation to the differential equations were of a very restricted kind. In fact, for the case of a single differential equation,

$$D(x, y, y', \ldots, y^{(n)}) = 0, \tag{4}$$

he only considered the possibility of transforming the solution of (4) by infinitesimal transformations of the kind

$$\bar{x} = x + \epsilon \eta(x, y), \quad \bar{y} = y + \epsilon \theta(x, y), \tag{5}$$

together with the obvious transformations induced by (5) on the successive derivatives $\bar{y}', \bar{y}'', \cdots$. In short, Lie only considered the symmetries of (4) *induced* by infinitesimal point transformations between the x and y variables.

On the other hand, the symmetries that we are speaking about have nothing to do, in general, with pointlike transformations, since they have as domain the functional space of the solutions of (4).

Before proceeding, it is useful to give examples of differential equations admitting symmetry transformations that cannot be reduced to the infinitesimal form (5): The linear differential equation

$$y' = F(x) \cdot y, \tag{6}$$

$F(x)$ being a periodic function of period T, obviously admits the global symmetry

$$s_T \circ y(x) = y(x + T),$$

which, of course, has no infinitesimal counterpart.

A much more significant example is furnished by the transformation

$$s_d \circ y(x) = y'(x),$$

which is, of course, a symmetry possessed by the solutions of any linear and homogeneous differential equation with constant coefficients a_n,

$$\sum_{n=1}^{N} a_n y^{(n)} = 0. \tag{7}$$

One can, of course, be led (for a concrete differential equation) to a very complicated prescription giving the analytical description of a symmetry. For instance, a recipe such as

$$s_K \circ y(x) = \int_0^x K(x, y(x+1), y'(x-2), \ldots, y^{(n}(x^2))dx\,, \quad (8)$$

could arise as a possible symmetry of the solutions of a particular Eq. (4). In order to avoid the difficulties of treatment of the global symmetries like (8) we shall confine our attention to the study of a particular subclass S_d of symmetries: the local differential symmetries.

III. THE LOCAL DIFFERENTIAL SYMMETRIES

In the following we shall consider that (3) can be written in the canonical form,

$$y_i^{(s}i = F_i(x; y_1, \ldots, y_1^{(s_1-1)}; \ldots; y_n; \ldots, y_n^{(s_n-1)})\,, \quad i = 1, \ldots, n\,. \quad (9)$$

A symmetry of (9) will be called local differential if the prescriptions used in order to transform any solution of (9) have the form

$$\bar{y}_i = f_i(x; y_1, \ldots y_1^{(s_1-1)}; \ldots; y_n; \ldots, y_n^{(s_n-1)})\,, \quad \bar{x} = f_0(x; y_1, \ldots, y_1^{(s_1-1)}; \ldots; y_n, \ldots, y_n^{(s_n-1)})\,. \quad (10)$$

That is, given any solution $\varphi_1(x), \ldots, \varphi_n(x)$ of (9) we have to substitute, in (10), $y_1, y_1', \ldots, y_n^{(s_n-1}$ by $\varphi_1(x)$, $\varphi_1'(x), \ldots, \varphi_n^{(s_n-1)}(x)$; in that way we get

$$\bar{y}_i = {}^{\varphi}f_i(x), \quad \bar{x} = {}^{\varphi}f_0(x)\,. \quad (11)$$

The transformed solution $\Psi_1(\bar{x}), \ldots, \Psi_n(\bar{x})$ is now obtained by elimination of the x variable between Eqs. (11).

At this point the reason why we have not introduced higher derivatives of $y_1, \ldots, y_n$ into (10) should be clear: Had we done this, the differential equation (9) would reduce this apparently more general prescription to the prescription given by Eqs. (10), just by substituting $y_i^{(t}i$ (for any $t_i \geq s_i$) by its value obtained from (9).

In order to get the transformation formulas for the derivatives $d\bar{y}_1/d\bar{x}, \ldots, d\bar{y}_n/d\bar{x}, \ldots, d^n\bar{y}_i/d\bar{x}^n$, one simply has to differentiate Eqs. (10), taking into account the dynamical equations (9) every time that in the process of differentiation a derivative $y_i^{(t}i$ $(t_i \geq s_i)$ appears.

IV. MATHEMATICAL CONSEQUENCES OF THE LOCAL DIFFERENTIAL SYMMETRIES

It is well known[6] that every first order differential equation

$$y' = F(x, y) \quad (12)$$

admits infinitely many monoparametric groups of transformations having the infinitesimal representation

$$\bar{x} = x + \epsilon\eta(x, y), \quad \bar{y} = y + \epsilon\theta(x, y)\,, \quad (13)$$

so that by means of (13) any solution of (12) is transformed into another solution of (12).

We point out that the monoparametric group structure of (13) is not necessary. In fact, a monoparametric monoid structure is sufficient.[7] The reason for always speaking, in this context, of monoparametric groups of transformations is that it can be shown[7] that any monoparametric monoid of pointlike transformations such as (13) contains, automatically, the inverse of every transformation contained in (13).

In contrast with the above result concerning the first order differential equations, a differential equation of order $n > 1$ only exceptionally admits continuous groups of transformations. In particular, if a second order differential equation admits continuous groups of symmetries like (13), these monoparametric groups generate a continuous Lie group having no more than eight parameters.[8] This is what happens to the differential equation:

$$y'' = 0\,,$$

admitting the projective group of the (x, y) plane as a group of symmetries which possesses just eight *essential* parameters.

A second order differential equation not admitting an infinitesimal transformation like (13) is

$$y'' = x^2 + y^2\,. \quad (14)$$

Indeed, we get from (13), to the first order in ϵ,

$$\begin{aligned}\bar{y}' &= \frac{d\bar{y}}{d\bar{x}} = y' + \epsilon[\theta_{,x} + y'(\theta_{,y} - \eta_{,x}) - y'^2(x^2 + y^2)] \\ &= y' + \epsilon\alpha(x, y, y')\,, \\ \bar{y}'' &= \frac{d\bar{y}'}{d\bar{x}} \\ &= y'' + \epsilon[\theta_{,xx} + (2\theta_{,xy} - \eta_{,xx})y' + (\theta_{,yy} - 2\eta_{,xy})y'^2 \\ &\quad - \eta_{,yy}\cdot y'^3 - 3\eta_{,y}\cdot y'\cdot y'' + (\theta_{,y} - 2\eta_{,x})\cdot y''] \\ &= y'' + \epsilon\beta(x, y, y', y'')\,,\end{aligned} \quad (15)$$

$\theta_{,x}, \theta_{,y}, \theta_{,xx}, \cdots$ being the partial derivatives $\partial\theta/\partial x$, $\partial\theta/\partial y$, $\partial^2\theta/\partial x^2, \cdots$.

If (13) is a symmetry, we must have

$$\bar{y}'' = \bar{x}^2 + \bar{y}^2\,. \quad (16)$$

To the first order in ϵ, Eq. (16) gives

$$\beta = 2x\eta + 2y\theta\,. \quad (17)$$

By substituting (15) into (17), taking (14) into account, and equating the different powers of y' to zero, we immediately get:

$$\begin{aligned}&\eta_{,yy} = 0\,, \\ &\theta_{,yy} - 2\eta_{,xy} = 0\,, \\ &2\theta_{,xy} - \eta_{,xx} - 3(x^2 + y^2)\eta_{,y} = 0\,, \\ &\theta_{,xx} + (\theta_{,y} - 2\eta_{,x})\cdot(x^2 + y^2) - 2x\eta - 2y\cdot\theta = 0\,.\end{aligned} \quad (18)$$

It is now very easy to show[9] that Eqs. (18) only possess the unique trivial solutions:

$$\theta(x, y) = \eta(x, y) \equiv 0\,.$$

Note that (13) is nothing more than the infinitesimal form of a monoparametric monoid of transformations of

the type in (10) for the particular case of $n=1$ and $s_1=1$. Similarly, the differential equation in (14) is a particular case of (9), obtained with $n=1$ and $s_1=2$. For this particular case, the infinitesimal form of a monoparametric monoid of transformations of symmetry is suggested by Eq. (10) with $n=1$ and $s_1=2$, that is

$$x=x+\epsilon\eta(x,y,y'),\quad y=y+\epsilon\theta(x,y,y'). \tag{19}$$

From (14) and (19) we easily get

$$\begin{aligned}
&\frac{d\bar{y}}{d\bar{x}}=y'+\epsilon\left(\frac{d\theta}{dx}-y'\frac{d\eta}{dx}\right)=y'+\epsilon\alpha(x,y,y'),\\
&\frac{d^2\bar{y}}{d\bar{x}^2}=\frac{d\bar{y}'}{d\bar{x}}=y''+\epsilon\left(\frac{d\alpha}{dx}-y''\frac{d\eta}{dx}\right)\\
&\qquad=y''+\epsilon\beta(x,y,y'),
\end{aligned} \tag{20}$$

where

$$\begin{aligned}
&\frac{d\theta}{dx}=\theta_{,x}+\theta_{,y}\cdot y'+(x^2+y^2)\cdot\theta_{,y'},\\
&\frac{d\eta}{dx}=\eta_{,x}+\eta_{,y}\cdot y'+(x^2+y^2)\cdot\eta_{,y'},\\
&\frac{d\alpha}{dx}=\frac{d}{dx}\Big(\theta_{,x}+\theta_{,y}\cdot y'+(x^2+y^2)\cdot\theta_{,y'}\\
&\quad-(x^2+y^2)\cdot\frac{d\eta}{dx}-y'\cdot\frac{d}{dx}\eta_{,x}+\eta_{,y}\cdot y'\\
&\quad+\eta_{,y'}\cdot(x^2+y^2)\Big).
\end{aligned} \tag{21}$$

To the first order in ϵ Eq. (16) reduces to

$$\beta(x,y,y')=2x\cdot\eta(x,y,y')+2y\cdot\theta(x,y,y'), \tag{22}$$

$\beta(x,y,y')$ being defined by Eqs. (20) and (21).

In contrast with (17), Eq. (22) is now a *single* partial differential equation of the second order relating the two functions $\eta(x,y,y')$ and $\theta(x,y,y')$. It is obvious that a similar partial differential equation for η and θ can be obtained for *any* differential equation of second order,

$$y''=F(x,y,y'). \tag{23}$$

Since a second order partial differential equation like (22) has infinitely many solutions, we get, for arbitrary second order differential equations like (23), a result similar to Lie's[3,6] concerning the first order differential equations: The existence of infinitely many monoparametric monoids of local differential symmetries for any second order differential equation. The only difference, in relation with Lie's results, is that the infinitesimal transformations admitted here are of the type in (19) instead of the particular type in (13).

The above conclusion for an arbitrary second order differential equation can be obviously extended to system (9) by substituting (10) by the infinitesimal counterpart,

$$\begin{aligned}
&\bar{y}_i=y_i+\epsilon\eta_i(x,y_1,\ldots,y_1^{(s_1-1)};\ldots;y_n,\ldots,y_n^{(s_n-1)}),\\
&\bar{x}=x+\epsilon\theta(x,y_1,\ldots,y_1^{s_1-1)};\ldots;y_n,\ldots,y_n^{(s_n-1)}).
\end{aligned} \tag{24}$$

From all this follows a similar behavior for any kind of differential equation or system of differential equations (with independence of its orders) in relation to the number of symmetries admitted by the set S of its solutions. In fact, as we have briefly explained, any system (9) admits infinitely many monoparametric monoids of transformations (24) acting on S. The differential systems with $s_1=\cdots=s_n=1$ are not, therefore, privileged in relation to the number of symmetries admitted by S. The only privilege is that (24) reduces to a monoparametric group of *pointlike* transformations when $s_i=1$.

V. THE LOCAL DIFFERENTIAL SYMMETRIES IN THEORETICAL MECHANICS

It is known that the differential equations of many problems of classical mechanics can be cast in the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{q}_i}\right)=\frac{\partial L}{\partial q_i}, \tag{25}$$

n being the number of degrees of freedom and $L(q_i,\dot{q}_i,t)$ being the Lagrangian.

Writing (25) explicitly, we have

$$\sum_{j=1}^{n}\frac{\partial^2 L}{\partial\dot{q}_i\partial\dot{q}_j}\ddot{q}_j+\frac{\partial^2 L}{\partial\dot{q}_i\partial q_j}\dot{q}_j+\frac{\partial^2 L}{\partial\dot{q}_i\partial t}=\frac{\partial L}{\partial q_i}, \tag{26}$$

and assuming that in a certain region

$$\det\left(\left(\frac{\partial^2 L}{\partial\dot{q}_i\partial\dot{q}_j}\right)\right)\neq 0,$$

we can write Eqs. (26) in the canonical form

$$\ddot{q}_i=F_i(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n,t). \tag{27}$$

Therefore, as we have explained in Sec. IV, the local differential symmetries associated with (27) will adopt the global form

$$\begin{aligned}
&\bar{q}_i=f_i(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t),\\
&\bar{t}=f_0(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t),
\end{aligned} \tag{28}$$

or the infinitesimal form

$$\begin{aligned}
&\bar{q}_i=q_i+\epsilon\eta_i(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t),\\
&\bar{t}=t+\epsilon\theta(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t),
\end{aligned} \tag{29}$$

which are usually referred to (by the physicists) as dynamical symmetries.

Two important kinds of transformations like (29) seem to have been used in the context of the Lagrange formulation of classical mechanics. Indeed, it is known that for the isotropic harmonic oscillator and for the Kepler problem, the additional first integrals

$$M_{\substack{kl\\k,l=1,2,3}}=\dot{q}_k\dot{q}_l+q_kq_l$$

and

$$\dot{\mathbf{q}}\wedge(\mathbf{q}\wedge\dot{\mathbf{q}})-\mathbf{q}/q,\quad \mathbf{q}=(q_1,q_2,q_3) \tag{30}$$

are related[10] with the gauge symmetries of L under the transformations

$$\bar{q}_j=q_j+(\epsilon/2)(\dot{q}_k\delta_{jl}+\dot{q}_l\delta_{jk}),\quad \bar{t}=t \tag{31a}$$

and

$$\bar{q}_i=q_i+\epsilon(\dot{q}_iq_k-\tfrac{1}{2}q_i\dot{q}_k-\tfrac{1}{2}\mathbf{q}\cdot\dot{\mathbf{q}}\delta_{ik}),\quad \bar{t}=t, \tag{31b}$$

which give, for the corresponding Lagrangians

$$L=\dot{\mathbf{q}}^2/2-\mathbf{q}^2/2\,,\quad L'=\dot{\mathbf{q}}^2/2+1/q\,, \tag{32}$$

the variations

$$\delta L=-\epsilon\frac{d}{dt}(q_iq_k)\,,\quad \delta L'=\epsilon\frac{d}{dt}\left(\frac{q_k}{q}\right). \tag{33}$$

What is still more important, it can be shown[11] that if $F(t,q_1\ldots,q_n,\dot{q}_1,\ldots,\dot{q}_n)$ is a first integral of (25), then the family of transformations

$$\bar{q}_i=q_i+\epsilon\eta_i(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t)\,,\quad \bar{t}=t \tag{34}$$

is a gauge symmetry of the corresponding L, where the functions η_i are given by the linear system

$$\sum_{i=1}^{n}\frac{\partial^2L}{\partial\dot{q}_i\partial\dot{q}_j}\eta_i=-\partial F/\partial\dot{q}_j\,. \tag{35}$$

This is, indeed, a very nice result connecting the first integrals of systems of differential equations like (25) (with $|\partial^2L/\partial\dot{q}_i\partial\dot{q}_j|\neq0$) with the infinitesimal gauge symmetries of L. This result shows, as well, that the symmetries of the formal type (29) are very useful to physicists. Nevertheless two warnings should be taken into account:

1. A gauge symmetry of L, like the one given by (29), is *not* in general a dynamical symmetry of the differential equations (25). In the particular case of systems like (25) it is known that every pointlike transformation,

$$q_i=g_i(Q_i,t),\quad t=t\,, \tag{36}$$

acting on the points of the configuration space, converts the differential equations (25) into another differential system equivalent to the Lagrangian system associated to L'.

If, in addition, transformation (36) is such that

$$L'(Q_i,\dot{Q}_i,t)=L(Q_i,Q_i,t)+\frac{d}{dt}\wedge(Q_i,t)\,, \tag{37}$$

then it is obvious that (36) is *not only* a gauge symmetry of any L, but it is, as well, a dynamical symmetry of the *solutions* of (25). Indeed, in the new coordinates $Q_1,\ldots,Q_n$ the new equations of motion are equivalent to

$$\frac{d}{dt}\left(\frac{\partial L'}{\partial\dot{Q}_i}\right)-\frac{\partial L'}{\partial Q_i}=0\,, \tag{38}$$

and by (37) the Eqs. (38) take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{Q}_i}\right)-\frac{\partial L}{\partial Q_i}=0\,,$$

as we desired to prove.

This particular coincidence of a gauge symmetry and a dynamical symmetry should *not* induce us to think in the existence of a general relation connecting these two completely different kinds of symmetry; the dynamical symmetries are not pointlike transformations [although in particular cases, such as (36) when Eq. (37) is fulfilled, they can be induced by pointlike transformations], but functional transformations acting on the set S of all the solutions of a given system of differential equations. They, therefore, have nothing to do at all with the gauge symmetries of a certain function L (in general not observable, neither in classical nor in quantum mechanics) existing only for a very restricted class of differential equations. These essential differences, frequently not clearly stated, do justify why the relationships (if any) connecting the monoparametric monoids of local differential symmetries and the first integrals of a *general* system of differential equations are still unknown.

2. We should not be induced to confuse the transformations (29), together with the transformations,

$$\bar{\dot{q}}_i=\dot{q}_i+\epsilon\alpha_i(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n)\,,\quad i=1,\ldots,n\,, \tag{39}$$

induced by (29) on $d\bar{q}_1/d\bar{t},\ldots,d\bar{q}_n/d\bar{t}$, with pointlike transformations on the "t-phase space" of coordinates $(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t)$. The canonical and canonoid transformations[12] are, as well, pointlike transformations acting on the standard phase space of classical mechanics and should, accordingly, be clearly distinguished from the prescriptions (29) and (39), having no meaning at all when acting on a single point $(q_1,\ldots,q_n;\dot{q}_1,\ldots,\dot{q}_n;t)$. This essential difference, as we have explained, does not preclude the possibility that a particular symmetry of Eq. (27) could be induced by a canonoid transformation,

$$Q_i=f_i(q_i,p_i,t)\,,\quad P_i=g_i(q_i,p_i,t)\,, \tag{40}$$

in the case that the new Hamiltonian K, obtained from (40), is equal [up to a function $f(t)$] to the old Hamiltonian $H(q,p,t)$ when written in the form $H(Q,P,t)$; that is,

$$K(Q,P,t)=H(Q,P,t)+f(t)\,. \tag{41}$$

In this case it is obvious that the canonoid transformation (40) changes any solution $q(t),p(t)$ of

$$\dot{q}_i=\frac{\partial H}{\partial p_i}\quad \dot{p}_i=-\frac{\partial H}{\partial\dot{q}_i}\,, \tag{42}$$

into a new solution $Q(t),P(t)$ of the same Eqs. (42), since we have

$$\dot{Q}_i=\frac{\partial K}{\partial P_i}\,,\quad \dot{P}_i=-\frac{\partial K}{\partial Q_i} \tag{43}$$

and because of (41), Eqs. (43) take the form

$$\dot{Q}_i=\frac{\partial H}{\partial P_i}\,,\quad \dot{P}_i=-\frac{\partial H}{\partial Q_i}\,. \tag{44}$$

At this point it is very easy to understand why the connection between monoparametric groups of dynamical symmetries and first integrals of *general* systems of differential equations is, geometrically at least, far from being clear. Indeed, the t-phase space associated with a general (not necessarily Lagrangian) system of differential equations like (27) has topological dimension $2n+1$. The first integrals of (27) are manifolds of dimension $2n$. On the other hand, given a solution $q_i(t)$ of (27) and a monoparametric group of dynamical symmetries of (27), the action of this group on appropriate solution $q_i(t)$ generates a partition of the t-phase space into two-dimensional manifolds, that can be labeled by the local parameters t and the parameter ϵ labeling the elements of the monoparametric group of symmetries. Therefore, unless $n=1$, the partitions of the t-phase space $(q_1,\ldots,q_n;q_1,\ldots,q_n;t)$, defined by giving a first

integral of a *general* system of differential equations and by giving a monoparametric group of dynamical symmetries admitted by the system, have *nothing to do* with each other. The reciprocity existing, at least for the Lagrangian systems, between first integrals and dynamical symmetries,[11] appears, therefore, as a very striking fact which should be studied more in order to further clarify this striking connection.

In the case $n=1$ (mechanical system with one degree of freedom) the dimension of the partitions defined by a first integral and by a monoparametric group of symmetries of the differential equation,

$$\ddot{q}_1 = F(q_1, \dot{q}_1, t)\,, \tag{45}$$

coincide. It is very interesting to remark that this seems to be *not* a purely fortuitous fact, of purely geometrical nature, having no other dynamical consequences. In fact, Havas[13] has shown that any differential equation like (45) is equivalent (has the same solutions) to other differential equations that can be obtained from a Lagrangian.

VI. CONCLUSIONS AND FINAL REMARKS

The natural way in which the local differential symmetries arise, in the framework of the differential systems, has been stressed. We have pointed out the conceptual differences between a dynamical symmetry and an ordinary pointlike transformation in configuration or phase space. We have given examples (Sec. II) showing that the richness of the dynamical symmetries is by no means exhausted by the local differential symmetries. The unifying role played by the local differential symmetries, as providing us with infinitely many transformations for *any* differential equation of arbitrary order (and not just for the *first* order differential equations, like in Lie's approach) has been emphasized. The relevance of the local differential symmetries in the framework of the Lagrangian systems of differential equations has been discussed. We have seen that there are still some questions to be answered: the connection, for general systems of differential equations, between monoparametric families of dynamical symmetries and the existence, or not, of an associated first integral. This connection should be clarified, not only in the context of the local differential symmetries, but for more general monoparametric monoids of general dynamical symmetries, *not* implementable by transformations such as (29). In this general case, one obviously needs a previous generalization of the concept of first integral, as a functional F such that for any solution $q_1(t), \ldots, q_n(t)$ of (27) and for any time t, the number obtained by the prescription F defining the functional,

$$F(q_1(t), \ldots, q_n(t); t)\,,$$

is a constant. That is,

$$\frac{d}{dt} F(q_1(t), \ldots, q_n(t); t) = 0\,.$$

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to M. C. Hidalgo-Brinquis for the constant help and encouragement given, and to J. McCleary for correcting the manuscript.

[1] R. Anderson, S. Kumei, and C. Wulfman, Phys. Rev. Lett. **28**, 988 (1972).

[2] G. Lamb, J. Math. Phys. **15**, 2157 (1974); M. Wadati, H. Sanuki, and K. Konno, Prog. Theor. Phys. Kyoto **53**, 419 (1975), and references therein.

[3] S. Lie, *Vorlesungen uber differentialgleichungen* (Teubner, Leipzig, 1891), and the references given in Ref. 1.

[4] S. Lang, *Algebra* (Addison-Wesley, Reading, Massachusetts, 1965), Chap. I.

[5] L. Bianchi, *Lezioni sulla Teoria dei Gruppi Continvi Finiti di Transformazioni* (Zanichelli, Bologna, 1928), Chap. X.

[6] See Ref. 5, Chap. XII, p. 436.

[7] L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, New Jersey, 1966), 6th ed., Chap. VI, p. 221–3.

[8] See Ref. 5, pp. 475–8.

[9] J. E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups* (Chelsea, New York, 1966), Chap. II, pp. 26–7.

[10] J. M. Levy-Leblond, Am. J. Phys. **39**, 502 (1971); K. Mariwalla, Phys. Rep. C **20**, 287 (1975) and the reference therein.

[11] E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971), pp. 83–5.

[12] D. G. Currie and E. J. Saletan, Nuovo Cimento B **9**, 143 (1972).

[13] P. Havas, Nuovo Cimento Suppl. **5**, 363 (1957).

Boost matrix elements and Clebsch-Gordan coefficients of the homogeneous Lorentz group

M. K. F. Wong

Fairfield University, Fairfield, Connecticut 06430

Hsin-Yang Yeh

Moorhead State University, Moorhead, Minnesota 56560
(Received 2 August 1976; revised manuscript received 3 March 1977)

It is shown that the boost matrix elements of SO(3,1) obtained by Smorodinskii and Shepelev can be written as a sum of two Fourier series, whose coefficients are Clebsch–Gordan coefficients of SO(3) with complex angular momenta and magnetic quantum numbers. Moreover, the second term is equal to zero for all representations of the principal series except the most degenerate case. The connection between our expression and the spherical functions of Dolginov and other authors is explained. It is also noted that there are two ways of writing the boost matrix elements of SO(3,1), differing from each other by a phase factor. A proof for the orthogonality relation of the representation functions of SO(3,1) is given in the Appendix. The Clebsch–Gordan coefficients of SO(3,1) for the principal series in the general case $(\sigma_1\nu_1)\times(\sigma_2\nu_2)\to(\sigma_3\nu_3)$ are obtained as X functions with complex angular momenta. An integral representation for these X functions is obtained. It is shown that the CG coefficients of SO(3,1) have a multiplicity-two problem. A solution to the multiplicity-two problem is presented.

1. INTRODUCTION

The subject of the representation function, or more simply the boost matrix elements, and the Clebsch–Gordan coefficients (CGC) of SO(3, 1) has been treated by many authors. We offer here a list, which is by no means exhaustive, of previous work done on these two subjects. For the boost matrix elements of SO(3, 1), work has been done by Ström,[1-3] Duc and Hieu,[4,5] Verdiev and Dadashev,[6] Sciarrino and Toller,[7] Anderson, Raczka, Rashid, and Winternitz,[8] Makarov and Shepelev,[9] Smorodinskii and Shepelev,[10] and Smorodinskii and Huszar.[11] For the Clebsch–Gordan coefficients, work has been done by Naimark,[12-14] Dolginov and Toptygin,[15] Dolginov and Moskalev,[16] Anderson *et al.*,[8,17] Klink,[18] Gavrilik,[19] Bisiacchi and Fronsdal,[20] Bamberg,[21] Verdiev,[22] and Verdiev, Kerimov, and Smorodinskii.[23]

In the case of the boost matrix elements of SO(3, 1) there are basically two ways to calculate the d matrix. These two ways are what Ström calls the infinitesimal method and the global method. The first one is to express the infinitesimal generators as differential operators and then solve a partial differential equation for a particular representation. Then one can use lowering or raising operators,[24] or solve recurrence relations, to obtain the general representation matrix. The second method is to write the d matrix as an integral representation and perform the integration. The final results so far obtained are all expressible as a summation over two variables. With the exception of Smorodinskii and Shepelev,[10] there are two points about these expressions which we think can be improved upon. The first point is that these expressions are unrelated to the d matrix of SO(4), which, as Freedman and Wang[25] have shown, is expressible as Fourier series with SU(2) CG coefficients. The boost matrix elements of SO(3, 1) should be connected to those of SO(4) by analytic continuation, but this point had not been made clear until Smorodinskii and Shepelev's work. The second point is that the various expressions lack an underlying structure. Thus they cannot be easily transformed from one to the other. They are unrelated to angular momentum theory, which is well understood, and as a result it is very difficult to extract symmetry properties from these expressions unless one applies complicated mathematical theories about hypergeometric functions. With regard to the first point, we think that Smorodinskii and Shepelev have made progress in showing that the boost matrix elements of SO(3, 1) can be analytically continued to the SO(4) d matrix of Freedman and Wang. However, it is our purpose to show that the expression obtained by Smorodinskii and Shepelev can be further improved upon, thus making closer contact between the boost matrix elements of SO(3, 1) and angular momentum theory with complex j and m. With regard to the second point, we show that the boost matrix elements can be written in terms of CGC of SU(2) with complex j and m, which is a direct analytic continuation of the CGC obtained by Racah.[26] Thus the symmetry properties of the boost matrix elements can be easily derived from the known symmetry properties of the CGC of SU(2).

In summary we offer the following three points as advantages possessed by our expression. 1. The boost matrix elements of SO(3, 1) are expressed as a Fourier series whose coefficients are CGC of SU(2) with complex j and m. This shows immediately that it is an analytic continuation from SO(4). It reduces the summation variable to one, simplifies the expression, and gives a meaning to the summation index, i.e., as the summation index for a Fourier series. 2. Its symmetry properties can be easily read off from the well-known symmetry properties of the CGC of SU(2), or the analytic continuation thereof. 3. We can use the expression of the d matrix of SO(3, 1) to prove that the CGC of SO(3, 1) are analytic continuations of X functions. This will be demonstrated in Sec. 4. We are thus able to show that there is a multiplicity-two problem connected with the CGC of SO(3, 1), which, as far as we know, has not been mentioned by previous authors. A solution to the multiplic-

ity-two problem is presented in Sec. 5.

For the CGC of SO(3, 1), the basic work was done by Naimark.[12-14] However, his expression is in terms of the basis functions of SL(2C), $F_{\sigma\nu lm}(u)$ or $f_{\sigma\nu lm}(z)$, and not in terms of angles, and therefore it is not clear how the CGC of SO(3, 1) are connected with X functions with complex angular momenta. Dolginov and Toptygin[15] and Dolginov and Moskalev[16] succeeded in re-expressing Naimark's result in terms of the spherical functions of SO(3, 1) and thus obtained the CGC of SO(3, 1) for the tensor product $(\sigma_1 0)\times(\sigma_2 0)\to(\sigma_3 0)$, in terms of X functions with complex angular momenta. Other authors, such as Bisiacchi and Fronsdal,[20] Bamberg,[21] and Domokos,[27] have treated the same particular case, for $\nu_1=\nu_2=\nu_3=0$. Anderson, Raczka, Rashid, and Winternitz have obtained general expression for the CGC of SO(3, 1). Their formula is quite complicated, the most general one containing a summation over 11 variables. Moreover, the expressions they obtain are products of two CGC and are therefore doubly lengthy. Gavrilik[19] used the same technique to obtain CGC for the principal and supplementary series. Verdiev, Kerimov, and Smorodinskii[23] obtained general expressions for a single CGC. However, their expression is in terms of "spinor basis" with labels $\rho\lambda\tau m$. In order to transform the spinor basis to the canonical basis $\rho\lambda JM$, one has to multiply the results further by CGC of SO(3) with complex j and m. We have obtained an integral representation for the CGC of SO(3, 1), using techniques similar to those of Dolginov and Moskalev[16] and Verdiev *et al.*,[23] but directly in terms of the canonical basis, i.e., in the decomposition $SO(3,1)\supset SO(3)\supset SO(2)$. Moreover, we have simplified the original expressions of Naimark which were used by Verdiev *et al.*[23] by omitting a phase factor, thereby reducing the CGC expression to a simpler form. Also Verdiev *et al.* did not succeed in obtaining the CGC as a product of CGC of SU(2), while we have succeeded in doing so.

The previous authors have not shown that the CGC of SO(3, 1), in the general case, is an analytic continuation of X functions from SO(4), though they have indicated that this can be done by looking at the recurrence relations of the CGC of SO(3, 1). We shall given an explicit demonstration in Sec. 4 that the CGC of SO(3, 1) are analytic continuations of X functions from SO(4), based on the results of Sec. 2. Finally none of the previous authors, including Naimark, have indicated the possibility of a multiplicity-two problem in the CGC of SO(3, 1). However, we show that the multiplicity-two problem exists in the decomposition of two tensor products of the irreducible representations of the principal series for the following reasons. 1. Just as in the case of SO(2, 1) there is a multiplicity two problem in the tensor product $j_1\times j_2\to j_3$, where j_1, j_2, j_3 all belong to the continuous series, because of the existence of two unitarily equivalent representations, so also in the case of SO(3, 1) there should be a multiplicity-two problem in the tensor product $(\sigma_1\nu_1)\times(\sigma_2\nu_2)\to(\sigma_3\nu_3)$ because of the existence of two unitarily equivalent representations $(\sigma\nu)$ and $(-2-\sigma,-\nu)$. In fact the multiplicity problem is the rule, rather than the exception, in the tensor products of unitary representations of noncompact groups. 2. The previous authors considered mainly the case $\nu_1+\nu_2+\nu_3$ = nonnegative integer. However, for the most general case, one should allow ν to take negative values as well. 3. Looking at the formula for the CGC of SO(3, 1), we see that it is obviously a complex quantity with $i\rho$ appearing at many places. We shall then show that the multiplicity-two problem is connected with the complex conjugate of CGC expressions.

In Sec. 2 we show that the boost matrix elements in the canonical decomposition $SO(3,1)\supset SO(3)\supset SO(2)$ can be written as two Fourier series whose coefficients are CGC of SU(2) with complex arguments. However, with the exception of the most degenerate case, i.e., $d^{\sigma 0}_{000}(\theta)$, the second term vanishes for all representations of the principal series. The continuation to SO(4) according to Freedman and Wang is indicated. For the spherical functions of SO(3, 1) we show how our expression is connected with the work of Dolginov and others. We also note that there are two ways of writing the boost matrix elements, differing from each other by a phase factor. This corresponds to two ways of writing the matrix elements of the generators of $SO(n, 1)$, as pointed out by Wong and Yeh.[28] The proof for the orthogonality and completeness relations of the boost matrix elements is given in the Appendix.

In Sec. 3 we obtain an integral representation for the single CGC of SO(3, 1), based on the work of Naimark[12-14] and Dolginov,[15,16] and similar to the technique used by Verdiev *et al.*[23] but with some simplifications. In Sec. 4 we show that the CGC of SO(3, 1) are analytic continuations of X functions from SO(4). In Sec. 5 we solve the multiplicity-two problem for these coefficients.

2. THE BOOST MATRIX ELEMENTS OF SO(3, 1)

Since the expression we obtain is a modification of Smorodinskii and Shepelev's, we use the same notation for the labels of the boost matrix elements. Thus the irreducible representations are characterized by σ and ν where $\sigma=-1+i\rho$, ρ continuous real and ν integral or half-integral. The representations (σ,ν) and $(-\sigma-2,-\nu)$ are unitarily equivalent, while (σ,ν) and $(-\sigma-2,-\nu)$ are the complex conjugates of each other.

Derivation of $d^{\sigma\nu}_{jmj'}(\theta)$

Start with Eq. (4) of Smorodinskii and Shepelev[10] and complete the contour with an infinite semicircle in the right half-plane of t. Using Stirling's formula for the asymptotic behavior of the gamma function, one can easily check that the integral vanishes on the infinite semicircle. The contour from $-k-i\infty$ to $-k+i\infty$ should be so chosen that all poles of the form $\Gamma(a+t)$ lie on the left of the contour. We then obtain two series of poles, with the result

$$
\begin{aligned}
&d^{\sigma\nu}_{Jm J'}(\theta) \\
&= (-1)^{2J+2m}\left[\frac{(2J+1)(2J'+1)(J-m)!\,(J-\nu)!\,(J'-m)!\,(J'-\nu)!}{(J+m)!\,(J+\nu)!\,(J'+m)!\,(J'+\nu)!}\right]^{1/2} \frac{e^{\theta(\sigma-\nu-m)}}{\Gamma(-J+m)\Gamma(-J'+m)(J-m)!\,(J'-m)!}\ \frac{1}{2\pi i} \\
&\times\int_{-k-i\infty}^{-k+i\infty} dt\ \Gamma(m+\nu+1+t)\Gamma(-t)e^{-2\theta t}\sum_{d,d'}\Gamma(-J+m+d)\Gamma(J+m+d+1)\Gamma(d+\sigma-\nu+1-t)\Gamma(-J'+m+d') \\
&\times\Gamma(d'+m-\sigma+t)\Gamma(J'+m+1+d')[d!\,\Gamma(d+m-\nu+1)\Gamma(d+m+\sigma+2)d'!\,\Gamma(d'+m-\nu+1)\Gamma(d'+m-\sigma)]^{-1} \\
&= \frac{1}{[(m-\nu)!]^2}\left[\frac{(2J+1)(2J'+1)(J+m)!\,(J-\nu)!\,(J'+m)!\,(J'-\nu)!}{(J-m)!\,(J+\nu)!\,(J'-m)!\,(J'+\nu)!}\right]^{1/2} \frac{e^{\theta(\sigma-\nu-m)}}{\Gamma(m-\sigma)\Gamma(m+\sigma+2)} \\
&\times\left\{\sum_{t=0}^{\infty}\Gamma(m+\nu+1+t)\Gamma(m-\sigma+t)\,\frac{(-1)^t}{t!}\,e^{-2\theta t}\Gamma(\sigma-\nu+1-t)\,{}_3F_2(\sigma-\nu+1-t,-J+m,J+m+1;m-\nu+1,m+\sigma+2)\right. \\
&\times{}_3F_2(m-\sigma+t,-J'+m,J'+m+1;m-\nu+1,m-\sigma) \\
&+\sum_{t'=0}^{\infty}\Gamma(m+\sigma+2+t')\Gamma(m-\nu+1+t')\,\frac{(-1)^{t'}}{t'!}\,e^{-2\theta(t'+\sigma-\nu+1)}\Gamma(-t'-\sigma+\nu-1) \\
&\left.\times{}_3F_2(-J+m,J+m+1,-t';m-\nu+1,m+\sigma+2)\,{}_3F_2(-J'+m,J'+m+1,m+t'-\nu+1;m-\nu+1,m-\sigma)\right\}. \qquad (2.1)
\end{aligned}
$$

Next we define CG coefficients from Racah's expression,[26] i.e.,

$$
\begin{aligned}
C\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix} &= (-1)^{j_1-m_1}\left[\frac{(2j+1)(j_1+j_2-j)!\,(j_2-m_2)!\,(j+m)!\,(j_1+m_1)!}{(j_1+j_2+j+1)!\,(j_1-j_2+j)!\,(-j_1+j_2+j)!\,(j_2+m_2)!\,(j-m)!\,(j_1-m_1)!}\right]^{1/2} \\
&\times\left[\frac{(j+j_2-m_1)!}{(j_2-j+m_1)!}\right]{}_3F_2(j_1+m_1+1,-j+m,-j_1+m_1;-j-j_2+m_1,j_2-j+m_1+1) \qquad (2.2)
\end{aligned}
$$

so that

$$
\begin{aligned}
C\begin{pmatrix} J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) \\ m & t-\frac{1}{2}(\sigma-\nu) & t+m-\frac{1}{2}(\sigma-\nu) \end{pmatrix} &= (-1)^{J-m}\left[\frac{(\sigma+\nu+1)(J-\nu)!\,(\sigma-\nu-t)!\,(t+m+\nu)!\,(J+m)!}{(J+\sigma+1)!\,(J+\nu)!\,(\sigma-J)!\,t!\,(\sigma-t-m)!\,(J-m)!}\right]^{1/2}\left[\frac{(\sigma-m)!}{(m-\nu)!}\right] \\
&\times{}_3F_2(J+m+1,m+t-\sigma,-J+m;m-\sigma,m-\nu+1) \qquad (2.3)
\end{aligned}
$$

[note that Barut and Wilson[29] recently derived Eq. (2.2). However, there is a factor $(l_1+m_1)!/(l_1-m_1)!$ missing on the right-hand side of their Eq. (II.1.8)] and

$$
\begin{aligned}
C\begin{pmatrix} J & \frac{1}{2}(-\sigma+\nu)-1 & -\frac{1}{2}(\sigma+\nu)-1 \\ m & t+\frac{1}{2}(\sigma-\nu)+1 & t+m+\frac{1}{2}(\sigma-\nu)\ \ +1 \end{pmatrix} &= (-1)^{J-m}\left[\frac{(-\sigma-\nu-1)(J+\nu)!\,(-\sigma+\nu-t-2)!}{(J-\sigma-1)!\,(J-\nu)!\,(-\sigma-J-2)!\,t!}\ \frac{(t+m-\nu)!\,(J+m)!}{(-\sigma-t-m-2)!\,(J-m)!}\right]^{1/2} \\
&\times\left[\frac{(-\sigma-2-m)!}{(m+\nu)!}\right]{}_3F_2(J+m+1,m+t+\sigma+2,-J+m;m+\sigma+2,m+\nu+1)\,. \qquad (2.4)
\end{aligned}
$$

By Eq. (A.9) of Ref. 10

$$
\begin{aligned}
&{}_3F_2(J+m+1,m+t+\sigma+2,-J+m;m+\sigma+2,m+\nu+1) \\
&\qquad = (-1)^{J-m}\left[\frac{\Gamma(m+\nu+1)\Gamma(J+1-\nu)}{\Gamma(m-\nu+1)\Gamma(J+\nu+1)}\right]{}_3F_2(-t,J+m+1,-J+m;m+\sigma+2,m-\nu+1)\,. \qquad (2.5)
\end{aligned}
$$

Therefore we can express the boost matrix elements as a sum of two Fourier series:

$$
\begin{aligned}
d^{\sigma\nu}_{Jm J'}(\theta) &= (-1)^{(J-J')/2}\left[\frac{(2J+1)(2J'+1)(J-i\rho)!\,(J'+i\rho)!}{(J'-i\rho)!\,(J+i\rho)!}\right]^{1/2} \\
&\times\left\{\frac{1}{(\sigma+\nu+1)}\sum_{t=0}^{\infty}e^{-2\theta[t-(\sigma-\nu)/2]-m\theta}\,C\begin{pmatrix} J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) \\ m & t-\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu) \end{pmatrix} C\begin{pmatrix} J' & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) \\ m & t-\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu) \end{pmatrix}\right. \\
&+\frac{1}{(-\sigma-\nu-1)}\sum_{t'=0}^{\infty}e^{-2\theta[t'+(\sigma-\nu)/2+1]-m\theta} \\
&\left.\times C\begin{pmatrix} J & -\frac{1}{2}(\sigma-\nu)-1 & -\frac{1}{2}(\sigma+\nu)-1 \\ m & t'+\frac{1}{2}(\sigma-\nu)+1 & m+t'+\frac{1}{2}(\sigma-\nu)+1 \end{pmatrix} C\begin{pmatrix} J' & -\frac{1}{2}(\sigma-\nu)-1 & -\frac{1}{2}(\sigma+\nu)-1 \\ m & t'+\frac{1}{2}(\sigma-\nu)+1 & m+t'+\frac{1}{2}(\sigma-\nu)+1 \end{pmatrix}\right\}\,. \qquad (2.6)
\end{aligned}
$$

However, we can show that, with the exception of the most degenerate case, i.e., $d^{\sigma 0}_{000}(\theta)$, the second term always vanishes. This is because, from the same equation (4) of Smorodinskii and Shepelev [our Eq. (2.1)], we have the expression

$$\Gamma(\sigma-\nu+1-t)\,{}_3F_2(\sigma-\nu+1-t,-J+m,J+m+1;m-\nu+1,m+\sigma+2)\,, \tag{2.7}$$

which is equal to

$$[\Gamma(t+1)\Gamma(m-\nu+1)\Gamma(m+\sigma+2)/\Gamma(t+1-J+m)\Gamma(t+2+J+m)]\,{}_3F_2(t+1,m-\sigma+t,m+1+\nu+t;t+1-J+m,t+2+J+m) \tag{2.8}$$

by means of Eq. (A3) of Smorodinskii and Shepelev, i.e.,

$$\Gamma(a)\,{}_3F_2(abc;ef)=[\Gamma(s)\Gamma(e)\Gamma(f)/\Gamma(s+b)\Gamma(s+c)]\,{}_3F_2(s,e-a,f-a;s+b,s+c)\,, \tag{2.9}$$

where

$$a=\sigma-\nu+1-t,\quad b=-J+m,\quad c=J+m+1,\quad e=m-\nu+1,\quad f=m+\sigma+2,\quad s=e+f-a-b-c=t+1\,. \tag{2.10}$$

Substituting (2.8) for (2.7) into the contour integral (2.1) we find that the second series of poles have disappeared. The contribution from the first series of poles remains the same. Thus we conclude that the second term in Eq. (2.6) vanishes in all cases except when (2.7) cannot be transformed into (2.8). This exception occurs only in one case, i.e., for the most degenerate representation $d^{\sigma 0}_{000}(\theta)$. There the ${}_3F_2$ function reduces, after some transformation, to

$$[(-\sigma-1)/(-t)]\,{}_2F_1(1,1-t+\sigma;-t+1)=(\sigma+1)\Gamma(-t+1)\Gamma(-\sigma-1)/t\Gamma(-t)\Gamma(-\sigma)=1\,. \tag{2.11}$$

Alternatively, we can also see that for $d^{\sigma 0}_{000}(\theta)$

$$C\begin{pmatrix}0 & -\frac{1}{2}\sigma-1 & -\frac{1}{2}\sigma-1\\ 0 & t'+\frac{1}{2}\sigma+1 & t'+\frac{1}{2}\sigma+1\end{pmatrix}=1\,. \tag{2.12}$$

Therefore, the second term in (2.6) does not vanish. In this case we can easily calculate from (2.6)

$$d^{\sigma 0}_{000}(\theta)=\sum_{t=0}^{\infty}\frac{e^{\theta\sigma}}{\sigma+1}\,(e^{-2\theta t}-e^{-2\theta(i\rho+t)})=\frac{\sin\rho\theta}{\rho\sinh\theta}\,, \tag{2.13}$$

which is the correct result.

Thus we conclude that the boost matrix elements can be expressed by Eq. (2.6) where the second term is always zero except for the most degenerate case, i.e., $\nu=J=J'=m=0$. In what follows we shall neglect the second term, when it is obvious that the most degenerate representation is not involved.

For the sake of future computation, let us note that $d^{\sigma\nu}_{Jm J'}(\theta)$ can also be written in the following form

$$d^{\sigma\nu}_{Jm J'}(\theta)=(-1)^{(J'-J)/2+J+J'-2\nu}\left[\frac{(2J+1)(2J'+1)}{(\nu+\sigma+1)^2}\right]^{1/2}\left[\frac{(J-i\rho)!\,(J'+i\rho)!}{(J'-i\rho)!\,(J+i\rho)!}\right]^{1/2}\sum_{t=0}^{\infty}e^{-2\theta[t-(\sigma-\nu)/2]-m\theta}$$
$$\times C\begin{pmatrix}J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu)\\ -m & -t+\frac{1}{2}(\sigma-\nu) & -t-m+\frac{1}{2}(\sigma-\nu)\end{pmatrix}C\begin{pmatrix}J' & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu)\\ -m & -t+\frac{1}{2}(\sigma-\nu) & -t-m+\frac{1}{2}(\sigma-\nu)\end{pmatrix}\,. \tag{2.14}$$

The CG coefficients can also be written in different ways as follows

$$C\begin{pmatrix}J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu)\\ m & t-\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu)\end{pmatrix}=(-1)^{J-\nu}C\begin{pmatrix}J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu)\\ -m & -t+\frac{1}{2}(\sigma-\nu) & -m-t+\frac{1}{2}(\sigma-\nu)\end{pmatrix}$$
$$=C\begin{pmatrix}\frac{1}{2}(\sigma-\nu) & J & \frac{1}{2}(\sigma+\nu)\\ -t+\frac{1}{2}(\sigma-\nu) & -m & -m-t+\frac{1}{2}(\sigma-\nu)\end{pmatrix}$$
$$=(-1)^t\left[\frac{(\sigma+\nu+1)}{(2J+1)}\right]^{1/2}C\begin{pmatrix}\frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) & J\\ -t+\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu) & m\end{pmatrix}$$
$$=(-1)^t C\begin{pmatrix}\frac{1}{2}(\sigma+\nu) & \frac{1}{2}(\sigma-\nu) & J\\ -m-t+\frac{1}{2}(\sigma-\nu) & t-\frac{1}{2}(\sigma-\nu) & -m\end{pmatrix}\left[\frac{(\sigma+\nu+1)}{(2J+1)}\right]^{1/2}\,. \tag{2.15}$$

We shall not list all the different ways that the CG coefficients can be expressed. However, we shall indicate the rule according to which CG coefficients can be obtained. The rule is: All 288 ways of expressing the CG coefficients are allowed except where the phase factor does not give discrete values.

The symmetry properties of the boost matrix elements can be found in Ref. 10 and Rühl.[30]

From the fact that $d^{\sigma\nu}_{Jm J'}(0) = \delta_{JJ'}$ we obtain the orthonormality equation for the CG coefficients with complex arguments, i.e.,

$$\sum_{t=0}^{\infty}\left[\frac{(2J+1)(2J'+1)}{(\sigma+\nu+1)^2}\right]^{1/2} C\begin{pmatrix} J & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) \\ m & t-\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu)\end{pmatrix} C\begin{pmatrix} J' & \frac{1}{2}(\sigma-\nu) & \frac{1}{2}(\sigma+\nu) \\ m & t-\frac{1}{2}(\sigma-\nu) & m+t-\frac{1}{2}(\sigma-\nu)\end{pmatrix} = \delta_{JJ'} . \tag{2.16}$$

Connection with Freedman and Wang[25]

With the substitution $\theta \to -i\delta$, $\mu = -m - t + \frac{1}{2}(\sigma-\nu)$, $m \to -\lambda$ we get Freedman and Wang's result[25] exactly except the phase factor $[(J-i\rho)!\,(J'+i\rho)!/(J'-i\rho)!\,(J+i\rho)!]^{1/2}$, which we will discuss shortly.

Connection with Dolginov *et al.*[15,16]

Dolginov *et al.* have obtained the spherical functions $d^{\sigma 0}_{J 00}(\theta)$ and $d^{\sigma 0}_{J J'm}(\theta)$ in terms of CG coefficients with complex angular momenta. Comparing with our result, we find

$$d^{\sigma 0}_{l00}(\theta) = (-1)^{-l/2}[(2l+1)(l-i\rho)!\,(i\rho)!/\rho^2(-i\rho)!\,(l+i\rho)!]^{1/2}\Pi_l(n,\alpha),$$

where $\alpha = -\theta$, $\mu = t - \frac{1}{2}\sigma$, $\frac{1}{2}\sigma = J$ with

$$\Pi_l(n,\alpha) = \sum_{\mu} e^{2\mu\alpha} C\begin{pmatrix} \frac{1}{2}\sigma & l & \frac{1}{2}\sigma \\ \mu & 0 & \mu\end{pmatrix} \tag{2.17}$$

as obtained by Dolginov and

$$d^{\sigma 0}_{ll'm}(\theta) = (-1)^{(l'-l)/2}[(l-i\rho)!\,(l'+i\rho)!/(l'-i\rho)!\,(l+i\rho)!]^{1/2} Q^{ll'}_{JJk}(\psi)$$

with

$$\Lambda = m + t - \tfrac{1}{2}\sigma, \quad \lambda = t - \tfrac{1}{2}\sigma, \quad k = -m, \quad \psi = -\theta, \quad J = \tfrac{1}{2}\sigma,$$

where

$$Q^{ll'}_{JJk}(\psi) = \sum_{\Lambda\lambda} C^{lk}_{J-\Lambda J\lambda} C^{l'k}_{J-\Lambda J\lambda} e^{(\Lambda+\lambda)\psi} . \tag{2.18}$$

We see therefore that the summation index according to Dolginov's expression is a complex number but summed over integral intervals. This is a point which is not obvious when one continues from the compact group to the noncompact group. We only learn about this complex number when we evaluate the d function of the noncompact group directly. We may also mention that in a way it is good to have a complex number for the magnetic quantum number, because then the phase factor $(-1)^{j+m}$ or $(-1)^{j-m}$ in the formula for the CG coefficients can be suitably defined so that either $j+m$ or $j-m$ is an integer.

The phase factor $[(J-i\rho)!\,(J'+i\rho)!/(J'-i\rho)!\,(J+i\rho)!]^{1/2}$ in Eq. (2.6) is due to the fact that the matrix elements of the generators of SO(n, 1) can be written as complex, as Chakrabarti[31] has shown. However, we[28] have also shown in a previous paper that the matrix elements of the generators can be written as pure real or pure imaginary. In that case the phase factor in Eq. (2.6) will disappear. We thus conclude that there are two ways of writing the d functions of SO(n, 1), one with the phase factor, which is equal to the inverse of ω_1 of Maekawa,[32] corresponding to complex matrix elements of the generators, and the other without the phase factor, corresponding to pure real or pure imaginary matrix elements of the generators.

In the Appendix we give a proof for the orthogonality and completeness relations of the representation functions of SO(3, 1).

3. AN INTEGRAL REPRESENTATION FOR THE CG COEFFICIENTS OF SO(3, 1)

In this section we wish to obtain the CG coefficients of SO(3,1) for the principal series in the most general case: $(\sigma_1\nu_1)\times(\sigma_2\nu_2) \to (\sigma_3\nu_3)$. For the time being, we shall ignore the multiplicity-two problem and follow Naimark's work,[12-14] used by Dolginov and Moskalev[16] to obtain the CGC of SO(3,1). We have, from Naimark,

$$f_{(\sigma_3\nu_3)}(u_3) = \int du_1 \int du_2 f_{(\sigma_1\nu_1)}(u_1) f_{(\sigma_2\nu_2)}(u_2) \times b_{(\sigma_1\sigma_2\sigma_3)}(u_1u_2u_3), \tag{3.1}$$

$$f_{(\sigma_1\nu_1)}(u_1) f_{(\sigma_2\nu_2)}(u_2) = (4\pi^4)^{-1}\int_0^{\infty} d\rho_3 \sum_{\nu_3}(\nu_3^2+\rho_3^2) \times \int du_3 f_{(\sigma_3\nu_3)}(u_3)\overline{b_{(\sigma_1\sigma_2\sigma_3)}}(u_1u_2u_3), \tag{3.2}$$

where

$$b_{(\sigma_1\sigma_2\sigma_3)} = \pi^{3/2}(\sin\tfrac{1}{2}\Omega_{12})^{-i\rho_1-i\rho_2-i\rho_3-1}(\sin\tfrac{1}{2}\Omega_{13})^{-1-i\rho_1+i\rho_2+i\rho_3} \times (\sin\tfrac{1}{2}\Omega_{23})^{i\rho_1-i\rho_2+i\rho_3-1}, \tag{3.3}$$

$$\cos\Omega_{ik} = \cos\theta_i\cos\theta_k + \sin\theta_i\sin\theta_k\cos(\phi_i-\phi_k).$$

Strictly speaking, there should be a phase $e^{i(\beta+\phi/2-\pi/2)\nu}$ attached to each of the three terms in (3.3), since

$$|\xi_2\eta_1-\xi_1\eta_2|^{2\nu-2i\rho-1}(\xi_2\eta_1-\xi_1\eta_2)^{-2\nu}$$
$$=(\sin\tfrac{1}{2}\Omega_{12})^{\nu-i\rho-1}(\sin\tfrac{1}{2}\Omega_{12})^{-\nu}e^{i(\beta+\phi/2-\pi/2)\nu},$$

where

$$\phi=\phi_2^1-\phi_2^2-\phi_1^1-\phi_1^2,$$

$$\tan\beta=\frac{\sin\frac{1}{2}\theta_2\cos\frac{1}{2}\theta_1\sin\phi}{\sin\frac{1}{2}\theta_1\cos\frac{1}{2}\theta_2-\sin\frac{1}{2}\theta_2\cos\frac{1}{2}\theta_1\cos\phi},$$

$$u_i=\begin{pmatrix}\bar\eta_i & -\bar\xi_i\\ \xi_i & \eta_i\end{pmatrix}=\begin{pmatrix}\cos\frac{1}{2}\theta_i e^{i\phi_i/2} & i\sin\frac{1}{2}\theta_i e^{i\phi_i'/2}\\ -i\sin\frac{1}{2}\theta_i e^{-i\phi_i'/2} & \cos\frac{1}{2}\theta_i e^{-i\phi_i/2}\end{pmatrix}, \tag{3.4}$$

$$\phi_i=\phi_1^i+\phi_2^i,\quad \phi_i'=\phi_1^i-\phi_2^i+\pi.$$

But the phase cannot change the absolute value of the matrix elements of the generators of the group. As we have shown in a previous paper,[28] the extra phase corresponds to writing the matrix elements of the generators as complex. We can therefore put $\phi=0$ in Eq. (3.4), and deal with the absolute value only as given by Eq. (3.3).

Now we follow Dolginov and Moskalev[16] and write

$$(\sin\tfrac{1}{2}\Omega_{12})^{2a-2}=\sum_{L_1}F_{L_1}(a)(2L_1+1)d^{L_1}_{\nu_1 0}(\cos\Omega_{12}),$$
$$(\sin\tfrac{1}{2}\Omega_{13})^{2b-2}=\sum_{L_2}F_{L_2}(b)(2L_2+1)d^{L_2}_{0\nu_3}(\cos\Omega_{13}), \tag{3.5}$$
$$(\sin\tfrac{1}{2}\Omega_{23})^{2c-2}=\sum_{L_3}F_{L_3}(c)(2L_3+1)d^{L_3}_{\nu_2 0}(\cos\Omega_{23}).$$

Then

$$F_L(g)=\int_0^\pi(\sin\tfrac{1}{2}\Omega)^{2\alpha-2}\bar d^L_{m',m}(\cos\Omega)\sin\Omega\,d\Omega, \tag{3.6}$$

where

$$d^L_{m',m}(\cos\Omega)=[(L+m')!(L-m)!/(L-m')!(L+m)!]^{1/2}$$
$$\times(\cos\tfrac{1}{2}\Omega)^{m+m'}(\sin\tfrac{1}{2}\Omega)^{m'-m}P^{m'-m,m'+m}_{L-m'}(\cos\Omega). \tag{3.7}$$

From Eq. (3), p. 284, Vol. 2 of Erdelyi *et al.*,[33] we have

$$F_L(g)=\int_{-1}^1 2^{-\rho'-\sigma'}(1-x)^{\rho'}(1+x)^{\sigma'}p_n^{\alpha\beta}(x)dx$$
$$=[2\Gamma(\rho'+1)\Gamma(\sigma'+1)/\Gamma(\sigma'+\rho'+2)]$$
$$\times{}_3F_2(-n,\alpha+\beta+n+1,\rho'+1;\alpha+1,\rho'+\sigma'+2,1), \tag{3.8}$$

where

$$a=\tfrac{1}{2}-\tfrac{1}{2}i(\rho_1+\rho_2+\rho_3),\quad b=\tfrac{1}{2}+\tfrac{1}{2}i(-\rho_1+\rho_2+\rho_3),$$
$$c=\tfrac{1}{2}+\tfrac{1}{2}i(\rho_1-\rho_2+\rho_3),\quad \rho_1'=a-1+\nu_1,\quad \sigma_1'=\nu_1,\quad n_1=L_1-\nu_1,$$
$$\alpha_1=\nu_1,\quad \beta_1=\nu_1,\quad \rho_2'=b-1-\nu_3,\quad \sigma_2'=\nu_3,\quad n_2=L_2,$$
$$\alpha_2=-\nu_3,\quad \beta_2=\nu_3,\quad \rho_3'=c-1+\nu_2,\quad \sigma_3'=\nu_2,\quad n_3=L_3-\nu_2,$$
$$\alpha_3=\nu_2,\quad \beta_3=\nu_2. \tag{3.9}$$

Next we have

$$d^{L_1}_{\nu_1 0}(\cos\Omega_{12})=\sum_{m_1'}d^{L_1}_{\nu_1 m_1'}(\cos\theta_1)d^{L_1}_{m_1' 0}(\cos\theta_2),$$
$$d^{L_2}_{0\nu_3}(\cos\Omega_{13})=\sum_{m_2'}d^{L_2}_{0m_2'}(\cos\theta_1)d^{L_2}_{m_2'\nu_3}(\cos\theta_3), \tag{3.10}$$
$$d^{L_3}_{\nu_2 0}(\cos\Omega_{23})=\sum_{m_3'}d^{L_3}_{\nu_2 m_3'}(\cos\theta_2)d^{L_3}_{m_3' 0}(\cos\theta_3).$$

Therefore,

$$d^{L_1}_{\nu_1 0}(\cos\Omega_{12})d^{L_2}_{0\nu_3}(\cos\Omega_{13})d^{L_3}_{\nu_2 0}(\cos\Omega_{23})$$
$$=\sum_{\substack{l_1 l_2 l_3\\ m_1 m_2 m_3}}d^{l_1}_{\nu_1 m_1}(\cos\theta_1)C\begin{pmatrix}L_1 & L_2 & l_1\\ \nu_1 & 0 & \nu_1\end{pmatrix}C\begin{pmatrix}L_1 & L_2 & l_1\\ m_1' & m_2' & m_1\end{pmatrix}$$
$$\times(-1)^{m_1}d^{l_2}_{\nu_2 m_2}(\cos\theta_2)C\begin{pmatrix}L_1 & L_3 & l_2\\ 0 & \nu_2 & \nu_2\end{pmatrix}C\begin{pmatrix}L_1 & L_3 & l_2\\ m_1' & m_3' & m_2\end{pmatrix}$$
$$\times(-1)^{m_3-\nu_3}d^{l_3}_{\nu_3 m_3}(\cos\theta_3)C\begin{pmatrix}L_2 & L_3 & l_3\\ m_2' & m_3' & m_3\end{pmatrix}C\begin{pmatrix}L_1 & L_3 & l_3\\ \nu_3 & 0 & \nu_3\end{pmatrix}, \tag{3.11}$$

where $m_1'=\frac{1}{2}(m_1+m_2-m_3)$, $m_2'=\frac{1}{2}(m_1+m_3-m_2)$, $m_3'=\frac{1}{2}(m_2+m_3-m_1)$.

Corresponding to Dolginov's[16] $f_{nlm}(u)$, we use

$$f_{\sigma\nu lm}(u_i)=(2l+1)^{1/2}u^l_{\nu m}(u_i). \tag{3.12}$$

Thus Dolginov's phase factor $A_l(\sigma)$ does not appear in our equations. From (3.3) the b can be expressed as

$$b_{\sigma_1\sigma_2\sigma_3}=\sum_{\substack{l_1 l_2 l_3\\ m_1 m_2}}[(2l_1+1)(2l_2+1)]^{1/2}B_{l_1 l_2 l_3}C\begin{pmatrix}l_1 & l_2 & l_3\\ m_1 & m_2 & m_3\end{pmatrix}$$
$$\times\overline{f_{\sigma_1\nu_1 l_1 m_1}}(u_1)\overline{f_{\sigma_2\nu_2 l_2 m_2}}(u_2)f_{\sigma_3\nu_3 l_3 m_3}(u_3), \tag{3.13}$$

where

$$B_{l_1 l_2 l_3}=\sum_{L_1 L_2 L_3}\pi^{3/2}(-1)^{L_2+l_1-m_1+m_2-\nu_3}(2l_1+1)^{-1/2}$$
$$\times(2l_2+1)^{-3/2}(2l_3+1)^{1/2}F_{L_1}(a)F_{L_2}(b)F_{L_3}(c)$$
$$\times(2L_1+1)(2L_2+1)(2L_3+1)C\begin{pmatrix}L_1 & L_2 & l_1\\ \nu_1 & 0 & \nu_1\end{pmatrix}$$
$$\times C\begin{pmatrix}L_1 & L_3 & l_2\\ 0 & \nu_2 & \nu_2\end{pmatrix}C\begin{pmatrix}L_2 & L_3 & l_3\\ \nu_3 & 0 & \nu_3\end{pmatrix}W(L_1L_2l_2l_3;l_1L_3). \tag{3.14}$$

Next we define

$$\Psi^{\sigma\nu}_{\nu\nu lm}(a)=(\rho^2+\nu^2)^{1/2}D^{\sigma\nu}_{\nu\nu lm}(a)$$
$$=\sum_{m'm''}u^\nu_{\nu m'}(u)d^{\sigma\nu\nu}_{\nu lm'm''}(\varepsilon)u_{m''m}(u')(\rho^2+\nu^2)^{1/2}, \tag{3.15}$$

where $a=u\varepsilon u'$. Now use the integral representation of the boost

$$d^{\sigma\nu}_{\nu lm'm''}(\varepsilon)$$
$$=\int du_1(2\nu+1)^{1/2}(2l+1)^{1/2}\,\bar u^\nu_{\nu m'}(u_1)[V(\varepsilon)u_{\nu m''}(u_1)]$$
$$=\int du_1(2\nu+1)^{1/2}(2l+1)^{1/2}[\bar V(\varepsilon^{-1})\bar u^\nu_{\nu m'}(u_1)]u_{m''}(u_1). \tag{3.16}$$

Substituting (3.16) into (3.15) and summing over m' and m'', we get

$$\Psi^{\sigma\nu}_{\nu\nu lm}(a) = \int du_1 (2\nu+1)^{1/2}(2l+1)^{1/2}\lambda(u_1 u^{-1}\varepsilon^{-1})^{-2-2i\rho} \times \overline{u^{\nu}_{\nu\nu}}(uu^{-1}\varepsilon^{-1}) u^{l}_{\nu m}(u_1 u')(\rho^2+\nu^2)^{1/2}. \tag{3.17}$$

Using the definition of $f_{\sigma\nu lm}(u)$ in Eq. (3.12), we can write (3.17) as

$$\Psi^{\sigma\nu}_{\nu\nu lm}(a) = \int F_{\sigma\nu}(a,u_1) f_{\sigma\nu lm}(u_1)\, du_1 . \tag{3.18}$$

Multiply both sides of Eq. (3.18) by $\overline{f_{\sigma\nu lm}(u_1)}$ and sum over l, m, using the completeness relations of the basis functions $f_{\sigma\nu lm}(u_1)$; we get

$$F_{\sigma\nu}(a,u_1) = \sum_{lm} \Psi^{\sigma\nu}_{\nu\nu lm}(a)\overline{f_{\sigma\nu lm}}(u_1). \tag{3.19}$$

Substituting (3.12) and (3.17) into (3.19), we get

$$\begin{aligned} F_{\sigma\nu}(a,u_1) &= \int du_1'(2\nu+1)^{1/2}[\lambda(u'u^{-1}\varepsilon^{-1})]^{-2-2i\rho} \\ &\quad \times \overline{u}^{\nu}_{\nu\nu}(u_1' u^{-1}\overline{\varepsilon^{-1}})(\rho^2+\nu^2)^{1/2}\delta(u_1 - u_1' u') \\ &= (\rho^2+\nu^2)^{1/2}(2\nu+1)^{1/2}[\lambda(u_1 u'^{-1}u^{-1}\varepsilon^{-1})]^{-2-2i\rho} \\ &\quad \times \overline{u}^{\nu}_{\nu\nu}(u_1 u'^{-1}u^{-1}\overline{\varepsilon^{-1}}) \\ &= (\rho^2+\nu^2)^{1/2}[\lambda(u_1\hat{u}^{-1}\varepsilon^{-1})]^{-2-2i\rho} \\ &\quad \times \overline{f_{\sigma\nu\nu\nu}}(u_1\hat{u}^{-1}\overline{\varepsilon^{-1}}), \end{aligned} \tag{3.20}$$

where $\hat{u} = uu'$.

Substituting the value of $F_{\sigma\nu}(a,u_1)$ in (3.20) into (3.18), we obtain the following result:

$$\begin{aligned} \Psi^{\sigma\nu}_{\nu\nu lm}(a) &= (\rho^2+\nu^2)^{1/2}(V_{\hat{u}^{-1}\varepsilon^{-1}} f_{\sigma\nu\nu\nu}(u_1),\ f_{\sigma\nu lm}(u_1)) \\ &= (\rho^2+\nu^2)^{1/2}(f_{\sigma\nu\nu\nu}(u_1),\ V_{\varepsilon\hat{u}} f_{\sigma\nu lm}(u_1)). \end{aligned} \tag{3.21}$$

It is clear that the results are still valid if, instead of $\Psi^{\sigma\nu}_{\nu\nu lm}$, we start with $\Psi^{\sigma\nu}_{i'm'lm}(a)$. Thus we have obtained the following result which we shall state as a theorem.

Theorem: The transformation between the basis vectors $\Psi^{\sigma\nu}_{i'm'lm}(a)$ and $f_{\sigma\nu lm}(u_1)$ is done through the following two equivalent equations:

$$\begin{aligned} &\Psi^{\sigma\nu}_{i'm'lm}(a) \\ &\quad = (\rho^2+\nu^2)^{1/2}(V_{\hat{u}^{-1}\varepsilon^{-1}} f_{\sigma\nu l'm'}(u_1),\ f_{\sigma\nu lm}(u_1)) \\ &\quad = (\rho^2+\nu^2)^{1/2}(f_{\sigma\nu l'm'}(u_1),\ V_{\varepsilon\hat{u}} f_{\sigma\nu lm}(u_1)), \end{aligned} \tag{3.22}$$

where

$$a = u\varepsilon u' \quad \text{and} \quad \hat{u} = uu' .$$

Using (3.18), we have

$$\begin{aligned} &\Psi^{\sigma_1\nu_1}_{\nu_1\nu_1 l_1 m_1}(a) \times \Psi^{\sigma_2\nu_2}_{\nu_2\nu_2 l_2 m_2}(a) \\ &\quad = \iint du_1\, du_2\, F_{\sigma_1\nu_1}(a,u_1) F_{\sigma_2\nu_2}(a,u_2) f_{\sigma_1\nu_1 l_1 m_1}(u_1) \\ &\quad\quad \times f_{\sigma_2\nu_2 l_2 m_2}(u_2) \\ &\quad = (4\pi^4)^{-1}\int d\rho_3 \sum_{\nu_3}(\rho_3^2+\nu_3^2)\int du_1\, du_2\, du_3\, F_{\sigma_1\nu_1}(a,u_1) \\ &\quad\quad \times F_{\sigma_2\nu_2}(a,u_2)\overline{b_{\sigma_1\sigma_2\sigma_3}}(u_1u_2u_3) f_{\sigma_3\nu_3}(u_3), \end{aligned} \tag{3.23}$$

where (3.2) has been used in the last step. From (3.1) and (3.13) we have

$$\begin{aligned} f_{\sigma_3\nu_3}(u_3) &= \sum_{l_3}(2l_1+1)^{1/2}(2l_2+1)^{1/2}B_{l_1 l_2 l_3} \\ &\quad \times C\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} f_{\sigma_3\nu_3 l_3 m_3}(u_3). \end{aligned} \tag{3.24}$$

Now define

$$\begin{aligned} I = \iiint du_1\, du_2\, du_3\, F_{\sigma_1\nu_1}(a,u_1) F_{\sigma_2\nu_2}(a,u_2) \\ \times \overline{b_{\sigma_1\sigma_2\sigma_3}}(u_1u_2u_3) f_{\sigma_3\nu_3 l_3 m_3}(u_3). \end{aligned} \tag{3.25}$$

Next define

$$u_i' = u_i\overline{\hat{u}^{-1}\varepsilon^{-1}}. \tag{3.26}$$

Then following the same argument as in the Appendix of Dolginov and Moskalev,[16] we have

$$[\lambda(u\hat{u}^{-1}\varepsilon^{-1})]^{-2-2i\rho_i} = [\lambda(u_i'\varepsilon\hat{u})]^{2+2i\rho_i}, \tag{3.27}$$

$$\begin{aligned} &\overline{b_{\sigma_1\sigma_2\sigma_3}}(u_i') \\ &\quad = \overline{b_{\sigma_1\sigma_2\sigma_3}}(u_i')[\lambda(u_i'\varepsilon\hat{u})]^{-2i\rho_1+2}[\lambda(u_2'\varepsilon\hat{u})]^{-2i\rho_2+2} \\ &\quad\quad \times [\lambda(u_3'\varepsilon\hat{u})]^{-2i\rho_3+2}, \end{aligned} \tag{3.28}$$

$$du_i = du_i'/[\lambda(u_i'\varepsilon\hat{u})]^4. \tag{3.29}$$

The arguments go through because all the quantities involved are independent of ν, as one can see from (3.3). Therefore,

$$\begin{aligned} I &= (\rho_1^2+\nu_1^2)^{1/2}(\rho_2^2+\nu_2^2)^{1/2}(2\nu_1+1)^{1/2}(2\nu_2+1)^{1/2}\iiint du_1'\, du_2'\, du_3' \\ &\quad \times \overline{b_{\sigma_1\sigma_2\sigma_3}}(u_1'u_2'u_3')[\lambda(u_3'\varepsilon\hat{u})]^{-2+2i\rho_3} f_{\sigma_3\sigma_3 l_3 m_3}(u_3'\overline{\varepsilon\hat{u}}) \\ &\quad \times \overline{u^{\nu_1}_{\nu_1\nu_1}}(u')\overline{u^{\nu_2}_{\nu_2\nu_2}}(u_2'). \end{aligned} \tag{3.30}$$

But

$$\overline{u^{\nu_i}_{\nu_i\nu_i}}(u_i') = (2\nu_i+1)^{-1/2}\overline{f_{\sigma_i\nu_i\nu_i\nu_i}}(u_i'). \tag{3.31}$$

Multiplying both sides of Eq. (3.13) by $\overline{u^{\nu_1}_{\nu_1\nu_1}}(u_1')\overline{u^{\nu_2}_{\nu_2\nu_2}}(u_2')$ and integrating over u_1', u_2', we obtain

$$\begin{aligned} &\iint \overline{b_{\sigma_1\sigma_2\sigma_3}}(u_1'u_2'u_3')\overline{u^{\nu_1}_{\nu_1\nu_1}}(u_1')\overline{u^{\nu_2}_{\nu_2\nu_2}}(u_2')\, du_1'\, du_2' \\ &\quad = \overline{B_{\nu_1\nu_2\nu_3}} C\begin{pmatrix} \nu_1 & \nu_2 & \nu_3 \\ \nu_1 & \nu_2 & \nu_1+\nu_2 \end{pmatrix} \overline{f_{\sigma_3\nu_3\nu_3\nu_1+\nu_2}}(u_3'). \end{aligned} \tag{3.32}$$

Substituting (3.32) into (3.30), we have

$$\begin{aligned} I &= (\rho_1^2+\nu_1^2)^{1/2}(\rho_2^2+\nu_2^2)^{1/2}(2\nu_1+1)^{1/2}(2\nu_2+1)^{1/2}\int du_3' \\ &\quad \times [\lambda(u_3'\varepsilon\hat{u})]^{-2+2i\rho_3} f_{\sigma_3\nu_3 l_3 m_3}(u_3'\overline{\varepsilon\hat{u}})\overline{B_{\nu_1\nu_2\nu_3}} \\ &\quad \times C\begin{pmatrix} \nu_1 & \nu_2 & \nu_3 \\ \nu_1 & \nu_2 & \nu_1+\nu_2 \end{pmatrix} \overline{f_{\sigma_3\nu_3\nu_3\nu_1+\nu_2}}(u_3') \\ &= (\rho_1^2+\nu_1^2)^{1/2}(\rho_2^2+\nu_2^2)^{1/2}(2\nu_1+1)^{1/2}(2\nu_2+1)^{1/2}(\rho_3^2+\nu_3^2)^{-1/2} \\ &\quad \times \overline{B_{\nu_1\nu_2\nu_3}} C\begin{pmatrix} \nu_1 & \nu_2 & \nu_3 \\ \nu_1 & \nu_2 & \nu_1+\nu_2 \end{pmatrix} \Psi^{\sigma_3\nu_3}_{\nu_3\nu_1+\nu_2 l_3 m_3}(a). \end{aligned} \tag{3.33}$$

From (3.23), we have

$$\Psi^{\sigma_1\nu_1}_{\nu_1\nu_1 l_1 m_1}(a)\times\Psi^{\sigma_2\nu_2}_{\nu_2\nu_2 l_2 m_2}(a)$$

$$=(4\pi^4)^{-1}\int_0^\infty d\rho_3\sum_{\nu_3}(\rho_3^2+\nu_3^2)(2\nu_1+1)^{1/2}(2\nu_2+1)^{1/2}$$

$$\times I\sum_{l_3}B_{l_1l_2l_3}C\begin{pmatrix} l_1 & l_2 & l_3\\ m_1 & m_2 & m_3\end{pmatrix}$$

$$=(4\pi^4)^{-1}\int_0^\infty d\rho_3\sum_{\nu_3}(\rho_1^2+\nu_1^2)^{1/2}(\rho_2^2+\nu_2^2)^{1/2}(\rho_3^2+\nu_3^2)^{1/2}$$

$$\times(2\nu_1+1)^{1/2}(2\nu_2+1)^{1/2}(2l_1+1)^{1/2}(2l_2+1)^{1/2}\sum_{l_3}B_{l_1l_2l_3}$$

$$\times C\begin{pmatrix} l_1 & l_2 & l_3\\ m_1 & m_2 & m_3\end{pmatrix}\overline{B_{\nu_1\nu_2\nu_3}}C\begin{pmatrix}\nu_1 & \nu_2 & \nu_3\\ \nu_1 & \nu_2 & \nu_1+\nu_2\end{pmatrix}$$

$$\times\Psi^{\sigma_3\nu_3}_{\nu_3\nu_1+\nu_2 l_3 m_3}. \tag{3.34}$$

Thus we obtain the CG coefficients of SO(3, 1) in the tensor products of the basis functions $\Psi^{\sigma_i\nu_i}_{\nu_i\nu_i l_i m_i}(a)$, $i=1,2$. The functions $B_{l_1l_2l_3}$ are the integral representations of X functions with complex angular momenta:

$$B_{l_1l_2l_3}=(2\pi^2)^{-1}\phi(\sigma_1\sigma_2\sigma_3)$$

$$\times X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}, \tag{3.35}$$

where $\phi(\sigma_1\sigma_2\sigma_3)$ is the same phase factor as obtained by Dolginov *et al.*[15,16] since, again, it is independent of ν:

$$\phi(\sigma_1\sigma_2\sigma_3)$$

$$=\tfrac{1}{2}\pi^{-1/2}(\rho_1\rho_2\rho_3)^{1/2}\Gamma[\tfrac{1}{2}-\tfrac{1}{2}i(\rho_1+\rho_2+\rho_3)]$$

$$\times\Gamma[\tfrac{1}{2}+\tfrac{1}{2}i(-\rho_1+\rho_2+\rho_3)]\,\Gamma[\tfrac{1}{2}+\tfrac{1}{2}i(\rho_1-\rho_2+\rho_3)]$$

$$\times\Gamma[\tfrac{1}{2}+\tfrac{1}{2}i(-\rho_1-\rho_2+\rho_3)]\,[\Gamma(1-i\rho_1)\Gamma(1-i\rho_2)\Gamma(1+i\rho_3)]^{-1} \tag{3.36}$$

and

$$|\phi(\sigma_1\sigma_2\sigma_3)|^2=\sinh\pi\rho_1\sinh\pi\rho_2\sinh\pi\rho_3[4\cosh^2\tfrac{1}{2}\pi(\rho_1+\rho_2+\rho_3)$$

$$\times\cosh^2\tfrac{1}{2}\pi(\rho_1+\rho_2-\rho_3)\cosh^2\tfrac{1}{2}\pi(\rho_1-\rho_2+\rho_3)$$

$$\times\cosh^2\tfrac{1}{2}\pi(-\rho_1+\rho_2+\rho_3)]^{-1}. \tag{3.37}$$

One more remark about the expansion (3.34). It is easy to see that the expansion (3.34) is not limited to the basis function $\Psi^{\sigma_i\nu_i}_{\nu_i\nu_i l_i m_i}$. It is still valid for an arbitrary basis function $\Psi^{\sigma_i\nu_i}_{l_i'm_i'l_im_i}$. All one has to do is to replace the corresponding quantities (ν_i and ν_i) by l_i' and m_i', e.g., $\overline{B_{\nu_1\nu_2\nu_3}}\to\overline{B_{l_1'l_2'l_3'}}$ and

$$C\begin{pmatrix}\nu_1 & \nu_2 & \nu_3\\ \nu_1 & \nu_2 & \nu_1+\nu_2\end{pmatrix}\to C\begin{pmatrix} l_1' & l_2' & l_3'\\ m_1' & m_2' & m_3'\end{pmatrix},\quad\text{etc.}$$

We have thus obtained an expression for the single CGC of SO(3, 1) in a form simpler than those obtained by previous authors.[8,17,23] Moreover, it is expressed as a product of $3j$ and $6j$ symbols of SU(2), since from Eq. (3.8) one can change $F_L(a)$ into a $3j$ symbol too.

4. CG COEFFICIENTS OF SO(3, 1) AS X FUNCTIONS WITH COMPLEX ANGULAR MOMENTA

In this section we shall give a justification why the CGC of SO(3, 1) obtained in the previous section are X functions with complex angular momenta. This is based on the results of Sec. 2, where the boost matrix elements have been expressed as a Fourier series with CGC of SU(2) with complex j and m. Of course, this result can also be inferred from analytic continuation from SO(4) to SO(3, 1), and by recurrence relations of $B_{l_1l_2l_3}$ with different values of l, showing that they obey the same equations as the X functions with complex angular momenta. However, we prefer to justify our statement from a more direct, group theoretical point of view.

First we note that the Lorentz group with generators M_i and N_i, $i=1,2,3$, satisfying the commutation relations

$$[M_i,M_k]=i\epsilon_{ikl}M_l,$$
$$[N_i,N_k]=-i\epsilon_{ikl}M_l\,[M_i,N_k]=i\epsilon_{ikl}N_l \tag{4.1}$$

can also be written in the following form:

$$[J_i,J_k]=i\epsilon_{ikl}J_l,$$
$$[K_i,K_k]=i\epsilon_{ikl}K_l\,[J_i,K_k]=0, \tag{4.2}$$

where $J_k=\frac{1}{2}(M_k+iN_k)$, $K_k=\frac{1}{2}(M_k-iN_k)$.

As shown by Smorodinskii and Huszar,[11] the representations of the Lorentz group can be considered under the two parameter subgroup $H=\mathrm{SO}(2)\times\mathrm{SO}(1,1)$, with J_i and K_i as generators. Since J_i and K_k commute, it is obvious that the representation can be considered as a direct product of the two groups generated by J_i and K_i. Moreover, it is easy to see from Smorodinskii and Huszar's work and the form of the d function obtained by us in Sec. 2 that the transformation from one basis to another is effected by the CGC with complex j and m, i.e.,

$$\left|\begin{matrix}\sigma & \nu\\ l & \\ m & \end{matrix}\right\rangle=\sum_k\left|\begin{matrix}\frac{1}{2}(\sigma+\nu)\\ m-k\end{matrix}\right\rangle\times\left|\begin{matrix}\frac{1}{2}(\sigma-\nu)\\ k\end{matrix}\right\rangle$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma+\nu) & \frac{1}{2}(\sigma-\nu) & l\\ m-k & k & m\end{pmatrix}, \tag{4.3}$$

where k is a complex number, but summed over integral intervals: $k=t-\frac{1}{2}(\sigma-\nu)$, $t=0,1,\ldots,\infty$.

Because of the Burchnall–Chaundy[34] formula, the CGC with complex arguments defined by us in terms of ${}_3F_2$ functions with unit argument is indeed the coupling coefficient for two representation functions with J_i and K_i as generators which, as shown by Smorodinskii and Huszar, are expressible as hypergeometric functions.

Since the CGC with complex angular momenta and magnetic quantum numbers exist, the following expansion is meaningful, where the summation over continuous σ_3 is to be understood as an integral:

$$\left|\begin{matrix}\sigma_1\ \nu_1\\ l_1\\ m_1\end{matrix}\right\rangle\times\left|\begin{matrix}\sigma_2\ \nu_2\\ l_2\\ m_2\end{matrix}\right\rangle=\sum_{k_1k_2}\left|\begin{matrix}\frac{1}{2}(\sigma_1+\nu_1)\\ k_1\end{matrix}\right\rangle\times\left|\begin{matrix}\frac{1}{2}(\sigma_1-\nu_1)\\ m_1-k_1\end{matrix}\right\rangle C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ k_1 & m_1-k_1 & m_1\end{pmatrix}\times\left|\begin{matrix}\frac{1}{2}(\sigma_2+\nu_2)\\ k_2\end{matrix}\right\rangle\times\left|\begin{matrix}\frac{1}{2}(\sigma_2-\nu_2)\\ m_2-k_2\end{matrix}\right\rangle$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ k_2 & m_2-k_2 & m_2\end{pmatrix}$$

$$=\sum_{(\sigma_3+\nu_3)/2,\ (\sigma_3-\nu_3)/2\atop k_1,k_2}\left|\begin{matrix}\frac{1}{2}(\sigma_3+\nu_3)\\ k_1+k_2\end{matrix}\right\rangle C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_3+\nu_3)\\ k_1 & k_2 & k_1+k_2\end{pmatrix}\left|\begin{matrix}\frac{1}{2}(\sigma_3-\nu_3)\\ m_1+m_2-k_1-k_2\end{matrix}\right\rangle$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_1-\nu_1) & \frac{1}{2}(\sigma_2-\nu_2) & \frac{1}{2}(\sigma_3-\nu_3)\\ m_1-k_1 & m_2-k_2 & m_1+m_2-k_1-k_2\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ k_1 & m_1-k_1 & m_1\end{pmatrix}$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ k_2 & m_2-k_2 & m_2\end{pmatrix}$$

$$=\sum_{\sigma_3\nu_3\atop k_1k_2l_3}\left|\begin{matrix}\sigma_3\ \nu_3\\ l_3\\ m_1+m_2\end{matrix}\right\rangle C\begin{pmatrix}\frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\\ k_1+k_2 & m_1+m_2-k_1-k_2 & m_1+m_2\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_3+\nu_3)\\ k_1 & k_2 & k_1+k_2\end{pmatrix}$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_1-\nu_1) & \frac{1}{2}(\sigma_2-\nu_2) & \frac{1}{2}(\sigma_3-\nu_3)\\ m_1-k_1 & m_2-k_2 & m_1+m_2-k_1-k_2\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ k_1 & m_1-k_1 & m_1\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ k_2 & m_2-k_2 & m_2\end{pmatrix}$$

$$=\sum_{\sigma_3\nu_3\atop l_3}\left|\begin{matrix}\sigma_3\ \nu_3\\ l_3\\ m_1+m_2\end{matrix}\right\rangle X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}C\begin{pmatrix}l_1 & l_2 & l_3\\ m_1 & m_2 & m_1+m_2\end{pmatrix}(2l_1+1)^{1/2}(2l_2+1)^{1/2}(\rho_3^2+\nu_3^2)^{1/2}$$

$$=\sum_{\nu_3l_3}\int_0^\infty d\rho_3(\rho_3^2+\nu_3^2)^{1/2}(2l_1+1)^{1/2}(2l_2+1)^{1/2}X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}C\begin{pmatrix}l_1 & l_2 & l_3\\ m_1 & m_2 & m_3\end{pmatrix}\left|\begin{matrix}\sigma_3\ \nu_3\\ l_3\\ m_1+m_2\end{matrix}\right\rangle. \tag{4.4}$$

Since the state in SO(3, 1) differs from the state in SO(4) by a phase factor $[(J-i\rho)!(J'+i\rho)!/(J'-i\rho)!(J+i\rho)!]^{1/2}$, the CG coefficients of SO(3, 1) will also differ from the CG coefficients of SO(4) by a corresponding phase factor, which has been calculated by Dolginov and Toptygin[15] and Dolginov and Moskalev.[16]

$$|\phi(\sigma_1\sigma_2\sigma_3)|^2=\sinh\pi\rho_1\sinh\pi\rho_2\sinh\pi\rho_3[4\cosh\tfrac{1}{2}\pi(\rho_1+\rho_2+\rho_3)\cosh\tfrac{1}{2}\pi(\rho_1+\rho_2-\rho_3)$$

$$\times\cosh\tfrac{1}{2}\pi(\rho_1-\rho_2+\rho_3)\cosh\tfrac{1}{2}\pi(-\rho_1+\rho_2+\rho_3)]^{-1}. \tag{4.5}$$

The orthogonality and completeness relations for the X functions can be easily deduced. From (3.1) and (3.2) we have

$$\int\int b_{\sigma_1\sigma_2\sigma_3}(u_1u_2u_3)\overline{b_{\sigma_1\sigma_2\sigma_3'}}(u_1u_2u_3')du_1du_2=4\pi^4(\rho_3^2+\nu_3^2)^{-2}\delta(\sigma_3-\sigma_3')\delta_{\nu_3\nu_3'}\delta(u_3-u_3'), \tag{4.6}$$

$$\sum_{\nu_3}\int\int b_{\sigma_1\sigma_2\sigma_3}(u_1u_2u_3)\overline{b_{\sigma_1\sigma_2\sigma_3}}(u_1'u_2'u_3)(\rho_3^2+\nu_3^2)^2d\rho_3du_3=4\pi^4\delta(u_1-u_1')\delta(u_2-u_2'). \tag{4.7}$$

From (3.13), (4.6), and (4.7) we obtain

$$|\phi(\sigma_1\sigma_2\sigma_3)|^2\sum_{l_1l_2}(\rho_3^2+\nu_3^2)(2l_1+1)(2l_2+1)X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\bar{X}\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3'+\nu_3') & \frac{1}{2}(\sigma_3'-\nu_3') & l_3\end{vmatrix}=\delta(\sigma_3-\sigma_3')\delta_{\nu_3\nu_3'}, \tag{4.8}$$

$$\sum_{\nu_3 l_3}\int(\rho_3^2+\nu_3^2)d\rho_3\,|\phi(\sigma_1\sigma_2\sigma_3)|^2(2l_1+1)(2l_2+1)X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\overline{X}\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1'\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2'\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}$$

$$\times C\begin{pmatrix}l_1 & l_2 & l_3'\\ m_1 & m_2 & m_3\end{pmatrix}C\begin{pmatrix}l_1' & l_2' & l_3'\\ m_1' & m_2' & m_3'\end{pmatrix}=\delta_{l_1l_1'}\delta_{l_2l_2'}\delta_{m_1m_1'}\delta_{m_2m_2'},\tag{4.9}$$

$$\sum_{\nu_3}\int|\phi(\sigma_1\sigma_2\sigma_3)|^2(2l_1+1)(2l_2+1)(\rho_3^2+\nu_3^2)\left|X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\right|^2 d\rho_3=1.\tag{4.10}$$

5. CG COEFFICIENTS OF SO(3, 1) WITH MULTIPLICITY TWO

Just as in SO(2,1) there is a case of multiplicity two occurring in the decomposition $j_1\times j_2-j_3$, where all three j's are continuous, so also in SO(3,1) there is a case of multiplicity two occurring in the decomposition of two principal representations into a third one. This is because of the equivalence of the two representations (σ,ν) and $(-\sigma-2,-\nu)$. In fact, this statement is true for the decomposition of tensor products of the principal series of all SU(n,1) and SO(n,1). The solution is not difficult to find. First let us define the X function as follows:

$$X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & j_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & j_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & j_3\end{vmatrix}=[(2j_1+1)(2j_2+1)(\rho_3^2+\nu_3^2)]^{-1/2}\sum_{\substack{m_1\ m_2\\ t_1\ t_2}}C\begin{pmatrix}j_1 & j_2 & j_3\\ -m_1 & -m_2 & -m_1-m_2\end{pmatrix}$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_1-\nu_1) & \frac{1}{2}(\sigma_2-\nu_2) & \frac{1}{2}(\sigma_3-\nu_3)\\ -t_1+\frac{1}{2}(\sigma_1-\nu_1) & -t_2+\frac{1}{2}(\sigma_2-\nu_2) & -t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2)\end{pmatrix}$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_3+\nu_3)\\ -m_2-t_2+\frac{1}{2}(\sigma_2-\nu_2) & -m_1-t_1+\frac{1}{2}(\sigma_1-\nu_1) & -m_1-m_2-t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2)\end{pmatrix}$$

$$\times C\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & j_1\\ -m_1-t_1+\frac{1}{2}(\sigma_1-\nu_1) & t_1-\frac{1}{2}(\sigma_1-\nu_1) & -m_1\end{pmatrix}C\begin{pmatrix}\frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & j_2\\ -m_2-t_2+\frac{1}{2}(\sigma_2-\nu_2) & t_2-\frac{1}{2}(\sigma_2-\nu_2) & -m_2\end{pmatrix}$$

$$\times C\begin{pmatrix}j_3 & \frac{1}{2}(\sigma_3-\nu_3) & \frac{1}{2}(\sigma_3+\nu_3)\\ -m_1-m_2 & -t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2) & -m_1-m_2-t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2)\end{pmatrix}$$

$$\times(2j_3+1)^{1/2}(\sigma_3+\nu_3+1)^{-1/2}.\tag{5.1}$$

The advantage of Eq. (5.1) is that we have now

$$\overline{X}\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & j_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & j_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & j_3\end{vmatrix}=X\begin{vmatrix}-\frac{1}{2}(\sigma_1+\nu_1)-1 & -\frac{1}{2}(\sigma_1-\nu_1)-1 & j_1\\ -\frac{1}{2}(\sigma_2+\nu_2)-1 & -\frac{1}{2}(\sigma_2-\nu_2)-1 & j_2\\ -\frac{1}{2}(\sigma_3+\nu_3)-1 & -\frac{1}{2}(\sigma_3-\nu_3)-1 & j_3\end{vmatrix}.\tag{5.2}$$

This is because

$$\overline{C}\begin{pmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & j_1\\ -m_1-t_1+\frac{1}{2}(\sigma_1-\nu_1) & t_1-\frac{1}{2}(\sigma_1-\nu_1) & -m_1\end{pmatrix}=C\begin{pmatrix}-\frac{1}{2}(\sigma_1+\nu_1)-1 & -\frac{1}{2}(\sigma_1-\nu_1)-1 & j_1\\ -m_1-t_1'-\frac{1}{2}(\sigma_1-\nu_1)-1 & t_1'+\frac{1}{2}(\sigma_1-\nu_1)+1 & -m_1\end{pmatrix},$$

$$\text{where}\quad -t_1=t_1'+m_1+\nu_1+1,\tag{5.3}$$

$$\overline{C}\begin{pmatrix}j_3 & \frac{1}{2}(\sigma_3-\nu_3) & \frac{1}{2}(\sigma_3+\nu_3)\\ -m_1-m_2 & -t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2) & -m_1-m_2-t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2)\end{pmatrix}(-\sigma_3+\nu_3-1)^{-1/2}$$

$$=C\begin{pmatrix}j_3 & -\frac{1}{2}(\sigma_3-\nu_3)-1 & -\frac{1}{2}(\sigma_3+\nu_3)-1\\ -m_1-m_2 & -t_1'-t_2'-\frac{1}{2}(\sigma_1-\nu_1)-\frac{1}{2}(\sigma_2-\nu_2)-2 & -m_1-m_2-t_1'-t_2'-\frac{1}{2}(\sigma_1-\nu_1)-\frac{1}{2}(\sigma_2-\nu_2)-2\end{pmatrix}(-\sigma_3-\nu_3-1)^{-1/2},$$

$$\text{where}\quad -t_1=t_1'+m_1+\nu_1+1,\quad -t_2=t_2'+m_2+\nu_2+1,\tag{5.4}$$

$$C\begin{pmatrix} \frac{1}{2}(\sigma_1-\nu_1) & \frac{1}{2}(\sigma_2-\nu_2) & \frac{1}{2}(\sigma_3-\nu_3) \\ -t_1+\frac{1}{2}(\sigma_1-\nu_1) & -t_2+\frac{1}{2}(\sigma_2-\nu_2) & -t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2) \end{pmatrix}$$

$$= C\begin{pmatrix} -\frac{1}{2}(\sigma_2+\nu_2)-1 & -\frac{1}{2}(\sigma_1+\nu_1)-1 & -\frac{1}{2}(\sigma_3+\nu_3)-1 \\ -m_2-t_2'-\frac{1}{2}(\sigma_2-\nu_2)-1 & -m_1-t_1'-\frac{1}{2}(\sigma_1-\nu_1)-1 & -m_1-m_2-t_1'-t_2'-\frac{1}{2}(\sigma_1-\nu_1)-\frac{1}{2}(\sigma_2-\nu_2)-2 \end{pmatrix},$$

where $-t_1 = t_1' + m_1 + \nu_1 + 1, \quad -t_2 = t_2' + m_2 + \nu_2 + 1,$ (5.5)

$$C\begin{pmatrix} \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_3+\nu_3) \\ -m_2-t_2+\frac{1}{2}(\sigma_2-\nu_2) & -m_1-t_1+\frac{1}{2}(\sigma_1-\nu_1) & -m_1-m_2-t_1-t_2+\frac{1}{2}(\sigma_1-\nu_1)+\frac{1}{2}(\sigma_2-\nu_2) \end{pmatrix}$$

$$= C\begin{pmatrix} -\frac{1}{2}(\sigma_1-\nu_1)-1 & -\frac{1}{2}(\sigma_2-\nu_2)-1 & -\frac{1}{2}(\sigma_3-\nu_3)-1 \\ -t_1'-\frac{1}{2}(\sigma_1-\nu_1)-1 & -t_2'-\frac{1}{2}(\sigma_2-\nu_2)-1 & -t_1'-t_2'-\frac{1}{2}(\sigma_1-\nu_1)-\frac{1}{2}(\sigma_2-\nu_2)-2 \end{pmatrix},$$

where $-t_1 = t_1' + m_1 + \nu_1 + 1, \quad -t_2 = t_2' + m_2 + \nu_2 + 1.$ (5.6)

Equations (5.3)–(5.6) can be easily checked from the definition of CGC as ${}_3F_2$ functions with unit argument, as defined in Sec. 2. Let us also remember that the CGC can be written in different ways as we have shown in Sec. 2. In particular, the change of variables from t to t', so that $t = -t' - m - \nu - 1$, corresponds to completing the contour in Eq. (2.1), in the left half-plane instead of the right half-plane.

The fact that the CG coefficients of $(-\sigma_1 - 2, -\nu_1) \times (-\sigma_2 - 2, -\nu_2) \to (-\sigma_3 - 2, -\nu_3)$ is equal to the complex conjugate of the CGC of $(\sigma_1\nu_1) \times (\sigma_2\nu_2) \to (\sigma_3\nu_3)$ can also be seen from the original equation of Naimark, Eq. (3.3). It is clear from Eq. (3.3) that

$$\overline{b_{\sigma_1\sigma_2\sigma_3}} = b_{-\sigma_1-2,-\sigma_2-2,-\sigma_3-2}$$ (5.7)

since the ν's are not involved.

One can also see from Eq. (3.5) that this must be so. The complex conjugate of Eq. (3.5) is

$$(\sin\tfrac{1}{2}\Omega_{12})^{-2a} = \sum_{L_1} \overline{F_{L_1}}(a)(2L_1+1)d^{L_1}_{\nu_1 0}(\cos\Omega_{12})$$ (5.8)

and similar expressions for b and c. For the unitarily equivalent representations, it is

$$(\sin\tfrac{1}{2}\Omega_{12})^{-2a} = \sum_{L_1} F^u_{L_1}(a)(2L_1+1)d^{L_1}_{-\nu_1 0}(\cos\Omega_{12}).$$ (5.9)

Since

$$d^{L_1}_{\nu_1 0}(\cos\Omega_{12}) = (-1)^{\nu_1} d^{L_1}_{-\nu_1 0}(\cos\Omega_{12}),$$ (5.10)

$F^u_{L_1}(a)$ can differ from $\overline{F_{L_1}}(a)$ by at most a sign. From (2.14) and (2.35) we conclude that

$$\overline{X}\begin{vmatrix} \frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & j_1 \\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & j_2 \\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & j_3 \end{vmatrix}$$

can differ from

$$X\begin{vmatrix} \frac{1}{2}(-\sigma_1-\nu_1)-1 & \frac{1}{2}(-\sigma_1+\nu_1)-1 & j_1 \\ \frac{1}{2}(-\sigma_2-\nu_2)-1 & \frac{1}{2}(-\sigma_2+\nu_2)-1 & j_2 \\ \frac{1}{2}(-\sigma_3-\nu_3)-1 & \frac{1}{2}(-\sigma_3+\nu_3)-1 & j_3 \end{vmatrix}$$

by at most a sign. That the sign is + can be seen from the other arguments, as well as from substituting particular values into the expressions and checking them.

Thus the Clebsch–Gordan expansion of SO(3, 1) for the principal series can be written as

$$\left|\begin{matrix} \sigma_1 & \nu_1 \\ l_1 & \\ m_1 & \end{matrix}\right\rangle \times \left|\begin{matrix} \sigma_2 & \nu_2 \\ l_2 & \\ m_2 & \end{matrix}\right\rangle = \sum_{i=1,2} \sum_{\nu_3 l_3} \int_{-\infty}^{\infty} d\rho_3 \left|\begin{matrix} \sigma_3 & \nu_3 \\ l_3 & \\ m_3 & \end{matrix}\right\rangle C\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} C_i\begin{pmatrix} \sigma_1\nu_1 & \sigma_2\nu_2 & \sigma_3\nu_3 \\ l_1 & l_2 & l_3 \end{pmatrix},$$ (5.11)

where

$$C_1\begin{pmatrix}\sigma_1\nu_1 & \sigma_2\nu_2 & \sigma_3\nu_3\\ l_1 & l_2 & l_3\end{pmatrix}=N_1[(\rho_3^2+\nu_3^2)(2l_1+1)(2l_2+1)]^{1/2}\left[\phi(\sigma_1\sigma_2\sigma_3)X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\right.$$
$$\left.+\overline{\phi}(\sigma_1\sigma_2\sigma_3)\overline{X}\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\right], \tag{5.12}$$

where

$$N_1^2=1/(2+\phi^2+\overline{\phi}^2),\quad \phi^2=\rho_1\rho_2\rho_3\Gamma^2(a)\Gamma^2(b)\Gamma^2(c)\Gamma^2(a+b+c-1)(4\pi)^{-1}[\Gamma(a+b)\Gamma(a+c)\Gamma(b+c)]^{-2},$$
$$\overline{\phi}^2=\rho_1\rho_2\rho_3\Gamma^2(1-a)\Gamma^2(1-b)\Gamma^2(1-c)\Gamma^2(2-a-b-c)(4\pi)^{-1}[\Gamma(2-a-b)\Gamma(2-a-c)\Gamma(2-b-c)]^{-2}, \tag{5.13}$$

$$C_2\begin{pmatrix}\sigma_1\nu_1 & \sigma_2\nu_2 & \sigma_3\nu_3\\ l_1 & l_2 & l_3\end{pmatrix}=N_2[(2l_1+1)(2l_2+1)(\rho_3^2+\nu_3^2)]^{1/2}\left[\phi(\sigma_1\sigma_2\sigma_3)X\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}-e\overline{\phi}\overline{X}\begin{vmatrix}\frac{1}{2}(\sigma_1+\nu_1) & \frac{1}{2}(\sigma_1-\nu_1) & l_1\\ \frac{1}{2}(\sigma_2+\nu_2) & \frac{1}{2}(\sigma_2-\nu_2) & l_2\\ \frac{1}{2}(\sigma_3+\nu_3) & \frac{1}{2}(\sigma_3-\nu_3) & l_3\end{vmatrix}\right] \tag{5.14}$$

where

$$N_2^2=1/(2-\bar{e}\phi^2-e\overline{\phi}^2), \tag{5.15}$$
$$e=(1+\phi^2)/(1+\overline{\phi}^2), \tag{5.16}$$
$$\bar{e}=1/e. \tag{5.17}$$

Upon continuation to SO(4), by setting the phase factor $\phi=1$, C_1 becomes the X function of SO(4), and C_2 becomes zero. Moreover, it is easy to see that C_1 and C_2 are orthogonal, because both X and $\overline{X}$ are normalized according to their definition in (5.1). Since the X function is connected to $B_{l_1l_2l_3}$ by Eq. (3.35), the complex conjugate $\overline{X}$ is connected to $\overline{B}_{l_1l_2l_3}$ by taking the complex conjugate on both sides of Eq. (3.35).

ACKNOWLEDGMENT

We wish to thank Moorhead State University for financial support of this work.

APPENDIX: ORTHOGONALITY AND COMPLETENESS RELATIONS FOR THE REPRESENTATION FUNCTIONS OF SO (3, 1)

We give here a proof for the orthogonality and completeness relations of the d functions of SO(3, 1) for the following reasons. 1. We have not seen in the literature a direct proof of the orthogonality relations for the d functions. 2. Verdiev[35] gave, we think, an erroneous expression for the normalization constant in the orthogonality relation of the d functions of SO(3, 1). 3. Our method can be extended to other groups. Thus Gel'fand and Graev[36] have obtained the Plancherel formula for unimodular complex groups. One can then use our method to prove the orthogonality and completeness for the representation functions of unimodular complex groups based on Gel'fand and Graev's result.

The completeness relation is basically contained in Rühl,[30] Eq. (4.32). We shall therefore give a proof for the orthogonality relation only. The orthogonality relation is

$$\int D^{\sigma_1\nu_1}_{j_1m_1j_1'm_1'}(a)\overline{D}^{\sigma_2\nu_2}_{j_2m_2j_2'm_2'}(a)d\mu(a)$$
$$=8\pi^4(\nu_1^2+\rho_1^2)^{-1}\delta(\sigma_1-\sigma_2)\,\delta_{\nu_1\nu_2}\delta_{j_1j_2}\delta_{j_1'j_2'}\delta_{m_1m_2}\delta_{m_1'm_2'}, \tag{A1}$$

where

$$a=u_1du_2,$$
$$d\mu(a)=(4\pi)^{-1}\,d\mu(u_1)d\mu(u_2)\sinh^2\theta\,d\theta,$$
$$\int d\mu(u_i)=8\pi^2.$$

The limits of integration over θ are from 0 to ∞.

We need the following equations for the proof.

$$K(uu_1m\rho)=\pi\int x(u^{-1}ku_1)\,\alpha_{m\rho}(k)\,d_lk, \tag{A2}$$

$$x(a)=(2\pi)^{-4}\sum_{m=-\infty}^{\infty}\int_0^{\infty}\left[\int K(uu\bar{a}m\rho)\frac{\overline{\alpha(ua)}}{\overline{\alpha(u\bar{a})}}du\right]\times(m^2+\rho^2)\,d\rho, \tag{A3}$$

$$\langle j_1q_1|\,T_x^{m\rho}\,|j_2q_2\rangle=\int x(a)\,D^{m\rho}_{j_1q_1j_2q_2}(a)\,d\mu(a), \tag{A4}$$

$$u\bar{a}=K^{-1}ua, \tag{A5}$$

$$\alpha_{m\rho}(k)=|\lambda|^{i\rho-m-2}\lambda^m, \tag{A6}$$

$$m=2\nu,\quad i\rho=2(\sigma+1)=2i\rho_0. \tag{A7}$$

Equation (A2) is taken from Naimark,[37] Eq. (5), p. 205. Equation (A3) is the main content of Plancherel theorem for SO(3, 1) and is taken from Naimark,[37] Eq. (16), p.

232. Equation (A4) is a definition. Equation (A5) is also a definition, taken from Naimark,[37] p. 158. Equation (A6) is the multiplier, taken from Naimark,[37] p. 149. Equation (A7) just expresses the relation between different notations.

Proof of orthogonality, Eq. (A1)

Substituting Eq. (A3) in Eq. (A2), we get

$$K(u_1u_2m\rho)=\pi\int x(u_1^{-1}ku_2)\,\alpha_{m\rho}(k)\,d_lk$$

$$=\pi(2\pi)^{-4}\sum_{m'=-\infty}^{\infty}\int_0^{\infty}d\rho'(m'^2+\rho'^2)$$

$$\times\iint\alpha_{m\rho}(k)\,d_lkK(u,u\overline{u_1^{-1}ku_2}m'\rho')$$

$$\times\frac{\overline{\alpha_{m'\rho'}(uu_1^{-1}ku_2)}}{\overline{\alpha_{m'\rho'}(uu_1^{-1}ku_2)}}du\,. \tag{A8}$$

Now put $u=u_1$, then $uu_1^{-1}ku_2=k^{-1}u_1u_1^{-1}ku_2=u_2$, and

$$\frac{\overline{\alpha_{m'\rho'}(uu_1^{-1}ku_2)}}{\overline{\alpha_{m'\rho'}(u\overline{u^{-1}ku_2})}}=\frac{\overline{\alpha_{m'\rho'}(ku_2)}}{\overline{\alpha_{m'\rho'}(u_2)}}=\overline{\alpha_{m'\rho'}(k)}\,. \tag{A9}$$

Thus we get from (A8)

$$\pi(32\pi^4)^{-1}\iint\alpha_{m\rho}(k)\,\overline{\alpha_{m'\rho'}(k)}$$

$$\times d_lkdu=(m^2+\rho^2)^{-1}\delta(\rho-\rho')\,\delta_{mm'}\,. \tag{A10}$$

Now we start with Eq. (A4)

$$\langle j_1q_1|\,T_x^m\,|j_2q_2\rangle=\int x(a)\,D^m_{j_1q_1j_2q_2}(a)\,d\mu(a)$$

$$=\iint\overline{\phi^{j_1}_{(1/2)mq_1}}(u_1)\,K(u_1u_2m\rho)$$

$$\times\phi^{j_2}_{(1/2)mq_2}(u_2)\,du_1\,du_2$$

$$=\iiint\overline{\phi^{j_1}_{(1/2)mq_1}}(u_1)\,\pi x(u_1^{-1}ku_2)\,\alpha_{m\rho}(k)\,d_lk$$

$$\times\phi^{j_2}_{(1/2)mq_2}(u_2)\,du_1\,du_2\,, \tag{A11}$$

where

$$\phi^j_{(1/2)mq}(u)=(2j+1)^{1/2}u^j_{(1/2)mq}(u)\,.$$

Now put $x(a)=\overline{D^{m'\rho'}_{j_1'q_1'j_2'q_2'}}(a)$ on both sides of Eq. (A11) while writing

$$u_1^{-1}ku_2=u_a\,du_b\,,\quad d=\begin{pmatrix}e^{\theta/2}&0\\0&e^{-\theta/2}\end{pmatrix}. \tag{A12}$$

Then

$$x(u_1^{-1}ku_2)=\overline{D^{m'\rho'}_{j_1'q_1'j_2'q_2'}}(u_a\,du_b)$$

$$=\sum_{q'q''}\overline{u^{j_1'}_{q_1'q'}}(u_a)\,\overline{d^{m'\rho'}_{j'j'q'q''}}(\theta)\,\overline{u^{j_2'}_{q''q_2'}}(u_b)\,. \tag{A13}$$

Now use the integral representation of the boost matrix

$$d^{m'\rho'}_{j_1'j_2'q_1'q_2''}(\theta)$$

$$=\int\overline{u^{j_1'}_{(1/2)m'q_1'}}(u)\,|\lambda(u,d)|^{i\rho'-m'-2}[\lambda(u,d)]^{m'}$$

$$\times u^{j_2'}_{(1/2)m'q_2''}(u\bar{d})\,du(2j_1'+1)^{1/2}(2j_2'+1)^{1/2}\,. \tag{A14}$$

Substituting (A14) into (A13), we obtain

$$\overline{D^{m'\rho'}_{j_1'q_1'j_2'q_2'}}(u_a\,du_b)$$

$$=\int u^{j_1'}_{(1/2)m'q_1'}(uu_a^{-1})\,|\lambda(u,d)|^{-i\rho'-m'-2}\overline{\lambda(u,d)}^{m'}$$

$$\times\overline{u^{j_2'}_{(1/2)m'q_2'}}(u\bar{d}u_b)\,du(2j_1'+1)^{1/2}(2j_2'+1)^{1/2}\,. \tag{A15}$$

We now set

$$u_a=u_1^{-1}u\,. \tag{A16}$$

Then from (A12) we have

$$u_1^{-1}ku_2=u_au^{-1}ku_2=u_adu_b$$

or

$$u^{-1}ku_2=du_b\quad\text{or}\quad u_2=k^{-1}udu_b=u\bar{d}u_b\,. \tag{A17}$$

Substituting u_a and u_b from (A16) and (A17) into (A11) and integrating over u_1,u_2, remembering the orthogonality of SU(2) representation functions, we get

$$\int D^{m\rho}_{j_1q_1j_2q_2}(a)\,\overline{D^{m'\rho'}_{j_1'q_1'j_2'q_2'}}(a)\,d\mu(a)$$

$$=\pi\iint\alpha_{m\rho}(k)\,\overline{\alpha_{m'\rho'}}(k)\,d_lkdu$$

$$\times\delta_{j_1j_1'}\delta_{j_2j_2'}\delta_{q_1q_1'}\delta_{q_2q_2'}\delta_{mm'}\,, \tag{A18}$$

where we have used (A6). Finally, applying (A10) to (A18), we obtain

$$\int D^{m\rho}_{j_1q_1j_2q_2}(a)\,\overline{D^{m'\rho'}_{j_1'q_1'j_2'q_2'}}(a)\,d\mu(a)$$

$$=32\pi^4(m^2+\rho^2)^{-1}\delta(\rho-\rho')\,\delta_{mm'}\delta_{j_1j_1'}\delta_{q_1q_1'}\delta_{j_2j_2'}\delta_{q_2q_2'} \tag{A19}$$

$$=8\pi^4(\nu^2+\rho_0^2)^{-1}\delta(\rho_0-\rho_0')\,\delta_{\nu\nu'}\delta_{j_1j_1'}\delta_{q_1q_1'}\delta_{j_2j_2'}\delta_{q_2q_2'}\,. \tag{A20}$$

This completes the proof of the orthogonality relation, Eq. (A1). It is easy to obtain from (A20) the orthogonality relation for the boost matrix elements, since the SU(2) representation functions are themselves orthogonal. Thus we have

$$\int_0^{\infty}\sum_m d^{\sigma\nu}_{j_1j_2m}(\theta)\,\overline{d^{\sigma'\nu'}_{j_1j_2m}}(\theta)\sinh^2\theta\,d\theta$$

$$=(\tfrac{1}{2})\pi(2j_1+1)(2j_2+1)(\nu^2+\rho_0^2)^{-1}\delta(\rho_0-\rho_0')\,\delta_{\nu\nu'}\,. \tag{A21}$$

[1] S. Ström, Arkiv Fysik **29**, 467 (1965).
[2] S. Ström, Arkiv Fysik **33**, 465 (1966).
[3] S. Ström, Arkiv Fysik **38**, 373 (1968).
[4] D. W. Duc and N. V. Hieu, Ann. Inst. H. Poincaré **6**, 77 (1967).
[5] D. W. Duc and N. V. Hieu, Dokl. Akad. Nauk SSSR **172**, 1281 (1967) [Sov. Phys. Dokl. **12**, 312 (1967)].
[6] I. A. Verdiev and L. A. Dadashev, Yad. Fiz. **6**, 1094 (1967) [Sov. J. Nucl. Phys. **6**, 795 (1968)].
[7] A. Sciarrino and M. Toller, J. Math. Phys. **9**, 1252 (1967).
[8] R. L. Anderson, R. Rączka, M. A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1050 (1970).
[9] A. A. Makarov and G. I. Shepelev, Yad. Fiz. **12**, 1092 (1970) [Sov. J. Nucl. Phys. **12**, 596 (1971)].
[10] Ya. A. Smorodinskii and G. I. Shepelev, Yad. Fiz. **13**, 441 (1971) [Sov. J. Nucl. Phys. **13**, 248 (1971)].
[11] Ya. A. Smorodinskii and M. Huszar, Teor. Mat. Fiz. **4**, 328 (1970) [Theor. Math. Phys. **4**, 867 (1970)].
[12] M. A. Naimark, Tr. Mosk. Mat. Obsc. **8**, 121 (1959) [Amer. Math. Soc. Transl., Ser. 2, **36**, 101 (1964)].
[13] M. A. Naimark, Tr. Mosk. Mat. Obsc. **9**, 237 (1960) [Amer. Math. Soc. Transl., Ser. 2, **36**, 137 (1964)].

[14]M. A. Naimark, Tr. Mosk. Mat. Obsc. **10**, 181 (1961) [Amer. Math. Soc. Transl., Ser. 2, **36**, 189 (1964)].
[15]A. Z. Dolginov and I. N. Toptygin, Zh. Eksp. Teor. Fiz. **37**, 1441 (1959) [Sov. Phys.—JETP **5**, 1022 (1960)].
[16]A. Z. Dolginov and A. N. Moskalev, Zh. Eksp. Teor. Fiz. **37**, 1697 (1959) [Sov. Phys.—JETP **6**, 1202 (1960)].
[17]R. L. Anderson, R. Rączka, M. A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1059 (1970).
[18]W. Klink, J. Math. Phys. **10**, 1477 (1969).
[19]A. M. Gavrilik, Teor. Mat. Fiz. **11**, 56 (1972) [Theor. Math. Phys. **11**, 342 (1972)].
[20]G. Bisiacchi and C. Fronsdal, Nuovo Cimento **41**, 35 (1965).
[21]P. G. Bamberg, Proc. Roy. Soc. (London) A **300**, 337 (1967).
[22]I. A. Verdiev, Teor. Mat. Fiz. **16**, 360 (1973) [Theor. Math. Phys. **16**, 895 (1973)].
[23]I. A. Verdiev, G. A. Kerimov, and Ya. A. Smorodinskii, Yad. Fiz. **21**, 1351 (1975) [Sov. J. Nucl. Phys. **21**, 697 (1975)].
[24]M. K. F. Wong, J. Math. Phys. **15**, 25 (1974).
[25]D. Z. Freedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).
[26]G. Racah, Phys. Rev. **62**, 438 (1942).
[27]G. Domokos, Phys. Rev. **159**, 1387 (1967).
[28]M. K. F. Wong and H. Y. Yeh, J. Math. Phys. **16**, 800 (1975).
[29]A. O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[30]W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[31]A. Chakrabarti, J. Math. Phys. **9**, 2087 (1968).
[32]T. Maekawa, J. Math. Phys. **16**, 2372 (1975).
[33]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).
[34]J. L. Burchnall and T. W. Chaundy, Proc. London Math. Soc. **50**, 72 (1944).
[35]I. A. Verdiev, Zh. Eksp. Teor. Fiz. **55**, 1773 (1968) [Sov. Phys.—JETP **28**, 935 (1969)].
[36]I. M. Gel'fand and M. I. Graev, Tr. Mosk. Mat. Obsc. **4**, 375 (1955) [Amer. Math. Soc. Transl., Ser. 2, **9**, 123 (1958)].
[37]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, London, 1964).

Conformal Killing tensors in reducible spaces

G. J. Weir

Department of Mathematics, University of Canterbury, New Zealand
(Received 28 October 1976)

It is shown that the dimension of the vector space of second order, trace-free conformal Killing tensors (CKT's) in a Riemannian space of dimension $n(\geq 3)$ is bounded above by $(1/12)(n-1)(n+2)(n+3)(n+4)$ and that this is attained in flat space. The discussion is eventually restricted to four-dimensional spaces which admit a two-dimensional, Abelian, orthogonally transitive symmetry group, as well as one nonredundant CKT. A sufficient condition is given for an empty space to be Type D.

1. INTRODUCTION

A symmetric tensor, $Q_{\alpha\beta}$, satisfying

$$Q_{(\alpha\beta;\gamma)} = Q_{(\alpha} g_{\beta\gamma)} , \tag{1.1}$$

where Q_α is some vector and $g_{\alpha\beta}$ is the metric tensor, is called a conformal Killing tensor (CKT for short) of order two. If Q_α is zero, then it is a Killing tensor (KT) or order two. By definition, $Q_{\alpha\beta}$ is redundant when it may be written as a linear sum, with constant coefficients, of a multiple of the metric tensor and symmetrized products of conformal Killing vectors (CKV's). If $Q_{\alpha\beta}$ is a redundant KT, then it is a linear sum, with constant coefficients, of the metric tensor and symmetrized products of Killing vectors (KV's). For the rest of this paper, only second order KT's and CKT's will be discussed.

Equation (1.1) is the necessary and sufficient condition[1] for the scalar $Q_{\alpha\beta}P^\alpha P^\beta$ to be a constant along the null geodesic P^α. If Q_α is zero, this scalar is a constant along any geodesic, P^α. The main difference between a KT and CKT is that the latter are assumed trace-free, since the trace does not contribute to the scalar $Q_{\alpha\beta}P^\alpha P^\beta$ along a null geodesic.

Although CKT's appear in the literature on separation of variables for the Hamilton–Jacobi equation,[2] they have only recently come into prominence in general relativity when Walker and Penrose showed that Carter's[3] KT for empty, nonradiating Type D spaces does not generalize to all empty Type D spaces. Instead, the natural object which appeared was a second order CKT.

In Sec. 2 we shall show that the linear vector space formed from constant real sums of CKT's is bounded above. We closely follow the work of Hauser and Malhiot[4] (hereafter referred to as HM), who derived a similar result for KT's. In Sec. 3 we consider spaces admitting both CKV's and CKT's, while in Sec. 4 we specialize to four-dimensional reducible spaces with two commuting KV's. Throughout this paper we shall assume that the dimension of our space is greater than two.

2. THE VECTOR SPACE OF CKT's

In this section we shall consider a CKT in an n-dimensional Riemannian space. The following definitions:

$$L_{\alpha\beta\gamma} = 2Q_{\gamma[\beta;\alpha]} , \tag{2.1}$$

$$M_{\alpha\beta\gamma\delta} = \tfrac{1}{2}(L_{\alpha\beta[\gamma;\delta]} + L_{\gamma\delta[\alpha;\beta]}) , \tag{2.2}$$

$$H_{\alpha\beta} = Q_{[\alpha;\beta]} , \tag{2.3}$$

$$H_{\alpha\beta\gamma} = H_{\alpha\beta;\gamma} , \tag{2.4}$$

$$I_{\alpha\beta} = Q^\gamma{}_{;\gamma;(\alpha;\beta)} , \tag{2.5}$$

$$2\Theta_{\alpha\beta\gamma} = 3Q_{(\alpha} g_{\beta\gamma)} , \tag{2.6}$$

$$M_{\alpha\beta} = M_{\alpha\gamma\beta}{}^\gamma , \tag{2.7}$$

$$M = M_\alpha{}^\alpha , \tag{2.8}$$

together with

$$Q^\alpha{}_\alpha = 0 , \tag{2.9}$$

$$g^\alpha{}_\alpha = n , \tag{2.10}$$

will enable us to deduce in flat space that

$$L_{[\alpha\beta\gamma]} = 0 , \tag{2.11}$$

$$H_{[\alpha\beta\gamma]} = 0 , \tag{2.12}$$

$$M_{\alpha[\beta\gamma\delta]} = 0 , \tag{2.13}$$

$$I^\alpha{}_\alpha = 0 , \tag{2.14}$$

where $Q_{\alpha\beta}$ and Q_α are defined in Eq. (1.1). Our immediate aim is to verify Eq. (2.14) (the other symmetries are proved in HM), and then to derive the structural equations[5] for second order CKT's,

$$\nabla_\mu B = BA , \tag{2.15}$$

where B is a row vector formed from the elements of $Q_{\alpha\beta}, L_{\alpha\beta\gamma}, M_{\alpha\beta\gamma\delta}, H_{\alpha\beta}, H_{\alpha\beta\gamma}, I_{\alpha\beta}$; and the members of A are concomitants of the metric tensor, Riemann tensor, and and their finite covariant derivatives. Since Eq. (2.15) is a tensor equation, it is sufficient to verify it in flat space. Explicitly, we shall show that the first, second, third, and fourth derivatives of $Q_{\alpha\beta}$ are a linear sum of the variables $L_{\alpha\beta\gamma}$, $M_{\alpha\beta\gamma\delta}$ and $H_{\alpha\beta}$, $H_{\alpha\beta\gamma}$, and $I_{\alpha\beta}$, respectively in flat space, when $I_{\alpha\beta}$ is constant.

Assuming for now that $\partial_\alpha g_{\beta\gamma} = 0$, the following equations may be derived directly:

$$\tfrac{3}{2} Q_{\alpha\beta,\gamma} = L_{\gamma(\alpha\beta)} + \Theta_{(\alpha\beta\gamma)} , \tag{2.16}$$

$$L_{\alpha\beta\gamma,\delta} = M_{\alpha\beta\gamma\delta} + 2\Theta_{\gamma\delta[\beta,\alpha]} . \tag{2.17}$$

The integrability condition of Eq. (2.17),

$$M_{\alpha\beta\gamma[\delta,\mu]} = \Theta_{\gamma\delta[\alpha,\beta],\mu} + \Theta_{\gamma\mu[\beta,\alpha],\delta} ,$$

allows an expression to be found for the derivative of $M_{\alpha\beta\gamma\delta}$,

$$\begin{aligned} M_{\alpha\beta\gamma\delta,\mu} &= M_{\alpha\beta\gamma[\delta,\mu]} + M_{\alpha\beta\mu[\delta,\gamma]} + M_{\alpha\beta\delta[\mu,\gamma]} \\ &= \delta^{\phi\chi}_{\alpha\beta}\,\delta^{\psi\omega}_{\gamma\delta}\,\Theta_{\phi\omega\mu,\chi\psi} , \end{aligned} \tag{2.18}$$

where $\delta^{\phi\chi}_{\alpha\beta} = \delta^{\phi}_{\alpha}\delta^{\chi}_{\beta} - \delta^{\phi}_{\beta}\delta^{\chi}_{\alpha}$.

From Eqs. (2.9), (2.10), (2.1), and (1.1),

$$(n+2)Q_{\alpha} = 2L^{\beta}{}_{\alpha\beta}\,, \tag{2.19}$$

while Eqs. (2.17) and (2.19) imply

$$(n+1)Q_{(\alpha,\beta)} = M_{\alpha\beta} + Mg_{\alpha\beta}/n+2\,, \tag{2.20}$$

and so the bivector components, $Q_{[\alpha,\beta]}$, are the first derivatives of Q_{α} to be considered.

Because of Eq. (2.17), it is necessary to discuss a general second order derivative of Q_{α} when inspecting the third order derivatives of $Q_{\alpha\beta}$. From the identity

$$Q_{\alpha,\beta\gamma} \equiv Q_{(\alpha,\beta\gamma)} + \tfrac{2}{3}\left(Q_{[\alpha,\beta]\gamma} + Q_{[\alpha,\gamma]\beta}\right), \tag{2.21}$$

it suffices to show that $Q_{(\alpha,\beta\gamma)}$ is a linear sum of $H_{\alpha\beta\gamma}$. By operating on Eq. (2.17) with $g^{\alpha\gamma}\partial_{\mu}$,

$$nQ_{(\beta,\delta\mu)} = Q^{\alpha}{}_{,\alpha(\beta}g_{\delta\mu)} - Q_{(\beta}\,g_{\delta\mu)}{}^{,\alpha}{}_{\alpha}\,,$$

while the trace over β and δ in this equation gives

$$2(n+1)Q_{\mu}{}^{,\alpha}{}_{\alpha} + (n-2)Q^{\alpha}{}_{,\alpha\mu} = 0\,, \tag{2.22}$$

which, with the expression $2g^{\alpha\beta}H_{\alpha\mu\beta} = Q^{\alpha}{}_{,\alpha\mu} - Q_{\mu}{}^{,\alpha}{}_{\alpha}$, allows us to write Eq. (2.21) as a linear sum of the $H_{\alpha\beta\gamma}$;

$$nQ_{(\alpha,\beta\gamma)} = 2g_{(\alpha\beta}H^{\mu}{}_{\gamma)\mu} = \frac{3n}{2(n+1)}\,Q^{\mu}{}_{,\mu(\alpha}g_{\beta\gamma)}\,. \tag{2.23}$$

Consequently, the fourth order derivatives of $Q_{\alpha\beta}$ are spanned by terms of the form $Q_{[\alpha,\beta]\gamma\delta}$. These expressions may be written as a sum involving only the $I_{\alpha\beta}$, which are defined in Eq. (2.5). This is verified by differentiating Eq. (2.23) w.r.t. ϵ, and antisymmetrizing over the γ and ϵ components. Then operating on Eq. (2.22) with $g^{\mu\rho}\partial_{\rho}$ gives $I_{\alpha\beta}$ as trace-free,

$$I^{\alpha}{}_{\alpha} = 0\,.$$

From Eqs. (2.23) and (2.22),

$$0 = Q_{(\alpha,\beta\gamma)}{}^{\mu}{}_{\mu} = \left((2-n)n/2(n+1)\right)Q^{\mu}{}_{,\mu\alpha\beta\gamma}\,,$$

and so the $I_{\alpha\beta}$ components are constant in flat space whenever $n \geq 3$, establishing Eq. (2.15). We shall now drop the assumption above that the space is flat.

Theorem 1: Let R be a Riemannian space of dimension n, where $n \geq 3$. Then in R, the dimension of the real vector space formed from trace-free, second order conformal Killing tensors is at most N_n, where $N_n = (n-1)(n+2)(n+3)(n+4)/12$.

Proof: The symmetries in Eqs. (2.9) and (2.11)–(2.14) imply that $Q_{\alpha\beta}, L_{\alpha\beta\gamma}, M_{\alpha\beta\gamma\delta}, H_{\alpha\beta}, H_{\alpha\beta\gamma}, I_{\alpha\beta}$ contribute $\frac{1}{2}(n-1)(n+2), \frac{1}{3}(n-1)n(n+1), \frac{1}{12}(n-1)n^2(n+1), \frac{1}{2}(n-1)n, \frac{1}{3}(n-1)n(n+1), \frac{1}{2}(n-1)(n+2)$ elements, respectively, to the B in Eq. (2.15). Their sum is N_n.

We have not proved Eq. (2.14) for a nonflat space, but nonzero contributions can result only from the noncommutivity of covariant differentiation, and so the trace $I^{\alpha}{}_{\alpha}$ will depend linearly on the other members of B. Finally, the CKT, $Q_{\alpha\beta}$, is a linear sum of the fundamental solutions of Eq. (2.15), proving Theorem 1.

Having established an upper bound for the number of linearly independent CKT's in a Riemannian space, we shall show that this is attained in flat space. The general solution for a CKV, ξ^{α}, in a space with constant metric $\eta^{\alpha\beta}$ is

$$\xi^{\alpha} = A_{\theta}x^{\theta}x^{\alpha} - \tfrac{1}{2}A^{\alpha}x_{\theta}x^{\theta} + x^{\theta}\omega_{\theta}{}^{\alpha} + \phi_{0}x^{\alpha} + B^{\alpha}\,, \tag{2.24}$$

where $\xi^{(\alpha;\beta)} = (A_{\theta}x^{\theta} + \phi_{0})\eta^{\alpha\beta}$; and A_{α}, $\omega_{\alpha\beta} = \omega_{[\alpha\beta]}$, ϕ_{0}, and B_{α} are real constants. The inner product, $\xi^{\alpha}P_{\alpha}$, of a CKV and a null geodesic, P_{α}, is constant along P_{α}. From Eq. (2.24), the terms

$$P_{\alpha}, L_{\alpha\beta} = x_{[\alpha}P_{\beta]}\,, \quad \phi = x_{\alpha}P^{\alpha}\,, \quad c_{\alpha} = \phi x_{\alpha} - P_{\alpha}x_{\theta}x^{\theta}\,, \tag{2.25}$$

are therefore constant along P_{α}, and so the expressions

$$P_{\alpha}P_{\beta},\ L_{\alpha\beta}L_{\gamma\delta},\ \phi^2,\ c_{\alpha}c_{\beta},\ L_{\alpha\beta}P_{\gamma},\ \phi P_{\alpha},\ c_{\alpha}P_{\beta},\ \phi L_{\alpha\beta},\ L_{\alpha\beta}c_{\gamma},\ \phi c_{\alpha} \tag{2.26}$$

are second order CKT's, and we shall show that N_n of them are linearly independent. The equations

$$\begin{aligned}
&\eta^{\alpha\beta}P_{\alpha}P_{\beta} = 0\,,\quad L_{[\alpha\beta}P_{\gamma]} = 0\,,\quad L_{\alpha[\beta}L_{\gamma\delta]} = 0\,,\ L_{[\alpha\beta}c_{\gamma]} = 0\,,\\
&\eta^{\alpha\beta}c_{\alpha}c_{\beta} = 0\,,\quad \phi^2 = 2\eta^{\alpha\beta}c_{\alpha}P_{\beta} + 2\eta^{\alpha\beta}\eta^{\gamma\delta}L_{\alpha\gamma}L_{\beta\delta}\,,\\
&\phi L_{\alpha\beta} = c_{[\alpha}P_{\beta]}\,,\quad 2c_{(\alpha}P_{\beta)} = \eta^{\gamma\delta}L_{\alpha\gamma}L_{\delta\beta}\,,\\
&\phi P_{\alpha} = \eta^{\beta\gamma}L_{\alpha\beta}P_{\gamma}\,,\quad \phi c_{\alpha} = \eta^{\beta\gamma}L_{\alpha\beta}c_{\gamma}\,,
\end{aligned} \tag{2.27}$$

split Eq. (2.26) into the five sets $\{P_{\alpha}P_{\beta}\}, \{L_{\alpha\beta}P_{\gamma}\}, \{L_{\alpha\beta}L_{\gamma\delta}, c_{[\alpha}P_{\beta]}\}, \{L_{\alpha\beta}c_{\gamma}\}, \{c_{\alpha}c_{\beta}\}$, and as these are directly analogous to the sets $\{Q_{\alpha\beta}\}, \{M_{\alpha\beta\gamma\delta}, H_{\alpha\beta}\}, \{H_{\alpha\beta\gamma}\}, \{I_{\alpha\beta}\}$ which were mentioned in Theorem 1, there are at most N_n linearly independent members in Eq. (2.26).

The only relationship among the members of the set $\{P_{\alpha}P_{\beta}\}$ is $\eta^{\alpha\beta}P_{\alpha}P_{\beta} = 0$, which is given in Eq. (2.27). To find the linear dependencies among the members of the set $\{L_{\alpha\beta}P_{\gamma}\}$, we shall consider the equation $A_{\alpha\beta\gamma}x^{\alpha}P^{\beta}P^{\gamma} = 0$, where $A_{\alpha\beta\gamma} = A_{[\alpha\beta]\gamma}$ is a constant tensor. Then

$$A_{\alpha\beta\gamma} = D_{\alpha[\beta\gamma]} + B_{\alpha}\eta_{\beta\gamma}\,, \tag{2.28}$$

where $D_{\alpha\beta\gamma}$ and B_{α} are constant tensors. Taking the trace over α and β gives $B_{\alpha} = D_{\beta\alpha}{}^{\beta}$, while symmetrizing Eq. (2.28) over α and β and taking the trace over β and γ gives $(n+1)B_{\alpha} + D_{\beta\alpha}{}^{\beta} = 0$, whence B_{α} is zero, and $A_{\alpha\beta\gamma}$ is antisymmetric in all of its indices. This restriction is given in Eq. (2.27), and so there can be no more independent conditions on $A_{\alpha\beta\gamma}$. The sets $\{L_{\alpha\beta}L_{\gamma\delta}, c_{[\alpha}P_{\beta]}\}$ and $\{L_{\alpha\beta}c_{\gamma}\}$ are treated similarly. The final equation is $A_{\alpha\beta}c^{\alpha}c^{\beta} = 0$, where $A_{\alpha\beta} = A_{(\alpha\beta)}$ is a constant tensor. We see that

$$\eta^{\rho\sigma}\eta^{\mu\nu}\partial_{\rho}\partial_{\sigma}\partial_{\mu}\partial_{\nu}(A_{\alpha\beta}c^{\alpha}c^{\beta}) = n(n-2)A_{\alpha\beta}P^{\alpha}P^{\beta}\,,$$

and so for $n \geq 3$, $A_{\alpha\beta} = \lambda\eta_{\alpha\beta}$, establishing

Theorem 2: Let F be a flat Riemannian space of dimension n, where $n \geq 3$. Then in F, the dimension of the linear vector space formed from second order, trace-free conformal Killing tensors is precisely $(n-1)(n+2)(n+3)(n+4)/12$.

Having shown that 84 linearly independent CKT's are possible in four dimensions, it is natural to enquire precisely how many are present in a given four-dimensional Lorentzian space. An indication of the difficulty of this problem can be gauged from that en-

countered with the integrability conditions for KV's, CKV's, and second order KT's when it is necessary to consider third, fourth, and fourth order derivatives, respectively, of the basic variables. We must inspect the sixth order derivatives of $Q_{\alpha\beta}$ when looking at the integrability conditions for a CKT, and so this appears an intractible problem with the techniques used so far.

3. SPACES WITH CKT's AND CKV's

In this section we shall derive canonical forms for the metric and CKT in spaces which admit one nonredundant CKT and either one or two commuting CKV's. This task is simplified by using some results of Nijenhuis.[6]

Let $\mathfrak{S}^q(M)$ be the space of qth order contravariant symmetric tensors. If $P\in\mathfrak{S}^p(M)$, $Q\in\mathfrak{S}^q(M)$, we define

$$P\cap Q = P^{(\alpha_1\cdots\ \alpha_p}Q^{\alpha_{p+1}\cdots\alpha_{p+q})}\,, \tag{3.1}$$

$$[P,Q] = pP^{\rho(\alpha_1\cdots\alpha_{p-1}}\partial_\rho Q^{\alpha_{p+1}\cdots\alpha_{p+q})} - qQ^{\rho(\alpha_1\cdots\alpha_{q-1}}\partial_\rho P^{\alpha_{q+1}\cdots\alpha_{q+p})}\,. \tag{3.2}$$

These operations satisfy the following equations:

$$[P,[Q,R]]+[Q,[R,P]]+[R,[P,Q]]=0\,, \tag{3.3}$$

$$[P,Q\cap R]=[P,Q]\cap R+Q\cap[P,R]\,, \tag{3.4}$$

$$[V,Q]=\vec{\mathfrak{L}}_V Q,\quad V\in\mathfrak{S}^1(M)\,, \tag{3.5}$$

$$[G,Q]=2Q^{(\alpha_1\cdots\alpha_q;\rho)},\quad Q\in\mathfrak{S}^q(M)\,, \tag{3.6}$$

where G is the metic tensor in contravariant form, and P, Q, and R are contravariant symmetric tensors. From Eq. (3.6), a CKV, K, and a CKT, Q, satisfy $[G,K]=\phi G$, $[G,Q]=L\cap G$, where $\phi\in\mathfrak{S}^0(M)$, $L\in\mathfrak{S}^1(M)$. KV's and KT's have ϕ and L respectively zero.

A. One CKV

We shall assume that the space allows one nonredundant CKT, Q, and one CKV, V,

$$[G,K]=\phi G\,, \tag{3.7}$$

$$[G,Q]=L\cap G\,, \tag{3.8}$$

and that any other CKT, $\tilde{Q}$, may be written as

$$\tilde{Q}=\psi G+aK\cap K+bQ\,, \tag{3.9}$$

where a and b are constants, ψ and ϕ belong to $\mathfrak{S}^0(M)$, and $L\in\mathfrak{S}^1(M)$. (If Q were redundant, b would be unnecessary.) From Eqs. (3.3) and (3.4),

$$[G,[K,Q]]=([K,L]+[\phi,Q])\cap G\,, \tag{3.10}$$

and so $[K,Q]$ is another CKT, implying from Eq. (3.9) that $[K,Q]=\psi G+aK\cap K+bQ$. The quantities ψ, a, b are not invariant, however, and if we define

$$K'=cK,\quad Q'=qQ+pK\cap K+\eta G\,, \tag{3.11}$$

then

$$[K',Q']=\psi' G+a'K'\cap K'+b'Q'\,, \tag{3.12}$$

where

$$\psi'=c([K,\eta]+q\psi-\eta(\phi+b))\,,$$

$$a'=(aq-bp)/c\,,$$

$$b'=bc\,,$$

and c,q,p are constants, $\eta\in\mathfrak{S}^0(M)$. Consequently, η can always be chosen so that $\psi'=0$, and if $b\neq 0$, we can set $b'=1$, $a'=0$. If $b=0$, we may choose $a'=0$ or 1, and so the canonical forms for $[K,Q]$ are (i) $[K,Q]=Q$; (ii) $[K,Q]=K\cap K$; (iii) $[K,Q]=0$. We shall choose $K=K^\mu\partial_\mu=\partial_1$, whence

$$G=\psi\tilde{G}(x^2,\ldots,x^n)\,,\quad [G,K]=-(\partial_1\log\psi)G\,,$$

(i) $Q=e^{x^1}\tilde{Q}(x^2,\ldots,x^n)$,

(ii) $Q=x^1K\cap K+\tilde{Q}(x^2,\ldots,x^n)$,

(iii) $Q=Q(x^2,\ldots,x^n)$.

If K and Q are KV's and KT's respectively, then the ψ and η terms in Eq. (3.9) and Eq. (3.11) are constants, and in general we cannot set ψ' in Eq. (3.12) to zero. The canonical forms are

$$K=\partial_1,\quad G=G(x^2,\ldots,x^n)\,,$$

$$Q=e^{x^1}\tilde{Q}(x^2,\ldots,x^n)\,,\quad b\neq 0\,,$$

$$Q=(a_1K\cap K+a_2G)x^1+\tilde{Q}(x^2,\ldots,x^n)\,,\quad b=0\,,$$

where a_1 and a_2 are constants.

B. Two commuting CKV's

We shall end this section by extending the previous analysis to the case of one nonredundant CKT, Q, and two commuting CKV's, K_i $(i=1,2)$,

$$[G,K_i]=\phi_i G,\quad [G,Q]=L\cap G\,,\quad [K_i,K_j]=0\,, \tag{3.13}$$

$$[K_i,Q]=\psi_i G+a_iQ+b_i{}^{st}K_s\cap K_t\,, \tag{3.14}$$

where $\psi_i\in\mathfrak{S}^0(M)$ and $a_i,b_i{}^{st}$ are constants. From Eqs. (3.13) and (3.3),

$$[K_j,[K_i,Q]]=[K_i,[K_j,Q]]\,, \tag{3.15}$$

and substituting Eq. (3.14) into Eq. (3.15),

$$K_s\cap K_t(a_ib_j{}^{st}-a_jb_i{}^{st})+([K_j,\psi_i]-[K_i,\psi_j] +\psi_j(\phi_i+a_i)-\psi_i(\phi_j+a_j))G=0\,. \tag{3.16}$$

In spaces of dimension greater than two the coefficients of $K_s\cap K_t$ and G in Eq. (3.16) must both be zero, whence

$$b_j{}^{st}=a_jb^{st}\,, \tag{3.17}$$

for some b^{st} [provided $(a_1,a_2)\neq(0,0)$],

$$[K_j,\psi_i]-[K_i,\psi_j]+\psi_j(\phi_i+a_i)-\psi_i(\phi_j+a_j)=0\,, \tag{3.18}$$

while Eqs. (3.3) and (3.13) imply

$$[K_i,\phi_j]=[K_j,\phi_i]\,. \tag{3.19}$$

We shall now derive the canonical forms for Q and G. Under the transformation

$$K_{i'}=c_{i'}{}^iK_i\,,\quad Q'=qQ+p^{st}K_s\cap K_t+\eta G\,, \tag{3.20}$$

the coefficients in Eq. (3.14) become

$$\begin{aligned}
&\psi_{i'}=c_{i'}{}^i(q\psi_i+[K_i,\eta]-\eta(\phi_i+a_i))\,,\\
&a_{i'}=c_{i'}{}^ia_i\,,\\
&b_{i'}{}^{s't'}=c_{i'}{}^ic^{s'}{}_sc^{t'}{}_t(qb_i{}^{st}-a_ip^{st})\,,\\
&c_{i'}{}^ic^{i'}{}_j=\delta^i_j\,,
\end{aligned} \tag{3.21}$$

and so the integrability condition for $\psi_{i'}=0$ is $K_j[K_i,\eta]$

$= K_i[K_j, \eta]$, or

$$0 = K_j(q\psi_i - \eta(\phi_i + a_i)) - K_i(q\psi_j - \eta(\phi_j + a_j)), \quad (3.22)$$

where

$$q\psi_i + [K_i, \eta] - \eta(\phi_i + a_i) = 0. \quad (3.23)$$

However, Eq. (3.22) is satisfied because of Eqs. (3.23), (3.19), and (3.18), and so we shall transform $\psi_{i'} = 0$. If $(a_1, a_2) \neq (0,0)$, the transformation $K_{i'} = c_{i'}{}^i K_i$ may be used to set $(a_1, a_2) = (1, 0)$. From Eq. (3.17), we may set $b_{i'}{}^{s't'}$ to zero by choosing $p^{st} = qb^{st}$ in Eq. (3.20), and so $[K_1, Q] = Q$, $[K_2, Q] = 0$ when $(a_1 a_2) \neq (0,0)$.

The canonical forms are

$$K_i = K_i{}^\mu \partial_\mu = \delta_i^\mu \partial_\mu,$$

$$G = e^{-\phi} G(x^3, \ldots, x^n), \quad [G, K_i] = \phi_{,i} G,$$

$$Q = e^{x^1} \tilde{Q}(x^3, \ldots, x^n), \quad (a_1, a_2) \neq (0,0),$$

$$Q = x^i b_i{}^{st} K_s \cap K_t + \tilde{Q}(x^3, \ldots, x^n), \quad (a_1, a_2) = (0,0).$$

Since we have not assumed that $\phi_{,i}$ is zero, these results hold when K_i are both KV's. This completes the solution for the CKT. If Q is a KT, and K_i are two KV's, we find as our canonical forms

$$[K_1, Q] = Q, \quad [K_2, Q] = 0, \quad (a_1, a_2) \neq (0,0),$$

$$[K_i, Q] = \psi_i G + b_i{}^{st} K_s \cap K_t, \quad (a_1, a_2) = (0,0).$$

The author has not investigated CKT's which depend on the ignorable Killing coordinates. In the next section we shall consider only CKT's which are independent of the ignorable coordinates.

4. CKT's IN QUASIDIAGONAL SPACES

A reducible space[7] is one for which the metric tensor allows r commuting KV's, ∂_a, and s nonignorable coordinates, $x^{\mathfrak{a}}$, to be chosen such that $g^{a\mathfrak{a}} = 0$. A quasidiagonal space (qd for short) is a four-dimensional reducible space with two commuting KV's, i.e., $r = 2$, $s = 2$. (The KV's are ∂_1 and ∂_2; $a, b, \cdots \in \{1, 2\}$; $\mathfrak{A}, \mathfrak{B}, \cdots \in \{3, 4\}$.) We shall briefly discuss reducible spaces, and then specialize to qd spaces, assuming in the future that the CKT is independent of the ignorable coordinates, x^a,

$$g^{a\mathfrak{a}} = 0, \quad \overset{\rightarrow}{\Omega}_{\partial_a} Q^{\alpha\beta} = 0,$$

where α runs from 1 to n, and $n = r + s$.

From Eq. (1.2), the equations to be solved,

$$g^{\mathfrak{a}(\mathfrak{b}} \partial_{\mathfrak{a}} Q^{\mathfrak{c}\mathfrak{d})} - Q^{\mathfrak{a}(\mathfrak{b}} \partial_{\mathfrak{a}} g^{\mathfrak{c}\mathfrak{d})} = Q^{(\mathfrak{b}} g^{\mathfrak{c}\mathfrak{d})}, \quad (4.1a)$$

$$g^{\mathfrak{a}\mathfrak{b}} \partial_{\mathfrak{a}} Q^{ab} - Q^{a\mathfrak{b}} \partial_{\mathfrak{a}} g^{ab} = Q^{\mathfrak{b}} g^{ab}, \quad (4.1b)$$

$$-Q^{\mathfrak{a}(a} \partial_{\mathfrak{a}} g^{bc)} = Q^{(a} g^{bc)}, \quad (4.2a)$$

$$g^{\mathfrak{a}(\mathfrak{b}} \partial_{\mathfrak{a}} Q^{\mathfrak{c}\,a)} - Q^{\mathfrak{a}(\mathfrak{b}} \partial_{\mathfrak{a}} g^{\mathfrak{c}a)} = \tfrac{1}{3} Q^a g^{\mathfrak{b}\mathfrak{c}}, \quad (4.2b)$$

split into two sets, since Eqs. (4.1) involves only $Q^{a\mathfrak{b}}$, Q^{ab}, and $Q^{\mathfrak{a}}$ while Eqs. (4.2) involve only $Q^{\mathfrak{a}a}$ and Q^a. Equations (4.1a) represents the general solution of a CKT in an s-dimensional space, while from Eq. (3.7), Eq. (4.2b) states that the r vectors, $Q^{a\mathfrak{a}}\partial_{\mathfrak{a}}$, are CKV's for the s-dimensional metric tensor, $g^{\mathfrak{a}\mathfrak{b}}$. These are CKV's for the whole space whenever $-Q^{\mathfrak{a}a}\partial_{\mathfrak{a}} g^{bc} = Q^a g^{bc}$, and so the CKT $Q^{\mathfrak{a}a}\partial_{(\mathfrak{a}} \otimes \partial_{a)}$ is redundant whenever $r = 1$, being the symmetrized product of a CKV, $Q^{a\mathfrak{a}}\partial_{\mathfrak{a}}$, and a KV, ∂_a.

We shall look for canonical forms for the pair $(g^{a\mathfrak{b}}, Q^{a\mathfrak{b}})$, by inspecting the eigenvalue equation

$$(Q^{\mathfrak{a}}{}_{\mathfrak{b}} - \lambda \delta^{\mathfrak{a}}_{\mathfrak{b}}) y^{\mathfrak{b}} = 0. \quad (4.3)$$

If this equation has distinct eigenvalues, then both $Q^{a\mathfrak{b}}$ and $g^{a\mathfrak{b}}$ can be diagonalized, but in general the mutually orthogonal eigenvectors will not form coordinate surfaces, and so cannot be chosen as coordinates. This has led in the past to technical difficulties which have only been circumvented by imposing additional conditions[8] on the eigenvectors of Eq. (4.3). Consequently, we shall not consider the general case further, but restrict the discussion to qd spaces. Then the eigenvectors can be chosen as coordinates, and Eq. (4.3) can be subdivided into the three cases when the eigenvalues are complex (Case 1a), real and distinct (Case 1b), and degenerate (Case 1c). The solution of the $Q^{a\mathfrak{a}}$ components will be denoted by Case 2. We ignore the case of $Q^{\mathfrak{a}\mathfrak{b}} = 0$, since this leads to a redundant CKT.

Case 1a: λ a complex eigenvalue, $y^{\mathfrak{a}}$, $\overline{y^{\mathfrak{a}}}$ eigenvectors

Let us use complex coordinates, $z^{\mathfrak{a}}$, such that $\delta^{\mathfrak{a}}_3 = y^{\mathfrak{a}}$, $\delta^{\mathfrak{a}}_4 = \overline{y^{\mathfrak{a}}}$. Then $Q^{\mathfrak{a}\mathfrak{b}}$ and $g^{\mathfrak{a}\mathfrak{b}}$ are diagonal, and from Eq. (4.1a),

$$Q^3 = g^{33}\partial_3 b, \quad Q^4 = -g^{44}\partial_4 a,$$

$$g^{\mathfrak{a}\mathfrak{b}}\partial_{(\mathfrak{a}} \otimes \partial_{\mathfrak{b})} = (R_3 \partial_3 \otimes \partial_3 + X_4 \partial_4 \otimes \partial_4)/a + b,$$

$$Q^{\mathfrak{a}\mathfrak{b}}\partial_{(\mathfrak{a}} \otimes \partial_{\mathfrak{b})} = (bR_3 \partial_3 \otimes \partial_3 - aX_4 \partial_4 \otimes \partial_4)/a + b,$$

where

$$Q^{33} = bg^{33}, \quad Q^{44} = -ag^{44},$$

and R_3 and X_4 are analytic functions of z^3 and z^4, respectively. The reality of G and Q imply that $X_4 = \overline{R_3}$, $b = \overline{a}$. We have assumed that $a + b \neq 0$, otherwise

$$Q = bG + r_1 K_1 \cap K_1 + r_2 K_1 \cap K_2 + r_3 K_2 \cap K_2, \quad (4.4)$$

where b is a real function of z^3 and z^4, r_1, r_2, and r_3 are real constants, and K_1 and K_2 are the KV's ∂_1 and ∂_2, respectively. This is a redundant CKT. If Q is a KT, then b is a real constant, and again $a + b = 0$ leads to a redundant KT. Ignoring the possibility that $a + b = 0$, we find from Eq. (4.1b),

$$\partial_3(Q^{cd} - bg^{cd}) = 0, \quad \partial_4(Q^{cd} + ag^{cd}) = 0,$$

and so

$$G = g^{\alpha\beta}\partial_{(\alpha} \otimes \partial_{\beta)}$$
$$= ((c_3 + d_4)\partial_1 \otimes \partial_1 + 2(e_3 + f_4)\partial_{(1} \otimes \partial_{2)}$$
$$+ (g_3 + h_4)\partial_2 \otimes \partial_2 + R_3 \partial_3 \otimes \partial_3 + X_4 \partial_4 \otimes \partial_4)/a + b, \quad (4.5)$$

$$Q = Q^{\alpha\beta}\partial_{(\alpha} \otimes \partial_{\beta)}$$
$$= ((bc_3 - ad_4)\partial_1 \otimes \partial_1 + 2(be_3 - af_4)\partial_{(1} \otimes \partial_{2)}$$
$$+ (bg_3 - ah_4)\partial_2 \otimes \partial_2 + bR_3 \partial_3 \otimes \partial_3 - aX_4 \partial_4 \otimes \partial_4))/a + b, \quad (4.6)$$

$$Q^3 = g^{33}\partial_3 b, \quad Q^4 = -g^{44}\partial_4 a, \quad (4.7)$$

where R_3, c_3, e_3, g_3 and X_4, d_4, f_4, h_4 are analytic functions of z^3 and z^4 respectively, and $\overline{b} = a$, $X_4 = \overline{R_3}$, $d_4 = \overline{c_3}$, $f_4 = \overline{e_3}$, $h_4 = \overline{g_3}$. If Q is a KT, then a and b are analytic

functions of z^3 and z^4, respectively, and there is essentially no difference between the given KT and the metric, since the KT in Eq. (4.6) may be written in the form of the metric in Eq. (4.5) by an obvious redefinition of the functions $a,b,c_3,d_4,e_3,f_4,R_3,X_4$, and interchanging the Killing vectors. In fact the symmetric form of G and Q in Eq. (1.1) when Q_α is zero shows that if Q is regarded as the metric, then G is its KT.

The metric in Eq. (4.5) may be written in terms of real coordinates by defining $z^3 = x^3 + ix^4$, where x^3 and x^4 are real. Then

$$G = (A\partial_1 \otimes \partial_1 + 2B\partial_{(1}\otimes\partial_{2)} + C\partial_2\otimes\partial_2 + D\partial_3\otimes\partial_3 + 2E\partial_{(3}\otimes\partial_{4)} - D\partial_4\otimes\partial_4)/\psi ,$$

where A,B,C,D,E are real harmonic functions of x^3 and x^4, and ψ is harmonic iff the CKT is a KT. The corresponding Q in Eq. (4.5) is pure imaginary, but as any CKT is only defined up to a multiplicative constant, we can always make Q real.

Case 1b: Two real eigenvalues

Let us choose our coordinates so that the linearly independent eigenvectors are $\delta_3^\mu\partial_\mu$ and $\delta_4^\mu\partial_\mu$. Then g^{ab} and Q^{ab} are diagonal, and the preceding analysis of Case 1a applies, so that the metric and CKT are given in Eq. (4.5) and Eq. (4.6), respectively, but now R_3, c_3,e_3,g_3, and X_4,d_4,f_4,h_4 are (not necessarily analytic) real functions of x^3 and x^4, respectively.

Case 1c: A single eigenvalue

Nonredundant CKT's exist only if signature $(g^{ab}) = 0$, and so we shall choose as our nonignorable coordinates the two real null directions for g^{ab}. Then the condition for a single eigenvalue is $Q^{33}Q^{44} = 0$, and without loss of generality we take $Q^{33} = 0$. If Q^{44} is zero also, we have the redundant solution of Eq. (4.4), and so we shall assume that Q^{44} is nonzero. Our canonical forms are

$$g^{ab} = \begin{pmatrix} 0 & g^{34} \\ g^{34} & 0 \end{pmatrix}, \quad Q^{ab} = \begin{pmatrix} 0 & Q^{34} \\ Q^{34} & Q^{44} \end{pmatrix}, \quad Q^{44} \neq 0 ,$$

and the coordinate freedom is $x^{3'} = x^{3'}(x^3)$, $x^{4'}(x^4)$. From Eq. (4.1a), $\partial_3 Q^{44} = 0$, and so the coordinate transformation $x^{4'} = x^{4'}(x^4)$ will be used to set $Q^{44} = \pm 1$. The complete solution for the metric and CKT is

$$G = ((x^4\partial_3 e_3^{cd} + f_3^{cd})\partial_{(c}\otimes\partial_{d)} + 2Q^{44}\partial_{(3}\otimes\partial_{4)})/(b - x^4\partial_3 a) ,$$

$$Q = aG + e_3^{cd}\partial_{(c}\otimes\partial_{d)} + Q^{44}\partial_4\otimes\partial_4 ,$$

where $Q^{34} = ag^{34}$, $b = x^4\partial_3 a + (Q^{34}/g^{34})$, and e_3^{cd} and f_3^{cd} are functions of x^3, $Q^{44} = \pm 1$, and Q is a KT iff a and b are functions of x^3.

Case 2: The Q^{aa} components

If the determinant of Q^{aa} is zero, then $Q^{1a} = fQ^{2a}$ for some f, and either f is constant, in which case a linear transformation among the KV's will set Q^{1a} to zero (and so Q is redundant), or f is nonconstant. This is only possible when $Q^{aa}\partial_a$ are both null, but then Eq. (4.2a) implies that g^{ab} is zero, or f is constant. Consequently, we assume that $\det(Q^{aa}) \neq 0$.

Specialized solutions of Eq. (4.2) may be found, for example

$$G = \phi(a(x^4)^2 + 2cx^4 + d)\partial_1\otimes\partial_1 - 2(ax^3x^4 + bx^4 + cx^3 + f)\partial_{(1}\otimes\partial_{2)} + (a(x^3)^2 + 2bx^3 + e)\partial_{(2}\otimes\partial_{2)} + G_0^{ab}\partial_{(a}\otimes\partial_{b)} ,$$

$$Q^1 = -\partial_3\log\phi , \quad Q^2 = -\partial_4\log\phi ,$$

where G_0^{ab} is a constant matrix, a,b,c,d,e,f are constants, and ϕ is an arbitrary function of x^3 and x^4. However, the author was unable to solve Eq. (4.2a) completely, and so we shall not discuss the Q^{aa} components further.

We end this section by solving the null geodesic equations in spaces defined by Eq. (4.5). If the corresponding Q is a KT, the discussion applies to all geodesics. For this metric there are four independent first integrals,

$$L_1 = \delta_1^\alpha P_\alpha , \quad L_2 = \delta_2^\alpha P_\alpha , \quad L_3 = g^{\alpha\beta}P_\alpha P_\beta , \quad L_4 = Q^{\alpha\beta}P_\alpha P_\beta ,$$

where $P^\alpha = dx^\alpha/d\lambda$ is a null geodesic with affine parameter λ, and L_1,L_2,L_3,L_4 are constant along P^α. From Eqs. (4.5) and (4.6),

$$R_3P_3^2 = aL_3 - c_3L_1^2 - 2e_3L_1L_2 - g_3L_2^2 + L_4 ,$$

$$X_4P_4^2 = bL_3 - d_4L_1^2 - 2f_4L_1L_2 - h_4L_2^2 - L_4 ,$$

and so when Q is a CKT ($L_3 = 0$), or a KT [$a = a(3)$, $b = b(4)$], we see that P_3 and P_4 are functions of x^3 and x^4, respectively. This allows the null geodesic equations to be solved by quadrature,[9]

$$dx^3/R_3P_3 = dx^4/X_4P_4 ,$$

$$dx^1 = (L_1c_3 + L_2e_3)dx^3/R_3P_3 + (L_1d_4 + L_2f_4)dx^4/X_4P_4 ,$$

$$dx^2 = (L_1e_3 + L_2g_3)dx^3/R_3P_3 + (L_1f_4 + L_2h_4)dx^4/X_4P_4 ,$$

$$d\lambda = adx^3/R_3P_3 + bdx^4/X_4P_4 , \tag{4.8}$$

and we still have the coordinate freedom $x^{3'} = x^{3'}(x^3)$, $x^{4'} = x^{4'}(x^4)$, and a linear transformation among the Killing vectors, which may simplify Eqs. (4.8).

An example of such a space is the solution (belonging to Case 1b) of Plebanski and Demianski for the charged Kinnersley metric with a cosmological constant

$$a = 2p^2/(1-pq)^2 , \quad b = 2q^2/(1-pq)^2 ,$$

$$c_3 = -p^4/g(p) , \quad d_4 = q^4/h(q) ,$$

$$e_3 = -p^2/g(p) , \quad f_4 = -q^2/h(q) ,$$

$$g_3 = -1/g(p) , \quad h_4 = 1/h(q) ,$$

$$x^3 = p , \quad x^4 = q ,$$

where g and h are certain real quartic functions[10-13] of the real variables p and q, respectively. If the perfect square condition, $c_3g_3 = e_3^2$, $d_4h_4 = f_4^2$, holds for the Case 1b metric of Eq. (4.5), and $R_3,X_4,c_3,e_3,g_3,-d_4,-f_4,-h_4$ are positive on some coordinate patch, with g_3 and h_4 being nonzero, then Eq. (4.5) may be written as

$$ds^2 = (a+b)((m_3/p_3 + q_4)^2(dx^1 + q_4dx^2)^2 - (r_4/p_3 + q_4)^2(dx^1 - p_3dx^2)^2 + (dx^3/\sqrt{R_3})^2 + (dx^4/\sqrt{X_4})^2) , \tag{4.9}$$

where $m_3^2 g_3 = 1$, $n_4^2 h_4 = 1$, $p_3^2 g_3 = c_3$, $q_4^2 h_4 = d_4$. The two real null vectors

$$\begin{pmatrix} \omega^+ \\ \omega^- \end{pmatrix} = \frac{dx^4}{\sqrt{X_4}} \pm \left(\frac{n_4}{p_3 + q_4}\right)(dx^1 - p_3\, dx^2)$$

are geodesic and shear free, and so the empty space subclass of Eq. (4.9) are the Type D metrics. The corresponding CKT is that found by Walker and Penrose.

CONCLUSION

We have shown that two second order, trace-free symmetric tensors, two third order tensors (satisfying $L_{(\alpha\beta)\gamma} = 0 = L_{[\alpha\beta\gamma]}$), a bivector, and a tensor with the Riemann symmetries arise in a derivation of the structural equations for CKT's. An analysis of their first integrability conditions, which result from identities among the sixth order derivatives of $Q_{\alpha\beta}$, seemed impractical and so additional restrictions were imposed.

We assumed that the space was quasidiagonal and that the CKT was independent of the ignorable Killing coordinates. The complete solution for the metric was given when the CKT is also qd and the subclass containing the empty Type D spaces discussed.

A different approach was adopted recently by Hauser and Malhiot[14] who showed that the existence of a certain canonical form for the Killing tensor is sufficient to impose a two-dimensional, Abelian symmetry group on the space. A similar result was found by Sommers.[15]

ACKNOWLEDGMENT

The author gratefully acknowledges the encouragement and many suggestions offered by R. P. Kerr.

[1] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[2] P. Havas, J. Math. Phys. **16**, 2476 (1975).
[3] B. Carter, Phys. Rev. **174**, 1559 (1968).
[4] I. Hauser and R. J. Malhiot, P.O.R.S.I.I.T.-15, Illinois Inst. of Tech., Chicago, Illinois, 60616 (1974).
[5] I. Hauser and R. J. Malhiot, J. Math. Phys. **16**, 150 (1975).
[6] R. Geroch, J. Math. Phys. **11**, 1955 (1970).
[7] H. W. Guggenheimer, *Differential Geometry* (McGraw-Hill, New York, 1963), p. 347.
[8] L. P. Eisenhart, Ann. Math. **35**, 284 (1934).
[9] N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[10] J. Plebanski, in *Grav. Rad. and Grav. Collapse, First Copernican Symposium, Warsaw 1973,* Inst. Astro. Union Symposium No. 64 (Inst. Astro. Union, Warsaw, 1974), p. 188.
[11] G. J. Weir and R. P. Kerr, "Diverging type D metrics," to appear in Proc. R. Soc. London.
[12] G. J. Weir, Ph.D. thesis, Canterbury University, New Zealand.
[13] J. Plebanski and M. Demianski, Ann. Phys. **98**, 98 (1976).
[14] I. Hauser and R. J. Malhiot, J. Math. Phys. **17**, 1306 (1976).
[15] P. Sommers, J. Math. Phys. **14**, 787 (1973).

Maxwell's equations in an expanding universe*

S. Malin

Department of Physics and Astronomy, Colgate University, Hamilton, New York 13346
(Received 19 August 1976)

Schrödinger's classical solutions of Maxwell's equations in an expanding universe with positive spatial curvature are reformulated in terms of group theory. Euler angles are used as coordinates in spherical space; the equations satisfied by the components of the complex electromagnetic field tensor are then given in terms of Euler angles. It is shown that if the fundamental modes of the electromagnetic field are appropriately chosen, certain components of the tensor are given by matrix elements of the irreducible representations of the group SU(2).

I. INTRODUCTION

The problem of finding the fundamental modes of the electromagnetic field in an expanding universe with positive spatial curvature was first solved by Schrödinger,[1] using analytical methods. In the present work the problem is formulated and solved using group theoretical methods. The expression of the fundamental modes of the electromagnetic field in group theoretical terms is of intrinsic interest; we also find that some derivations are considerably simplified.

An additional motivation for the present work stems from considering the quantization of the electromagnetic field in curved space–time. In the case of flat space–time, an analysis of Maxwell's equations based on the group SU(2) led to an elegant method of gauge-free quantization.[2] The present work is also based on the group SU(2) and may lead to a similar quantization formalism for homogeneous, spatially isotropic curved space–times.

Section II contains the group theoretical framework of the present approach. It is based on the observation that spherical space of radius unity is also the space of the group SU(2); Euler angles can be used, therefore, as coordinates in the space. By using a well-known completeness theorem, all square-integrable quantities defined over the space can be uniquely expanded in terms of the matrix elements of the irreducible representations of the group SU(2).

Schrödinger's formulation of Maxwell's equations in an expanding universe with positive spatial curvature is briefly summarized in Sec. III, and some of the results are stated. He uses cylindrical coordinates in spherical space, which are related to Euler angles by simple linear transformations.

The group analysis is given in Sec. IV. The components of the complex electromagnetic field tensor and the equations they satisfy are transformed from cylindrical to Euler angles coordinates. A complete solution of these equations is derived in terms of the matrix elements $T^j_{pq}(u)$, $u \in \mathrm{SU}(2)$, of the irreducible representation of the group SU(2). It is shown that if the fundamental modes are appropriately chosen, some of the tensor components are simply given by $[S(t)]^{-1}T^j_{pq}(u)$ where t is the cosmic time and $S(t)$ is the Robertson–Walker expansion function.

II. SPHERICAL SPACE AND THE GROUP SU (2)

If the universe is assumed to be isotropic, spatially homogeneous, and contains a congruence of fundamental world lines which fills the whole of space–time, then it is possible to choose a canonical coordinate system (t, x_1, x_2, x_3) such that the metric tensor in the coordinate system is of the form[3]

$$ds^2 = dt^2 - S^2(t)\frac{dx_1^2 + dx_2^2 + dx_3^2}{(1 + \frac{1}{4}kr^2)^2}, \tag{2.1}$$

where

$$r = (x_1^2 + x_2^2 + x_3^2)^{1/2}, \tag{2.2}$$

and on the fundamental world lines $ds^2 = dt^2$. The coordinate t in the canonical frame of reference is called "cosmic time."

The isotropic, spatially homogeneous cosmological models are classified into the following three well-known types:

(i) $k = +1$ (spherical space). The hypersurfaces t = const have constant positive curvature.

(ii) $k = 0$ (Euclidean space). The hypersurfaces k = const have zero curvature.

(iii) $k = -1$ (pseudospherical space). The hypersurfaces t = const have constant negative curvature.

The present paper deals with the case of spherical space. Following Schrödinger's notation,[1] let $x^1 = \omega$, $x^2 = \phi$, $x^3 = \psi$ be cylindrical coordinates in spherical space, having the range $0 < \phi, \psi < 2\pi$ and $0 \leq \omega < \pi$.

The possibility of applying group theoretical methods to the analysis of equations in spherical space arises through the following observation: Spherical space with curvature unity is also the space of the group SU(2), and, if the following transformation of coordinates is employed,

$$\bar{x}^1 = \beta = \tfrac{1}{2}\omega, \quad \bar{x}^2 = \alpha = \tfrac{1}{2}(\phi + \psi), \quad \bar{x}^3 = \gamma = \tfrac{1}{2}(\phi - \psi), \tag{2.3}$$

then α, β, γ are the Euler angles, which are the most commonly used parameters of the group SU(2).[4]

Since the matrix elements $T^j_{pq}(u)$ of all the irreducible representations of the group SU(2) form a complete

orthogonal set over the group, any function $F(u)$, $u \in SU(2)$, which satisfies

$$\int |F(u)|^2 du < \infty, \tag{2.4}$$

where $du = (1/16)\pi^{-2} \sin\beta \, d\alpha \, d\beta \, d\gamma$ is the invariant measure over SU(2), normalized so that $\int du = 1$, can be uniquely expanded in the $T^j_{pq}(u)$:

$$F(u) = \sum_j \sum_{p=-j}^{j} \sum_{q=-j}^{j} a^j_{pq} T^j_{pq}(u). \tag{2.5}$$

The coefficients are given by

$$a^j_{pq} = (2j+1) \int f(u) T^{j*}_{pq}(u) \, du, \tag{2.6}$$

where T^{j*}_{pq} is the complex conjugate of T^j_{pq}. Equation (2.6) follows from the orthogonality relations of the $T^j_{pq}(u)$:

$$\int T^j_{pq}(u) T^{j'*}_{p'q'}(u) \, du = (2j+1)^{-1} \delta_{jj'} \delta_{pp'} \delta_{qq'}. \tag{2.7}$$

The $T^j_{pq}(u)$ satisfy the equation[5]

$$\left[\frac{\partial^2}{\partial\beta^2} + \cot\beta \frac{\partial}{\partial\beta} + \frac{1}{\sin^2\beta}\left(\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\gamma^2} - 2\cos\beta \frac{\partial^2}{\partial\alpha \, \partial\gamma}\right) + j(j+1)\right] T^j_{pq}(\alpha, \beta, \gamma) = 0. \tag{2.8}$$

They can be expressed in the form

$$T^j_{pq}(\alpha, \beta, \gamma) = \exp(ip\alpha) \, d^j_{pq}(\beta) \exp(iq\gamma), \tag{2.9}$$

where $d^j_{pq}(\beta) \equiv T^j_{pq}(0, \beta, 0)$ satisfies the equation

$$\left[\frac{d^2}{d\beta^2} + \cot\beta \frac{d}{d\beta} - \frac{p^2 + q^2 - 2pq\cos\beta}{\delta u^2 \beta} + j(j+1)\right] \times d^j_{pq}(\beta) = 0. \tag{2.10}$$

III. MAXWELL'S EQUATIONS IN SPHERICAL SPACE-TIME

Let the electromagnetic field tensor be defined in locally Minkowskian coordinates by

$$\phi^{0i} = E_i, \quad \phi^{ij} = \epsilon^{ijk} B_k, \quad \phi^{\alpha\beta} = -\phi^{\beta\alpha}, \tag{3.1}$$

where E_i, B_k, are the usual electric and magnetic field components: Latin indices take the values 1, 2, 3, Greek indices take the values 0, 1, 2, 3, and ϵ^{ijk} is the totally symmetric Levi-Civita symbol with $\epsilon^{123} = 1$. As usual, let the covariant components of $\phi^{\alpha\beta}$ be given by $\phi_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} \phi^{\alpha\beta}$, and let $\hat{\phi}_{\alpha\beta} = \frac{1}{2}\sqrt{-g}\epsilon_{\alpha\beta\mu\nu} \phi^{\mu\nu}$ be the dual tensor of $\phi_{\alpha\beta}$. Let

$$F_{\alpha\beta} \equiv \phi_{\alpha\beta} + i\hat{\phi}_{\alpha\beta}, \tag{3.2}$$

be the complex electromagnetic field tensor.[6]

Maxwell's equations in curved space–time can now be expressed in the form

$$\frac{\partial F_{\alpha\beta}}{\partial x^\gamma} + \frac{\partial F_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial F_{\gamma\alpha}}{\partial x^\beta} = 0, \tag{3.3}$$

and the following results were obtained by Schrödinger[1]:

(i) F_{23}, F_{31}, and F_{12} can be expressed in terms of F_{12}, F_{13}, F_{14} as follows:

$$F_{23} = iS \sin\omega \cos\omega F_{14}, \quad F_{31} = iS \cot\omega F_{24}, \quad F_{12} = iS \tan\omega F_{31}. \tag{3.4}$$

$S = S(t)$ is the expansion function (see Sec. II).

(ii) In terms of a new variable $\tau \equiv \int^t dt/S(t)$ the F_{k4} can be written in the form

$$F_{k4} = S^{-1} \exp[i(\nu\tau + n\phi + m\psi)] f_{k4}(\omega), \tag{3.5}$$

where ν, n, m are integers and $\nu \geq 2$. A fairly complex set of relationships between the allowed values of ν, n, m was derived. These will not be reproduced here; the group approach yields equivalent relationships rather simply (see Sec. IV).

(iii) f_{14} can be expressed in terms of f_{24}, f_{34} by the equation

$$\nu \sin\omega \cos\omega f_{14} = imf_{34} - imf_{24}, \tag{3.6}$$

and f_{24}, f_{34} satisfy the equations

$$\begin{aligned} &\nu \sin\omega \cos\omega \frac{df_{34}}{d\omega} + nmf_{34} - m^2 f_{24} + \nu^2 \cos^2\omega f_{24} = 0, \\ &\nu \sin\omega \cos\omega \frac{df_{24}}{d\omega} - mnf_{24} + n^2 f_{34} - \nu^2 \sin^2\omega f_{34} = 0. \end{aligned} \tag{3.7}$$

If f, g are now defined as

$$f = f_{34} + f_{24}, \quad g = f_{34} - f_{24}, \tag{3.8}$$

then the following equations are obtained:

$$\begin{aligned} &2\nu \sin\omega \cos\omega \frac{df}{d\omega} + (n^2 - m^2 + \nu^2 \cos^2\omega - \nu^2 \sin^2\omega) f \\ &\quad + (2mn + m^2 - n^2 - \nu^2) g = 0, \\ &2\nu \sin\omega \cos\omega \frac{dg}{dt} + (2mn - m^2 - n^2 - \nu^2) f \\ &\quad + (m^2 + n^2 - \nu^2 \cos^2\omega + \nu^2 \sin^2\omega) g = 0. \end{aligned} \tag{3.9}$$

IV. GROUP ANALYSIS

Transformation of the tensor $F_{\alpha\beta}$ from cylindrical coordinates x^μ to Euler angle coordinates $\bar{x}^\mu$ yields

$$\begin{aligned} &\bar{F}_{01} = 2F_{01}, \quad \bar{F}_{02} = F_{02} + F_{03}, \quad \bar{F}_{03} = F_{02} - F_{03}, \\ &\bar{F}_{12} = 2(F_{12} + F_{13}), \quad \bar{F}_{23} = -2F_{23}, \quad \bar{F}_{31} = -2(F_{12} + F_{31}). \end{aligned} \tag{4.1}$$

Because of Eqs. (3.5) and (3.8) we obtain the following form for $\bar{F}_{02}$, $\bar{F}_{03}$:

$$\begin{aligned} &\bar{F}_{02} = S^{-1} \exp[i(\nu s + p\alpha + q\gamma)] f(\beta), \\ &\bar{F}_{03} = S^{-1} \exp[i(\nu\tau + p\alpha + q\gamma)] g(\beta), \end{aligned} \tag{4.2}$$

where

$$p = \tfrac{1}{2}(m+n), \quad q = \tfrac{1}{2}(m-n). \tag{4.3}$$

The set of equations satisfied by f and g is obtained by applying the transformation (2.3) to Eqs. (3.9). The calculation yields

$$\begin{aligned} &2\nu \sin\beta \frac{df}{d\beta} + (\nu^2 \cos\beta - 4pq) f + (4p^2 - \nu^2) g = 0, \\ &2\nu \sin\beta \frac{dg}{d\beta} + (4pq - \nu^2 \cos\beta) g - (4q^2 - \nu^2) f = 0. \end{aligned} \tag{4.4}$$

By eliminating $f(\beta)$ from Eqs. (4.4), the following

equation is obtained for $g(\beta)$:

$$\frac{d^2g}{d\beta^2}+\cot\beta\frac{dg}{d\beta}+(\tfrac{1}{4}\nu^2+\tfrac{1}{2}\nu)g - \frac{p^2+q^2-2pq\cos\beta}{\sin^2\beta}g=0, \tag{4.5}$$

and $f(\beta)$ is given in terms of $g(\beta)$ by the equation

$$f(\beta)=\frac{2\nu\sin\beta\, dg/d\beta+(4pq-\nu^2\cos\beta)g}{4q^2-\nu^2}. \tag{4.6}$$

Alternatively, by eliminating $g(\beta)$, one obtains

$$\frac{d^2f}{d\beta^2}+\cot\beta\frac{df}{d\beta}+(\tfrac{1}{4}\nu^2-\tfrac{1}{2}\nu)f - \frac{p^2+q^2-2pq\cos\beta}{\sin^2\beta}f=0, \tag{4.7}$$

and $g(\beta)$ is given in terms of $f(\beta)$ by

$$g(\beta)=\frac{2\nu\sin\beta\, df/d\beta+(\nu\cos\beta-4pq)f}{\nu^2-4p^2}. \tag{4.8}$$

If the functions F_+ and F_- are now defined as

$$\begin{aligned}F_+(\alpha,\beta,\gamma)&\equiv\exp[i(p\alpha+q\gamma)]f(\beta),\\ F_-(\alpha,\beta,\gamma)&\equiv\exp[i(p\alpha+q\gamma)]g(\beta),\end{aligned} \tag{4.9}$$

it follows from Eqs. (4.5) and (4.7) that they satisfy

$$\left[\frac{\partial^2}{\partial\beta^2}+\cot\beta\frac{\partial}{\partial\beta}+(\tfrac{1}{4}\nu^2\mp\tfrac{1}{2}\nu)+\frac{1}{\sin^2\beta}\times\left(\frac{\partial^2}{\partial\alpha^2}+\frac{\partial^2}{\partial\gamma^2}-2\cos\beta\frac{\partial^2}{2\alpha\,\partial\gamma}\right)\right]F_\pm(\alpha,\beta,\gamma)=0. \tag{4.10}$$

Since the functions $F_+(\alpha,\beta,\gamma)$ and $F_-(\alpha,\beta,\gamma)$ are defined over the space of the group SU(2), they can be uniquely expanded according to Eq. (2.5). If we now substitute these expansions in Eqs. (4.10) and denote

$$j_+=\begin{cases}\tfrac{1}{2}\nu-1 & \text{if } \nu\geq 0,\\ -\tfrac{1}{2}\nu & \text{if } \nu<0,\end{cases} \tag{4.11}$$

and

$$j_-=\begin{cases}\tfrac{1}{2}\nu & \text{if } \nu\geq 0,\\ -\tfrac{1}{2}\nu-1 & \text{if } \nu<0,\end{cases} \tag{4.12}$$

then in all cases

$$\tfrac{1}{2}\nu^2\mp\tfrac{1}{2}\nu=j_\pm(j_\pm+1),\quad j_+,\ j_-\geq 0, \tag{4.13}$$

and the following conclusions follow from Eq. (2.10) and (4.5)—(4.8):

(i) Solutions of Eqs. (4.10) exist only for j_+, j_-, p, q integer or half-integer; ν is, therefore, a (positive or negative) integer. Since ν is real, it follows from Eq. (4.12) that $|\nu|\geq 2$. If ν is odd, j_+, j_-, p, q are all half-integers; if ν is even, they are all integers.

(ii) Let ν be positive, and consider Eqs. (4.5), (4.6). Because of Eq. (2.6), $|q|=\frac{1}{2}\nu=j_-$ is excluded; there are ν^2-1 solutions[7] of the form[8]

$$F_-(\alpha,\beta,\gamma)=a^{j_-}_{pq}T^{j_-}_{pq}(\alpha,\beta,\gamma), \tag{4.14}$$

where $a^{j_-}_{pq}$ are arbitrary constants and the allowed values of p,q are $p=-2j_--1,\ -2j_-,\ldots,2j_-+1$, $q=-2j_-,\ldots,2j_-$.

(iii) Let ν be negative and apply similar considerations to Eqs. (4.7), (4.8). One obtains again ν^2-1 solutions; they are

$$F_+(\alpha,\beta,\gamma)=a^{j_+}_{pq}T^{j_+}_{pq}(\alpha,\beta,\gamma), \tag{4.15}$$

with $j_+=-\frac{1}{2}\nu$, $a^{j_+}_{pq}$ arbitrary constants and p, q having the allowed values $p=-2j_+,\ldots,2j_+$, $q=-2j_+-1$, $-2j_+,\ldots,2j_++1$.

It follows now from Eqs. (3.5), (4.14), and (4.15) that all the fundamental modes of the electromagnetic field in spherical space can be described as follows: Corresponding to every positive integer ν satisfying $\nu\geq 2$ there are ν^2-1 modes given by

$$F_{34}(\alpha,\beta,\gamma,t)=[S(t)]^{-1}T^{\nu/2}_{pq}(\alpha,\beta,\gamma), \tag{4.16}$$

with the corresponding $F_{24}(\alpha,\beta,\gamma,t)$ determined by Eqs. (4.6) and (3.5); and corresponding to every negative integer satisfying $\nu\leq -2$ there are ν^2-1 modes given by

$$F_{24}(\alpha,\beta,\gamma,t)=[S(t)]^{-1}T^{-\nu/2}_{pq}(\alpha,\beta,\gamma), \tag{4.17}$$

with the corresponding $F_{34}(\alpha,\beta,\gamma,t)$ given by Eqs. (4.8) and (3.5). For all allowed values of ν, once F_{24} and F_{34} are given, the other components of the tensor $F_{\alpha\beta}$ are determined by Eqs. (3.4)—(3.6).

It follows, therefore, that for any given electromagnetic field in spherical space-time the expansion of F_{34} (for $\nu\geq 2$) or F_{24} (for $\nu\leq -2$) in terms of the matrix elements of the irreducible representations of the group SU(2) can be physically interpreted as an expansion in terms of the fundamental modes of the field.

ACKNOWLEDGMENTS

This work was completed at the Center for Radiophysics and Space Research, Cornell University, in summer 1976. It is a pleasure to thank Professor Kenneth Greisen and his colleagues at the Center, and in particular Professor Victor Mansfield, for their warm hospitality.

*This work was supported in part by the Colgate Research Council.

[1]E. Schrödinger, Proc. Roy. Irish Soc. 46, 25 (1940).
[2]M. Carmeli, Nuovo Cimento B 67, Ser. X 103 (1970).
[3]H. P. Robertson, Astrophys. J. 82, 284 (1935). For a recent observational analysis of isotropy, see A. Sandage, Astrophys. J. 202, 563 (1975).
[4]The Euler angles are, however, not the only parameters used; see H. E. Moses, Ann. Phys. (N.Y.) 37, 224 (1966).
[5]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N.J., 1957).
[6]This is analogous to using the complex vector field $\mathbf{V}=\mathbf{E}+i\mathbf{B}$ in flat space—time. The use of V is essential for group analysis of Maxwell equations [M. Carmeli, J. Math. Phys. 10, 569 (1969)].
[7]The multiplicity ν^2-1 was derived by Schrödinger[1] using analytical method.
[8]When Eqs. (4.6), (4.9) are used to obtain $F_+(\alpha,\beta,\gamma)$ in terms of $F_-(\alpha,\beta,\gamma)$, the F_+ which correspond to $p=\pm j_-$ turn out to be zero; not all solutions of Maxwell's equations are obtained, therefore, by using Eq. (4.7) instead of (4.5) as the fundamental one. The opposite happens for $\nu<0$. Equation (4.7) has to be used as fundamental in order to obtain all solutions.

Fibre-bundle structure of thermodynamic states

Awele Maduemezia

Department of Physics, University of Ibadan, Ibadan, Nigeria
(Received 4 February 1977)

It is shown how the requirement that the Gibbs' ensemble average $\langle A\rangle = \mathrm{Tr}(e^{2Hb}A)/\mathrm{Tr}A$, $b=-(1/2)kT$, of any physical quantity A be formally expressible as an expectation value $(Pb|A|bP)/(Pb|bP)$ over a thermodynamic state $|Pb)$, naturally leads to the realization of $(Pb|$, as a cross section of a fibre bundle $\theta(G)$ with fibre G, over a manifold M of pressure states P as *base* space, where G is the infinite-dimensional Lie group $\{\exp(Hb+u_nP_n);\ -\infty<b,\ u_n<\infty\}$ generated by H (Hamiltonian) and $\{P_n;\ n=1,2,\cdots\}$, the sequence of projectors on the eigenvector subspaces of H. The group G is thus partly parametrized by the temperature variable b.

1. INTRODUCTION

The recent attempt of Takahashi and Umezawa[1] to formulate quantum statistical mechanics at finite temperatures in the form of a quantum field theory with a temperature-dependent "vacuum" raises some issues regarding the true mathematical structure of the object $|P,T)$, which can be said to represent the state of system in thermodynamic equilibrium.

In the paper of Takahashi and Umezawa, a so-called temperature-dependent vacuum is constructed in the form

$$|\beta\rangle = Z^{-1/2}(\beta)\sum_n \exp(-\beta E_n/2)\,|n,\tilde{n}\rangle,$$

where Z is the partition function, $\beta=1/kT$, and $|n,\tilde{n}\rangle$ is the direct product of two kets $|n\rangle$, $|\tilde{n}\rangle$, which themselves are defined by the two eigenvalue equations

$$H|n\rangle = E_n|n\rangle,$$
$$\tilde{H}\,|\tilde{n}\rangle = E_n|\tilde{n}\rangle,$$

the former equation referring to the real physical system, and the latter to an analogous fictitious system whose Hamiltonian $\tilde{H}$ is supposed to have the same spectrum as the actual Hamiltonian, H. With the aid of this construction, the authors then proceeded to construct Fock operators and other related quantities.

The above construction for $|\beta\rangle$ arose from the requirement that the statistical average of a quantity A, over the canonical ensemble at temperature T, viz.,

$$\langle A\rangle = \mathrm{tr}(Ae^{-\beta H})/Z,$$
$$\beta = 1/kT,$$
$$Z = \mathrm{tr}(e^{-\beta H}),$$

be expressible as an expectation value

$$\langle A\rangle = \langle\beta|A|\beta\rangle/\langle\beta|\beta\rangle,$$

where $|\beta\rangle$ is a suitable temperature-dependent state. The main object of this article is to demonstrate that, keeping the last-mentioned prescription in mind, one can elucidate the precise mathematical nature of the (P,T) states of a system in equilibrium. We shall henceforth call such states *thermodynamic states*, and denote them by $(P,b|$, where P is pressure, T is temperature, and $b=-1/2kT$.

The laws of thermodynamics provide a direct mathematical relationship between the finely structured (i.e. detailed) microscopic states on the one hand, and the much coarser space of thermodynamic states $(P,b|$, on the other hand. The relationship is provided by the above-mentioned averaging process, which we shall show is tantamount to the canonical projection of the manifold of microscopic states, viewed as the *total* space of a *fibre-bundle*, on the manifold M of pressure states (P) taken as the *base* space.

2. THERMODYNAMIC STATES

A nonmagnetic physical system in thermodynamic equilibrium is describable in terms of the pairs of variables: (P,T) (T,V), (V,S), or (S,P), with corresponding potential functions G (Gibbs' potential), F (free energy), U (internal energy), and H (enthalpy), respectively, where P is pressure, T, Kelvin temperature, V, volume and S, entropy. The pair that is chosen determines the potential that is minimized at equilibrium. We shall in this article choose the pair (P,T). Furthermore, we shall allow for the fact that (metastable) quantum electronic systems can be shown to exhibit (formal) thermodynamic properties with temperature values in the negative range[2] of the temperature scale. Thus we shall use as the temperature parameter

$$b=-1/2kT,$$

where k is the Boltzmann constant, and $-\infty<b<\infty$. Suppose the physical system C is described by a Hamiltonian $H(C)$. Let $H(C)$ have a spectrum $E_n(C)$ defined by the eigenvalue equation

$$H(C)|n\rangle = E_n(C)|n\rangle,$$

where the eigenvectors $|n\rangle$ are complete and orthonormal. The significance of the *system parameter* C written explicitly in $E_n(C)$ is that since b can be positive, the partition function

$$Z=\mathrm{tr}e^{2Hb}$$

may be infinite unless we ensure that the upper bound of $E_n(C)$ is finite for systems C corresponding to positive values of b. Thus the system parameter C will have essentially two values, which we denote as + and −. C is + if and only if $b\geq 0$ (e.g., for metastable quantum electronic systems) and C is − if $b<0$, i.e., for normal physical systems. The set $\{E_n(-)\}$ need not have an upper bound, but $\{E_n(+)\}$ must have an upper bound. With this proviso on the boundedness of the spectrum

of H, we shall assume that b takes arbitrary values on the entire real line. We shall also drop explicit mention of the system parameter C.

3. BUNDLE OF LINEAR FRAMES

Let M denote the positive real line defined by the pressure states $P \geq 0$. M is a differentiable manifold.[3] At each point of M, i.e., for each pressure value, P, let $L(a,P)$ denote the ordered set $\{\langle n,p|a_n\colon n=1,2,\ldots\}$ of eigenvectors of H (acting from the right),

$$\langle n,P|a_nH=\langle n,P|a_nE_n,$$

where $\{a_n\}$ is any sequence of real numbers for which the series $\sum_{n=1}^{\infty} a_n^2$ converges. $L(a,P)$ will be called a *linear frame* at P, and will be represented as a *row* matrix with nth entry $\langle n,P|a_n$. The frame $L(a,P)$ spans, i.e., is basis for, the *Hilbert space* at P, which we denote by E_p. Now let $L(M)$ be the set of all such linear frames at all points P of M, i.e., $L(M)$ is a product space consisting of pairs $(P,L(a,P))$. Let $\mathrm{GL}(E_p)$ denote the group of automorphisms of E_p. $\mathrm{GL}(E_p)$ is the infinite general linear group with real entries. Let Π be the mapping of $L(M)$ onto M which maps a linear frame at P into P,

$$\Pi : L(a,P) \to P.$$

The group $\mathrm{GL}(E_p)$ acts on $L(M)$ on the right as follows: If $L(a,P)$ is a linear frame at P, and $A \in \mathrm{GL}(E_p)$, then $L(a',P)=L(a,P)A$ is also a linear frame at P, where $L(a',P)$ is the row matrix with nth entry $\sum_m \langle m|a_mA_{mn}$. This action of $\mathrm{GL}(E_p)$ on $L(M)$ is *effective* in that only the identity element of $\mathrm{GL}(E_p)$ leaves any point of $L(M)$ fixed. Thus the triplet $\xi=(L(M),\Pi,M)$ is a *principal* $\mathrm{GL}(E_p)$-*bundle* which we may call the bundle of frames over M.[4]

We shall now consider the subgroup K of $\mathrm{GL}(E_p)$ consisting of diagonal matrices, k, with real entries k_n, $n=1,2,\cdots$. In general, k would be semi-infinite in dimension and we must impose the condition that the series $\sum_{n=1}^{\infty} k_n^2$ converges. The group K acts effectively on $L(M)$ from the right, $K : L(M) \to L(M)$. It thus defines the subbundle $\phi=(L(M),\Pi,K)$ as a reduced bundle. [We note that it is not required that K be a closed subgroup of $\mathrm{GL}(E_p)$, which it is not, since the diagonal entries are not equal.] ϕ will be called a *principal* K-*bundle* and denoted by $\phi(K)$. The space M is isomorphic with the quotient space of $L(M)$ by the equivalence relation induced by $K : M \approx L(M)/K$.

The elements of M are in one-to-one correspondence with the sets of equivalent row matrices $L(a,P)$ with nth entries $\langle n,P|a_n$, two such matrices $L(a,P)$, $L(a',P)$ being equivalent if they are connected to each other by a right action of some $k \in K$. Since all the row matrices for a given P are so related, we see that the mapping Π defined earlier coincides with the canonical map

$$\Pi : L(M) \to L(M)/K.$$

This map associates each K-equivalence class in $L(M)$ with a point P of M. From this right action of K on $L(M)$, we may also obtain a representation $k \to \exp(k_nP_n)$ of K as a Lie group of transformations of $L(M)$ at P, where k_n is real and P_n is the *projection operator* onto the eigenvector subspace corresponding tothe eigenvalue E_n of the Hamiltonian H. This representation of K will be useful later.

For each value of pressure $P \in M$, the space $\Pi^{-1}(P)$ is the *fibre* over P. Each point y of $\Pi^{-1}(P)$ is called a *cross section* and is representable by the *graph* $(P,g(y))$, where $g(y)$ is the element of K with the aid of which y is reached by right translation from some fixed row vector $L(a_0,P)$.

4. THE BUNDLE OF THERMODYNAMIC STATES

Now let us consider again the eigenvalue equation

$$H|n\rangle = |n\rangle E_n,$$

where

$$\langle m|n\rangle = \delta_{mn}.$$

We also put

$$H=\sum_n a_nP_n,$$

where a_n are real numbers subject only to the condition that $\sum_{n=1}^{\infty} a_n^2$ converges, and P_n is, as before, the projection operator onto the eigenvector subspace corresponding to the eigenvalue E_n of H. It follows that

$$[H,P_n]=0.$$

Also since $P_nP_m=0$ for $m \neq n$, we have

$$[P_n,P_m]=0$$

for all m, n.

Furthermore H and $\{P_n\}$ are generators of an infinite-dimensional Lie group G, whose elements g are given by

$$g=e^X;$$

$$X=bH+u_nP_n,$$

where $b=-1/2kT$, and u_n are real numbers subject to the Hilbert space convergence condition imposed earlier. *The Einstein summation convention over repeated indices is also assumed.*

The group K acts on G from the left according to

$$kg=e^W \in G,$$

$$W=bH+(u_n+k_n)P_n,$$

where $g \in G$, and $k=\exp(k_nP_n) \in K$. We have here used the representation

$$k \to \exp(k_nP_n)$$

of K as a Lie group of transformations on the space of $\{L(a,P)\}$. Now, associated with the principal K-bundle ϕ is a fibre bundle $\theta(G)$ constructed as follows: We consider the *left action* of K on the Lie group G as given above. For every $k \in K$, $x \in L(M)$, $g \in G$, the relation

$$k : (x,g) \equiv (x,g)k \to (xk,k^{-1}g)$$

defines a (right) K-structure on $L(M)\times G$.

Furthermore, the map

$$f : (L(M)\times G) \to L(M)G \equiv \theta,$$

under which the action of K on $(L(M)\times G)$ goes to the quotient, defines the quotient space

$$\theta = f((L(M)\times G)K)$$

which is isomorphic with $(L(M)\times G)$ mod$K$.

Let $\Pi_\theta : \theta \to M$ be the factorization of the composition

$$L(M)\times G \xrightarrow{\pi_p} L(M) \xrightarrow{\pi} M$$

by the projection $f: (L(M)\times G) \to \theta$, so that the following diagram is commutative:

$$\begin{array}{ccc} L(M) & \xleftarrow{\pi_p} & L(M)\times G \\ \Big\downarrow \Pi & & \Big\downarrow f \\ M & \xleftarrow{\pi_\theta} & \theta \end{array}$$

$\theta = \theta(G)$ is the *fibre bundle* over M with *fibre* G (viewed as a K-module) and *associated* principal K bundle ϕ. The group K is the structure group of the fibre bundle θ. For every point $P \in M$, the fibre $\Pi_\theta^{-1}(P)$ is a space homeomorphic with G. It has as points the objects $(P,b,k) = L(1,P)e^X$; $X = bH + k_n P_n$, where $L(1,P)$ is the row matrix with nth entry $\langle P,n|$, so that (P,b,k) is a row matrix with nth entry $\langle P,n|\exp(bE_n + k_n)$, k_n being arbitrary real numbers subject only to the already stated convergence conditions. The objects (P,b,k) are the *cross sections* of the fibre. They are parametrized by real numbers $\{b,k_n,n=1,2,\cdots\}$. The *family of sections* for which $k_n = 0$ for all n are denoted by $(P,b|$, and are precisely the thermodynamic states needed to express the average of any physical quantity in the form

$$\langle A\rangle = (P,b|A|b,P)/(P,b|b,P).$$

5. CONCLUSION

We have shown that the requirement that the Gibbs expression for the statistical average of a physical quantity in the canonical ensemble be written as an expectation value between thermodynamic states $|b,P)$ naturally leads to $(P,b|$ being realized as a cross section of a fibre bundle θ, with *standard fibre* G, where G is the infinite-dimensional Lie group generated by the Hamiltonian H and the infinite sequence of projection operators P_n onto the eigenvector subspace corresponding to the eigenvalue E_n of H. The *base* space M of this bundle is the positive half-line $P \geq 0$, where P is the pressure of the system. If Π_θ denotes the mapping from θ onto M, and if $P \in M$, then the fibre $\Pi_\theta^{-1}(P)$ of which $(P,b|$ is a cross section, is homeomorphic to the group G, which does not have a vector space structure. The space of the thermodynamic states $(P,b|$ is therefore not a vector space. This is in agreement with the superselection rule according to which no temperature state can be obtained as a linear superposition of two or more temperature states.

Referring to the paper of Takahashi and Umezawa,[1] which motivated this work, one can see that their fictitious operator $\tilde{H}$ is very much akin to our direct sum $\tilde{H} = \sum_n a_n P_n$ of projection operators, and these latter are by no means fictitious. Having obtained the states as sections of a fibre bundle, it is a straightforward matter to realize the relationship between one (P,T) state and another in terms of the usual morphisms of the fibre bundle.[5]

ACKNOWLEDGMENTS

Part of this work was done while the author was a Senior Associate Member of the International Centre for Theoretical Physics, Trieste, Italy. The author wishes to thank the Director of the Centre, Professor Abdus Salam for hospitability at the Centre. The author also wishes to acknowledge useful discussions he has had with Professor M. A. Rashid and Dr. Akin Ojo, both of the University of Ibadan, Department of Physics.

[1] Y. Takahashi and H. Umezawa, Collect. Phenom. 2, 55 (1975).
[2] G. Weinreich, *Fundamental Thermodynamics* (Addison-Wesley, Reading, Massachusetts, 1968).
[3] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[4] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1.
[5] D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).

On the blowing up of solutions to the Cauchy problem for nonlinear Schrödinger equations*

R. T. Glassey

Department of Mathematics, Indiana University, Bloomington, Indiana 47401
(Received 17 December 1976; revised manuscript received 21 March 1977)

Solutions to the Cauchy problem for the equation $iu_t = \Delta u + F(|u|^2)u$ ($x \in \mathbb{R}^n$, $t > 0$), $u(x,0) = \varphi(x)$, are considered. Conditions on φ and F are given so that, for solutions with nonpositive energy, the following obtains: There exists a *finite* time T, estimable from above, such that $\|\text{grad}u(t)\|_{L^2(\mathbb{R}^n)} \to +\infty$ as $t \to T^-$. It is also shown that other L_q-norms of a solution (including $q = \infty$) blow up in finite time.

INTRODUCTION

Consider the Cauchy problem for the equation

$$iu_t = \Delta u + g|u|^{p-1}u, \quad x \in \mathbb{R}^n, \quad t > 0$$
$$u(x,0) = \varphi(x). \tag{1}$$

Here $g > 0$, $p > 1$, and φ is assumed to be smooth and small at infinity. We investigate conditions under which solutions to (1) (and to equations with more general nonlinearities) may blow up in finite time. To motivate the result, we consider the problem of finding an *a priori* bound on the $H_1(\mathbb{R}^n)$ norm of a solution. As is well known, a sufficiently regular solution satisfies

$$\|u(t)\|_2 \equiv \left(\int_{\mathbb{R}^n} |u(x,t)|^2 dx\right)^{1/2} = \text{const} = \|\varphi\|_2 \tag{2}$$

and

$$\int_{\mathbb{R}^n} \left(|\nabla u|^2 - \frac{2g}{p+1}|u|^{p+1}\right) dx = \text{const} \equiv E_0. \tag{3}$$

We estimate the norm $\|u(t)\|_{p+1}$ appearing here by the Sobolev inequality[1] and obtain (assuming $p+1 < 2n/(n-2)$ if $n \geq 3$, p arbitrary for $n = 1, 2$)

$$\|u(t)\|_{p+1} \leq \text{const}\|u(t)\|_2^{1-\theta}\|\nabla u(t)\|_2^{\theta},$$

where

$$\frac{1}{p+1} = \theta\left(\frac{1}{2} - \frac{1}{n}\right) + \frac{1-\theta}{2},$$

i.e., for $\theta = n(p-1)/2(p+1)$. Since $\|u(t)\|_2$ is uniformly bounded by (2), this gives the estimate

$$\|u(t)\|_{p+1}^{p+1} \leq \text{const}\|\nabla u(t)\|_2^{n(p-1)/2}.$$

Then, assuming $|E_0| < \infty$, we find from (3) the inequality

$$\|\nabla u(t)\|_2^2 \leq |E_0| + g\,\text{const}\|\nabla u(t)\|_2^{n(p-1)/2}.$$

Clearly an uniform bound on $\|\nabla u(t)\|_2$ results, provided $(n/2)(p-1) < 2$; that is, provided $p < 1 + 4/n$. For larger values of p, such a bound could still be obtained under a "smallness condition" (e.g., if g were sufficiently small); however, the global existence of "large" solutions would then be in doubt. Indeed, our blow-up theorem, as applied to this case, gives a class of initial values φ for which the norm $\|\nabla u(t)\|_2$ of a solution u to (1) blows up in finite time, provided $E_0 \leq 0$ and $p > 1 + 4/n$.

Strauss[2,3] has shown that "small" solutions to (1) (regardless of the sign of g) exist globally and decay as $t \to \infty$, provided p is sufficiently large. When $n = 3$ and $g < 0$, Lin[4] has determined the asymptotic behavior of "large" solutions to (1) for $3 \leq p < 5$. When $n = 1$ and $g > 0$, (1) possesses solitary wave solutions.[2,3,5] In fact, a special consequence of the procedure in Ref. 5 is global existence for (1) in the case $n = 1$, $p = 3$.

In the past few years a dramatic increase of interest in such blow-up results has occurred. However, none of the presently known methods seems to apply to Schrödinger-type equations. We include in the bibliography a representative (but not complete) list of known results (cf. Refs. 6–21). Of course for solutions to quasilinear hyperbolic equations, similar singular behavior (the pointwise blowing up of derivatives of solutions) is well known[19,21] in one space dimension.

1. THE BLOW-UP THEOREM

The Cauchy problem to be considered has the form

$$iu_t = \Delta u + F(|u|^2)u, \quad x \in \mathbb{R}^n, \quad t > 0, \quad u(x,0) = \varphi(x). \tag{4}$$

For simplicity, we shall assume that $\varphi \in \mathcal{S}$. In addition, we assume that the real-valued function F is smooth enough so that a unique, local classical solution to (4) exists, which lies in a Sobolev space of sufficiently high order. This can be done using the existence theorem of Segal.[22]

We begin with a lemma giving several identities which will be used in the proof. We also define

$$G(u) = \int_0^u F(s)ds.$$

Lemma: Let u be a solution of (4) (as above) on an interval $0 \leq t < t_1$. Then

(i) $\|u(t)\|_2 = \|\varphi\|_2$,

(ii) $\int [|\nabla u|^2 - G(|u|^2)]dx = \text{const} \equiv E_0$,

(iii) $\frac{d}{dt}\int |x|^2|u|^2 dx = -4\,\text{Im}\int r\bar{u}u_r dx, \quad r = |x|$,

(iv) $\frac{d}{dt}\,\text{Im}\int r\bar{u}u_r dx = -2\int |\nabla u|^2 dx + n\int [|u|^2F(|u|^2) - G(|u|^2)]dx$.

(All integrals are taken over $\mathbb{R}^n$, and $\bar{u}$ denotes the complex conjugate of u.)

Proof: We first multiply (4) by $2\bar{u}$ and take the imag-

inary part of the result to get

$$\frac{\partial}{\partial t}|u|^2 = \nabla \cdot (2\,\mathrm{Im}\bar{u}\nabla u).$$

Thus (i) follows. Now multiplying this identity by $|x|^2$, integrating over $\mathbb{R}^n$, and integrating the right-hand side by parts, we obtain (iii). To establish (ii) we multiply (4) by $2\bar{u}_t$, integrate, and take the real part of the resulting expression. It remains to derive (iv). For this purpose, multiply (4) by $2r\bar{u}_r$: $2ir\bar{u}_r u_t = 2r\bar{u}_r \Delta u + 2F(|u|^2) ru\bar{u}_r$. We integrate the real part of this expression over $\mathbb{R}^n$; the result can be written as

$$I = II + III,$$

where

$$I = \mathrm{Re}[i\int \sum_k x_k(\bar{u}_{x_k} u_t - u_{x_k}\bar{u}_t)dx],$$

$$II = 2\mathrm{Re}\int r\bar{u}_r \Delta u dx,$$

$$III = 2\mathrm{Re}\int F(|u|^2) ru\bar{u}_r dx.$$

We integrate by parts and find directly that

$$II = (n-2)\int |\nabla u|^2 dx,$$

$$III = -n\int G(|u|^2)dx.$$

Finally, I can be written as

$$I = \mathrm{Re}\left[i\int \sum_k x_k \left(\frac{\partial}{\partial t}(\bar{u}_{x_k} u) - \frac{\partial}{\partial x_k}(u\bar{u}_t)\right) dx\right]$$

$$= \frac{d}{dt}\,\mathrm{Re}[i\int ru\bar{u}_r dx] + n\mathrm{Re}[i\int u\bar{u}_t dx].$$

The last expression can be evaluated from (4) itself. Thus we obtain

$$I = \frac{d}{dt}\,\mathrm{Im}\int r\bar{u}u_r dx + n\int |\nabla u|^2 dx - n\int |u|^2 F(|u|^2)dx$$

from which (iv) follows.

With these identities we can establish our principal result:

Theorem: Let u be a classical solution to the Cauchy problem (4) with $\varphi \in \mathcal{S}$. Assume that

(1°) $E_0 \leq 0$;

(2°) $\mathrm{Im}\int r\bar{\varphi}\varphi_r dx > 0$;

(3°) there exists a constant $c_n > 1 + 2/n$ such that $sF(s) \geq c_n G(s)$ for all $s \geq 0$.

Then there exists a *finite* time T, estimable from above, such that

$$\lim_{t\to T^-} \|\nabla u(t)\|_2 = +\infty.$$

Proof: Wherever u exists we put

$$y(t) = \mathrm{Im}\int r\bar{u}u_r dx.$$

By hypothesis (2°), $y(0) > 0$ and by (iv) of the lemma we have

$$\dot{y}(t) = -2\int |\nabla u|^2 dx + n\int [|u|^2 F(|u|^2) - G(|u|^2)]dx.$$

Using hypothesis (3°) we find

$$\dot{y}(t) \geq -2\int |\nabla u|^2 dx + n(c_n - 1)\int G(|u|^2)dx.$$

We substitute (ii) of the lemma for the last term here on the right; this gives

$$\dot{y}(t) \geq -2\int |\nabla u|^2 dx + n(c_n - 1)[\int |\nabla u|^2 dx - E_0]$$

$$= [n(c_n - 1) - 2]\int |\nabla u|^2 dx - n(c_n - 1)E_0.$$

Now $E_0 \leq 0$ by (1°). We put

$$k_n = n(c_n - 1) - 2$$

so that $k_n > 0$ by (3°). Then the above inequality implies

$$\dot{y}(t) \geq k_n \|\nabla u(t)\|_2^2. \tag{5}$$

Since $y(0) > 0$ and $k_n > 0$, the function $y(t)$ is increasing wherever u exists; thus $y(t) > 0$ there. From (iii) of the lemma we have

$$\frac{d}{dt}\int r^2 |u|^2 dx = -4\mathrm{Im}\int r\bar{u}u_r dx = -4y(t) \leq 0.$$

Therefore, wherever u exists we have

$$\int r^2 |u|^2 dx \leq \int r^2 |\varphi|^2 dx \equiv d_0^2 < \infty.$$

The Schwarz inequality then yields

$$|y(t)| = y(t) \leq (\int r^2 |u|^2 dx)^{1/2} (\int |u_r|^2 dx)^{1/2}$$

$$\leq d_0 \|\nabla u(t)\|_2.$$

Hence from (5) we have the differential inequality

$$\dot{y}(t) \geq \frac{k_n}{d_0^2} y^2(t)$$

with

$$y(0) > 0.$$

It follows that on the interval

$$0 \leq t < \frac{d_0^2}{k_n y(0)}$$

we have the estimate

$$y(t) \geq \frac{y(0)d_0^2}{d_0^2 - k_n y(0)t};$$

thus we have the estimate

$$\|\nabla u(t)\|_2 \geq \frac{y(0)d_0}{d_0^2 - k_n y(0)t}$$

there too. Hence

$$\lim_{t\to T^-} \|\nabla u(t)\|_2 = +\infty$$

for some $T \leq T_0 < \infty$, where

$$T_0 = \frac{d_0^2}{k_n y(0)}.$$

This completes the proof.

Under an additional assumption, we can also conclude that $\sup_x |u|$ blows up. We state this as

Corollary 1: Suppose in addition to (1^0)—(3^0) that there is a constant $\sigma > 0$ such that

$$s^{-1}G(s) \quad \text{const } s^{\sigma} \text{ for all } s \geq 0.$$

Then $\|u(t)\|_\infty$ blows up in finite time.

Proof: We have $E_0 \leq 0$ by (1^0); hence by part (ii) of the lemma,

$$\|\nabla u(t)\|_2^2 \leq \int \frac{G(|u|^2)}{|u|^2} \, |u|^2 dx$$

$$\leq \text{const } \|u(t)\|_\infty^{2\sigma} \|\varphi\|_2^2.$$

Thus $\|u(t)\|_\infty$ blows up whenever $\|\nabla u(t)\|_2$ does.

The proof of the theorem does not seem to generalize easily. There is, however, one simple case which can be directly handled. Suppose that F depends on both x and $|u|^2$, $F = F(x, |u|^2)$. We put

$$G = G(x,u) = \int_0^u F(x,s) ds.$$

We calculate again expression *III* in the proof of the lemma:

$$III = \int \sum_k x_k F(x, |u|^2) \frac{\partial}{\partial x_k} |u|^2 dx$$

$$= \int \sum_k x_k \left(\frac{\partial G}{\partial x_k}(x, |u|^2) - \int_0^{|u|^2} F_{x_k}(x,s) ds \right) dx$$

$$= -n \int G(x, |u|^2) dx - \int \left(\int_0^{|u|^2} r \frac{\partial F}{\partial r}(x,s) ds \right) dx.$$

Hence if $F_r \leq 0$, we can "throw away" the extra term. We state this as

Corollary 2: Let the F in Eq. (4) depend on x and $|u|^2$,

$$F = F(x, |u|^2).$$

Assume (1^0) and (2^0) of the theorem as well as:

$(3^{0\prime})$ there is a constant $c_n > 1 + 2/n$ such that

$$sF(x,s) \geq c_n G(x,s) \text{ for all } s \geq 0, \; x \in \mathbb{R}^n,$$

$(4^{0\prime})$ $(\partial F/\partial r)(x,s) \leq 0$ for all $s \geq 0$, $x \in \mathbb{R}^n$ $(r = |x|)$.

Then the conclusion of the theorem holds.

Before specializing, we consider briefly the achievement of hypothesis (2^0) of the theorem. Consider for example an initial function $\varphi(x)$ of the form

$$\varphi(x) = \exp(i|x|^2)\psi(x),$$

where $\psi(x)$ is any real-valued nontrivial element of $\mathcal{S}$. Then a direct calculation gives

$$\text{Im} \int r \bar{\varphi} \varphi_r dx = 2 \int r^2 |\psi|^2 dx$$

so that (2^0) always holds in this case.

Finally, we consider Eq. (4) with $F(s) = s^{(p-1)/2}$ $(p > 1)$, i.e., the equation

$$iu_t = \Delta u + |u|^{p-1}u, \quad x \in \mathbb{R}^n, \quad t > 0, \quad u(x,0) = \varphi(x). \tag{6}$$

In this case we have

$$G(|u|^2) = \int_0^{|u|^2} s^{(p-1)/2} ds$$

$$= \frac{2}{p+1} |u|^{p+1}.$$

Then hypothesis (3^0) of the theorem requires that

$$s \cdot s^{(p-1)/2} \geq c_n \left(\frac{2}{p+1} s^{(p+1)/2} \right)$$

for some constant $c_n > 1 + 2/n$ and for all $s \geq 0$. This will be satisfied by choosing $c_n = (p+1)/2$ provided

$$(p+1)/2 = c_n > 1 + 2/n,$$

that is, provided $p > 1 + 4/n$. This is precisely the complementary condition mentioned in the Introduction. Hence we have

Corollary 3: For the Cauchy problem for Eq. (6) assume that

(1^0) $E_0 \equiv \int \left(|\nabla \varphi|^2 - \frac{2}{p+1} |\varphi|^{p+1} \right) dx \leq 0,$

(2^0) $\text{Im} \int r \bar{\varphi} \varphi_r dx > 0,$

(3^0) $p > 1 + 4/n.$

Then $\|\nabla u(t)\|_2$ and $\|u(t)\|_\infty$ blow up in finite time.

In this special case, $F(s) = s^{(p-1)/2}$, we can also establish that other L_q norms of the solution blow up in finite time. This we state as

Corollary 4: Consider Eq. (6) and assume (1^0)—(3^0) of Corollary 3. Then:

(a) for every $q \geq p+1$, $\|u(t)\|_q \to +\infty$ in finite time.

(b) Let q satisfy

$$\frac{n}{2}(p-1) < q < p+1.$$

If $n \geq 3$, assume also that $p < (n+2)/(n-2)$. Then $\|u(t)\|_q \to +\infty$ in finite time.

Proof: To prove (a), we note that, since $|E_0| = \text{const} < \infty$, $\|u(t)\|_{p+1}$ blows up in finite time since $\|\nabla u(t)\|_2$ does by Corollary 3. For $q > p+1$, we have, since $E_0 \leq 0$,

$$\|\nabla u(t)\|_2^2 \leq \frac{2}{p+1} \|u(t)\|_{p+1}^{p+1}$$

$$\leq \frac{2}{p+1} \left[\|u(t)\|_2^{\theta} \|u(t)\|_q^{1-\theta} \right]^{p+1},$$

where

$$\frac{1}{p+1} = \frac{\theta}{2} + \frac{1-\theta}{q}.$$

This proves (a). To prove (b), we assume that q is in

the indicated range and apply the Sobolev inequality[1] to get

$$\|u(t)\|_{p+1} \leq \text{const} \|\nabla u(t)\|_2^{\theta} \|u(t)\|_q^{1-\theta}, \tag{7}$$

where

$$\frac{1}{p+1} = \theta\left(\frac{1}{2} - \frac{1}{n}\right) + \frac{1-\theta}{q},$$

i.e., where

$$\theta = \frac{2n(p+1-q)}{(p+1)[2n-q(n-2)]}.$$

Now $E_0 \leq 0$; hence part (ii) of the lemma yields the estimate

$$\|\nabla u(t)\|_2 \leq \text{const} \|u(t)\|_{p+1}^{(p+1)/2}.$$

Thus, from (7) we have

$$\|u(t)\|_{p+1} \leq \text{const} \|u(t)\|_q^{1-\theta} \|u(t)\|_{p+1}^{(p+1)\theta/2}.$$

It follows that $\|u(t)\|_q$ blows up in finite time, provided

$$(p+1)\theta/2 < 1.$$

Using the expression for θ given above, we see that this last condition is equivalent to

$$q > (n/2)(p-1).$$

CONCLUDING REMARKS

In a recent paper, Ball[20] points out that arguments similar to those employed here, while proving nonexistence of global solutions, do not by themselves establish that nonexistence occurs by "blow up." For the Schrödinger equations considered here, we can show, by estimating the relevant integral equation, that nonexistence must occur by blow up, at least for $n \leq 3$. The invariance of the L_2 norm of a solution is helpful in this regard.

It remains to be seen if a similar blow-up method will apply to certain semilinear equations of relativistic quantum mechanics, e.g., to particular classes of nonlinear Dirac equations.

ACKNOWLEDGMENTS

The author is especially grateful to Professor W. Strauss, of Brown University, for encouragement and helpful suggestions which improved the original manuscript. We also thank Professor J. Rauch, Univeristy of Michigan, who showed (private correspondence) that Corollary 1 can be strengthened by omitting the extra hypothesis.

*Supported in part by NSF Grant MCS73-08567A3.

[1]A. Friedman, *Partial Differential Equations* (Holt, Rinehart, and Winston, New York, 1969).
[2]W. Strauss, "Nonlinear Scattering Theory," in *Scattering Theory in Mathematical Physics* (Reidel, Holland, 1974), pp. 53—78.
[3]W. Strauss, Arch. Rat. Mech. Anal. 55, 86—92 (1974).
[4]J.E. Lin, thesis, Brown University, 1976.
[5]V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. 61, 118 (1971) [JETP-Sov. Phys. 34, 62—9 (1972)].
[6]J.B. Keller, Commun. Pure Appl. Math. 10, 523—30 (1957).
[7]R. Knops and B. Strauchan, Sympos. Trends of Appl. of Pure Math. to Mechanics, Univ. of Lecce, (1975), pp. 187—206.
[8]H. Levine, Trans. Amer. Math.Soc. 192, 1—21 (1974).
[9]H. Levine, Arch. Rat. Mech. Anal. 51, 371—386 (1973).
[10]H. Levine, Math. Ann. 214, 205—20 (1975).
[11]H. Fujita, J. Fac. Sci. Univ. Tokyo, Sec. IA, 13, 109—24 (1966).
[12]R.T. Glassey, Math.Z. 132, 182—203 (1973).
[13]S. Kaplan, Commun. Pure Appl. Math. 16, 305—30 (1957).
[14]H. Kawarada, RIMS Kyoto University 10, 729—36 (1975).
[15]L. Payne. *Improperly Posed Problems in Partial Differential Equations*, SIAM Regional Conf. Ser. Appl. Math. 22 (Soc. Ind. Appl. Math., Philadelphia, 1975).
[16]L.E. Payne and D.H. Sattinger, Israel J. Math. 22, 273—03 (1975).
[17]M. Tsutsumi, RIMS Kyoto Univ. 8, 211—29 (1972).
[18]M. Tsutsumi, Math. Jpn. 17, 173—93 (1972).
[19]P. Lax, J. Math. Phys. 5, 611—13. (1964).
[20]J.M. Ball, "Remarks on blow-up and nonexistence theorems for nonlinear evolution equations," to appear.
[21]F. John, Commun. Pure Appl. Math. 27, 377—405 (1974).
[22]I. Segal, Ann. Math. 78, 339—64 (1963).

Instantaneous Cauchy surfaces, topology change, and exploding black holes[a)]

Robert H. Gowdy[b)]

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland

(Received 18 August 1976; revised manuscript received 12 May 1977)

Instantaneous Cauchy surfaces are defined and several of their properties are given. Instantaneous Cauchy surfaces are achronal surfaces whose Cauchy development interiors are maximal on the set of all achronal surfaces. It is shown that topology changes in such surfaces always result in nonempty future Cauchy horizons and departures from global hyperbolicity. It is also shown that the structure which is usually assumed for the background spacetime of a self-consistent exploding black hole implies a topology change in instantaneous Cauchy surfaces.

I. INTRODUCTION

If one quantizes fields in a curved spacetime background, then one must weaken cosmic censorship[1,2] enough to allow the naked singularities that are believed to accompany black hole explosions.[3,4] This prospect raises a serious problem of principle: One is forced to consider spacetimes that do not have Cauchy surfaces. What, then should one use for the "regular spacelike or null surfaces" that most versions of quantum field theory require for defining commutation relations and obtaining complete sets of observables? This paper introduces a new type of achronal surface which may help to resolve this problem. The proposed surface will be called an "instantaneous Cauchy surface" and is intermediate between a partial Cauchy surface and a Cauchy surface. This type of surface will be defined, some examples will be given to clarify the definition, and several properties which follow from the definition will be discussed.

Section II of this paper introduces the instantaneous Cauchy surface. Section III applys this concept to a type of singular background spacetime that has played a role in past speculations about quantum general relativity: the topology changing spacetimes discussed by Wheeler.[5] The basic result of Sec. III is the rather obvious one that, with suitable definitions, topology changes require naked singularities. The new result is that the instantaneous Cauchy surface concept supplies the "suitable definitions." Furthermore, with these definitions, the exploding black hole spacetimes that are of current interest because of the Hawking process[3,4] turn out to be examples of topology change.

The notation and terminology of this paper are chosen to agree with the text by Hawking and Ellis.[1] One slight departure from the Hawking and Ellis conventions is that here Cauchy surfaces and partial Cauchy surfaces are only required to be achronal and need not be acausal. Another slight departure is that the interior Cauchy development $\mathrm{int}\tilde{D}(S)$ of a surface S is represented here by $D^0(S)$. In this paper, a *spacetime* (M, g) is a connected, Hausdorff C^∞ 4-manifold M together with a C^2 Lorentz-signature $(-+++)$ metric g which renders M time orientable.[6] The C^2 differentiability condition amounts to the assumption that any shock waves or dust caustics have been removed by a "smoothing process" which replaces piecewise C^2 metrics by C^2 metrics.[7]

II. INSTANTANEOUS CAUCHY SURFACES

This paper is concerned with singularities which disrupt attempts to predict the future from initial data on "regular surfaces." To define and classify such singularities, one must define "regular." Certainly a Cauchy surface deserves to be called regular, but spacetimes which admit Cauchy surfaces are rather uninteresting. A Cauchy surface is too much to ask for. A partial Cauchy surface, on the other hand, is too little to ask for. Figure 1 shows a well-known example of a bad choice of partial Cauchy surface in Minkowski space. This surface is the set of points at a constant proper-time interval to the past of an event O. It is an unfortunate choice because its Cauchy development interior is just $I^-(O)$ when there are other surfaces such as $t=0$ whose Cauchy development interiors are larger. The implication here is that surfaces should be judged according to the interiors of their Cauchy developments: If $D^0(S)$ is a proper subset of $D^0(S')$, then S' is a better surface than S. The reason for comparing interiors is that it is possible for a null-surface part of S to lie outside of $D^0(S)$ and therefore be useless for the purpose of evolving initial data. Define an *instantaneous Cauchy surface* to be an achronal surface S whose Cauchy development interior is maximal on the set of all achronal surfaces: If S' is an achronal surface and $D^0(S) \subset D^0(S')$, then $D^0(S) = D^0(S')$. An immediate consequence of this definition together with the well-known result that globally hyperbolic spacetimes admit Cauchy surfaces[8] is:

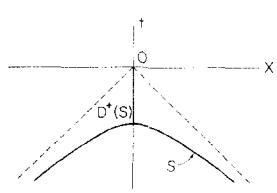


FIG. 1. An unsatisfactory partial Cauchy surface in Minkowski space.

a)Supported in part by the National Science Foundation, Grant Number MPS 74-18386-A01.
b)Supported in part by the Alfred P. Sloan Foundation.

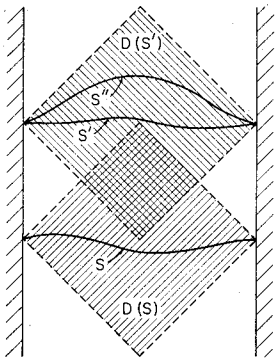


FIG. 2. The universal covering of anti-deSitter space. In this conformal diagram, 2-spheres are represented by points and infinity is mapped into a pair of vertical lines (not included in the manifold). Any two acausal partial Cauchy surfaces either have the same Cauchy development as with S' and S'' or fail to be comparable as with S and S'. Thus, in this spacetime, every acausal partial Cauchy surface is an instantaneous Cauchy surface.

Property 1: In a globally hyperbolic spacetime, the instantaneous Cauchy surfaces are just the Cauchy surfaces.

The example described by Fig. 1 shows that a partial Cauchy surface may fail to be an instantaneous Cauchy surface even when there are no singularities. One can easily show that surfaces which are generated by inextendable null geodesics without conjugate points fail to be instantaneous Cauchy surfaces in strongly causal spacetimes because such surfaces have empty Cauchy development interiors. However, there are spacetimes in which all partial Cauchy surfaces which are not of this degenerate type are instantaneous. One such spacetime, the universal covering of anti-deSitter space, is shown in Fig. 2.[9] In the closed Robertson–Walker spacetime shown in Fig. 3, the "big bang" singularity is responsible for some nondegenerate or even acausal partial Cauchy surfaces failing to be instantaneous. It is also possible to have a spacetime with no instantaneous Cauchy surfaces at all. For example, the Gödel "rotating" universe[10] has closed timelike lines through every point and therefore can have no instantaneous Cauchy surfaces simply because it has no achronal surfaces. It is natural to ask "when and where do instantaneous Cauchy surfaces exist?" Work currently in progress indicates that a variety of existence theorems can be proven. For example, strong causality is enough to guarantee the existence of instantaneous Cauchy surfaces through every point of a spacetime. This work will be described in a following paper.

III. TOPOLOGY CHANGE

One source of interest in topology change is Wheeler's speculation that a fully quantized theory of general relativity may, in some sense, require topology changes in the 3-manifolds that are used to label complete sets of field observables. This speculation cannot be dealt with at the present time because there is, as yet, no fully satisfactory quantum theory of general relativity. However, one can consider situations in which quantized nongravitational fields react back on a spacetime which is not quantized and is defined in a semiclassical mean-field sense. Wheeler's speculation then leads one to wonder if topology change can become important in these supposedly more manageable situations. How would one recognized a background spacetime that is describing a topology change? One would define a kind of surface which supports complete or at least maximal sets of field observables and one would then look for topology changes in these surfaces. The instantaneous Cauchy surfaces are obvious candidates for these surfaces.

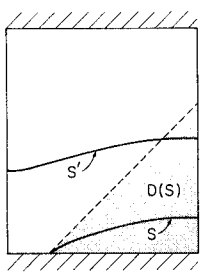


FIG. 3. A conformal diagram of a closed Robertson–Walker spacetime. The lower horizontal boundary is the "big bang." The partial Cauchy surface S cannot be an instantaneous Cauchy surface because $D^0(S) = D(S)$ is a proper subset of $D^0(S')$.

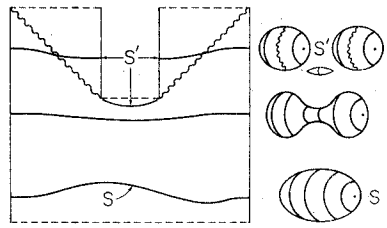


FIG. 4. Topology change by spherically symmetric fission. Each point in the left-hand figure represents a 2-sphere orbit of spherical symmetry. The areas of these spheres go to zero at the edges of the figure. Radial light rays travel at ±45° to the vertical. The vertical boundaries are world lines of centers of spherical symmetry and are not singular. The bottom boundary, shown dashed because it is not in the manifold, is the big bang singularity while the top boundaries are final collapse singularities. The diagonal wavy lines represent the future Cauchy horizon $H^+(S)$. The figures on the right display three successive instantaneous Cauchy surfaces with an angle coordinate restored so that 2-spheres now appear as circles instead of points.

In order to understand what it means to define topology change in terms of instantaneous Cauchy surfaces, consider some examples. Figure 4 shows the Penrose diagram of a spacetime that is often cited as an example of topology change, a closed universe which undergoes fission.[11] Figure 5 shows a similar topology change in which a closed universe buds off from an open one. These examples show that it will not do to define topology change in terms of partial Cauchy surfaces because one could take just one component of the final surface S' in each case and note that it is homeomorphic to S. The surfaces S and S' in Figs. 4 and 5 are instantaneous Cauchy surfaces. Notice that S' has an extra piece which is needed to supply initial data for the region that falls into the singularity that is associated with the topology change. Thus, the "fission" process produces

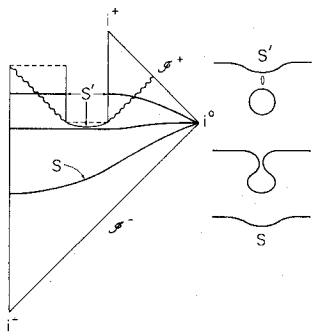


FIG. 5. A closed universe budding from an open universe. The conventions are as in Fig. 4 except that a conformal transformation has mapped the null infinities into diagonal boundaries and the figures at the right are to be rotated about their vertical symmetry axes in order to restore the angle coordinate.

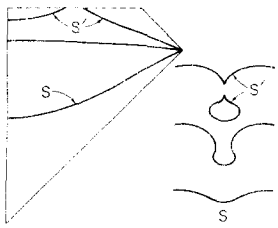


FIG. 6. A bogus topology change due to improper choice of partial Cauchy surfaces. The conventions are as in Figs. 4 and 5.

two regular 3-manifolds and a sort of "left-over scrap." Figure 6 shows another example that is sometimes described as a topology change, the collapse of a star to form a classical black hole. However, the Penrose diagram of this spacetime makes it quite clear that the topology change in this case is an artifact of the choice of partial Cauchy surfaces. Instantaneous Cauchy surfaces in this globally hyperbolic spacetime are just the Cauchy surfaces and cannot change topology.[8]

It is fairly obvious at an intuitive level that topology changes require spacetime singularities. With topology change defined in terms of instantaneous Cauchy surfaces, this idea finds precise expression in terms of the properties of these surfaces. From Property 1 and Geroch's result that all Cauchy surfaces are homeomorphic to one another,[8] one can easily show the following property:

Property 2: If S and S' are instantaneous Cauchy surfaces in a spacetime M and either $H(S)$ is empty or M is globally hyperbolic, then S' is homeomorphic to S.

If one chooses to regard any departure from global hyperbolicity as a naked singularity,[2] then the property states that a topology change is necessarily accompanied by a naked singularity which causes a breakdown in the unique evolution of fields that obey causal wave equations. The difficulty with this property is that it says nothing about the location of the singularities and ties them to the topology change only by a kind of "circumstantial evidence." For example, the property leaves open the possibility that the singularity occurs long before the topology change becomes manifest or perhaps long afterward. Fortunately, a more specific property can be proven without too much trouble:

Property 3: If S and S' are instantaneous Cauchy surfaces in a spacetime with S acausal, $S' \subset I^+(S)$, and either $H^+(S)$ empty or $I^+(S)$ globally hyperbolic, then S' is homeomorphic to S.

Proof: Suppose $H^+(S)$ is empty and consider the submanifold $K := I^+(\tilde{D}(S))$. Because S is acausal, its Cauchy development includes a neighborhood of each of its points. Thus, S is in the interior of $\tilde{D}(S)$ and is a subset of K. Regard K as a spacetime and note that $H(S, K) = H^+(S, K)$. From the way in which K has been defined, $H^+(S, K) = H^+(S)$ so that $H(S, K)$ is empty. Thus S is a Cauchy surface for K which is therefore globally hyperbolic and thus $K = \tilde{D}(S, K) \subset \tilde{D}(S)$. But, $S' \subset I^+(S) \subset K \subset \tilde{D}(S)$ so that $D^0(S') \subset D^0(S)$. Since S' is an instantaneous Cauchy surface, it follows that $D^0(S') = D^0(S)$. Thus, S' is also a Cauchy surface for K and, from a result of Geroch,[8] S' is homeomorphic to S.

Now suppose that $I^+(S)$ is globally hyperbolic. It then includes a Cauchy surface S'', an achronal surface such that $I^+(S) \subset \tilde{D}(S'')$.[8] But every inextendable timelike curve through a point of the acausal surface S enters $I^+(S)$ and must therefore intersect S''. Consequently $S \subset \tilde{D}(S'')$ so that $D^0(S) \subset D^0(S'')$. Because S is an instantaneous Cauchy surface, $D^0(S) = D^0(S'')$. But then $I^+(S) \subset D^0(S'') \subset D^0(S)$ and thus $S' \subset D^0(S)$ which leads to the same conclusion as before.

Because the whole point of the instantaneous Cauchy surface concept is that these surfaces can be defined in spacetimes that behave badly, the useful form of Property 3 is the negative one:

Property 3': If S and S' are instantaneous Cauchy surfaces in a spacetime with S acausal, $S' \subset I^+(S)$ and S' is not homeomorphic to S, then $H^+(S)$ is not empty and $I^+(S)$ is not globally hyperbolic.

Thus, a topology change to the future of S induces a naked singularity to the future of S.

IV. EXPLODING BLACK HOLES

Topology change has, until now, been a highly speculative subject with only the most tenuous connection to known physical processes. With the definition of topology change in terms of instantaneous Cauchy surfaces, it is now possible to recognize that a plausible example of topology change is already known and under intensive investigation. Hawking's exploding black hole is a topology change. Figure 7 shows the Penrose conformal diagram of spherically symmetric exploding black hole spacetime. This is the spacetime which most current workers (with notable exceptions[12]) expect to result from a fully self-consistent black hole explosion. Without a precise definition of topology change one would not think of this spacetime as an example of topology change. The surfaces S and S' shown in the figure appear to be "regular" and are homeomorphic to each other. However, the surface S' is not an instantaneous Cauchy surface because the surface $\dot{I}^-(S')$ has a larger Cauchy development. To obtain an instantaneous Cauchy surface which includes S' one must add an extra piece which may be taken to be the surface of the black hole. The Hawking particle creation calculation makes it quite clear that final particle states are to be constructed from operators defined on both S' and the surface of the black hole. Thus, the instantaneous Cauchy surface leads one to make the right choice of final surface for the quantum field theory calculation. Because the final surface $\dot{I}^-(S')$ consists of two disjoint components, it is not homeomorphic to S and there is a topology change.

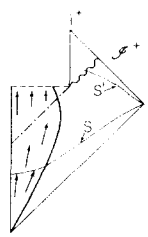


FIG. 7. A possible background spacetime for a self-consistent exploding black hole. The conventions are as in Figs. 4–6. The wavy line shows the future Cauchy horizon that is demanded by the proposition in the text and the dot-dash line shows the surface of the black hole which must be combined with S' to make an instantaneous Cauchy surface.

By making full use of the instantaneous surface concept one can do somewhat better than recognizing one speculative example of an exploding black hole as a topology change. One can define a general class of disappearing black hole spacetimes and show that they are all topology changes. The usual definition of a black hole, exploding or otherwise, requires weak asymptotic simplicity so that an exterior future null infinity $\mathcal{I}^+$ and a corresponding black hole surface $\dot{J}^-(\mathcal{I}^+)$ can be defined. However, if the black hole disappears, there is no reason to wait forever to see the last ray that escapes from it. In Fig. 7, one can just wait until the surface S' and define the black hole surface to be just $\dot{I}^-(S') - S'$. Here S' is a "regular" surface for the external part of the spacetime. A natural generalization of this surface is the *externally instantaneous Cauchy surface* which is defined to be a surface S which is an instantaneous Cauchy surface for the spacetime $I^-(S) \cup S \cup I^+(S)$. A generalization of the situation described by Fig. 7 is then the following proposition.

Proposition: If a spacetime contains a closed, edgeless, externally instantaneous Cauchy surface S' and an acausal instantaneous Cauchy surface S such that

(1) $\dot{I}^-(S') - S'$ is not empty,

(2) $\dot{I}^-(S')$ is an instantaneous Cauchy surface,

(3) $\dot{I}^-(S') \subset I^+(S)$,

(4) S is homeomorphic to S',

then the instantaneous Cauchy surfaces S and $I^-(S')$ are not homeomorphic, $H^+(S)$ is not empty, and $I^+(S)$ is not globally hyperbolic.

Proof: The closed, edgeless nature of S' ensures that $\dot{I}^-(S')$ consists of at least two disjoint components, S' and the set of black hole surfaces that make up $\dot{I}^-(S') - S'$. The rest follows from Property 3'.

The disappearing black hole proposition is not quite as general as one might wish. It applies to a spacetime which contains black holes [from (1)] that have all formed from gravitational collapse [from (2) and (3)] and an external region that shows no topology change [from (3) and (4)]. Its applicability to primordial black holes depends on the details of their formation. Requirements (2) and (3) could easily be violated for black holes associated with very early density fluctuations in the big bang. However, for primordial black holes that can actually be said to "form from density fluctuations" in the sense that there is an instantaneous Cauchy surface between them and the initial singularity, the proposition applies and says that the disappearance of such black holes is equivalent to a topology change in instantaneous Cauchy surfaces.

ACKNOWLEDGMENTS

I am grateful for the hospitality of the Mathematical Institute at Oxford where this work was begun. I am also grateful to Dieter Brill and Charles Misner for their help and encouragement and to all of the Maryland relativity group for reading and commenting on earlier drafts of this paper. I am particularly indebted to Mark Gotay and Robert Budic who provided key insights and many helpful comments.

[1]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).

[2]Strong cosmic censorship asserts that no point on the causal boundary of the strongly causal region of a spacetime can have both a past and a future. For details about causal boundaries see R.P. Geroch, E.H. Kronheimer, and R. Penrose, Proc. Roy. Soc. (London) A 327, 545 (1972). A naked singularity is regarded as a causal boundary point with both a past and a future. That strong cosmic censorship is essentially equivalent to assuming global hyperbolicity is shown by R. Budic and R.K. Sachs, J. Math. Phys. 15, 1302—9 (1974). Expecially see Theorem 6.2 on page 1308 of that article. Additional material on naked singularities may be found in R. Penrose, article in *Gravitational Radiation and Gravitational Collapse*, edited by C.M. DeWitt-Morette, IAU Symposium 64, 82 (1974); article in *Confrontation of Cosmological Theories with Observational Data*, edited by M.S. Longair (Reidel, Dordrecht-Holland, Boston, 1974), see pp. 263, 266.

[3]S.W. Hawking, Nature 248, 20 (1974); Commun. Math. Phys. (Germany) 43, 199 (1975); Cal. Tech. Preprint OAP-420 (1975); J.B. Hartle and S.W. Hawking, Phys. Rev. D 13 2188—2203 (1976); R. Wald, Commun. Math. Phys. 45, 9 (1975); Phys. Rev. D 13, 3176 (1976).

[4]B.S. DeWitt, Phys. Rep. 19, 295 (1975).

[5]For reviews on quantum general relativity, see D.R. Brill and R.H. Gowdy, Rep. Prog. Phys. 33, 423 (1970); A. Ashtekhar and R. Geroch, *ibid.* 37, 1211—56 (1974); C.J. Isham, *Quantum Gravity, an Oxford Symposium*, edited by C.J. Isham, R. Penrose, and D.W. Sciama (Clarendon, Oxford, 1975), pp. 1—65; J.A. Wheeler in *Relativity, Groups, and Topology*, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, Amsterdam, 1968), p. 242. For a discussion of topology change see D.R. Brill in *Magic Without Magic*, edited by J.R. Klauder (Freeman, San Francisco, 1972), pp. 309—16.

[6]A discussion of the time-orientability requirement may be found in Hawking and Ellis, Ref. 1, p. 181.

[7]Examples of dust-caustic naked singularities are given in P. Yodzis, H.J. Seifert, and H. Muller zum Hagen, Commun. Math. Phys. 34, 135 (1973). Topologies for the space of spacetime metrics are discussed in R.P. Geroch, *ibid.* 13, 180 (1969) and in an article by R.P. Geroch in *Relativity*, edited by M. Carmeli, S.I. Fickler, and L. Witten (Plenum, New York, London, 1970), p. 259.

[8]R.P. Geroch, J. Math. Phys. 11, 437—49 (1970). Note that, as in this paper, Geroch allows Cauchy surfaces to be achronal and does not require them to be acausal. Also see Hawking and Ellis, Ref. 1, p. 212 (Proposition 6.6.8).

[9]Hawking and Ellis, Ref. 1. See pp. 131—4 for a discussion of anti-deSitter space.

[10]N. Gödel, Rev. Mod. Phys. 21, 447 (1949); Proc. Intern. Congr. Math. (Cambridge, Mass.) 1, 175 (1952); Hawking and Ellis, Ref. 1, p. 168; F.J. Tipler, thesis, University of Maryland (1976).

[11]A similar example of topology change by fission, but for a different symmetry, was treated by R.H. Gowdy, Ann. Phys. (N.Y.) 83, 203 (1974). See especially p. 218.

[12]D.G. Boulware, Phys. Rev. D 13, 2169—87 (1976); U. Gerlach, Phys. Rev. D 14, 1479—1508 (1976).

Pairing in the limit of a large number of particles

R. W. Richardson

Physics Department, New York University, New York, New York 10003
(Received 12 October 1976)

Expressions for the ground and excited state energies of a system of nucleons interacting through pairing forces are given as a power series in inverse powers of the number of particles. The expressions are valid for systems with superfluid ground states and either $J=0$ or $L=0$ pairing. The first three terms in the expansion are given explicitly, and they exhibit excitations with both vibrational and rotational (in isospin space) character. Analytical and numerical results are given for a model system with a two-level single-particle spectrum.

I. INTRODUCTION

Exact expressions for the energies and wavefunctions of the ground and excited states of various pairing Hamiltonians of interest in nuclear structure studies have been developed over the past few years.[1-3] These results complement the many approximate calculations[4] that have been applied to these Hamiltonians and they have also provided the basis for numerical calculations.[5] The approximate calculations have generally been based on techniques developed in the BCS theory of superconductivity.[6] These techniques are asymptotically exact in the limit of a large system. The numerical studies, on the other hand, are useful for studying small systems. The analytical connection between these two approaches is given in this paper. We give explicit expressions for the energies of the states and the occupation probabilities of the single-particle levels as expansions in inverse powers of the number of particles. We show that the leading term in this expansion is just the BCS expression for the energy and the higher order terms in the expansion can be obtained as functions of the energy-gap and chemical potential that characterize the BCS expression. These results are obtained for the ground state and pairing-vibration excited states as well as the isospin "rotational" states built upon these band heads. We are therefore able to give explicit expressions for the parameters characterizing these modes of collective excitation and their interactions. States with single-particle as well as collective excitations, i.e., unpaired nucleons, will be treated in a subsequent paper.

It is well known that a system of nucleons interacting through pairing forces can make a discontinuous transition to or from a superfluid state as one varies the number of nucleons. This observation seems to imply that the expansion we have just described does not exist since it can not represent such nonanalytic behavior. However, our expansion has a different interpretation. It must be viewed as the sequence of corrections to the thermodynamic limit that are due to the finite size of the system. That is, the expansion of the energy is to be viewed as an expansion in inverse powers of the number of particles taken at fixed, nonvanishing, energy-gap and fixed chemical potential. The point at which this interpretation is imposed on our results will be indicated in what follows. Thus, our expansion will be shown to be valid as long as the leading term exhibits superfluidity. The interesting transition region still remains an open problem.

We now review the results that form the starting point for this work. We consider $A=2P$ nucleons contained in a potential well whose single-particle levels are labeled by the index k and have energies ϵ_k and spatial degeneracies Ω_k. These nucleons interact with each other through a pairing interaction that is effective in one of the following two-body states: (1) $J=0$, $T_z=1$, Ref. 1; (2) $J=0$, $T=1$, Ref. 2; or (3) $L=0$, $S+T=1$, Ref. 3. Here, L and J refer to the orbital and total angular momentum of the pair and S and T refer to its total spin and isospin. In these three cases the energies of the states of the system are given by

$$E=2\sum_{i=1}^{P} e(i), \tag{1}$$

where the $e(i)$ are, in general complex, roots of the system of algebraic equations

$$\frac{-A}{\kappa G}+\sum_{j=1}^{P}{}' \frac{1}{e(i)-e(j)}-\frac{1}{\kappa}\sum_{k}\frac{\Omega_k}{e(i)-\epsilon_k}=0,$$
$$i=1,\ldots,P, \tag{2}$$

where the pairing interaction strength is taken to be G/P so that the energy is an extensive quantity for fixed G in the thermodynamic limit. Furthermore, we assume that the sum over k is of order A, in order to obtain a sensible thermodynamic limit, i.e., we are assuming that the product of the interaction strength and the level density are a constant in this limit. Each state of the many-nucleon system corresponds to a different set of roots of this system of equations and the parameter κ distinguishes between the three different interactions given above taking on the values

$$\begin{aligned}
&\kappa=2, && \text{case (1)},\\
&=\frac{P(P-3)+T(T+1)}{P(P-1)}, && T=P,P-2,\ldots,0 \text{ or } 1, \text{ case (2)},\\
&=\frac{P(P-6)+p(p+4)}{P(P-1)}, && p=P,P-2,\ldots,0 \text{ or } 1, \text{ case (3)}.
\end{aligned} \tag{3}$$

Case (1) is the case treated by BCS and is of course the same as case (2) with $T=P$. In order to make our discussions explicit, we will always assume that κ takes on the values given in case (2) unless otherwise stated.

The occupation probabilities of the single-particle levels for a given state of the system may be obtained from (1) by differentiating E with respect to the appropriate single-particle energy. More details on these equations can be found in the references cited.

Progress in understanding Eqs. (2) was made by Gaudin[7] who, guided by an electrostatic analogy, transformed the equations into integral equations equivalent to the BCS equations. However, his technique was limited to obtaining the leading contribution to the energy, in powers of $1/A$, and therefore was not applicable to excited states which are degenerate with the ground state to this order. In this paper, we use the electrostatic analogy to recast Eqs. (2) in a form that is amenable to an expansion in powers of $1/A$. This is done and applied to the ground state and its associated band of isospin rotational states in Sec. II. In Sec. III, we apply the expansion to pairing-vibration excited states and their isospin rotational states. In Sec. IV, we consider some analytical and numerical examples of our results and we conclude with a general discussion in Sec. V. The reader is urged to look forward to the examples in Sec. IV as an aid in following the analysis presented in Secs. II and III.

II. GENERAL THEORY AND THE GROUND STATE

We seek a method for representing the solutions of Eqs. (2) as a power series in the small parameter $1/A$. In developing this method, we will neglect the A dependence of the parameter κ and therefore each term in the expansion will be a function of κ as well as the other parameters that characterize the system and the particular state of the system under consideration. The method is suggested by an electrostatic analogy[7] to Eqs. (2). In this analogy, one interprets the solution of these equations as a problem in two-dimensional electrostatics in which the roots $e(i)$ are the locations, in the complex plane, of P free charges (actually lines of charge perpendiuclar to the plane) of unit strength. Eqs. (2) then describe the equilibrium configuration of these charges under the influence of a uniform external field, $-A/G$, and the field of fixed charges of strength, $-\Omega_k/\kappa$ located at the points ϵ_k on the real axis. The equations then state that the force on the ith charge is zero under the influence of these external fields plus the mutual Coulomb repulsion of the various free charges, i.e., they describe the equilibrium configuration of the free charges. In this section we exploit this analogy and develop an exact differential equation for the electrostatic field produced by the charges which can be solved as a power series in $1/A$. Various multipole moments of the field are shown to give physically important information such as the energy of the state and these are also obtained as a power series in $1/A$. Finally, the κ dependence of the coefficients is shown to give the parameters characterizing the collective states, e.g., a "variable moment of inertia in isospin space."

In order to recast Eqs. (2) in a form that is amenable to a power series expansion in $1/A$, we introduce the electrostatic field associated with the charge distribution

$$F(z)=\sum_{i=1}^{P}\frac{1}{z-e(i)}-\frac{1}{\kappa}\sum_{k}\frac{\Omega_k}{z-\epsilon_k}-\frac{A}{\kappa G}. \tag{4}$$

Two expansions of this field are useful. A multipole expansion enables us to relate various physical quantities to the multipole moments of F and a power series expansion in powers of $1/A$ enables us to solve the field equation for F developed below. We first consider the multipole expansion and define the multipole moments of F by

$$F(z)=\sum_{m=0} F^{(m)}z^{-m}, \tag{5}$$

where the expansion is valid in the domain $|z|>\max[\epsilon_k, e(i)]$. Note that superscripts are used to denote the various multipole moments of F. Direct expansion of (4) yields the following expressions for the first three multipole moments:

$$F^{(0)}=-A/\kappa G, \tag{6}$$

$$F^{(1)}=P-\Omega/\kappa, \tag{7}$$

$$F^{(2)}=\sum_{i=1}^{P}e(i)-(1/\kappa)\sum_{k}\Omega_k\epsilon_k, \tag{8}$$

where $\Omega=\sum_k\Omega_k$ is the total spatial degeneracy of the system. We will not need the higher moments in what follows. We will call Eq. (6) the generalized gap equation and we rewrite (7) in the form

$$P=\tfrac{1}{2}A=F^{(1)}+\Omega/\kappa \tag{9}$$

and call it the generalized equation for the chemical potential. We use (1) to rewrite (8) as

$$E=2\left[F^{(2)}+(1/\kappa)\sum_{k}\Omega_k\epsilon_k\right] \tag{10}$$

and call this the generalized expression for the energy. These expressions are generalized in the sense that they are exact extensions of the BCS equations. We will show that the leading terms in an expansion in powers of $1/A$ (terms of order A) of the multipole moments in Eqs. (6), (9), and (10) reproduces the corresponding equations of the BCS theory. Furthermore, we will show that the leading terms in Eqs. (6) and (9) saturate these equations and they have no higher order corrections. Thus, in accord with the general concepts of statistical mechanics, the energy gap and chemical potential are determined by equating extensive quantities in the thermodynamic limit, and they have no higher order corrections. On the other hand, the expression for the energy (10) will have higher order corrections thereby modifying the role played by the energy gap and chemical potential in determining the ground state energy and the excitation spectrum of the system.

We now turn to the field equation which will provide the basis for the $1/A$ expansion. Using Eqs. (2) and (4), one can show that the field F identically satisfies the equation

$$\frac{dF}{dz}+F^2=\frac{1}{\kappa}\sum_{k}\frac{\Omega_k}{(z-\epsilon_k)^2}+\frac{1}{\kappa^2}\left(\sum_{k}\frac{\Omega_k}{(z-\epsilon_k)}+\frac{A}{G}\right)^2-\frac{2}{\kappa}\sum_{k}\frac{\Omega_k H(k)}{z-\epsilon_k}, \tag{11}$$

where $H(k)$ is the field produced by the free charges at the locations of the fixed charge, ϵ_k,

$$H(k)=\sum_{i=1}^{P}\frac{1}{\epsilon_k-e(i)}=\frac{1}{2\pi i}\oint_C\frac{F(z)\,dz}{\epsilon_k-z}, \tag{12}$$

where the contour C is defined to include only those singularities of F that arise from the free charges located at the points $e(i)$.

It will prove useful to rewrite the chemical potential and energy equations (9) and (10) in terms of the values of $H(k)$. This may be done by solving (11) for large z to obtain expressions for the multipole moments of F. Substituting these results into (9) yields

$$\frac{A}{2}=\frac{G}{A}\sum_k\Omega_k H(k) \tag{13}$$

and (10) becomes

$$E=-G[\Omega-(P-1)\kappa/2]+(G/P)\sum_k\Omega_k\epsilon_k H(k). \tag{14}$$

Thus, the values of the field due to the free charges at the locations of the fixed charges determine the energy of the state and, by differentiation with respect to ϵ_k, the occupation probabilities of the single-particle levels.

Our solution of the above equations will be in the form of a power series expansion of E, F, and H in powers of $1/A$, e.g.,

$$F(z)=\sum_{n=0}^{\infty}F_n(z), \tag{15}$$

where F_n is of order A^{1-n}. Throughout this paper we will use subscripts to denote the order with this convention. Substituting these expansions into (11), expanding the right-hand side (recall that Ω_k is assumed to be of order A for the sake of this expansion), and equating the terms of the same order gives for the first three terms

$$[F_0(z)]^2=\left[\frac{1}{\kappa}\sum_k\frac{\Omega_k}{(z-\epsilon_k)^2}+\frac{A}{\kappa G}\right]^2-\frac{2}{\kappa}\sum_k\frac{\Omega_k H_0(k)}{z-\epsilon_k}, \tag{16}$$

$$2F_0(z)F_1(z)=\frac{1}{\kappa}\sum_k\frac{\Omega_k}{(z-\epsilon_k)^2}-\frac{2}{\kappa}\sum_k\frac{\Omega_k H_1(k)}{z-\epsilon_k}-\frac{dF_0}{dz}, \tag{17}$$

$$2F_0(z)F_2(z)=-\frac{2}{\kappa}\sum_k\frac{\Omega_k H_2(k)}{z-\epsilon_k}-[F_1(z)]^2-\frac{dF_1(z)}{dz}, \tag{18}$$

where

$$H_n(k)=\frac{1}{2\pi i}\oint_C\frac{F_n(z)\,dz}{\epsilon_k-z}. \tag{19}$$

We will not consider higher order terms in the expansion here. As they stand, these equations, which provide the basis for all subsequent developments in this paper, look like a set of integral equations that has to be solved at each order for $F_n(z)$. The dipole moment of F_n would then give the nth order contribution to the energy. Fortunately, we will show that the $H_n(k)$ can be determined directly from self-consistency requirements on $F_n(z)$ and at each order we have only algebraic equations to solve.

The equations derived thus far apply equally well to all states of the system. We obtain results for a particular state of the system by considering a particular solution of the equations. For the remainder of this section, we will concentrate on the ground state of the system. Excited states will be treated in the next section.

For a state in the ground state band, we first need to solve (16) for the F_0 that describes this state. This solution can then be used to obtain F_1 and F_2 from (17) and (18). Our technique is to use physical arguments to construct F_0 and then show that it gives the results of the BCS theory and satisfies (16). In order to construct F_0, we need to know what the limiting distribution of the free charges is as their number becomes infinite and the strengths of the fixed charges and external field increase proportionately. We assume that they coalesce to form a line of charge (or a sheet if one thinks in three dimensions). This assumption is suggested by numerical studies which indicate this behavior. The result is that the poles of F arising from the first term in (4) merge to form a branch cut. We assume that this branch cut extends from the point a in the complex plane to the point a^* since complex roots of Eqs. (2) occur in complex conjugate pairs, i.e., the charge distribution is symmetric under reflection through the real axis. Since there are P free charges of unit strength, the total charge on the line is P or

$$\frac{1}{2\pi i}\oint_C F_0(z)\,dz=P, \tag{20}$$

where, as before, the contour C encloses the singularities of F that arise from the first term of (4) which, by assumption, form a branch cut. We can also require that F_0 describe the fixed charges exactly, i.e., that it have poles at the points $z=\epsilon_k$ with residues $-\Omega_k/\kappa$, or

$$\lim_{z\to\epsilon_k}(z-\epsilon_k)F_0(z)=-\Omega_k/\kappa, \tag{21}$$

which is consistent with our assumption that the Ω_k are to be treated formally as being of order A. We also require that F_0 describe the external field exactly, i.e.,

$$\lim_{z\to\infty}F_0(z)=-A/\kappa G. \tag{22}$$

Thus the field F_0 describes the external and monopole fields of F exactly. The field with these properties is

$$F_0(z)=-\frac{Z}{\kappa}\sum_k\frac{\Omega_k}{\eta_k(z-\epsilon_k)}, \tag{23}$$

where

$$\begin{aligned}Z&=[(z-a)(z-a^*)]^{1/2}=Z(z),\\ \eta_k&=|\epsilon_k-a|=[(\epsilon_k-\lambda)^2+\Delta^2]^{1/2},\end{aligned} \tag{24}$$

and where $a=\lambda+i\Delta$ is yet to be determined.

The parameters λ and Δ in F_0 may be obtained from the values of the external field and monopole moment, Eqs. (6) and (7), and then the leading contribution to the energy is obtained from the dipole moment (10). Expanding (23) in powers of $1/z$ and substituting the resulting multipole moments into Eqs. (6), (9), and (10) yields the BCS gap equation

$$\frac{A}{G}=\sum_k\frac{\Omega_k}{\eta_k}, \tag{25}$$

the BCS equation for the chemical potential

$$A=\frac{2}{\kappa}\sum_k \Omega_k\left(1-\frac{\epsilon_k-\lambda}{\eta_k}\right), \tag{26}$$

and the BCS expression for the ground state energy

$$E_0=\frac{2}{\kappa}\sum_k \Omega_k\epsilon_k\left(1-\frac{\epsilon_k-\lambda}{\eta_k}\right)-\frac{A}{\kappa G}\Delta^2. \tag{27}$$

Note, that for a strict comparison with BCS, we must set $\kappa=2$, see Eq. (3). We therefore interpret the real and imaginary parts of a, λ, and Δ, as the BCS chemical potential and energy gap. Furthermore, since there are no higher order corrections to the first two moments of F, these parameters are fixed by F_0.

We have shown that the field F_0 given by (23) reproduces the results of the BCS theory. We now show that it is also a leading order solution of the field equation (16). This is done by using the expression (23) to calculate $H_0(k)$ from (12) and substituting the result into (16) to verify that it is satisfied. This requires a certain amount of straightforward algebra and we will only reproduce here the expression

$$H_0(k)=\frac{A}{\kappa G}-\frac{1}{\kappa}\sum_{k'}\frac{\Omega_{k'}(\epsilon_k+\epsilon_{k'}-2\lambda)}{\eta_{k'}(\eta_k+\eta_{k'})}. \tag{28}$$

Substitution of this result into (16) does indeed show that (23) is a solution of (16) and we conclude that (23) is the correct leading order contribution to the field for a state in the ground state band.

We now make some comments on the interpretation of this leading order solution. It should first be noted that, in this theory, λ and Δ are functions of intensive parameters only and are not subject to higher order corrections. Indeed the higher order corrections will be functions of λ and Δ which have been determined by (25) and (26) and our expansion is done at fixed λ and Δ. A closely related point is that our theory is unable to treat systems with a normal ground state. This follows from the fact that (23) is no longer valid for such states. However, our equations are valid for normal systems and it is only our inability to generate an appropriate F_0 that stops us from treating such systems.

We now turn to the solution of Eq. (17) for F_1. This solution is straightforward apart from the determination of $H_1(k)$. We determine H_1 by requiring that the field F_1 introduce no new charges into the system that are located on the real axis. That this must be so follows from the fact that (21) is an exact condition satisfied by F as well as F_0. Therefore, F_0 describes all the charges on the real axis and higher order terms can not introduce any new charges there. Once we have formulated this condition, we can apply it to any order and we have a purely algebraic prescription for calculating $F_n(z)$.

Inspection of (17) shows that $F_1(z)$ will have poles on the real axis at the zeros of $F_0(z)$ for arbitrary $H_1(k)$. This reflects the fact that to this order one can place additional charges at the places where the field vanishes. The modification of the field due to the added charges is a higher order effect. We will see that the possibility of placing charges at these points plays an important role for excited states. However, for the ground state there cannot be any charges at these points since all the free charges were assumed to form a line of charge. Therefore, we must determine $H_1(k)$ so that the right-hand side of (17) vanishes at the real zeros of F_0. The real zeros of F_0 are the roots of the equation

$$\sum_k \frac{\Omega_k}{\eta_k(x-\epsilon_k)}=0. \tag{29}$$

If there are M terms in the sum over k, then there are $M-1$ roots of (29) which we denote by $x_1,\ldots,x_{M-1}$. We then have, from the vanishing of the right-hand side of (17) at $z=x_l$,

$$\sum_k \frac{\Omega_k H_1(k)}{x_l-\epsilon_k}=\tfrac{1}{2}\sum_k\frac{\Omega_k}{(x_l-\epsilon_k)^2}-\frac{\kappa}{2}\frac{dF_0(z=x_l)}{dz}, \tag{30}$$

$$l=1,\ldots,M-1.$$

This yields $M-1$ equations for the M quantities $H_1(k)$. We obtain one more equation for H_1 from (13) which is exactly satisfied by H_0. We therefore have

$$\sum_k \Omega_k H_1(k)=0 \tag{31}$$

which is just the requirement that the field F_1 have no monopole moment.

Equations (30) and (31) completely determine $H_1(k)$. The solution of these equations and similar ones that occur in higher order equations is facilitated by making the expansion

$$H(k)=\frac{h(0)}{\eta_k}+\sum_{l=1}^{M-1}\frac{h(l)}{\eta_k(x_l-\epsilon_k)}. \tag{32}$$

Such an expansion is always possible. We note that Eq. (31) applies to H_1 and all higher terms in H and on substituting (32) we see that this implies that $h_0(0)=P$ and all higher order contributions to $h(0)$ are zero. We therefore drop the $h(0)$ term in (32) in all but leading order expressions which can be calculated directly from (28). Before turning to the solution of (31) and (32) it is useful to obtain an expression for the energy in terms of h by substituting (32) into (14),

$$E=A\lambda-(P+1)\frac{\kappa G}{2}-2\sum_{l=1}^{M-1}h(l), \tag{33}$$

where we have used the known value of $h(0)=P$ and Eqs. (25) and (26) to simplify the expression.

We now substitute the expansion (32) into (30) to solve for $h_1(l)$ with the result

$$h_1(l)=\frac{1}{2}\left(\frac{N_l}{D_l}-\zeta_l\right), \tag{34}$$

where

$$\zeta_l=|x_l-a|=[(x_l-\lambda)^2+\Delta^2]^{1/2},$$
$$N_l=\sum_k\frac{\Omega_k}{(x_l-\epsilon_k)^2},\quad D_l=\sum_k\frac{\Omega_k}{\eta_k(x_l-\epsilon_k)^2}, \tag{35}$$

and where we have used (23) and (29) to evaluate dF_0/dz at $z=x_l$. It is useful for subsequent work to have an explicit expression for $F_1(z)$. This may be obtained from (17) using the known expressions for F_0 and H_1 with the result, after some algebraic manipulation (we use Liouville's theorem to simplify the ratios of sums),

$$F_1(z)=\frac{1}{2Z}\left(\sum_k \frac{z+\epsilon_k-2\lambda}{Z+\eta_k}-\sum_l \frac{z+x_l-2\lambda}{Z+\zeta_l}-\frac{z-\lambda}{Z}\right). \quad (36)$$

Note that the sum on k has M terms and the sum on l has $M-1$ terms so that F_1 has no monopole moment. Note also that the singularities of F_1 are located on the branch cut of the square root Z and that it has simple poles at the branch points a and a^*. It therefore represents a redistribution of the free charges on the line of charge.

The first order correction to the energy can be obtained from (33) and (34),

$$E_1=-\frac{\kappa G}{2}+\sum_l(\zeta_l-N_l/D_l). \quad (37)$$

We shall see that the term involving ζ_l in (37) is just the zero-point energy of the pairing vibrational states while the term involving the ratio of two sums is approximately equal to minus this zero-point energy in the independent quasiparticle approximation. The term $-\kappa G/2$ arises from our formal treatment of κ as a free parameter independent of A and reflects the incorrect treatment of the Pauli principle that is characteristic of the BCS approximation.

The results just described can be obtained alternatively by observing that the field F_1 must arise from the free charges and not the fixed ones. Therefore, (12) implies that $H_1(k)=F_1(\epsilon_k)$ which is a result that can also be obtained from (17) by equating the residues of the pole terms at $z=\epsilon_k$. This eliminates H_1 from (17) which can now be solved for F_1.

Higher order corrections to the energy may now be obtained in the same way as the first order term. We list here the terms of order $1/A$ which will be needed in what follows

$$h_2(l)=-\frac{\kappa}{2D_l}\left[[F_1(x_l)]^2+\frac{dF_1(x_l)}{dz}\right]. \quad (38)$$

These results will be used in discussing some examples in Sec. IV.

Thus far we have an expansion of the ground state energy in which κ is treated as an arbitrary parameter. If we put in the expression for κ [Eq. (3), case (2)], then we can perform a further expansion in powers of $T(T+1)/P(P+1)$ for small T. In this way we can obtain a rotational expression for the energies of the states of the ground-state band and a $1/A$ expansion of the associated moment of inertia. Alternatively, one can keep the full κ dependence of the energy and then obtain a variable-moment of inertia[8] for these isorotational states. We exhibit the first term in this expansion

$$E_0(T)-E_0(T=0)$$
$$=8\left(\frac{1}{A}\sum_k \Omega_k[\eta_k-(\epsilon_k-\lambda)]-\frac{\Delta^2}{2G}\right)\frac{T(T+1)}{A}, \quad (39)$$

where λ and Δ are obtained from (25) and (26) with $T=0$.

III. EXCITED STATES

The procedure for treating excited states is much the same as that used in the previous section for the ground state. The one important difference is in the choice of the input field which determines the higher order terms in the expansion. In this section we treat pairing vibrational states with p pairs excited out of the condensate. After some general discussion, we give detailed results for $p=1, 2$. Throughout this section, we use primes to distinguish excited state quantities from their ground state analogs. We also treat κ as an arbitrary parameter so that our expressions are valid for isorotational bands built upon the vibrational states.

We consider an excited state with p pairs excited out of the condensate, where p is always assumed to be of order one. It is clear that the limiting charge distribution for such a state should have $P-p$ charges on a line plus p isolated charges. The field produced by such a distribution can be separated into that part which comes from the line plus external field plus that part which comes from the isolated charges, i.e.,

$$F'(z)=\bar{F}'(z)+\sum_{i=1}^{p}\frac{1}{z-e(i)}, \quad (40)$$

where $\bar{F}'$ is the field produced by the line of charge and external field and $e(i)$, $i=1,\ldots,p$, are the locations of the isolated charges. The field F' is a solution of Eq. (11), however it is useful to treat the two parts of F' differently. We determine the $e(i)$ from Eqs. (2) and then we use Eq. (11) to determine $\bar{F}'$.

We can write Eqs. (2) for the free charges in a more transparent form if we first split the sum on j into one part over the line of charge and another part over the isolated charges. The sum over the line of charges can then be represented as a contour integral around the branch cut in $\bar{F}'$ and this integral can be evaluated by expanding the contour to infinity and using the known analytic properties of F'. The result of this procedure is the set of equations

$$\bar{F}'(e(i))+\sum_{j=1}^{p}\frac{1}{e(i)-e(j)}=0, \quad i=1,\ldots,p, \quad (41)$$

which states that the isolated charges must be located at those points where the total field vanishes. The equations to be solved for the excited states are Eqs. (41) for the isolated charges and Eqs. (11) and (40) for the line of charge. We now turn to the details of the cases with $p=1, 2$.

For $p=1$, Eq. (41) becomes $\bar{F}'(e(1))=0$. If we now expand $\bar{F}'$ and $e(1)$ as power series in $1/A$, $e(1)=e_1+e_2$, we obtain

$$\bar{F}_0'(e_1)=0,$$
$$e_2=-\left[\bar{F}_1'\left(\frac{dF_0}{dz}\right)^{-1}\right]_{z=e_1}. \quad (42)$$

Since we have only moved one charge, we expect that the line of charge is unaffected to leading order, i.e.,

$$\bar{F}_0'(z)=F_0(z). \quad (43)$$

Then e_1 is one of the roots of (29) which we denote by x_m. This illustrates the general point that we can calculate the location of the isolated charges to one order higher than we know the field of the line of charges and this decouples the two aspects of the problem.

Equation (42) implies that the energy of the excited state equals that of the ground state to leading order, i.e., the excitation energies are of order one. We therefore obtain the leading contribution to the excitation energy from F_1' which we now calculate from (17). Using the above solution for e_1, $e_1 = x_m$, we write F_1' as in (40)

$$F_1'(z) = \bar{F}_1'(z) + \frac{1}{z - x_m}, \tag{44}$$

where we know that $\bar{F}_1'$ has no poles on the real axis. This last requirement leads to the system of equations, the analog of (30),

$$\sum_k \frac{\Omega_k H_1'(k)}{x_l - \epsilon_k} = \tfrac{1}{2} \sum_k \frac{\Omega_k}{(x_l - \epsilon_k)^2} - \frac{\kappa}{2}(1 + \delta_{lm}) \frac{dF_0(z = x_l)}{dz}, \tag{45}$$
$$l = 1, \ldots, M - 1.$$

Retracing the steps between (30) and (37), we obtain

$$h_1'(l) = h_1(l) - \delta_{lm}\zeta_m \tag{46}$$

and the excitation energy of the state $E_1' - E_1 = 2\zeta_m$ which is identical with the result obtained using the quasiboson[9] approximation. Thus, the quasiboson approximation gives the leading term in an expansion in powers of $1/A$. The change in the field is given by

$$F_1'(z) - F_1(z) = \zeta_m / Z(z - x_m) \tag{47}$$

which, on decomposition, shows clearly that it represents the addition of a charge at the point x_m and the subtraction of one unit of charge from the line of charge.

The first order correction to the location of the isolated charge may be calculated from (42) using (36), (42), and (47) with the result

$$e_2 = \frac{\kappa}{\zeta_m D_m} \left[\frac{x_m - \lambda}{\zeta_m^2} - F_1(x_m) \right], \tag{48}$$

where F_1 is given by (36).

We now turn to the first order correction to the excitation energy which is obtained from the difference $E_2' - E_2$. We first write

$$F_2'(z) = \bar{F}_2' + e_2/(z - x_m)^2,$$

where we have expanded the field of the fixed charge to first order in e_2 to maintain the consistency of the expansion. Substituting F_2' into (18), requiring $\bar{F}_2'$ to have no poles on the real axis, and retracing the steps that led to (38) gives

$$E_2' - E_2 = \frac{\kappa}{D_m} \left[(\bar{F}_1')^2 - (F_1)^2 + \frac{d}{dz}(3\bar{F}_1' - F_1) + e_2 \frac{d^2 F_0}{dz^2} \right]_{z = x_m} + \kappa \sum_{l \neq m} \frac{1}{D_l} \left[(\bar{F}_1')^2 - (F_1)^2 + \frac{d}{dz}(\bar{F}_1' - F_1) + \frac{2\bar{F}_1'}{x_l - x_m} \right]_{z = x_l}, \tag{49}$$

where F_1 and $\bar{F}_1'$ are given by (36), (44), and (47). This result is somewhat cumbersome due to its complete generality. We therefore defer discussion of it until Sec. IV where we consider some specific examples.

The states with $p = 2$ can be treated in much the same way as those with $p = 1$ with the one additional complication that the interaction between the two isolated charges must be treated with care. We denote the locations of the two isolated charges by $e(1)$ and $e(2)$ and expand in powers of $1/A$. Equations (41), taken to leading order, then imply that the $e_1(i)$ are two zeros of F_0 which we denote by x_m and $x_{m'}$. To proceed further, the two cases of m equal to unequal to m' must be treated separately. In the cases with $m \neq m'$, the interaction between the isolated charges is weak and the expansion proceeds as in the cases with $p = 1$. The excitation energy is the sum of the energies of the separate excitations to leading order and the first order corrections have minor modifications. We will only treat here the more interesting cases with $m = m'$ in which there is a strong interaction between the two isolated charges and the expansion in powers of $1/A$ for $e(1)$ and $e(2)$ breaks down. We will show that in these cases they must be expanded in powers of $A^{-1/2}$. However, we will also show that the expansion of the field and therefore the energy of the states in powers of $1/A$ is still valid. We shall put two primes on the field in order to distinguish it from previous cases.

We first consider the solution of Eqs. (41). In order to simplify the notation, we let $e(1) = e$ and $e(2) = e'$. Then under the assumption that $\lim_{A \to \infty} e = \lim_{A \to \infty} e' = x_m$, we find that (41) implies that the corrections to these limiting values are a power series in $A^{-1/2}$ rather than A^{-1}. Substituting power series of the form

$$e = e_1 + e_{3/2} + e_2 + \cdots$$

into (41) then yields the following results:

$$\begin{aligned} &e_1 = e_1' = x_m, \\ &e_{3/2} = -e_{3/2}' = q, \\ &e_2 = e_2' = r, \end{aligned} \tag{50}$$

where

$$q^2 = -\left[2 \frac{dF_0(x_m)}{dz} \right]^{-1},$$
$$r = 2\bar{F}_1''(x_m) q^2 + \frac{d^2 F_0(x_m)}{dz^2} q^4, \tag{51}$$

where $\bar{F}_1''$ is defined below.

We now turn to the solution of the field equations (17) and (18) under the assumption that the leading order term is unchanged by the relocation of two units of charge. We separate the field as in (40) and expand in powers of $1/A$ to obtain

$$F_0''(z) = \bar{F}_0''(z) = F_0(z), \tag{52}$$

$$F_1''(z) = \bar{F}_1''(z) + \frac{2}{z - x_m}, \tag{53}$$

$$F_2''(z) = \bar{F}_2''(z) + \frac{2r}{(z - x_m)^2} + \frac{2q^2}{(z - x_m)^3}. \tag{54}$$

Terms proportional to odd powers of $A^{-1/2}$ all cancel. The second equality in (52) is our assumption that the field is unchanged to leading order. Comparing (53) with (44) and observing that (17) is a linear equation for F_1 yields the result that the change in the field for

this $p=2$ state is given by twice the change for the $p=1$ state, Eq. (47). Furthermore, to this order, the excitation energy of the state is $4\zeta_m$, i.e., the energies of the excitations are additive. It should be noted that r in Eq. (51) is given in terms of known functions as a result of this determination of $\bar{F}_1''$.

The first correction to the excitation energy of this state is obtained by substituting (54) into (18) and requiring that $\bar{F}_2''$ be regular on the real axis. This yields the expression

$$E_2''-E_2=-2\sum_l[h_2''(l)-h_2(l)]$$
$$=\frac{\kappa}{D_m}\left[(\bar{F}_1'')^2-(F_1)^2+\frac{d}{dz}(5\bar{F}_1''-F_1)+2\frac{d^2F_0}{dz^2}r\right.$$
$$\left.+\frac{2}{3}\frac{d^3F_0}{dz^3}q^2\right]_{z=x_m}+\kappa\sum_{l\neq m}\frac{1}{D_l}\left[(\bar{F}_1'')^2-(F_1)^2\right.$$
$$\left.+\frac{d}{dz}(\bar{F}_1''-F_1)+\frac{4\bar{F}_1''}{x_l-x_m}+\frac{2}{(x_l-x_m)^2}\right]_{z=x_l}. \quad (55)$$

Again this leads to a rather cumbersome expression for the energy due to its complete generality. We shall discuss it in the context of specific models in the next section.

IV. EXAMPLES

In this section we present some analytical and numerical results for simple examples of the expansion developed in the preceding sections. These results are presented as an aid to the understanding of the expansion and one should not draw the conclusion that our formalism is limited to such simple examples. Indeed, our formalism is quite general and the most difficult step in applying it is the first one, i.e., the solution of the BCS equations, which must be done numerically for any realistic model. Our examples are therefore dictated by the requirement that the single-particle spectrum be such that the BCS equations are easily solved. This leads us to the model in which the single-particle energies take on two values with the total degeneracy of each level being equal to the number of particles. We first treat this model for identical nucleons in which $\kappa=2$ and give analytical results for the terms in our expansion. We then go on to discuss the cases in which $\kappa<2$ for which we give numerical results. We conclude this section with a brief comment on the continuum model which is the limit in which the single-particle spectrum becomes continuous.

We first consider identical nucleons, $\kappa=2$, in a system with a two-level single-particle spectrum. We denote the two levels by $k=\pm$ and choose $\epsilon_\pm=\pm\epsilon$ and $\Omega_\pm=A/2$. The BCS equations (25) and (26) are easily solved for such a system with the results

$$\lambda=0, \quad \Delta=(G^2-\epsilon^2)^{1/2}, \quad G>\epsilon, \quad (56)$$

which imply $\eta_\pm=\eta=G$. The simplicity of this model is somewhat offset by its artificiality which leads to the seeming contradiction that for $G<\epsilon$ there exist solutions to the equations for two more or less particles than A but not for the precise number chosen. That is, there is a discontinuous behavior of the leading term in the expansion as we go from $A-2$ to A to $A+2$ particles which is in apparent conflict with our assumption of an expansion in powers of $1/A$. However, the expansion remains valid as long we start with the proper leading term *and* the system has well developed superfluidity. Therefore, we can treat the systems with $A\pm2$ particles for all $G>0$ but the system with A particles only for $G>\epsilon$ and, for $G>\epsilon$, there is no discontinuous behavior. Since the systems with $A\pm2$ particles are particle-hole images of each other, they are in this sense identical and the discontinuity for $G<\epsilon$ exists only in the limit of a self conjugate system. In what follows, we will assume that $G>\epsilon$ and that Δ is of order one.

The leading order term for the ground state energy is obtained by substituting (56) into (27),

$$E_0=-\frac{A}{2G}(\epsilon^2+G^2), \quad (57)$$

from which we can obtain the usual BCS occupation probabilities by differentiation with respect to ϵ,

$$n_0(\pm)=\frac{A}{2}\left(1\mp\frac{\epsilon}{G}\right). \quad (58)$$

The leading order field is obtained from (23),

$$F_0(z)=-\frac{AZ}{2G}\left(\frac{z}{z^2-\epsilon^2}\right), \quad (59)$$

where $Z=(z^2+\Delta^2)^{1/2}$. From (59) we see that F_0 has one real zero at $z=0$ and therefore $x_1=0$ and $\zeta_1=\Delta$, see (35). We also obtain from (35) that $N_1/D_1=\eta$ and the first order correction to the energy is, from (37),

$$E_1=-G+\zeta-\eta=\Delta-2G. \quad (60)$$

The term $-G$ comes from the usual Pauli principle correction to the BCS results and the term $\zeta-\eta$ is the difference between the zero-point energies of the pairing vibrational mode as calculated in the quasiboson and quasiparticle approximations. Differentiating (60) with respect to ϵ gives the first-order correction to the occupation probabilities

$$n_1(\pm)=\mp\epsilon/2\Delta. \quad (61)$$

The first-order field is obtained from (36),

$$F_1(z)=\frac{z}{Z}\left(\frac{1}{Z+G}-\frac{1}{2(Z+\Delta)}-\frac{1}{2Z}\right), \quad (62)$$

which can be used to evaluate the second-order correction to the energy,

$$E_2=-\frac{G(3G-\Delta)(G-\Delta)}{2A\Delta^2} \quad (63)$$

and to the occupation probabilities

$$n_2(\pm)=\mp\frac{\epsilon G^2(3G-2\Delta)}{2A\Delta^4}. \quad (64)$$

It should be noted that these expansions have the appearance of being expansions in $1/(A\Delta)$ rather than just $1/A$. This has to do with the singular nature of the perturbation theory on Eq. (11) and we shall discuss this in the next section. Note also that the "Fermi surface" is tightened up by the higher order terms, i.e., the occupation of the upper (lower) level is lowered (raised) by the correction terms. This feature has been noted in numerical calculations.[5]

For the excited state with $p=1$, we have from (46)–(50),

$$E'-E=2\Delta-(2G/A\Delta^2)(1+2G^2-2G\Delta) \tag{65}$$

and for the states with $p=2$, from (51)–(55),

$$E''-E=4\Delta-(2G/A\Delta^2)(3+6G^2-4G\Delta). \tag{66}$$

The deviation from additivity of the energies of the separate excitations is given by

$$E''-E-2(E'-E)=-(2G/A\Delta^2)(2G^2+1).$$

Thus there is a weak attractive interaction between the excitations.

As a second example, we consider a system with the same single-particle spectrum as above but with both neutrons and protons present. The total isospin for these states takes on the values $T=P, P-2, \ldots, 0$ or 1 and the corresponding range for κ is $2\geq\kappa\geq(P-3)/(P-1)\approx 1$, see (3). In this case we can not solve the BCS equations analytically. However, we can reduce the problem to the solution of a simple algebraic equation for the zero of the leading order field and express all quantities in terms of this root. We include some numerical results for typical values of the parameters in Tables I–VI. Before proceeding, it should be pointed out that the BCS equations (25) and (26) have the formal appearance of those describing $\kappa A/2$ identical nucleons in the same potential. Therefore, as long as $\kappa<2$, which we will assume in what follows, there is no discontinuous behavior for $G>0$.

We first consider the solution of the leading order problem, Eqs. (20)–(30). The BCS equations for $\kappa<2$ can be written as (we set $\epsilon=1$)

$$\frac{1}{\eta_-}+\frac{1}{\eta_+}=\frac{2}{G}, \tag{67}$$

$$\frac{1}{\eta_-}-\frac{1}{\eta_+}=\kappa-2-\frac{2\lambda}{G}, \tag{68}$$

where $\eta_\pm=[(1\mp\lambda)^2+\Delta^2]^{1/2}$. Before considering the solution of these equations we turn to the field F_0 which we write as

$$F_0(z)=-\frac{AZ(z-x)}{\kappa G(z^2-1)}, \tag{69}$$

where we have one real zero at $z=x_1\equiv x$, where

$$x=\frac{G}{2}\left(\frac{1}{\eta_-}-\frac{1}{\eta_+}\right)=-\lambda-\left(1-\frac{\kappa}{2}\right)G, \tag{70}$$

where we have used (68). Solving (67), (68), and (70) gives

$$\eta_\pm=G/(1\mp x). \tag{71}$$

TABLE I. Energies of the $T=0$, $p=0$ states. The energies are given relative to their values at $G=0$.

A	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	0	−1.870(5.2)	−4.184(4.5)	−9.817(3.8)
	1	−1.962(0.5)	−4.374(0.1)	−10.189(0.1)
	2	−1.962(0.5)	−4.374(0.1)	−10.188(0.1)
	Exact	−1.972	−4.379	−10.204
16	0	−2.905(6.8)	−6.612(6.0)	−16.216(4.8)
	1	−3.103(0.5)	−7.027(0.1)	−17.020(0.1)
	2	−3.103(0.5)	−7.026(0.1)	−17.020(0.1)
	Exact	−3.118	−7.037	−17.036
32	0	−5.156(4.6)	−11.810(4.2)	−29.662(3.2)
	1	−5.397(0.1)	−12.319(0.1)	−30.642(0.1)
	2	−5.397(0.1)	−12.318(0.1)	−30.642(0.1)
	Exact	−5.404	−12.325	−30.661

TABLE II. Excitation energies of the $T=0$, $p=1$, 2, states. The energies are given relative to the $T=0$, $p=0$ states.

A	p	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	1	1	3.946(1.1)	4.150(2.1)	5.245(1.9)
		2	3.992(0.0)	4.235(0.1)	5.342(0.0)
		Exact	3.992	4.238	5.344
	2	1	7.893(2.1)	8.300(3.4)	10.489(2.6)
		2	8.056(0.1)	8.568(0.3)	10.712(0.6)
		Exact	8.062	8.591	10.771
16	1	1	3.731(1.2)	3.739(2.7)	4.752(2.0)
		2	3.785(0.0)	3.844(0.1)	4.846(0.1)
		Exact	3.784	3.842	4.850
	2	1	7.462.(2.6)	7.478(4.5)	9.504(2.7)
		2	7.652(0.1)	7.805(0.3)	9.695(0.7)
		Exact	7.660	7.831	9.768
32	1	1	3.640(0.9)	3.557(1.8)	4.550(1.2)
		2	3.674(0.0)	3.625(0.0)	4.603(0.0)
		Exact	3.674	3.624	4.605
	2	1	7.279(1.7)	7.114(3.1)	9.101(1.4)
		2	7.400(0.1)	7.325(0.3)	9.198(0.4)
		Exact	7.406	7.345	9.234

TABLE III. Energies of the $T=2$, $p=0$ states. The energies are given relative to their value at $G=0$.

A	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	0	−1.323(15)	−3.028(14)	−7.568(11)
	1	−1.555(0.3)	−3.516(0.1)	−8.509(0.1)
	2	−1.554(0.3)	−3.513(0.1)	−8.509(0.1)
	Exact	−1.559	−3.518	−8.513
16	0	−2.668(8.3)	−6.120(7.3)	−15.252(5.7)
	1	−2.897(0.5)	−6.600(0.0)	−16.179(0.0)
	2	−2.897(0.5)	−6.599(0.0)	−16.179(0.0)
	Exact	−2.910	−6.601	−16.182
32	0	−5.041(4.9)	−11.574(4.4)	−29.222(3.3)
	1	−5.290(0.2)	−12.099(0.0)	−30.230(0.0)
	2	−5.290(0.2)	−12.098(0.0)	−30.230(0.0)
	Exact	−5.300	−12.104	−30.231

We obtain an equation for x by using (70), (71), and the relation $4\lambda = \eta_-^2 - \eta_+^2$,

$$x = -\left(1 - \frac{\kappa}{2}\right)G + \frac{G^2 x}{(1-x^2)^2}. \tag{72}$$

The singular nature of the limit $\kappa \to 2$ is clearly evident. Equation (72) may be simplified by regarding it as a quadratic equation for G and then obtaining G as a function of x,

$$G = \frac{(1-x^2)^2}{2x}\left\{\left(1-\frac{\kappa}{2}\right) + \left[\left(1-\frac{\kappa}{2}\right)^2 + \frac{4x^2}{(1-x^2)^2}\right]^{1/2}\right\}. \tag{73}$$

The right-hand side of (73) diverges like $(1-\kappa/2)/x$ as $x \to 0$, goes to zero like $2(1-x)$ as $x \to 1$, and behaves smoothly between these limits. Therefore, for any positive value of G there is a root of (73) in the interval $0 < x < 1$. The energy gap may also be obtained in terms of x by using $\Delta^2 = \frac{1}{2}(\eta_-^2 + \eta_+^2) - (1+\lambda^2)$,

$$\Delta^2 = \left(1-\frac{\kappa}{2}\right)G\left[\frac{1}{x} - x - \left(1-\frac{\kappa}{2}\right)G\right]. \tag{74}$$

The energy, to leading order, is then given by

$$E_0 = -(A/\kappa G)[2 + 2\lambda x + \Delta^2] \tag{75}$$

which can be evaluated using the above expressions for λ and Δ^2.

The first-order correction to the ground state energy is given by

$$E_1 = -\frac{\kappa G}{2} + \zeta - G\frac{(1+x^2)}{(1-x^2)}, \tag{76}$$

where $\zeta = [(x-\lambda)^2 + \Delta^2]^{1/2}$. The second-order correction can be obtained from (38) which, while easily evaluated numerically, does not lead to a simple algebraic expression. The same is true for the corrections to the excitation energies and we will not record the expressions here.

TABLE IV. Excitation energies of the $T=2$, $p=1$, 2 states. The energies are given relative to the $T=2$, $p=0$ states.

A	p	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	1	1	3.660(3.3)	3.598(6.4)	4.594(4.8)
		2	3.791(0.2)	3.856(0.3)	4.800(0.6)
		Exact	3.786	3.843	4.828
	2	1	7.320(5.9)	7.196(10)	9.189(6.8)
		2	7.779(0.0)	7.998(0.0)	9.581(2.8)
		Exact	7.781	8.028	9.863
16	1	1	3.667(1.7)	3.612(3.3)	4.610(2.2)
		2	3.732(0.1)	3.738(0.1)	4.713(0.1)
		Exact	3.730	3.735	4.717
	2	1	7.334(3.1)	7.224(5.5)	9.221(3.0)
		2	7.560(0.1)	7.618(0.4)	9.417(0.9)
		Exact	7.567	7.647	9.507
32	1	1	3.624(1.0)	3.526(1.9)	4.517(1.2)
		2	3.660(0.0)	3.597(0.0)	4.570(0.1)
		Exact	3.660	3.596	4.573
	2	1	7.249(1.8)	7.051(3.3)	9.035(1.4)
		2	7.374(0.1)	7.272(0.3)	9.131(0.4)
		Exact	7.380	7.292	9.168

TABLE V. Energies of the $T=4$, $p=0$ states. The energies are given relative to their value at $G=0$.

A	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	0	0.0	0.0	−2.000(54)
	1	–	–	−4.268(1.6)
	2	–	–	−4.316(0.5)
	Exact	−0.577	−1.395	−4.338
16	0	−2.119(13)	−4.926(11)	−12.942(8.6)
	1	−2.420(0.2)	−5.566(0.1)	−14.160(0.1)
	2	−2.420(0.2)	−5.563(0.1)	−14.161(0.1)
	Exact	−2.423	−5.570	−14.169
32	0	−4.774(5.6)	−11.025(4.8)	−28.127(3.8)
	1	−5.040(0.3)	−11.586(0.0)	−29.204(0.1)
	2	−5.040(0.3)	−11.585(0.0)	−29.204(0.1)
	Exact	−5.056	−11.588	−29.225

Numerical results are given in Tables I–VI for the following values of the parameters: $A = 8, 16, 32$; $G = 0.5, 1.0, 2.0$; $T = 0, 2, 4$; $p = 0, 1, 2$. All energies are given in units with $\epsilon = 1$ and the energies of the states with $p = 0$ are given with respect to the energy of the corresponding state of the noninteracting system, $-A$. The energies of the states with $p = 1, 2$ are given as excitation energies relative to the $p = 0$ state with the same value of T. In these tables, we compare the energies of these states, as calculated by our expansion, with the exact energy. We give both the numerical value of the energies and the percentage error, in parentheses, of the various terms. We see from these tables that, with one exception ($A = 8$, $G = 2$, $T = 4$, $p = 2$ in Table VI), the convergence of our expansion is very rapid with the first two terms being essentially exact. The one exception is in a case where the convergence is not expected to be fast and this one case out of 81 confirms this expectation. The fluctuations in the rates of convergence of the expansions for the energies of the other states are due to the particular values of the parameters used.

We conclude this section with some brief comments on the continuum limit of our equations. For this limit, we consider a system whose single-particle spectrum has only spin and isospin degeneracy, i.e., $\Omega_k = 1$, and whose levels are distributed evenly on the interval $(-1, 1)$, i.e., $\epsilon_k = k/P$, $k = 0, \pm 1, \ldots, \pm P$. We then replace sums over k by integrals over ϵ_k

$$\sum_k \to P\int_{-1}^{+1} d\epsilon_k. \tag{77}$$

The BCS equations for A particles in this system can then be solved for λ and Δ as functions of G and κ with the results

$$\lambda = -\left(1-\frac{\kappa}{2}\right)\coth\frac{1}{G}, \tag{78}$$

TABLE VI. Excitation energies of the $T=4$, $p=1$, 2 states. The energies are given relative to the $T=4$, $p=0$ states.

A	p	Order	$G=0.5$	$G=1.0$	$G=2.0$
8	1	1	4.0	4.0	3.464(7.2)
		2	–	–	3.149(2.5)
		Exact	3.256	2.679	3.230
	2	1	8.0	8.0	6.928(2.4)
		2	–	–	4.738(33)
		Exact	7.094	6.579	7.100
16	1	1	3.512(2.4)	3.290(5.2)	4.280(2.8)
		2	3.603(0.1)	3.480(0.2)	4.385(0.4)
		Exact	3.600	3.472	4.403
	2	1	7.025(4.4)	6.581(8.5)	8.560(3.5)
		2	7.341(0.1)	7.164(0.4)	8.713(1.8)
		Exact	7.347	7.195	8.872
32	1	1	3.589(1.0)	3.452(2.2)	4.440(1.3)
		2	3.627(0.0)	3.530(0.0)	4.494(0.1)
		Exact	3.627	3.528	4.497
	2	1	7.178(1.9)	6.903(3.7)	8.880(1.5)
		2	7.313(0.1)	7.145(0.3)	8.972(0.4)
		Exact	7.319	7.167	9.013

$$\Delta^2=\left(\frac{\kappa}{2}\right)^2\coth^2\frac{1}{G}+2\left(1-\frac{\kappa}{2}\right)\coth\frac{1}{G}-\frac{\kappa}{2}\left(2-\frac{\kappa}{2}\right),$$

and the leading contribution to the energy is

$$E_0=-A\left(1-\frac{\kappa}{4}\right)\coth\frac{1}{G} \tag{79}$$

which yields a pure rotational spectrum.

Corrections to (79) will come from the higher order terms in our expansion plus corrections to the approximate evaluation of sums in (77). These two types of correction get mixed up as can be seen from the evaluation of the first-order correction to E_0. The roots x_l of (29) lie between the values of ϵ_k and therefore the values of x_l are the same as those of ϵ_k, excluding $k=-P$, up to terms of order $1/P$. Furthermore, (29) has A distinct roots x_l. Both N_l and D_l of Eq. (35) will be of order P^2 and their ratio is $N_l/D_l=\eta_k$ to leading order, where k is the value such that $x_l\approx\epsilon_k$. Substituting this into (37), $\zeta_l-N_l/D_l=\zeta_l-\eta_k$, gives a contribution that seems to be of order A to E_1. However, to leading order, $\zeta_l=\eta_k$, canceling the leading order term and the sum over l in (37) is of order-one preserving the consistency of our expansion. However the evaluation of the term E_1 requires a careful treatment of the sums as well as terms that arise from corrections to the use of (77) in the evaluation of E_0. We will not pursue these issues any further here.

V. CONCLUSION

We have developed an expansion, in inverse powers of the number of particles, for the energy of a many-fermion system interacting through pairing forces. We have used an electrostatic analogy to the basic equations and our expansion starts with an exact equation for the electrostatic field which is valid for systems with normal or superfluid ground states. However, our expansion is initiated by inserting a leading term that is only appropriate for systems with superfluid ground states and our results are limited by this constraint. Our results are therefore not valid for systems with zero energy gap Δ and the convergence of the expansion becomes slow as Δ becomes small. The origin of this limitation is the singular nature of the perturbation in Eq. (11). In solving this equation, we have treated dF/dz as one of the terms in the perturbation which is assumed to be small compared to F^2. This is not true in the neighborhood of the branch points of F_0. As $\Delta\to 0$ the branch points approach the real axis and the values of F on the real axis determine the energy of the system. Thus, the break down of the expansion can be understood in terms of the approach to the real axis of the branch points of F_0. One can solve the equations near the branch points by using the stretched coordinate techniques of hydrodynamic boundary layer theory. This leads to the replacement of the branch point singularities by simple poles as must be the case from the definition of F. We have not pursued these ideas to develop an expansion that is appropriate for systems with normal ground states although the theory is ripe for such a development.

For systems with superfluid ground states, we have shown in examples that our expansion converges more rapidly than one would expect *a priori* with the first two terms being almost exact for $A>8$. We expect the convergence of our expansion to be equally rapid for systems with more complicated single-particle spectra.

Our expansion shows that the collective excitations are of a rotational and/or vibrational character and explicit expressions for the energies of these excitations and their interactions may be obtained by a further expansion in powers of $T(T+1)/P(P-1)$ and p/P. We have reserved the discussion of states with single-particle excitations for a future publication.

[1]R.W. Richardson and N. Sherman, Nucl. Phys. 52, 221 (1964).
[2]R.W. Richardson, Phys. Rev. **144**, 874 (1966).
[3]R.W. Richardson, Phys. Rev. **159**, 792 (1967).
[4]S.T. Belyaev, *Proceedings of the International School of Physics LIII* (Academic, New York, 1973), p. 60 and the references cited therein.
[5]R.W. Richardson, Phys. Rev. **141**, 949 (1966); J. Bang and J. Krumlinde, Nucl. Phys. A **141**, 18 (1970).
[6]J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. 108, 1175 (1957).
[7]M. Gaudin, unpublished.
[8]M.A.J. Mariscotti, G. Scharff-Goldhaber, and B. Buck, Phys. Rev. **178**, 1864 (1969).
[9]J. Hogaasen-Feldman, Nucl. Phys. 28, 258 (1961).

Twistors and induced representations of SU(2,2) *

David E. Lerner

University of Kansas, Lawrence, Kansas 66045
(Received 18 October 1976)

We give an explicit realization of a series of representations of SU(2,2) induced by $\mathbf{R}^+ \otimes SL(2,\mathbf{C})$. Vectors in these representation spaces are homogeneous spinor-valued functions of two twistor variables. They may also be realized, in a frame-dependent way, either as conformally invariant fields in Minkowski space or as homogeneous spinor-valued functions on the O(2,4) null cone. The conformal invariance of the massless free fields is discussed from this point of view, and the twistor version of the field equations is derived. Finally, irreducible twistors are shown to correspond to conformally invariant fields satisfying the generalized twistor equation.

1. INTRODUCTION

Among the most important representations of the conformal group of space–time are those which give rise to physical fields in Minkowski space. These fields can be constructed by the method of induced representations; they are derived by means of a certain projection operation from the cross sections of homogeneous SU(2,2) vector bundles. In suitably chosen bases, these cross sections assume the familiar form of homogeneous spinor-valued functions on the O(2,4) null cone; the actual physical fields are obtained by projecting (or restricting) these functions to Minkowski space.[1–4]

An alternative realization of some of these fields has been obtained by Penrose,[5–7] who shows that to each set $\{\varphi_r: 0 \leq r \leq n\}$ of holomorphic functions of two twistor variables U and V satisfying

$$U^\alpha \frac{\partial \varphi_r}{\partial U^\alpha} = -(r+1)\varphi_r, \quad V^\alpha \frac{\partial \varphi_r}{\partial V^\alpha} = -(n-r+1)\varphi_r, \quad (1.1)$$

and

$$\frac{\partial \varphi_r}{\partial U^\alpha} + \frac{\partial \varphi_{r+1}}{\partial V^\alpha} = 0, \quad (1.2)$$

there corresponds a unique real analytic solution to the zero rest-mass free field equations[8]

$$\nabla^{AA'} \varphi_{A'B'\cdots C'} = 0 \quad (1.3)$$

in Minkowski space. Conversely, it can be shown (Sec. 5, see also Ref. 9 for an altogether different approach) that any real analytic solution to (1.3) may be described by holomorphic twistor functions.

More generally, it turns out that *any* representation of of SU(2,2) induced by $\mathbf{R}^+ \otimes SL(2,\mathbf{C})$ may be realized on a space of vector-valued functions of two twistor variables (Secs. 3 and 4). If the representation can be analytically continued to the complex O(2,4) null cone, then the real analytic fields determined by the representation will correspond to holomorphic twistor functions. It is a decided advantage of this formalism that the twistor functions, independent of their analyticity properties, transform as scalars under SU(2,2). The reason for this, as might be expected, is that the pair (U,V) is actually a generalized "frame" in an appropriate principal bundle, and the main purpose of this paper is to examine this fact in some detail. As a by-product, we are able to give a unified treatment of massless fields, conformally weighted densities, twistor functions, and homogeneous functions on the O(2,4) null cone, objects which often appear in somewhat different settings.

Some of the representations considered here are quite well known in one form or another, and certain of their properties have been considered at length in the literature. The Lie algebra version of much of this, together with some applications to quantum field theory, may be found in Ferrara *et al.*[3] Invariant scalar products are discussed by Gross[10] and Penrose.[5,7] These subjects will not be considered here; in particular, we do not discuss the topology of the representations.

The first part of the paper is concerned with establishing the existence of a principal bundle homorphism from $SU(2,2) \to \mathcal{N}$ onto the set of twistor dyads over $\mathcal{N}$, $\mathcal{N}$ being the O(2,4) null cone. It then follows directly that the representations above referred to can all be realized on spaces of twistor functions, different representations being characterized by the homogeneity properties of their respective twistor functions. Those representations which can be analytically continued to the complex O(2,4) null cone are labeled by an ordered pair of integers (n_1, n_2).

We then show that the massless free fields of helicity $n/2$ belong to an invariant subspace of the $(n,1)$ representation, and that, in the real analytic case, (1.3) implies (1.1) and (1.2). As a simple application of the formalism introduced, we conclude by showing the direct connection between finite-dimensional representations of SU(2,2) and solutions to the generalized twistor equation.

2. PRELIMINARIES

Notation and conventions:

$\mathbf{M}^4$: Minkowski space; $\eta_{\mu\nu} = \mathrm{diag}\{1,-1,-1,-1\}$

$\mathcal{R}$: $\mathbf{R}^6$ together with the quadratic form $h_{ab} = \mathrm{diag}\{1,-1,-1,-1,-1,1\}$

$\mathcal{N}$: The null cone (minus the vertex) of $\mathcal{R}$

$\tilde{\mathcal{M}}$: The space of rays of $\mathcal{N}$

$\mathcal{M}$: The space of lines of $\mathcal{N}$, the conformal compactification of $\mathbf{M}^4$

$\mathcal{T}$: Twistor space; $\mathbf{C}^4$ together with the Hermitian form $Q = \mathrm{diag}\{1,1,-1,-1\}$

$\mathbb{C}\mathcal{N}$: The complex O(2,4) null cone (minus the vertex), defined by the zeroes of h in $\mathbf{C}^6$

Points of $\mathbf{M}^4$ will be denoted either by x^μ or $x^{AB'}$, where[8]

$$(x^{AB'}) = \begin{pmatrix} x^{00'} & x^{01'} \\ x^{10'} & x^{11'} \end{pmatrix} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} x^0 + x^1 & x^2 + ix^3 \\ x^2 - ix^3 & x^0 - x^1 \end{pmatrix}.$$

The mapping σ: $\mathbf{M}^4 \to \mathcal{N}$ given by

$$\sigma(x^\mu) = y^a(x) \equiv (x^\mu, \tfrac{1}{2}[1 + x_\nu x^\nu], \tfrac{1}{2}[1 - x_\nu x^\nu]) \tag{2.1}$$

embeds Minkowski space onto the section of $\mathcal{N}$ given by the null hyperplane $y^4 + y^5 = 1$. A useful local coordinate system in $\mathcal{N}$ is obtained by letting x^μ label the generator containing $\sigma(x^\mu)$ and defining a fifth coordinate

$$\kappa = y^4 + y^5, \quad \kappa > 0. \tag{2.2}$$

Any *slice* of $\mathcal{N}$ (a four-dimensional submanifold nowhere tangent to the generators) may be given locally by specifying κ as a function of the x^μ. Using the x^μ as local coordinates in the slice, the induced metric is given by

$$g_{\mu\nu}(x) = \kappa^2(x)\eta_{\mu\nu}. \tag{2.3}$$

Thus all conformal rescalings of Minkowski space[6,8] may be realized as slices of $\mathcal{N}$; in particular, taking the slice $\kappa = 1$, it follows that σ is an isometric embedding, and that $x^\mu \in \mathbf{M}^4$ can be identified with $(x^\mu, 1) \in \mathcal{N}$.

We may also obviously regard σ as mapping complex Minkowski space onto the $y^4 + y^5 = 1$ slice of $\mathbb{C}\mathcal{N}$, and a local coordinate system in $\mathbb{C}\mathcal{N}$ is obtained by allowing the x^μ to take on arbitrary complex values and κ to range over the complex plane minus the negative real axis.

As usual,[6,7] we choose a basis in $\mathcal{T}$ in which Q takes the form

$$Q = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}. \tag{2.4}$$

Let $B \in \mathcal{T} \wedge \mathcal{T}$, and define $y(B) \in \mathbf{C}^6$ by

$$\begin{aligned} y^0(B) &= \frac{-i}{\sqrt{2}}(B^{03} - B^{12}), \quad y^3(B) = \frac{1}{\sqrt{2}}(B^{13} + B^{02}), \\ y^1(B) &= \frac{-i}{\sqrt{2}}(B^{03} + B^{12}), \quad y^4(B) = \tfrac{1}{2}(B^{23} - 2B^{01}), \\ y^2(B) &= \frac{-i}{\sqrt{2}}(B^{13} - B^{02}), \quad y^5(B) = \tfrac{1}{2}(B^{23} + 2B^{01}). \end{aligned} \tag{2.5}$$

This establishes a one-to-one correspondence between simple twistor bivectors and points of $\mathbb{C}\mathcal{N}$. Notice that $y(B)$ is real (lies on $\mathcal{N}$) only when B is real; i.e., only when $\overline{B}_{\alpha\beta} = B_{\alpha\beta} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta}B^{\gamma\delta}$. Two twistors are said to meet at $x = x^{AA'}$ if they can be written in the form $U^\alpha = (ix^{AA'}\pi_{A'}, \pi_{A'})$, $V^\alpha = (ix^{AA'}\omega_{A'}, \omega_{A'})$. (see Refs. 11 and 12 for the geometrical implications.) In this case, if $B = 2U \wedge V$,

$$y(B) = (\pi_{A'}, \omega^{A'})\sigma(x). \tag{2.6}$$

Note, in particular, that U and V may meet at a real point without $2U \wedge V$ determining a point of $\mathcal{N}$.

If $y \to B(y)$ is the inverse of (2.5) and $B(y)$ is regarded as a skew matrix, the action of SU(2,2) on $\mathcal{N}$ is given by

$$T(y) = y(TB(y)T^t), \quad T \in \mathrm{SU}(2,2). \tag{2.7}$$

The subgroup of SU(2,2) leaving $\sigma(0)$ fixed is given by

$$\mathcal{W} \equiv \left\{ \begin{bmatrix} \alpha & 0 \\ iA\alpha & \alpha^{*-1} \end{bmatrix} : \alpha \in \mathrm{SL}(2,\mathbb{C}),\ A \text{ Hermitian} \right\}. \tag{2.8}$$

Clearly, $\mathcal{W}$ is the semidirect product $\mathrm{SL}(2,C) \cdot \mathfrak{A}$, where $\mathfrak{A}$ is the four-parameter Abelian group of special conformal transformations.

Finally, if $[y]$ denotes the equivalence class of $y \in \mathcal{N}$ under the relation $y \sim ry$, $r \in \mathbf{R}^+$, the action of SU(2,2) on $\tilde{\mathcal{M}}$ is given by

$$T([y]) = [T(y)], \tag{2.9}$$

where y is any representative of $[y]$. The subgroup of SU(2,2) fixing $\sigma(0)$ is $\mathcal{V} = \{\mathbf{R}^+ \otimes \mathrm{SL}(2,\mathbf{C})\} \circ \mathfrak{A}$.

3. REPRESENTATIONS

We are going to be concerned with representations of SU(2,2) induced by representations of either $\mathcal{W}$ or $\mathcal{V}$ with the property that $\mathfrak{A}$ is in the kernel of the inducing representation. Vectors in these representation spaces, regarded as functions on SU(2,2), will thus be constant on the cosets of $\mathfrak{A}$. It turns out that a particularly useful parametrization of these cosets is provided by pairs of twistors (U, V) satisfying $y(U \wedge V) \in \mathcal{N}$.

To see this explicitly, regard a matrix $T \in \mathrm{SU}(2,2)$ as being composed of four twistors

$$T = [E \mid F \mid G \mid H] \tag{3.1}$$

subject to $\overline{E}_\alpha G^\alpha = \overline{F}_\alpha H^\alpha = 1$, all other products vanishing, and $E^{[\alpha}F^\beta G^\gamma H^{\delta]} = \epsilon^{\alpha\beta\gamma\delta}$. In the principal bundle $\mathrm{SU}(2,2) \to \mathrm{SU}(2,2)/\mathcal{W} = \mathcal{N}$, T lies in the fiber over

$$y = y(TB(\sigma(0))T^t). \tag{3.2}$$

Since $TB(\sigma(0))T^t = 2G \wedge H$, the third and fourth columns of T determine the fiber in which it lies. Moreover, for

$$\mathbf{A} = \begin{bmatrix} I & 0 \\ iA & I \end{bmatrix} \in \mathfrak{A},$$

$T\mathbf{A}$ has the form $[* \mid * \mid G \mid H]$, so that these same columns are invariant under $\mathfrak{A}$ and may be thought of as labeling the elements of $\mathrm{SU}(2,2)/\mathfrak{A}$.

We formalize this as follows. Put

$$\mathcal{S}^* \equiv \{(U, V) \in \mathcal{T} \times \mathcal{T} : U \neq \lambda V,\ \lambda \in \mathbb{C}\}. \tag{3.3}$$

Define a projection P: $\mathcal{S}^* \to \mathbb{C}\mathcal{N}$ by

$$P(U, V) = y(2U \wedge V) \tag{3.4}$$

and let

$$\mathcal{S} = P^{-1}(\mathcal{N}). \tag{3.5}$$

Notice that for any $y \in \mathbb{C}\mathcal{N}$, $P^{-1}(y) \tilde{=} \mathrm{SL}(2,\mathbf{C})$ by virtue of (3.4). It will be convenient to label U and V by spinor indices,

$$U_{A'} = (U_{0'}, U_{1'}) = (U, V). \tag{3.6}$$

A right action of $\mathrm{SL}(2,\mathbf{C})$ on $\mathcal{S}^*$ is given by

$$U_{A'} \cdot \alpha \equiv \bar{\alpha}_{A'}{}^{B'} U_{B'}, \tag{3.7}$$

where $\bar{\alpha}_{A'}{}^{B'} = \overline{\alpha_A{}^B}$ and $\mathcal{S}^* \to \mathbb{C}\mathcal{N}$ (resp. $\mathcal{S} \to \mathcal{N}$) becomes a principal SL(2,C) bundle. SU(2,2) acts on $\mathcal{S}^*$ (resp. $\mathcal{S}$) in the obvious way,

$$T \cdot U_{A'} = T_\beta^\alpha U_{A'}^\beta . \tag{3.8}$$

Let $\tilde{\rho}$: $(\alpha, A) \to \alpha$ be the homorphism from $\mathcal{W}$ onto SL(2,C), and define a mapping ρ: SU(2,2) $\to \mathcal{S}$ by

$$\rho : [E \mid F \mid G \mid H] \to (H, -G). \tag{3.9}$$

Evidently, ρ is fiber preserving, and a simple calculation shows that

$$\rho(RTW) = R \circ \rho(T) \cdot \tilde{\rho}(W), \quad R, T \in \mathrm{SU}(2,2), \quad W \in \mathcal{W}. \tag{3.10}$$

Thus, ρ is a principal bundle homorphism which commutes with the action of SU(2,2).

We come now to the induced representations. Let τ: SL(2,C) $\to$ aut(V) be a representation of SL(2,**C**) on the finite-dimensional vector space V. τ may be regarded as a representation of $\mathcal{W}$ as well, by putting $\tilde{\tau}(\alpha, A) = \tau(\alpha)$. It is an immediate consequence of (3.10) that the two associated homogeneous vector bundles SU(2,2) $\times_{\tilde{\tau}} V \to \mathcal{N}$ and $\mathcal{S} \times_\tau V \to \mathcal{N}$ are *isomorphic*: thus the representation of SU(2,2) induced by τ can be realized on the vector space of cross sections of $\mathcal{S} \times_\tau V \to \mathcal{N}$. (It is not required that τ be unitary.)

Recall that an element of $\mathcal{S} \times_\tau V$ is an equivalence class $\{(U_{A'}, v)\}$, with $U_{A'} \in \mathcal{S}$, $v \in V$, where the equivalence relation is defined by

$$(U_{A'}, v) \sim (U_{A'}, v) \circ \alpha = (U_{B'} \bar{\alpha}^{B'}{}_{A'}, \tau^{-1}(\alpha) \cdot v). \tag{3.11}$$

Let Φ be a cross section of $\mathcal{S} \times_\tau V$, and let $U_{A'} \in \mathcal{S}$ with $P(U_{A'}) = y \in \mathcal{N}$. Then $\Phi(y)$ has a local representative of the form $(U_{A'}, \hat{\Phi}(U_{A'}))$, where $\hat{\Phi}(U_{A'})$ are the "components" of the field $\Phi(y)$ in the twistor dyad $U_{A'}$. The function $\hat{\Phi} : \mathcal{S} \to V$ so defined has the property

$$\hat{\Phi}(U_{A'} \circ \alpha) = \tau^{-1}(\alpha) \cdot \hat{\Phi}(U_{A'}), \tag{3.12}$$

reflecting the local transformation law of the field, and transforms under $T \in$ SU(2,2) as

$$\widehat{T\Phi}(TU_{A'}) = \hat{\Phi}(U_{A'}). \tag{3.13}$$

(See, for example, Hermann.[13]) Conversely, any V-valued function on $\mathcal{S}$ satisfying (3.12) may be thought of as a cross section of $\mathcal{S} \times_\tau V$. We therefore have the result: Any representation of SU(2,2) induced by SL(2,**C**) in the manner described above may be realized on a space of vector-valued functions of two twistor variables $U_{A'}$, where $U_{A'} \in \mathcal{S}$.

4. HOMOGENEOUS TWISTOR FUNCTIONS

The representations constructed in the above manner are all highly reducible. The simplest way to see this is to consider $\mathcal{S}$ as a bundle over $\tilde{\mathcal{M}}$ with structure group $\mathbf{R}^+ \otimes$ SL(2,**C**) acting in the obvious way,

$$U_{A'} \cdot r\alpha = r U_{B'} \bar{\alpha}^{B'}{}_{A'}, \quad r \in \mathbf{R}^+, \quad \alpha \in \mathrm{SL}(2,\mathbf{C}). \tag{4.1}$$

Let ρ': $\mathcal{V} \to \mathbf{R}^+ \otimes$SL(2;**C**) send $(r\alpha, A)$ to $r\alpha$ and let ρ: SU(2,2) $\to \mathcal{S}$ be given by (3.9). Then, as before, ρ is a principal bundle homorphism (of bundles over $\tilde{\mathcal{M}}$ this time) commuting with the action of SU(2,2). So any representation induced by $\mathbf{R}^+ \otimes$ SL(2,**C**) may also be realized on a space of functions of two twistor variables.

In particular, if τ is any representation of SL(2,**C**) on V, and λ is any complex number, we can define a representation τ_λ: $\mathbf{R}^+ \otimes$ SL(2,**C**) $\to$ aut(V) by

$$\tau_\lambda(r\alpha)(v) = r^\lambda \tau(\alpha)(v). \tag{4.2}$$

If Ψ is a cross section of $\mathcal{S} \times_{\tau_\lambda} V \to \tilde{\mathcal{M}}$, the associated twistor function satisfies

$$\hat{\Psi}(U_{A'} \cdot r\alpha) = r^{-\lambda} \tau^{-1}(\alpha) \cdot \hat{\Psi}(U_{A'}). \tag{4.3}$$

For $r = 1$, this agrees with (3.12), and we may therefore regard Ψ (or $\hat{\Psi}$) as belonging to an invariant subspace of the representation induced by τ. In short, the mapping $\Phi \to \hat{\Phi}$ intertwines the τ_λ representation with the τ representation. The net effect of the intertwining operator, as can be seen by taking $\alpha = I$ in (4.3), is to remove the $\mathbf{R}^+$ dependence (or local scaling behavior) of the original field (defined on $\tilde{\mathcal{M}}$) by spreading it out in an appropriate homogeneous fashion on the generators of $\mathcal{N}$.

Example: Let $V(n)$ be the space of symmetric spinors with n primed indices. Let $\tau(n)$ be the representation of SL(2,**C**) on $V(n)$ given by

$$\tau(n)(\alpha) : \xi_{A' \ldots B'} \to \bar{\alpha}_{A'}^{-1\,C'} \cdots \bar{\alpha}_{B'}^{-1\,D'} \xi_{C' \ldots D'}. \tag{4.4}$$

If Φ is a cross section of $\mathcal{S} \times_{\tau(n)} V(n) \to \mathcal{N}$, then $\hat{\Phi}$ takes values in $V(n)$ and may be written, somewhat redundantly, as $\hat{\Phi}_{A' \ldots B'}(U_{C'})$. There are $n+1$ independent components here,

$$\varphi_r(U_{C'}) = \hat{\Phi}_{\underbrace{1' \ldots 1'}_{r} \underbrace{0' \ldots 0'}_{n-r}}(U_{C'}), \quad 0 \le r \le n. \tag{4.5}$$

If, in addition, $\hat{\Phi}$ satisfies (4.3), the functions φ_r satisfy

$$\varphi_r(tU_{0'}, U_{1'}) = t^{(n-\lambda-2r)/2} \varphi_r(U_{0'}, U_{1'}), \tag{4.6a}$$

$$\varphi_r(U_{0'}, tU_{1'}) = t^{(-n-\lambda+2r)/2} \varphi_r(U_{0'}, U_{1'}) \tag{4.6b}$$

for $t > 0$. [To obtain (4.6a), for example, note that $(tU_{0'}, U_{1'}) = (U_{0'}, U_{1'}) \circ (t^{1/2}\alpha)$, with $\alpha = \begin{bmatrix} t^{1/2} & 0 \\ 0 & t^{-1/2} \end{bmatrix}$, and use (4.3) and (4.5).]

Though no assumption has yet been made concerning the smoothness of Φ, it is natural within the twistor formalism to consider the possibility that the functions given in (4.6) are actually the restrictions to $\mathcal{S}$ of single-valued holomorphic functions defined on open sets of $\mathcal{S}^*$. A necessary precondition for this is that the exponents appearing in (4.6) be integers; if λ is set equal to $n + 2k$, this means that k must be an integer. In this case, independent of analyticity properties, $\hat{\Phi}$ is called a *function of type* (n,k). It will be shown in the next section that a massless free field of helicity $n/2$ is a function of type $(n,1)$.

Clearly, both $\mathcal{S}^*$ and $\mathcal{S}^* \times_{\tau(n)} V(n)$ are holomorphic bundles over $\mathbb{C}\mathcal{N}$, and a holomorphic (local) cross section of $\mathcal{S}^* \times_{\tau(n)} V(n)$ determines a holomorphic $V(n)$-valued function of $U_{A'}$ satisfying (3.12). In this context, if a $V(n)$-valued function $\hat{\Phi}$ [of type (n,k), say] is the restriction to $W \subseteq \mathcal{S}$ of a function holomorphic in some open neighborhood of W in $\mathcal{S}^*$, we call the field Φ determined by $\hat{\Phi}$ *real analytic* on $P(W) \subseteq \mathcal{N}$. It will be evident in the next section that this corresponds to the usual

notion of real analyticity whenever it makes sense. Notice that fields of type (n,k) which are real analytic on the whole of $\mathcal{N}$ are mapped onto themselves by the action of SU(2,2). Examples of such fields can be obtained by contour integration from Penrose's "elementary states."[6]

To continue, we now wish to show that if $\hat{\Psi}$ satisfies (4.3), the components of Ψ (regarded as a cross-section of $\mathcal{S}\times_{\tau}V\to\mathcal{N}$, will actually be homogeneous functions of r in suitably chosen bases. Let $y\to G(y)$ be a local cross section of SU(2,2) over $\mathcal{N}$. G is called *elementary* if, whenever y and ry $(r>0)$ are in the domain of G, $G(ry)=G(y)D(r)$, with

$$D(r)=\begin{bmatrix} r^{-1/2}I & 0 \\ 0 & r^{1/2}I \end{bmatrix}. \tag{4.7}$$

Correspondingly, $\rho(G(y))$ is called an elementary cross section of $\mathcal{S}$. Denoting $\rho(G(y))$ by $U_{A'}(y)$, we have

$$U_{A'}(ry)=r^{1/2}U_{A'}(y). \tag{4.8}$$

Thus, except for a scale factor, such cross sections are constant on the generators of $\mathcal{N}$.

As was mentioned before, a cross section of $\mathcal{S}\times_{\tau}V\to\tilde{\mathcal{H}}$ may be regarded as a cross section of $\mathcal{S}\times_{\tau}V\to\mathcal{N}$ as well. The components of $\Psi(y)$ in the dyad $U_{A'}(y)$ are given by

$$\psi(y)=\hat{\Psi}(U_{A'}(y)), \tag{4.9}$$

and by construction, it is immediate that (compare Refs. 1–3)

$$\psi(ry)=r^{-\lambda/2}\psi(y). \tag{4.10}$$

From a slightly different point of view, the same cross section may be realized as a conformally weighted density in $\mathbf{M}^4$ (compare Refs. 8, 14, and 15). When expressed in terms of x^{μ} and κ, the components of ψ are are homogeneous of degree 0 in the x^{μ} and of degree $-\lambda/2$ in κ. Thus, if M is a slice of $\mathcal{N}$ given by $\kappa(x)$, the components of ψ on M are related to those on $\mathbf{M}^4$ by

$$\psi(x^{\mu},\kappa(x))=[\kappa(x)]^{-\lambda/2}\psi(x^{\mu},1). \tag{4.11}$$

Identifying M with $\mathbf{M}^4$ via $(x,\kappa(x))\leftrightarrow(x,1)$, this can be written as

$$\psi'(x)=[\kappa(x)]^{-\lambda/2}\psi(x). \tag{4.12}$$

Together with (2.3), this gives the local transformation law for a density of conformal weight $-\lambda/2$. Of course, if the cross section of SU(2,2) is not elementary, the right-hand side of (4.12) will contain an x-dependent SL(2,C) matrix acting on the indices of ψ.

5. THE ZERO REST-MASS FIELDS

Consider a solution to the helicity $n/2$ free field equations in $\mathbf{M}^4$,

$$\nabla^{AA'}\varphi_{A'B'\dots C'}=0 \tag{5.1}$$

with $\varphi_{A'\dots C'}$ being the components, in the standard frame, of a spinor field with n indices. We can associate with φ a twistor function of type $(n,1)$ by declaring that $\varphi_{A'\dots C'}$ are to be the components at $(x,1)\in\mathcal{N}$ of $\hat{\Phi}$ in a particular basis.

Define the *standard* cross section $\hat{U}_{A'}(x,\kappa)$ of $\mathcal{S}\to\mathcal{N}$ to be the image, under ρ, of

$$G(x,\kappa)=T(x)D(\kappa)=\begin{bmatrix} I & ix^{AA'} \\ 0 & I \end{bmatrix}\begin{bmatrix} \kappa^{-1/2}I & 0 \\ 0 & \kappa^{1/2}I \end{bmatrix}. \tag{5.2}$$

(In $\mathbf{M}^4$, this is the normal basis in which the operators P^{μ} do not act on the indices of the fields.) We thus have

$$\begin{aligned}\hat{U}_{J'}(x^{\mu},\kappa)&=(\kappa)^{1/2}(ix^{CD'}o_{D'},o_{D'};\ ix^{CD'}\iota_{D'},\iota_{D'})\\&=(\kappa)^{1/2}\hat{U}_{J'}(x^{\mu},1),\end{aligned} \tag{5.3}$$

where $o_{D'}=(0,1)$ and $\iota_{D'}=(-1,0)$ are the components of the standard spin-frame at x. Allowing x to take on arbitrary complex values and κ to range over the complex plane minus the negative real axis, we may extend this to a holomorphic local cross section of $\mathcal{S}^*\to\mathbb{C}\mathcal{N}$ (sitting over an open neighborhood of $\mathbf{M}^4\times\mathbf{R}^+$ in $\mathbb{C}\mathcal{N}$) by demanding that (5.3) continue to hold.

Restricting our attention to $\mathcal{S}\to\mathcal{N}$ for the moment, we now define $\hat{\Phi}$ by

$$\hat{\Phi}_{A'\dots B'}(U_{K'}(x,\kappa))=\kappa^{-n/2-1}\varphi_{A'\dots B'}(x), \tag{5.4}$$

the value of $\hat{\Phi}$ in any other twistor dyad over (x,κ) being given by (3.12). Notice that if

$$U_{K'}=\bar{\alpha}_{K'}{}^{J'}\hat{U}_{J'}(x,1) \tag{5.5}$$

is an arbitrary twistor dyad at x, it has the form

$$U_{0'}=(ix^{CD'}\tilde{o}_{D'},\tilde{o}_{D'}),\quad U_{1'}=(ix^{CD'}\tilde{\iota}_{D'},\tilde{\iota}_{D'}), \tag{5.6}$$

where

$$\tilde{o}_{D'}=(\bar{\alpha}^{-1})_{D'}{}^{C'}o_{C'},\ \tilde{\iota}_{D'}=(\bar{\alpha}^{-1})_{D'}{}^{C'}\iota_{D'}. \tag{5.7}$$

Using (3.12), one sees that $\hat{\Phi}(U_{K'})$ are the components at x of the original field in the spin-frame $\{\tilde{o},\tilde{\iota}\}$. Thus the dyad $U_{K'}$ determines both the point x and a spin-frame at x,[5] while the transformation properties of $\hat{\Phi}$ correctly mirror the behavior of φ under local SL(2,C) transformations.[16]

To transform the original field φ under SU(2,2) is now a simple matter. If $T\in\mathrm{SU}(2,2)$, we define

$$(T\varphi)_{A'\dots B'}(x)=\hat{T\Phi}_{A'\dots B'}(\hat{U}_{K'}(x,1)). \tag{5.8}$$

It is now easy to verify that

$$\nabla^{AA'}(T\varphi)_{A'B'\dots C'}=0, \tag{5.9}$$

i.e., that Eq. (5.1) is conformally invariant. Poincaré and dilation invariance is immediate, so the only thing that needs to be checked is invariance under $\mathfrak{A}$. Since $\mathfrak{A}$ is conjugate to the translations under $H=\left[\begin{smallmatrix}0 & I\\ I & 0\end{smallmatrix}\right]$, the proof reduces to showing that $H\varphi$ satisfies (5.1) if φ does. The calculation involved is straightforward, but quite lengthy, and is omitted, although it should be noted that this is the point at which the *precise* degree of homogeneity of $\hat{\Phi}$ plays a crucial role.

Conversely, it is clear that any $\hat{\Phi}$ of type $(n,1)$ which satisfies (5.1) in the standard cross section gives rise to a zero rest-mass field in $\mathbf{M}^4$. In a general cross section (restricted to $\mathbf{M}^4$), Eq. (5.1) will of course involve the spinor covariant derivative.

We now confine our attention to solutions of (5.1) which are real analytic on some open subset of $\mathbf{M}^4$. These admit holomorphic extensions to open subsets of

$\mathbf{CM}^4$ which also satisfy (5.1). Using (5.4) with x and κ complex, we obtain a $\hat{\Phi}$ of type $(n,1)$, holomorphic on some domain $\mathcal{D} \subseteq \mathcal{S}^*$, where $\mathcal{D} \cap \mathcal{S}$ is open in $\mathcal{S}$. Thus, from (5.4),

$$\nabla^{AA'}\hat{\Phi}_{A'B'\ldots C'}(\hat{U}_{J'}(x,\kappa)) = 0. \tag{5.10}$$

The two equations $(A' = 0', 1')$ reduce to

$$\frac{\partial \varphi_r}{\partial U^0} + \frac{\partial \varphi_{r+1}}{\partial V^0} = \frac{\partial \varphi_r}{\partial U^1} + \frac{\partial \varphi_{r+1}}{\partial V^1} = 0. \tag{5.11}$$

Although these equations apparently hold only on the standard cross section of $\mathcal{S}^*$, they are in fact valid on the entire domian of $\hat{\Phi}$. To see this, let $\alpha \in \mathrm{SL}(2,C)$ and consider the cross section $\tilde{U}_{J'}(x,\kappa) = \bar{\alpha}_{J'}{}^{K'}\hat{U}_{K'}(x,\kappa)$. Then (3.12) together with the local $\mathrm{SL}(2,\mathbf{C})$ invariance of the fields, implies

$$\tilde{\nabla}^{AA'}\hat{\Phi}_{A'B'\ldots C'}(\tilde{U}_{J'}(x,\kappa)) = 0, \tag{5.12}$$

where $\tilde{\nabla}^{AA'} = \bar{\alpha}^{-1}{}_{C'}{}^{A'}\nabla^{AC'}$. Expanding (5.12) then gives (5.11) for $U_{J'} = \tilde{U}_{J'}$. Since any dyad over (x,κ) may be obtanied in this fashion, (5.11) is valid on an open subset of $\mathcal{S}^*$; thus, by analytic continuation, (5.11) holds whenever it makes sense.

Now let H act on $\hat{\Phi}$: this interchanges U^0 with U^2 and U^1 with U^3 (similarly for V^α). Since $H\varphi$ satisfies (5.10), proceeding as above we obtain the additional equations

$$\frac{\partial \varphi_r}{\partial U^2} + \frac{\partial \varphi_{r+1}}{\partial V^2} = \frac{\partial \varphi_r}{\partial U^3} + \frac{\varphi_{r+1}}{\partial V^3} = 0. \tag{5.13}$$

We therefore have

$$\frac{\partial \varphi_r}{\partial U^\alpha} + \frac{\partial \varphi_{r+1}}{\partial V^\alpha} = 0, \quad r = 0,1,\ldots,n-1. \tag{5.14}$$

The homogeneity conditions (4.6) may be written

$$U^\alpha \frac{\partial \varphi_r}{\partial U^\alpha} = -(r+1)\varphi_r, \quad V^\alpha \frac{\partial \varphi_r}{\partial V^\alpha} = -(n-r+1)\varphi_r, \tag{5.15}$$

and these are the twistor versions of the zero rest-mass free field equations.[5-7] The auxilliary conditions

$$\frac{\partial^2 \varphi_r}{\partial U^{[\alpha}\partial V^{\beta]}} = 0 \tag{5.16}$$

follows immediately from (5.14).

Conversely, Eqs. (5.14)—(5.16) are manifestly conformally invariant, and it is a simple matter to check that any solution to them yields, in the manner described above, a real analytic solution to the zero rest-mass field equations in Minkowski space. If the functions φ_r are holomorphic on a neighborhood of the whole of $\mathcal{S}$ in $\mathcal{S}^*$, the resultant field will be *single-valued* and vanish in a characteristic fashion at null infinity. Specifically, it will exhibit the "peeling-off" property; see Ref. 8 and the other references cited there.

Remark: The center of $\mathrm{SU}(2,2)$ is generated by powers of $J = \left[\begin{smallmatrix} iI & 0 \\ 0 & iI \end{smallmatrix}\right]$. For twistor functions of type $(n,1)$,

$$J\varphi_r = (-i)^{-n-2}\varphi_r. \tag{5.17}$$

Thus, if $n \equiv 2 (\mathrm{mod}\ 4)$, $J\varphi_r = \varphi_r$, and we have a representation of the conformal group. If $n \equiv 0 (\mathrm{mod}\ 4)$, we have a representation of $\mathrm{SO}_0(2,4)$, while if $n \equiv 1$ or $3 (\mathrm{mod}\ 4)$, we have a representation of $\mathrm{SU}(2,2)$.

6. COMPLEX CONJUGATE REPRESENTATIONS

The complex conjugates of the representations considered above are most usefully realized on the bundle of dual twistor dyads. Let

$$\mathcal{S}'^* = \{(U_\alpha, V_\alpha): U_\alpha \neq \lambda V_\alpha,\ \lambda \in \mathbf{C}\} \tag{6.1}$$

with U and V transforming under the adjoint representation of $\mathrm{SU}(2,2)$. Letting $\bar{U}^\alpha = Q^{\alpha\beta}U_\beta$, define the projection from $\mathcal{S}'^*$ to $\mathbb{CN}$ by

$$P'(U_\alpha, V_\alpha) = P(\bar{U}^\alpha, \bar{V}^\alpha). \tag{6.2}$$

Denote the pair (U_α, V_α) by $U_A (A = 0,1)$, and define the action of $\mathrm{SL}(2,C)$ by

$$U_A \cdot \alpha = \alpha_A{}^B U_B. \tag{6.3}$$

The bundle of dual twistor dyads over $\mathcal{N}$ is defined by $\mathcal{S}' \equiv P'^{-1}(\mathcal{N})$.

Just as above, the representation of $\mathrm{SU}(2,2)$ induced by the representation

$$\pi(n)(\alpha): \xi_{A\ldots B} \to \alpha_A^{-1C}\cdots\alpha_B^{-1D}\xi_{C\ldots D} \tag{6.4}$$

of $\mathrm{SL}(2,\mathbf{C})$ may be realized as a space of functions of the two variables U_A. It is a simple matter to verify that the homogeneous vector bundle constructed in this fashion is (isomorphic to) the complex conjugate of $\mathcal{S} \times_{\tau(n)} V(n)$. There is an analogous decomposition of the representation into homogeneous twistor functions. In particular, corresponding to a real analytic solution of the field equations

$$\nabla^{AA'}\psi_{AB\ldots C} = 0, \tag{6.5}$$

there is a holomorphic twistor version,[6,7]

$$\frac{\partial \psi_r}{\partial U_\alpha} + \frac{\partial \psi_{r+1}}{\partial V_\alpha} = 0$$
$$U_\alpha \frac{\partial \psi_r}{\partial U_\alpha} = -(r+1)\psi_r, \quad V_\alpha \frac{\partial \psi_r}{\partial V_\alpha} = -(n-r+1)\psi_r. \tag{6.6}$$

The field $\bar{\psi}_{A'\ldots B'} = \overline{\psi_{A\ldots B}}$ determines (see Sec. 5) a set of twistor fucntions $\{\varphi_r(U^\alpha, V^\alpha)\}$ satisfying (5.14)—(5.16). They are related to the ψ_r's by

$$\varphi_r(\bar{U}^\alpha, \bar{V}^\alpha) = \overline{\psi_r(U_\alpha, V_\alpha)}. \tag{6.7}$$

7. FINITE-DIMENSIONAL REPRESENTATIONS

In certain applications,[6,7,11] it is convenient to regard a twistor $W_\alpha = (w_A, \pi^{A'})$ of valence $[{}^0_1]$ as defining a spinor field in Minkowski space via

$$\sigma_{A'}(x) \equiv ix_{AA'}w^A - \pi_{A'}. \tag{7.1}$$

This field satisfies the *twistor equation*[11]

$$\nabla_{B(C'}\sigma_{A')} = 0. \tag{7.2}$$

Conversely, it can be shown that any global solution to (7.2) is of the form (7.1), thereby uniquely defining a twistor W_α.

Equations such as (7.1) arise quite naturally in the formalism given in this paper. For consider the constant, totally symmetric twistor $S_{\alpha\beta\ldots\gamma}$ of valence $[{}^0_n]$. This can be used to define a symmetric spinor-valued function $\Sigma_{A'\ldots B'}$ on $\mathcal{S}^*$,

$$\Sigma_{A'\dots B'} = S_{\alpha\dots\beta} U^{\alpha}_{A'} \cdots U^{\beta}_{B'} \ . \tag{7.3}$$

Upon writing out the independent components of Σ, one sees that it is a function of type $(n, -n)$; thus this representation contains a finite-dimensional invariant subspace. If we use Σ to define a conformally invariant field in M^4 by putting

$$\tau_{A'B'\dots K'}(x) \equiv \Sigma_{A'B'\dots K'}(\hat{U}_{J'}(x, 1)), \tag{7.4}$$

then an easy calculation gives

$$\nabla_{E(E'} \tau_{A'B'\dots K')} = 0, \tag{7.5}$$

a solution to the generalized twistor equation. Conversely, using the methods of Sec. 5, it may be shown that any solution to (7.5) uniquely determines a symmetric twistor of valence $\left[{}^{0}_{n}\right]$.

CONCLUSION

We have established a correspondence between a certain series of induced representations of SU(2,2) and different space of homogeneous functions of two twistor variables. We have shown, in a concrete way, that the zero rest-mass fields, homogeneous functions on $\mathcal{N}$, and conformally weighted spinor densities, all of which naturally occur in discussions of the conformal group, should be regarded as representing different aspects of the same geometric objects, namely cross sections of homogeneous vector bundles. We have also tried to show that, even in the nonanalytic case, these objects are most naturally described by means of twistors. The advantages of the twistor description are particularly evident in the case of analytic fields.

As a final remark, we should note that while twistor theory is manifestly conformally invariant, it is a simple matter to "break conformal invariance" and thereby obtain a Poincare invariant theory. It is only necessary to restrict the bundles $\mathcal{S}$ and $\mathcal{S}'$ to the slice $\kappa = 1$ of $\mathcal{N}$ and it is achieved in practice by incorporating the "infinity twistor" into the equations. For this, and other matters related to quantization, the reader is referred particularly to Ref. 7.

*Partially supported by the General Research Fund of the University of Kansas.

[1]P.A.M. Dirac, Ann. Math. 37, 429 (1936).
[2]G. Mack and A. Salam, Ann. Phys. 53, 174 (1969).
[3]S. Ferrara, R. Gatto, and A.F. Grillo, Springer Tracts in Modern Physics, Vol. 67 (1973).
[4]T.H. Go and D.H. Mayer, Rep. Math. Phys. 5, 187 (1974).
[5]R. Penrose, Int. J. Theor. Phys. 1, 61 (1968).
[6]R. Penrose and M.A.H. MacCallum, Phys. Rep. 6C, 242 (1973).
[7]R. Penrose in *Quantum Gravity: An Oxford Symposium* (Oxford U.P., 1975).
[8]R. Penrose, Proc. Roy. Soc. London A 284, 159 (1965).
[9]B.D. Bramson, R. Penrose, and G.A.J. Sparling, "A Derivation of the Twistor Description of Free Fields," to appear.
[10]L. Gross, J. Math. Phys. 5, 687 (1964).
[11]R. Penrose, J. Math. Phys. 8, 345 (1967).
[12]R. Hansen and E.T. Newman, Gen. Rel. Grav. 6, 361 (1975).
[13]R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[14]H.A. Buchdahl, Nuovo Cimento **11**, 496 (1959).
[15]R. Penrose in *Group Theory in Nonlinear Problems*, edited by A. Barut (Reidel, Holland, 1974).
[16]The identification of a twistor dyad with an SL(2,C) spin-frame can only be made over simply connected regions of $\mathcal{N}$. Any attempt to extend this identification further leads to inconsistencies (the "Grgin discontinuities"[7]).
[17]E. Grgin, Ph.D. Thesis, Syracuse University (1966).
[18]D. Lerner, "Global Properties of Conformally Invariant Fields," to appear.

Post-Newtonian two-body and n-body problems with electric charge in general relativity

B. M. Barker

Department of Physics and Astronomy, The University of Alabama, University, Alabama 35486

R. F. O'Connell

Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803
(Received 16 February 1977)

Starting with the Bażański two-body post-Newtonian Lagrangian with electric charge in general relativity, we construct a coordinate transformation (not involving center-of-mass coordinates) with two arbitrary parameters and obtain a Hamiltonian which is in agreement with one derived from quantum field theory. The field theory Hamiltonian corresponds to using an arbitrary parameter x_p in the photon propagator as well as an arbitrary parameter x_g in the graviton propagator. These results are also generalized to the case of n bodies. The condition for static balance $e_i = \pm G^{1/2} m_i$ is found to hold both for the exact Reissner-Nordstrøm "one-body" problem and for the post-Newtonian n-body problem. An alternate condition for static balance $e_i = \pm (G m_1 m_2)^{1/2}$ is found to hold for the post-Newtonian two-body problem. The precession of the perihelion for the post-Newtonian two-body problem is given along with four special cases, one of which is the two-body generalization of the "one-body" special relativity result of Sommerfeld. Post-Newtonian two-body equations of motion (in center-of-mass coordinates) with the condition of static balance are also examined.

INTRODUCTION

Bażański[1,2] has given the two-body post-Newtonian equations of motion[1] and Lagrangian[2] for two charged bodies in general relativity. His Lagrangian contains both the potential-energy terms of the Einstein–Infeld–Hoffman[3] Lagrangian (G, Gvv, and G^2 terms) and the potential-energy terms of the Darwin[4] Lagrangian (e^2 and e^2vv terms) as well as some additional mixed potential-energy terms (Ge^2 terms).

In Sec. I we write the Bażański Lagrangian and Hamiltonian in the more convenient notation[5] of Landau and Lifshitz.[6]

We then make a coordinate transformation from the coordinate system used by Bażański, Einstein–Infeld–Hoffmann, and Darwin to a new arbitrary coordinate system characterized by two arbitrary dimensionless parameters α_g and α_p in the transformation equations. We then obtain (in the new coordinate system) the Bażański Lagrangian and Hamiltonian which contain the parameters α_g and α_p. The coordinate transformation used is new, in two respects, compared to what has been done before by Hiida and Okamura[7] and by ourselves.[8] First, the transformation is made before going to center-of-mass coordinates and, second, it contains the additional parameter α_p which means that the gravitational and electromagnetic aspects of this paper are treated in a similar manner.

We then show how one can derive the potential-energy terms containing α_g and α_p in the Bażański Hamiltonian from quantum field theory and mention which terms have (and which have not) been so derived.

Next, we introduce center-of-mass coordinates in our Lagrangian. Then, after making the large-mass large-charge approximation, we compare our result with the Reissner–Nordstrøm[9] Lagrangian.

In Sec. II we discuss the conditions of static balance[5] where the electric and gravitational forces cancel out when the two particles are at rest.

In Sec. III we give the post-Newtonian equation of motion in a center-of-mass coordinate system and then find the precession of the orbit (perihelion precession) which agrees with the special cases of Robertson[10] and Sommerfeld.[11] Solutions to the post-Newtonian equations of motion for the two-body problem in center-of-mass coordinates, satisfying the condition of static balance are found.

In Sec. IV we generalize our results to the case of n-bodies and in Sec. V we present our conclusions.

I. LAGRANGIAN AND HAMILTONIAN

The two-body post-Newtonian Lagrangian[2] and equations of motion[1] for the case of charged particles in general relativity have been given by Bażański. The Lagrangian in Bażański (B) coordinates [same coordinates as used by Einstein–Infeld–Hoffmann[3] (EIH) and Darwin[4] (D)] can be written as

$$\begin{aligned} L(\mathbf{r}_{1B}, \mathbf{v}_{1B}; \mathbf{r}_{2B}, \mathbf{v}_{2B}) &= \tfrac{1}{2} m_1 v_{1B}^2 + \tfrac{1}{2} m_2 v_{2B}^2 + \tfrac{1}{8} m_1 v_{1B}^4/c^2 + \tfrac{1}{8} m_2 v_{2B}^4/c^2 \\ &+ \frac{G m_1 m_2}{r_B}\left[1 + \tfrac{3}{2}(v_{1B}^2 + v_{2B}^2)/c^2 - \tfrac{7}{2}(\mathbf{v}_{1B}\cdot\mathbf{v}_{2B})/c^2\right. \\ &\left. - \tfrac{1}{2}(\mathbf{v}_{1B}\cdot\mathbf{r}_B)(\mathbf{v}_{2B}\cdot\mathbf{r}_B)/c^2 r_B^2\right] - \frac{e_1 e_2}{r_B}\left[1 - \tfrac{1}{2}(\mathbf{v}_{1B}\cdot\mathbf{v}_{2B})/c^2\right. \\ &\left. - \tfrac{1}{2}(\mathbf{v}_{1B}\cdot\mathbf{r}_B)(\mathbf{v}_{2B}\cdot\mathbf{r}_B)/c^2 r_B^2\right] - \frac{G^2 m_1 m_2 (m_1 + m_2)}{2c^2 r_B^2} \\ &+ \frac{G e_1 e_2 (m_1 + m_2)}{c^2 r_B^2} - \frac{G(e_1^2 m_2 + e_2^2 m_1)}{2c^2 r_B^2}, \end{aligned} \tag{1}$$

where c and G are the speed of light and gravitational

constant respectively, $\mathbf{r}_B = \mathbf{r}_{1B} - \mathbf{r}_{2B}$ and thus $\mathbf{v}_B = \mathbf{v}_{1B} - \mathbf{v}_{2B}$. By the usual standard procedure we obtain the Hamiltonian

$$
\begin{aligned}
&H(\mathbf{r}_{1B}, \mathbf{P}_{1B}; \mathbf{r}_{2B}, \mathbf{P}_{2B}) \\
&= \frac{P_{1B}^2}{2m_1} + \frac{P_{2B}^2}{2m_2} - \frac{P_{1B}^4}{8m_1^3c^2} - \frac{P_{2B}^4}{8m_2^3c^2} - \frac{Gm_1m_2}{r_B} \\
&\times \left[1 + \frac{3}{2}\left(\frac{P_{1B}^2}{m_1^2c^2} + \frac{P_{2B}^2}{m_2^2c^2}\right) - \frac{7}{2}\frac{(\mathbf{P}_{1B}\cdot\mathbf{P}_{2B})}{m_1m_2c^2} \right. \\
&\left. - \frac{1}{2}\frac{(\mathbf{P}_{1B}\cdot\mathbf{r}_B)(\mathbf{P}_{2B}\cdot\mathbf{r}_B)}{m_1m_2c^2r_B^2}\right] + \frac{e_1e_2}{r_B}\left[1 - \frac{1}{2}\frac{(\mathbf{P}_{1B}\cdot\mathbf{P}_{2B})}{m_1m_2c^2}\right. \\
&\left. - \frac{1}{2}\frac{(\mathbf{P}_{1B}\cdot\mathbf{r}_B)(\mathbf{P}_{2B}\cdot\mathbf{r}_B)}{m_1m_2c^2r_B^2}\right] + \frac{G^2m_1m_2(m_1+m_2)}{2c^2r_B^2} \\
&- \frac{Ge_1e_2(m_1+m_2)}{c^2r_B} + \frac{G(e_1^2m_2 + e_2^2m_1)}{2c^2r_B^2}. \qquad (2)
\end{aligned}
$$

A. Coordinate transformation

We shall now make the coordinate transformation (see Appendix)

$$\mathbf{r}_{1B} = \mathbf{r}_1 - \mathbf{r}\left(\alpha_g\frac{Gm_2}{c^2r} - \alpha_p\frac{e_1e_2}{m_1c^2r}\right), \qquad (3)$$

$$\mathbf{r}_{2B} = \mathbf{r}_2 + \mathbf{r}\left(\alpha_g\frac{Gm_1}{c^2r} - \alpha_p\frac{e_1e_2}{m_2c^2r}\right), \qquad (4)$$

which implies that

$$\mathbf{r}_B = \mathbf{r}\left(1 - \alpha_g\frac{GM}{c^2r} + \alpha_p\frac{e_1e_2}{\mu c^2r}\right), \qquad (5)$$

$$\mathbf{v}_{1B} = \mathbf{v}_1 - \left(\alpha_g\frac{Gm_2}{c^2r} - \alpha_p\frac{e_1e_2}{m_1c^2r}\right)\left[\mathbf{v} - \frac{(\mathbf{v}\cdot\mathbf{r})\mathbf{r}}{r^2}\right], \qquad (6)$$

$$\mathbf{v}_{2B} = \mathbf{v}_2 + \left(\alpha_g\frac{Gm_1}{c^2r} - \alpha_p\frac{e_1e_2}{m_2c^2r}\right)\left[\mathbf{v} - \frac{(\mathbf{v}\cdot\mathbf{r})\mathbf{r}}{r^2}\right], \qquad (7)$$

$$
\begin{aligned}
\mathbf{P}_{1B} &= \mathbf{P}_1 + \left(\alpha_g\frac{Gm_2}{c^2r} - \alpha_p\frac{e_1e_2}{m_1c^2r}\right)\left[\mathbf{P}_1 - \frac{(\mathbf{P}_1\cdot\mathbf{r})\mathbf{r}}{r^2}\right] \\
&- \left(\alpha_g\frac{Gm_1}{c^2r} - \alpha_p\frac{e_1e_2}{m_2c^2r}\right)\left[\mathbf{P}_2 - \frac{(\mathbf{P}_2\cdot\mathbf{r})\mathbf{r}}{r^2}\right], \qquad (8)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}_{2B} &= \mathbf{P}_2 + \left(\alpha_g\frac{Gm_1}{c^2r} - \alpha_p\frac{e_1e_2}{m_2c^2r}\right)\left[\mathbf{P}_2 - \frac{(\mathbf{P}_2\cdot\mathbf{r})\mathbf{r}}{r^2}\right] \\
&- \left(\alpha_g\frac{Gm_2}{c^2r} - \alpha_p\frac{e_1e_2}{m_1c^2r}\right)\left[\mathbf{P}_1 - \frac{(\mathbf{P}_1\cdot\mathbf{r})\mathbf{r}}{r^2}\right], \qquad (9)
\end{aligned}
$$

where α_g and α_p are arbitrary dimensionless parameters, and $\mu = m_1m_2/(m_1+m_2)$ and $M = m_1 + m_2$ are the reduced mass and total mass, respectively.

The Hamiltonian of Eq. (2) in the new coordinates is

$$
\begin{aligned}
H(\alpha_g, \alpha_p) &= H_0 + V_{1(\mathrm{EIH})}(\alpha_g) + V_{2(\mathrm{EIH})}(\alpha_g) \\
&+ V_{1(\mathrm{D})}(\alpha_p) + V_{2(\mathrm{D})}(\alpha_p) + V_{2(\mathrm{B})}(\alpha_g, \alpha_p), \qquad (10)
\end{aligned}
$$

where

$$H_0 = \frac{P_1^2}{2m_1} + \frac{P_2^2}{2m_2} - \frac{P_1^4}{8m_1^3c^2} - \frac{P_2^4}{8m_2^3c^2}, \qquad (11)$$

$$
\begin{aligned}
&V_{1(\mathrm{EIH})}(\alpha_g) \\
&= -\frac{Gm_1m_2}{r}\left\{1 + (\tfrac{3}{2} - \alpha_g)\left(\frac{P_1^2}{m_1^2c^2} + \frac{P_2^2}{m_2^2c^2}\right)\right. \\
&+ (2\alpha_g - \tfrac{7}{2})\frac{(\mathbf{P}_1\cdot\mathbf{P}_2)}{m_1m_2c^2} - (\tfrac{1}{2} + 2\alpha_g)\frac{(\mathbf{P}_1\cdot\mathbf{r})(\mathbf{P}_2\cdot\mathbf{r})}{m_1m_2c^2r^2} \\
&\left. + \alpha_g\left[\frac{(\mathbf{P}_1\cdot\mathbf{r})^2}{m_1^2c^2r^2} + \frac{(\mathbf{P}_2\cdot\mathbf{r})^2}{m_2^2c^2r^2}\right]\right\}, \qquad (12)
\end{aligned}
$$

$$V_{2(\mathrm{EIH})}(\alpha_g) = (\tfrac{1}{2} - \alpha_g)\frac{G^2m_1m_2(m_1+m_2)}{c^2r^2}, \qquad (13)$$

$$
\begin{aligned}
&V_{1(\mathrm{D})}(\alpha_p) \\
&= \frac{e_1e_2}{r}\left\{1 - \alpha_p\left(\frac{P_1^2}{m_1^2c^2} + \frac{P_2^2}{m_2^2c^2}\right) + (2\alpha_p - \tfrac{1}{2})\frac{(\mathbf{P}_1\cdot\mathbf{P}_2)}{m_1m_2c^2}\right. \\
&\left. - (\tfrac{1}{2} + 2\alpha_p)\frac{(\mathbf{P}_1\cdot\mathbf{r})(\mathbf{P}_2\cdot\mathbf{r})}{m_1m_2c^2r^2} + \alpha_p\left[\frac{(\mathbf{P}_1\cdot\mathbf{r})^2}{m_1^2c^2r^2} + \frac{(\mathbf{P}_2\cdot\mathbf{r})^2}{m_2^2c^2r^2}\right]\right\}, \qquad (14)
\end{aligned}
$$

$$V_{2(\mathrm{D})}(\alpha_p) = -\alpha_p\frac{e_1^2e_2^2}{\mu c^2r^2}, \qquad (15)$$

$$V_{2(\mathrm{B})}(\alpha_g, \alpha_p) = (\alpha_g + \alpha_p - 1)\frac{Ge_1e_2(m_1+m_2)}{c^2r^2} + \frac{G(e_1^2m_2 + e_2^2m_1)}{2c^2r^2}. \qquad (16)$$

The coordinate system is B—EIH—D if both $\alpha_g = 0$ and $\alpha_p = 0$. In a coordinate system where $\alpha_g = \frac{1}{2}$ we have $V_{2(\mathrm{EIH})}(\frac{1}{2}) = 0$, while for a coordinate system where $\alpha_p = 0$ we have $V_{2(\mathrm{D})}(0) = 0$.

The special cases of the Einstein—Infeld—Hoffman Hamiltonian (pure gravitation, i.e., $e_1 = e_2 = 0$) and the Darwin Hamiltonian (pure electromagnetism, i.e., $G = 0$) can be written, respectively, as

$$H_{\mathrm{EIH}}(\alpha_g) = H_0 + V_{1(\mathrm{EIH})}(\alpha_g) + V_{2(\mathrm{EIH})}(\alpha_g), \qquad (17)$$

$$H_{\mathrm{D}}(\alpha_p) = H_0 + V_{1(\mathrm{D})}(\alpha_p) + V_{2(\mathrm{D})}(\alpha_p). \qquad (18)$$

The Lagrangian corresponding to Eq. (10) can be written as

$$
\begin{aligned}
L(\alpha_g, \alpha_p) &= L_0 - V_{1(\mathrm{EIH})}(\alpha_g) - V_{2(\mathrm{EIH})}(\alpha_g) \\
&- V_{1(\mathrm{D})}(\alpha_p) - V_{2(\mathrm{D})}(\alpha_p) - V_{2(\mathrm{B})}(\alpha_g, \alpha_p), \qquad (19)
\end{aligned}
$$

where

$$L_0 = \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2 + \tfrac{1}{8}m_1v_1^4/c^2 + \tfrac{1}{8}m_2v_2^4/c^2, \qquad (20)$$

and $m_1\mathbf{v}_1$ and $m_2\mathbf{v}_2$ replace $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively, in the potential energy terms of Eqs. (12) and (14).

B. Hamiltonian from quantum field theory

The potential-energy terms $V_{1(\mathrm{EIH})}(\alpha_g)$ and $V_{2(\mathrm{EIH})}(\alpha_g)$ can be derived from the one-graviton exchange interaction and the two-graviton exchange interaction,[12] respectively. The potential-energy terms $V_{1(\mathrm{D})}(\alpha_p)$ and $V_{2(\mathrm{D})}(\alpha_p)$ can be derived from the one-photon exchange interaction and the two-photon exchange interaction,[12] respectively. The potential-energy term $V_{2(\mathrm{B})}(\alpha_g, \alpha_p)$ can be derived from the one-graviton one-photon exchange interaction.[12]

Both the graviton and the photon propagators are proportional to $1/(\mathbf{k}^2 - k_0^2)$, [see Eqs. (12)—(18) of Ref. 8], where k_0^2 can be written in a form that contains an arbitrary dimensionless parameter x. For the graviton case let this parameter be x_g and for the photon case let this parameter be x_p. It is convenient to introduce two other dimensionless parameters α_g and α_p which are given by

$$\alpha_g = -\tfrac{1}{4}(1 - x_g) \quad \text{and} \quad \alpha_p = -\tfrac{1}{4}(1 - x_p). \qquad (21)$$

Thus, the potential-energy terms derived from quantum

field theory will contain two arbitrary parameters α_g and α_p. The proper interpretation of the Hamiltonians with arbitrary parameters α_g and α_p is that the Hamiltonians are related to each other by the coordinate transformation given by Eqs. (3) and (4) whose form was carefully chosen to be consistent with the field theory results.

Iwasaki[13] gave the field theory derivation for $V_{1(\mathrm{EIH})}(0)$ and $V_{2(\mathrm{EIH})}(0)$, while Hiida and Okamura[7] gave the field theory derivation [see their Eqs. (1.3) and (3.20)] for $V_{1(\mathrm{EIH})}(\alpha_g)$ and $V_{2(\mathrm{EIH})}(\alpha_g)$ which agree with our Eqs. (12) and (13). We have verified that the field theory derivation for $V_{1(\mathrm{EIH})}(\alpha_g)$ and $V_{1(\mathrm{D})}(\alpha_p)$ is in agreement with Eqs. (12) and (14). To our knowledge, a quantum field theory derivation of $V_{2(\mathrm{D})}(\alpha_p)$ and $V_{2(\mathrm{B})}(\alpha_g, \alpha_p)$ has not been made.

C. Center-of-mass coordinates

Going to center-of-mass coordinates we put $\mathbf{P} = \mathbf{P}_1 = -\mathbf{P}_2$ in Eq. (10) to obtain

$$
\begin{aligned}
&H(\alpha_g, \alpha_p; \mathbf{r}, \mathbf{P}) \\
&= \frac{1}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) P^2 - \frac{1}{8}\left(\frac{1}{m_1^3} + \frac{1}{m_2^3}\right)\frac{P^4}{c^2} \\
&- \frac{Gm_1m_2}{r}\left\{1 + \left[\frac{1}{2} + \left(\frac{3}{2} - \alpha_g\right)\frac{M}{\mu}\right]\frac{P^2}{m_1m_2c^2} \right. \\
&\left. + \left(\frac{1}{2} + \alpha_g\frac{M}{\mu}\right)\frac{(\mathbf{P}\cdot\mathbf{r})^2}{m_1m_2c^2r^2}\right\} + \frac{e_1e_2}{r} \\
&\times\left\{1 + \left(\frac{1}{2} - \alpha_p\frac{M}{\mu}\right)\frac{P^2}{m_1m_2c^2} + \left(\frac{1}{2} + \alpha_p\frac{M}{\mu}\right)\right. \\
&\left.\times\frac{(\mathbf{P}\cdot\mathbf{r})^2}{m_1m_2c^2r^2}\right\} + \left(\frac{1}{2} - \alpha_g\right)\frac{G^2\mu M^2}{c^2r^2} - \alpha_p\frac{e_1^2e_2^2}{\mu c^2r^2} \\
&+ (\alpha_g + \alpha_p - 1)\frac{Ge_1e_2M}{c^2r^2} + \frac{G(e_1^2m_2 + e_2^2m_1)}{2c^2r^2}. \qquad (22)
\end{aligned}
$$

Equation (22) also could have been obtained by starting with Eq. (2), going to center-of-mass coordinates, and then making the transformation of Eq. (5) which implies that

$$
\mathbf{v}_B = \mathbf{v} - \left(\alpha_g\frac{GM}{c^2r} - \alpha_p\frac{e_1e_2}{\mu c^2r}\right)\left[\mathbf{v} - \frac{(\mathbf{v}\cdot\mathbf{r})\,\mathbf{r}}{r^2}\right], \qquad (23)
$$

$$
\mathbf{P}_B = \mathbf{P} + \left(\alpha_g\frac{GM}{c^2r} - \alpha_p\frac{e_1e_2}{\mu c^2r}\right)\left[\mathbf{P} - \frac{(\mathbf{P}\cdot\mathbf{r})\,\mathbf{r}}{r^2}\right]. \qquad (24)
$$

The pure gravitational part of Eq. (22) [G, GPP, and G^2 terms] has been given by Hiida and Okamura[7] and by ourselves.[8] The one-graviton exchange part of Eq. (22) [G and GPP terms] with $\alpha_g = -\frac{1}{2}\mu/M$ has been derived by Barker, Gupta, and Haracz[14] using Gupta's[15] quantum theory of gravitation.

The Lagrangian corresponding to Eq. (22) can be written as

$$
\begin{aligned}
&L(\alpha_g, \alpha_p; \mathbf{r}, \mathbf{v}) \\
&= \tfrac{1}{2}\mu v^2 + \tfrac{1}{8}(1 - 3\mu/M)\frac{\mu v^4}{c^2} + \frac{G\mu M}{r}\left[1 + \left(\frac{3}{2} - \alpha_g + \frac{1}{2}\frac{\mu}{M}\right)\frac{v^2}{c^2}\right. \\
&\left. + \left(\alpha_g + \frac{1}{2}\frac{\mu}{M}\right)\frac{(\mathbf{v}\cdot\mathbf{r})^2}{c^2r^2}\right] - \frac{e_1e_2}{r}\left[1 + \left(-\alpha_p + \frac{1}{2}\frac{\mu}{M}\right)\frac{v^2}{c^2}\right. \\
&\left. + \left(\alpha_p + \frac{1}{2}\frac{\mu}{M}\right)\frac{(\mathbf{v}\cdot\mathbf{r})^2}{c^2r^2}\right] + \left(\alpha_g - \frac{1}{2}\right)\frac{G^2\mu M^2}{c^2r^2} + \alpha_p\frac{e_1^2e_2^2}{\mu c^2r^2} \\
&+ (1 - \alpha_g - \alpha_p)\frac{Ge_1e_2M}{c^2r^2} - \frac{G(e_1^2m_2 + e_2^2m_1)}{2c^2r^2}. \qquad (25)
\end{aligned}
$$

In the coordinate system where $\alpha_g = \alpha_p = -\mu/2M$ we can write Eq. (25) as

$$
\begin{aligned}
&L(\alpha_g = \alpha_p = -\mu/2M; \mathbf{r}, \mathbf{v}) \\
&= \tfrac{1}{2}\mu v^2 + G'\mu M/r + \tfrac{1}{8}k_1\mu v^4/c^2 \\
&+ \tfrac{3}{2}k_2G'\mu Mv^2/c^2r - \tfrac{1}{2}k_3G'^2\mu M^2/c^2r^2, \qquad (26)
\end{aligned}
$$

where

$$
G' = G - e_1e_2/m_1m_2, \qquad (27)
$$

$$
k_1 = 1 - 3\mu/M, \qquad (28)
$$

$$
k_2 = 1 + \tfrac{2}{3}\mu/M + Z_a, \qquad (29)
$$

$$
k_3 = 1 + \mu/M + Z_b + Z_aZ_b - Z_a^2, \qquad (30)
$$

and

$$
Z_a = \frac{e_1e_2}{G'm_1m_2}, \qquad (31)
$$

$$
Z_b = \frac{e_1^2m_2 + e_2^2m_1}{G'm_1m_2M}. \qquad (32)
$$

As the precession of the orbit (see Sec. III) is independent of the parameters α_g and α_p it is convenient (for the derivation of the precession of the orbit) to choose the Lagrangian in as simple a form as possible.[8] By setting $\alpha_g = \alpha_p = -\mu/2M$, we eliminate the $(\mathbf{v}\cdot\mathbf{r})^2$ potential energy terms in Eq. (25). We can also eliminate the v^2 potential energy terms in Eq. (25) by setting $\alpha_g = \frac{3}{2} + \frac{1}{2}\mu/M$ and $\alpha_p = \frac{1}{2}\mu/M$.

D. Large-mass, large-charge approximation

Applying the large-mass approximation ($m_2 \gg m_1$) and the large-charge approximation ($e_2 \gg e_1$) with the condition $e_2^2m_1 \gg e_1^2m_2$ to Eq. (25) we obtain the so-called "one-body" Lagrangian (there are actually two bodies) which is

$$
\begin{aligned}
&L(\alpha_g, \alpha_p; \mathbf{r}, \mathbf{v}) \\
&= \tfrac{1}{2}m_1v^2 + \tfrac{1}{8}m_1\frac{v^4}{c^2} + \frac{Gm_1m_2}{r}\left[1 + \left(\tfrac{3}{2} - \alpha_g\right)\frac{v^2}{c^2}\right. \\
&\left. + \alpha_g\frac{(\mathbf{v}\cdot\mathbf{r})^2}{c^2r^2}\right] - \frac{e_1e_2}{r}\left[1 - \alpha_p\frac{v^2}{c^2} + \alpha_p\frac{(\mathbf{v}\cdot\mathbf{r})^2}{c^2r^2}\right] \\
&+ \left(\alpha_g - \frac{1}{2}\right)\frac{G^2m_1m_2^2}{c^2r^2} + \alpha_p\frac{e_1^2e_2^2}{m_1c^2r^2} \\
&+ (1 - \alpha_g - \alpha_p)\frac{Ge_1e_2m_2}{c^2r^2} - \frac{Ge_2^2m_1}{2c^2r^2}. \qquad (33)
\end{aligned}
$$

The Lagrangian for a test charged particle (mass m_1 and charge e_1) in the field of a heavy large-charged particle (mass m_2 and charge e_2), where again $m_2 \gg m_1$ and $e_2 \gg e_1$ and $e_2^2m_1 \gg e_1^2m_2$, is given by

$$
L = -m_1c[-g_{00}c^2 - g_{ij}\dot{x}^i\dot{x}^j]^{1/2} - e_1A_0. \qquad (34)
$$

The Reissner–Nordström[9] solution for g_{00}, g_{ij}, and A_0 in this "one-body" Lagrangian is

$$
g_{00} = -\left(1 - \frac{2Gm_2}{c^2r} + \frac{Ge_2^2}{c^4r^2}\right), \qquad (35)
$$

$$
g_{ij} = \delta_{ij} - \left(1 + \frac{1}{g_{00}}\right)\frac{x^ix^j}{r^2}, \qquad (36)
$$

$$
A_0 = e_2/r. \qquad (37)
$$

The post-Newtonian expansion of Eq. (34) (apart from the rest energy term $-m_1c^2$) is Eq. (33) with $\alpha_g = 1$ and $\alpha_p = 0$. The coordinates are Schwarzschild coordinates.[8]

II. CONDITION OF STATIC BALANCE

For the Lagrangians of Eqs. (25) and (34) (which are in the center-of-mass system), we can write Lagrange's equations as

$$\mathbf{F} = \dot{\mathbf{P}}, \tag{38}$$

where

$$\mathbf{F} \equiv \frac{\partial L}{\partial \mathbf{r}} \quad \text{and} \quad \mathbf{P} \equiv \frac{\partial L}{\partial \mathbf{v}}. \tag{39}$$

If we now set[5]

$$e_i = \pm G^{1/2} m_i, \quad i = 1, 2 \tag{40}$$

(the notation is meant to imply that the + sign holds for all i or the − sign holds for all i) we find that $(L)_{\mathbf{v}=0} \equiv 0$, $-m_1c^2$ for Eqs. (25) and (34), respectively, and hence, $(\mathbf{F})_{\mathbf{v}=0} \equiv 0$. Thus, if the particles are at rest they will remain at rest. We note that if $\mathbf{v}$ is not zero the force will not be zero. The condition for static balance[5] of Eq. (40) holds for both the exact "one-body" problem of Eq. (34) and the post-Newtonian two-body problem of Eq. (25).

Let us now look more carefully at the post-Newtonian two-body problem of Eq. (25). In order to have static balance for all r the static $1/r$ terms and the static $1/r^2$ terms must independently cancel out in the Lagrangian of Eq. (25). We thus must have

$$e_1 e_2 = G m_1 m_2, \tag{41}$$

$$e_1^2 m_2 + e_2^2 m_1 = 2e_1 e_2 (m_1 + m_2) - G m_1 m_2 (m_1 + m_2). \tag{42}$$

Note that the $1/r^2$ terms proportional to α_g and α_p cancel out due to Eq. (41). Using Eq. (41) in Eq. (42) we obtain

$$m_1 e_2 (e_2 - e_1) = m_2 e_1 (e_2 - e_1), \tag{43}$$

which gives us the solution of Eq. (40) and also another solution for static balance, namely,

$$e_i = \pm (G m_1 m_2)^{1/2}, \quad i = 1, 2. \tag{44}$$

In the special case where $m_1 = m_2$ the two solutions of Eq. (40) and (44) become the same.

It should be noted that Eq. (44) is not a solution to the exact "one-body" problem of Eq. (34). This is as expected since we must have $e_2 \gg e_1$ and $m_2 \gg m_1$ for Eq. (34) to be valid.

We shall return to our discussion of static balance in Sec. IV.

III. EQUATION OF MOTION AND PRECESSION OF THE ORBIT

Lagrange's equations of motion for the Lagrangian of Eq. (26) may be written as

$$\dot{\mathbf{v}} + \frac{G'M\mathbf{r}}{r^3} = \frac{G'M}{c^2 r^3}\left[4k_4 G'M\mathbf{r}/r - k_5 v^2 \mathbf{r} + 4k_6(\mathbf{v}\cdot\mathbf{r})\,\mathbf{v}\right], \tag{45}$$

where

$$k_4 = \tfrac{3}{4}k_2 + \tfrac{1}{4}k_3 = 1 + \tfrac{3}{4}\mu/M + \tfrac{3}{4}Z_a + \tfrac{1}{4}(Z_b + Z_a Z_b - Z_a^2), \tag{46}$$

$$k_5 = \tfrac{3}{2}k_2 - \tfrac{1}{2}k_1 = 1 + \tfrac{5}{2}\mu/M + \tfrac{3}{2}Z_a, \tag{47}$$

$$k_6 = \tfrac{1}{4}k_1 + \tfrac{3}{4}k_2 = 1 - \tfrac{1}{4}\mu/M + \tfrac{3}{4}Z_a. \tag{48}$$

The secular results for the precession of the orbit can be written as[8]

$$\dot{E}_{av} = 0, \quad \dot{\mathbf{L}}_{av} = \boldsymbol{\Omega}^* \times \mathbf{L}, \quad \mathbf{A}_{av} = \boldsymbol{\Omega}^* \times \mathbf{A}, \tag{49}$$

where $E = \mu(\frac{1}{2}v^2 - G'M/r)$, $\mathbf{L} = \mu\mathbf{r}\times\mathbf{v}$, and $\mathbf{A} = \mu[\mathbf{v}\times(\mathbf{r}\times\mathbf{v}) - G'M\mathbf{r}/r]$ are the Newtonian energy, orbital angular momentum, and Runge–Lenz vector, respectively, and

$$\boldsymbol{\Omega}^* = \frac{(4k_6 - 2k_4 + k_5)\, G'M\bar{\omega}}{c^2 a(1-e^2)}\,\mathbf{n}, \tag{50}$$

where a is the semimajor axis, e is the eccentricity, $\bar{\omega}$ is the average orbital angular velocity, and $\mathbf{n}$ is a unit vector in the $\mathbf{L}$ direction. To insure a bound orbit we must have $G' > 0$. The result for $\boldsymbol{\Omega}^*$ can be cast into different forms by using the relations

$$\frac{L/\mu}{a^2(1-e^2)^{1/2}} = \left(\frac{G'M}{a^3}\right)^{1/2} = \frac{2\pi}{T} = \bar{\omega}, \tag{51}$$

where T is the orbital period.

Unfortunately, the expression

$$4k_6 - 2k_4 + k_5 = \tfrac{1}{2}k_1 + 3k_2 - \tfrac{1}{2}k_3 = (1 + Z_a)(3 - \tfrac{1}{2}Z_b) + \tfrac{1}{2}Z_a^2 \tag{52}$$

contained in $\boldsymbol{\Omega}^*$ cannot be expressed in a simpler form.

However, we do have four special cases in which the form of Eq. (50) and (52) can be put in a very simple form.

A. First case: $e_1 = e_2 = 0$

We then have $G' = G$, $Z_a = Z_b = 0$, $4k_6 - 2k_4 + k_5 = 3$ and thus

$$\boldsymbol{\Omega}^* = \frac{3GM\bar{\omega}}{c^2 a(1-e^2)}\,\mathbf{n}, \tag{53}$$

which is the result of Robertson.[8,10]

B. Second case: $G = 0$

We then have $G' = -e_1e_2/m_1m_2$, $Z_a = -1$, $Z_b = -(e_1^2 m_2 + e_2^2 m_1)/e_1e_2M$, $4k_6 - 2k_4 + k_5 = \frac{1}{2}$ and thus

$$\boldsymbol{\Omega}^* = \frac{|e_1e_2|\,\bar{\omega}/\mu}{2c^2 a(1-e^2)}\,\mathbf{n}, \tag{54}$$

where $e_1e_2 < 0$ because the orbit must be bound. Equation (54) is the two-body generalization of the "one-body" Sommerfeld[11] result and reduces to the Sommerfeld result under the large mass approximation $m_2 \gg m_1$.

C. Third case: $e_1 = 0$

We then have $G' = G$, $Z_a = 0$, $Z_b = e_2^2/Gm_2M$, $4k_6 - 2k_4 + k_5 = 3 - e_2^2/2Gm_2M$ and thus

$$\boldsymbol{\Omega}^* = \frac{(3GM - e_2^2/2m_2)\,\bar{\omega}}{c^2 a(1-e^2)}\,\mathbf{n}. \tag{55}$$

We also have a similar result for the case $e_2 = 0$.

D. Fourth case: $e_i = \pm(-1)^i G^{1/2} m_i$ or $e_i = \pm(-1)^i (G m_1 m_2)^{1/2}$

In these cases both the Newtonian gravitational and electric forces are equal and attractive. We then have $G' = 2G$, $Z_a = -\frac{1}{2}$, $Z_b = \frac{1}{2}$, $4k_6 - 2k_4 + k_5 = \frac{3}{2}$ and thus

$$\boldsymbol{\Omega}^* = \frac{3GM\bar{\omega}}{c^2 a(1-e^2)}\,\mathbf{n}. \tag{56}$$

The results of the first and fourth cases as given by Eqs. (53) and (56) are only superficially the same since $G'M = \bar{\omega}^2 a^3$ [see Eq. (51)] is different in each of these two cases.

E. Equations of motion with $e_i = \pm G^{1/2} m_i$ or $e_i = \pm (G m_1 m_2)^{1/2}$

Using any one of these conditions of static balance which imply $G' = 0$ and $G' Z_a = G' Z_b = G$, in Eq. (45) we obtain

$$\dot{\mathbf{v}} = \frac{GM}{c^2 r^3}\left[-\tfrac{3}{2}v^2\mathbf{r} + 3(\mathbf{v}\cdot\mathbf{r})\,\mathbf{v}\right], \tag{57}$$

which shows the acceleration $\dot{\mathbf{v}}$ will be zero when $\mathbf{v}$ is zero. Equation (57) has a solution for a bound circular orbit with any velocity ($v \ll c$ to be consistent with the post-Newtonian approximation) at $r = 3GM/2c^2$. However, Eq. (57) being a post-Newtonian approximation is *not* valid for r of the order of GM/c^2.

Equation (57) is in a form similar to the equation of motion of a photon (neglecting spin effects) in the gravitational field of the sun which can be written as

$$\dot{\mathbf{v}} = \frac{Gm_\odot}{c^2 r^3}\left[-2c^2\mathbf{r} + 4(\mathbf{v}\cdot\mathbf{r})\,\mathbf{v}\right], \tag{57p}$$

which gives the well-known deflection angle of

$$\theta_p = 4Gm_\odot/c^2 R, \tag{58p}$$

where $m_\odot$ is the mass of the sun and R is the distance of closest approach. Hence, the deflection angle for the similar problem with Eq. (57) would be

$$\theta = 3GM/c^2 R, \tag{58}$$

with the conditions that $R \gg GM/c^2$ and $v \ll c$.

IV. GENERALIZATION TO n BODIES

The post-Newtonian n-body Lagrangian with electric charge has also been given by Bażański.[16] The n-body generalization of the two-body coordinate transformation of Eqs. (3) and (4) is

$$\mathbf{r}_{iB} = \mathbf{r}_i - \sum_{\substack{j=1\\ i\neq j}}^{n} \mathbf{r}_{ij}\left(\alpha_g \frac{Gm_j}{c^2 r_{ij}} - \alpha_p \frac{e_i e_j}{m_i c^2 r_{ij}}\right), \tag{59}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and thus $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. Transforming the Bażański Lagrangian[16] which is a function of $\mathbf{r}_{iB}$ and $\mathbf{v}_{iB}$ into the new variables $\mathbf{r}_i$ and $\mathbf{v}_i$ we obtain

$$\begin{aligned}
L(n\text{-body};\alpha_g,\alpha_p) = {} & \sum_{i=1}^{n}\left\{\tfrac{1}{2}m_i v_i^2 + \tfrac{1}{8}m_i v_i^4/c^2\right\} + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{\frac{Gm_i m_j}{r_{ij}}\left[1 - \frac{(\mathbf{v}_i\cdot\mathbf{v}_j)}{2c^2} - \frac{(\mathbf{v}_i\cdot\mathbf{r}_{ij})(\mathbf{v}_j\cdot\mathbf{r}_{ij})}{2c^2 r_{ij}^2} + \left(\frac{3}{2} - \alpha_g\right)\frac{v_{ij}^2}{c^2}\right.\right. \\
& \left. + \alpha_g \frac{(\mathbf{v}_{ij}\cdot\mathbf{r}_{ij})^2}{c^2 r_{ij}^2}\right] - \frac{e_i e_j}{r_{ij}}\left[1 - \frac{(\mathbf{v}_i\cdot\mathbf{v}_j)}{2c^2} - \frac{(\mathbf{v}_i\cdot\mathbf{r}_{ij})(\mathbf{v}_j\cdot\mathbf{r}_{ij})}{2c^2 r_{ij}^2} - \alpha_p\frac{v_{ij}^2}{c^2} + \alpha_p\frac{(\mathbf{v}_{ij}\cdot\mathbf{r}_{ij})^2}{c^2 r_{ij}^2}\right] \\
& + \left(\alpha_g - \frac{1}{2}\right)\frac{G^2 m_i m_j M_{ij}}{c^2 r_{ij}^2} + \alpha_p\frac{e_i^2 e_j^2}{\mu_{ij}c^2 r_{ij}^2} + (1-\alpha_g-\alpha_p)\frac{Ge_i e_j M_{ij}}{c^2 r_{ij}^2} - \left.\frac{G(e_i^2 m_j + e_j^2 m_i)}{2c^2 r_{ij}^2}\right\} + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n} \\
& \times\left\{\frac{G^2 m_i m_j m_k}{c^2}\left[-\frac{1}{r_{ij}r_{ik}} - \frac{1}{r_{ji}r_{jk}} - \frac{1}{r_{ki}r_{kj}} + \alpha_g\left(\frac{1}{r_{ij}^2} + \frac{1}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\cdot\mathbf{r}_{ik}}{r_{ij}r_{ik}} + \alpha_g\left(\frac{1}{r_{ji}^2} + \frac{1}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}}{r_{ji}r_{jk}}\right.\right. \\
& \left. + \alpha_g\left(\frac{1}{r_{ki}^2} + \frac{1}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}r_{kj}}\right] + \alpha_p\left[\frac{e_i^2 e_j e_k}{m_i c^2}\left(\frac{1}{r_{ij}^2} + \frac{1}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\cdot\mathbf{r}_{ik}}{r_{ij}r_{ik}} + \frac{e_i e_j^2 e_k}{m_j c^2}\left(\frac{1}{r_{ji}^2} + \frac{1}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}}{r_{ji}r_{jk}}\right. \\
& \left. + \frac{e_i e_j e_k^2}{m_k c^2}\left(\frac{1}{r_{ki}^2} + \frac{1}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}r_{kj}}\right] + \frac{Gm_i e_j e_k}{c^2}\left[-\frac{1}{r_{ij}r_{ik}} + \frac{1}{r_{ji}r_{jk}} + \frac{1}{r_{ki}r_{kj}} - (\alpha_g+\alpha_p)\left(\frac{1}{r_{ij}^2} + \frac{1}{r_{ik}^2}\right)\frac{\mathbf{r}_{ij}\cdot\mathbf{r}_{ik}}{r_{ij}r_{ik}}\right] \\
& + \frac{Ge_i m_j e_k}{c^2}\left[+\frac{1}{r_{ij}r_{ik}} - \frac{1}{r_{ji}r_{jk}} + \frac{1}{r_{ki}r_{kj}} - (\alpha_g+\alpha_p)\left(\frac{1}{r_{ji}^2} + \frac{1}{r_{jk}^2}\right)\frac{\mathbf{r}_{ji}\cdot\mathbf{r}_{jk}}{r_{ji}r_{jk}}\right] \\
& + \frac{Ge_i e_j m_k}{c^2}\left[+\frac{1}{r_{ij}r_{ik}} + \frac{1}{r_{ji}r_{jk}} - \frac{1}{r_{ki}r_{kj}} - (\alpha_g+\alpha_p)\left(\frac{1}{r_{ki}^2} + \frac{1}{r_{kj}^2}\right)\frac{\mathbf{r}_{ki}\cdot\mathbf{r}_{kj}}{r_{ki}r_{kj}}\right]\Bigg\},
\end{aligned} \tag{60}$$

where $M_{ij} = m_i + m_j$ and $\mu_{ij} = m_i m_j / M_{ij}$. To obtain the above Lagrangian in B—EIH—D coordinates (the particular form given by Bażański[16]) one merely has to set $\alpha_g = \alpha_p = 0$. The above type of summation is not quite as compact as the form given by Bażański[16] but it does have the nice feature of avoiding permutations in the various combinations so that there is much less summing to do. For example, for the case of $n = 3$ there are only three i,j combinations in the double sum and one i,j,k combination in the triple sum. In the triple sum the terms proportional to $G^2 m_1 m_2 m_3$ agree with the field theory result of Hiida and Okamura[7] [see their Eqs. (3.3), (3.8), and (3.18)]. The velocity dependent terms in Eq. (60) are combined in the manner of Khan and O'Connell[5] while the velocity dependent terms in Eq. (1) are combined in the manner of Landau and Lifshitz.[6]

A. Condition of static balance

The generalization of Eq. (40) to n bodies is

$$e_i = \pm G^{1/2} m_i, \quad i = 1, \ldots, n. \tag{61}$$

Note that Eq. (44) cannot be generalized to n bodies. Using Eq. (61) in Eq. (60) we obtain

$$\begin{aligned} &\mathcal{L}(e_k = \pm G^{1/2} m_k, \text{ all } k) \\ &= \sum_{i=1}^{n} \left\{ \tfrac{1}{2} m_i v_i^2 + \tfrac{1}{8} m_i v_i^4 / c^2 \right\} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{G m_i m_j}{r_{ij}} \\ &\quad \times \left[\left(\tfrac{3}{2} - \alpha_g + \alpha_p \right) \frac{v_{ij}^2}{c^2} + (\alpha_g - \alpha_p) \frac{(\mathbf{v}_{ij} \cdot \mathbf{r}_{ij})^2}{c^2 r_{ij}^2} \right] . \end{aligned} \tag{62}$$

We also have

$$\mathbf{F}_i = \dot{\mathbf{P}}_i, \tag{63}$$

where

$$\mathbf{F}_i \equiv \frac{\partial L}{\partial \mathbf{r}_i} \quad \text{and} \quad \mathbf{P}_i \equiv \frac{\partial L}{\partial \mathbf{v}_i} . \tag{64}$$

In the case where all the velocities are equal ($\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_n$) we have

$$\mathbf{F}_i (e_k = \pm G^{1/2} m_k, \text{ all } k;\ \mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_n) = 0, \tag{65}$$

and thus the n bodies will move with constant velocity. A similar argument for the two-body problem (with $\alpha_g = \alpha_p = 0$) has been given by Khan and O'Connell.[5]

Das[17] has shown that $\sigma(\mathbf{r}) = \pm G^{1/2} \rho(\mathbf{r})$ [where $\sigma(\mathbf{r})$ is charge density and $\rho(\mathbf{r})$ is mass density] leads to the exact static solutions of Weyl,[18] Curzon,[19] Majumdar,[20] and Papapetrou.[21] We conclude that the static solution of Eq. (61) is exact to all orders. We do not know whether or not the static solution of Eq. (44) is exact beyond the post-Newtonian approximation. For exact stationary solutions see Refs. 22 and 23 and for exact solutions with magnetic charge see Refs. 24 and 25.

V. CONCLUSION

Starting with the Bażański post-Newtonian two body Lagrangian, we have shown that it is possible to construct a coordinate transformation (before going to center-of-mass coordinates) containing two arbitrary parameters α_g and α_p in such a manner as to produce a Hamiltonian consistent with that derived from quantum field theory and containing the same arbitrary parameters. In particular, we have introduced the new parameter x_p in the photon propagator which is treated in the same manner as the graviton propagator[7,8] with x_g. Thus, the electromagnetic and gravitational aspects of this paper are treated symmetrically. We have also generalized the above results to n bodies.

The condition for static balance,[5] $e_i = \pm G^{1/2} m_i$ has been examined and found to hold for the Reissner–Nordström "one-body" problem and also for the post-Newtonian n-body problem. We also know that this solution is exact to all orders. For the post-Newtonian two-body problem an alternate condition for static balance $e_i = \pm (G m_1 m_2)^{1/2}$ has been found. We do not know if this condition is exact beyond the post-Newtonian approximation. *It would be interesting to check this conditions in the post-post-Newtonian two-body problem.*

We have found the precession of the perihelion for the post-Newtonian two-body problem with charge and have looked at four special cases: The first case agrees with the result of Robertson[10] and the large-mass approximation of the second case agrees with the result of Sommerfeld.[11]

We have also looked at the post-Newtonian two-body equations of motion, where the condition of static balance has been used, and found a solution mathematically similar to the well-known one for the gravitational deflection of light by the sun.

ACKNOWLEDGMENT

One of us (R. F. O'C.) would like to express his thanks to Professor I. W. Roxburgh and to Professor J. T. Lewis for their kind hospitality at Queen Mary College and the Dublin Institute for Advanced Studies, respectively, where he stayed during some of the period of execution of this work. He would also like to thank the SRC for the award of a Senior Visiting Fellowship at Queen Mary College.

APPENDIX

When we first sought to find the correct form of Eqs. (3) and (4) we knew from previous work[7,8] that Eq. (5) had to be in the form

$$\mathbf{r}_B = \mathbf{r} \left(1 - \alpha_g \frac{GM}{c^2 r} + \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) . \tag{5}$$

We first tried the transformations

$$\mathbf{r}_{1B} = \mathbf{r}_1 \left(1 - \alpha_g \frac{GM}{c^2 r} + \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) , \tag{3a}$$

$$\mathbf{r}_{2B} = \mathbf{r}_2 \left(1 - \alpha_g \frac{GM}{c^2 r} + \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) , \tag{4a}$$

which are consistent with Eq. (5). However, besides being *inconsistent* with the field theory results (see Sec. I, part B), the Bażański Lagrangian in these new coordinates $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{v}_1$, $\mathbf{v}_2$ had a peculiar form which could *not* be expressed as a function of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{r}$. If we now were to change to a new coordinate system where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, $\mathbf{r}_{CM} = \nu_1 \mathbf{r}_1 + \nu_2 \mathbf{r}_2$ with $\nu_1 + \nu_2 = 1$ ($\mathbf{r}_1 = \mathbf{r}_{CM} + \nu_2 \mathbf{r}$, $\mathbf{r}_2 = \mathbf{r}_{CM} - \nu_1 \mathbf{r}$ in inverse form) the Lagrangian $\mathcal{L}(\mathbf{r}, \mathbf{v}; \mathbf{r}_{CM}, \mathbf{v}_{CM})$ would not be cyclic for the coordinate $\mathbf{r}_{CM}$. The corresponding momentum relations are $\mathbf{P} = \nu_2 \mathbf{P}_1 - \nu_1 \mathbf{P}_2$, $\mathbf{P}_{CM} = \mathbf{P}_1 + \mathbf{P}_2$ ($\mathbf{P}_1 = \nu_1 \mathbf{P}_{CM} + \mathbf{P}$, $\mathbf{P}_2 = \nu_2 \mathbf{P}_{CM} - \mathbf{P}$ in inverse form). Since this particular Lagrangian contains the coordinate $\mathbf{r}_{CM}$, its canonical momentum $\mathbf{P}_{CM}$ does *not* satisfy the relation $\dot{\mathbf{P}}_{CM} = 0$. However, $\mathbf{P}_{CM}$ *does* satisfy the relation

$$\dot{\mathbf{P}}_{CM} = \left(\alpha_g \frac{GM}{c^2 r} - \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) \frac{(\mathbf{v} \cdot \mathbf{r})}{r^2} \mathbf{P}_{CM}, \tag{A1}$$

so that we can still introduce center-of-mass coordinates where $\mathbf{P}_{CM} = 0$.

We next tried the transformations

$$\mathbf{r}_{1B} = \mathbf{r}_1 - \tfrac{1}{2} \mathbf{r} \left(\alpha_g \frac{GM}{c^2 r} - \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) , \tag{3b}$$

$$\mathbf{r}_{2B} = \mathbf{r}_2 + \tfrac{1}{2} \mathbf{r} \left(\alpha_g \frac{GM}{c^2 r} - \alpha_p \frac{e_1 e_2}{\mu c^2 r} \right) , \tag{4b}$$

which are again consistent with Eq. (5). While the Lagrangian resulting from these transformations *could* be expressed as a function of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{r}$, it was, nevertheless, *inconsistent* with the field theory results.

The correct transformations, Eqs. (3) and (4), which are consistent with Eq. (5), were next tried. The Lagrangian resulting from these transformations *can* be expressed as a function of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{r}$ and *is* consistent with the field theory results.

[1]S. Bażański, Acta Phys. Pol. **15**, 363 (1956).
[2]S. Bażański, Acta Phys. Pol. **16**, 423 (1957).
[3]A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. **39**, 65 (1938).
[4]C. G. Darwin, Philos. Mag. **39**, 537 (1920).
[5]A. R. Khan and R. F. O'Connell, Nature **261**, 480 (1976).
[6]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, fourth revised English edition (Pergamon, New York, 1975), see Secs. 65 and 106.
[7]K. Hiida and H. Okamura, Prog. Theor. Phys. **47**, 1743 (1972).
[8]B. M. Barker and R. F. O'Connell, Phys. Rev. D **12**, 329 (1975).
[9]H. Reissner, Ann. Phys. Leipzig **50**, 106 (1916); G. Nordstrøm, Proc. Kon. Ned. Akad. Wet **20**, 1238 (1918); see also J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), p. 398.
[10]H. P. Robertson, Ann. Math. **39**, 101 (1938).
[11]A. Sommerfeld, *Atomic Structure and Spectral Lines* (Dutton, New York, 1934), pps. 251—5.
[12]Technically speaking, these interactions are more involved than two-particle exchanges. See diagrams on p. 1594 of Ref. 13.
[13]Y. Iwasaki, Prog. Theor. Phys. **46**, 1587 (1971).
[14]B. M. Barker, S. N. Gupta, and R. D. Haracz, Phys. Rev. **149**, 1027 (1966).
[15]S. N. Gupta, Proc. Phys. Soc. London A**65**, 161, 608 (1952); Phys. Rev. **96**, 1683 (1954); Rev. Mod. Phys. **29**, 334 (1957); in *Recent Developments in General Relativity* (Pergamon, New York, 1962), p. 251; Phys. Rev. **172**, 1303 (1968); D **14**, 2596 (1976).
[16]S. Bażański, in *Recent Developments in General Relativity* (Pergamon, New York, 1962), p. 137. See n-body Lagrangian on p. 149.
[17]A. Das, Proc. R. Soc. London Ser. A **267**, 1 (1962).
[18]H. Weyl, Ann. Phys. **56**, 117 (1917).
[19]H. E. J. Curzon, Proc. Lond. Math. Soc. **23**, 477 (1925).
[20]S. D. Majumdar, Phys. Rev. **72**, 390 (1947).
[21]A. Papapetrou, Proc. R. Irish. Acad. **51**, 191 (1947).
[22]Z. Perjés, Phys. Rev. Lett. **27**, 1668 (1971).
[23]W. Israel and G. A. Wilson, J. Math. Phys. **13**, 865 (1972).
[24]L. Parker, Phys. Rev. Lett. **34**, 412 (1975).
[25]R. J. Adler, Phys. Rev. D **14**, 392 (1976).

On the bulk distribution functions and fluctuation theorems[a)]

Robert M. Ziff

Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545
(Received 11 April 1977)

A new integral relation linking the bulk distribution functions of a one phase fluid is given. It is related to the Ornstein–Zernike compressibility formula, which is also discussed and generalized. In addition, an expression is found for the probability that a region of volume V in an infinite system of given density and temperature contains exactly N particles. This enables the fluctuation–compressibility relation and its generalization to be calculated, without the use of the grand canonical ensemble.

I. INTRODUCTION

This paper is concerned with some general properties of the *bulk* distribution functions $\bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) = \bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s; \rho, T)$, especially their interrelation and their connection with the compressibility and also the local fluctuations in the density. These functions give the probability density that any s particles are at positions $\mathbf{r}_1 \cdots \mathbf{r}_s$, in an infinite (bulk) system in equilibrium at a given temperature T and average number density ρ. Assuming that the system is unordered (fluid) and homogeneous so that there is no separation of phases, it is well known that there are certain formal relations between the $\bar{n}_s$.[1] First of all there is the so-called product property, which states that if the s particles are divided into two groups (say $\mathbf{r}_1 \cdots \mathbf{r}_t$ and $\mathbf{r}_{t+1} \cdots \mathbf{r}_s$) that are separated infinitely far from each other, then $\bar{n}_s$ becomes exactly the product of the distribution functions of the two groups:

$$\bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) \to \bar{n}_t(\mathbf{r}_1 \cdots \mathbf{r}_t)\bar{n}_{s-t}(\mathbf{r}_{t+1} \cdots \mathbf{r}_s). \tag{1}$$

Especially removing one particle infinitely far from the rest causes $\bar{n}_s$ to become

$$\bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) \to \rho \bar{n}_{s-1}(\mathbf{r}_1 \cdots \mathbf{r}_{s-1}), \tag{2}$$

since $\bar{n}_1(\mathbf{r}) = \rho$. This gives therefore a procedure to find n_{s-1} knowing n_s. Secondly there are the so-called fluctuation theorems. These refer to the cluster (or irreducible) distribution functions $\bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s)$ which are defined in terms of the $\bar{n}_s$ according to the Ursell development:

$$\begin{aligned}
&\bar{n}_1(\mathbf{r}_1) = \bar{\chi}_1(\mathbf{r}_1), \\
&\bar{n}_2(\mathbf{r}_1\mathbf{r}_2) = \bar{\chi}_2(\mathbf{r}_1\mathbf{r}_2) + \bar{\chi}_1(\mathbf{r}_1)\bar{\chi}_1(\mathbf{r}_2), \\
&\bar{n}_3(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) = \bar{\chi}_3(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) + \bar{\chi}_2(\mathbf{r}_1\mathbf{r}_2)\bar{\chi}_1(\mathbf{r}_3) + \bar{\chi}_2(\mathbf{r}_2\mathbf{r}_3)\bar{\chi}_1(\mathbf{r}_1) \\
&\qquad + \bar{\chi}_2(\mathbf{r}_3\mathbf{r}_1)\bar{\chi}_1(\mathbf{r}_2) + \bar{\chi}_1(\mathbf{r}_1)\bar{\chi}_1(\mathbf{r}_2)\bar{\chi}_1(\mathbf{r}_3), \quad \text{etc.}
\end{aligned} \tag{3}$$

The product property for the $\bar{n}_s$ implies that the $\bar{\chi}_s$ vanish when the particles are divided into two or more groups that are separated infinitely far from each other—the so-called cluster property. Thus the integral of $\bar{\chi}_s$ over all space and all but one particle exists,[2] and according to the fluctuation theorems is assigned the following thermodynamic meaning:

$$\int d\mathbf{r}_2 \cdots \int d\mathbf{r}_s \bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) = \left[\prod_{n=1}^{s-1}\left(kT\rho\frac{\partial}{\partial p} - n\right)\right]\rho(p, T). \tag{4}$$

It evidently depends upon the isothermal compressibility $(\partial\rho/\partial p)_T/\rho$ and its higher derivatives. When $s = 2$, this is the Ornstein—Zernike compressibility formula; the above generalization to all s is due to Hemmer.[3]

Now, (4) implies a constraint on the interrelation of the $\bar{\chi}_s$, but only indirectly. It has commonly been believed that no direct integral relation exists between the s-particle functions and the $(s-1)$-particle ones, as there is in the case of a finite system. However, this is not so, since, as will be shown, the $\bar{\chi}_s$ satisfy

$$\left(kT\rho\frac{\partial}{\partial p} - s\right)\bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) = \int \bar{\chi}_{s+1}(\mathbf{r}_1 \cdots \mathbf{r}_{s+1})\, d\mathbf{r}_{s+1}, \tag{5}$$

where $\bar{\chi}_s$ is also a function of p and T. This appears to be a new result, although it turns out that it is related to a formula given by Lebowitz and Percus[4] that interrelates the distribution functions of a finite grand canonical system. When extended to the bulk limit their formula implies that the $\bar{n}_s$ satisfy

$$\begin{aligned}
&\left(kT\rho\frac{\partial}{\partial p} - s\right)\bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) \\
&\quad = \int \{\bar{n}_{s+1}(\mathbf{r}_1 \cdots \mathbf{r}_{s+1}) - \bar{n}_s(\mathbf{r}_1 \cdots \mathbf{r}_s)\bar{n}_1(\mathbf{r}_{s+1})\}\, d\mathbf{r}_{s+1}.
\end{aligned} \tag{6}$$

For $s = 1$ this is identical to (5), and for $s > 1$ they can be shown to be equivalent, using (3). Note that iterating (5) $(s-1)$ times gives (4), since $\bar{\chi}_1(\mathbf{r}_1) = \rho$.

Thus the pressure dependence of $\bar{\chi}_s$ is directly related to the integral of $\bar{\chi}_{s+1}$. Note that if $\bar{\chi}_s$ is known (in an interval of the pressure) but not $\bar{\chi}_{s+1}$, then this puts a constraint on the latter (and the succeding $\bar{\chi}_s$) that is more stringent than does (4). The cluster property gives absolutely no information about $\bar{\chi}_{s+1}$ from the $\bar{\chi}_1 \cdots \bar{\chi}_s$ (except, of course, that it must satisfy the cluster property itself).

Additional comments on (5) will be given at the end of the paper. In the following sections it will be proven in two ways. First it will be shown that (5) can be verified directly using the known fugacity expansion of the $\bar{\chi}_s$. However, such expansions are only believed to be

[a)]Work performed under the auspices of the U.S.E.R.D.A.

valid for low density (gaseous) systems, while it seems that (4) and (5) should apply to liquids as well. In the third section a derivation of (5) will be given which does not make use of these expansions, but requires the introduction of the grand canonical ensemble, following the usual derivation of (4). But, to derive (5) and (4) this way, a step must be made that involves the interchange of the order of two limiting processes, and, until it can be justified, this proof is not complete either. Thus the proper derivation of (4)—(6) for a liquid remains an open problem.

The relation (4) for $s=2$,

$$\int \bar{\chi}_2(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_2 = kT\rho\left(\frac{\partial\rho}{\partial p}\right)_T - \rho, \tag{7}$$

is actually a combination of the fluctuation—correlation relation (due to Ornstein and Zernike)[3]

$$\langle N^2\rangle - \langle N\rangle^2 - \langle N\rangle \sim V \int \bar{\chi}_2(\mathbf{r}_1\mathbf{r}_2)\, d\mathbf{r}_2 \tag{8}$$

(for $V\to\infty$), and the fluctuation—compressibility relation

$$\langle N^2\rangle - \langle N\rangle^2 \sim V kT\rho\left(\frac{\partial\rho}{\partial p}\right)_T, \tag{9}$$

which is a basic thermodynamic result that goes back to Gibbs.[5] Here $\langle N\rangle$ and $\langle N^2\rangle$ are the average and mean square average of the number of particles in a region of volume V that is part of an infinite (bulk) system. In the fourth section a direct derivation of both these relations will be given, without invoking the grand canonical ensemble, and will be generalized to all moments of the density distribution. For this calculation the following interesting expression for the probability that the region contains exactly N particles will be derived:

$$P(N, V) \sim \frac{1}{N!}\left[\left(\frac{\partial}{\partial u}\right)^N \exp\{(V/kT)[p(uz) - p(z)]\}\right]_{u=0}, \tag{10}$$

where z is the fugacity (or activity) defined in terms of the chemical potential μ by

$$z = \lambda^3 \exp(\mu/kT) \tag{11}$$

and λ is the thermal wavelength $\lambda^2 \equiv 2\pi\hbar^2/mkT$. It is a generalization to all N of the expression for $P(0, V)$ found by Green, and Kac and Luttinger.[6] More aspects of this formula will be discussed in that section.

II. FIRST PROOF OF (5)

Equation (5) can be verified directly using the following expression (fugacity expansion) of the bulk $\bar{\chi}_s$ [7]:

$$\bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) = \sum_{l=s}^{\infty} \frac{z^l}{(l-s)!} \int d\mathbf{r}_{s+1} \cdots \int d\mathbf{r}_l\, U(\mathbf{r}_1 \cdots \mathbf{r}_l), \tag{12}$$

where $U(\mathbf{r}_1 \cdots \mathbf{r}_l)$ are the Ursell functions which are defined in terms of the l-body interaction $\Phi(\mathbf{r}_1 \cdots \mathbf{r}_l)$ as follows: First define

$$W(\mathbf{r}_1 \cdots \mathbf{r}_l) = \exp[-\Phi(\mathbf{r}_1 \cdots \mathbf{r}_l)/kT)] \tag{13}$$

for a classical system, and

$$W(\mathbf{r}_1 \cdots \mathbf{r}_l) \equiv N!\, \lambda^{3N} \sum_E \exp(-E/kT) \left|\Psi_E(\mathbf{r}_1 \cdots \mathbf{r}_l)\right|^2 \tag{14}$$

(the "Slater sum") for a quantum one, for $l=1, 2, 3, \cdots$. The $U(\mathbf{r}_1 \cdots \mathbf{r}_l)$ are the cluster functions of these, defined as in (3):

$$W(\mathbf{r}_1) = U(\mathbf{r}_1), \quad W(\mathbf{r}_1\mathbf{r}_2) = U(\mathbf{r}_1\mathbf{r}_2) + U(\mathbf{r}_1)U(\mathbf{r}_2), \quad \text{etc.} \tag{15}$$

These are the same U's that appear in the fugacity (or cluster) expansions of the thermodynamic functions. In particular the pressure and density are given by

$$\frac{p}{kT} = \sum_{l=1}^{\infty} b_l z^l, \tag{15a}$$

$$\rho = \sum_{l=1}^{\infty} l b_l z^l, \tag{15b}$$

where

$$b_l = b_l(T) \equiv \frac{1}{l!} \int d\mathbf{r}_2 \cdots \int d\mathbf{r}_l\, U(\mathbf{r}_1 \cdots \mathbf{r}_l) \tag{16}$$

are the so-called cluster integrals. Note that (15b) is identical to (12) for $s=1$. Now, from (12) one can verify that

$$\left(z\frac{d}{dz} - s\right)\bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s) = \int \bar{\chi}_{s+1}(\mathbf{r}_1 \cdots \mathbf{r}_{s+1})\, d\mathbf{r}_{s+1} \tag{17}$$

and since it follows from (11) that

$$z\frac{d}{dz} = kT\frac{d}{d\mu} = kT\rho\frac{d}{dp} \tag{18}$$

(at constant T), then (17) is identical to (5). Then also (4) follows by iteration.

The problem with the above proof, as mentioned earlier, is that the fugacity expansions are only valid for a low density system. Also, it seems unnecessary that explicit use must be made of the molecular interaction Φ. The following proof suffers neither of these objections (but has limitations of its own). It also has the virtue that the result is deductively derived rather than just verified.

III. SECOND PROOF OF (5)

This proof begins by considering a finite system. For such a system of N particles in a volume V in thermal equilibrium with a heat reservoir at temperature T, the spatial configuration of the particles is described by the canonical distribution function $n^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_N) = n^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_N; N, V, T)$.[8] It is proportional to $W(\mathbf{r}_1 \cdots \mathbf{r}_N)$ of (13) or (14), being normalized to unity as follows:

$$\int_V d\mathbf{r}_1 \cdots \int_V d\mathbf{r}_N\, n^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_N) = 1. \tag{19}$$

The probability density that *any* s particles are at $\mathbf{r}_1 \cdots \mathbf{r}_s$ is given by

$$n_s^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_s) = n_s^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_s; N, V, T) \equiv [N!/(N-s)!] \int_V d\mathbf{r}_{s+1} \cdots \int_V d\mathbf{r}_N\, n^{\text{can}}(\mathbf{r}_1 \cdots \mathbf{r}_N) \tag{20}$$

when $V\to\infty$ with $N/V=\rho$, these become the bulk functions $\bar{n}_s$ mentioned in the introduction. It follows that the $n_s^{\rm can}$ are normalized according to

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\, n_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s)=N!/(N-s)! \tag{21}$$

and interrelated by

$$(N-s)n_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s)=\int_V n_{s+1}^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}. \tag{22}$$

It is an expression of this type that is desired for the bulk functions. However, (22) does not lead to any useful expression when $N\to\infty$ (with $N=\rho V$) as both sides diverge.

The distribution functions in the grand canonical ensemble (which allows for particles to exchange between V and the reservoir) are related to the canonical ones by

$$\begin{aligned} n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)&=n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s;\,z,\,V,\,T)\\ &\equiv\sum_{N=0}^{\infty}\frac{z^N Z(N)}{Z^{\rm gr}(z)}n_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s;\,N,\,V,\,T), \end{aligned} \tag{23}$$

where $Z(N)=Z(N,\,V,\,T)$ is the canonical partition function ($\times\lambda^{3N}$), and

$$Z^{\rm gr}(z)=Z^{\rm gr}(z,\,V,\,T)\equiv\sum_{N=0}^{\infty}z^N Z(N) \tag{24}$$

is the grand partition function. From (21) and (22) it follows that the $n_s^{\rm gr}$ are normalized according to

$$\begin{aligned} &\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\, n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)\\ &=\sum_{N=0}^{\infty}\frac{N!}{(N-s)!}\frac{z^N Z(N)}{Z^{\rm gr}(z)}\\ &\equiv\left\langle\frac{N!}{(N-s)!}\right\rangle^{\rm gr}, \end{aligned} \tag{25}$$

and interrelated by

$$\left(z\frac{d}{dz}+\langle N\rangle^{\rm gr}-s\right)n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)=\int_V n_{s+1}^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}. \tag{26}$$

First note that the fluctuation theorem (4) can be derived from (25). One can write

$$\begin{aligned} \left\langle\frac{N!}{(N-s)!}\right\rangle^{\rm gr}&=\langle(N-s+1)(N-s+2)\cdots(N-1)N\rangle^{\rm gr}\\ &=\frac{1}{Z^{\rm gr}}\left[\left(z\frac{d}{dz}-s+1\right)\left(z\frac{d}{dz}-s+2\right)\cdots\right.\\ &\quad\left.\times\left(z\frac{d}{dz}-1\right)z\frac{d}{dz}\right]Z^{\rm gr}, \end{aligned} \tag{27}$$

and this implies that

$$\begin{aligned} &\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\chi_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)\\ &=\left[\left(z\frac{d}{dz}-S+1\right)\cdots\left(z\frac{d}{dz}-1\right)z\frac{d}{dz}\right]\log Z^{\rm gr}, \end{aligned} \tag{28}$$

where $\chi_s^{\rm gr}$ are the cluster functions of the $n_s^{\rm gr}$, defined analogous to (3). That (28) follows from (25) and (27) can be verified directly for small s; a general proof is given in the Appendix. Since

$$\log Z^{\rm gr}\sim pV/kT \tag{29}$$

and

$$z\to\lambda^3\exp(\mu/kT) \tag{30}$$

as $V\to\infty$ with $\langle N\rangle^{\rm gr}/V=\rho$, where p and μ are respectively the pressure and chemical potential of a *bulk* system of density ρ and temperature T, and *assuming* that

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\chi_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)\sim V\int d\mathbf{r}_2\cdots\int d\mathbf{r}_s\,\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s) \tag{31}$$

as $V\to\infty$, then (4) follows from (28), using also the thermodynamic identity (18). On the right-hand side of (31), the integrals are over all space, and the bulk functions $\bar{\chi}_s$ are given by $\chi_s^{\rm gr}$ in the limit $V\to\infty$ with $\langle N\rangle^{\rm gr}/V=\rho$.

Similarly (5) can be derived from (26). Introducing the $\chi_s^{\rm gr}$, (26) implies that

$$\left(z\frac{d}{dz}-s\right)\chi_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)=\int_V\chi_{s+1}^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}, \tag{32}$$

as can be verified directly for small s.[9] Here there is no explicit appearance of the $\langle N\rangle^{\rm gr}$, and both sides remain finite when $V\to\infty$. Assuming that as $V\to\infty$,

$$\int_V\chi_{s+1}^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}\to\int\bar{\chi}_{s+1}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}, \tag{33}$$

and again using (18), then (5) follows.

The two assumptions made in (31) and (33) involve an interchange of the order of two limiting processes, namely the extension of the integral to cover all space, and the replacement of $\chi_s^{\rm gr}$ by $\bar{\chi}_s$. If the validity of this step could be proven then this proof would be complete. Note, however, that if the same interchange were applied to the canonical distribution functions, then the results would differ! For in that ensemble one has, from (21) and (22),

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\,\chi_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s)=(-1)^{s-1}(s-1)!N \tag{34}$$

and

$$-s\chi_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s)=\int\chi_{s+1}^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})\,d\mathbf{r}_{s+1}, \tag{35}$$

and when $V\to\infty$ [interchanging the order of the limits analogous to (31) and (33)], then (4) and (5) do not follow.[10] In particular the $\partial/\partial p$ terms are lacking. From the previous proof it seems that (4) and (5) must be correct, at least in certain circumstances, for which the above interchange must be valid in the grand canonical ensemble but not in the canonical one. Why this should be so is an intriguing (and apparently unanswered) question.[11]

The emphasis here has been on the introduction of the cluster functions χ_s. In fact this is not necessary. As has been shown by Lebowitz and Percus,[4] (26) can

be rewritten in the form

$$\left(z\frac{d}{dz}-s\right)n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)$$
$$=\int_V[n_{s+1}^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_{s+1})-n_s^{\rm gr}(\mathbf{r}_1\cdots\mathbf{r}_s)n_1^{\rm gr}(\mathbf{r}_{s+1})]\,d\mathbf{r}_{s+1} \quad (36)$$

since the integral over V of $n_1^{\rm gr}$ is just $\langle N\rangle^{\rm gr}$. Clearly both sides remain finite when $V\to\infty$, and this leads one to propose (6) for a bulk system. However, once again one must interchange two limits, and again doing the same in the canonical ensemble gives a different result. Of course, (6) is formally equivalent to (5). The virtue of (6) is that it applies directly to the $\bar{n}_s$, without invoking the $\bar{\chi}_s$.

IV. FLUCTUATIONS

In this section the general relation between the correlations (as expressed by $\bar{\chi}_s$) and the fluctuations of density in a region of a bulk system will be explored. The result will be a generalization of (8) and (9). Again, if the grand canonical ensemble were used, then such relations would follow almost immediately from (25) or (28), but these relations would be in terms of the integrals of the $\chi_s^{\rm gr}$ instead of $\bar{\chi}_s$. To avoid this difficulty, one can proceed as follows:

For a system of M particles in a region of volume Ω, described by the canonical ensemble, the probability that a subregion of volume V contains exactly N particles is given in terms of $n^{\rm can}$ and $n_s^{\rm can}$ by[12]

$$P(N,V;M,\Omega)=\frac{M!}{(M-N)!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_N\int_\Omega d\mathbf{r}_{N+1}\cdots$$
$$\times\int_\Omega d\mathbf{r}_M\left(\prod_{i=N+1}^{M}(1-\theta(\mathbf{r}_i))\right)$$
$$\times n^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_N;M,\Omega,T)$$
$$=\sum_{s=N}^{M}\frac{(-1)^{s-N}}{(s-N)!\,N!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s$$
$$\times n_s^{\rm can}(\mathbf{r}_1\cdots\mathbf{r}_s;M,\Omega,T), \quad (37)$$

where $\theta(\mathbf{r})$ is unity when $r\in V$ and zero otherwise. Note that for $s=0$ (when $N=0$), the "integral" must be taken to be unity. Let $\Omega\to\infty$ with $M/\Omega=\rho$; then one gets

$$P(N,V)=\sum_{s=N}^{\infty}\frac{(-1)^{s-N}}{(s-N)!\,N!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{n}_s(\mathbf{r}_1\cdots\mathbf{r}_s;\rho,T), \quad (38)$$

which gives the probability that V contains N particles in a bulk (infinite) system of average density ρ. This can also be written

$$P(N,V)=\frac{1}{N!}\left[\left(\frac{\partial}{\partial t}\right)^N\left(1+\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\right.\right.$$
$$\left.\left.\times\int_V d\mathbf{r}_s\bar{n}_s(\mathbf{r}_1\cdots\mathbf{r}_s)\right]\right]_{t=-1}, \quad (39)$$

and since the generating functions of the integrals of the integrals of the $\bar{n}_s$ and those of the $\bar{\chi}_s$ are related by the general relation[13]

$$1+\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{n}_s(\mathbf{r}_1\cdots\mathbf{r}_s)$$
$$=\exp\left(\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)\right) \quad (40)$$

then $P(N,V)$ is also given by

$$P(N,V)=\frac{1}{N!}\left[\left(\frac{\partial}{\partial t}\right)^N\exp\left(\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\right.\right.$$
$$\left.\left.\times\int_V d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)\right)\right]_{t=-1}. \quad (41)$$

For the generating function of $P(N,V)$, (39) and (41) lead to the two equivalent expressions

$$\sum_{N=0}^{\infty}\xi^N P(N,V)$$
$$=1+\sum_{s=1}^{\infty}\frac{(\xi-1)^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{n}_s(\mathbf{r}_1\cdots\mathbf{r}_s) \quad (42a)$$
$$=\exp\left(\sum_{s=1}^{\infty}\frac{(\xi-1)^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)\right). \quad (42b)$$

From the first of these it follows that

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{n}_s(\mathbf{r}_1\cdots\mathbf{r}_s)$$
$$=\left[\left(\frac{\partial}{\partial\xi}\right)^s\sum_{N=0}^{\infty}\xi^N P(N,V)\right]_{\xi=1}$$
$$=\sum_{N=0}^{\infty}\frac{N!}{(N-s)!}P(N,V)\equiv\left\langle\frac{N!}{(N-s)!}\right\rangle, \quad (43)$$

and from the second

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)$$
$$=\left[\left(\frac{\partial}{\partial\xi}\right)^s\log\sum_{N=0}^{\infty}\xi^N P(N,V)\right]_{\xi=1}. \quad (44)$$

The difference between these and the corresponding expressions in the grand canonical ensemble are strikingly little, as is particularly evident when contrasting (43) with (25): To find the average of $N!/(N-s)!$ in a region of volume V, one integrates the $n_s^{\rm gr}$ over V when the properties of that region are described by the grand canonical ensemble, and $\bar{n}_s$ when it is a piece of an infinite system.[1] Thus the grand canonical system and the region of an infinite system are only asymptotically identical as $V\to\infty$, since only in that limit is $n_s^{\rm gr}$ identical to $\bar{n}_s$.

For $s=1$, (43) or (44) gives $\langle N\rangle=\rho V$, since $\bar{\chi}_1(\mathbf{r})=\bar{n}_1(\mathbf{r})=\rho$. For $s=2$, they give

$$\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2\bar{\chi}_2(\mathbf{r}_1\mathbf{r}_2)=\langle N^2\rangle-\langle N\rangle^2-\langle N\rangle=\langle(\Delta N)^2\rangle-\langle N\rangle, \quad (45)$$

exactly for a region of volume V. As $V\to\infty$,

$$\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)\sim V\int d\mathbf{r}_2\cdots\int d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s) \quad (46)$$

[note in contrast to (31) it is $\bar{\chi}_s$ which appears on both sides], and so (8) follows. Then combining (8) with (7), the fluctuation—compressibility relation (9) follows.

The corresponding formulas for the integrals of the χ_s for higher s are implied by (44), but there seems to be no simple means of expressing it. To find the generalized fluctuation—compressibility relation, one can proceed as follows:

By virtue of the identity

$$\sum_{s=1}^{\infty}\frac{t^s}{s!}\int d\mathbf{r}_2\cdots\int d\mathbf{r}_s\bar{\chi}_s(\mathbf{r}_1\cdots\mathbf{r}_s)$$
$$=\sum_{s=1}^{\infty}\frac{t^s}{s!}\sum_{l=s}^{\infty}\frac{z^l}{(l-s)!}l!\,b_l$$
$$=\sum_{l=1}^{\infty}z^l b_l\sum_{s=1}^{l}\frac{l!\,t^s}{s!\,(l-s)!}$$
$$=\sum_{l=1}^{\infty}z^l b_l[(1+t)^l-1]$$
$$=[p(z+tz)-p(z)]/kT \tag{47}$$

[where use has been made of (12), (15a) and (16)], Eqs. (41) and (46) imply that

$$P(N,V)\sim\frac{1}{N!}\left[\left(\frac{\partial}{\partial t}\right)^N\exp\{(V/kT)[p(z+tz)-p(z)]\}\right]_{t=-1}$$
$$=\frac{1}{N!}\left[\left(\frac{\partial}{\partial u}\right)^N\exp\{(V/kT)[p(uz)-p(z)]\}\right]_{u=0} \tag{48}$$

as $V\to\infty$, as has been stated in (10). Indeed, this exactly agrees with the grand canonical ensemble since in the latter

$$P^{\mathrm{gr}}(N,V)=\frac{z^N Z(N)}{Z^{\mathrm{gr}}(z)}=\frac{1}{N!\,Z^{\mathrm{gr}}}\left[\left(\frac{\partial}{\partial u}\right)^N Z^{\mathrm{gr}}(uz)\right]_{u=0}, \tag{49}$$

and as $V\to\infty$, $\log Z^{\mathrm{gr}}(z)\sim p(z)V/kT$. Therefore, all of the following implications of (48) apply asymptotically to the grand canonical ensemble as well.

For $N=0$, (48) gives

$$P(0,V)\sim\exp(-pV/kT), \tag{50}$$

which is the result of Green, and Kac and Luttinger.[6] The expression for the probability that there is one particle in V,

$$P(1,V)\sim z\left(\frac{\partial p}{\partial z}\right)_{z=0}\frac{V}{kT}\exp(-pV/kT)$$
$$=b_1 z\,V\exp(-pV/kT), \tag{51}$$

involves the first cluster integral $b_1=1$. Likewise the probability for $N=2$ involves b_2 and can be written

$$2\frac{P(2,V)}{P(0,V)}-\left(\frac{P(1,V)}{P(0,V)}\right)^2\sim 4b_2 z^2 V. \tag{52}$$

The generalization can be found from the generating function of (48),

$$\sum_{N=0}^{\infty}\xi^N P(N,V)\sim\exp\{(V/kT)[p(\xi z)-p(z)]\}, \tag{53}$$

from which it follows that the *cumulants* Q_k of $N!\,P(N,V)/P(0,V)$, defined by

$$P(0,V)+\xi P(1,V)+\xi^2 P(2,V)\cdots$$
$$=P(0,V)\exp(\xi Q_1+\tfrac{1}{2}\xi^2 Q_2\cdots) \tag{54}$$

or

$$1+\sum_{N=1}^{\infty}\frac{\xi^N}{N!}\frac{N!\,P(N,V)}{P(0,V)}=\exp\sum_{k=1}^{\infty}\frac{\xi^k}{k!}Q_k \tag{55}$$

so that $Q_1=(51)$, $Q_2=(52)$, etc., are given by

$$Q_k\sim\left[\left(\frac{\partial}{\partial\xi}\right)^k\frac{p(\xi z)V}{kT}\right]_{\xi=0}=k\cdot k!\,b_k z^k V. \tag{56}$$

Thus a physical interpretation has been given to the cluster integrals b_l.

The above formulas describe the behavior of $P(N,V)$ for small N and large V. Consider now the behavior about the mean number ρV, which can itself be calculated from the generating function (53):

$$\langle N\rangle\sim\left[\xi\frac{d}{d\xi}\exp\{(V/kT)[p(\xi z)-p(z)]\}\right]_{\xi=1}$$
$$=\frac{Vz}{kT}\left(\frac{\partial p}{\partial z}\right)=\rho V. \tag{57}$$

In the same manner for the fluctuations one finds (9) directly. The generalization which also follows from (53) is that the cumulants C_k of the $\langle N^s\rangle$, defined by

$$1+\sum_{s=1}^{\infty}\frac{\xi^s}{s!}\langle N^s\rangle=\exp\left(\sum_{s=1}^{\infty}\frac{\xi^s}{s!}C_s\right) \tag{58}$$

so that $C_1=\langle N\rangle$, $C_2=\langle N^2\rangle-\langle N\rangle^2$, $C_3=\langle N^3\rangle-3\langle N^2\rangle\langle N\rangle+2\langle N\rangle^3$, etc., are given by

$$C_k\sim V\left(kT\rho\frac{\partial}{\partial p}\right)^{k-1}\rho(p,T). \tag{59}$$

This relates directly the compressibility and its higher derivatives with the distribution of particles in V.

For an ideal gas, where $p=\rho kT$ and $z=\rho$, (48) correctly predicts that $P(N,V)$ is given by the Poisson distribution

$$P(N,V)=(1/N!)(\rho V)^N\exp(-\rho V) \tag{60}$$

and (59) correctly gives that the cumulants C_k are all equal to ρv. Thus, (48) can be viewed as the corrections to the Poisson distribution for an interacting gas. The deviations are best described by using the relation (59) between the compressibility and the cumulants of the distribution.

V. FURTHER REMARKS

1. Again it should be emphasized that (4)—(6) have only been properly proven for a low density (gaseous) system. To complete the proof for a liquid, either the fugacity expansion of the $\bar{\chi}_s$ must be validated (or generalized) for that system, or, for the grand canonical proof, (33) must be justified. Likewise the derivation of the relations concerning the fluctuations and also the comments about the asymptotic equivalence of the grand canonical ensemble and open system in the fourth section depended upon the fugacity expansions and therefore have only been proven for the low density system while it appears that they too should apply to a liquid as well. Note that the identity (47) can also be proven directly using (4), so that if (4) is valid for a liquid then the fluctuation theorems are valid too, and *vice versa*.

2. In a two phase region (first order transition) the relations cannot hold, since $(\partial/\partial p)_T$ is infinite and the correlation integrals such as (8) are also infinite since $\bar{\chi}_s$ does not satisfy the cluster property (due to the separation of the phases).[14]

3. For simplicity only the distribution functions have been considered, but in fact (5) can be generalized to the reduced phase space distribution functions $f_s(\mathbf{r}_1 \cdots \mathbf{r}_s, \mathbf{p}_1 \cdots \mathbf{p}_s)$ for a classical system, and the reduced density matrices $\rho_s(\mathbf{r}_1' \cdots \mathbf{r}_s', \mathbf{r}_1'' \cdots \mathbf{r}_s'')$ for a quantum system. The former are defined in terms of the complete distribution function by

$$f_s(\mathbf{r}_1 \cdots \mathbf{r}_s, \mathbf{p}_1 \cdots \mathbf{p}_s) = \frac{N!}{(N-s)!} \int d\mathbf{r}_{s+1} \cdots \int d\mathbf{r}_N \times \int d\mathbf{p}_{s+1} \cdots \int d\mathbf{p}_N f(\mathbf{r}_1 \cdots \mathbf{r}_N, \mathbf{p}_1 \cdots \mathbf{p}_N), \tag{61}$$

and the latter are defined in terms of the complete density matrix by

$$\rho_s(\mathbf{r}_1' \cdots \mathbf{r}_s', \mathbf{r}_1'' \cdots \mathbf{r}_s'') = \frac{N!}{(N-s)!} \int_V d\mathbf{r}_{s+1} \cdots \int_V d\mathbf{r}_N \times \rho(\mathbf{r}_1' \cdots \mathbf{r}_s' \mathbf{r}_{s+1} \cdots \mathbf{r}_N; \mathbf{r}_1'' \cdots \mathbf{r}_s'' \mathbf{r}_{s+1} \cdots \mathbf{r}_N). \tag{62}$$

For the f_s introduce the cluster functions as follows:

$$f_1(\mathbf{r}_1\mathbf{p}_1) = \chi_1(\mathbf{r}_1\mathbf{p}_1),$$
$$f_2(\mathbf{r}_1\mathbf{r}_2\mathbf{p}_1\mathbf{p}_2) = \chi_2(\mathbf{r}_1\mathbf{r}_2\mathbf{p}_1\mathbf{p}_2) + \chi_1(\mathbf{r}_1\mathbf{p}_1)\chi_1(\mathbf{r}_2\mathbf{p}_2), \quad \text{etc.}, \tag{63}$$

and let $\bar{\chi}_s$ represent the functions in the bulk limit. Then in the same way as for (5) one can show that they are interrelated by

$$\left(kT\rho\frac{\partial}{\partial p} - s\right)\bar{\chi}_s(\mathbf{r}_1 \cdots \mathbf{r}_s, \mathbf{p}_1 \cdots \mathbf{p}_s) = \int d\mathbf{r}_{s+1} \int d\mathbf{p}_{s+1}\, \bar{\chi}_{s+1}(\mathbf{r}_1 \cdots \mathbf{r}_{s+1}, \mathbf{p}_1 \cdots \mathbf{p}_{s+1}). \tag{64}$$

Likewise define the cluster functions of the ρ_s as follows

$$\rho_1(\mathbf{r}_1', \mathbf{r}_1'') = \chi_1(\mathbf{r}_1', \mathbf{r}_1''),$$
$$\rho_2(\mathbf{r}_1'\mathbf{r}_2', \mathbf{r}_1''\mathbf{r}_2'') = \chi_2(\mathbf{r}_1'\mathbf{r}_2', \mathbf{r}_1''\mathbf{r}_2'') + \chi_1(\mathbf{r}_1', \mathbf{r}_1'')\chi_1(\mathbf{r}_2', \mathbf{r}_2''), \quad \text{etc.} \tag{65}$$

Then the bulk functions $\bar{\chi}_s$ are interrelated by

$$\left(kT\rho\frac{\partial}{\partial p} - s\right)\bar{\chi}_s(\mathbf{r}_1' \cdots \mathbf{r}_s', \mathbf{r}_1'' \cdots \mathbf{r}_s'') = \int d\mathbf{r}_{s+1}\, \bar{\chi}_{s+1}(\mathbf{r}_1' \cdots \mathbf{r}_s'\mathbf{r}_{s+1}, \mathbf{r}_1'' \cdots \mathbf{r}_s''\mathbf{r}_{s+1}). \tag{66}$$

For the ideal quantum mechanical gases (Bose and Fermi) the $\bar{\chi}_s(\mathbf{r}_1' \cdots \mathbf{r}_s', \mathbf{r}_1'' \cdots \mathbf{r}_s'')$ can be found, and one can verify that the above relation holds, and consequently that (4)—(6) also hold. Of course, for the Bose system this is just for the one phase region.

4. For the *condensed* ideal Bose gas there are many peculiarities that relate to these formulas.[15] Since there is no separation of the phases, the $\bar{\chi}_s$ (calculated by means of the canonical ensemble) satisfy the cluster property, yet they go to zero so slowly that their integrals do not exist. Thus both sides of (4) are infinite. On the other hand, the fluctuations in the density $\langle(\Delta N)^2\rangle/V^2$ [calculated by means of (45)] go to zero. The fluctuation—compressibility relation (9) therefore cannot be true. Since the integrals of the $\bar{\chi}_s$ do not exist, the identity (47) cannot be used and therefore the ensuing derivation of (9) is not valid. Furthermore, the asymptotic equivalence of the open subregion and the grand canonical ensemble is not anymore true, as the χ_s^{gr} and χ_s^{can} do not become the same function $\bar{\chi}_s$ when $V \to \infty$ (for $s \geq 2$)!

5. The usual way that $P^{\text{gr}}(N, V) \equiv z^N Z(N, V)/Z^{\text{gr}}(z)$ is discussed for large V is by writing

$$z \to \lambda^3 \exp[-\mu(\rho_0)/kT],$$
$$Z(N, V) \to \lambda^{-3N} \exp[-F(N, V)/kT], \tag{67}$$
$$Z^{\text{gr}}(z) \to \exp[p(\rho_0)V/kT],$$

where ρ_0 is the average density of particles in the reservoir and F is the Helmholtz free energy.[16] Then also taking $F \sim N\mu(\rho) - Vp(\rho)$, where $\rho = N/V$, one finds

$$P^{\text{gr}}(N, V) \approx \exp[-(V/kT)\{\rho[\mu(\rho) - \mu(\rho_0)] - p(\rho) + p(\rho_0)\}] = \exp[-(V/kT)\int_{\rho_0}^{\rho}[\mu(\rho') - \mu(\rho_0)]\,d\rho'] = \exp[-(N/kT)\int_{1/\rho_0}^{1/\rho}[p(\rho') - p(\rho_0)]\,d(1/\rho')]. \tag{68}$$

Apparently this is not the same as (48), which can be derived from P^{gr} by the procedure indicated in (49). First of all (68) has to be normalized; but then it is still different. For example, for an ideal gas it does not quite give the Poisson distribution. The problem with (67) is that $\mu(\rho_0)$ and $p(\rho_0)$ represent a finite grand canonical system, while $F = N\mu(\rho) - Vp(\rho)$ refers to a canonical system, and these p's and μ's are not the same functions. However, they should be the same asymptotically, and in fact one can show that to leading order in V the logarithm of the generating function of (68) agrees with (53) and consequently the cumulants of the distribution (68) agree with (59) to leading order in V.

6. Finally it should be noted that the Kirkwood superposition approximation

$$\bar{n}_3(\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3) \approx (1/\rho^3)\bar{n}_1(\mathbf{r}_1)\bar{n}_1(\mathbf{r}_2)\bar{n}_1(\mathbf{r}_3), \tag{69}$$

introduced to "cut off" the hierarchy, does not in general satisfy (5).

ACKNOWLEDGMENTS

The author would like to thank the following people for helpful discussions and suggestions concerning this work: Dr. Louis Goldstein, Dr. Albert Migliori, Professor E.G.D. Cohen, Professor Mark Kac, Professor Elliott Lieb, Professor George Stell, and Professor George E. Uhlenbeck.

APPENDIX

Proof of (28):[17] First note that

$$\left(z\frac{d}{dz}-s+1\right)\left(z\frac{d}{dz}-s+2\right)\cdots\left(z\frac{d}{dz}-1\right)z\frac{d}{dz} = z^s\left(\frac{d}{dz}\right)^s, \tag{A1}$$

which can be proven by induction. The generating functions of the integrals of the n_s^{gr} and χ_s^{gr} are related by

$$1+\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s n_s^{gr}(\mathbf{r}_1\cdots\mathbf{r}_s) = \exp\left\{\sum_{s=1}^{\infty}\frac{t^s}{s!}\int_V d\mathbf{r}_1\cdots\int_V d\mathbf{r}_s\chi_s^{gr}(\mathbf{r}_1\cdots\mathbf{r}_s)\right\} \tag{A2}$$

[compare with (40)]. Therefore, (28) is equivalent to (25) and (27) if

$$1+\sum_{s=1}^{\infty}\frac{t^s}{s!}\frac{1}{Z^{gr}(z)}z^s\left(\frac{d}{dz}\right)^s Z^{gr}(z) = \exp\left[\sum_{s=1}^{\infty}\frac{t^s}{s!}z^s\left(\frac{d}{dz}\right)^s\log Z^{gr}(z)\right]. \tag{A3}$$

But this is simply the identity

$$\frac{Z^{gr}(z+tz)}{Z^{gr}(z)} = \exp[\log Z^{gr}(z+tz)-\log Z^{gr}(z)], \tag{A4}$$

assuming that $Z^{gr}(z)$ is a continuous function of z between z and $z+tz$.

[1]See, for example, G.E. Uhlenbeck, P.C. Hemmer, and M. Kac, J. Math. Phys. 4, 229 (1963), especially p. 231.

[2]Of course, this assumes that $\bar{\chi}_s \to 0$ fast enough. Normally the range of the correlations is expected to be similar to the interaction length, so there is no problem for short range forces.

[3]P.C. Hemmer, Physica Norvegica 3, 9 (1968); L.S. Ornstein and F. Zernike, Proc. Akad. Sci. (Amsterdam) 17, 793—806 (1914). See also Ref. 1. An expression essentially equivalent to (4) has been derived by D.J. Vezzetti, J. Math. Phys. 16, 31 (1975).

[4]J.L. Lebowitz and J.K. Percus, J. Math. Phys. 4, 116 (1963), esp. p. 123.

[5]*The Collected Works of J. Willard Gibbs* (Yale U.P., New Haven, 1948), Vol. 2, pt. 1, p. 201. For other derivations of this formula, see H.A. Lorentz, *Les théories statistiques en thermodynamique* (Teubner, Leipzig, 1916), pp. 86—97; M. von Smoluchowski, Ann. Physik (Leipz.) 25, 205 (1908), and L.D. Landau and E.M. Lifshitz, *Statistical Physics* (Pergamon, London, 1959), pp. 350—55.

[6]M. Kac and J.M. Luttinger, J. Math. Phys. 14, 583 (1973); M.S. Green, "Some Applications of the Generating Functional of the Molecular Distribution Functions," in *Lectures in Theoretical Physics*, edited by Brittin, Downs, and Downs (Interscience, New York, 1961), Vol. III, pp. 195—220.

[7]See G.E. Uhlenbeck and G.W. Ford, "The Theory of Linear Graphs and Applications to the Theory of Virial Development of the Properties of Gases," in *Studies in Statistical Mechanics* (North-Holland, Amsterdam, 1962), p. 139, and also *Lectures in Statistical Mechanics* (Am. Math. Soc., Providence, R.I., 1963). It should be noted that for a finite system (12) holds exactly in the grand canonical ensemble (with $\bar{\chi}_s$ replaced by χ_s^{gr} and the integrals restricted to V) but that in the canonical ensemble for finite V there is no simple expression.

[8]See, for example, J. de Boer, Rep. Prog. Phys. 12, 305—74 (1948), Chap. II.

[9]A general proof that (32) follows from (26) can be made using the fugacity expansion of the χ_s^{gr} (see Footnote 7), although the formal equality of (26) and (32) does not depend upon the existence of such an expansion.

[10]Note that there is no doubt that for most systems χ_s^{can} becomes χ_s when $V\to\infty$, the same as found from χ_s^{gr}. One exception is the condensed ideal Bose—Einstein gas, for $s \geq 2$ (see Sec. V).

[11]See Hemmer, Ref. 3, and also Ref. 1, p. 234.

[12]This has been given by Vezzetti, Ref. 3. See also R.M. Ziff, G.E. Uhlenbeck, and M. Kac, Phys. Rep. C. (to be published), Sec. 2. A related development has also been given by Green, Ref. 6.

[13]This is the so-called first theorem of Mayer and follows from the definition of χ_s (3). See either article in Ref. 7; also Kac and Luttinger, Ref. 6, Eq. (1.8).

[14]Ref. 1, p. 239.

[15]See Ziff, Uhlenbeck, and Kac, Ref. 12, Sec. 2.

[16]See, for example, K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), p. 172. Note that (68) appears in the paper of Smoluchowski, Ref. 5, p. 210.

[17]Compare this with the arguments of Hemmer and Vezzetti, Ref. 3.

Markov systems of imprimitivity[a)]

Etang Chen

Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium[b)]
(Received 6 April 1977)

It is shown that Nelson's Markoff system is a Markov system of imprimitivity; and conversely, an ergodic Markov system of imprimitivity determines a Nelson's Markoff system. Markovian dynamics are also constructed on some Markov systems of imprimitivity.

1. INTRODUCTION

The constructive field theory has been highly developed in recent years. In particular, the probabilistic approach, after Nelson's work,[1] has been proved to be a powerful and fruitful way for Euclidean field theory.

In fact, probabilistic methods in quantum theory and quantum field theory have been introduced by many people, e.g., Kac, Mackey, Segal, etc. Among them, Mackey's probabilistic ideas for quantum field theory are less known. Indeed, Mackey showed us to look at Euclidean covariant theory of quantum fields as a problem of finding a general ergodic action of the Euclidean group on a probability space with an invariant probability measure.[2] In contrast to the transitivity for a finite system, where the imprimitivity theorem[3,4] is applicable, he has advocated the metric transitivity (or ergodicity as it is now commonly called) for an infinite system.

In the present paper, we shall investigate the relation between Mackey's scheme and Nelson's approach. Roughly speaking, we study Mackey's system of imprimitivity with Nelson's Markoff property, called the Markov system of imprimitivity. It turns out that a Nelson's Markoff system is indeed a Markov system of imprimitivity; and conversely, an ergodic Markov system of imprimitivity determines a Nelson's Markoff system. In analogy with Nelson's program, we construct a Markovian dynamic on a Markov system of imprimitivity. For an ergodic system, even a self-adjoint contraction semigroup can be established.

This paper is organized as follows. In Sec. 2 we first recall some notions on Borel spaces, and then define a Markov system of imprimitivity. We give an abstraction of Nelson's Markoff system in Sec. 3, and show its relation with the Markov system of imprimitivity (Propositions 3.1 and 3.3). In Sec. 4, we construct a Markovian dynamic on an arbitrary Markovian system of imprimitivity (Theorem 4.1). We show some properties of an ergodic Markov system of imprimitivity in Sec. 5. Finally, in Sec. 6, for an ergodic system, we construct a Markov process, similar to Euclidean field theory; the associated semigroup is self-adjoint and contraction (Theorem 6.2). We also show in this section the equivalence of Markoff property and the boundary positivity for an ergodic Markov system of imprimitivity (Theorem 6.1).

2. MARKOV SYSTEM OF IMPRIMITIVITY

We begin with some notions on Borel spaces. For details we refer to Varadarajan's book.[5]

By a *Borel space*, we mean a pair (S,Σ) of a set S and a Borel structure Σ, which is a σ-algebra of subsets of S. The elements of Σ are called *Borel sets*. We shall also refer to S itself as a Borel space.

Given two Borel spaces (S_1,Σ_1) and (S_2,Σ_2), map f of S_1 into S_2 is a *Borel map* if $f^{-1}(\Sigma_2)\subset\Sigma_1$. If f is bijective and $f^{-1}(\Sigma_2)=\Sigma_1$, then f is a *Borel isomorphism*; (S_1,Σ_1) and (S_2,Σ_2) are *Borel isomorphic*. In particular, if $S_1=S_2$, then an isomorphism f is called an *automorphism*.

A Borel space is *standard* if it is isomorphic with the subspace associated with a Borel subset of some complete separable metric space.

Suppose that G is a group which is at the same time a Borel space. G is a *Borel group* if the map $(g_1,g_2)\to g_1g_2^{-1}$ of $G\times G$ into G is Borel. G is a standard Borel group if G is a standard Borel group. A separable locally compact group is a standard Borel group.[5]

Given a Borel space S and a separable locally compact group G, S is a *G-space* if for each $g\in G$, there exists a Borel automorphism β_g of S, such that, if we write $\beta_g(s)=[s]g$:

(i) $[s]e=s$ for $s\in S$; here e is the identity of G.

(ii) $[s]g_1g_2=[[s]g_1]g_2$ for all $s\in S$ and all $g_1,g_2\in G$. S is a *Borel G-space* if the map $(s,g)\to[s]g$ of $S\times G$ into S is Borel.

Suppose (S,Σ) is a standard Borel space. A *projection valued measure* on S is a mapping $P:\Delta\to P_\Delta$ of the Borel subsets $\Delta\in\Sigma$ into the projections on some separable Hilbert space $\mathcal{H}(P)$ such that

(i) $P_{\Delta_1\cap\Delta_2}=P_{\Delta_1}P_{\Delta_2}$ for all $\Delta_1,\Delta_2\in\Sigma$,

(ii) $P_S=I$, $P_\phi=0$,

(iii) $P_\Delta=\sum_{i=1}^{\infty}P_{\Delta_i}$ whenever $\Delta=\cup_{i=1}^{\infty}\Delta_i$ with Δ_i mutually disjoint.

Two projection valued measures P and Q are *equivalent* if there exists a unitary map W of $\mathcal{H}(P)$ onto $\mathcal{H}(Q)$ such that $WP_\Delta W^{-1}=Q_\Delta$ for all $\Delta\in\Sigma$. For each P, there is an element $\xi\in\mathcal{H}(P)$ such that $P_\Delta\xi=0$ if and only if $P_\Delta=0$. Thus, the measure $\Delta\to(P_\Delta\xi,\xi)$ has exactly the null sets as the sets Δ for which $P_\Delta=0$. Therefore, every P has associated with it a unique measure class, which will be called the *measure class of P* and will be denoted by C^P.

[a)] Dedicated to Professor Günther Ludwig for the occasion of his 60th birthday.

[b)] Present address: Dept. de Matematica, Pontificia Universidade Catolica, Rio de Janeiro ZC-20, Brazil.

Given a separable locally compact group G, let $U: g \to U_g$ be a strongly continuous unitary representation of G on a Hilbert space $\mathcal{H}(U)$. We note that for the homomorphism $g \to U_g$ to be strongly continuous it is sufficient to show that $g \to (U_g \xi, \eta)$ are measurable on G for all $\xi, \eta \in \mathcal{H}(U)$.

Let S be a standard Borel space. We relate G and S by assuming that S is a Borel G-space. A *system of imprimitivity* on a separable Hilbert space $\mathcal{H}$ is a pair (P, U), consisting of a projection valued measure P on the class of Borel subsets of S with $\mathcal{H}(P) = \mathcal{H}$, and a strongly continuous representation U of G with $\mathcal{H}(U) = \mathcal{H}$, such that the relations

$$U_g P_\Delta U_g^{-1} = P_{[\Delta]g^{-1}} \tag{2.1}$$

are satisfied for all $g \in G$ and all $\Delta \in \Sigma$. (S, Σ) is called the *base* for the system of imprimitivity.

The measure class C^P of P is called the *measure class of the system*. If $P_\Delta = 0$, then $P_{[\Delta]g} = U_{g^{-1}} P_\Delta U_{g^{-1}}^{-1} = 0$ for all $g \in G$. Hence the measure class C^P of the system is invariant under the action of G on S.

Two systems of imprimitivity (P, U) acting on $\mathcal{H}$ and (P', U') acting on $\mathcal{H}'$ based on the same G-space S are *equivalent*, if there exists a unitary isomorphism W of $\mathcal{H}$ onto $\mathcal{H}'$ such that

$$U'_g = W U_g W^{-1} \quad \text{and} \quad P'_\Delta = W P_\Delta W^{-1}$$

for all $g \in G$ and all $\Delta \in \Sigma$.

Let X be a topological space, $\mathcal{J}$ a family of closed subsets of X containing X. We shall use a pair $(X, \mathcal{J})$ to denote a topological space. A Borel space (S, Σ) is *indexed* by $(X, \mathcal{J})$, if the Borel structure Σ is indexed by $\mathcal{J}$ in the following sense: For each $\Lambda \in \mathcal{J}$ there is $\Sigma_\Lambda \subset \Sigma$ such that, (a) for $\Lambda_1, \Lambda_2 \in \mathcal{J}$, $\Lambda_1 \subset \Lambda_2$ implies $\Sigma_{\Lambda_1} \subset \Sigma_{\Lambda_2}$, and (b) $\Sigma_X = \Sigma$. A system of imprimitivity (P, U) on $\mathcal{H}$ is indexed by a topological space $(X, \mathcal{J})$, if its base (S, Σ) is indexed by $(X, \mathcal{J})$.

Let $\Lambda \in \mathcal{J}$, we define $M_\Lambda = \{P_\Delta; \Delta \in \Sigma_\Lambda\}''$, which is an Abelian von Neumann algebra on $\mathcal{H}$.

Notice that $M = M_X$, and $M_\Lambda \subset M$. Let E_Λ be the conditional expectation from M onto M_Λ, i.e., E_Λ is a faithful normal projection of norm 1 from M onto M_Λ.

For each $\Lambda \in \mathcal{J}$, the complement (resp. the closure, the boundary) of Λ is denoted by Λ^c (resp. $\overline{\Lambda}, \partial\Lambda$).

A *Markov system of imprimitivity* (P, U) acting on $\mathcal{H}$ is a system of imprimitivity indexed by a topological space $(X, \mathcal{J})$ such that to each $\Lambda \in \mathcal{J}$

$$E_{\overline{\Lambda^c}}(P_\Delta) = E_{\partial\Lambda}(P_\Delta) \tag{2.2}$$

$\forall\, \Delta \in \Sigma_\Lambda$; here $P_\Delta \in M_\Lambda$. (2.2) is the *Markov property* of the system.[1] This is the property which enables the system to have Markovian behavior.

3. MARKOV SYSTEM OF IMPRIMITIVITY AND NELSON'S MARKOFF SYSTEM

In this section, we shall study a close connection between the Markov system of imprimitivity (MSI) defined in Sec. 2 and Nelson's Markoff system (NMS) in the constructive field theory.[1] First, we give an abstraction of NMS in the context of standard Borel spaces.

By a *measure* on a Borel space we mean a nonnegative set function which is countably additive on its Borel structure, and $+\infty$ is an allowed value. As usual, the *finiteness* and *σ-finiteness* of a measure on a Borel space can also be defined. The notions of *measurable sets* and *measurable functions* can also be defined in an ordinary way.

A measure μ on (S, Σ) is a *probability measure*, if $\mu(S) = 1$. A probability space is denoted by (S, Σ, μ) where (S, Σ) can be a Borel space, standard Borel space, G-space, etc. By *random variables* we mean complex valued Borel functions defined on (S, Σ). The set of all bounded Borel functions on (S, Σ) will be denoted by $\mathcal{B}(\Sigma)$, which is in fact the set of all bounded Σ-measureable functions for μ-almost everywhere. Similarly, for $\Sigma_1 \subset \Sigma$, $\mathcal{B}(\Sigma_1)$ is the set of all bounded Σ_1-measurable functions.

Given a separable locally compact group G and a Borel G-space (S, Σ), let μ be a G-invariant probability measure on (S, Σ), i.e., $\mu(\beta_g(\Delta)) = \mu([\Delta]g) = \mu(\Delta)$ for all $\Delta \in \Sigma$ and $g \in G$, which means that β_g is a measure preserving Borel automorphism of S. In the probability space (S, Σ, μ), if (S, Σ) is indexed by a topological space $(X, \mathcal{J})$, then for each $\phi \in \mathcal{B}(\Sigma)$ there is a unique conditional expectation $\tilde{E}_\Lambda(\phi)$ in $\mathcal{B}(\Sigma_\Lambda)$ for a $\Lambda \in \mathcal{J}$ such that $\int_S \tilde{E}_\Lambda(\phi)\psi d\mu = \int_S \phi\psi d\mu$ for all positive ψ in $\mathcal{B}(\Sigma_\Lambda)$. A *Nelson's Markoff system* (NMS) based on (S, Σ, μ) is a pair (ϕ, β), consisting of a Borel function ϕ in $\mathcal{B}(\Sigma)$ and a representation β of G by a measure preserving Borel automorphism β_g $(g \in G)$ of Borel G-space S, indexed by $(X, \mathcal{J})$, such that

$$(\phi \circ \beta_g)(s) = \phi([s]g) \tag{3.1}$$

for all $s \in S$ and $g \in G$, and

$$\tilde{E}_{\overline{\Lambda^c}}(\phi) = \tilde{E}_{\partial\Lambda}(\phi) \tag{3.2}$$

for all $\phi \in \mathcal{B}(\Sigma_\Lambda)$; here $\overline{\Lambda^c}$ and $\partial\Lambda$ are defined in Sec. 2.

In Nelson's original formulation, S is the Sobolev space $\mathcal{H}^{-1}(\mathbf{R}^d)$ or $\mathcal{D}(R^d)$, and $(X, \mathcal{J}) = (R^d, \mathcal{J})$.[1,6]

The following results show the relation between a MSI and a NMS.

Proposition 3.1: Each NMS determines a MSI.

Proof: Let (ϕ, β) be a NMS based on (S, Σ, μ). For each $\phi \in L^2(S, \Sigma, \mu)$, we define $P_\Delta \phi = \chi_\Delta \phi$, where χ_Δ is the characteristic function of Δ. Then P_Δ is a projection on $L^2(S, \Sigma, \mu)$. For $\Delta \in \Sigma_\Lambda$, χ_Δ is Σ_Λ-measurable, so is P_Δ. Therefore, by (3.2) we have for all $\Delta \in \Sigma_\Lambda$

$$E_{\overline{\Lambda^c}}(P_\Delta) = E_{\partial\Lambda}(P_\Delta),$$

which is (2.2); here E_Λ is the restriction of $\tilde{E}_\Lambda$ on $M_\Lambda = \{\chi_\Delta; \Delta \in \Sigma_\Lambda\}''$.

Moreover, we define $(U_g\phi)(s) = (\phi \circ \beta_g)(s) = \phi([s]g)$ for all $\phi \in L^2(S, \Sigma, \mu)$, $g \in G$. By routine calculation, $U_{g_1 g_2} = U_{g_1} U_{g_2}$, $U_e = 1$, and it is also easy to check that

$$U_g P_\Delta U_g^{-1} = P_{[\Delta]g^{-1}}$$

holds for $g \in G$, $\Delta \in \Sigma$. This is exactly (2.1).

By the definition of NMS, μ is G-invariant, hence U_g is unitary. As S is a Borel G-space, for $\phi_1, \phi_2 \in L^2(S, \Sigma, \mu)$, the map $(g,s) \to \phi_1([s]g)\phi_2(s)^*$ is Borel on $G \times S$, hence the map

$$g \to \langle U_g \phi_1, \phi_2 \rangle = \int_S \phi_1([s]g)\phi_2(s)^* d\mu(s)$$

is Borel on G. Thus, U is a strongly continuous representation of G on $L^2(S, \Sigma, \mu)$. Therefore, (P, U) is a system of imprimitivity acting on $L^2(S, \Sigma, \mu)$ and satisfying Markov property (2.2), i.e., (P, U) is a MSI.

In order to show the converse of Proposition 3.1, we need an additional notion on a system of imprimitivity.

Let (P, U) be a system of imprimitivity acting on H with the base (S, Σ). A vector $\Omega \in H$ is said to be *cyclic for* P if $\{P_\Delta \Omega; \Delta \in \Sigma\}$ is dense in H. For a system of imprimitivity (P, U) with a cyclic vector, we have the following lemma.

Lemma 3.2: If there is a cyclic vector of P, then $\{P_\Delta; \Delta \in \Sigma\}''$ is equivalent to $L^\infty(S, \Sigma, \nu)$ for some ν in the measure class C^P.

Proof: Since there is a cyclic vector for P, (P, U) is equivalent to a system of imprimitivity (P', U') acting on $H' = L^2(S, \Sigma, \nu)$ for some $\nu \in C^P$ [Ref. 5, Theorem 9.11]. Hence, P is equivalent to P', $P'_\Delta \phi = \chi_\Delta \phi$ for $\phi \in L^2(S, \Sigma, \nu)$; here χ_Δ is a characteristic function for the Borel set $\Delta \in \Sigma$ (modulo ν-null sets). Thus, each P_Δ for $\Delta \in \Sigma$ is equivalent to a characteristic function χ_Δ on $L^2(S, \Sigma, \nu)$. But, each projection in $L^\infty(S, \Sigma, \nu)$ is corresponding to a characteristic function for some Borel set (modulo null sets) in Σ. Therefore, P_Δ for $\Delta \in \Sigma$, is equivalent to a projection in $L^\infty(S, \Sigma, \nu)$. As $L^\infty(S, \Sigma, \nu)$ is generated by all its projections, we have, consequently, $\{P_\Delta; \Delta \in \Sigma\}''$ is equivalent to $L^\infty(S, \Sigma, \nu)$.

We note that for an ergodic system there is always a cyclic vector for P (see Sec. 6).

Furthermore, a similar result can also be obtained if $\{P_\Delta; \Delta \in \Sigma$ (modulo ν-null sets)$\}$ is full on probability space (S, Σ, ν) in the sense that Σ (modulo ν-null sets) is the smallest σ-algebra with respect to which each P_Δ is measurable.[6]

Proposition 3.3: Given a MSI (P, U) acting on H, if there is a cyclic vector for P, then (P, U) determines a NMS.

Proof: Let $M = \{P_\Delta; \Delta \in \Sigma\}''$ by the given assumption for P and Lemma 3.2. We may consider

$$M = L^\infty(S, \Sigma, \nu) \tag{3.3}$$

for some probability measure ν in C^P; indeed, $\nu_{\xi,\eta}(\Delta) = \langle P_\Delta \xi, \eta \rangle$ for $\xi, \eta \in H$, and $\Delta \in \Sigma$.

For each $\phi \in \mathcal{B}(\Sigma)$, we define a map

$$\phi \to A_\phi = \int_S \phi(s) d\nu_{\xi,\eta}(s). \tag{3.4}$$

Then $A_\phi \in L^\infty(S, \Sigma, \nu)$. From (3.3), A_ϕ is a weak-limit point of $\{P_{\Delta_i}\}(\Delta_i \in \Sigma)$. Furthermore, E_Λ is normal in the sense that $E_\Lambda(\text{w-lim} P_{\Delta_i}) = \text{w-lim}_i E_\Lambda(P_{\Delta_i})$. Here w-lim means the weak-limit. Hence $E_\Lambda(A_\phi) = \text{w-lim}_i E_\Lambda(P_{\Delta_i})$. By Markov property of (P, U), we thus have

$$E_{\overline{\Lambda^c}}(A_\phi) = E_{\partial\Lambda}(A_\phi) \tag{3.5}$$

for all $A_\phi \in M_\Lambda = \{P_\Delta; \Delta \in \Sigma_\Lambda\}''$.

Moreover, for $\phi \in \mathcal{B}(\Sigma)$ and $\xi, \eta \in H$,

$$\begin{aligned}\langle E_\Lambda(A_\phi)\xi, \eta \rangle &= \int \phi(s) d\langle E_\Lambda(P_\Delta)\xi, \eta \rangle \\ &= \int \phi(s) d(\nu_{\xi,\eta} \restriction \Sigma_\Lambda)(s),\end{aligned} \tag{3.6}$$

where $\nu \restriction \Sigma_\Lambda$ denotes the restriction of ν to Σ_Λ. On the other hand, by (3.4),

$$\langle A_{\tilde{E}_\Lambda(\phi)}\xi, \eta \rangle = \int (\tilde{E}_\Lambda \phi)(s) d(\nu_{\xi,\eta} \restriction \Sigma_\Lambda)(s);$$

here $\tilde{E}_\Lambda(\phi) \in \mathcal{B}(\Sigma_\Lambda)$ is the conditional expectation of $\phi \in \mathcal{B}(\Sigma)$. But, from the definition of conditional expectation,

$$\tilde{E}_\Lambda(\phi) = d\nu_\phi / d(\nu_{\xi,\eta} \restriction \Sigma_\Lambda),$$

here $\nu_\phi(\Delta) = \int_\Delta \phi d\nu_{\xi,\eta}$ for $\Delta \in \Sigma_\Lambda$. Hence,

$$\langle A_{\tilde{E}_\Lambda(\phi)}\xi, \eta \rangle = \int d\nu_\phi = \int \phi(s) d(\nu_{\xi,\eta} \restriction \Sigma_\Lambda)(s)$$

which is exactly (3.6). Therefore, we have shown

$$E_\Lambda(A_\phi) = A_{\tilde{E}_\Lambda(\phi)} \tag{3.7}$$

for all $\phi \in \mathcal{B}(\Sigma)$. Furthermore, $A_\phi = 0$ iff $\phi = 0$ for ν-almost everywhere. Thus, (3.5) implies

$$\tilde{E}_{\overline{\Lambda^c}}(\phi) = \tilde{E}_{\partial\Lambda}(\phi)$$

for all $\phi \in \mathcal{B}(\Sigma_\Lambda)$, i.e., ϕ satisfies Markoff property (3.2).

As (S, Σ) is a Borel G-space, there is a Borel automorphism β_g ($g \in G$) of S, such that $(\phi \circ \beta_g)(s) = (U_g \phi)(s) = \phi([s]g)$ for all $s \in S$ and $\phi \in \mathcal{B}(\Sigma)$.

Consequently, the pair (ϕ, β) shown above is a NMS.

4. MARKOVIAN DYNAMIC ON MSI (I)

In analogy with Nelson's theory,[1] we shall construct a Markovian dynamics on a given MSI.

Given a MSI (P, U) on H indexed by $(X, \mathcal{J})$, for a fixed closed subset $\Lambda \in \mathcal{J}$, let $M_\Lambda = \{P_\Delta; \Delta \in \Sigma_\Lambda\}''$, $M = \{P_\Delta; \Delta \in \Sigma\}''$, and let E_Λ be the conditional expectation from M onto M_Λ. Suppose that the time evolution of the system is given by a one-parameter group of automorphisms $\{\alpha_t; t \in \mathbb{R}\}$ of M. We denote by $\alpha_t P_\Delta = P_{[\Delta]t^{-1}}$, and Σ_{Λ_t} the σ-subalgebra of Borel sets $[\Delta]t^{-1}$ for $\Delta \in \Sigma$. Let M_{Λ_t} be the corresponding von Neumann algebra generated by P_Δ for $\Delta \in \Sigma_{\Lambda_t}$, and E_{Λ_t} the corresponding conditional expectation of M onto M_{Λ_t}. We define $\gamma_t P_\Delta$, for $0 \leqslant t < \infty$ and $P_\Delta \in M_{\partial\Lambda}$, by

$$\gamma_t P_\Delta = E_{\partial\Lambda} \alpha_t P_\Delta. \tag{4.1}$$

Theorem 4.1: Given a MSI (P, U) and γ_t defined by (4.1), then $\{\gamma_t; t \geqslant 0\}$ is a semigroup on $M_{\partial\Lambda}$ if $\Sigma_{\Lambda_s} \subset \Sigma_\Lambda$ for $s \geqslant 0$.

Proof: For $s, t \geqslant 0$, $\Delta \in \Sigma_{\partial\Lambda}$, we have

$$\gamma_s \gamma_t P_\Delta = E_{\partial\Lambda} \alpha_s [\gamma_t P_\Delta] \tag{4.2}$$

$$= E_{\partial\Lambda} \alpha_s [E_{\partial\Lambda} \alpha_t P_\Delta] \tag{4.3}$$

$$= E_{\partial\Lambda} E_{\partial\Lambda_s} [\alpha_{s+t} P_\Delta] \tag{4.4}$$

$$= E_{\overline{\Lambda^c}} [E_{\partial\Lambda_s}(\alpha_{s+t} P_\Delta)] \tag{4.5}$$

$$= E_{\overline{\Lambda^c}} E_{\overline{\Lambda_s^c}}[\alpha_{s+t} P_\Delta] \tag{4.6}$$

$$= E_{\overline{\Lambda^c}}[\alpha_{s+t} P_\Delta] \tag{4.7}$$

$$= E_{\partial\Lambda}[\alpha_{s+t} P_\Delta] \tag{4.8}$$

$$= \gamma_{s+t} P_\Delta . \tag{4.9}$$

Here (4.2) and (4.3) are due to (4.1); (4.4) is from the covariance of $M_{\partial\Lambda}$; (4.5) and (4.6) hold by the hypotheses and Markov property (2.2); (4.7) is due to $\Sigma_{\overline{\Lambda^c}} \subset \Sigma_{\overline{\Lambda_s^c}}$, (4.8) is again from Markov property (2.2), and (4.9) holds by (4.1).

The assumption in Theorem 4.1 holds exactly in Nelson's theory; namely, let $\Lambda = \{(\mathbf{x}, t) \in \mathbf{R}^{d-1} \times \mathbf{R}; t \geq 0\}$, $\Lambda^c = \{(\mathbf{x}, t) \in \mathbf{R}^{d-1} \times \mathbf{R};\ t < 0\}$, then $\partial\Lambda = \mathbf{R}^{d-1}$, and $\Lambda_s \subset \Lambda$ for $s \geq 0$.

Some other sufficient conditions for γ_t to be a semigroup have also been studied on noncommutative systems.[7]

As a remark, we note that γ_t defined in this way is a contraction semigroup, and it is strongly continuous if α_t is. However, the self-adjointness of γ_t cannot be obtained for an arbitrary MSI. Nevertheless, we shall show in Sec. 6, a self-adjoint semigroup can be established for some special MSI.

5. ERGODICITY OF MSI

A system of imprimitivity (P, U) is *G-ergodic* if every G-invariant Borel subset of S is either a set of P measure zero or the complement of such a set. We say a MSI is *G-ergodic* if its system of imprimitivity is G-ergodic. Hence, ergodic properties of a system of imprimitivity can be carried over to a MSI. In particular, a characterization of ergodicity by Macky[4] can be used to a MSI:

Proposition 5.1: A MSI, (P, U), is not G-ergodic if and only if there exists a Borel subset $\Delta_0 \in \Sigma$ such that P_{Δ_0} is neither zero nor one, and $P_{\Delta_0} U_g = U_g P_{\Delta_0}$ for all $g \in G$.

For a given closed subset Λ of $(X, \mathcal{J})$, let M_Λ, E_Λ be th the corresponding von Neumann algebra and conditional expectation from M onto M_Λ as given before. We define

$$V_g P_\Delta V_g^{-1} = E_\Lambda(U_g P_\Delta U_g^{-1}) \tag{5.1}$$

for all $\Delta \in \Sigma_\Lambda$.

Proposition 5.2: For a G-ergodic MSI, if $V_g P_\Delta V_g^{-1} = P_\Delta$ for $\Delta \in \Sigma_\Lambda$, then P_Δ is either zero or one.

Proof: By assumption and (5.1), we have $E_\Lambda(U_g P_\Delta U_g^{-1}) = P_\Delta$, which implies $E_\Lambda(P_{[\Delta]g^{-1}}) = P_\Delta = E_\Lambda(P_\Delta)$, since $\Delta \in \Sigma_\Lambda$. Hence $E_\Lambda(P_{[\Delta]g^{-1}} - P_\Delta) = 0$ and $P_{[\Delta]g^{-1}} = P_\Delta$ (by the faithfulness of E_Λ). Thus, from Proposition 5.1, $P_\Delta = 0$ or I.

A similar result for a R-ergodic NMS can also be found in Nelson's theory.[6] The R-ergodicity in NMS is the property which yields the uniqueness of vacuum in quantum field theory. The significance of an ergodic action on a system with infinite degree of freedom has been always emphasized by Mackey.[2]

As another application of Proposition 5.1, we show the preservation of G-ergodicity in a MSI. Similar to Sec. 4, the dynamic of a MSI, (P, U), is given by $\{\alpha_t; t \in \mathbf{R}\}$ a one-parameter group of automorphism of M. We assume that

$$\alpha_t(P_\Delta) = V_t P_\Delta V_t^{-1} \tag{5.2}$$

for all $\Delta \in \Sigma$ and $t \in \mathbf{R}$, i.e., α_t is unitarily implementable. Furthermore, we denote by (P_t, U) a new MSI, where the measure class of P is translated by time t in the sense of (5.2). Moreover, we say a MSI (P, U) is *dynamical invariant* if $V_t U_g = U_g V_t$ for all $g \in G$ and $t \in R$.

Proposition 5.3: Given a dynamical invariant MSI (P, U), if (P, U) is G-ergodic, so is (P_t, U) for all $t \in R$.

Proof: Suppose (P_t, U) is not G-ergodic, then by Proposition 5.1, there is a $\Delta_0 \in \Sigma$ such that

$$(V_t P_{\Delta_0} V_t^{-1}) U_g = U_g (V_t P_{\Delta_0} V_t^{-1})$$

for all $g \in G$ and all $t \in R$. By the hypothesis of dynamical invariance, $V_t P_{\Delta_0} U_g V_t^{-1} = V_t U_g P_{\Delta_0} V_t^{-1}$ for all $t \in R$ and all $g \in G$. Hence, $P_{\Delta_0} U_g = U_g P_{\Delta_0}$ for all $g \in G$. This is impossible by the G-ergodicity of (P, U).

The dynamical invariance holds in some physical systems, e.g., quantum and classical lattice systems[8,9]; in particular, one can show the existence of Markovian dynamics in these systems, and the Z-ergodicity of the systems is also preserved by this dynamic in both quantum and classical systems.

Moreover, under the assumption of dynamical invariance, the preservation of G-ergodicity has also been proved in a Markovian subdynamical system induced from a quantum dynamical system.[7]

6. MARKOVIAN DYNAMIC ON MSI (II)

As we have mentioned in Sec. 4, it is impossible to construct a self-adjoint semigroup under the general setting of a MSI. We shall show in this section, however, that a more concrete Markov dynamic can be established, if a MSI is G-ergodic.

It is known, if a MSI (P, U) on H is G-ergodic, then P is the direct sum of k replicas ($k = \infty, 1, 2, \cdots$) of some projection valued measures with a cyclic vector.[4] For simplicity, we shall take only $k = 1$ for a G ergodic MSI in this section, i.e., $H = L^2(S, \Sigma, \mu)$ for some $\mu \in C^P$, and P_Δ is the characteristic function on H for a Borel set Δ (c.f. Lemma 3.2 and Proposition 3.3).

For a fixed $\Lambda \in \mathcal{J}$, we shall construct a Markov process acting on a subspace of $\mathcal{E} = H \cap \mathcal{B}(\Sigma_\Lambda)$.

Let the time evolution of MSI (P, U) be given in such a way that S is also a Borel R-space, i.e., $t \to \beta_t$ ($t \in \mathbf{R}$) is a representation of R by Borel automorphism of the Borel R-space S. Suppose μ is invariant under time evolution, the unitary operator induced by β_t on H will be denoted by V_t.

Furthermore, we define a *ϑ-involution* on a MSI (P, U) based on (S, Σ) as follows: ϑ is an automorphism of S onto S such that

(i) $\vartheta^2 = 1$,

(ii) $\vartheta\beta_t = \beta_{-t}$ for $t \in \mathbf{R}$, and

(iii) $s \in \Sigma_\Lambda$ implies $\vartheta(s) \in \Sigma_{\overline{\Lambda}^c}$.

μ is also assumed to be ϑ-invariant, and the corresponding unitary operator on H will be denoted by ϑ again, i.e., for $\phi \in H$, $\vartheta\phi(s) = \phi(\vartheta(s))$, and notice that $\vartheta V_t \phi(s) = V_{-t}\phi(\vartheta(s))$.

We are now able to show the equivalence of Markov property (3.2) and the boundary positivity on a MSI.

Theorem 6.1: For all $\phi \in \mathcal{E}$, the boundary positivity

$$\int \overline{\vartheta\phi} \cdot \phi \, d\mu = \int |\tilde{E}_{\partial\Lambda}\phi|^2 \, d\mu \tag{6.1}$$

holds if and only if

$$\tilde{E}_{\overline{\Lambda}^c}(\phi) = \tilde{E}_{\partial\Lambda}(\phi) \tag{6.2}$$

and

$$\vartheta\tilde{E}_{\partial\Lambda}(\phi) = \tilde{E}_{\partial\Lambda}(\vartheta\phi) = \tilde{E}_{\partial\Lambda}(\phi). \tag{6.3}$$

Proof: We apply methods given by Nelsen and Hegerfeld.[10,11] Suppose (6.2) and (6.3) hold, by the definition of ϑ-involution, we have for $\phi \in \beta(\Sigma_\Lambda)$ implies $\overline{\vartheta\phi} \in \beta(\Sigma_{\overline{\Lambda}^c})$. Thus, for $\phi \in \mathcal{E}$

$$\begin{aligned}
\int (\overline{\vartheta\phi})\phi \, d\mu &= \int (\overline{\vartheta\phi})\tilde{E}_{\overline{\Lambda}^c}(\phi) d\mu \\
&= \int (\overline{\vartheta\phi})\tilde{E}_{\partial\Lambda}(\phi) d\mu \qquad \text{[by (6.2)]} \\
&= \int \tilde{E}_{\partial\Lambda}(\overline{\vartheta\phi})\tilde{E}_{\partial\Lambda}(\phi) d\mu \\
&= \int [\tilde{E}_{\partial\Lambda}(\overline{\phi})][\tilde{E}_{\partial\Lambda}(\phi)] d\mu \qquad \text{[by (6.3)]} \\
&= \int |\tilde{E}_{\partial\Lambda}(\phi)|^2 d\mu \geq 0,
\end{aligned}$$

which is the boundary positivity (6.1).

Conversely, suppose (6.2) holds, then for $\phi \in \mathcal{E}$,

$$\langle \vartheta\phi, \phi \rangle = \langle E_{\partial\Lambda}\phi, E_{\partial\Lambda}\phi \rangle \tag{6.4}$$

where $\langle \, , \, \rangle$ is the inner product on H, $E_{\partial\Lambda}$ is the corresponding projection of $\tilde{E}_{\partial\Lambda}$ on H with range $H \cap \beta(\Sigma_{\partial\Lambda})$. Similarly, E_Λ and $E_{\overline{\Lambda}^c}$ will also denote the corresponding projectors with range $\mathcal{E} = H \cap \beta(\Sigma_\Lambda)$ and $H \cap \beta(\Sigma_{\overline{\Lambda}^c})$, respectively.

From (6.4), we have $\langle \vartheta E_\Lambda \phi, E_\Lambda \phi \rangle = \langle E_{\partial\Lambda}\phi, E_{\partial\Lambda}\phi \rangle$, hence $\langle E_\Lambda \vartheta E_\Lambda \phi, \phi \rangle = \langle E_{\partial\Lambda}\phi, \phi \rangle$ for all $\phi \in \mathcal{E}$. Thus,

$$E_\Lambda \vartheta E_\Lambda = E_{\partial\Lambda}. \tag{6.5}$$

By the definition of ϑ, we have $\vartheta E_\Lambda \vartheta = E_{\overline{\Lambda}^c}$. Hence from (6.5) $E_{\overline{\Lambda}^c}E_\Lambda = \vartheta E_{\partial\Lambda}$ and $E_\Lambda E_{\overline{\Lambda}^c} = E_{\partial\Lambda}\vartheta$. However, the range of $E_{\overline{\Lambda}^c}E_\Lambda$ is $H \cap \beta(\Sigma) \cap \beta(\Sigma_{\overline{\Lambda}^c})$, thus, we have

$$E_{\overline{\Lambda}^c}E_\Lambda = E_\Lambda E_{\overline{\Lambda}^c} = \vartheta E_{\partial\Lambda} = E_{\partial\Lambda}\vartheta. \tag{6.6}$$

Moreover, as $E_{\partial\Lambda}$ is a projector, from (6.5), we have $(E_\Lambda \vartheta E_\Lambda)(E_\Lambda \vartheta E_\Lambda) = E_\Lambda \vartheta E_\Lambda$. Hence, $E_\Lambda E_{\overline{\Lambda}^c} E_\Lambda = \vartheta E_{\overline{\Lambda}^c} E_\Lambda$, but $E_\Lambda \geq E_{\overline{\Lambda}^c}E_\Lambda$, we have then $E_{\overline{\Lambda}^c}E_\Lambda = \vartheta E_{\overline{\Lambda}^c}E_\Lambda$. However, $\vartheta E_{\overline{\Lambda}^c}E_\Lambda = E_{\partial\Lambda}$ from (6.6), thus

$$E_{\overline{\Lambda}^c}E_\Lambda = \vartheta E_{\overline{\Lambda}^c}E_\Lambda = E_{\partial\Lambda}. \tag{6.7}$$

Combining (6.6) and (6.7),

$$E_{\overline{\Lambda}^c}E_\Lambda = E_{\partial\Lambda} = \vartheta E_{\partial\Lambda} = E_{\partial\Lambda}\vartheta$$

which is (6.2) and (6.3); indeed for $\phi \in \mathcal{E}$,

$$\tilde{E}_{\partial\Lambda}(\phi) = \vartheta\tilde{E}_{\partial\Lambda}(\phi) = \tilde{E}_{\partial\Lambda}(\vartheta\phi).$$

In order to construct a strongly continuous self-adjoint semigroup on a subspace of $\mathcal{E}$, we have to impose one additional assumption on the time evolution

$$V_t\mathcal{E} \subset \mathcal{E} \quad \text{for all } t \geq 0. \tag{6.8}$$

We notice that $t \to V_t$ is strongly continuous. Indeed, as S is Borel-R-space and μ is R-invariant, $V_t\phi(s) = \phi([s]t)$, and $(t,s) \to \phi_1([s]t)\phi_2(s)^*$ is Borel on $\mathbf{R} \times S$. Hence,

$$t \to \langle V_t\phi_1, \phi_2 \rangle = \int_S \phi_1([s]t)\phi_2(s)^* d\mu(s)$$

is Borel on R.

Our construction is mainly based on Glimm and Jaffe's method in constructive field theory.[12] Given an ergodic MSI (P, U) with ϑ-involution and time-evolution as defined above, then (6.2) holds from Proposition 3.3. Suppose (6.3) also holds, then from Theorem 6.1, we can define a positive sesquelinear form on $\mathcal{E}$,

$$\langle \phi, \psi \rangle_K \equiv \langle \vartheta\phi, \psi \rangle_{\mathcal{E}} = \int \overline{\vartheta\phi}(s)\psi(s) d\mu(s) \quad \text{for } \phi, \psi \in \mathcal{E}$$

and the corresponding seminorm $\|\phi\|_K = \langle \phi, \phi \rangle_K^{1/2}$ on $\mathcal{E}$. Let $N = \{\phi \in \mathcal{E}; \|\phi\|_K = 0\}$, then N is stable under time evolution. Indeed, for $\phi \in N$, $\|V_t\phi\|_K^2 = \langle \phi, V_{2t}\phi \rangle_K \leq \|\phi\|_K \|V_{2t}\phi\|_K = 0$, hence $V_t N \subset N$ for $t \geq 0$. The quotient space $\mathcal{E}/N$, equipped with a positive definite inner product $\langle \, , \, \rangle_K$, defines a pre-Hilbert space. The completion of $\mathcal{E}/N$ in the norm $\|\cdot\|_K$ will be denoted by K.

Let $\delta : \mathcal{E} \to \mathcal{E}/N$. As μ is ϑ-invariant, $\|\vartheta\phi\|_{\mathcal{E}} = \|\phi\|_{\mathcal{E}}$ for $\phi \in \mathcal{E}$. Hence, $\|\delta\phi\|_K = \langle \delta\phi, \delta\phi \rangle_K^{1/2} = \langle \vartheta\phi, \phi \rangle_{\mathcal{E}}^{1/2} \leq \|\vartheta\phi\|_{\mathcal{E}}^{1/2} \|\phi\|_{\mathcal{E}}^{1/2} = \|\phi\|_{\mathcal{E}}$, i.e., the canonical map δ is contraction.

By the hypothesis for $t \geq 0$, $V_t\mathcal{E} \subset \mathcal{E}$, and by the fact $V_t N \subset N$ shown above, we can define

$$\gamma_t \delta\phi = \delta V_t \phi \tag{6.9}$$

for $t \geq 0$, $\phi \in \mathcal{E}$. $\{\gamma_t; t \geq 0\}$ is defined on $\delta\mathcal{E}$, which is dense in K.

Theorem 6.2: Given an ergodic MSI (P, U) with ϑ-involution and time-evolution as defined above. Suppose (6.3) and (6.8) hold for ϑ-involution and time-evolution, respectively. Then $\{\gamma_t; t \geq 0\}$ in (6.9) can be extended to a strongly continuous self-adjoint contraction semigroup e^{-tH} on K, for some positive self-adjoint H.

Proof: For $\phi \in \mathcal{E}$ and $s, t \geq 0$, we have

$$\gamma_s\gamma_t\delta\phi = \gamma_s[\delta V_t\phi] = \delta V_s V_t \phi = \delta V_{s+t}\phi = \gamma_{s+t}\delta\phi.$$

Hence, for $\phi \in K$, $\gamma_s\gamma_t\phi = \gamma_{s+t}\phi$, i.e., γ_t is a semigroup.

Moreover, for $\phi \in \mathcal{E}$, $t \geq 0$,

$$\begin{aligned}
\langle \gamma_t\delta\phi, \delta\phi \rangle_K &= \langle \delta V_t\phi, \delta\phi \rangle_K = \langle \vartheta V_t\phi, \phi \rangle_{\mathcal{E}} = \langle V_{-t}\vartheta\phi, \phi \rangle_{\mathcal{E}} \\
&= \langle \vartheta\phi, V_t\phi \rangle_{\mathcal{E}} = \langle \delta\phi, \delta V_t\phi \rangle_K = \langle \delta\phi, \gamma_t\delta\phi \rangle_K
\end{aligned}$$

which shows the self-adjointness of γ_t.

By applying the same method in Ref. 12, the contraction of γ_t can be shown. Indeed, for $\phi \in \mathcal{E}$,

$$\begin{aligned}
\|\gamma_t\delta\phi\|_K &= \langle \gamma_t\delta\phi, \gamma_t\delta\phi \rangle_K^{1/2} \\
&= \langle \gamma_{2t}\delta\phi, \delta\phi \rangle_K^{1/2} \\
&\leq \|\gamma_{2t}\delta\phi\|_K^{1/2} \|\delta\phi\|_K^{1/2}.
\end{aligned}$$

Applying Schwartz' inequality in this way for n times,

$$\|\gamma_t\delta\phi\|_K \leq \|\gamma_{2^n t}\delta\phi\|_K^{1/2^n}\|\delta\phi\|_K^{1-1/2^n}$$

$$= \|\delta V_{2^n t}\phi\|_K^{1/2^n}\|\delta\phi\|_K^{1-1/2^n}.$$

As δ is contraction and $\|V_t\| = 1$, we have

$$\|\gamma_t\delta\phi\|_K \leq \|\phi\|_{\mathcal{E}}^{1/2^n}\|\delta\phi\|_K^{1-1/2^n}$$

Hence, $\|\gamma_t\delta\phi\|_K \leq \|\delta\phi\|_K$ as $n \to \infty$.

We note that V_t is strongly continuous on $\mathcal{E}$, and δ is a contraction on $\mathcal{E}$. Therefore, γ_t is strongly continuous on $\delta\mathcal{E}$. Moreover, γ_t is a contraction, thus, γ_t is strongly continuous on K.

Consequently, the closure of γ_t in (6.10) is a strongly continuous, self-adjoint contraction semigroup on K. Therefore, it can be written as e^{-tH} for some positive self-adjoint H.

As a final remark, for the unit vector 1 in $\mathcal{E}$, we have $\gamma_t(\delta 1) = \delta(V_t 1) = \delta 1$, hence $\delta 1$ is in the domain of H and $H(\delta 1) = 0$. $\delta 1$ is called the ground state of H in Ref. 12.

ACKNOWLEDGMENTS

The author would like to thank Professor G. W. Mackey and Professor A. Verbeure for helpful discussion and kind hospitality at Harvard University and Universiteit Leuven, respectively.

[1]E. Nelson "Construction of Quantum Fields from Markoff Fields," J. Funct. Anal. **12**, 97–112 (1973).

[2]G. W. Mackey, "Ergodic Actions of Euclidean Group and the Canonical Commutation Relation," in *Proceedings of the 1967 International Conference on Particles and Fields* (Interscience, New York, 1967).

[3]G. W. Mackey, "Imprimitivity for Representations of Locally Compact Groups," Proc. Nat. Acad. Sci. U.S.A. **35**, 537–45 (1949).

[4]G. W. Mackey, "Unitary Representations of Group Extensions. I," Acta Math. **99**, 265–311 (1958).

[5]V. S. Varadarajan, *Geometry of Quantum Theory* (van Nostrand Reinhold, New York, 1970), Vol. II.

[6]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, New Jersey, 1974).

[7]E. Chen, "Markovian Subdynamics in Quantum Dynamical Systems," J. Math. Phys. **17**, 1785–90 (1976).

[8]E. Chen, "Markov Processes in Quantum Lattice Systems," Rep. Math. Phys. (to appear).

[9]R. Holly, "Markovian Interactions Processes with Finite Range Interactions," Ann. Math. Stat. **43**, 1961–7 (1972).

[10]E. Nelson, "Probability Theory and Euclidean Field Theory," in *Constructive Quantum Field Theory*, edited by G. Velo and A. S. Wightman, Lecture Notes in Physics No. 25 (Springer, New York, 1973).

[11]G. C. Hegerfeldt, "From Euclidean to Relativistic Fields and on the Notion of Markoff Fields," Commun. Math. Phys. **35**, 155–71 (1974).

[12]J. Glimm and A. Jaffe, "Quantum Field Theory," Harvard's Lecture 1976 (unpublished monograph).

A general theory of linear finite-difference equations with a few applications

Alain J. Phares[a)]

Department of Physics and Astronomy, University of Montana, Missoula, Montana 59812
(Received 3 January 1977; revised manuscript received 28 March 1977)

Recently a new formalism has been developed giving the solutions in explicit form of multiterm linear homogeneous recursion relations with nonconstant coefficients. The basic idea was to relate this problem to the one of partitioning an interval into parts of given lengths. This idea is extensively used here to obtain the solutions of linear inhomogeneous difference equations. The resulting method has the advantage of being general: It does not rely on any special device and does not assume any special form for the recursion relation. Applications of these techniques to physical problems are presented elsewhere. Here we show how the method works for first and second order equations. The three-term Legendre polynomial recursion relation with an arbitrary inhomogeneous term is discussed in detail. The Legendre polynomials are then viewed as special cases of combinatorics functions, $P(j,m;z)$, based on the partitions of an interval (j,m), $0\leq j\leq m$, into parts of lengths 1 and 2. $P(j,m;z)$ reduces to the Legendre polynomial, $P_m(z)$, for $j=0$. It is also interesting to note that $P(j,m;z)$ are not orthogonal in general. However, a new set of orthogonal functions can be constructed as the solutions of the inhomogeneous Legendre recursion relation with a suitable choice of the inhomogeneous term.

I. INTRODUCTION

In a recent publication[1] the solutions of linear homogeneous recursion relations were obtained via a new class of combinatorics functions. The recursion relation was conveniently written as

$$b_m=\sum_{k=1}^{N} f_{a_k}(m)b_{m-a_k}\,,\quad m>j_0\,, \tag{1.1}$$

where the mth term, b_m, is related to N terms of lower rank, $b_{m-a_1},\ldots,b_{m-a_N}$. The numbers $a_1,\ldots,a_N$ are positive integers assumed to be ordered

$$0<a_1<\cdots<a_N\,. \tag{1.2}$$

The coefficient multiplying the term b_{m-a_k} is denoted by $f_{a_k}(m)$ and is an arbitrary function of the rank m.

In general, Eq. (1.1) is valid for $m>j_0$ for some integer j_0. Then the values $b_{j_0},b_{j_0-1},\ldots,b_{j_0+1-a_N}$ are arbitrary initial conditions taken as

$$b_{j_0-i}=\lambda_i\,,\quad i=0,1,\ldots,a_N-1\,. \tag{1.3}$$

The solution of Eq. (1.1) with the arbitrary initial conditions (1.3) was shown[1] to be related to the problem of partitioning an interval (m_1,m_2) into parts of lengths $a_1,a_2,\ldots,$ and a_N.

Let us consider a possible partition, labeled by q, of (m_1,m_2) into n parts, $\delta_1,\delta_2,\ldots,\delta_n$,

$$\delta_i\in\{a_1,a_2,\ldots,a_N\} \tag{1.4a}$$

whose first part δ_1 is constrained according to

$$\delta_1>d\,. \tag{1.4b}$$

It is possible to represent this partition, (nq), by a functional, $F_n^q(m_1,m_2)$, depending on the set of functions $\{f_{a_k};k=1,\ldots,N\}$ in the following manner[1]:

$$F_n^q(m_1,m_2)=\prod_{i=0}^{n} f_{\delta_i}(s_i)\,. \tag{1.5a}$$

The s_i are related to the δ_i through

$$s_i-s_{i-1}=\delta_i\,,\quad s_0=m_1\,,\quad s_n=m_2 \tag{1.5b}$$

and

$$\delta_0=0\,,\quad f_0(m)\equiv 1\,. \tag{1.5c}$$

Whenever the interval (m_1,m_2) cannot be partitioned into n parts with the constraint (1.4b), then we set $q=0$, and the associated functional is set to be identically zero,[1]

$$F_n^0(m_1,m_2)\equiv 0\,. \tag{1.5d}$$

One step further is to associate all the constrained partitions of (m_1,m_2) into n parts, a quantity called the constrained combinatorics function of the first kind.[1] This function is obtained by summing the functionals $F_n^q(m_1,m_2)$ over all the partitions q into n parts, i.e.,

$$\bar{C}_1(m_1,m_2,n,d)=\sum_q F_n^q(m_1,m_2)\,, \tag{1.6}$$

where again d refers to the constraint on the first part.

On the other hand, summing the functionals $F_n^q(m_1,m_2)$ over all possible partitions, we obtain a representation of the interval (m_1,m_2) called the constrained combinatorics function of the second kind[1]

$$\bar{C}_2(m_1,m_2,d)=\sum_{n,q} F_n^q(m_1,m_2) \tag{1.7a}$$

or, according to Eq. (1.6),

$$\bar{C}_2(m_1,m_2,d)=\sum_n \bar{C}_1(m_1,m_2,n,d)\,. \tag{1.7b}$$

An interval (m_1,m_2) can be partitioned into n parts belonging to the set $\{a_k;k=1,\ldots,N\}$ provided there exists a set of nonnegative integers $\{p_k;k=1,\ldots,N\}$ satisfying the relations

$$\sum_{i=1}^{N} a_i p_i=m_2-m_1\,, \tag{1.8a}$$

$$\sum_{i=1}^{N} p_i=n\,. \tag{1.8b}$$

a) Present address: Department of Physics, Villanova University, Villanova, PA 19085.

When $m_2 = m_1$, Eq. (1.8a) can still be satisfied provided the p's are all zero. This leads through Eq. (1.8b) to $n = 0$. Thus, an interval of length zero can be partitioned into zero parts. This remark with the help of Eqs. (1.5), (1.6), and (1.7) leads to[1]

$$F_0^1(m_1, m_1) = \overline{C}_1(m_1, m_1, 0, d) = \overline{C}_2(m_1, m_1, d) = 1 . \quad (1.9)$$

Furthermore, if $m_2 < m_1$ there are no p's satisfying Eqs. (1.8a) and (1.8b), and we therefore have[1]

$$\overline{C}_1(m_1, m_2, n, d) = \overline{C}_2(m_1, m_2, d) = 0 \quad \text{for} \quad m_2 < m_1 . \quad (1.10)$$

We will now state the main results of Ref. 1: A particular solution of Eq. (1.1) is the combinatorics function $\overline{C}_2(l-j, m, j)$ corresponding to the initial condition $\lambda_i = \delta_{ij}$, and for $m > l$, we have

$$\overline{C}_2(l-j, m, j) = \sum_{k=1}^{N} f_{a_k}(m)\, \overline{C}_2(l-j, m-a_k, j) \quad \text{for} \quad m > l \quad (1.11)$$

thus leading to the general solution

$$b_m^{(h)} = \sum_{i=0}^{a_N - 1} \lambda_i \overline{C}_2(j_0 - i, m, i) . \quad (1.12)$$

These techniques were sucessfully applied by Antippa and Phares[2,3] to give the solutions for the eigenvalue problems related to the linear potential,[2] and the combined Coulomb and linear potential,[3] of importance in the charmonium model[4] of the ψ/J particles.[5] There, the partitioning subintervals have lengths 2 and 3, or 1, 2, and 3, thus corresponding to a three- or four-term recursion relation. Also in Refs. 2 and 3, the energy eigenvalue problem is related to the coefficient b_m expressed in terms of higher order coefficients taken as initial conditions. In other words, we write Eq. (1.1) as

$$b_m = \sum_{k=1}^{N} g_{c_k}(m) b_{m-c_k} , \quad 0 < c_1 < \cdots < c_N . \quad (1.13)$$

The c's are simply related to the a's through

$$c_N = a_N , \quad c_{N-k} = a_N - a_k \quad \text{for} \quad k = 1, \ldots, N-1 \quad (1.14)$$

and the g's are related to the f's according to

$$g_{c_N}(m) = 1/f_{a_N}(m + a_N) , \quad (1.15a)$$

$$g_{c_k}(m) = -f_{a_{N-k}}(m + a_N)/f_{a_N}(m + a_N) \quad \text{for } k = 1, \ldots, N-1 . \quad (1.15b)$$

In Appendix C of Ref. 2, b_m is shown to be given in terms of higher order coefficients, say, $b_{k_0}, b_{k_0+1}, \ldots, b_{k_0+a_N-1}$, for $k_0 > m$, as

$$b_m^{(h)} = \sum_{i=0}^{a_N - 1} \overline{C}_2^*(k_0 + i, m, i) b_{k_0+i} . \quad (1.16)$$

$\overline{C}_2^*(m_2, m_1, d)$ were called conjugate combinatorics functions of the second kind. These are related to the g's in a manner analogous to that relating the $\overline{C}_2$'s to the f's

$$\overline{C}_2^*(m_2, m_1, d) = \sum_{n,q} G_n^q(m_2, m_1) \quad (1.17a)$$

and

$$G_n^q(m_2, m_1) = \prod_{i=1}^{n} g_{\delta_i'}(s_i') , \quad (1.17b)$$

where

$$s_0' = m_2 , \quad s_i' - s_{i-1}' = -\delta_i' , \quad s_n' = m_1 \quad (1.17c)$$

and

$$\delta_i' > d , \quad \delta_i' \in \{c_k; k = 1, \ldots, N\} . \quad (1.17d)$$

The number d always refers to the constraint on the first part, starting from the lower or upper end of the interval (m_1, m_2). The constraints $\delta_1 > d$ or $\delta_1' > d$ are automatically satisfied for $a_1 > d$ or $c_1 > d$, in which cases the corresponding combinatorics functions become unconstrained. In the subsequent sections we will be dealing with the special case $d = 0$. Thus, for shorthand writing, it is convenient to set

$$\overline{C}_2(m_1, m_2, d = 0) \equiv C_2(m_1, m_2) , \quad (1.18a)$$

$$\overline{C}_1(m_1, m_2, n, d = 0) \equiv C_1(m_1, m_2, n) . \quad (1.18b)$$

Analogous notations will also be used for the unconstrained conjugate combinatorics functions.

The purpose of this work is to further extend the techniques of Refs. 1 and 2 and obtain the solutions of inhomogeneous linear multiterm recursion relations with *nonconstant* coefficients. This is done in Secs. II and III by adding to the general solution of the homogeneous relation a particular solution of the complete difference equation. Two expressions for a general term b_m are obtained. The first one connecting b_m to lower order coefficients taken as initial conditions, the second one connecting b_m to higher order terms.

The particular solution of the complete inhomogenous relation we search for is the one with initial conditions corresponding to the vanishing of all the lowest or highest order coefficients that are chosen to be $\{b_{j_0-i}; i = 0, 1, \ldots, a_N - 1\}$, or $\{b_{k_0+i}; i = 0, 1, \ldots, c_N - 1\}$, respectively. This particular solution is shown to be related in a very simple manner to the special (unconstrained) combinatorics functions of the second kind $(d = 0)$.

In Sec. IV we show how our method works in the simple case of the complete linear equation of the first order, and apply it to the cases of constant coefficients and Bernouilli and Euler polynomial recursion relations.

Finally in Sec. V a detailed discussion is carried out for one second-order equation, the Legendre polynomial recursion relation with an arbitrary inhomogeneous term.

II. INHOMOGENEOUS DIFFERENCE EQUATIONS

A. Particular solution

The general multiterm inhomogeneous linear difference equation with nonconstant coefficients is obtained by adding to the right-hand side of Eq. (1.1) an arbitrary function of the rank $m, I(m)$, thus leading to

$$b_m = I(m) + \sum_{k=1}^{N} f_{a_k}(m) b_{m-a_k} , \quad m > j_0 . \quad (2.1)$$

In a first step we search for the particular solution of Eq. (2.1) satisfying the initial conditions

$$b_{j_0-i} = 0 \quad \text{for } i = 0, 1, \ldots, a_N - 1 . \quad (2.2)$$

This solution is given in the form of a theorem.

Theorem 1: The solution of Eq. (2.1) satisfying the initial conditions (2.2) is

$$b_m^{(I)} = \sum_{i=j_0+1}^{m} I(i) C_2(i,m)\,, \tag{2.3}$$

where $C_2(i,m)$ are the special (unconstrained) combinatorics functions of the second kind

$$C_2(i,m) \equiv \bar{C}_2(i,m,0)\,. \tag{2.4}$$

Proof of Theorem 1: Three cases will be treated separately, $m \leq j_0$, $m = j_0+1$, and $m > j_0+1$.

(i) *The case* $m \leq j_0$: In this case since i varies in the range $[j_0+1, m]$, then $m < i$. Consequently, according to Eq. (1.10),

$$C_2(i,m) = 0 \quad \text{for } i \in [j_0+1, m] \tag{2.5}$$

thus implying that

$$b_m^{(I)} = 0 \quad \text{for} \quad m \leq j_0\,. \tag{2.6}$$

Equation (2.6) shows that (2.3) does satisfy the initial conditions given by Eq. (2.2).

(ii) *The case* $m = j_0+1$: In this case Eq. (2.3) reduces to

$$b_{j_0+1}^{(I)} = I(j_0+1) C_2(j_0+1, j_0+1)\,. \tag{2.7}$$

According to Eqs. (1.9) and (2.4), we have

$$C_2(m,m) = 1 \quad \text{for any } m \tag{2.8}$$

thus leading to

$$b_{j_0+1}^{(I)} = I(j_0+1)\,. \tag{2.9}$$

Equation (2.9) simply shows that Theorem 1 is trivially satisfied.

(iii) *The case* $m > j_0+1$: In this case, it is possible, in Eq. (2.3), to single out the term corresponding to $i = m$, so that

$$b_m^{(I)} = I(m) + \sum_{i=j_0+1}^{m-1} I(i) C_2(i,m)\,. \tag{2.10}$$

The condition $m > l$, for which Eq. (1.11) holds, is satisfied by the speical combinatorics functions, $C_2(i,m)$, that appear in Eq. (2.10). This follows from Eq. (2.4), the fact that $l = i$ and $j = 0$ [l and j as defined in Eq. (1.11)], and that i is strictly less than m. Thus, combining Eqs. (2.4), (2.10), and (1.11), we obtain

$$b_m^{(I)} = I(m) + \sum_{i=j_0+1}^{m-1} I(i) \sum_{k=1}^{N} f_{a_k}(m) C_2(i, m-a_k)\,. \tag{2.11}$$

We then exchange the summations over i and k and obtain $b_m^{(I)}$ as

$$b_m^{(I)} = I(m) + \sum_{k=1}^{N} f_{a_k}(m) \sum_{i=j_0+1}^{m-1} C_2(i, m-a_k)\,. \tag{2.12}$$

Now, according to Eq. (1.2), we have

$$m-1 \geq m-a_k \quad \text{for} \quad k = 1, \ldots, N\,. \tag{2.13}$$

The equality in Eq. (2.13) may only occur at $k = 1$ if and only if $a_1 = 1$. Excluding for the moment this special case, it is then possible to split the summation over i appearing in Eq. (2.12) into two parts as follows:

$$b_m^{(I)} = I(m) + \sum_{k=1}^{N} f_{a_k}(m) \Bigg\{ \sum_{i=j_0+1}^{m-a_k} C_2(i, m-a_k) + \sum_{i=m-a_k+1}^{m-1} C_2(i, m-a_k) \Bigg\}\,. \tag{2.14}$$

When i is in the range $[m-a_k+1, m-1]$, then we necessarily have $m - a_k < i$, and, according to Eq. (1.10), $C_2(i, m-a_k) = 0$. Thus, Eq. (2.14) reduces to

$$b_m^{(I)} = I(m) + \sum_{k=1}^{N} f_{a_k}(m) \sum_{i=j_0+1}^{m-a_k} C_2(i, m-a_k)\,. \tag{2.15}$$

It is obvious that Eq. (2.15) also holds for $a_1 = 1$. Finally, using the hypothesis (2.3), the summation over i in Eq. (2.15) is nothing but $b_{m-a_k}^{(I)}$, thus proving that $b_m^{(I)}$ is a solution of the difference equation (2.1). This completes the proof of Theorem 1.

B. General solution

The problem at hand is to give an explicit expression for the coefficients b_m, in terms of the arbitrary functions $f_{a_k}(m)$, and an arbitrary set of parameters $\{\lambda_i; i = 0, 1, \ldots, a_N - 1\}$ specifying arbitrary initial conditions

$$b_{j_0-i} = \lambda_i \quad \text{for} \quad i = 0, 1, \ldots, a_N - 1\,. \tag{2.16}$$

We will give the general solution of Eq. (2.1) satisfying the initial conditions (2.16) again in the form of a theorem.

Theorem 2: The solution of the difference equation (2.1) satisfying the arbitrary initial conditions (2.16) is obtained by adding to the particular solution $b_m^{(I)}$, Eq. (2.3), the solution $b_m^{(h)}$ of the associated homogeneous difference equation with the same initial conditions (2.16),

$$b_m^{(h)} = \sum_{i=0}^{a_N-1} \lambda_i \bar{C}_2(j_0 - i, m, i)\,. \tag{2.17}$$

Proof of Theorem 2: The proof of this theorem is done in two steps. We first show that $b_m = (b_m^{(h)} + b_m^{(I)})$ satisfies the recursion relation (2.1), and then that it also satisfies the initial conditions (2.16).

(i) By hypothesis $b_m^{(h)}$ and $b_m^{(I)}$ satisfy the difference equations

$$b_m^{(h)} = \sum_{k=1}^{N} f_{a_k}(m)\, b_{m-a_k}^{(h)}\,, \tag{2.18}$$

$$b_m^{(I)} = \sum_{k=1}^{N} f_{a_k}(m)\, b_{m-a_k}^{(I)} + I(m)\,. \tag{2.19}$$

Adding up Eqs. (2.18) and (2.19) we obtain

$$b_m^{(h)} + b_m^{(I)} = \sum_{k=1}^{N} f_{a_k}(m)(b_{m-a_k}^{(h)} + b_{m-a_k}^{(I)}) + I(m)\,. \tag{2.20}$$

Accordingly, $b_m^{(h)} + b_m^{(I)}$ is a solution of Eq. (2.1).

(ii) Let $b_m = b_m^{(h)} + b_m^{(I)}$ and set $m = j_0 - j$ with $j = 0, 1, 2, \ldots, a_N - 1$. According to Theorem 4 of Ref. 1, we have

$$b_{j_0-j}^{(h)} = \lambda_j \tag{2.21}$$

and, by construction,

$$b_{j_0-j}^{(I)} = 0 \tag{2.22}$$

so that

$$b_{j_0-j}=b^{(h)}_{j_0-j}+b^{(I)}_{j_0-j}=\lambda_j\,. \tag{2.23}$$

Therefore, $b_m=b^{(h)}_m+b^{(I)}_m$, i.e.,

$$b_m=\sum_{i=0}^{a_N-1}\lambda_i\bar{C}_2(j_0-i,m,i)+\sum_{i=j_0+1}^{m}I(i)C_2(i,m)\,, \tag{2.24}$$

is the solution of the inhomogeneous difference equation (2.1), satisfying the initial conditions (2.16). This completes the proof of Theorem 2.

III. INVERTED INHOMOGENEOUS DIFFERENCE EQUATIONS

The inverted inhomogeneous linear difference equation is obtained by adding to the right-hand side of Eq. (1.13) an arbitrary function of the rank m, $J(m)$,

$$b_m=J(m)+\sum_{k=1}^{N}g_{c_k}(m)b_{m+c_k}\,,\quad 0<c_1<c_2<\cdots<c_N\,. \tag{3.1}$$

The higher order terms to be considered as initial conditions are

$$b_{k_0+i}=\lambda'_i\ \text{ for }\ k_0>m\ \text{ and }\ i=0,1,\ldots,c_N-1\,. \tag{3.2}$$

Again, the general solution of Eq. (3.1) satisfying the initial conditions (3.2) will be given in the form of two theorems.

Theorem 3: A particular solution of Eq. (3.1) satisfying the set of initial conditions

$$\lambda'_i=0\ \text{ for }\ i=0,1,\ldots,c_N-1 \tag{3.3}$$

is given by

$$b^{(J)}_m=\sum_{i=m}^{k_0-1}J(i)C^*_2(i,m)\,, \tag{3.4}$$

where $C^*_2(i,m)$ are the special (unconstrained) conjugate combinatorics functions of the second kind defined to be

$$C^*_2(i,m)\equiv\bar{C}^*_2(i,m,0)\,. \tag{3.5}$$

Theorem 4: The solution of the recursion relation (3.1) satisfying the arbitrary initial conditions (3.2) is obtained by adding to the particular solution $b^{(J)}_m$, Eq. (3.4), the solution $b^{(h)}_m$ of the associated homogeneous recursion relation with the same initial conditions (3.2)

$$b^{(h)}_m=\sum_{i=0}^{c_N-1}\lambda'_i\bar{C}^*_2(k_0+i,m,i),\quad m<k_0\,. \tag{3.6}$$

The proofs of Theorems 3 and 4 are very similar to those of Theorems 1 and 2. This is why we shall skip them, beyond remarking that they are based on the following properties[1,2]:

$$\text{(i)}\ \bar{C}^*_2(m_2,m_1,d)=0\ \text{ for }\ m_1>m_2\,, \tag{3.7}$$

$$\text{(ii)}\ C^*_2(m,m)=1\ \text{ for any }\ m\,, \tag{3.8}$$

(iii) Theorem II, Appendix C of Ref. 2: "Given an interval (j,m), $j>k_0$, and a set of partitioning subintervals $\{c_i;i=1,\ldots,N\}$, then

$$\bar{C}^*_2(j,m,j-k_0)=\sum_{i=1}^{N}g_{c_i}(m)\bar{C}^*_2(j,m-c_i,j-k_0) \tag{3.9}$$

for $m<k_0$."

Based on Theorems 3 and 4 the general solution of Eq. (3.1) can then be written as

$$b_m=\sum_{i=0}^{c_N-1}\lambda'_i\bar{C}^*_2(k_0+i,m,i)+\sum_{i=m}^{k_0-1}J(i)C^*_2(i,m)\,. \tag{3.10}$$

IV. FIRST ORDER EQUATIONS

The purpose of this section is to show the equivalence between our method and the well established standard methods for first order equations. As it will become apparent, although the equivalence is a trivial one, the standard methods for first order cannot be generalized to higher order, while our method provides a straightforward generalization.

With no loss of generality, it is always possible to reduce a first order equation to an equation relating the level x to the level $x-1$, thus leading to an equation of the form

$$u(x)=p(x)u(x-1)+q(x)\,. \tag{4.1}$$

Assuming Eq. (4.1) to hold for $x>x_0$, then $u(x_0)$ can be considered as an initial condition

$$u(x_0)=\lambda\,. \tag{4.2}$$

As it is apparent from the general techniques of Sec. II, the combinatorics functions associated with this problem are constructed out of the function $p(x)$ and based on the partitions of an interval (y,x) into parts of length 1. In order that such partitions be possible $(x-y)$ has to be a positive integer. Obviously, there is only one way of partitioning (y,x) into parts of length 1. Furthermore the combinatorics involved are unconstrained, and will therefore be called $C_2(y,x)$. These are given through

$$C_2(y,x)=\prod_{y+1}^{x}p(t)\,. \tag{4.3}$$

In order to make the connection with the standard solution of Eq. (4.1), we write Eq. (4.3) as

$$C_2(y,x)=\exp\Big[\sum_{y+1}^{x}\ln p(t)\Big]\,. \tag{4.4}$$

Finally, according to Theorem 1, the solution of Eq. (4.1) is given by

$$u(x)=\lambda C_2(x_0,x)+\sum_{y=x_0+1}^{x}q(y)\,C_2(y,x)\,. \tag{4.5}$$

When combining Eqs. (4.4) and (4.5) we obtain

$$u(x)=\lambda\exp\Big[\sum_{x_0+1}^{x}\ln p(t)\Big]+\sum_{y=x_0+1}^{x}I(y)\exp\Big[\sum_{y+1}^{x}\ln p(t)\Big]$$

$$=\left[\lambda+\sum_{y=x_0+1}^{x}\left(\frac{q(y)}{\exp\Big[\sum_{x_0+1}^{y+1}\ln p(t)\Big]}\right)\right]\exp\Big[\sum_{x_0+1}^{x}\ln p(t)\Big]\,. \tag{4.6}$$

This equation is the same as the one derived using the standard methods. The derivation given by Milne-Thomson,[6] for example, is based on what is known as the method of "variation of parameters." This technique relies on the knowledge of the solution, $u_1(x)$, of the

associated homogeneous equation, then the solution of the complete equation is obtained as

$$u(x) = v(x)\, u_1(x)\,. \tag{4.7}$$

The equivalence between the two methods is now clearly established.

Before going any further we would like to comment about a particular solution due to Milne-Thomson[6] obtained as an "ascending continued fraction"[7] equivalent to the infinite series

$$u(x) = \frac{q(x+1)}{p(x+1)} + \frac{q(x+2)}{p(x+1)p(x+2)} + \frac{q(x+3)}{p(x+1)p(x+2)p(x+3)} + \cdots . \tag{4.8}$$

We expect to obtain such a solution using our general formalism. This is better seen by looking at the inverted inhomogeneous equation whose solution has been discussed in Sec. III. The first order equation (4.1) is inverted as

$$u(x) = \frac{u(x+1)}{p(x+1)} + \frac{q(x+1)}{p(x+1)}\,, \quad x \geq x_0\,. \tag{4.9}$$

The conjugate combinatorics function of the second kind associated with Eq. (4.9) is

$$C_2^*(x,y) = \prod_{t=x+1}^{y} \left[\frac{1}{p(t)}\right], \quad y > x\,, \tag{4.10}$$

which satisfies the required normalization condition

$$C_2^*(x,x) = 1\,. \tag{4.11}$$

Theorem 3 provides a particular solution of Eq. (4.9). Setting $k_0 = \infty$ in Eq. (3.4) and using the results (4.10) and (4.11), we obtain as a particular solution

$$u(x) = \frac{q(x+1)}{p(x+1)} + \sum_{y=x+1}^{\infty} \frac{q(y+1)}{p(y+1)} \prod_{t=x+1}^{y} \frac{1}{p(t)}\,. \tag{4.12}$$

This is precisely Milne-Thomson's result mentioned above in Eq. (4.8).

A. First order equation with constant coefficients

The coefficients $p(x)$ and $q(x)$ of Eq. (4.1) are assumed to be constant

$$p(x) = p\,, \quad q(x) = q\,. \tag{4.13}$$

In this case the solution of (4.1) as given by (4.6) is inadequate. We prefer to use the equivalent expression (4.5) where, according to Eqs. (4.3) and (4.13), the combinatorics function $C_2(y,x)$ is given by

$$C_2(y,x) = p^{x-y}\,, \quad x \geq y\,. \tag{4.14}$$

Thus Eq. (4.5) specialized to the case of constant coefficients gives the solution in explicit form as

$$u(x) = \lambda p^{x-x_0} + q \sum_{y=x_0+1}^{x} p^{x-y}\,, \quad x > x_0\,, \tag{4.15}$$

or in closed form as

$$u(x) = \lambda p^{x-x_0} + \frac{q}{1-p}\,(1 - p^{x-x_0})\,, \quad p \neq 1\,, \tag{4.16a}$$

$$u(x) = \lambda + q(x - x_0)\,, \quad p = 1\,. \tag{4.16b}$$

Here again λ and x_0 specify the initial conditions. This result could have been obtained by solving the characteristic equation. The purpose of discussing this simple example was to show how our method gives a unified treatment to many special devices appropriate to special kinds of equations.

B. Bernouilli and Euler polynomials

In increasing difficulty the next case to be considered is

$$p(x) = \text{constant} = p\,, \quad q(x) \text{ variable}\,. \tag{4.17}$$

Equation (4.10) is thus replaced by

$$u(x) = \lambda p^{x-x_0} + \sum_{y=x_0+1}^{x} q(y) p^{x-y}\,. \tag{4.18}$$

Two interesting applications of Eq. (4.18) are those corresponding to Bernouilli and Euler polynomials, $B_n(x)$ and $E_n(x)$, respectively. They are defined through the following generating functions[8]:

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{t^n}{n!}\,, \quad |t| < 2\pi\,, \tag{4.19}$$

$$\frac{2e^{xt}}{e^t + 1} = \sum_{n=0}^{\infty} E_n(x) \frac{t^n}{n!}\,, \quad |t| < \pi\,, \tag{4.20}$$

and satisfy the first order difference equations:

$$B_{n+1}(x+1) = B_{n+1}(x) + (n+1)x^n\,, \tag{4.21}$$

$$E_n(x+1) = -E_n(x) + 2x^n\,. \tag{4.22}$$

Let us choose as initial point x_0,

$$x_0 = x - [x]\,, \tag{4.23}$$

where $[x]$ refers to the integer part of x. x_0 is obviously in the range

$$0 \leq x_0 < 1\,. \tag{4.24}$$

The corresponding initial values will be $B_{n+1}(x_0)$ and $E_{n+1}(x_0)$.

In the Bernouilli case we have $p = 1$ and $q(x+1) = (n+1)x^n$, and Eq. (4.18) yields

$$B_{n+1}(x+1) = B_{n+1}(x_0) + \sum_{y=x_0+1}^{x} (n+1)\, y^n\,. \tag{4.25}$$

Making the shift $y = x_0 + k$, where k is now a positive integer, we deduce that

$$\sum_{k=1}^{[x]} (k + x_0)^n = \frac{B_{n+1}(x+1) - B_{n+1}(x_0)}{n+1}\,. \tag{4.26}$$

This result holds for $[x] \geq 1$ and $n = 1, 2, \cdots$. Specializing to integer values of x, $x = m$, we obtain the known result[8]

$$\sum_{k=0}^{m} k^n = \frac{B_{n+1}(m+1) - B_{n+1}(0)}{n+1}\,. \tag{4.27}$$

In the Euler case we set $p = -1$ and $q(x+1) = 2x^n$, and Eq. (4.18) yields

$$E_n(x+1) = (-1)^{x-x_0} E_n(x_0) + \sum_{y=x_0+1}^{x} 2y^n (-1)^{x-y}\,. \tag{4.28}$$

Again with the help of the shift $y = x_0 + k$ we deduce that

$$\sum_{k=1}^{[x]} (-1)^{[x]-k}(k+x_0)^n = \frac{E_n(x+1)+(-1)^{[x]}E_n(x_0)}{2} . \tag{4.29}$$

This result holds for $[x] \geq 1$ and $n = 1, 2, \cdots$. Specializing to integer values of x, $x = m$, we obtain the known result[8]

$$\sum_{k=1}^{m} (-1)^{m-k}k^n = \frac{E_n(m+1)+(-1)^m E_n(0)}{2} . \tag{4.30}$$

V. THE INHOMOGENEOUS LEGENDRE POLYNOMIAL RECURSION RELATION

The Legendre polynomials $P_m(z)$ (m nongegative integer) are polynomials of order m containing even (odd) powers of z for even (odd) values of m. The expansion coefficients c_k and the polynomials P_m satisfy the recursion relations[9]:

$$(k+1)(k+2)c_{k+2} = -[m(m+1) - k(k+1)]c_k \tag{5.1}$$

and

$$P_m(z) = \frac{2m-1}{m} z P_{m-1}(z) - \frac{m-1}{m} P_{m-2}(z) \quad \text{for} \quad m > 0 . \tag{5.2}$$

The initial conditions on c_k are

$$c_{-2} = 0 \quad \text{for } m \text{ even}, \tag{5.3a}$$

$$c_{-1} = 0 \quad \text{for } m \text{ odd}, \tag{5.3b}$$

and on P_m

$$P_0(z) = 1, \quad P_1(z) = z . \tag{5.4}$$

In this section we would like to solve the inhomogeneous equation obtained by adding to the right-hand side of Eq. (5.2) a term that depends in general on the rank m and may as well have an explicit z dependence,

$$b_m = \frac{2m-1}{m} z b_{m-1} - \frac{m-1}{m} b_{m-2} + I(m,z), \quad m > 0 . \tag{5.5}$$

Here again our formalism can be used in a straightforward manner to give the solution of Eq. (5.5). The comparison between Eqs. (2.1) and (5.5) shows that the partition problem to be considered involves parts of lengths

$$a_1 = 1, \quad a_2 = 2, \quad \text{with } j_0 = 0 . \tag{5.6}$$

Furthermore the representations of the partitions of an interval into parts of lengths 1 and 2 are obtained upon using the functions $f_{a_k}(m)$ ($k = 1, 2$) given by

$$f_1(m) = \frac{2m-1}{m} z, \quad f_2(m) = -\frac{m-1}{m} . \tag{5.7}$$

The initial conditions, Eq. (2.16), reduce to

$$b_0 = 1, \quad b_{-1} \text{ arbitrary but finite.} \tag{5.8}$$

Here we take b_0 to be unity to agree with the standard normalization of the Legendre polynomials, Eq. (5.4). Furthermore, since $f_2(1)b_{-1} = 0$, the general solution, Eq. (2.24), applied to the case of Eq. (5.5), is independent of b_{-1}, leading to

$$b_m = C_2(0,m) + \sum_{j=1}^{m} I(j,z) C_2(j,m) . \tag{5.9}$$

Before going any further, let us note that, when no inhomogeneous term is present, $I(j,z) = 0$, the solution of Eq. (5.5),

$$b_m = C_2(0,m) \quad \text{for } I(j,z) = 0 \tag{5.10}$$

is, of course, the Legendre polynomial $P_m(z)$. Therefore, $P_m(z)$ can be viewed as a combinatorics function of the second kind,

$$P_m(z) = C_2(j,m)\big|_{j=0} . \tag{5.11}$$

What is even more interesting is the explicit evaluation of the combinatorics function, $C_2(j,m)$, appropriate to the present situation, for any value of j in the range $0 \leq j \leq m$. This will give us, indeed, the solution of Eq. (2.1) for an arbitrary inhomogeneous term $I(m,z)$.

A. The partition problem

In this section we will make apparent the polynomial structure in the parameter z of the combinatorics function of the second kind $C_2(j,m)$, by introducing two functions $g_1(m)$ and $g_2(m)$,

$$f_1(m) = z\, g_1(m), \quad f_2(m) = -g_2(m), \tag{5.12}$$

$$g_1(m) = \frac{2m-1}{m}, \quad g_2(m) = \frac{m-1}{m} . \tag{5.13}$$

According to Eqs. (1.4), (1.5), (1.7), (5.6), and (5.7), the appropriate combinatorics functions, $C_2(j,m)$, to be considered are those based on the functionals $F_n^q(j,m)$ of Eq. (1.5a) with

$$\delta_i \in \{1, 2\} \tag{5.14}$$

and the functions f_1 and f_2 as given by Eq. (5.7). In Fig. 1 we give an example on how to write down the combinatorics $C_2(j,m)$ and $C_1(j,m,n)$ for $j = 2$, $m = 7$, and $n = 3$. More generally, let us consider all possible partitions of (j,m) into n parts, and let p be the number of parts of length 2. Then $(n-p)$ is the number of parts of length 1. We therefore have

$$2p + (n-p) = \bar{m}, \quad \bar{m} = m - j \tag{5.15}$$

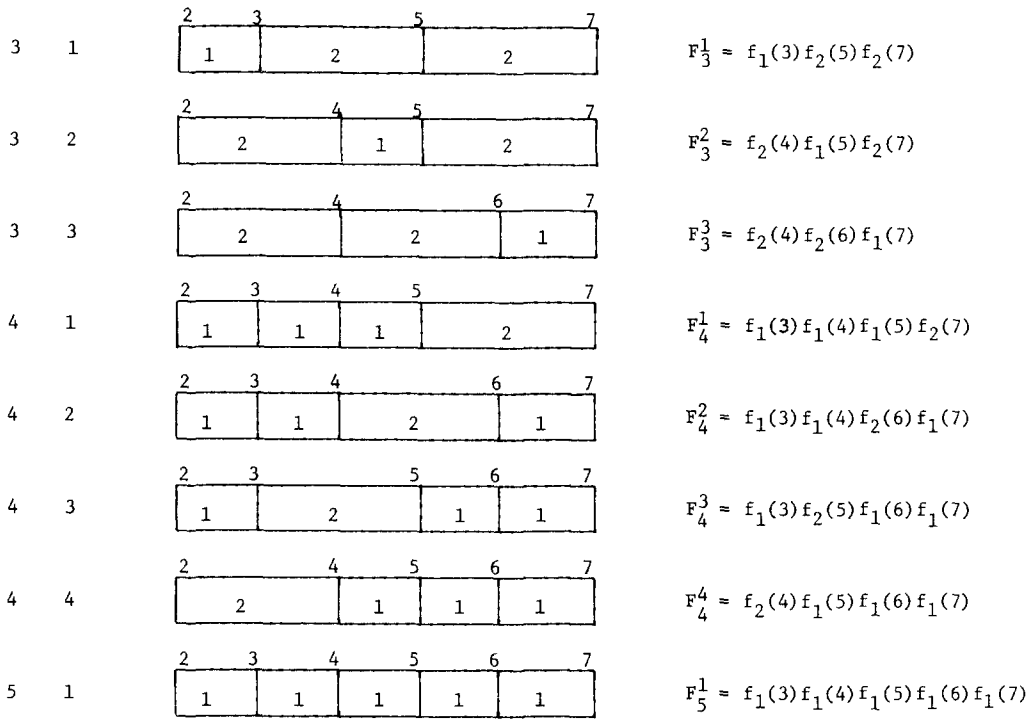
or

$$p = \bar{m} - n, \quad n - p = 2n - \bar{m} . \tag{5.16}$$

Thus, all partitions of (j,m) into n parts have the same number of parts of length 1, and the same number of parts of length 2, given by Eq. (5.16). From the latter equation, it is seen that the only possible values of n are in the range

$$n_{\min} = \bar{m} - [\bar{m}/2] \leq n \leq n_{\max} = \bar{m} , \tag{5.17}$$

where the square brackets refers to integer division. n decreases from $\bar{m}$ to $\bar{m} - [\bar{m}/2]$ by successively exchanging two parts of length 1 by one part of length 2. Or, as apparent from Eq. (5.16), holding $\bar{m} = m - j$ fixed, while n changes by one unit, the number of parts of lengths 2 and 1 will change by steps of 1 and 2, respectively.



$$C_2(2,7) = \sum_{n,q} F_n^q = F_3^1 + F_3^2 + F_3^3 + F_4^1 + F_4^2 + F_4^3 + F_4^4 + F_5^1$$

$$C_1(2,7,3) = \sum_{q} F_3^q = F_3^1 + F_3^2 + F_3^2$$

FIG. 1. Partitioning (2, 7) into $\{1, 2\}$.

B. Reduction of the combinatorics

According to Eqs. (1.5a), (5.7), and (5.14), in the product defining the functional $F_n^q(j,m)$, the function f_1 appears $(2n-\bar{m})$ times, while f_2 appears $(\bar{m}-n)$ times, for all partitions, q, into n parts. Consequently, using Eqs. (5.12) and (5.13), we obtain

$$F_n^q(j,m) = \prod_{i=1}^{n} f_{\delta_i}(s_i) = z^{n-p}(-1)^p \prod_{i=1}^{n} g_{\delta_i}(s_i), \tag{5.18}$$

where the s_i are related to the δ_i as in Eq. (1.5b), with $m_1 = j$, $m_2 = m$, and p as given by Eq. (5.16).

Summing $F_n^q(j,m)$ over all values of q, we obtain the combinatorics function of the first kind, $C_1(j,m,n)$,

$$C_1(j,m,n) = \sum_q F_n^q(j,m) = (-1)^{\bar{m}-n} z^{2n-\bar{m}} \beta(j,m,m), \tag{5.19}$$

where the "structure" function $\beta(j,m,n)$ is easily identified to be given by

$$\beta(j,m,n) = \sum_q \prod_{i=1}^{n} g_{\delta_i}(s_i). \tag{5.20}$$

β is nothing but a combinatorics function of the first kind based on the functions g_1 and g_2 instead of f_1 and f_2.

On the other hand, using the relationship (1.7b), and summing $C_1(j,m,n)$ over all possible values of the number of parts n of Eq. (5.17), we obtain the combinatorics function of the second kind

$$C_2(j,m) = \sum_{n=n_{\min}}^{n_{\max}} (-1)^{\bar{m}-n} z^{2n-\bar{m}} \beta(j,m,n). \tag{5.21}$$

Equation (5.21) shows explicitly that $C_2(j,m)$ are polynomials in z of order $\bar{m}$ containing even (odd) powers of z if $\bar{m}$ is even (odd). This justifies the notation

$$C_2(j,m) \equiv P(j,m;z). \tag{5.22}$$

Finally, it is convenient to replace in Eq. (5.21) the summation over n by a summation over p, the number of parts of length 2, Eq. (5.16), so that $P(j,m;z)$ can be written as

$$P(j,m;z) = \sum_{p=0}^{[\bar{m}/2]} z^{\bar{m}-2p}(-1)^p \beta(j,m,\bar{m}-p). \tag{5.23}$$

Making explicit the z dependence of $C_2(j,m)$, the solution of Eq. (5.5) is given by

$$b_m = P(0,m;z) + \sum_{j=1}^{m} I(j,z) P(j,m;z). \tag{5.24}$$

The next step is to study the structure function $\beta(j,m,\bar{m}-p)$, where p varies in the range $0 \leq p \leq [\bar{m}/2]$. A good check of our calculation would be to show explicitly that

$$P(0,m;z)=\sum_{p=0}^{[m/2]} z^{m-2p}(-1)^p \beta(0,m,m-p)=P_m(z)\,. \tag{5.25}$$

A detailed analysis of the structure function $\beta(j,m,\bar{m}-p)$ is carried out in Appendix A.

An interesting question to investigate is whether the polynomials $P(j,m;z)$ are orthogonal. Such a study goes far beyond the scope of this paper and we will leave it for a separate publication, beyond remarking that they are not orthogonal in general. We do not know as yet if orthogonality occurs for $j=0$ only. Nevertheless, according to Eq. (5.24) a new set of orthogonal functions, $b_m(z)$, can be constructed, in principle, by adjusting the inhomogeneous term $I(m,z)$.

VI. CONCLUSION

This work gives a unified treatment of linear inhomogeneous finite-difference equations. It gives an alternative way of looking at the solutions of these equations through the representations of the partitions of an interval into parts of given lengths. In many respects, special devices for solving special types of equations may be more suitable or direct; still, as it has been shown in a few simple examples, these devices can be derived from our general theory. It is our hope that the ideas first introduced by Antippa and Phares[1,2,3] can be extended to linear equations of partial differences or other more complicated equations, where no general theory exists.

APPENDIX A: THE STRUCTURE FUNCTION $\beta(j,m,\bar{m}\text{-}p)$

The definition of the structure function $\beta(j,m,\bar{m}-p)$ is given by Eq. (5.20). It is a combinatorics function of the first kind[1] based on the functions

$$g_1(m)=\frac{2m-1}{m}\,,\quad g_2(m)=\frac{m-1}{m} \tag{A1}$$

and on the partitions of the interval (j,m) into $(\bar{m}-p)$ parts of lengths 1 and 2. Here $\bar{m}=m-j$ and p is the number of parts of length 2.

It is easy to show that p satisfies a linear equation of partial differences, namely,

$$\beta(j,m,\bar{m}-p)=g_1(m)\beta(j,m-1,\bar{m}-p-1)+g_2(m)\beta(j,m-2,\bar{m}-p-1)\,. \tag{A2}$$

This result also follows from Theorem 1 of Ref. 1 on combinatorics functions of the first kind.

A. The case $p = 0$

Since p represents the number of parts of length 2 in any partition of (j,m), the case $p=0$ corresponds to the partition of (j,m) into parts of length 1, thus there is only one way of performing such a partition. The corresponding combinatorics function representing this partition is

$$\beta(j,m,\bar{m})=\prod_{i=j+1}^{m} g_1(i)=\prod_{i=j+1}^{m}\frac{2i-1}{i}\,. \tag{A3}$$

Equation (A3) can be written in closed form in terms of Γ functions as

$$\beta(j,m,\bar{m})=2^{\bar{m}}\frac{j!}{m!}\frac{\Gamma(m+\frac{1}{2})}{\Gamma(j+\frac{1}{2})}\,, \tag{A4}$$

while Eq. (A3) holds only for $m>j$, Eq. (A4) continues to be meaningful even for $m=j$, reproducing the correct normalization for combinatorics functions, Eq. (1.9), namely,

$$\beta(m,m,0)=1\,. \tag{A5}$$

B. The case $p = 1$

This case corresponds to the exchange in the preceeding partition ($p=0$) of two consecutive parts of length 1 by one part of length 2. The resulting effect on the structure function is shown in Fig. 2, where i may run from $i=(j+1)$ to $i=m$. The relationship between $\beta(j,m,\bar{m}-1)$ and $\beta(j,m,\bar{m})$ follows as

$$\beta(j,m,\bar{m}-1)=\beta(j,m,\bar{m})\sum_{i=j+1}^{m-1}\frac{i}{4i^2-1}\,. \tag{A6}$$

It is easy to show by induction that

$$\sum_{i=1}^{m-1}\frac{i^2}{4i^2-1}=\frac{m(m-1)}{2(2m-1)}\,. \tag{A7}$$

Combining (A6) and (A7) we obtain

$$\beta(j,m,\bar{m}-1)=\beta(j,m,\bar{m})\left[\frac{m(m-1)}{2(2m-1)}-\frac{j(j+1)}{2(2j+1)}\right]\,. \tag{A8}$$

Closed form expressions of the structure function can be obtained for $p=2,3,\ldots,[\bar{m}/2]$ using techniques similar to the one presented here. Before going any further, we would like to first discuss the case, $j=0$. This is an important step and deserves separate treatment as a consistency check of Eq. (5.25).

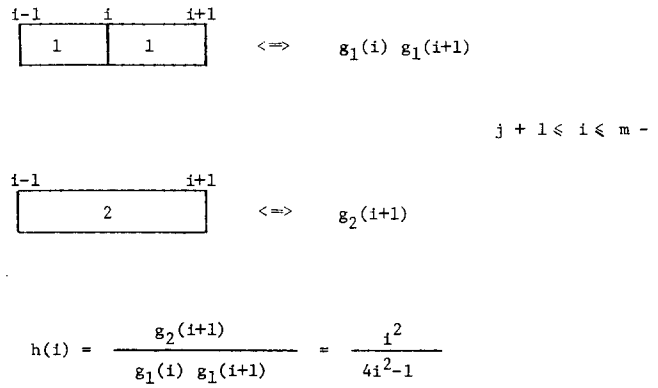


FIG. 2. Exchange of two parts of length 1 by one part of length 2.

C. The structure function $\beta(0, m, m\text{-}p)$

The claim is that $\beta(0,m,m-p)$ is related to $\beta(0,m,m)$ according to

$$\beta(0,m,m-p)=\beta(0,m,m)\prod_{k=1}^{p}\frac{(m-2k+1)(m-2k+2)}{2k(2m-2k+1)}. \tag{A9}$$

The proof is done by induction with the aid of the difference equation (A2). In Eq. (A2) we set $j=0$ and make the shift $m\to m+2$,

$$g_2(m+2)\beta(0,m,m-p+1)=\beta(0,m+2,m+2-p)-g_1(m+2)\beta(0,m+1,m+1-p). \tag{A10}$$

Assuming Eq. (A9) to hold for a given value of p, we show that it also holds for $p-1$ using Eq. (A10). This step is straightforward. In Eq. (A10) we replace $\beta(0,m+2,m+2-p)$ and $\beta(0,m+1,m+1-p)$ by their assumed values compuated from Eq. (A9). Then, using Eqs. (A1) and (A3), we find

$$\frac{m+1}{m+2}\beta(0,m,m-p+1)=\frac{2m+1}{m+1}\frac{2m+3}{m+2}\beta(0,m,m)\times\left[\prod_{k=1}^{p}\frac{(m-2k+3)(m-2k+4)}{2k(2m-2k+5)}-\prod_{k=1}^{p}\frac{(m+2k+2)(m+2k+3)}{2k(2m-2k+3)}\right]. \tag{A11}$$

The above equation is equivalently written as

$$\beta(0,m,m-p+1)=\frac{(2m+1)(2m+3)}{(m+1)^2}\beta(0,m,m)\left[1-\frac{2m+3}{m+2}\frac{m-2p+2}{2m-2p+3}\right]\times\prod_{k=1}^{p}\frac{(m-2k+3)(m-2k+4)}{2k(2m-2k+5)}. \tag{A12}$$

The next step is to use the following relations:

$$\prod_{k=1}^{p}\frac{1}{2k}=\frac{1}{2p}\prod_{k=1}^{p-1}\frac{1}{2k}, \tag{A13a}$$

$$\prod_{k=1}^{p}(m-2k+3)=(m+1)\prod_{k=1}^{p-1}(m-2k+1), \tag{A13b}$$

$$\sum_{k=1}^{p}(m-2k+4)=(m+2)\sum_{k=1}^{p-1}(m-2k+2), \tag{A13c}$$

$$\prod_{k=1}^{p}\frac{1}{(2m-2k+5)}=\frac{(2m-2p+3)}{(2m+1)(2m+3)}\prod_{k=1}^{p-1}\frac{1}{(2m-2k+1)}. \tag{A13d}$$

Finally, combining Eqs. (A12) and (A13), we obtain

$$\beta(0,m,m-p+1)=\beta(0,m,m)\prod_{k=1}^{p-1}\frac{(m-2k+1)(m-2k+2)}{2k(2m-2k+1)}. \tag{A14}$$

In order to complete the proof by induction we still have to show that Eq. (A9) holds for the highest possible value of p, $p=[m/2]$, where the square brackets refer to integer division. This is done by comparing the expressions of

$$\beta(0,m,[m/2]) \text{ and}$$

$$\beta(0,m,m)\prod_{k=1}^{[m/2]}\frac{(m-2k+1)(m-2k+2)}{2k(2m-2k+1)}.$$

A direct calculation of these two expressions involves the use of Eq. (5.20) and rearrangements analogous to those of Eqs. (A13). The result is that the two expressions are equal, and this completes the proof by induction.

We would like to make a consistency check and verify that $P(0,m;z)$,

$$P(0,m;z)=\sum_{p=0}^{[m/2]}z^{m-2p}(-1)^p\beta(0,m,m-p), \tag{A15}$$

is indeed equal to the Legendre polynomial $P_m(z)$ whose expansion coefficients, c_k, satisfy the recursion relation (5.1).

According to Eq. (A9), the structure functions $\beta(0,m,m-p)$ satisfy the recursion relation

$$\beta(0,m,m-p)=\frac{2p(2m-2p+1)}{(m-2p+1)(m-2p+2)}\times\beta(0,m,m-p+1). \tag{A16}$$

If we now set in Eq. (A16)

$$m-2p=k, \tag{A17}$$

we obtain

$$\beta(0,m,m-p)=\frac{(m-k)(m+k+1)}{(k+1)(k+2)}\beta(0,m,m-p+1). \tag{A18}$$

Comparing Eqs. (5.1) and (A18) we immediately establish the relationship

$$C_{k+2}=-\beta(0,m,(m-k)/2) \tag{A19}$$

where k and m must have the same parity as evidenced in Eq. (A17). This completes the check and shows that

$$P(0,m;z)=P_m(z). \tag{A20}$$

D. The general case: $0\leqslant p\leqslant[m/2]$ and $j\neq 0$

It is convenient to define a quantity $\sigma(j,m,p)$ through

$$\beta(j,m,\bar{m}-p)=\sigma(j,m,p)\beta(j,m,\bar{m}), \tag{A21}$$

where p couples of the type shown in Fig. 2 are replaced by p parts of length 2. As evidenced by Fig. 2, the effect of a single exchange of two consecutive parts of length 1 by one part of length 2 is the appearance of a factor $h(i)$ given by

$$h(i)=\frac{i^2}{4i^2-1}. \tag{A22}$$

Let G_n^q be the functional constructed with the functions g_1 and g_2, and associated with a partition, q, of (j,m)

into parts of lengths 1 and 2. Thus a representation G_n^q, $n=\overline{m}-p$, of a partition, q, of (j,m) into $\overline{m}-p$ parts differs from $\beta(j,m,\overline{m})$ by the multiplicative factor

$$h(i_1)h(i_2)\cdots h(i_p)\,. \tag{A23}$$

Here $i_1, i_2, \ldots, i_p$ correspond to the location of the p couples of the type shown in Fig. 2, that have been replaced by parts of length 2. Another partition is obtained by varying the values of $i_1, i_2, \ldots, i_p$. A straightforward analysis shows that the only possible values of $i_k\,(1\leq k\leq p)$ are in the range

$$i_{k-1}+2\leq i_k\leq m-2(p-k)+1 \tag{A24a}$$

with the initial value, i_0, given by

$$i_0=j-1\,. \tag{A24b}$$

Since the structure function $\beta(j,m,\overline{m}-p)$ is obtained by summing the G_n^q over all values q, or partitions into a given number of parts $n=\overline{m}-p$, we deduce that

$$\sigma(j,m,p)=\sum_{i=j+1}^{m-2p+1}\cdots\sum_{i_k=i_{k-1}+2}^{m-2(p-k)-1}\cdots \sum_{i_p=i_{p-1}+2}^{m-1} h(i_1)\cdots h(i_k)\cdots h(i_p)\,. \tag{A25}$$

Comparing Eqs. (A9) and (A21) we immediately get $\sigma(0,m,p)$ as

$$\sigma(0,m,p)=\prod_{k=1}^{p}\frac{(m-2k+1)(m-2k+2)}{2k(2m-2k+1)} \tag{A26a}$$

or in closed form as

$$\sigma(0,m,p)=\frac{1}{p!}\,\frac{\Gamma((m+1)/2)}{\Gamma((m+1)/2-p)}\,\frac{\Gamma((m+2)/2)}{\Gamma((m+2)/2-p)}\times\frac{\Gamma(m+\frac{1}{2}-p)}{\Gamma(m+\frac{1}{2})}\,. \tag{A26b}$$

While Eq. (A26a) holds for $p\geq 1$, Eq. (A26b) continues to be meaningful for $p=0$, and reproduces the required result,

$$\sigma(0,m,0)=1\,. \tag{A27a}$$

As a matter of fact, according to Eq. (A21), we must have the initial value condition

$$\sigma(j,m,0)=1\,,\quad 0\leq j\leq m\,. \tag{A27b}$$

Equation (A25) can be written differently when leaving out the summation over i_1,

$$\sigma(j,m,p)=\sum_{i_1=j+1}^{m-2p+1}h(i_1)\sum_{i_2=i_1+2}^{m-2(p-1)+1}\cdots\sum_{i_p=i_{p-1}+2}^{m-1}h(i_2)\cdots h(i_p)$$
$$=\sum_{i_1=j+1}^{m-2p+1}h(i_1)\sigma(i_1+1,m,p-1)\,,\quad p\geq 1\,. \tag{A28}$$

The above equation is basically a linear equation of partial differences, that we plan to solve using the result of the preceding subsection. The summation over i_1 in Eq. (A28) can be broken into two parts,

$$\sigma(j,m,p)=\sum_{i_1=1}^{m-2p+1}h(i_1)\sigma(i_1+1,m,p-1)-\sum_{i_1=0}^{j}h(i_1)\sigma(i_1+1,m,p-1)\,. \tag{A29}$$

Here we used the property $h(0)=0$. According to Eq. (A28), the first term in the right-hand side of Eq. (A29) is $\sigma(j,m,p)$ evaluated at $j=0$, thus

$$\sigma(j,m,p)=\sigma(0,m,p)-\sum_{i_1=0}^{j}h(i_1)\sigma(i_1+1,m,p-1)\,. \tag{A30}$$

On the other hand, $\sigma(i_1+1,m,p-1)$ can be replaced by its equivalent expression computed from Eq. (A30) itself with j replaced by i_1+1, and p replaced by $p-1$, i.e.,

$$\sigma(i_1+1,m,p-1)=\sigma(0,m,p-1)-\sum_{i_2=0}^{i_1+1}h(i_2)\sigma(i_2+1,m,p-2)\,. \tag{A31}$$

We repeat the same procedure p times, on $\sigma(i_2+1,m,p-2)$, etc. After p iterations Eq. (A30) becomes

$$\sigma(j,m,p)=\sum_{k=0}^{p}(-1)^k I_k(j)\sigma(0,m,p-k)\,, \tag{A32}$$

where the expansion coefficient $I_k(j)$ is given by

$$I_k(j)=\sum_{i_1=0}^{j}h(i_1)\sum_{i_2=0}^{i_1+1}h(i_2)\cdots\sum_{i_k=0}^{i_{k-1}+1}h(i_k)\quad\text{for } k\geq 1 \tag{A33a}$$

and

$$I_k(j)=1\quad\text{for } k=0 \text{ and any } j\,. \tag{A33b}$$

The problem of calculating $\sigma(j,m,p)$ has now been reduced to determining the explicit expressions of the expansion coefficients $I_k(j)$ for $k\neq 0$. This is done by deriving a recursion relation for the I_k's.

E. Recursion relation for $I_k(j)$

We first set $j=0$ in Eq. (A28)

$$\sigma(0,m,p)=\sum_{i_1=1}^{m-2p+1}h(i_1)\sigma(i_1+1,m,p-1)\,,\quad p\geq 1 \tag{A34}$$

and then replace in Eq. (A34), $\sigma(i_1+1,m,p-1)$ by its value computed from (A32), where j is set to be i_1+1 and p is shifted to $p-1$,

$$\sigma(0,m,p)=\sum_{i_1=1}^{m-2p+1}h(i_1)\sum_{k=0}^{p-1}(-1)^k I_k(i_1+1)\times\sigma(0,m,p-k-1)\quad\text{for } p\geq 1\,. \tag{A35}$$

Equation (A33a) can also be written as

$$I_k(j)=\sum_{i_1=0}^{j}h(i_1)I_{k-1}(i_1+1)\,. \tag{A36}$$

This is a recursion relation for $I_k(j)$ but it is a useless one because we do not know how to solve it. We will be using it to derive another relation more suitable to our purpose.

In Eq. (A35), we perform the summation over i_1 first,

and make use of the property (A36), where j is replaced by $(m-2p+1)$ and k shifted to $(k+1)$. The result of these operations is

$$\sigma(0,m,p)=\sum_{k=0}^{p-1}(-1)^k I_{k+1}(m-2p+1) \times\sigma(0,m,p-k-1) \quad \text{for } p\geq 1 . \tag{A37}$$

Finally, in the above equation, we set $m=j+2p-1$, and obtain

$$\sigma(0,j+2p-1,p)=-\sum_{k=1}^{p}(-1)^k I_k(j)\sigma(0,j+2p-1,p-k) \quad \text{for } p\geq 1 . \tag{A38}$$

Using the property (A33b) we arrive at the desired form,

$$\sum_{k=0}^{p}(-1)^k I_k(j)\sigma(0,j+2p-1,p-k)=0 \quad \text{for } p\geq 1 . \tag{A39}$$

This equation is used to calculate the closed form expression of $I_k(j)$ with the help of Eq. (A26). Setting $p=1$ and 2 in Eq. (A39) we obtain

$$I_1(j)=\sigma(0,j+1,1)=\frac{j(j+1)}{2(2j+1)} \tag{A40}$$

$$I_2(j)=I_1(j)\sigma(0,j+3,1)-\sigma(0,j+3,2) . \tag{A41a}$$

Combining Eqs. (A40), (A41), and (A26a), we find that $I_2(j)$ can also be written as

$$I_2(j)=\sigma(0,j+3,2) . \tag{A41b}$$

Anticipating the final result, we generalize to

$$I_k(j)=\sigma(0,j+2k-1,k) . \tag{A42}$$

The proof that $I_k(j)$ is given by (A42) can be done by induction in the following way: We assume Eq. (A42) to hold for $k=0,1,\ldots,p-1$ and prove that it also holds for $k=p$ by using the recursion relation (A39). Since the proof does not present any major difficulty we shall skip it.

Combining Eqs. (A21), (A32), and (A42) we have at hand an explicit expression for the structure function $\beta(j,m,\bar{m}-p)$,

$$\beta(j,m,\bar{m}-p)=\beta(j,m,\bar{m})\sum_{k=0}^{p}(-1)^k\sigma(0,j+2k-1,k) \times\sigma(0,m,p-k) . \tag{A43}$$

Using Eq. (A26b) it is worthwhile noting that $\sigma(0,j+2k-1,k)$ is given in closed form as

$$\sigma(0,j+2k-1,k) =\frac{1}{k!}\frac{\Gamma(k+j/2)}{\Gamma(j/2)}\frac{\Gamma(k+(j+1)/2)}{\Gamma((j+1)/2)}\frac{\Gamma(j+k-\frac{1}{2})}{\Gamma(j+2k-\frac{1}{2})} . \tag{A44}$$

As special cases of Eq. (A44), we have

$$\sigma(0,j+2k-1,k)=0 \quad \text{for } j=0,k\neq 0 , \tag{A45a}$$

$$\sigma(0,j+2k-1,k)=1 \quad \text{for } k=0 . \tag{A45b}$$

Thus Eq. (A9) is reproduced by setting $j=0$ in Eq. (A43). Likewise Eq. (A8) can be derived from the general formula (A43) by setting $p=1$.

This completes the study of the structure function $\beta(j,m,\bar{m}-p)$ for all possible values of p and j.

[1]A. F. Antippa and A. J. Phares, J. Math. Phys. **18**, 173 (1977).
[2]A. F. Antippa and A. J. Phares, "The linear potential: An exact solution," submitted to J. Math. Phys.
[3]A. F. Antippa and A. J. Phares, "The combined Coulomb and linear potential: An exact solution," in preparation.
[4]T. Appelquist and H. D. Politzer, Phys. Rev. Lett. **34**, 43 (1975); A. De Rujula and S. L. Glashow *ibid.* **34**, 46 (1975); T. Appelquist *et al.*, *ibid.* **34**, 365 (1975); E. Eichten *et al.*, *ibid.* **34**, 370 (1975); B. J. Harrington *et al.*, *ibid.* **34**, 168, 706 (1975).
[5]J. J. Aubert *et al.*, Phys. Rev. Lett. **33**, 1404 (1974); J.-E. Augustin *et al.*, *ibid.* **33**, 1406 (1974); C. Bacci *et al.*, *ibid.* **33**, 1408 (1974); G. S. Abrams *et al.*, *ibid.* **33**, 1453 (1974); J.-E. Augustin *et al.*, *ibid.* **34**, 764 (1975); G. J. Feldman *et al.*, *ibid.* **35**, 821 (1975).
[6]L. M. Milne-Thomson, *The Calculus of Finite Differences* (St. Martin's, New York, 1951).
[7]L. M. Milne-Thomson, Proc Edinburgh Math. Soc. **2**, 3 (1933).
[8]See, for example, *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, (National Bureau of Standards, Washington, D.C., 1964), p. 804.
[9]See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1963), pp. 176—8.

Spin-weighted angular spheroidal functions

Edward D. Fackerell and Robert G. Crossman*

Department of Applied Mathematics, The University of Sydney, N.S.W. 2006, Australia
(Received 20 August 1976)

The analytic properties of the spin-weighted angular spheroidal functions introduced by Teukolsky are investigated by means of a series involving Jacobi polynomials. This approach facilitates the numerical determination of eigenvalues, particularly in the case of complex frequencies.

1. INTRODUCTION

Generalizations of angular spheroidal wavefunctions known as spin-weighted angular spheroidal functions were introduced by Teukolsky[1] in a paper dealing with perturbations of rotating black holes. These functions, which we shall denote by ${}_sS_{lm}(\gamma;x)$ are defined as the eigenfunctions of the differential equation

$$(1-x^2)\frac{d^2S}{dx^2} - 2x\frac{dS}{dx} + \left(\gamma^2x^2 - \frac{m^2+s^2}{1-x^2} - \frac{2msx}{1-x^2} - 2\gamma sx + {}_sE_l^m(\gamma)\right)S = 0, \tag{1}$$

where ${}_sE_l^m(\gamma)$ is the eigenvalue. Clearly, when $s=0$, Eq. (1) reduces to the usual spheroidal wave equation. In a later paper Press and Teukolsky[2] used a perturbation technique for small real γ to obtain solutions of Eq. (1) in the form

$${}_sS_{lm}(\gamma;x) = \sum_{l'} {}_sA_{ll'}^m(\gamma)\,{}_sY^m{}_{l'}(x'). \tag{2}$$

For larger values of the parameter γ Press and Teukolsky used a continuation technique due to Wasserstrom.[3] This technique is extremely powerful for the purposes of numerical computation. However, these methods are not suitable for the elucidation of the analytic properties of Eq. (1).

In the present paper the analytic properties of the eigenfunctions of Eq. (1) are investigated by means of expansions of the form

$${}_sS_{lm}(\gamma;x) = \exp(\gamma x)\left(\frac{1-x}{2}\right)^{|m+s|/2}\left(\frac{1+x}{2}\right)^{|m-s|/2} \times \sum_{r=0}^{\infty} {}_sA_{lm}^{(r)}(\gamma)P_r^{(|m+s|,\,|m-s|)}(x) \tag{3}$$

and

$${}_sS_{lm}(\gamma;x) = \exp(-\gamma x)\left(\frac{1-x}{2}\right)^{|m+s|/2}\left(\frac{1+x}{2}\right)^{|m-s|/2} \times \sum_{r=0}^{\infty} {}_sB_{lm}^{(r)}(\gamma)P_r^{(|m+s|,\,|m-s|)}(x), \tag{4}$$

where $P_r^{(\alpha,\beta)}(x)$ is the Jacobi polynomial and the coefficients ${}_sA_{lm}^{(r)}(\gamma)$ and ${}_sB_{lm}^{(r)}(\gamma)$ satisfy separate three-term recurrence relations. Both of these recurrence relations give rise to a certain transcendental equation involving a continued fraction for the determination of the eigenvalues ${}_sE_l^m(\gamma)$.

2. TRANSFORMATIONS OF THE DIFFERENTIAL EQUATION

In order to investigate the eigenfunctions of the differential equation (1), we transform this equation by taking out the appropriate behavior of the eigenfunctions at the singular points of the differential equation, namely $x=-1$, $x=+1$, and $x=\infty$. We find that near $x=-1$ the dominant behavior is $(1+x)^{|m-s|/2}$, near $x=+1$ the dominant behavior is $(1-x)^{|m+s|/2}$, and at the point at infinity the behavior is either $\exp(\gamma x)$ or $\exp(-\gamma x)$. We therefore introduce the abbreviations

$$\alpha = |m+s|, \tag{5}$$

$$\beta = |m-s|, \tag{6}$$

and introduce new functions ${}_sU_{lm}(\gamma;x)$ and ${}_sV_{lm}(\gamma;x)$ by means of the equations

$${}_sS_{lm}(\gamma;x) = \exp(\gamma x)\left(\frac{1-x}{2}\right)^{\alpha/2}\left(\frac{1+x}{2}\right)^{\beta/2}{}_sU_{lm}(\gamma;x) \tag{7}$$

and

$${}_sS_{lm}(\gamma;x) = \exp(-\gamma x)\left(\frac{1-x}{2}\right)^{\alpha/2}\left(\frac{1+x}{2}\right)^{\beta/2}{}_sV_{lm}(\gamma;x). \tag{8}$$

Note that this implies

$${}_sV_{lm}(\gamma;x) = \exp(2\gamma x)\,{}_sU_{lm}(\gamma;x). \tag{9}$$

We readily find that ${}_sU_{lm}(\gamma;x)$ satisfies the differential equation

$$(1-x^2)\frac{d^2U}{dx^2} + [\beta-\alpha-(2+\alpha+\beta)x + 2\gamma(1-x^2)]\frac{dU}{dx} + \left[{}_sE_l^m(\gamma) + \gamma^2 - \frac{\alpha+\beta}{2}\left(\frac{\alpha+\beta}{2}+1\right) + \gamma(\beta-\alpha) - \gamma(\alpha+\beta+2+2s)x\right]U = 0 \tag{10}$$

and that ${}_sV_{lm}(\gamma;x)$ satisfies the differential equation

$$(1-x^2)\frac{d^2V}{dx^2} + [\beta-\alpha-(2+\alpha+\beta)x - 2\gamma(1-x^2)]\frac{dV}{dx} + \left[{}_sE_l^m(\gamma) + \gamma^2 - \frac{\alpha+\beta}{2}\left(\frac{\alpha+\beta}{2}+1\right) - \gamma(\beta-\alpha) + \gamma(\alpha+\beta+2-2s)x\right]V = 0. \tag{11}$$

3. EIGENFUNCTIONS AS SERIES OF JACOBI POLYNOMIALS

The differential equations (10) and (11) are closely related to the differential equation for Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$. These polynomials, defined by the

Rodrigues' formula

$$P_n^{(\alpha,\beta)}(x)=\frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\left(\frac{d}{dx}\right)^n [(1-x)^{\alpha+n}(1+x)^{\beta+n}], \tag{12}$$

satisfy the differential equation

$$(1-x^2)\frac{d^2y}{dx^2}+[\beta-\alpha-(\alpha+\beta+2)x]\frac{dy}{dx}+n(n+\alpha+\beta+1)y=0 \tag{13}$$

to which both Eqs. (10) and (11) reduce when $\gamma=0$, provided we make the identification

$${}_sE_l^m(0)=[n+(\alpha+\beta)/2][n+(\alpha+\beta)/2+1] \tag{14}$$

or

$${}_sE_l^m(0)=l(l+1), \tag{15}$$

where

$$l=n+(\alpha+\beta)/2=n+\max(|m|,|s|). \tag{16}$$

Further, Jacobi polynomials also satisfy the recurrence relation

$$xP_n^{(\alpha,\beta)}(x)=\frac{2(n+1)(n+\alpha+\beta+1)}{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)}P_{n+1}^{(\alpha,\beta)}(x)-\frac{(\alpha^2-\beta^2)}{(2n+\alpha+\beta)(2n+\alpha+\beta+2)}P_n^{(\alpha,\beta)}(x)+\frac{2(n+\alpha)(n+\beta)}{(2n+\alpha+\beta)(2n+\alpha+\beta+1)}P_{n-1}^{(\alpha,\beta)}(x) \tag{17}$$

and the differentiation formula

$$(1-x^2)\frac{d}{dx}P_n^{(\alpha,\beta)}(x)=\frac{-2n(n+1)(n+\alpha+\beta+1)}{(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)}P_{n+1}^{(\alpha,\beta)}(x)+\frac{(\alpha-\beta)2n(n+\alpha+\beta+1)}{(2n+\alpha+\beta)(2n+\alpha+\beta+2)}P_n^{(\alpha,\beta)}(x)+\frac{2(n+\alpha)(n+\beta)(n+\alpha+\beta+1)}{(2n+\alpha+\beta)(2n+\alpha+\beta+1)}P_{n-1}^{(\alpha,\beta)}(x). \tag{18}$$

Consequently, if we expand ${}_sU_{lm}(\gamma;x)$ and ${}_sV_{lm}(\gamma;x)$ as infinite series of Jacobi polynomials,

$${}_sU_{lm}(\gamma;x)=\sum_{r=0}^{\infty}{}_sA_{lm}^{(r)}(\gamma)P_r^{(\alpha,\beta)}(x) \tag{19}$$

and

$${}_sV_{lm}(\gamma;x)=\sum_{r=0}^{\infty}{}_sB_{lm}^{(r)}(\gamma)P_r^{(\alpha,\beta)}(x), \tag{20}$$

we obtain three-term recurrence relations for the coefficients. Specifically, we find that

$$\left[{}_sE_l^m(\gamma)+\gamma^2-\frac{\alpha+\beta}{2}\left(\frac{\alpha+\beta}{2}+1\right)+\frac{2(\alpha-\beta)s\gamma}{\alpha+\beta+2}\right]{}_sA_{lm}^{(0)}(\gamma)+\frac{4\gamma(\alpha+1)(\beta+1)((\alpha+\beta)/2+1-s)}{(\alpha+\beta+2)(\alpha+\beta+3)}{}_sA_{lm}^{(1)}(\gamma)=0, \tag{21}$$

and, for $r=1,2,\cdots$,

$$\frac{4\gamma(r+\alpha+1)(r+\beta+1)(r+(\alpha+\beta)/2+1-s)}{(2r+\alpha+\beta+2)(2r+\alpha+\beta+3)}{}_sA_{lm}^{(r+1)}(\gamma)+\left[{}_sE_l^m(\gamma)+\gamma^2-\left(r+\frac{\alpha+\beta}{2}\right)\left(r+\frac{\alpha+\beta}{2}+1\right)+\frac{2\gamma s(\alpha-\beta)(\alpha+\beta)}{(2r+\alpha+\beta)(2r+\alpha+\beta+2)}\right]{}_sA_{lm}^{(r)}(\gamma)-\frac{4\gamma r(r+\alpha+\beta)(r+(\alpha+\beta)/2+s)}{(2r+\alpha+\beta-1)(2r+\alpha+\beta)}{}_sA_{lm}^{(r-1)}(\gamma)=0; \tag{22}$$

similarly the equations for the ${}_sB_{lm}^{(r)}(\gamma)$ are

$$\left[{}_sE_l^m(\gamma)+\gamma^2-\frac{\alpha+\beta}{2}\left(\frac{\alpha+\beta}{2}+1\right)+\frac{2(\alpha-\beta)s\gamma}{(\alpha+\beta+2)}\right]{}_sB_{lm}^{(0)}(\gamma)-\frac{4\gamma(\alpha+1)(\beta+1)((\alpha+\beta)(2+1-s))}{(\alpha+\beta+2)(\alpha+\beta+3)}{}_sB_{lm}^{(1)}(\gamma)=0 \tag{23}$$

and, for $r=1,2,\cdots$,

$$\frac{-4\gamma(r+\alpha+1)(r+\beta+1)(r+(\alpha+\beta)/2+1+s)}{(2r+\alpha+\beta+2)(2r+\alpha+\beta+3)}{}_sB_{lm}^{(r+1)}(\gamma)+\left[{}_sE_l^m(\gamma)+\gamma^2-\left(r+\frac{\alpha+\beta}{2}\right)\left(r+\frac{\alpha+\beta}{2}+1\right)+\frac{2\gamma s(\alpha-\beta)(\alpha+\beta)}{(2r+\alpha+\beta)(2r+\alpha+\beta+2)}\right]{}_sB_{lm}^{(r)}(\gamma)+\frac{4\gamma r(r+\alpha+\beta)(r+(\alpha+\beta)/2-s)}{(2r+\alpha+\beta-1)(2r+\alpha+\beta)}{}_sB_{lm}^{(r-1)}(\gamma). \tag{24}$$

At first sight it would seem from either Eqs. (21) and (22) or Eqs. (23) and (24) that a well-behaved solution of the original differential equation (1) exists for any value of the constant ${}_sE_l^m(\gamma)$, since, given ${}_sA_{lm}^{(0)}(\gamma)$, all of the other ${}_sA_{lm}^{(r)}$ may be determined uniquely via Eqs. (21) and (22). However, unless ${}_sE_l^m(\gamma)$ is chosen appropriately, the ${}_sA_{lm}^{(r)}$ increase without bound and the series (19) does not converge. It may be shown directly from Eq. (22) that for sufficiently large r, either

$${}_sA_{lm}^{(r)}(\gamma)\sim\frac{\text{const}(-\gamma)^r}{\Gamma(r+(\alpha+\beta+3)/2-s)} \tag{25}$$

or

$${}_sA_{lm}^{(r)}(\gamma)\sim\text{const}(\gamma)^{-r}\Gamma\left(r+\frac{\alpha+\beta+1}{2}+s\right). \tag{26}$$

For the case of Eq. (26) the coefficients increase in absolute value without bound and the series (19) diverges for all values of x. In the case of Eq. (25), the series (19) converges uniformly in the finite complex x plane.[4] By an argument analogous to that given by Flammer[5] (cf. also Meixner and Schäfke,[6] Sec. 1–8), this convergence requires that ${}_sE_l^m(\gamma)$ satisfy a certain transcendental equation, which is most conveniently formulated in terms of continued fractions.

4. DETERMINATION OF THE EIGENVALUE

In order to obtain the equation which determines the eigenvalues ${}_sE_l^m(\gamma)$, we find it convenient to define quantities ${}_sN_r^{lm}(\gamma)$, and ${}_sK_r^{lm}(\gamma)$, and ${}_sL_r^{lm}(\gamma)$ by

$$ {}_sN_r^{lm}(\gamma)=\frac{2\gamma(r+\alpha)(r+\beta)(2r+\alpha+\beta-2s)}{(2r+\alpha+\beta)(2r+\alpha+\beta+1)}\ \frac{{}_sA_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(r-1)}(\gamma)}\ , \tag{27} $$

$$ {}_sK_r^{lm}(\gamma)=\left(r+\frac{\alpha+\beta}{2}\right)\left(r+\frac{\alpha+\beta}{2}+1\right)-\gamma^2-\frac{2\gamma s(\alpha-\beta)(\alpha+\beta)}{(2r+\alpha+\beta)(2r+\alpha+\beta+2)}\ , \tag{28} $$

$$ {}_sL_r^{lm}(\gamma)=\frac{2\gamma^2(r+\alpha)(r+\beta)(r+\alpha+\beta)(2r+\alpha+\beta+2s)(2r+\alpha+\beta-2s)}{(2r+\alpha+\beta)(2r+\alpha+\beta+3)(2r+\alpha+\beta-1)}\ , \tag{29} $$

where we have made the assumption that none of the ${}_sA_{lm}^{(r)}$ vanish.

Equation (21) may now be written as

$$ {}_sN_1^{lm}(\gamma)-{}_sK_0^{lm}(\gamma)+{}_sE_l^m(\gamma)=0 \tag{30} $$

and (22) may be written as

$$ {}_sN_r^{lm}(\gamma)={}_sL_r^{lm}(\gamma)/[{}_sE_l^m(\gamma)-{}_sK_r^{lm}(\gamma)+{}_sN_{r+1}^{lm}(\gamma)] \tag{31} $$

or as

$$ {}_sN_{r+1}^{lm}(\gamma)={}_sK_r^{lm}(\gamma)-{}_sE_l^m(\gamma)+{}_sL_r^{lm}(\gamma)/{}_sN_r^{lm}(\gamma). \tag{32} $$

These expressions suggest that our treatment should be based upon continued fractions. Define quantities ${}_sN_r'^{lm}(\gamma)$ in terms of infinite continued fractions by

$$ {}_sN_r'^{lm}(\gamma)=\frac{{}_sL_r^{lm}(\gamma)}{{}_sE_l^m(\gamma)-{}_sK_r^{lm}(\gamma)}\ \frac{{}_sL_{r+1}^{lm}(\gamma)}{{}_sE_l^m(\gamma)-{}_sK_{r+1}^{lm}(\gamma)}\ +\ \cdots\ . \tag{33} $$

From the analysis given by Perron,[7] this infinite continued fraction is convergent for all values of ${}_sE_l^m(\gamma)$. Consequently, from the definition (33) we have

$$ {}_sN_r'^{lm}(\gamma)={}_sL_r^{lm}(\gamma)/[{}_sE_l^m(\gamma)-{}_sK_r^{lm}(\gamma)+{}_sN_{r+1}'^{lm}(\gamma)] \tag{34} $$

so that we can satisfy Eq. (31) by taking

$$ {}_sN_r^{lm}(\gamma)={}_sN_r'^{lm}(\gamma), \tag{35} $$

with ${}_sN_r'^{lm}(\gamma)$ given by (33). But by iteration of Eq. (32) with the condition ${}_sA_{lm}^{(r)}(\gamma)=0$ for $r<0$ we also have

$$ {}_sN_{r+1}^{lm}(\gamma)={}_sK_r^{lm}(\gamma)-{}_sE_l^m(\gamma)+\frac{{}_sL_r^{lm}(\gamma)}{{}_sK_{r-1}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ \frac{{}_sL_{r-1}^{lm}(\gamma)}{{}_sK_{r-2}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ +\ \cdots\ +\ \frac{{}_sL_1^{lm}(\gamma)}{{}_sK_0^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ . \tag{36} $$

Equating this continued fraction with the infinite convergent continued fraction ${}_sN_{r+1}'^{lm}(\gamma)$ obtained by putting $r=r+1$ in Eq. (33) gives rise to a transcendental equation for the determination of the eigenvalue ${}_sE_l^m(\gamma)$; specifically

$$ {}_sE_l^m(\gamma)={}_sK_r^{lm}(\gamma)\ \frac{{}_sL_r^{lm}(\gamma)}{{}_sK_{r-1}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ \frac{{}_sL_{r-1}^{lm}(\gamma)}{{}_sK_{r-2}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ +\ \cdots\ +\ \frac{{}_sL^{lm}(\gamma)}{{}_sK_0^{lm}(\gamma)-{}_sE_l^m(\gamma)}+\frac{{}_sL_{r+1}^{lm}(\gamma)}{{}_sK_{r+1}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ \frac{{}_sL_{r+2}^{lm}(\gamma)}{{}_sK_{r+2}^{lm}(\gamma)-{}_sE_l^m(\gamma)}\ +\ \cdots\ , \tag{37} $$

where r is less than or equal to the least value of l.

We wish to obtain a series expansion of the form

$$ {}_sE_l^m(\gamma)=\sum_{p=0}^{\infty}{}_sf_p^{lm}\gamma^p. \tag{38} $$

In light of Eqs. (15) and (16), it is convenient here to put $r=l-\max(|m|,|s|)$ in Eq. (37). The coefficients ${}_sf_p^{lm}$ are then obtained by substituting Eq. (38) into Eq. (37) and successively raising each denominator up to its associated numerator by binomial expansion and then equating coefficients. (Note that the smallest eigenvalue has $l=\max(|m|,|s|)$. If we let

$$ H(l)=\frac{[l^2-(\alpha+\beta)^2/4][l^2-s^2][l^2-(\alpha-\beta)^2/4]}{2[l-1/2]l^3[l+1/2]}, \tag{39} $$

where

$$ \frac{\alpha+\beta}{2}=\max(|m|,|s|)\ \text{ and }\ \frac{\alpha-\beta}{2}=\frac{ms}{\max(|m|,|s|)}\ , \tag{40} $$

then the first seven coefficients are given explicitly by

$$ {}_sf_0^{lm}=l(l+1), \tag{41a} $$

$$ {}_sf_1^{lm}=-2s^2m/l(l+1), \tag{41b} $$

$$ {}_sf_2^{lm}=H(l+1)-H(l)-1, \tag{41c}[8] $$

$$ {}_sf_3^{lm}=2s^2m\left[\frac{H(l)}{(l-1)l^2(l+1)}-\frac{H(l+1)}{(l+2)(l+1)^2l}\right], \tag{41d} $$

$$ \begin{aligned} {}_sf_4^{lm}=&\,4s^4m^2\left[\frac{H(l+1)}{(l+2)^2(l+1)^4l^2}-\frac{H(l)}{(l-1)^2l^4(l+1)^2}\right]\\ &+\frac{1}{2}\left[\frac{H^2(l+1)}{(l+1)}+\frac{H(l+1)H(l)}{(l+1)l}-\frac{H^2(l)}{l}\right]\\ &+\frac{1}{4}\left[\frac{[l-1]H(l)H(l-1)}{l(l-1/2)}-\frac{[l+2]H(l+1)H(l+2)}{(l+1)(l+3/2)}\right], \end{aligned} \tag{41e} $$

$$ \begin{aligned} {}_sf_5^{lm}=&\,8s^6m^3\left[\frac{H(l)}{(l-1)^3l^6(l+1)^3}-\frac{H(l+1)}{(l+2)^3(l+1)^6l^3}\right]\\ &+s^2m\left[\frac{3H^2(l)}{(l-1)l^3(l+1)}-\frac{[7l^2+7l+4]H(l)H(l+1)}{(l-1)l^3(l+1)^3(l+2)}\right.\\ &-\frac{3H^2(l+1)}{(l+2)(l+1)^3l}+\frac{1}{2}\left(\frac{[3l+7]H(l+1)H(l+2)}{(l+3)(l+3/2)(l+1)^3l}\right.\\ &\left.\left.-\frac{[3l-4]H(l)H(l-1)}{(l-2)(l-1/2)l^3(l+1)}\right)\right], \end{aligned} \tag{41f} $$

$$ \begin{aligned} {}_sf_6^{lm}=&\,16s^8m^4\left[\frac{H(l+1)}{(l+2)^4(l+1)^8l^4}-\frac{H(l)}{(l-1)^4l^8(l+1)^4}\right]\\ &+4s^4m^2\left[\frac{3H^2(l+1)}{(l+2)^2(l+1)^5l^2}\right.\\ &+\frac{[11l^4+22l^3+31l^2+20l+6]H(l)H(l+1)}{(l-1)^2l^5(l+1)^5(l+1)^2}\\ &-\frac{3H^2(l)}{(l-1)^2l^5(l+1)^2}+\frac{1}{2}\left(\frac{[3l^2-8l+6]H(l)H(l-1)}{(l-2)^2(l-1)(l-1/2)l^5(l+1)^2}\right. \end{aligned} $$

$$- \frac{[3l^2+14l+17]H(l+1)H(l+2)}{(l+3)^2(l+2)(l+3/2)(l+1)^5l^2}\Bigg)\Bigg]$$
$$+\frac{1}{4}\Bigg[\frac{2H^3(l+1)}{(l+1)^2}+\frac{[2l^2+4l+3]H^2(l)H(l+1)}{l^2(l+1)^2}$$
$$-\frac{[2l^2+1]H^2(l+1)H(l)}{(l+1)^2l^2}-\frac{2H^3(l)}{l^2}$$
$$+\frac{[l+2][3l^2+2l-3]H(l)H(l+1)H(l+2)}{4(l+3/2)^2(l+1)^2l}$$
$$-\frac{[l-1][3l^2+4l-2]H(l+1)H(l)H(l-1)}{4(l-1/2)^2l^2(l+1)}$$
$$+\frac{[l+2]H^2(l+2)H(l+1)}{4(l+3/2)^2(l+1)^2}-\frac{[l-1]^2H^2(l-1)H(l)}{4(l-1/2)^2l^2}$$
$$+\frac{[l-1][7l-3]H^2(l)H(l-1)}{4(l-1/2)^2l^2}$$
$$-\frac{[l+2][7l+10]H^2(l+1)H(l+2)}{4(l+3/2)^2(l+1)^2}$$
$$+\frac{[l+3]H(l+1)H(l+2)H(l+3)}{12(l+3/2)^2(l+1)}$$
$$-\frac{[l-2]H(l)H(l-1)H(l-2)}{12(l-1/2)^2l}\Bigg]. \tag{41g}$$

These coefficients may be used to give a first approximation to the eigenvalue. Successively closer approximations, together with accurate values of the ${}_sN_r^{lm}(\gamma)$ may be made by use of the method of Blanch[9] and Bouwkamp.[10] Finally the value of ${}_sA_{lm}^{(r)}(\gamma)/{}_sA_{lm}^{(0)}(\gamma)$ is given by

$$\frac{{}_sA_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}=\frac{((\alpha+\beta)/2+1)_r((\alpha+\beta+1)/2+1)_r}{\gamma^r(\alpha+1)_r(\beta+1)_r((\alpha+\beta)/2-s+1)_r}$$
$$\times\prod_{j=1}^{r}{}_sN_j^{lm}(\gamma). \tag{42}$$

The quantities ${}_sB_{lm}^{(r)}$ may be handled in a similar fashion. By defining the quantities ${}_sM_r^{lm}$, for integral $r\geq 1$ by

$${}_sM_r^{lm}(\gamma)=-\frac{\gamma(r+\alpha)(\alpha+\beta)[r+(\alpha+\beta)/2+s]}{[r+(\alpha+\beta)/2][r+(\alpha+\beta+1)/2]}\frac{{}_sB_{lm}^{(r)}(\gamma)}{{}_sB_{lm}^{(r-1)}(\gamma)}, \tag{43}$$

we find that ${}_sM_r^{lm}(\gamma)$ satisfies precisely the same equations as ${}_sN_r^{lm}(\gamma)$. Again the requirement that the series (20) should converge leads to precisely the same eigenvalue for the solution (20) as that obtained from the series (19). Consequently, ${}_sM_r^{lm}(\gamma)={}_sN_r^{lm}(\gamma)$.

From Eqs. (27) and (42) we find that

$$\frac{{}_sB_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(r)}(\gamma)}=\frac{{}_sB_{lm}^{(0)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}(-1)^r\frac{((\alpha+\beta)/2-s+1)_r}{((\alpha+\beta)/2+s+1)_r}. \tag{44}$$

However, provided that $2s$ is a positive integer,

$$\frac{((\alpha+\beta)/2-s+1)_r}{((\alpha+\beta)/2+s+1)_r}=\frac{((\alpha+\beta)/2-s+1)_{2s}}{((\alpha+\beta)/2-s+r+1)_{2s}}. \tag{45}$$

Consequently,

$$\frac{{}_sB_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(r)}(\gamma)}=(-1)^r\frac{{}_sB_{lm}^{(0)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}\frac{((\alpha+\beta)/2-s+1)_{2s}}{((\alpha+\beta)/2-s+r+1)_{2s}}, \tag{46}$$

and in a similar fashion we also have

$$\frac{{}_{-s}B_{lm}^{(r)}(\gamma)}{{}_sB_{lm}^{(r)}(\gamma)}=\frac{((\alpha+\beta)/2-s+r+1)_{2s}}{((\alpha+\beta)/2-s+1)_{2s}}\frac{{}_{-s}B_{lm}^{(0)}(\gamma)}{{}_sB_{lm}^{(0)}(\gamma)} \tag{47a}$$

and

$$\frac{{}_{-s}A_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(r)}(\gamma)}=\frac{((\alpha+\beta)/2-s+1)_{2s}}{((\alpha+\beta)/2-s+r+1)_{2s}}\frac{{}_{-s}A_{lm}^{(0)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}. \tag{47b}$$

5. NORMALIZATION

So far we have been able to show that the convergence of either (19) or (20) leads to an equation for the determination of the eigenvalues ${}_sE_l^m(\gamma)$ and that, given a correctly determined eigenvalue, the two sets of coefficients ${}_sA_{lm}^{(r)}(\gamma)$ and ${}_sB_{lm}^{(r)}(\gamma)$ are determined for positive integral values of r in terms of ${}_sA_{lm}^{(0)}(\gamma)$ and ${}_sB_{lm}^{(0)}(\gamma)$. We now come to the problem of the normalization of the solution. Clearly, two conditions are required to determine the two unknowns, ${}_sA_{lm}^{(0)}(\gamma)$ and ${}_sB_{lm}^{(0)}(\gamma)$. One of these conditions may be obtained by substituting $x=1$ in Eqs (9), (19), and (20) to obtain

$${}_sB_{lm}^{(0)}(\gamma)\sum_{r=0}^{\infty}\frac{{}_sB_{lm}^{(r)}(\gamma)}{{}_sB_{lm}^{(0)}(\gamma)}(r+\alpha)(r+\alpha-1)$$
$$=\exp(2\gamma)\,{}_sA_{lm}^{(0)}(\gamma)\sum_{r=0}^{\infty}\frac{{}_sA_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}(r+\alpha)(r+\alpha-1). \tag{48}$$

Provided neither of the infinite series vanishes, we obtain an equation for ${}_sB_{lm}^{(0)}/{}_sA_{lm}^{(0)}$. An equation for ${}_sA_{lm}^{(0)}(\gamma){}_sB_{lm}^{(0)}(\gamma)$ is provided by the normalization requirement

$$\int_{-1}^{1}[{}_sS_{lm}(\gamma;x)]^2\,dx=1 \tag{49}$$

since we can use (3) and (4) to write this equation as

$${}_sA_{lm}^{(0)}(\gamma)\,{}_sB_{lm}^{(0)}(\gamma)\int_{-1}^{1}\left(\frac{1-x}{2}\right)^{\alpha}\left(\frac{1+x}{2}\right)^{\beta}\left[\sum_{n=0}^{\infty}\frac{{}_sA_{lm}^{(n)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}P_n^{(\alpha,\beta)}(x)\right]$$
$$\times\left[\sum_{r=0}^{\infty}\frac{{}_sB_{lm}^{(r)}(\gamma)}{{}_sB_{lm}^{(0)}(\gamma)}P_r^{(\alpha,\beta)}(x)\right]dx=1, \tag{50}$$

which, in view of the uniform convergence of (19) and (20), becomes, on using Eq. (44),

$$\sum_{r=0}^{\infty}\frac{2(-1)^r\Gamma(r+\alpha+1)\Gamma(r+\beta+1)(1+(\alpha+\beta)/2-s)_r}{(2r+\alpha+\beta+1)r!\,\Gamma(r+\alpha+\beta+1)(1+(\alpha+\beta)/2+s)_r}$$
$$\times\left[\frac{{}_sA_{lm}^{(r)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}\right]^2=\frac{1}{{}_sA_{lm}^{(0)}(\gamma)\,{}_sB_{lm}^{(0)}(\gamma)}. \tag{51}$$

Consequently, we have equations for both ${}_sA_{lm}^{(0)}(\gamma)\,{}_sB_{lm}^{(0)}(\gamma)$ and ${}_sA_{lm}^{(0)}(\gamma)/{}_sB_{lm}^{(0)}(\gamma)$. The final determination of ${}_sA_{lm}^{(0)}(\gamma)$ and ${}_sB_{lm}^{(0)}(\gamma)$ is made by the requirement that the real part of ${}_sA_{lm}^{(0)}(\gamma)$ is to be positive.

6. REVERSING THE SIGN OF THE SPIN WEIGHT. RAISING AND LOWERING OPERATORS

Recently Teukolsky and Press[11] have given an explicit local transformation between quantities of opposite spin weight. In this section we rederive their results using our series (3) and (4). We begin by defining operators ${}_nT_m^-$ and ${}_nT_m^+$ by the equations

$${}_nT_m^-Q=-(1-x^2)^{1/2}\left(\frac{d}{dx}-\frac{m}{1-x^2}-\frac{nx}{1-x^2}\right)Q \tag{52}$$

and

$$ {}_nT_m^{+}Q = -(1-x^2)^{1/2}\left(\frac{d}{dx} + \frac{m}{1-x^2} + \frac{nx}{1-x^2}\right)Q. \tag{53} $$

Straightforward though tedious application of the identities (17) and (18) together with

$$ \frac{d}{dx}P_n^{(\alpha,\beta)}(x) = \tfrac{1}{2}(n+\alpha+\beta+1)P_{n-1}^{(\alpha+1,\beta+1)}(x) $$

then gives for $s > 0$ the results

$$ {}_{(1-s)}T_m^{-}\,{}_{(2-s)}T_m^{-}\cdots{}_sT_m^{-}\left[\left(\frac{1-x}{2}\right)^{\alpha/2}\left(\frac{1+x}{2}\right)^{\beta/2}P_r^{(\alpha,\beta)}\right] $$

$$ = (-1)^{[s+(\beta-\alpha)/2]}\left(\frac{\alpha+\beta}{2} - s + r + 1\right)_{2s} $$

$$ \times\left(\frac{1-x}{2}\right)^{\beta/2}\left(\frac{1+x}{2}\right)^{\alpha/2}P_r^{(\beta,\alpha)} \tag{54} $$

and

$$ {}_{s-1}T_m^{+}\,{}_{s-2}T_m^{+}\cdots{}_sT_m^{+}\left[\left(\frac{1-x}{2}\right)^{\beta/2}\left(\frac{1+x}{2}\right)^{\alpha/2}P_r^{(\beta,\alpha)}\right] $$

$$ = (-1)^{[s+(\alpha-\beta)/2]}\left(\frac{\alpha+\beta}{2} - s + r + 1\right)_{2s} $$

$$ \times\left(\frac{1-x}{2}\right)^{\alpha/2}\left(\frac{1+x}{2}\right)^{\beta/2}P_r^{(\alpha,\beta)}. \tag{55} $$

Consequently, if we define the operator L_n by

$$ L_nQ = -(1-x^2)^{1/2}\left(\frac{d}{dx} - \frac{m}{1-x^2} + \gamma - \frac{nx}{1-x^2}\right)Q, \tag{56} $$

note that

$$ L_n(\exp(-\gamma x)Q) = \exp(-\gamma x)\,{}_nT_m^{-}Q \tag{57} $$

and use Eqs. (8), (20), (47a), and (54), we find that

$$ L_{-s+1}L_{-s+2}\cdots L_s\,{}_sS_{lm}(\gamma;x) $$

$$ = (-1)^{[s+(\beta-\alpha)/2]}\frac{{}_sB_{lm}^{(0)}(\gamma)}{{}_{-s}B_{lm}^{(0)}(\gamma)}\left(\frac{\alpha+\beta}{2} - s + 1\right)_{2s}\,{}_{-s}S_{lm}(\gamma;x) \tag{58} $$

$$ = C_s\,{}_{-s}S_{lm}(\gamma;x), \tag{59} $$

say. But by a result of Teukolsky and Press[11] the constant C_s can be determined for any positive integral $2s$. The only cases of interest are $s = \frac{1}{2}, 1, 2$ for which we have

$$ C_{1/2} = -(Q+\tfrac{1}{4})^{1/2}, \tag{60a} $$

$$ C_1 = (Q^2 + 4\gamma m - 4\gamma^2)^{1/2}, \tag{60b} $$

and

$$ C_2 = \{(Q^2+4\gamma m-4\gamma^2)[(Q-2)^2+36\gamma m-36\gamma^2] $$

$$ + (2Q-1)(96\gamma^2-48\gamma m) - 144\gamma^2\}^{1/2}, \tag{60c} $$

where

$$ Q = {}_sE_l^m(\gamma) + \gamma^2 - 2\gamma m. \tag{60d} $$

Thus

$$ {}_{-s}B_{lm}^{(0)}(\gamma) = (-1)^{[s+(\beta-\alpha)/2]}\frac{((\alpha+\beta)/2 - s + 1)_{2s}}{C_s}\,{}_sB_{lm}^{(0)}, \tag{61} $$

which enables us to determine ${}_{-s}B_{lm}^{(0)}(\gamma)$, once ${}_sB_{lm}^{(0)}(\gamma)$ is known. On using (47a) we also have

$$ {}_{-s}B_{lm}^{(r)}(\gamma) = (-1)^{[s+(\beta-\alpha)/2]}\frac{((\alpha+\beta)/2 - s + r + 1)_{2s}}{C_s}\,{}_sB_{lm}^{(r)}. \tag{62} $$

If we define $L_n^{\dagger}$ by

$$ L_n^{\dagger}Q = -(1-x^2)^{1/2}\left(\frac{d}{dx} + \frac{m}{1-x^2} - \gamma - \frac{nx}{1-x^2}\right)Q $$

and proceed in a similar fashion using Eqs. (7), (19), (47b), and (55), we obtain

$$ L_{-s+1}^{\dagger}L_{-s+2}^{\dagger}\cdots L_s^{\dagger}\,{}_{-s}S_{lm}(\gamma;x) $$

$$ = (-1)^{[s+(\alpha-\beta)/2]}\frac{{}_{-s}A_{lm}^{(0)}(\gamma)}{{}_sA_{lm}^{(0)}(\gamma)}\left(\frac{\alpha+\beta}{2} - s + 1\right)_{2s}\,{}_sS_{lm}(\gamma;x). \tag{63} $$

In order to make the normalizations consistent, this latter expression must also be equal to

$$ C_s\,{}_sS_{lm}(\gamma;x). $$

Thus

$$ {}_{-s}A_{lm}^{(0)}(\gamma) = C_s\left(\frac{\alpha+\beta}{2} - s + 1\right)_{2s}(-1)^{[s+(\alpha-\beta)/2]}\,{}_sA_{lm}^{(0)}(\gamma) \tag{64} $$

TABLE I. Typical eigenvalues for a range of $a\omega$.

$s=2$ $l=4$ $m=5$			
$a\omega$		Eigenvalue	
Real	Imag.	Real	Imag.
2.50	2.50	17.39773077	−5.66344689
2.50	2.00	17.01558312	−4.56332174
2.50	1.50	16.61410548	−3.45766590
2.50	1.00	16.28167956	−2.32605871
2.50	0.50	16.06600456	−1.16997224
2.50	0	15.99158250	0
2.00	2.50	18.57791007	−5.27323127
2.00	2.00	18.10954737	−4.13986680
2.00	1.50	17.69356737	−3.08480799
2.00	1.00	17.36799810	−2.05365496
2.00	0.50	17.16145520	−1.02705010
2.00	0	17.09077118	0
1.50	2.50	19.74556118	−4.69370520
1.50	2.00	19.15136680	−3.65113239
1.50	1.50	18.68072323	−2.69303775
1.50	1.00	18.33328939	−1.77884449
1.50	0.50	18.11917091	−0.88539869
1.50	0	18.04677552	0
1.00	2.50	20.76262438	−3.97052818
1.00	2.00	20.07601364	−3.08172634
1.00	1.50	19.54581684	−2.26033668
1.00	1.00	19.16479619	−1.48497071
1.00	0.50	18.93421146	−0.73644245
1.00	0	18.85693205	0
0.50	2.50	21.60196115	−3.16334833
0.50	2.00	20.84741563	−2.44354058
0.50	1.50	20.26302111	−1.78229220
0.50	1.00	19.84603024	−1.16507990
0.50	0.50	19.59561798	−0.57586109
0.50	0	19.51206790	0
0	2.50	22.26768762	−2.28909746
0	2.00	21.44873661	−1.74532034
0	1.50	20.81323599	−1.26012241
0	1.00	20.36086577	−0.81750859
0	0.50	20.09013160	−0.40216282

TABLE II. Typical expansion coefficients.

$s=2$	$l=4$	$m=5$

$a\omega = 1.5 + i2.0$ Eigenvalue $= 18.80496813 - i4.41751318$

	${}_sA_{lm}(r)(a\omega)$		${}_sB_{lm}(r)(a\omega)$	
r	Real	Imag.	Real	Imag.
0	0.912373602	−0.329432964	0.146796596	−6.114384929
1	−2.239413164	1.766013426	−1.755579943	−5.731174356
2	3.191155441	0.090524274	1.116787865	−2.649810011
3	−1.521163756	−1.668853332	0.948655825	−0.366449175
4	−0.124857053	1.194768320	0.267339123	0.137361234
5	0.426436151	−0.292867252	0.017802430	0.075593104
6	−0.180807050	−0.051354958	−0.010827490	0.014324403
7	0.023383874	0.054379069	−0.003767549	0.000134193
8	0.007732272	−0.014543782	−0.000474301	−0.000549990
9	−0.004040707	0.000744673	0.000024218	−0.000127115
10	0.000697596	0.000613857	0.000019049	−0.000009885
11	0.000011122	−0.000191956	0.000003030	0.000001384
12	−0.000030202	0.000020811	0.000000112	0.000000473
13	0.000006185	0.000001970	−0.000000042	0.000000052
14	−0.000000366	−0.000001007	−0.000000009	−0.000000000
15	−0.000000089	0.000000140	−0.000000001	−0.000000001
16	0.000000024	−0.000000002	0.000000000	−0.000000000
17	−0.000000002	−0.000000002	0.000000000	0.000000000
18	−0.000000000	0.000000000	0.000000000	0.000000000

and so

$$ {}_{-s}A_{lm}^{(r)}(\gamma) = C_s \frac{(-1)^{[s+(\alpha-\beta)/2]}}{((\alpha+\beta)/2 - s + r + 1)_{2s}} \, {}_sA_{lm}^{(0)}(\gamma). \tag{65} $$

Note that (62) and (64) require that this normalization be equivalent to

$$ {}_sA_{lm}^{(0)}(\gamma)\, {}_sB_{lm}^{(0)}(\gamma) = (-1)^{\alpha-\beta} \, {}_{-s}A_{lm}^{(0)}(\gamma)\, {}_{-s}B_{lm}^{(0)}(\gamma) \tag{66} $$

and, consequently,

$$ {}_sA_{lm}^{(r)}(\gamma)\, {}_sB_{lm}^{(r)}(\gamma) = (-1)^{\alpha-\beta} \, {}_{-s}A_{lm}^{(r)}\, {}_{-s}B_{lm}^{(r)}(\gamma). \tag{67} $$

7. NUMERICAL COMPUTATION

The analytic techniques developed in this paper have been used by one of us (R. G. C.) to implement a computer program for the calculation of the eigenvalues and the expansion coefficients for a given frequency and for given values of s, l, and m. A feature of this program is that the eigenvalues are obtained to high accuracy (more than eight significant figures for both real and complex values of γ). Some typical eigenvalues and expansion coefficients are given in Tables I and II. The program has been checked both by comparison with published results for the eigenvalue for real frequencies and also by the use of the raising and lowering identities (66) and (67) in the case of complex frequencies.

8. CONCLUSION

We have shown how the analytic properties of spin-weighted angular spheroidal functions may be investigated by the use of infinite series of Jacobi polynomials.

*Part of the work was carried out whilst the author was at Monash University, Clayton, Victoria.

[1]S.A. Teukolsky, Astrophys J. 185, 635 (1973).
[2]W.H. Press and S.A. Teukolsky, Astrophys. J. 185, 649 (1973).
[3]E. Wasserstrom, J. Comp. Phys. 9, 53 (1972).
[4]Remark (3) following Theorem 9.2.1 in G. Szego, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, Vol. XXIII (American Mathematical Society, Providence, R.I., 1939).
[5]Carson Flammer, *Spheroidal Wave Functions* (Stanford U. P., Stanford, California, 1957).
[6]J. Meixner and F.W. Schäfke, *Mathieusche Funktionen und Spharoidfunktionen* (Springer-Verlag, Berlin, 1954).
[7]O. Perron, *Die Lehre von den Kettenbruchen* (Teubner, Stuttgart, 1957), 3rd. ed., Vol. II.
[8]For $s=0$, Eq. (41d) gives the value $2[m^2 - l(l+1) + \frac{1}{2}]/(2l-1)(2l+3)$ which agrees with the corresponding expression for $s=0$ given by Flammer in Ref. 5. The expression given by Press and Teukolsky [Eq. (3)—(10) of Ref. 2], however, differs from this and is wrong.
[9]G. Blanch, J. Math. and Phys. 25, 1 (1946).
[10]C.J. Bouwkamp, J. of Math. and Phys. 26, 79 (1947); also, Philips Res. Rep. 5, 87 (1950).
[11]S.A. Teukolsky and W.H. Press, Astrophys. J. 193, 443 (1974); see also A.A. Starobinskiĭ and S.M. Churilvo, Zh. Eksp. Teor. Fiz. 65, 3 (1973) [Sov. Phys. JETP 38, 1 (1974)].

Generalization of the inversion equations and application to nonlinear partial differential equations. I

Henri Cornille

Cen-Saclay, B. P. N°2, 91190 Gif-sur-Yvette, France
(Received 17 December 1976)

The generalization of the inversion equations corresponding to a system of n linear coupled equations is obtained in two different cases: first for the second order differential linear operator $\mu^2(x)[\partial^2/\partial x^2 - V_0(x)]$ and second for the linear first order operator $\mu(x)\partial/\partial x$. For the latter (which is an extension of the 2×2 case of Ablowitz *et al.*) we consider only a twofold eigenvalue problem for the $n\times n$ system, leaving the more general case to a subsequent paper. The equations of inversion are obtained by simple algebraic method, without considering both the direct and inverse scattering problem as in the classical method and without investigating the properties of the Jost solutions in the complex eigenvalue plane. Always using pedestrian algebraic method, and assuming that the kernel of the integral equation satisfies a linear partial differential equation, we deduce, for the solution of the integral equation, the corresponding nonlinear part of the evolution equation. In this way we show that the explicit construction of the classical solvable nonlinear evolution equations can be extended in two cases: coupled equations and substitution of $\mu(x)\partial/\partial x$ for $\partial/\partial x$.

I. INTRODUCTION

In recent years, the method of inverse scattering has appeared as a useful tool in order to solve a class of physically interesting nonlinear partial differential equations[1] (n. l. p. d. e.). This method was first applied by Gardner, Greene, Kruskal and Miura[2a] to the KdV equation, reviewed by Lax[2b] in an alternative formulation which permitted Zakharov and Shabat[2c] to solve the Schrödinger cubic equation. Subsequently,[2d] the sine–Gordon being also reducible to this method led Ablowitz, Kaup, Newell, and Segur[2e] to the conclusion that all these cases can be solved by the study of an associated 2×2 system of first order linear differential equations for which the coefficients or "potentials" are the solution of the nonlinear evolution equations.

Thus there exist n. l. p. d. e. which are associated with a linear system (time independent)of eigenvalue equations whose spectrum remains invariant under the evolution in time of the coefficients; they are a completely integrable system. We are mainly faced with the problem of the reconstruction of the coefficients ("potentials") of these linear systems. The classical method[1-3] consists of the study of the direct and inverse scattering problem and of the investigation of the corresponding Jost functions in the complex eigenvalue plane.

However, the method of derivation of the inversion equation becomes more and more complicated as we can see for the second order eigenvalue problem (Ref. 2 for the 2×2 system and Manakov[3a] for the 3×3 system) as well as for the third order[3b] eigenvalue problem (Zakharov, Manakov, and Kaup). In the general nth order eigenvalue and $m\times m$ linear system[4] ($m \geq n$) even if one can write down the expected n. l. p. d. e. (see Ablowitz and Haberman, it is not obvious whether or not one can explicitly deduce the corresponding inversion equation using the classical method recalled above).

So we must find another entirely different method which could be easily generalized to an arbitrary nth order eigenvalue problem and arbitrary $m\times m$ linear system and this is the first aim of the present paper and of the subsequent one.[5]

However, for a second order differential equation, in the scalar case, a method giving the inversion equations, which does not at all require the study of the Jost function in the eigenvalue plane, was recently given.[6] We sketch here very briefly this method because it is the object of the present paper to generalize it (Sec. II) to second and first order (twofold eigenvalue) differential equations and the arbitrary $n\times n$ matrix case.

(i) We start from a second order differential operator

$$\Delta_0 = \mu^2(x)\left(\frac{\partial^2}{\partial x^2} - V_0(x)\right), \tag{1}$$

where μ and V_0 are arbitrary. We associate a kernel F satisfying

$$\begin{aligned} &O_d F = [\Delta_0(x) - \Delta_0(y)]F(x,y) = 0, \\ &\lim_{x\to a} F(x,y) = \lim_{y\to a} F(x,y) = 0, \end{aligned} \tag{2}$$

$$F = \sum \phi_n^0(x)\,\phi_n^0(y)\,C_n \tag{3}$$

and where ϕ_n^0 is an eigenvalue $\Delta_0\phi_n^0 = \delta_n\phi_n^0$ (the sum $\sum$ could be as well an integral or both). a is a fixed constant.

(ii) We assume that there exists a unique Fredholm type solution $K(x,y)$ of the integral equation

$$K(x,y) + F(x,y) + \int_x^a F(s,y)\,K(x,s)\,\mu^{-2}(s)\,ds \tag{4}$$

(except perhaps for discrete values $x_1, x_2, \cdots$).

(iii) We assume the following boundary condition:

$$\lim_{s\to a}\left\{K(x,s)\frac{\partial}{\partial s}F(s,y) - F(s,y)\frac{\partial}{\partial s}K(x,s)\right\} = 0. \tag{5}$$

(iv) We define a reconstructed second order operator

$$\begin{aligned} &\Delta = \Delta_0 + 2\mu\frac{d}{dx}\left(\frac{K(x,x)}{\mu(x)}\right) = \mu^2\left(\frac{\partial^2}{\partial x^2} - V(x)\right), \\ &V = V_0 - \frac{2}{\mu}\frac{d}{dx}\left(\frac{\hat{K}}{\mu}\right), \quad \hat{K} = K(x,x), \end{aligned} \tag{6}$$

where V is the reconstructed potential and V_0 the comparison potential. If (i), (ii), (iii) and (iv) are satisfied then we got the Property I:

$$[\Delta(x) - \Delta_0(y)]K(x,y) = 0. \tag{7}$$

(v) We consider an eigenfunction U_n^0 of Δ_0 such that $\Delta_0 U_n^0 = \gamma_n U_n^0$.

(vi) We define a function U_n such that

$$U_n(x) = U_n^0(x) + \int_x^a U_n^0(s)\,K(x,s)\,\mu^{-2}(s)\,ds. \tag{8}$$

(vii) We assume the boundary condition

$$\lim_{s\to a}\left\{K(x,s)\frac{\partial}{\partial s}U_n^0(s) - U_n^0(s)\frac{\partial}{\partial s}K(x,s)\right\} = 0. \tag{9}$$

If the assumptions leading to Theorem I as well as (v), (vi), (vii) are satisfied, then we have the Property II:

$$(\Delta - \gamma_n)U_n = 0, \quad \left[\mu^2\left(\frac{\partial^2}{\partial x^2} - V\right) - \gamma_n\right]U_n = 0. \tag{10}$$

Thus Eq. (4) is the integral inversion equation corresponding to the scalar, second order differential equation (10), and in order to obtain this result we use only pedestrian algebraic relations. Over the traditional method we also gain two other advantages. First as is widely discussed in Ref. 5, the kernel F and consequently K and the potential V could be eigenvalue dependent if V_0 (or Δ_0) contains a const μ^{-2} term). This open door gives the possibility of reconstructing energy dependent or angular momentum dependent potentials. (On the contrary we recall that in the traditional method the investigation of the direct problem is done with energy independent potentials). Secondly, we have at our disposal an arbitrary $\mu(x)$ function, and so we enlarge the class of inversion equation, encompassing, for instance, the inversion at fixed k or at fixed l.[6]

In Sec. III we extend the Ablowitz *et al.* method,[2d] giving the possibility of finding (without explicitly constructing the solutions) the n. l. p. d. e. associated with a linear differential system. We consider the system of linear operators $D(x) = \mu(x)\,\partial/\partial_x$, with twofold eigenvalues in the matrix case. The main result is that the classical solvable n. l. p. d. e. are members of a larger family of exactly solvable n. l. p. d. e., where $\partial/\partial x$ is replaced by $D_0(x) = \mu(x)\,\partial/\partial x$. In Sec. IV we consider alone the inversion equation which was established in Sec. II B for the linear matrix operator $\mu(x)\,\partial/\partial x$ with a second order eigenvalue problem. We introduce as usual the time as a parameter. Investigating the properties of the solutions when the kernels satisfy linear partial differential equations (l. p. d. e.) we show that it is possible to determine directly (and we do) the corresponding n. l. p. d. e. That this possibility may exist results from previous works by Gardner *et al.*[2a] for the KdV and by Zakharov and Shabat[4] for other cases. In this way we get classical coupled n. l. p. d. e., and we extend their properties in two directions. First we show that they can still be completely integrable if $\partial/\partial x$ is replaced by $D_0(x)$. Secondly from the solutions of the inversion equations depending on two coordinates variables x and y we deduce the n. l. p. d. e. satisfied by the three variables (x, y, t) and only at the end do we restrict to $y = x$.

It appears very convenient to define an algebra of matrices in such a way that our problem originally written in an $n\times n$ space is finally reduced to a 2×2 space. In this way, the n. l. p. d. e. are primarily obtained in a matrix formulation and finally written down with scalar quantities. In the last section (V) we derive some properties of the inversion equation (studied in Secs. II—IV) and, for particular degenerate kernels, explicitly construct the solutions of the corresponding n. l. p. d. e.

In conclusion, in order to construct explicitly the solutions of the exactly solvable n. l. p. d. e., it is sufficient that the kernels of the inversion equations satisfy well-defined l. p. d. e. (and boundary conditions) whereas any reference to the scattering problem, corresponding to the linear associate system, does not seem very crucial.

II. GENERALIZATION OF THE INVERSION EQUATIONS IN THE MATRIX CASE

We follow the scheme briefly sketched in the introduction for the scalar case and second order differential equation.[6]

A. Second order differential operators

(i) We define a scalar and a matrix second order differential operators

$$D_0(x) = \mu^2(x)\left(\frac{\partial^2}{\partial x^2} - V_0(x)\right), \quad \Delta_0 = ID_0 = \begin{pmatrix} D_0 & \cdot & 0 \\ & \cdot & \\ 0 & \cdot & D_0 \end{pmatrix}, \tag{11}$$

$$O_d = D_0(x) - D_0(y),$$

where μ and V_0 are arbitrary x functions, I the identity matrix; and we associate a set $\{F_j^i\}$ of scalar kernels such that for $i = 1, \ldots, n$, $j = 1, \ldots, n$,

$$O_d F_j^i(x,y) = 0, \quad \lim_{x\to a} F_j^i(x,y) = \lim_{y\to a} F_j^i(x,y) = 0, \tag{12}$$

where a is a fixed constant, for instance, $\pm\infty$, 0, or any other fixed value. There exists a general method of constructing such kernels. Let us consider the eigenfunctions $\Delta_0\phi_n^0 = \delta_n\phi_n^0$ and associate the kernel

$$F_j^i(x,y) = \sum_n \phi_n^0(x)\,\phi_n^0(y)\,C_j^i(n) \tag{13}$$

the $C_j^i(n)$ being constants. The sum $\sum_n$ in (13), which can be an integral $\int$ can include discrete terms as well as a continuum. If $\lim_{x\to a}\phi_n^0(x) = 0$, both conditions in (12) are satisfied.

(ii) We define a matrix $(n\times n)$ kernel $\mathcal{J}$ and a matrix $(n\times n)$ solution $\mathcal{K}$

$$\mathcal{J} = \begin{pmatrix} F_1^1 & \cdots & F_n^1 \\ \vdots & & \vdots \\ F_1^n & & F_n^n \end{pmatrix}, \quad \mathcal{K} = \begin{pmatrix} K_1^1 & \cdots & K_n^1 \\ \vdots & & \vdots \\ K_1^n & & K_n^n \end{pmatrix}$$

of a Fredholm type integral equation

$$\mathcal{K}(x,y) = \mathcal{J}(x,y) + \int_x^a \mathcal{J}(s,y)\,\mathcal{K}(x,s)\,\mu^{-2}(s)\,ds, \tag{14}$$

where the weight function μ^{-2} is a scalar. We assume that the Fredholm type solution of Eq. (14) exists and is unique (except perhaps for discrete values $x_1, x_2, \cdots$).

(iii) We consider the boundary conditions

$$\lim_{s\to a}\left\{\left(\frac{\partial}{\partial s}F_j^i(s,y)\right)K_i^j(x,s)-F_j^i(s,y)\frac{\partial}{\partial s}K_i^j(x,s)\right\}=0. \tag{15}$$

Applying the boundary condition (12) to Eq. (14), we get $\lim_{s\to a}K(x,s)=\lim_{x\to a}K(x,s)=0$. It follows that in general (15) is satisfied. In the following we always assume that this is the case

(iv) We define the transposed matrix K^T,

$$K^T=\begin{pmatrix}K_1^1\cdots K_1^n\\ \vdots\quad\vdots\\ K_n^1\cdots K_n^n\end{pmatrix};$$

for a scalar $f(x,y)$ we define $\hat{f}(x,x)$ and for a matrix $\hat{M}=(M_j^i(x,x))$. It follows that $\hat{K}^T=K^T(x,x)$. We define a matrix second order differential operator:

$$\Delta=\Delta_0+2\mu\frac{d}{dx}\left(\frac{\hat{K}^T}{\mu}\right)=\begin{pmatrix}D_0+2\mu(\hat{K}_1^1/\mu)_x\cdots 2\mu(\hat{K}_1^n/\mu)_x\\ \vdots\qquad\qquad\vdots\\ 2\mu(\hat{K}_n^1/\mu)_x\quad\cdots\quad D_0+2\mu(\hat{K}_n^n/\mu)_x\end{pmatrix} \tag{16}$$

$$\Delta=\mu^2\left(I\frac{\partial^2}{\partial x^2}-V\right),\quad V=IV_0-\frac{2}{\mu}\frac{d}{dx}\left(\frac{\hat{K}^T}{\mu}\right),$$

where V is the reconstructed matrix potential and V_0 the comparison scalar potential and I the identity matrix.

Property I: If (i), (ii), (iii), (iv) are satisfied, then we get for the matrix K:

$$[\Delta(x)-\Delta_0(y)]K^T(x,y)=0. \tag{17}$$

In the proof of this property we shall use the following lemma.

Lemma I: Let $F, L, M,$ be scalar functions with $O_dF=0$ and the coupled functions $\{F,M\}$ satisfying the boundary conditions (iii). Then if

$$L(x,y)=F(x,y)+\int_x^a F(u,y)\,M(x,u)\,\mu^{-2}(u)\,du,$$

it follows that

$$O_dL=\int_x^a\mu^{-2}F(O_dM)\,ds-2\mu(x)F(x,y)(\hat{M}/\mu)_x. \tag{18}$$

From the integral Eq. (14) and applying Lemma I, we get

$$O_dK=-2\mu\mathcal{J}(\hat{K}/\mu)_x+\int_x^a\mu^{-2}\mathcal{J}(O_dK)\,ds,$$
$$2\mu K(\hat{K}/\mu)_x=2\mu\mathcal{J}(\hat{K}/\mu)_x+\int_x^a\mu^{-2}(s)\mathcal{J}2\mu(x)K(\hat{K}/\mu)_x\,ds. \tag{19}$$

The homogeneous integral equation corresponding to Eq. (14) having only the trivial solution, we get

$$O_dK+2\mu K\frac{d}{dx}\left(\frac{\hat{K}}{\mu}\right)=0,$$
$$O_dK^T+2\mu\frac{d}{dx}\left(\frac{\hat{K}^T}{\mu}\right)K^T=0. \tag{17'}$$

(v) We consider an eigenfunction U_n^0 such that $D_0U_n^0=\gamma_nU_n^0$. We write for these eigenfunctions (γ_n, U_n^0) a different notation from that of the eigenfunctions (δ_n,ϕ_n^0) with which we build our kernel F_j^i because U_n^0 is not necessarily a function of the set $\{\phi_n^0\}$. We define a set of eigenfunctions $\{\psi_j^0\}$ such that

$$\psi_1^0=\begin{pmatrix}\delta_{1,1}U_n^0\\ 0\end{pmatrix},\quad \psi_j^0=\begin{pmatrix}0\\ \delta_{j,i}U_n^0\\ 0\end{pmatrix},$$
$$j=1,\dots,n,\quad i=1,\dots,n,\quad \delta_{i,i}=1,\quad \delta_{j,i}=0\ \text{for}\ j\neq i$$
$$(\Delta_0-\gamma_nI)(\psi_1^0,\psi_2^0,\dots,\psi_n^0)=(0). \tag{20}$$

(vi) We define a set of functions $\{\psi_j\}$, $j=1,\dots,n$:

$$\psi_1=\begin{pmatrix}\delta_{1,1}U_n^0(x)+\int_x^a\mu^{-2}(s)\,K_1^1(x,s)\,U_n^0(s)\,ds\\ \vdots\\ \int_x^a\mu^{-2}K_n^1U_n^0\,ds\end{pmatrix}$$

$$\psi_j=\begin{pmatrix}\int_x^a\mu^{-2}K_1^jU_n^0\,ds\\ \vdots\\ U_n^0\delta_{j,i}+\int_x^a\mu^{-2}K_i^j(x,s)\,U_n^0(s)\,ds\\ \vdots\end{pmatrix},\quad j>1. \tag{21}$$

(vii) We assume the following boundary condition:

$$\lim_{s\to a}\left\{K_i^j(x,s)\frac{\partial}{\partial s}U_n^0(s)-U_n^0(s)\frac{\partial}{\partial s}K_i^j(x,s)\right\}=0. \tag{22}$$

We recall that $\lim_{s\to a}K_i^j(x,s)=0$. If either $\lim_{s\to a}(\partial/\partial s)K_i^j(x,s)=0$ or $\lim_{s\to a}U_n^0(s)=0$, then in general Eq. (22) is satisfied.

Property II: If the assumptions leading to the Property I are satisfied and if further (v), (vi), (vii) are verified, then ψ_j is an eigenfunction of the differential operator Δ:

$$(\Delta-I\gamma_n)(\psi_1,\psi_2,\dots,\psi_n)=(0),$$
$$(\Delta-I\gamma_n)=\begin{pmatrix}D_0-\gamma_n+2\mu(\hat{K}_1^1/\mu)_x,\dots,2\mu(\hat{K}_1^n/\mu)_x\\ \vdots\\ 2\mu(\hat{K}_n^1/\mu)_x,\dots,D_0-\gamma_n+2\mu(\hat{K}_n^n/\mu)_x\end{pmatrix}; \tag{23}$$

and for the proof of Eq. (23) we shall use the following lemma.

Lemma II: If the assumptions of Property II are satisfied, then

$$[D_0(x)-\gamma_n]\int_x^a\mu^{-2}(s)U_n^0(s)\,K_i^j(x,s)\,ds$$
$$=-2\mu(x)U_n^0(x)(\hat{K}_i^j)_x+\int_x^a\mu^{-2}U_n^0O_dK_i^j\,ds. \tag{24}$$

Applying this lemma, we get

$$(\Delta-I\gamma_n)(\psi_1,\psi_2,\dots,\psi_n)$$
$$=\int_x^a U_n^0\mu^{-2}\left[O_dK^T+2\mu\frac{d}{dx}\left(\frac{\hat{K}^T}{\mu}\right)K^T\right]ds=(0),$$

where we have used Property I in the equivalent formulation (17'). *Application* of Properties I and II to a linear $(n\times n)$ system of second order differential equations. In order to summarize the results, we slightly change the notations. Let us define

$$\left[\mu^2(x)\left(\frac{\partial^2}{\partial x^2}-V_0(x)\right)-\gamma\right]\psi=Q\psi,\quad \psi=\begin{pmatrix}v_1\\ \vdots\\ v_n\end{pmatrix},$$

$$Q=\begin{pmatrix} q_1^1 \cdots q_1^n \\ \vdots \quad \vdots \\ q_n^1 \cdots q_n^n \end{pmatrix}$$

and consider

$$(D_0(x)-\gamma)\,U_0(\gamma,x)=0, \quad F_i^j=\int U_0(v,x)\,U_0(v,y)\,C_i^j(v)\,dv,$$

$$O_d\mathcal{J}=0,$$

$$K=\mathcal{J}+\int \mu^{-2}\,\mathcal{J}K\,ds, \quad \mathcal{J}=(F_i^j)$$

$$\psi_j=\begin{pmatrix} \\ \\ U_0(\gamma,x)\,\delta_{j,i}+\int \mu^{-2}(s)\,U_0(s)\,K_i^j(x,s)\,ds \\ \\ \\ \end{pmatrix}, \quad j=1,\ldots,n;$$

then, if the assumptions leading to Properties I and II are satisfied, we get

$$\left[\mu^2(x)\left(\frac{\partial^2}{\partial x^2}-V_0(x)\right)-\gamma\right]\psi_j=Q\psi_j,$$

$$q_i^j=-2\mu(x)\,\frac{d}{dx}\left(\frac{K_i^j(x,x)}{\mu(x)}\right). \tag{25}$$

B. First order linear operators and associate linear system with twofold eigenvalues

(i) We define a scalar and a matrix $((n+1)\times(n+1))$ first order differential operator

$$D_0(x)=\mu(x)\,\frac{\partial}{\partial x}, \quad O_d=D_0(x)-D_0(y),$$

$$\Delta_0=ID_0=\begin{pmatrix} D_0 & & 0 \\ & \ddots & \\ 0 & & D_0 \end{pmatrix}, \tag{26}$$

where μ is an arbitrary x function, and associate two sets $\{F_i\}$, $\{G_i\}$ of scalar kernels such that

$$O_d F_i=O_d G_i=0, \quad i=1,\ldots,n,$$

$$\lim_{x\to\infty} F_i(x,y)=\lim_{y\to\infty} F_i(x,y) \tag{27}$$

$$=\lim_{x\to\infty} G_i(x,y)=\lim_{y\to\infty} G_i(x,y)=0.$$

F_iG_i are constructed with the eigenfunctions $\Delta_0\phi_n^0=\delta_n\phi_n^0$:

$$F_i=\sum \phi_n^0(x)\,\phi_n^0(y)f_{n,i}, \quad G_i=\sum \phi_n^0(x)\,\phi_n^0(y)\,g_{n,i}, \tag{28}$$

$f_{n,i}$ and $g_{n,i}$ being constants; here also the sun can include discrete and continuous terms.

(ii) Let $\mathcal{M}$ (or $\mathcal{N}$) be any matrix that we separate in two parts. We define the first component (or matrix of type I), the remaining of the $\mathcal{M}$ matrix when we put equal to zero the first line first column element, all the elements of the second line (except the first one) being zero, all the elements of the third line (except the first one) being zero, all the elements of the third line (except the first one) being zero, ..., and so on

$$\begin{Bmatrix} \mathcal{M}_{\rm I} \\ \text{or} \\ \mathcal{N}_{\rm I} \end{Bmatrix} \Rightarrow \begin{pmatrix} 0 & \\ \text{▨} & 0 \end{pmatrix}. \tag{29}$$

Consequently, we define the second component (or matrix of type II) $\mathcal{M}_{\rm II}=\mathcal{M}-\mathcal{M}_{\rm I}$ as the remaining part of $\mathcal{M}$ when all the elements of the first line (except the first) are zero and all the elements of the first column (except the first) are zero:

$$\begin{Bmatrix} \mathcal{M}_{\rm II} \\ \text{or} \\ \mathcal{N}_{\rm II} \end{Bmatrix} \Rightarrow \begin{pmatrix} \square & \cdots 0 \cdots \\ \vdots & \\ 0 & \text{▨} \\ \vdots & \end{pmatrix}. \tag{30}$$

We easily get the rule for the product of two matrices.

$$(\mathcal{MN})_{\rm I}=\mathcal{M}_{\rm I}\mathcal{N}_{\rm II}+\mathcal{M}_{\rm II}\mathcal{N}_{\rm I},$$

$$(\mathcal{MN})_{\rm II}=\mathcal{M}_{\rm I}\mathcal{N}_{\rm I}+\mathcal{M}_{\rm II}\mathcal{N}_{\rm II}. \tag{31}$$

We define a matrix $(n+1\times n+1)$ kernel $\mathcal{J}\equiv\mathcal{J}_{\rm I}$ with only the first component and a matrix $(n+1\times n+1)$ solution K:

$$\mathcal{J}\equiv\mathcal{J}_{\rm I}=\begin{pmatrix} 0 & -G_1,\ldots,-G_n \\ F_1 & \\ \vdots & 0 \\ F_n & \end{pmatrix},$$

$$K=K_{\rm I}+K_{\rm II}=\begin{pmatrix} K_1^1 \cdots K_{n+1}^1 \\ \vdots \quad \vdots \\ K_1^{n+1} \cdots K_{n+1}^{n+1} \end{pmatrix}, \tag{32}$$

$$K_{\rm I}=\begin{pmatrix} 0 & K_2^1 \circ\circ\circ K_{n+1}^1 \\ K_1^2 & \\ \vdots & 0 \\ K_1^{n+1} & \end{pmatrix}, \quad K_{\rm II}=\begin{pmatrix} K_1^1 & 0 \\ & K_2^2 \circ\circ\circ K_{n+1}^2 \\ 0 & \vdots \quad \vdots \\ & K_2^{n+1} \circ\circ\circ K_{n+1}^{n+1} \end{pmatrix}$$

of a linear integral equation

$$K(x,y)=\mathcal{J}(x,y)+\int_x^\infty \mu^{-1}(s)\,\mathcal{J}(s,y)K(x,s)\,ds. \tag{33}$$

We still assume that the Fredholm type solution of Eq. (33) exists and is unique (except perhaps for discrete values $x_1,x_2,\cdots$).

(iii) We assume the boundary condition

$$\lim_{s\to\infty} K_j^i(x,s)\,F_k(s,y)=0, \quad \lim_{s\to\infty} K_j^i(x,s)\,G_k(s,y)=0, \tag{34}$$

which is in general satisfied due to the boundary condition (27).

(iv) We define the transposed matrix

$$K^T=K_{\rm I}^T+K_{\rm II}^T,$$

$$K^T=\begin{pmatrix} K_1^1 \cdots K_1^{n+1} \\ \vdots \quad \vdots \\ K_{n+1}^1 \cdots K_{n+1}^{n+1} \end{pmatrix}, \quad K_{\rm II}^T=\begin{pmatrix} K_1^1 & 0 \\ & K_2^2 \cdots K_1^{n+1} \\ 0 & \vdots \quad \vdots \\ & K_{n+1}^2 \circ\circ K_{n+1}^{n+1} \end{pmatrix}; \tag{35}$$

we define a matrix first order differential operator:

$$\Delta=\Delta_0+2\hat{K}_{\rm I}^T=\begin{pmatrix} D_0 & 2\hat{K}_2^1 \cdots 2\hat{K}_1^{n+1} \\ 2\hat{K}_2^1 & D_0 \ddots \quad 0 \\ \vdots & \ddots \ddots \\ 2\hat{K}_{n+1}^1 & 0 \quad \ddots \\ & \quad\quad\quad D_0 \end{pmatrix} \tag{36}$$

$$\Delta = \mu\left(I\frac{\partial}{\partial x} - V\right), \quad V = -2\hat{K}_{\mathrm{I}}^{T}, \quad O_s = D_0(x) + D_0(y),$$

$$\Delta(x) - \Delta_0(y) = O_d I + 2\hat{K}_{\mathrm{I}}^{T}, \quad \Delta(x) + \Delta_0(y) = O_s I + 2\hat{K}_{\mathrm{I}}^{T},$$

where V is the reconstructed matrix potential.

Property I: If (i), (ii), (iii), (iv) are satisfied, then we get for the K matrix

$$(\Delta(x) - \Delta_0(y))K_{\mathrm{I}}^{T} + (\Delta(x) + \Delta_0(y))K_{\mathrm{II}}^{T} = (0). \tag{37}$$

For the proof we shall use the following lemma:

Lemma III: Let L, H, M, N, be scalar functions with $O_d M = O_d N = 0$ and the coupled functions $\{N, H\}$ satisfying the boundary conditions (iii). Then, if

$$L(x,y) = M(x,y) + \int_x^{\infty} \mu^{-1}(s)\, N(s,y)\, H(x,s)\, ds,$$

it follows that

$$\begin{aligned} O_d L &= \int \mu^{-1} N (O_s H)\, ds, \\ O_s L &= O_s M - 2N\hat{H} + \int \mu^{-1} N (O_d H)\, ds. \end{aligned} \tag{38}$$

Applying the lemma for matrices $\mathcal{M} = (M_i^j)$, $\mathcal{N} = (N_i^j)$, $\mathcal{H} = (H_i^j)$, $\mathcal{L} = (L_i^j)$, we get that if $\mathcal{L} = \mathcal{M} + \int_x^{\infty} ds\, \mu^{-1} \mathcal{N}\mathcal{H}$, it follows that

$$\begin{aligned} O_d \mathcal{L} &= \int \mu^{-1} \mathcal{N} O_d \mathcal{H}\, ds, \\ O_s \mathcal{L} &= O_s \mathcal{M} - 2\mathcal{N}\hat{\mathcal{H}} + \int \mu^{-1} \mathcal{N} O_d \mathcal{H}\, ds. \end{aligned} \tag{38'}$$

Equation (33) can be written in a (2×2) matrix formulation where the first line element of the vector solution refers to component I and the second line to component II:

$$\begin{pmatrix} K_{\mathrm{I}} \\ K_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} \mathcal{J} \\ 0 \end{pmatrix} + \int_x^{\infty} \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} K_{\mathrm{I}} \\ K_{\mathrm{II}} \end{pmatrix} ds. \tag{33}$$

Applying Lemma III, we get

$$\begin{pmatrix} O_d\, K_{\mathrm{I}} \\ O_s\, K_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} 0 \\ -2\mathcal{J}\hat{K}_{\mathrm{I}} \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} O_d\, K_{\mathrm{I}} \\ O_s\, K_{\mathrm{II}} \end{pmatrix}, \tag{39}$$

$$\begin{pmatrix} O_d K_{\mathrm{I}} + 2K_{\mathrm{II}}\hat{K}_{\mathrm{I}} \\ O_s K_{\mathrm{II}} + 2K_{\mathrm{I}}\hat{K}_{\mathrm{I}} \end{pmatrix} = \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} O_d K_{\mathrm{I}} + 2K_{\mathrm{II}}\hat{K}_{\mathrm{I}} \\ O_s K_{\mathrm{II}} + 2K_{\mathrm{I}}\hat{K}_{\mathrm{I}} \end{pmatrix} ds. \tag{40}$$

The homogeneous part of Eq. (33) having only the trivial solution, we get

$$\begin{aligned} &O_s K_{\mathrm{II}} + 2K_{\mathrm{I}}\hat{K}_{\mathrm{I}} = 0, \quad \widehat{O_s K_{\mathrm{II}}} = D_0(x)\hat{K}_{\mathrm{II}} = -2\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}}, \\ &D_0\hat{K}_1^1 + 2\sum_{2}^{n+1} \hat{K}_j^1 \hat{K}_1^j = 0, \quad D_0\hat{K}_j^1 + 2\hat{K}_j^1\hat{K}_1^1 = 0 \ \text{ for } j > 1, \end{aligned} \tag{41a}$$

$$\begin{aligned} &O_d K_{\mathrm{I}} + 2K_{\mathrm{II}}\hat{K}_{\mathrm{I}} = 0, \\ &O_d K_{\mathrm{I}} + O_s K_{\mathrm{II}} + 2K\hat{K}_{\mathrm{I}} = 0, \quad O_d K_{\mathrm{I}}^{T} + O_s K_{\mathrm{II}}^{T} + 2\hat{K}_{\mathrm{I}}^{T} K^{T} = 0. \end{aligned} \tag{42b}$$

The last relation in (42b) is equivalent to Eq. (37).

(v) We consider two eigenvalues and eigenfunctions of D_0

$$D_0 U_0(\pm\varphi, x) = \pm i\varphi U_0(\pm\varphi, x) \tag{43}$$

and associate a set of eigenfunctions

$$\psi_1^0 = \begin{pmatrix} \delta_{1,1} U_0(-\varphi, x) \\ 0 \end{pmatrix}, \quad \psi_2^0 = \begin{pmatrix} 0 \\ \delta_{2,2} U_0(\varphi, x) \\ 0 \end{pmatrix},$$

$$\psi_j^0 = \begin{pmatrix} 0 \\ \delta_{j,i} U_0(\varphi, x) \\ 0 \end{pmatrix}, \quad j \geq 2, \quad i \geq 2$$

$$(\Delta_0 - iy\tilde{I})(\psi_1^0, \psi_2^0, \ldots, \psi_{n+1}^0) = (0),$$

$$\tilde{I} = \begin{pmatrix} -1 & & & \\ & 1 & & 0 \\ & & 1 & \\ 0 & & & \ddots \\ & & & & 1 \end{pmatrix}. \tag{44}$$

(vi) We define a set of functions $\{\psi_j\}$:

$$\psi_1 = \begin{pmatrix} \delta_{11} U_0(-\varphi, x) + \int_x^{\infty} \mu^{-1} U_0(-\varphi, t)\, K_1^1(x,t)\, dt \\ \vdots \\ \int \mu^{-1} U_0 K_{n+1}^1\, dt \end{pmatrix},$$

$$\psi_j = \begin{pmatrix} \vdots \\ \delta_{j,i} U_0(\varphi, x) + \int_x^{\infty} \mu^{-1} U_0(\varphi, t)\, K_i^j(x,t)\, dt \\ \vdots \end{pmatrix},$$

$$j \geq 2, \quad i \geq 2.$$

(vii) We assume the following boundary condition:

$$\lim_{s\to\infty} U_0(-\varphi, s)\, K_i^1(x,s) = 0, \quad \lim_{s\to\infty} U_0(\varphi, s)\, K_i^j(x,s) = 0,$$

$$j \geq 2, \quad i \geq 2. \tag{45}$$

Property II: If the assumptions leading to Property I as well as (v), (vi), (vii) are satisfied, then we get

$$(\Delta - i\varphi\tilde{I})(\psi_1, \psi_2, \ldots, \psi_{n+1}) = (0); \tag{46}$$

for the proof we shall use the following lemma:

Lemma IV: If the boundary condition (vii) is satisfied, one has

$$\begin{aligned} (D_0(x) \pm i\varphi) \int_x^{\infty} \mu^{-1} U_0(\mp\varphi, s)\, K_i^j(x,s)\, ds &= \int \mu^{-1} U_0 O_s K_i^j\, ds, \\ (D_0(x) \mp i\varphi) \int_x^{\infty} \mu^{-1} U_0(\mp\varphi, s)\, K_i^j(x,s)\, ds &= \int \mu^{-1} U_0 O_d K_i^j\, ds \\ &\quad - 2U_0(\mp\varphi, x)\hat{K}_i^j. \end{aligned} \tag{47}$$

Applying the lemma, one gets

$$(\Delta - i\varphi\tilde{I})(\psi_1, \psi_2, \ldots, \psi_{n+1}) = \int_x^{\infty} \mu^{-1}(O_d K_{\mathrm{I}}^{T} + O_s K_{\mathrm{II}}^{T} + 2\hat{K}_{\mathrm{I}}^{T} K^{T})$$

$$\times \begin{pmatrix} U_0(-\varphi, s) & & & \\ & U_0(\varphi, s) & & 0 \\ & & \ddots & \\ 0 & & & U_0(\varphi, s) \end{pmatrix} = (0). \tag{48}$$

Application to a linear $(n+1)\times(n+1)$ system of first order differential equations: Let us consider

$$\left(I\mu(x)\frac{\partial}{\partial x} - i\varphi\tilde{I} - Q\right)\psi = 0, \quad Q = \begin{pmatrix} q_1^1 & \cdots & q_1^{n+1} \\ \vdots & & \vdots \\ q_{n+1}^1 & \cdots & q_{n+1}^{n+1} \end{pmatrix}. \tag{49}$$

Let us choose for the ψ solutions of Eq. (49) the following representation:

$$\psi_1 = (\delta_{1,i} U_0(-\varphi, x) + \int \mu^{-1} U_0(-\varphi, s) K_i^1(x, s)\, ds),$$
$$\psi_j = (\delta_{j,i} U_0(\varphi, x) + \int \mu^{-1} U_0(\varphi, s) K_i^j(x, s)\, ds), \tag{50}$$

If we substitute these solutions in the differential system (49) and apply Lemma IV, we get that the K_i^j must satisfy constraint relations:

$$(ID_0(x) - i\varphi\tilde{I} - Q)(\psi_1, \ldots, \psi_{n+1})$$
$$= -(Q_{\mathrm{I}} + Q_{\mathrm{II}} + 2\hat{K}_{\mathrm{I}}^T) \begin{pmatrix} U_0(-\varphi, x) & & 0 \\ & U_0(\varphi, x) & \\ 0 & & \ddots \\ & & & U_0(\varphi, x) \end{pmatrix}$$
$$+ \int_x^a \mu^{-1}(O_d K_{\mathrm{I}}^T + O_s K_{\mathrm{II}}^T - Q K^T)$$
$$\times \begin{pmatrix} U_0(-\varphi, s) & & 0 \\ & U_0(\varphi, s) & \\ 0 & & \ddots \\ & & & U_0(\varphi, s) \end{pmatrix} ds = (0) \tag{51}$$

If follows that we must have

$$Q_{\mathrm{II}} = 0, \quad Q_{\mathrm{I}} + 2\hat{K}_{\mathrm{I}}^T = 0,$$
$$O_d K_{\mathrm{I}}^T + O_s K_{\mathrm{II}}^T + 2\hat{K}_{\mathrm{I}}^T K^T = 0. \tag{52}$$

Consequently, if K is the solution of Eq. (33), these constraints are satisfied if $q_1^j = -2\hat{K}_1^j$, $q_j^1 = -2\hat{K}_j^1$ $q_1^1 = 0$, $q_j^i = 0$ for either i or $j > 1$.

The fact that, in order to solve the linear system (49), we have necessarily $Q_{\mathrm{II}} \equiv 0$ is due to the fact that the system (49) has twofold eigenvalues $\pm i\varphi$. In the folllowing paper,[6] where we will not be restricted to a two-fold-eigenvalue problem we shall see how this condition $Q_{\mathrm{II}} = 0$ is modified.

Now we discuss the existence and the assumed uniqueness (except for discrete values $x_1, x_2, \cdots$) of the solution of the I. E. (inversion equations). In this twofold eigenvalue case $\mathcal{J}$ is symmetric [for instance, if $\mu = 1$, $\mathcal{J} = \mathcal{J}(x+y)$ in Eq. (33)]. As in the Marchenko scalar case or as in the 2×2 case (see Refs. 1–3) it is easy to satisfy the boundary condition at $x = \infty$ in order to have Fredholm type equations for the I. E. We recall that *x is a parameter* for the I. E., the variable being y. For the scalar differential second order case we recall[6] that $K(x,x) = \mu^2(x)(\partial/\partial x)\log D(x)$, where D is the Fredholm determinant of the I. E. It follows that, for the x_i values $D(x_i) = 0$, the potentials have at most second order poles and the possible bound states become ghosts. We have not established such general connection here, but partial results in Appendix A and B. We show for the 2×2 case of Eq. (33) that $D_0(x)\hat{K}_1^1 = \frac{1}{2}D_0(x)\{[D_0(x)D]D^{-1}\} = -2\hat{K}_2^1\hat{K}_1^2$, the corresponding potentials (solutions of MKdV or Shrödinger cubic) have poles of the first order (see also the soliton ghosts written down in Sec. V). We show also that for $x > 0$ the first eigenvalue appears at $x = 0$ (similarly as it was the case in N/D equations. Furthermore, in the MKdV case, the coupling of the nonlinear term must be negative in order that D vanishes. In the following and *always when we derive the n.l.p.d.e. we compare solutions of Eq.* (33) *with free terms differing by constants* (x dependent but y independent). If the solution is such that $D = 0$ for discrete values x_i, we must exclude these values for which in fact the solution is infinite. Excluding these x_i values, nevertheless, the solutions of Eq. (33) satisfy both the matrix differential system and the associated n. l. p. d. e. (as an illustration see what happens for the soliton ghost of Sec. V). The important point is that we do not have to restrict our derivation to the cases where $D(x)$ does not vanish for $x \in [-\infty, +\infty]$.

Property III: Integral equations for $O_s O_d K$, $(O_s)^p K_{\mathrm{I}}$, $(O_d)^p K_{\mathrm{II}}$, p integer > 0.

(a) From Eq. (33) and applying the Lemma IV, we get

$$O_d K = \int \mu^{-1} \mathcal{J} O_s K\, ds,$$
$$O_s O_d K = -2\widehat{\mathcal{J} O_s K} + \int \mu^{-1} \mathcal{J} O_s O_d K\, ds. \tag{53}$$

From our assumption that the solution of the inversion equation is unique we deduce

$$O_s O_d K + 2K D_0 \hat{K} = 0. \tag{54}$$

Taking the second matrix component of Eq. (54) and putting $x = y$, one has

$$O_s\widehat{O_d K_{\mathrm{II}}} + 2(\hat{K}_{\mathrm{I}} D_0 \hat{K}_{\mathrm{I}} + \hat{K}_{\mathrm{II}} D_0 \hat{K}_{\mathrm{II}}) = 0. \tag{55}$$

(b) From Eq. (33) written in the two components formulation and applying the Lemma IV, we obtain the following integral equations:

$$\begin{pmatrix} (O_s)^p K_{\mathrm{I}} \\ (O_d)^p K_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} (O_s)^p \mathcal{J} - 2\ell_p \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} (O_s)^p K_{\mathrm{I}} \\ (O_d)^p K_{\mathrm{II}} \end{pmatrix}, \tag{56}$$

where

$$\ell_1 = \mathcal{J}\hat{K}_{\mathrm{II}}, \quad \ell_2 = O_s(\mathcal{J}\hat{K}_{\mathrm{II}}) + \widehat{\mathcal{J} O_d K_{\mathrm{II}}},$$
$$\ell_3 = O_s^2(\mathcal{J}\hat{K}_{\mathrm{II}}) + O_s(\widehat{\mathcal{J} O_d K_{\mathrm{II}}}) + \widehat{\mathcal{J}(O_d^2 K_{\mathrm{II}})}, \tag{57}$$
$$\ell_p = O_s^{p-1}(\mathcal{J}\hat{K}_{\mathrm{II}}) + O_s^{p-2}(\widehat{\mathcal{J} O_d K_{\mathrm{II}}}) + \cdots + \widehat{\mathcal{J} O_d^{p-1} K_{\mathrm{II}}}.$$

Property IV: Particular properties of the solutions K_{I} and K_{II} when the scalar kernels G_i and F_i building $\mathcal{S}$ are linked.

First, due to the fact that the scalar kernels G_i and F_i are symmetric $G_i(x,y) = G_i(y,x)$, $F_i(x,y) = F_i(y,x)$, it follows that the matrix kernel $\mathcal{J}$ itself is symmetric with the meaning $\mathcal{J}(x,y) = \mathcal{J}(y,x)$. Secondly, if we iterate the solution of Eq. (33) and apply our rule for the multiplication of matrices of types I and II, we see that K_{I} has only an odd number of $\mathcal{J} \equiv \mathcal{J}_{\mathrm{I}}$ whereas K_{II} has an even number. We get for $x = y$

$$\hat{K}_{\mathrm{I}} = \hat{\mathcal{J}} + \sum_{p=1}^{\infty} \int_x^a \cdots \int_x^a \mu^{-1}(s_1) \cdots \mu^{-1}(s_{2p}) \times \prod_1^{2p} ds_i\, \mathcal{J}(s_1, x)\, \mathcal{J}(s_1, s_2) \cdots \mathcal{J}(x, s_{2p}),$$
$$\hat{K}_{\mathrm{II}} = \sum_{p=0}^{\infty} \int_x^a \cdots \int_x^a \mu^{-1}(s_1) \cdots \mu^{-1}(s_{2p+1})\, \mathcal{J}(s_1, x) \times \mathcal{J}(s_1, s_2) \cdots \mathcal{J}(x, s_{2p+1}) \prod_1^{2p+1} ds_i. \tag{58}$$

Let us define the Hermitian matrix $\mathcal{M}^* = (\mathcal{M}^T)^*$, and we get

$$\hat{K}_{\rm I}^{\dagger}=\hat{\mathcal{J}}^{\dagger}+\sum\int\cdots\int\mu^{-1}\cdots\mu^{-1}\prod ds_i\,\mathcal{J}^{\dagger}(x,s_{2p})\cdots\mathcal{J}^{\dagger}(s_1,x),$$
$$\hat{K}_{\rm II}^{\dagger}=\sum\int\cdots\int\mu^{-1}\cdots\mu^{-1}\prod ds_i\,\mathcal{J}^{\dagger}(x,s_{2p+1})\cdots\mathcal{J}^{\dagger}(s_1,x). \tag{59}$$

Property IVa: Let us assume that $G_i=\kappa F_i^*$, κ being a real constant, and define

$$(\chi)=\begin{pmatrix}\chi^{-1} & & 0\\ & \chi & \\ & & \ddots\\ 0 & & \chi\end{pmatrix},$$

it follows that the solutions of the inversion equation (33) satisfy

$$\hat{K}_{\rm I}^{\dagger}(\chi)=-\hat{K}_{\rm I},\quad \hat{K}_{\rm II}^{\dagger}=\hat{K}_{\rm II}, \tag{60a}$$

or equivalently

$$(\hat{K}_j^1)^*=-\chi\hat{K}_1^j,\quad \hat{K}_j^j \text{ real},\quad (\hat{K}_j^i)^*=\hat{K}_i^j \text{ for } i \text{ and } j>1,$$
$$D_0\hat{K}_1^1=2\chi^{-1}\sum_j|\hat{K}_j^1|^2,\quad D_0\hat{K}_j^j=2\chi^{-1}|\hat{K}_j^1|^2 \text{ for } j>1. \tag{60b}$$

For the proof, we first define

$$(\chi^{-1})=\begin{pmatrix}\chi & & 0\\ & \chi^{-1} & \\ & & \ddots\\ 0 & & \chi^{-1}\end{pmatrix}$$

and get

$$-(\chi^{-1})\mathcal{J}^{\dagger}=-\mathcal{J}^{\dagger}(\chi)=\mathcal{J},\quad (\chi)(\chi^{-1})=I=(\chi^{-1})(\chi) \tag{61a}$$

We multiply the first Eq. (59) on the right by (χ) and get

$$\hat{K}_{\rm I}^{\dagger}(\chi)=\hat{\mathcal{J}}^{\dagger}(\chi)+\sum\int\int\mu^{-1}\mu^{-1}\prod\mathcal{J}^{\dagger}(\chi)(\chi^{-1})\mathcal{J}^{\dagger}\mathcal{J}^{\dagger}(\chi)(\chi^{-1})\cdots(\chi)(\chi^{-1})\mathcal{J}^{\dagger}\mathcal{J}^{\dagger}(\chi),$$
$$\hat{K}_{\rm II}^{\dagger}=\sum\int\int\mu^{-1}\cdots\mu^{-1}\prod\mathcal{J}^{\dagger}(\chi)(\chi^{-1})\mathcal{J}^{\dagger}\mathcal{J}^{\dagger}\cdots\mathcal{J}^{\dagger}(\chi)(\chi^{-1})\mathcal{J}^{\dagger}. \tag{62}$$

If we substitute the relation (61a) into Eq. (62), then the result given by Eq. (60a) follows:

Property IVb: If $G_i=\chi F_i$, χ being a constant, it follows that

$$\hat{K}_{\rm I}^{T}(\chi)=-\hat{K}_{\rm I},\quad \hat{K}_{\rm II}^{T}=\hat{K}_{\rm II},\quad \hat{K}_j^1=-\chi\hat{K}_1^j,\hat{K}_j^i \text{ for } i \text{ and } j>1,$$
$$D_0\hat{K}_1^1=2\chi^{-1}\sum(\hat{K}_j^1)^2,\quad D_0\hat{K}_j^j=2\chi^{-1}(\hat{K}_j^1)^2.$$

For the proof let us remark that in this case:

$$-(\chi^{-1})\mathcal{J}^{T}=-\mathcal{J}^{T}(\chi)=\mathcal{J}.$$

The proof is the same as above if we substitute the transposed matrices instead of the Hermitian ones.

III. GENERALIZATION OF THE ABLOWITZ *ET AL.* METHOD WHEN ∂/∂_x IS REPLACED BY $\mu(x)\,\partial/\partial x$

In the remaining of the paper we do not consider the matrix second order differential system studied in Sec. II A. We are always concerned with the matrix formulation of the first order linear differential system $\mu(x)\,\partial/\partial x$ with twofold eigenvalues for which we have derived the inversion equations in Sec. II B.

In this method one assumes that there exist two linear first order differential systems with the same solution. One is related to the coordinate variable (as in the inversion equation) and the other with the time variable. The compatibility of both systems is obtained by cross differentiation and leads to constraints for the coefficients of the systems which thus must satisfy n. l. p. d. e. The $[2\times2]$ matrix case has been widely studied and Ablowitz and Haberman[4] have recently extended their method to the higher order matrix case. There is an ambiguity because if we consider the solution of the inversion equation studied in Sec. II B, where only the coordinate variable appears, it can happen that, in order to satisfy both linear systems, this solution must be multiplied by some t (time) dependent function. Let us consider $D_0=\mu(x)\,\partial/\partial x$, $\Delta_0=D_0 I$, and

$$\Delta_0\psi=X\psi,\quad \psi_t=Y\psi,$$
$$X=\begin{pmatrix}-i\varphi, q_1,\dots,q_n\\ r_1, i\varphi \quad 0\\ \vdots\quad\ddots\\ r_n \quad 0 \quad i\varphi\end{pmatrix},\quad Y=\begin{pmatrix}A & B_1\cdots B_n\\ C_1 & D_{11} \quad D_{1n}\\ \vdots & \vdots\quad\vdots\\ C_n & D_{n1} \quad D_{nn}\end{pmatrix}. \tag{63}$$

The fact that $X-i\varphi\tilde{I}$ is of the matrix type I is due to the fact that we consider in this paper a twofold eigenvalue problem (see Sec. II B. We shall leave this restriction in the following paper.[5] Cross differentiation of both systems in Eq. (63) leads to

$$D_0(x)\psi_t=\frac{\partial}{\partial t}D_0(x)\psi,\quad [X,Y]=D_0Y-X_t, \tag{64}$$

or

$$A+\sum_j D_{jj}=0,\quad D_0(D_{ij})=r_iB_j-q_jC_i,$$
$$q_{i,t}=D_0B_i+2i\varphi B_i+q_i(A-D_{ii})-\sum_{j\neq i}q_jD_{j,i}, \tag{64'}$$
$$r_{i,t}=D_0C_i-2i\varphi C_i-r_i(A-D_{ii})+\sum_{j\neq i}r_jD_{i,j}.$$

Application to the Schrödinger cubic coupled case

It can be verified that the following set verifies the first line of relations given in Eq. (64′)

$$A=-\frac{4n}{n+1}i\varphi^2-i\sum r_iq_i-i\sum\lambda_{jj},$$
$$D_{jj}=\frac{4i\varphi^2}{n+1}+ir_jq_j+i\lambda_{jj}, \tag{65}$$
$$B_j=iD_0q_j+2\varphi q_j,\quad C_j=-iD_0r_j+2\varphi r_j,$$
$$D_{ij}=i\lambda_{ij}+ir_iq_j,\quad \lambda_{ij}=\text{const}.$$

It follows that we have the n. l. p. d. e.

$$iq_{i,t}+D_0^2q_i-2q_i\sum q_jr_j-\sum q_j\bar{\lambda}_{j,i}=0,\quad \bar{\lambda}_{ij}=\text{const},$$
$$ir_{i,t}-D_0^2r_i+2r_i\sum r_jq_j-\sum r_j\bar{\lambda}_{ij}=0. \tag{66}$$

IV. N.L.P.D.E FOR THE SOLUTIONS OF THE INVERSION EQUATION WHEN THE KERNELS SATISFY L.P.D.E.

In this section we still consider the linear system with two eigenvalues and $D_0=\mu(x)\,\partial/\partial x$ for which the inversion equations have been established in Sec. II B and written down in Eq. (33)

$$\begin{pmatrix} K_{\rm I} \\ K_{\rm II} \end{pmatrix} = \begin{pmatrix} \mathcal{J} \\ 0 \end{pmatrix} + \int_x^\infty \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} K_{\rm I} \\ K_{\rm II} \end{pmatrix} ds, \quad \mathcal{J} \equiv \mathcal{J}_{\rm I}. \tag{33}$$

We assume that the kernel $\mathcal{J}$ (and consequently the solution K) depends on a parameter t: $\mathcal{J} = \mathcal{J}(x, y; t)$. As we shall see, if $\mathcal{J}$ satisfies an l. p. d. e., then $K_{\rm I}$ satisfies an n. l. p. d. e. with respect to x, y, t whose linear part is identical to that of $\mathcal{J}$. We first obtain the equations for x not necessarily equal to y and at the end put $x = y$. We do not try to investigate all possibilities and focuse our attention for two cases: generalizations of both Schrödinger cubic and KdV equations.

(i) We define for integer p values the operator

$$O_{\alpha,p} = I\alpha \frac{\partial}{\partial t} - (O_s)^p \tilde{I}_p, \quad \tilde{I}_p = \begin{pmatrix} (-1)^{p+1} & & \\ & 1 \cdot \cdot & 0 \\ & 0 & \cdot \cdot 1 \end{pmatrix},$$

$$O_s = D_0(x) + D_0(y). \tag{67}$$

We get $I = \tilde{I}_p \tilde{I}_p$, $\tilde{I}_{2p}/M$ changes the sign of the first line of M, $M\tilde{I}_{2p}$ changes the sign of the first column, $\tilde{I}_p M_{\rm I} \tilde{I}_p = (-1)^{p+1} M_{\rm I}$, for odd p values $\tilde{I}_p = I$ the identity matrix. As an application we see that Eq. (56) can be written

$$\begin{pmatrix} -\tilde{I}_p (O_s)^p K_{\rm I} \\ (-1)^p \tilde{I}_p (O_d)^p K_{\rm II} \end{pmatrix} = \begin{pmatrix} -O_s^p \tilde{I}_p \mathcal{J} + 2\tilde{I}_p \ell_p \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} -\tilde{I}_p O_s^p K_{\rm I} \\ (-1)^p \tilde{I}_p O_d^p K_{\rm II} \end{pmatrix}, \tag{68}$$

where ℓ_p is written down in Eq. (57).

(ii) We take the derivative with respect to t of Eq. (33):

$$\begin{pmatrix} \alpha K_{{\rm I},t} \\ \alpha K_{{\rm II},t} \end{pmatrix} = \begin{pmatrix} \alpha \mathcal{J}_t \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} \alpha K_{{\rm I},t} \\ \alpha K_{{\rm II},t} \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \alpha \mathcal{J}_t \\ \alpha \mathcal{J}_t & 0 \end{pmatrix} \begin{pmatrix} K_{\rm I} \\ K_{\rm II} \end{pmatrix}. \tag{69}$$

We define the operator $\bar{O}_{\alpha,p} = I\alpha\partial/\partial t + (-1)^p \tilde{I}_p O_d^p$ and add both Eqs. (68) and (69):

$$\begin{pmatrix} O_{\alpha,p} K_{\rm I} \\ \bar{O}_{\alpha,p} K_{\rm II} \end{pmatrix} = \begin{pmatrix} O_{\alpha,p} \mathcal{J} + 2\tilde{I}_p \ell_p \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \times \begin{pmatrix} O_{\alpha,p} K_{\rm I} \\ \bar{O}_{\alpha,p} K_{\rm II} \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \alpha \mathcal{J}_t \\ \alpha \mathcal{J}_t & 0 \end{pmatrix} \begin{pmatrix} K_{\rm I} \\ K_{\rm II} \end{pmatrix}. \tag{70}$$

This is all that can be done without any specification of the relation between the (x, y) dependence and the t dependence of $\mathcal{J}$. As we shall see below, we assume an l. p. d. e. for $\mathcal{J}$ of the kind $O_{\alpha,p}\mathcal{J} = 0$, we shall obtain, for some functional of K, an integral equation with the same kernel as that of the inversion equation (33) itself.

Theorem: Let us assume that

$$O_{\alpha,p}\mathcal{J} = 0, \tag{71}$$

$$\lim_{s \to a} [(D_0^{p-1}(s)\mathcal{J}(s,y)) K_{\rm II}(x,s) - (D_0^{p-2}\mathcal{J}) D_0(s) K_{\rm II}(x,s) \cdots + (-1)^{p-1} \mathcal{J}(D_0^{p-1}(s) K_{\rm II}(x,s))] = 0 \tag{72}$$

and define $\tilde{O}_{\alpha,p} = \alpha\partial/\partial t + (-1)^p \tilde{I}_p (O_d)^p - 2^p D_0^p(y) \tilde{I}_p$. Then we get

$$\begin{pmatrix} O_{\alpha,p} K_{\rm I} \\ \tilde{O}_{\alpha,p} K_{\rm II} \end{pmatrix} = \begin{pmatrix} \tilde{I}_p (2\ell_p + 2^p \beta_p) \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \times \begin{pmatrix} O_{\alpha,p} K_{\rm I} \\ \tilde{O}_{\alpha,p} K_{\rm II} \end{pmatrix}, \tag{73}$$

$$\beta_1 = -\mathcal{J}\hat{K}_{\rm II} \quad \beta_2 = -(D_0(x)\mathcal{J})\hat{K}_{\rm II} + \mathcal{J}\widehat{(D_0(s) K_{\rm II}(x,s))},$$
$$\beta_3 = -(D_0^2\mathcal{J})\hat{K}_{\rm II} + (D_0\mathcal{J})\widehat{(D_0 K_{\rm II})} - \mathcal{J}\widehat{D_0^2 K_{\rm II}},$$
$$\beta_p = -(D_0^{p-1}(x)\mathcal{J}(x,y))\hat{K}_{\rm II} + (D_0^{p-2}\mathcal{J})\widehat{D_0 K_{\rm II}} + \cdots + (-1)^p \mathcal{J}\widehat{D_0^{p-1} K_{\rm II}}. \tag{74}$$

For the proof we shall use the following lemma:

Lemma V: Let H, N be scalar functions such that

$$\lim_{s \to \infty} \{H(x,s) D_0^{p-1}(s) N(y,s) - (D_0^{p-2}(s) N(y,s))(D_0(s) H) + \cdots + (-1)^p N(D_0^{p-1}(s) H)\} = 0;$$

then one has

$$\int_x^a \mu^{-1}(s) H(x,s) D_0^p(s) N(y,s)\, ds = (-1)^p \int_x^a \mu^{-1} N(D_0^p(s) H(x,s))\, ds + B_p,$$
$$B_1 = -N\hat{H}, \quad B_2 = -(D_0 N)\hat{H} + \widehat{N D_0 H},$$
$$B_p = -(D_0^{p-1}N)\hat{H} + (D_0^{p-2}N)\widehat{D_0 H} + \cdots + (-1)^p \widehat{N D_0^{p-1} H}.$$

For the proof of the theorem we first remark that due to $O_{\alpha,p}\mathcal{J} = O_d\mathcal{J} = 0$, we have $\alpha \mathcal{J}_t = O_s^p \tilde{I}_p \mathcal{J} = 2^p D_0^p \tilde{I}_p \mathcal{J}$. Consequently, we have $\alpha \int \mu^{-1} \mathcal{J}_t K_{\rm I}\, ds = \tilde{I}_p 2^p D_0^p(y) \times \int_x^\infty \mu^{-1}(s) \mathcal{J}(s,y) K_{\rm I}(x,s)\, ds = \tilde{I}_p 2^p D_0^p(y) K_{\rm II}$, and thus the second line of Eq. (70) can be written $\tilde{O}_{\alpha,p} K_{\rm II}$ $\int \mu^{-1} \mathcal{J} O_{\alpha,p} K_{\rm I}\, ds$. Applying Lemma V, we get $\int \mu^{-1} (D_0^p \mathcal{J}) K_{\rm II}\, ds = (-1)^p \int \mu^{-1} \mathcal{J}(D_0^p K_{\rm II}) + \beta_p$. Consequently, we have $\alpha \int \mu^{-1} \mathcal{J} K_{\rm II}\, ds = 2^p \tilde{I}_p \beta_p + \int \mu^{-1} \mathcal{J} \times (-2^p D_0^p \tilde{I}_p K_{\rm II})\, ds$ and finally the first line of Eq. (70) can be written

$$O_{\alpha,p} K_{\rm I} - (\tilde{I}_p (2\ell_p + 2^p \beta_p)) = \int \mu^{-1} \mathcal{J}(\bar{O}_{\alpha,p} - 2^p D_0^p \tilde{I}_p) K_{\rm II}\, ds = \int \mu^{-1} \mathcal{J} \tilde{O}_{\alpha,p} K_{\rm II}\, ds.$$

In order to solve the equation integral (72) we must know the free term and thus calculate $2_{\ p} + 2^p \beta_p$. We quote the three first values:

$$p = 1: \quad \ell_1 + \beta_1 = 0,$$
$$p = 2: \quad 2(\ell_2 + 2\beta_2) = 4\mathcal{J}(D_0(x)\hat{K}_{\rm II}) = -8\mathcal{J}\hat{K}_{\rm I}\hat{K}_{\rm I}; \tag{75}$$
$$p = 3: \quad 2(\ell_3 + 4\beta_3) = 6[\mathcal{J}\widehat{(O_s O_d K_{\rm II})} + 2(D_0\mathcal{J})(D_0\hat{K}_{\rm II})].$$

For $p = 1$ the free term in Eq. (73) is zero; it follows, from the assumed uniqueness of the solution of Eq. (33), that $O_{\alpha,1} K_{\rm I} = 0$, $\tilde{O}_{\alpha,1} K_{\rm II} = 0$. Thus in this case both the kernel $\mathcal{J}$ and the solution $K_{\rm I}$ satisfy the same l. p. d. e. without nonlinear term for $K_{\rm I}$. For $p = 2$ we see that both free terms of Eq. (73) and (33) are proportional to a term $(\hat{K}_{\rm I})^2$ and thus this case will lead to a cubic

nonlinear term. For $p=3$ we have in Eq. (74) one term proportional to $\mathcal{J}$ leading, roughly speaking, to $K\hat{K}D_0\hat{K}$ [see Eq. (55)] and another one proportional to $D_0\mathcal{J}$ which is not of the same type as the free term of Eq. (33). For higher p values, terms of the type $D_0^q\mathcal{J}$ appear, and we see that if we except the $p=1$ case, the solutions of Eq. (72) will have nonlinear terms. Let us consider a sum of linear operators

$$\sum_{i=1}^{i_{\max}} \gamma_i O_{\alpha_i,p_i}, \quad \sum_{i=1}^{i_{\max}} \gamma_i \bar{O}_{\alpha_i,p_i},$$

γ_i being constants and $p_i=i$ positive integers, and go back to Eq. (70), where the corresponding integral equation is

$$\begin{pmatrix} \sum \gamma_i O_{\alpha_i,p_i} K_{\mathrm{I}} \\ \sum \gamma_i \bar{O}_{\alpha_i,p_i} K_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} \sum \gamma_i O_{\alpha_i,p_i}\mathcal{J} + 2\sum \gamma_i \tilde{I}_{p_i} \ell_{p_i} \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} \sum \gamma_i O_{\alpha_i,p_i} K_{\mathrm{I}} \\ \sum \gamma_i \bar{O}_{\alpha_i,p_i} K_{\mathrm{II}} \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0, \sum \gamma_i \alpha_i \mathcal{J}_t \\ \sum \gamma_i \alpha_i \mathcal{J}_t, 0 \end{pmatrix} \begin{pmatrix} K_{\mathrm{I}} \\ K_{\mathrm{II}} \end{pmatrix}.$$

If we assume $\sum_i \gamma_i O_{\alpha_i,p_i}\mathcal{J}=0$, use the relation $\sum \alpha_i\gamma_i \mathcal{J}_t = \sum \gamma_i (O_s)^{p_i} \tilde{I}_{p_i}$, and follow the same lines of proof as in the theorem, we get easily a generalization of it:

Corollary of the theorem: Let us assume

$$\text{(i)} \quad \sum_{i=1}^{i_{\max}} \gamma_i O_{\alpha_i,p_i}\mathcal{J}=0 \tag{71'}$$

(ii) for each integer value p_i, the same boundary condition as in Eq. (72), then we get

$$\begin{pmatrix} \sum \gamma_i O_{\alpha_i p_i} K_{\mathrm{I}} \\ \sum \gamma_i \tilde{O}_{\alpha_i p_i} K_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} \sum \gamma_i \tilde{I}_{p_i} (2\ell_{p_i} + 2^{p_i}\beta_{p_i}) \\ 0 \end{pmatrix} + \int \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} \sum \gamma_i O_{\alpha_i p_i} K_{\mathrm{I}} \\ \sum \gamma_i \tilde{O}_{\alpha_i p_i} K_{\mathrm{II}} \end{pmatrix}, \tag{73'}$$

where for each p_i the ℓ_{p_i} and β_{p_i} have been defined in Eq. (57) and Eq. (75). For instance, if $i_{\max}=3$, we get

$$\sum_1^3 \gamma_i \tilde{I}_{p_i}(2\ell_{p_i} + 2^{p_i}\beta_{p_i}) = 4\gamma_2 \tilde{I}\mathcal{J}D_0\hat{K}_{\mathrm{II}} + 6\gamma_3[\mathcal{J}\widehat{O_s O_d K}_{\mathrm{II}} + 2(D_i(x)\mathcal{J})(D_0\hat{K}_{\mathrm{II}})]. \tag{76}$$

n.l.p.d.e. corresponding to a linear combination of the p_i = 1, 2, 3 cases

We assume

$$\sum_{i=1}^{3} \gamma_i O_{\alpha_i,i}\mathcal{J}=0, \quad O_d\mathcal{J}=0 \tag{77}$$

and apply the results of the corollary. (The study is done in Appendix A, subsection A1.) We get

$$\sum_1^3 \gamma_i O_{\alpha_i,i} K_{\mathrm{I}} + 8\gamma_2 \tilde{I} K_{\mathrm{I}}\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}} + 12\gamma_3(K_{\mathrm{I}}\hat{K}_{\mathrm{I}}D_0\hat{K}_{\mathrm{I}} + (O_s K_{\mathrm{I}})\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}}) = 0,$$

$$\sum_1^3 \gamma_i \widehat{O_{\alpha_i,i}} K_{\mathrm{I}} + 8\gamma_2 \tilde{I}\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}} + 12\gamma_3(\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}}D_0\hat{K}_{\mathrm{I}} + (D_0\hat{K}_{\mathrm{I}})\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}}) = 0. \tag{78}$$

As usual $i=1$ does not give any nonlinear term, if $\gamma_1=\gamma_3=0$, we recognize the generalized Schrödinger cubic equation whereas if $\gamma_1=\gamma_2=0$ we have the generalized KdV equation Introducing the scalar functions $\hat{K}_j^1$ and $\hat{K}_1^j$, Eq. (78) becomes for $x=y$

$$\sum_1^3 \gamma_i \alpha_i \frac{\partial}{\partial t} - \gamma_1 D_0 + \gamma_2 D_0^2 - \gamma_3 D_0^3)\hat{K}_j^1 - 8\gamma_2 \hat{K}_j^1 \sum_2^{n+1} \hat{K}_1^1 \hat{K}_1^1 + 12\gamma_3 \sum_2^{n+1} (\hat{K}_j^1 \hat{K}_1^1 D_0 \hat{K}_1^1 + \hat{K}_1^1 \hat{K}_1^1 D_0 \hat{K}_j^1),$$

$$\left(\sum_1^3 \gamma_i \alpha_i \frac{\partial}{\partial t} - \gamma_1 D_0 - \gamma_2 D_0^2 - \gamma_3 D_0^3\right)\hat{K}_1^j + 8\gamma_2 \hat{K}_1^1 \sum_2^{n+1} \hat{K}_1^1 \hat{K}_1^j + 12\gamma_3 \sum_2^{n+1} (\hat{K}_1^1 \hat{K}_1^j D_0 \hat{K}_1^1 + \hat{K}_1^1 \hat{K}_1^1 D_0 \hat{K}_1^j) = 0. \tag{78'}$$

Application to the generalized Schrödinger cubic equation

We put in Eqs. (78) and (78') $\gamma_1=\gamma_3=0$, $\gamma_2=1$, and $\alpha_2=\alpha$. If

$$O_{\alpha,p=2}\mathcal{J}=0, \tag{79a}$$

it follows

$$\tilde{O}_{\alpha,2} K_{\mathrm{II}} + 4\tilde{I} K_{\mathrm{II}} D_0 \hat{K}_{\mathrm{II}} = 0,$$

$$\left[\alpha \frac{\partial}{\partial t} - \left(\mu(x)\frac{\partial}{\partial x} + \mu(y)\frac{\partial}{\partial y}\right)^2\right] K_{\mathrm{I}}(x,y;t) + 8\tilde{I} K_{\mathrm{I}}(x,y;t) K_{\mathrm{I}}(x,x;t) K_{\mathrm{I}}(x,x;t) = 0, \tag{79b}$$

$$\left[\alpha \frac{\partial}{\partial t} + O_s^2\right] K_i^1 - 8\hat{K}_i^1 \sum_2^{n+1} K_j^1 \hat{K}_1^1 = 0,$$

$$\left[\alpha \frac{\partial}{\partial t} - O_s^2\right] K_1^j + 8\hat{K}_1^j \sum_2^{n+1} K_j^1 \hat{K}_1^1 = 0. \tag{80}$$

At this stage we still have x not necessarily equal to y. If we put $x=y$, we have the relation (80) where the quantities K_i^1, K_j^1, $K^j{}_1$ must be replaced by $\hat{K}_i^1$, $\hat{K}_j^1$, $\hat{K}_1^j$. Let us further assume $\alpha=i$ and $G_i=\kappa F_i^*$. We know from Sec. II that $\hat{K}^{*1}_j = -\kappa \hat{K}_1^j$, $\hat{K}_1^1 = 2\kappa^{-1}\sum |\hat{K}_j^1|^2$, $\hat{K}_j^j = 2\kappa^{-1}|\hat{K}_j^1|^2$ and Eq. (80) can be written

$$\left(i\frac{\partial}{\partial t} + D_0^2(x)\right)\hat{K}_j^1 + 8\kappa^{-1}\hat{K}_j^1 \sum_{i=2}^{n+1} |\hat{K}_i^1|^2 = 0. \tag{81}$$

We recall that $\mathcal{J}$ must satisfy

$$O_d F_i = O_d G_i = \left(i\frac{\partial}{\partial t} - O_s^2\right) F_i = \left(i\frac{\partial}{\partial t} + O_s^2\right) G_i = 0, \quad G_i = \chi F_i^*, \tag{82}$$

as well as well-defined boundary conditions (see Sec. II B and the theorem of the present section).

For instance, if $\mathcal{J}$ is built with discrete terms, we take

$$F_i = \sum_m \alpha_{m,i} \exp[4i\nu_m^2 t + i\nu_m(\int_c^x \mu^{-1}(u)\,du + \int_c^y \mu^{-1}(u)\,du)],$$

$$G_i = \chi F_i^*, \tag{83}$$

where we assume $\mathrm{Im}\nu_m > 0$, $\mu(x) > 0$ for simplicity and $\lim_{x\to-\infty} \int_c^x \mu^{-1}(u)\,du = +\infty$. Of course, $\mathcal{J}$ can include a continuum part. More generally, F_i and G_i can be written

$$F_i = \int d\nu\, \alpha_i(\nu) \exp[i\nu[\int^x \mu^{-1}(u)\,du + \int^y \mu^{-1}(u)\,du] + 4i\nu^2 t], \quad G_i = \chi F_i^*.$$

Application to the generalized modified KdV equation

We put in Eqs. (78) and (78′) $\gamma_1 = \gamma_2 = 0$, $\gamma_3 = 1$, $\alpha_3 = \alpha$. If

$$O_{\alpha,\rho=3}\mathcal{J} = 0, \tag{84a}$$

it follows

$$\begin{aligned} &O_{\alpha,3}K_{\mathrm{I}} + 12[(O_s K_{\mathrm{I}})\hat{K}_{\mathrm{I}}\hat{K}_{\mathrm{I}} + K_{\mathrm{I}}\hat{K}_{\mathrm{I}}D_0\hat{K}_{\mathrm{I}}] = 0,\\ &\tilde{O}_{\alpha,3}K_{\mathrm{II}} + 6[2K_{\mathrm{II}}\hat{K}_{\mathrm{I}}D_0\hat{K}_{\mathrm{I}} - O_d K_{\mathrm{II}} D_0\hat{K}_{\mathrm{II}}] = 0. \end{aligned} \tag{84}$$

Introducing the scalar quantities $\hat{K}_j^1$, $\hat{K}_1^j$, we get

$$\begin{aligned} &\left[\alpha\frac{\partial}{\partial t} - \left(\mu\frac{\partial}{\partial x}\right)^3\right]\hat{K}_1^j + 12\left[\sum_{j=2}^{n+1}\hat{K}_1^1\hat{K}_j^1(D_0\hat{K}_1^j) + \hat{K}_1^j\hat{K}_j^1 D_0\hat{K}_1^1\right] = 0,\\ &\left[\alpha\frac{\partial}{\partial t} - \left(\mu\frac{\partial}{\partial x}\right)^3\right]\hat{K}_i^1 + 12\left[\sum_{j=2}^{n+1}\hat{K}_i^1(D_0\hat{K}_j^1)\hat{K}_1^j + \hat{K}_1^j\hat{K}_j^1 D_0\hat{K}_j^1\right] = 0, \end{aligned} \tag{85}$$

which is a KdV coupled ($\partial/\partial x$ replaced by $\mu(x)\,\partial/\partial x$) equation. Let us further assume $\alpha = 1$, $G_i = \kappa F_i$. We know from Sec. II that $\hat{K}_j^1 = -\kappa\hat{K}_1^1$, $D_0\hat{K}_1^1$, $D_0\hat{K}_1^1 = 2\kappa^{-1}\sum_j(\hat{K}_j^1)^2$ $D_0\hat{K}_j^1 = 2\kappa^{-1}(\hat{K}_j^1)^2$ and Eq. (85) can be written

$$\left(\frac{\partial}{\partial t} + D_0^3(x)\right)\hat{K}_j^1 + 12\chi^{-1}\left(\sum_{l=2}^{n+1}\hat{K}_j^1\hat{K}_l^1(D_0\hat{K}_l^1) + (\hat{K}_1^1)^2 D_0\hat{K}_j^1\right) = 0. \tag{86}$$

In order to explicity construct the kernel $\mathcal{J}$ we have to satisfy

$$O_d F_i = O_d G_i = \left(\frac{\partial}{\partial t} + O_s^3\right)F_i = \left(\frac{\partial}{\partial t} + O_s^3\right)G_i = 0, \quad G_i = \chi F_i.$$

For instance, the discrete part of $\mathcal{J}$ can be written $\nu_m > 0$, $F_i = \sum_m \alpha_{m,i}\exp[-\nu_m[\int^x \mu^{-1}\,du + \int^y \mu^{-1}\,du] + 8\nu_m^3 t]$; and the continuum part,

$$F_i = \int d\nu\alpha_i(\nu)\exp i\nu[\int^x \mu^{-1}\,du + \int^y \mu^{-1}\,du + 8\nu^2 t].$$

V. APPLICATIONS

A. Properties of the inversion Eq. (33)

1. Determination of the elements of the matrix solution

For the solution K we recall that $\hat{K}_j^1$ and $\hat{K}_1^j$ are the scalar solutions of the n. l. p. d. e. $D_0\hat{K}_j^j$ for $j > 1$ represents essentially their products and $D_0\hat{K}_1^1$ a sum of such products [see Eq. (41a)]. (On the contrary the other elements K_j^i, $i \neq j$, $i \neq 1$, $j \neq 1$ are less interesting because they are not directly connected with the n. l. p. d. e.). We get the following integral equations (see also Appendix A2):

$$K_j^1(x,y) = -G_{j-1}(x,y) - \int_x^\infty \mu^{-1}(u) \times[\int_x^\infty \mu^{-1}(s)\sum_1^n G_m(s,y)F_m(u,s)\,ds]K_j^1(x,u)\,du, \quad j > 1, \tag{87a}$$

$$K_1^1(x,y) = -\sum_{m=1}^n[\int_x^\infty \mu^{-1}G_m(s,y)F_m(x,s)\,ds - \int_x^\infty \mu^{-1}(s)\int_x^\infty \mu^{-1}(u)\,G_m(s,y)F_m(u,s) \times K_1^1(x,u)\,du\,ds]. \tag{87b}$$

2. K_1^1 for $n = 1$

We consider the Ablowitz *et al.* case 2×2. In Eq. (87) the sum has only one term $m = n = 1$. Let us define

$$\mu^{-1}(y)K_1^1(x,y) = \phi^x(y),$$
$$H^x(y,s) = \int_x^\infty \mu^{-1}(u)\,\mu^{-1}(y)\,G(u,y)\,F(s,u)\,du,$$

the integral equation corresponding to (87b) can be written (see Appendix A3)

$$\phi^x(y) = -H^x(y,x) - \int_x^\infty H^x(y,s)\,\phi^x(s)\,ds, \quad \phi^x(x) = \frac{1}{2}\frac{\partial}{\partial x}\log\Delta. \tag{88}$$

This equation has a structure similar to the classical Gel'fand—Levitan or Marchenko equations for which we know that the solution for $x = y$ is $(\partial/\partial x)\log\Delta$, Δ being the Fredholm determinant of Eq. (87b):

$$\hat{K}_1^1 = \tfrac{1}{2}(D_0(x)\Delta/\Delta), \quad D_0(x)\hat{K}_1^1 = \tfrac{1}{2}D_0(x)(D_0(x)\Delta/\Delta) = -2\hat{K}_1^2\hat{K}_2^1.$$

In Appendix A4 we extend, for the KdV and Schrödinger cubic cases, a previous property[7] established in N/D and inversion equations. If we consider $x \geq 0$, then the smallest eigenvalue of Eq. (88) (and consequently the first ghost) appears when $x \to 0$. Unfortunately, we have no such general result for $x < 0$ but we do have at least a partial result.

3. $\hat{K}_j^j$: case of degenerate kernels $\mathcal{J}$

Let us assume

$$F_m = \sum_{i=1}^{i_{\max}} f_i(x)f_i(y)\,\alpha_m^i, \quad G_m = \sum_{i=1}^{i_{\max}} g_i(x)g_i(y)\,\beta_m^i, \tag{89}$$

the kernel of (87) is degenerated and can be written

$$\sum_{i,k}^{i_{\max},k_{\max}} f_i(y)g_k(u)h_{k,i}\gamma_{ik},$$
$$h_{k,i} = \int \mu^{-1}g_k f_i\,ds, \quad \gamma_{ik} = \sum_m \alpha_m^i\beta_m^k.$$

In this case, Eq. (87a) is reduced to a system of linear algebraic equations and the spectrum associated with the kernel of (87) has a finite number of eigenvalues.

(3a) Case of $i_{\max} = 1$. In this case we call $f_1 = f$, $g_1 = g$, $\alpha_m^1 = \alpha_m$, $\beta_m^1 = \beta_m$, $\gamma_{ij} = \gamma$, $h_{ij} = h$. Starting from $F_m = f(x)f(y)\,\alpha_m\cdots$, we get

$$K_j^1(x,y) = -g(x)g(y)\beta_{j-1}\Delta^{-1}, \quad \hat{K}_j^1 = -g^2\beta_{j-1}\Delta^{-1},$$

$$\hat{K}_j^j = -\alpha_{j-1}\beta_{j-1}hgf\Delta^{-1}, \quad \hat{K}_j^j = \frac{\alpha_{j-1}\beta_{j-1}}{2\gamma}\left(\frac{D_0\Delta}{\Delta}\right), \tag{90}$$

$$D_0\hat{K}_j^j = \frac{\alpha_{j-1}\beta_{j-1}}{2\gamma}\, D_0\left(\frac{D_0\Delta}{\Delta}\right) = -2\hat{K}_j^1\hat{K}_1^j,$$

where $\Delta = 1 + h^2\gamma$ is the Fredholm determinant of Eq. (87)

(3b) The case of $i_{\max} = 2$ is studied in Appendix B.

B. Application to the N.L.P.D.E.

In both the coupled Schrödinger case and modified KdV equations we want to explicitly write down the solution when the kernel of the integral inversion equation is degenerated and corresponds in Eq. (89) to $i_{\max} = 1$ or $i_{\max} = 2$.

(1) $i_{\max} = 1$ *and Schrödinger cubic coupled case*: We consider

$$F = f(x)f(y)\alpha_m(t), \quad \alpha_m(t) = \alpha_m(0)\exp 4i\nu^2 t,$$

$$f(x) = \exp[i\nu\textstyle\int^x \mu^{-1}(u)\,du],$$

$$\mathrm{Im}\nu > 0, \quad G = xF^*, \quad \gamma(0) = \chi\textstyle\sum|\alpha_m(0)|^2,$$

$$\xi = 2\,\mathrm{Im}\nu[\textstyle\int^x \mu^{-1}\,du + 4(\mathrm{Re}\nu)\,t]$$

and the solution Eq. (90) can be written

$$\hat{K}_j^1 = \alpha_{j-1}^*(0)(\exp(i\phi))V, \quad j > 1$$

$$\phi = -4[(\mathrm{Re}\nu)^2 t - (\mathrm{Im}\nu)^2 t] - 2\mathrm{Re}\nu\textstyle\int^x \mu^{-1}\,du,$$

$$V = \frac{4\chi\,\mathrm{Im}\nu}{\sqrt{\gamma(0)}}\left\{\cosh\left[\xi - \tfrac{1}{2}\log\left(\frac{\gamma(0)}{4(\mathrm{Im}\nu)^2}\right)\right]\right\}^{-1}. \tag{91}$$

The set $\{\hat{K}_j^1\}$ given by (91) is a solution of the n. l. p. d. e. Eq. (81). For $\kappa > 0$, this solution is a soliton, whereas for $\kappa < 0$, it is a ghost having a pole at

$$\xi = \tfrac{1}{2}\log[-\kappa\textstyle\sum|\alpha_m(0)|^2/4(\mathrm{Im}\nu)^2].$$

We remark that all the solutions $\hat{K}_j^1$ $(j = 2, \ldots, n+1)$ differ only by the constant $\alpha_{j-1}(0)$.

(2) $i_{\max} = 1$ *and modified KdV* (86): We consider

$$F = f(x)f(y)\alpha_m(t), \quad \alpha_m(t) = \alpha_m(0)\exp 8\nu^3 t,$$

$$f(x) = \exp[-\nu\textstyle\int^x \mu^{-1}(u)\,du], \quad \nu > 0, \quad G = \chi F,$$

$$\gamma(0) = \chi\textstyle\sum\alpha_m^2(0), \quad \xi = 2\nu[\textstyle\int^x \mu^{-1}\,du + 4\nu^2 t],$$

and the solution or Eq. (90) can be written

$$\hat{K}_j^1 = \alpha_{j-1}(0)\,V, \quad V = 4x\nu/\sqrt{\gamma(0)}\,[\cosh\{\xi - \tfrac{1}{2}\log[\gamma(0)/4\nu^2]\}]^{-1}. \tag{92}$$

For $\kappa > 0$ we have a soliton and for $\kappa < 0$ a ghost.

(3) The application of degenerate kernels F_m, G_m, with $i_{\max} = 2$, to the coupled Schrödinger cubic case is done in Appendix B.

VI. CONCLUSION

In this paper we have treated mainly two problems. The first problem is the extension, to the matrix case, of a simple method giving the inversion equation corresponding to scalar second order differential equations. The extension in the matrix case has been obtained both for first order (twofold eigenvalues) and second order differential equations. This method is purely algebraic and, contrary to the one currently used, does not involve any study of the eigenvalue plane. It has also the advantage to be not restricted with purely differential operators $(\partial/\partial x)^p$ but can be enlarged to $(\mu(x)\,\partial/\partial x)^p$. (This means that we are not restricted to inverse Fourier transforms.) We find that the matrix kernel of the inversion integral equation must satisfy well-defined l. p. d. e. and boundary conditions.

In the two cases treated here, these l. p. d. e. being symmetric with respect to the two variables of the inversion equations, it follows that the kernels are symmetric with respect to these two variables. A remarkable fact is the similarity between the structure of the equations and conditions to satisfy when we go from scalar to matrix case as well as from first order to second order differential equations. The cases not treated here (more than a two eigenvalue problem) and which will be presented in the next paper[5] lead also to similar structures. Finally we emphasize that a great simplification of the formalism (in this twofold eigenvalue problem) has been achieved by the introduction of a two-component algebra which permits us to ignore practically the $(n \times n)$ original dimension of the problem.

The second problem was to extract directly from the integral inversion equations, the properties leading to the explicit determination of the n. l. p. d. e., the time variable being included now. A typical solution of the inversion equation has two variables x, y, and a parameter t. We find that, in order to get the n. l. p. d. e. for the solutions of the inversion equation, the kernels must satisfy well-defined l. p. d. e. with respect to the three variables (x, y, t). (This result was also obtained by Zakharov and Shabat[4]). Then, these solutions, written in a matrix formulation, satisfy also n. l. p. d. e. with respect to (x, y, t), the linear part of the equation being the same as the one of the kernels of the inversion equations generating the solutions. Restricting the solutions to $x = y$, one finds that they satisfy the coupled classical n. l. p. d. e., where the derivations $(\partial/\partial x)^p$ are replaced by $(\mu(x)\,\partial/\partial x)^p$. This means that the classical entirely integrable n. l. p. d. e. are particular members of a larger family of solvable n. l. p. d. e.

ACKNOWLEDGMENTS

I would like to thank D. Vautherin and P. Bonche for their interest in the work and discussions concerning n. l. p. d. e. in nuclear physics.

APPENDIX A

A1: Using very elementary algebra, we deduce a set of integral equations with the same kernel as that of the integral inversion Eq. (33) but with different free terms:

$$\begin{pmatrix} M_{\mathrm{I}} \\ N_{\mathrm{II}} \end{pmatrix}_j = \begin{pmatrix} M_{\mathrm{I}}^0 \\ N_{\mathrm{II}}^0 \end{pmatrix}_j + \int_x^\infty \mu^{-1} \begin{pmatrix} 0 & \mathcal{J} \\ \mathcal{J} & 0 \end{pmatrix} \begin{pmatrix} M_{\mathrm{I}} \\ N_{\mathrm{II}} \end{pmatrix}_j .$$

$j=1$:

$$\begin{pmatrix}\mathcal{M}_{\rm I}\\ \mathcal{N}_{\rm I}\end{pmatrix}=\begin{pmatrix}\sum_1^3 \gamma_i O_{\alpha_i,i} K_{\rm I}\\ \sum_1^3 \gamma_i \tilde{O}_{\alpha_i,i} K_{\rm II}\end{pmatrix},\quad \begin{pmatrix}\mathcal{M}_{\rm I}^0\\ \mathcal{N}_{\rm II}^0\end{pmatrix}=\begin{pmatrix}\sum_1^3 \gamma_i \tilde{I}_i(2l_i+2^i\beta_i)\\ 0\end{pmatrix}.$$

$j=2$:

$$\begin{pmatrix}\mathcal{M}_{\rm I}\\ \mathcal{N}_{\rm I}\end{pmatrix}=\begin{pmatrix}O_s K_{\rm I}+2K_{\rm I}\hat{K}_{\rm II}\\ O_d K_{\rm II}+2K_{\rm II}\hat{K}_{\rm II}\end{pmatrix},\quad \begin{pmatrix}\mathcal{M}_{\rm I}^0\\ \mathcal{N}_{\rm II}^0\end{pmatrix}=\begin{pmatrix}2D_0\mathcal{J}\\ 0\end{pmatrix}.$$

$j=3$:

$$\begin{pmatrix}\mathcal{M}_{\rm I}\\ \mathcal{N}_{\rm II}\end{pmatrix}=\begin{pmatrix}\sum \gamma_i O_{\alpha_i,i} K_{\rm I}-6\gamma_3(O_s K_{\rm I}+2K_{\rm I}\hat{K}_{\rm II})D_0\hat{K}_{\rm II}\\ \sum \gamma_i \tilde{O}_{\alpha_i,i} K_{\rm II}-6\gamma_3(O_d K_{\rm II}+2K_{\rm II}\hat{K}_{\rm II})D_0\hat{K}_{\rm II}\end{pmatrix},\quad \begin{pmatrix}\mathcal{M}_{\rm I}^0\\ \mathcal{N}_{\rm II}^0\end{pmatrix}=\begin{pmatrix}4\gamma_2\tilde{I}\,\mathcal{J}D_0\hat{K}_{\rm II}+6\gamma_3\mathcal{J}\widehat{O_s O_d K_{\rm II}}\\ 0\end{pmatrix}.$$

$j=4$:

$$\begin{pmatrix}\mathcal{M}_{\rm I}\\ \mathcal{N}_{\rm II}\end{pmatrix}=\begin{pmatrix}4\gamma_2\tilde{I}K_{\rm I}D_0\hat{K}_{\rm II}\\ -4\gamma_2\tilde{I}K_{\rm II}D_0\hat{K}_{\rm II}\end{pmatrix},\quad \begin{pmatrix}\mathcal{M}_{\rm I}^0\\ \mathcal{N}_{\rm II}^0\end{pmatrix}=\begin{pmatrix}4\gamma_2\tilde{I}\,\mathcal{J}D_0\hat{K}_{\rm II}\\ 0\end{pmatrix}.$$

$j=5$: We subtract the solution $j=4$ from the solution $j=3$:

$$\begin{pmatrix}\mathcal{M}_{\rm I}\\ \mathcal{N}_{\rm II}\end{pmatrix}=\begin{pmatrix}\sum \gamma_i O_{\alpha_i,i} K_{\rm I}-[6\gamma_3(O_s K_{\rm I}+2K_{\rm I}\hat{K}_{\rm II})+4\gamma_2\tilde{I}K_{\rm I}]D_0\hat{K}_{\rm II}\\ \sum \gamma_i \tilde{O}_{\alpha_i,i} K_{\rm II}-[6\gamma_3(O_d K_{\rm II}+2K_{\rm II}\hat{K}_{\rm II})-4\gamma_2\tilde{I}K_{\rm II}]D_0\hat{K}_{\rm II}\end{pmatrix},\quad \begin{pmatrix}\mathcal{M}_{\rm I}^0\\ \mathcal{N}_{\rm II}^0\end{pmatrix}=\begin{pmatrix}6\gamma_3\mathcal{J}\widehat{O_s O_d K_{\rm II}}\\ 0\end{pmatrix}.$$

$j=6$: We multiply Eq. (33) on the right by $6\gamma_3\widehat{O_s O_d K_{\rm II}}$ and subtract to the solution $j=5$, taking into account the relation (55). We are left finally with a identically zero free term. From the assumed uniqueness, except for discretes values, of the solution of the inversion equation (33), we obtain that the solution corresponding to this free term must also be identically zero, $D_0\hat{K}_{\rm II}=-2\hat{K}_{\rm I}\hat{K}_{\rm I}$:

$$\sum_1^3 \gamma_i O_{\alpha_i,i} K_{\rm I}-6\gamma_3(O_s K_{\rm I}D_0\hat{K}_{\rm II})+12\gamma_3 K_{\rm I}\hat{K}_{\rm I}D_0\hat{K}_{\rm I}-4\gamma_2\tilde{I}K_{\rm I}D_0\hat{K}_{\rm II}=0,$$

$$\sum_1^3 \gamma_i \tilde{O}_{\alpha_i,i} K_{\rm II}-6\gamma_3(O_d K_{\rm II}D_0\hat{K}_{\rm II})+12\gamma_3 K_{\rm II}\hat{K}_{\rm I}D_0\hat{K}_{\rm I}+4\gamma_2\tilde{I}K_{\rm II}D_0\hat{K}_{\rm II}=0.$$

A2: Determination of the elements of the matrix solution (33). We can decouple the two solutions $K_{\rm I}$ and $K_{\rm II}$ of Eq. (33):

$$K_{\rm I}=\mathcal{J}+\int\mu^{-1}(\int\mu^{-1}\mathcal{J}\mathcal{J})K_{\rm I},\quad K_{\rm II}=\int\mu^{-1}\mathcal{J}\mathcal{J}+\int\mu^{-1}(\int\mu^{-1}\mathcal{J}\mathcal{J})K_{\rm II}$$

with the following structure for the kernel [K_j^1 and K_1^1 are given in (87)]:

$$\int\mu^{-1}\mathcal{J}\mathcal{J}=-\int\mu^{-1}\begin{pmatrix}\sum G_i F_i & 0\\ 0 & \begin{matrix}F_1G_1\cdots F_1G_n\\ \vdots\quad\ \vdots\\ F_nG_1\cdots F_nG_n\end{matrix}\end{pmatrix},\quad \begin{pmatrix}K_1^2\\ \vdots\\ K_1^{n+1}\end{pmatrix}=\begin{pmatrix}F_1\\ \vdots\\ F_{n+1}\end{pmatrix}-\int_x^\infty\mu^{-1}\int_x^\infty\mu^{-1}\begin{pmatrix}F_1G_1\cdots F_1G_n\\ \vdots\quad\ \vdots\\ F_nG_1\cdots F_nG_n\end{pmatrix}\begin{pmatrix}K_1^2\\ \vdots\\ K_1^{n+1}\end{pmatrix},$$

$$\hat{K}_j^1=\int_x^\infty\mu^{-1}(s)\,F_{j-1}(s,x)\,K_j^1(x,s)\,ds,\quad j>1.$$

Whereas the equations for K_j^1 decouple, this property does not not hold for K_1^j. However, practically for the n. l. p. d. e. studied in this paper, we link $\hat{K}_j^1$ and $\hat{K}_1^j$, and so it is sufficient to consider K_j^1.

A3: We consider the integral equation, where a is a fixed constant [$a=+\infty$ in Eq. (88)]

$$\phi^x(y)+H^x(y,x)+\int_x^a H^x(y,s)\,\phi^x(s)\,ds=0 \tag{A1}$$

with the kernel

$$H^x(y,s)=\int_x^a\mu^{-1}(u)\,\mu^{-1}(y)\,G(u,y)\,F(s,u)\,du. \tag{A2}$$

Let us call Δ the Fredholm determinant of (A1), (A2). *We want to show*

$$\phi^x(x)=\frac{1}{2}\frac{\partial}{\partial x}\log\Delta. \tag{A3}$$

The Fredholm solution of (A1) can be written:

$$\mathcal{D}(\phi^x(x)+H^x(x,x))=\int_x^a N(x,t_1)\,H^x(t_1,x)\,dt_1,$$

$$\mathcal{D}(x)=1+\sum_1^\infty \frac{D_n(x)}{n!},\quad D_n(x)=\int_x^a dt_1\cdots\int_x^a dt_n\,E_n^x(t_1,\ldots,t_n),$$

$$N=\sum_0^\infty \frac{N_n(x,t_1)}{n!},\quad N_n(x,t_1)=\int_x^a dt_2\cdots\int_x^a dt_{n+1}\,H_n^x(x,t_1,\ldots,t_{n+1}),$$

$$E_n^x=\begin{vmatrix} & \cdot & \\ H^x(t_1,t_1)\cdots H^x(t_1,t_n) & & \\ \vdots & & \vdots \\ H^x(t_n,t_1)\cdots H^x(t_n,t_n) & & \end{vmatrix},\quad H_n^x=\begin{vmatrix} H^x(x,t_1)\,H^x(t_1,t_2)\cdots H^x(t_1,t_{n+1}) \\ H^x(t_2,t_1)\,H^x(t_2,t_2)\cdots H^x(t_1,t_{n+1}) \\ \vdots \\ H^x(t_{n+1},t_1)\,H^x(t_{n+1},t_2)\cdots H^x(t_{n+1},t_{n+1}) \end{vmatrix}.$$

In order to prove (A3) it is sufficient so show:

$$\frac{1}{2(n+2)}\,\frac{\partial}{\partial x}\,D_{n+2}=-H^x(x,x)\,D_{n+1}(x)+(n+1)\int_x^a N_n(x,t_1)\,H^x(t_1,x)\,dt_1\,. \tag{A4}$$

Let us define $(n+2)^{-1}(\partial/\partial x)\,D_{n+2}=A_{n+2}+B_{n+2}$ with

$$A_{n+2}=-\int_x^a dt_2\cdots\int_x^a dt_{n+2}\,E_{n+2}^x(x,t_2,\ldots,t_{n+2}),\quad B_{n+2}=(n+2)^{-1}\int_x^a dt_1\cdots\int_x^a dt_{n+2}\,\frac{\partial}{\partial x}\,E_{n+2}^x(t_1,\ldots,t_{n+2}).$$

We know that A_{n+2} is equal to the rhs of (A4). It follows that we must prove $A_{n+2}=B_{n+2}$ or $A_n=B_n$. (See, for instance, Ref. 7, where this property is proved in Appendix A.) Let us define:

$$E_{n,k}^x=\begin{vmatrix} H^x(t_1,t_1)\cdots H^x(t_1,t_n) \\ \cdots \\ \dfrac{\partial}{\partial x}H^x(t_k,t_1)\cdots\dfrac{\partial}{\partial x}H^x(t_k,t_n) \\ \cdots \\ H^x(t_n,t_1)\cdots H^x(t_n,t_n) \end{vmatrix}.$$

We have $B_n=\sum_{k=1}^n n^{-1}\int\cdots\int E_{n,k}^x$. Furthermore we have $\int\cdots\int E_{n,k}^x=\int\cdots\int E_{n,1}^x$. This last property resulting from the exchanges of the first and the kth line, the first and the $k$the column, and $t_1\rightleftharpoons t_k$. We get

$$B_n=\int\cdots\int\prod_1^n dt_i\,\mu^{-1}(t_1)\begin{vmatrix} \dfrac{\partial}{\partial x}\displaystyle\int_x^a G(t_1,t_1')\,F(t_1',t_1)\,\mu^{-1}(t_1')\,dt_1'\cdots & \dfrac{\partial}{\partial x}\displaystyle\int_x^a G(t_1,t_1')\,F(t_1',t_n)\,\mu^{-1}(t_1')\,dt \\ H^x\cdots & \cdots H^x \\ \cdots & \cdots \\ H^x\cdots & H^x \end{vmatrix},$$

$$B_n=-\int\cdots\int\prod_1^n \mu^{-1}(t_i)\,dt_i\,\mu^{-1}(x)\prod_2^n dt_i'\mu^{-1}(t_i')\begin{vmatrix} G(t_1,x)\,F(x,t_1),\ldots,G(t_1,x)\,F(x,t_n) \\ G(t_2,t_1')\,F(t_1',t_1),\ldots,G(t_2,t_2')\,F(t_2',t_n) \\ \cdots \\ G(t_n,t_n')\,F(t_n',t_1),\ldots,G(t_n,t_n')\,F(t_n',t_n) \end{vmatrix},$$

$$\vdots$$

$$B_n=-\int\cdots\int\prod_1^n \mu^{-1}(t_i)\,dt_i\prod_1^n dt_i'\,\mu^{-1}(t_i')\,\mu^{-1}(x)\,G(t_1,x)\,G(t_2,t_2')\cdots G(t_n,t_n')\,F_n,$$

with

$$F_n=\begin{vmatrix} F(x,t_1),\,F(x,t_2)\cdots F(x,t_n) \\ F(t_2',t_1),F(t_2',t_2)\cdots F(t_2',t_n) \\ \cdots \\ F(t_n',t_1),F(t_n',t_2)\cdots F(t_n',t_n) \end{vmatrix}=\begin{vmatrix} F(x,t_1)\,F(t_2',t_1)\cdots F(t_n',t_1) \\ F(x,t_2)\,F(t_2',t_2)\cdots F(t_n',t_2) \\ \cdots \\ F(x,t_n)\,F(t_2',t_n)\cdots F(t_n',t_n) \end{vmatrix},$$

We get

$$B_n=-\int\cdots\int\prod_1^n \mu^{-1}(t_i)\,dt_i\prod_1^n dt_i'\,\mu^{-1}(t_i')\,\mu^{-1}(x)\,E_n,$$

$$E_n=\begin{vmatrix} G(t_1,x)\,F(x,t_1),\,G(t_1,x)\,F(t_2',t_1)\cdots G(t_1,x)\,F(t_n',t_1) \\ G(t_2,t_2')\,F(x,t_2),\,G(t_2,t_2')\,F(t_2',t_2)\cdots G(t_2,t_2')\,F(t_n',t_2) \\ \cdots \\ G(t_n,t_n')\,F(x,t_n),\,G(t_n,t_n')\,F(t_2',t_n)\cdots G(t_n,t_n')\,F(t_n',t_n) \end{vmatrix},$$

$$B_n=-\int\cdots\int\prod_2^n dt_i'\begin{vmatrix}H^x(x,x), H^x(x,t_2')\cdots H^x(x,t_n')\\ H^x(t_2',x), H^x(t_2',t_2')\cdots H^x(t_2',t_n')\\ \cdots\\ H^x(t_n',x), H^x(t_n',t_2')\cdots H^x(t_n',t_n')\end{vmatrix},\quad B_n=-\int\cdots\int\prod_1^n dt_i\,E_n(x,t_2,\ldots,t_n)=A_n.$$

A4: Let us consider $\mu>0$ and either the KdV or the Schrödinger cubic case, where a is still a fixed constant:

$$F(x,y)=F(y,x),\quad G=\chi\tilde F,\quad \tilde F(x,y)=\tilde F(y,x),\quad \chi \text{ real},\quad \mu>0,$$
$$\tilde F\equiv F \text{ if } F \text{ is real},\quad \tilde F\equiv F^* \text{ if } F \text{ is complex}. \tag{A5}$$

The Fredholm determinant of (A1) and (A2) can be written

$$\mathcal{D}(\kappa,x)=\exp[-\textstyle\sum(-\kappa)^n n^{-1}A_n(x)],\quad A_n(x)=\int_x^a dt_1\cdots\int_x^a dt_{2n}\prod_1^{2n}\mu^{-1}(t_i)\,\tilde F(t_1,t_2)F(t_2,t_3)\tilde F\cdots F(t_{2n},t_1), \tag{A6}$$

$$A_n(x)=\int_x^a\int_x^a dt_1\,dt_{n+1}\,\mu^{-1}(t_1)\,\mu^{-1}(t_{n+1})\Big|\int\cdots\int\prod_2^n\mu^{-1}(t_i)\,\tilde F(t_1,t_2)F(t_2,t_3)\tilde F\cdots\tilde{\tilde F}(t_n,t_{n+1})\Big|^2,$$

where $\tilde{\tilde F}=\tilde F$ if n is odd and $\tilde{\tilde F}=F$ if n is even. As in Ref. 7, Appendix A2, we want to study the smallest modulus eigenvalue $\kappa_1(x)$ or the first zero of $\mathcal{D}$, or the radius of convergence of the series in (A6). We get

$$\mu(x)\frac{\partial A_n}{\partial x}=-2n\int\cdots\int\prod_1^{2n+1}dt_i\,\mu^{-1}(t_i)\,\tilde F(r,t_1)F(t_1,t_2)\tilde F\cdots F(t_{2n-1},r)$$
$$=-2n\int_x^a\mu^{-1}(t_n)\,dt_n\Big|\int_x^a\cdots\int_x^a\prod_1^{n-1}\mu^{-1}(t_i)\,dt_i\,\tilde F(r,t_1)F(t_1,t_2)\tilde F\cdots\tilde{\tilde F}(t_{n-1},t_n)\Big|^2 \tag{A7}$$

Let us consider $0\le x<a$. We get $A_n(x)>0$ and $\kappa_1(x)<0$, $\partial A_n/\partial x<0$ and consequently $A_n(x_1)\ge A_n(x_2)$ if $0\le x_1\le x_2$. It follows that $|\kappa_1(0)|\le|\kappa_1(x)|$ for $x>0$. We consider $a=+\infty$, then $\tilde F=F$ corresponds to the KdV case and $\tilde F=F^*$ to the Schrödinger cubic case. In both cases for $x\in[0,\infty]$, the first ghost appears at $x=0$ when κ crosses $\kappa_1(0)$ and for $|\kappa|<|\kappa_1(0)|$ there is no ghost. On the contrary for $x<0$, the above results do not apply and the simple examples exhibited in Eqs. (91)—(92) show that for any $\kappa<0$ value, finite but as small as we want, a ghost is present for some $x<0$ value. However, a result can still be obtained for $x<0$ and $\kappa>0$. We remark that the Fredholm determinant (defined on $[x,a]$) of the kernel $M^x(y,t)=-\kappa\int_x^a(\mu(t)\,\mu(y))^{-1/2}\tilde F(y,u)F(u,t)\,\mu^{-1}(u)\,du$ is also given by (A6). We consider the modified KdV case where $\tilde F=F$ is real and put $a=+\infty$. In this case M^x is the second iterated kernel of the symmetric kernel $\mu^{-1/2}(y)F(y,t)\,\mu^{-1/2}(t)$. It follows that the κ eigenvalues are negative. Consequently, even for $x<0$ but $\kappa>0$, $\mathcal{D}$ cannot vanish and there is no ghost.

APPENDIX B

B1: *Solution of the inversion equation* for K_j^1 [see Eq. (87a)] when (F_m,G_m) are degenerated kernels:

$$F_m(x,y)=\sum_1^{i_{\max}}f_i(x)f_i(y)\,\alpha_m^i,\quad G_m(x,y)=\sum_1^{i_{\max}}g_i(x)g_i(y)\,\beta_m^i. \tag{B1}$$

Let us define

$$\int\mu^{-1}g_if_l=h_{il},\quad \int\mu^{-1}f_l(u)K_j^1(x,u)\,du=H_{j,l},\quad \sum_m\alpha_m^i\beta_m^l=\gamma_{il}. \tag{B2}$$

We get

$$-K_j^1=\sum_i g_i(y)\Big[g_i(x)\beta_{j-1}^i+\sum_{m=1}^n\beta_m^i\sum_l\alpha_m^l h_{il}H_{j,l}\Big] \tag{B3}$$

$$-H_{j,l'}=\sum h_{il'}\Big[g_i\beta_{j-1}^i+\sum_m\beta_m^i\sum_l\alpha_m^l h_{il}H_{j,l}\Big]. \tag{B4}$$

We are reduced to a system of linear algebraic equations. *Taking* $i_{\max}=2$, *we get*

$$\begin{pmatrix}1+h_{11}^2\gamma_{11}+h_{21}^2\gamma_{12}, h_{12}h_{11}\gamma_{21}+h_{22}h_{21}\gamma_{22}\\ h_{12}h_{11}\gamma_{11}+h_{22}h_{21}\gamma_{12}, 1+h_{12}^2\gamma_{21}+h_{22}^2\gamma_{22}\end{pmatrix}\begin{pmatrix}H_{j,1}\\ H_{j,2}\end{pmatrix}+\begin{pmatrix}h_{11}g_1\beta_{j-1}^1+h_{21}g_2\beta_{j-1}^2\\ h_{12}g_1\beta_{j-1}^1+h_{22}g_2\beta_{j-1}^2\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

Solving this system and substituting in (B3), we get

$$-K_j^1(x,y)=N/\mathcal{D},\quad i=1,2,\quad k=1,2,\quad i\ne k,$$
$$\mathcal{D}=1+\sum_{i=1}^2h_{ii}^2\gamma_{ii}+h_{ik}^2\gamma_{ki}+(h_{11}h_{22}-h_{21}h_{12})(\gamma_{11}\gamma_{22}h_{11}h_{22}-\gamma_{12}\gamma_{21}h_{21}h_{12}),$$
$$N=\sum_{i=1}^2g_i(x)\beta_{j-1}^i\big[g_i(y)(1+h_{ki}^2\gamma_{ik}+h_{kk}^2\gamma_{kk})-g_k(y)(\gamma_{kk}h_{kk}h_{ik}+\gamma_{ik}h_{ki}h_{ii})\big], \tag{B5}$$

$\mathcal{D}$ being the Fredholm determinant of Eq. (87a).

B2: We introduce a t dependence in α_m^i and write down for $i_{\max}=2$ the solution of the n. l. p. d. e. corresponding to the Schrödinger cubic coupled $(D_0(x)=\mu(\partial/\partial x))$ case. We consider the particular frame (center-of-mass) introduced previously by Dolan[8] for the pure cubic case. We define $G_m=\chi F_m^*$, χ real $a>0$, and

$$f_i(x)=\exp[i\nu_i\int^x\mu^{-1}(u)\,du],\quad \nu_1=v/4+ia/2,\quad \nu_2=-v/4+ia/2,\quad \alpha_m^i=\alpha_m^i(0)\exp(4i\nu^2t). \tag{B6}$$

We get

$$\begin{aligned}&g_i=f_i^*,\quad \beta_m^i=x(\alpha_m^i)^*,\quad \gamma_{11}=\gamma_{11}(0)\exp(-2avt),\quad \gamma_{22}=\gamma_{22}(0)\exp(2avt),\\&\gamma_{ii}(0)=\chi\textstyle\sum|\alpha_m^i(0)|^2,\quad \gamma_{12}(t)=\gamma_{12}(0)=\chi\sum\alpha_m^1(\alpha_m^2)^*=\gamma_{21}^*,\quad h_{21}=h_{12}^*,\\&h_{ii}=a^{-1}\exp(-a\textstyle\int^x\mu^{-1}\,du),\quad h_{21}=i(v/2+ia)^{-1}\exp[i(v/2+ia)\int^x\mu^{-1}\,du].\end{aligned} \tag{B7}$$

We calculate first the Fredholm determinant given in (B5) and after $-\hat{K}_j^1\Delta=N$:

$$\begin{aligned}\Delta=1+(2/a^2)\exp(-2a\textstyle\int^x\mu^{-1}\,du)\{(\chi/|\chi|)\sqrt{\gamma_{11}(0)\gamma_{2i}(0)}\cosh(2avt+\tfrac{1}{2}\log[\gamma_{22}(0)/\gamma_{11}(0)]+\gamma_{12}(0)\,\mathrm{Re}\exp(ivx)[2ia/(v+2ia)]^2\}\\+[\exp(-4a\textstyle\int^x\mu^{-1}\,du)][v^2/(v^2+4a^2)^2]\{v^2\gamma_{11}(0)\gamma_{22}(0)+4a^2[\gamma_{11}(0)\gamma_{22}(0)-\gamma_{12}^2(0)]\},\end{aligned} \tag{B8}$$

$$\begin{aligned}N=x\exp[i(-v/4+a^2)t]\{&\exp(\tfrac{1}{2}iv\textstyle\int^x\mu^{-1}\,du)[\alpha_{j-1}^{2*}(0)\exp(-a\int^x\mu^{-1}\,du+avt)\\&+\{4x/[(v+2ia)^2]\}\{\alpha_{j-1}^{1*}(0)[\gamma_{12}(0)\,v/2ia]+\alpha_{j-1}^{2*}[\gamma_{11}(0)\,v/2a^2](v/2+ia)\}\exp[-(3a\textstyle\int^x\mu^{-1}\,du+avt)]]\\&+\exp(-\tfrac{1}{2}vi\textstyle\int^x\mu^{-1}\,du)\{\alpha_{j-1}^{1*}(0)\exp[-(a\int\mu^{-1}\,du+avt)]+[4\chi/(v-2ia)^2]\\&\times[\alpha_{j-1}^{1*}(0)[\gamma_{22}(0)\,v/2a^2](v/2-ia)+i\alpha_{j-1}^{2*}(0)[\gamma_{21}(0)\,v/2a]\exp(-3a\textstyle\int\mu^{-1}+avt)]]\}.\end{aligned} \tag{B9}$$

Let us assume $x>0$, α_m^i real and consider the denominator Δ in (B8). From the Schwarz inequality $\gamma_1(0)\gamma_{22}(0)-\gamma_{12}^2(0)\geq 0$ and the third term is positive. The bracket in the second term has a lower bound $\sqrt{\gamma_{11}(0)\gamma_{22}(0)}-\gamma_{12}(0)$ and so cannot be negative. It follows that in this case Δ is always positive and (B8) and (B9) represents a two-soliton solution.

[1]For a review articles and books see: A. C. Scott *et al.* IEEE **61**, 1443 (1973); *Dynamical Systems Theory and Applications*, edited by J. Moser, Lectures Notes in Physics 38 (Springer-Verlag, Berlin, 1975); *Linear and Nonlinear Waves*, edited by G. B. Whitham (Wiley, New York, 1974); *Studies of a Nonlinear Lattice*, edited by M. Toda [Phys. Rep. **18** (1975)]. For more recent references, see the last Calogero paper: Nuovo Cimento B **31**, 229 (1976).

[2](a) C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); Comm. Pure Appl. Math. **27**, 97 (1974). (b) P. D. Lax, Comm. Pure. Appl. Math. **21**, 467 (1968). (c) V. E. Z. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Phys. **61**, 118 (1971); **64**, 1627 (1973) [Sov. Phys. JETP **34**, 62 (1972); **37**, 823 (1973)]. (d) M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1962 (1973); G. L. Lamb, Phys. Rev. Lett. **31**, 197 (1973); D. J. Kaup, Stud. Appl. Math. **54**, 165 (1975). (e) M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973); Stud. Appl. Math. **53**, 249 (1974).

[3](a) S. V. Manakov, Zh. Eksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1974)]. (b) V. E. Zakharov and J. V. Manakov, Zh. Eksp. Teor. Fiz. Pis. Red. **18**, 413 (1973) [JETP Lett. **18**, 243 (1973)]; D. J. Kaup, Stud. Appl. Math. **55**, 9 (1976).

[4]V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **8**, 226 (1974); M. J. Ablowitz and R. Haberman, J. Math. Phys. **16**, 2301 (1975); A. B. Shabat, Funct. Anal. Appl. **9**, 75 (1975).

[5]H. Cornille, Rockefeller Preprint C00-2232B-124.

[6]H. Cornille, J. Math. Phys. **17**, 2143 (1976).

[7]H. Cornille, Nuovo Cimento A **14**, 141 (1973). We take the opportunity to complete a proof in Appendix A. 2 of that paper where the three following formulas are missing at the end of page 157:

$$\frac{dA_n}{dr}=-2n\int_r^a dt_1\cdots\int_r^a dt_{2n-1}F(r,t_1)F(t_1,t_2)\cdots F(t_{2n-2},t_{2n-1})\times F(t_{2n-1},r),$$

$$\frac{dA_n}{dr}=-2n\int_r^a dt_n\Big|\int_r^a dt_1\cdots\int_r^a dt_{n-1}F(r,t_1)\times F(t_1,t_2)\cdots F(t_{n-1},t_n)\Big|^2,$$

$$\frac{dA_n}{dr}<0\ \text{ if } 0\leq r<a,\quad \frac{dA_n}{dr}>0\ \text{ if } 0\leq a<r.$$

[8]L. Dolan, Phys. Rev. D **13**, 528 (1976).

U(n) lowering operators

Adam M. Bincer

Physics Department, University of Wisconsin, Madison, Wisconsin 53706
(Received 4 April 1977)

Lowering operators for the U(n) groups are given in product form.

An important concept in the study of representations of groups is that of lowering operators. By definition, the state of any weight in an irreducible representation may be obtained from the state of highest weight by application of the appropriate lowering operator. Thus the concept is useful not only in the study of unitary representations of compact groups (which necessarily have a state of highest weight), but also in the study of certain unitary representations of noncompact groups, namely, those in the discrete series.

In 1965 Nagel and Moshinsky[1] obtained explicit expressions for lowering operators for unitary groups, U(n). For the orthogonal groups, O(n), the construction of lowering operators was partially achieved by Pang and Hecht,[2] and completed by Wong.[3] To the best of my knowledge, lowering operators for the symplectic groups, Sp($2n$), do not appear in the literature.

I have devised a method for the construction of lowering operators that treats, with minor modifications, all classical groups on the same footing. It should, therefore, on the one hand fill the gap in the literature with respect to Sp($2n$) groups and, on the other hand, it should yield the O(n) operators in the elegant form obtained by Nagel and Moshinsky for the U(n) operators. The method relies heavily on the concept of vector operators as used by Okubo[4] for U(n) and generalized by Nwachuku and Rashid[5] for O(n) and Sp($2n$).

To test the method I have constructed the U(n) operators. The result is extremely simple, and I feel that workers in this field might be interested in it. Accordingly I present in this note just the result. The details of the derivation will appear in future publications dealing with the orthogonal and symplectic groups.

In the notation of Nagel and Moshinsky the lowering operator L_n^m is defined by

$$L_n^m \left| \begin{matrix} h_1 \cdots h_n \\ q_1 \cdots q_m \cdots q_{n-1} \end{matrix} \right\rangle \propto \left| \begin{matrix} h_1 \cdots h_n \\ q_1 \cdots q_m - 1 \cdots q_{n-1} \end{matrix} \right\rangle , \qquad 1 \leq m < n . \tag{1}$$

Our result for L_n^m is

$$L_n^m = \left\{ C \prod_{k=m+1}^{n-1} (C - c_k \mathbf{1}) \right\}_n^m , \tag{2}$$

$$c_k \equiv q_k - k + n - 1 , \tag{3}$$

and the product in (2) is defined according to Okubo's[4] convention for vector operators

$$(XY)_\beta^\alpha = \sum_{\rho=1}^{n-1} X_\beta^\rho Y_\rho^\alpha . \tag{4}$$

Also, for $m = n-1$,

$$\prod_{k=n}^{n-1} (C - c_k \mathbf{1}) \equiv \mathbf{1} . \tag{5}$$

Upon comparison of (2) with Nagel and Moshinsky, their Eq. (2.27), it is seen that even in the relatively simple case of U(n) the present method has the advantage of presenting L_n^m in a product form. It will be undoubtedly clear to the reader that the multiple sum over the μ_i's in Eq. (2.27) has not been really eliminated but rather absorbed into the notation by means of Okubo's convention. However, the one nontrivial sum in Eq. (2.27) (over the index p) has been eliminated.

In many applications one needs the normalized operators $L_{q_m - 1}^{q_m}$ defined by

$$L_{q_m - 1}^{q_m} \left| \begin{matrix} h_1 \cdots h_n \\ q_1 \cdots q_m \cdots q_{n-1} \end{matrix} \right\rangle = \left| \begin{matrix} h_1 \cdots h_n \\ q_1 \cdots q_m - 1 \cdots q_{n-1} \end{matrix} \right\rangle \tag{6}$$

We find

$$L_{q_m - 1}^{q_m} = \left\{ D \prod_{k=m+1}^{n-1} D(k) \right\}_n^m , \tag{7}$$

where

$$D_\beta^\alpha \equiv C_\beta^\alpha \left\{ (c_m - d_n - 1)(d_m - c_m + 1) \times \prod_{j=1}^{n-1} (d_j - c_m + 1)/(c_j - c_m + 1) \right\}^{-1/2} , \tag{8}$$

$$D(k)_\beta^\alpha \equiv (C_\beta^\alpha - c_k \delta_\beta^\alpha) \{ (c_k - c_m)(d_k - c_m + 1) \}^{-1/2} , \tag{9}$$

$$d_k \equiv h_k - k + n - 1 . \tag{10}$$

[1] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[2] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[3] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[4] S. Okubo, J. Math. Phys. **16**, 528 (1975).
[5] C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **17**, 1611 (1976).

Exact solution for the linear response of a hydrogenic atom to an external electromagnetic field. I. Frequencies below photoelectric threshold

Marshall Luban and Batya Nudler-Blum

Department of Physics, Bar-Ilan University, Ramat-Gan, Israel
(Received 17 January 1977)

We obtain an exact closed-form expression for the first-order correction $\delta\Psi(\mathbf{r},t)$ of the wavefunction of a hydrogenic atom of atomic number Z, initially in the ground state, due to a linearly polarized, monochromatic electromagnetic field treated in the dipole approximation. This has been obtained by employing the method of integral transformations to solve the inhomogeneous differential equation governing $\delta\Psi$. We find, for angular frequencies ω below the ionization frequency ω_c, that $\delta\Psi(\mathbf{r},t)$ is characterized by an exponential decrease for large r with a frequency-dependent range parameter $(a_0/Z)(1-\omega/\omega_c)^{-1/2}$, where $a_0=\hbar^2/me^2$ is the Bohr radius. As an application, starting with our result for $\delta\Psi$, we have calculated the frequency dependent polarizability $\alpha(\omega)$.

I. INTRODUCTION

In this series of two articles we derive an exact, closed-form expression for the first-order wavefunction of a hydrogenic atom of atomic number Z in the presence of a linearly polarized monochromatic electromagnetic field of angular frequency ω.[1] In particular, we deal with the case whereby in the absence of the electromagnetic field the atom is in its ground state. Furthermore, we assume that $\omega/\omega_c \ll 2/Z\alpha$, where $\hbar\omega_c = Z^2e^2/2a_0$ is the ionization energy, $a_0 \simeq 0.5$ Å is the usual Bohr radius, and $\alpha \simeq 1/137$ is the fine structure constant. For these values of ω one is justified in adopting the simplifying dipole approximation. In the present article we limit the discussion to the case that ω is less than ω_c. The corresponding problem where ω exceeds ω_c, but yet $\omega/\omega_c \ll 2/Z\alpha$, is dealt with in the second article of this series.

It is perhaps surprising that although the hydrogenic atom is probably one of the most exhaustively studied quantum-mechanical systems, an exact, closed-form expression for the first-order correction $\delta\Psi(\vec{\mathbf{r}},t)$ is known only for a *static* electric field.[2] To appreciate the complexity of the problem we recall that in standard first-order perturbation theory, the wavefunction is expanded in terms of the infinite set of discrete unperturbed bound state eigenfunctions, as well as the unperturbed continuum eigenfunctions. With considerable effort the relevant expansion coefficients can be calculated, but the remaining task of summing over the bound states and integrating over the continuum states degenerates to a fairly opaque and unrewarding exercise in numerical methods. Clearly, a more direct approach is required.

Such an approach was in fact suggested as far back as 1928 by Podolsky.[3] He showed that $\delta\Psi$ can be expressed in terms of two time-independent spatial functions which are solutions of a certain inhomogeneous second-degree partial differential equation [see Eqs. (4)–(7) below]. To obtain a solution in a useful form to these imposing differential equations is a problem of considerable complexity, and Podolsky was satisfied to expand $\delta\Psi$ in an infinite series of orthonormal functions involving the associated Laguerre polynomials. Unfortunately the unwieldy form of the infinite series prevents one from establishing any of the essential properties of $\delta\Psi$. However, Podolsky did succeed in obtaining the correct expression for the frequency dependent atomic polarizability for all frequencies up to the photoelectric threshold. Over the years this early paper was essentially forgotten except for an occasional reference by other workers[4,5] who succeeded in rederiving the atomic polarizability via some new techniques not involving $\delta\Psi$. The suggestion that one obtain the first-order perturbed wavefunction by solving directly the differential equation for $\delta\Psi$ appears to have been raised again only in 1951 by Kotani,[6] in connection with the problem of the hydrogen atom in a *static* electric field. Subsequently Dalgarno and Lewis[7] and Schwartz[8] extended this idea to a wider class of problems. We have solved the pair of Podolsky equations using the method of integral transformations[9] and thereby obtain $\delta\Psi$ for the case that the hydrogenic atom is in the ground state in the absence of the electromagnetic field.

We close this section with an outline of the format of the present article. Section II is devoted to establishing the relevant basic equations, and in particular to deriving the pair of Podolsky equations. We also show that the physically relevant solutions of the pair of Podolsky equations can be expressed in terms of the solutions of a single, ordinary second-order inhomogeneous differential equation [see Eq. (10)]. This differential equation is solved by the method of integral transformations in Sec. III. The solution of this equation is given in closed form and thus we obtain a closed-form integral representation of $\delta\Psi(\vec{\mathbf{r}},t)$. This form is particularly convenient for the purpose of establishing all essential properties of $\delta\Psi$.

Section IV is devoted to a simple application of our results for $\delta\Psi$. In particular we rederive the known expression for the frequency-dependent polarizability $\alpha(\omega)$. A new result given [see Eq. (46)] is a particularly useful decomposition of $\alpha(\omega)$, whereby the poles of α, occurring for the $1s$–np resonance frequencies, are described by a single term $\alpha_s(\omega)$. The remainder term $\bar{\alpha}(\omega)=\alpha(\omega)-\alpha_s(\omega)$ is a slowly varying function of ω. Fi-

nally, in Sec. V we summarize our present results.

In recent years a number of important physical quantities have been calculated for the hydrogenic atom starting from the nonrelativistic Coulomb Green's function.[10] These quantities include the cross sections for (i) the elastic scattering of photons including retardation effects; (ii) Compton scattering; (iii) various two-photon processes. Some of the most important papers are those listed in Refs. 5, 11–19. The pair of Podolsky equations must also be solvable using the nonrelativistic Coulomb Green's function. However, apart from the fact that the Green's function has a complex form, to complete the derivation of $\delta\Psi$ appears to be an extremely formidable mathematical problem. We prefer a simple, straightforward derivation of $\delta\Psi$, based on the method of integral transformations, which requires no prior knowledge of the Coulomb Green's function.

II. BASIC EQUATIONS

We consider a nonrelativistic hydrogenic atom of atomic number Z in the presence of an external, linearly polarized, electromagnetic wave of angular frequency ω. The scalar and vector potentials will be chosen in the Coulomb gauge and of the form

$$\phi = 0, \quad \mathbf{A} = -\hat{z}(F_0 c/\omega)\sin\omega(t - y/c). \tag{1}$$

Throughout the present work we assume that the electromagnetic field is so weak that we can restrict our attention to the linear response of the atom. We suppose further that in the remote past the atom was in its ground state $u_{100} = (Z^3/\pi a_0^3)^{1/2} e^{-Zr/a_0}$, where $a_0 = \hbar^2/me^2$ is the Bohr radius, and m is the reduced mass of the electron. Writing the wavefunction as

$$\Psi(\mathbf{r}, t) = e^{i\omega_c t}[u_{100} + \delta\psi(\mathbf{r}, t)], \tag{2}$$

the first-order time-dependent Schrödinger equation is

$$i\hbar\frac{\partial\Psi}{\partial t} = H_0\Psi + i(e\hbar F_0/m\omega) \times e^{i\omega_c t}\sin\omega(t - y/c)\frac{\partial u_{100}}{\partial z}, \tag{3}$$

where H_0 denotes the unperturbed Hamiltonian

$$H_0 = -(\hbar^2/2m)\nabla^2 - Ze^2/r, \tag{4}$$

the electron charge is denoted by $-e$ and $\hbar\omega_c = Z^2e^2/2a_0$ is the ionization energy. It is convenient to expand $\delta\psi(\mathbf{r}, t)$ as

$$\delta\psi(\mathbf{r}, t) = e^{i\omega t} f_+(\mathbf{r}; \omega) - e^{-i\omega t} f_-(\mathbf{r}; \omega), \tag{5}$$

so that the time-independent functions $f_\pm$ satisfy the pair of equations

$$[H_0 + \hbar(\omega_c \pm \omega)]f_\pm = -\left(\frac{e\hbar F_0}{2m\omega}\right)\left(\frac{\partial u_{100}}{\partial z}\right)e^{\mp i\omega y/c}. \tag{6}$$

To simplify these equations we shall adopt the dipole approximation, whereby the exponential on the right-hand side of Eq. (6) is replaced by unity. As pointed out earlier, this approximation is satisfied if $(\omega/c)(a_0/Z) \ll 1$, or, equivalently, if $\omega/\omega_c \ll 2/(Z\alpha)$, where α is the fine structure constant. Thus in the following we consider the pair of equations

$$[H_0 + \hbar(\omega_c \pm \omega)]f_\pm = -\left(\frac{e\hbar F_0}{2m\omega}\right)\left(\frac{\partial u_{100}}{\partial z}\right). \tag{7}$$

The first attempt to solve Eq. (7) directly was made by Podolsky[3] in 1928. The limitations of Podolsky's solution have been listed in the Introduction. Another early attempt to solve Eq. (6) was made by Sommerfeld and Schur[20] in their classic paper on the theory of the photoelectric effect for atomic hydrogen. Employing standard time-independent Schrödinger perturbation theory, those authors expanded the functions $f_\pm$ in terms of the unperturbed hydrogen bound state and continuum state eigenfunctions. For their purposes it was sufficient to calculate in closed form the leading term of the asymptotic expansion of $f_\pm$ for large values of r for the frequency range $\omega > \omega_c$. Apart from these early efforts we are not familiar with any other papers dealing with the calculation of $f_\pm$. In fact, in recent years the calculation of matrix elements, appearing in the amplitudes for physical processes, are performed circumventing the direct calculation of $\delta\psi(\mathbf{r}, t)$.[4,5,11–19]

In the following section we obtain the exact solution of Eq. (7) in closed form and thereby pave the way for establishing all essential properties of $\delta\psi$. As a first step in that direction we show now that the pair of partial differential equations of Eq. (7) can be reduced to a single, ordinary inhomogeneous differential equation. Choosing the z axis as polar axis and employing spherical coordinates we have $\partial u_{100}/\partial z = -(Z/a_0)(\cos\theta)u_{100}(r)$. Thus the angular dependence of $f_\pm(\mathbf{r}; \omega)$ will be described by a factor $\cos\theta$. We now express the electron–nucleus distance r in terms of a dimensionless variable x according to the relation

$$r = \mu(a_0/2Z)x, \tag{8}$$

where μ is a parameter to be chosen momentarily. It is then easy to show that the solutions of Eq. (7) can be written as

$$f_\pm(\mathbf{r}; \omega) = F\left(\frac{2Zr}{a_0\mu_\pm}; \mu_\pm\right)\cos\theta, \tag{9}$$

where the radial function $F(x; \mu)$ satisfies the equation

$$F'' + \frac{2}{x}F' + \left(-\frac{1}{4} + \frac{\mu}{x} - \frac{2}{x^2}\right)F = A(\mu)e^{-\mu x/2}, \tag{10}$$

$$A(\mu) = -\tfrac{1}{4}(eF_0 a_0/Z\hbar\omega)\mu^2(Z^3/\pi a_0^3)^{1/2}, \tag{11}$$

and the quantities $\mu_\pm$ are defined by

$$\mu_\pm = (1 \pm \omega/\omega_c)^{-1/2}. \tag{12}$$

We conclude then that the pair of functions $f_\pm$ are given in terms of a single function F satisfying (10). The relevant solution (10) must satisfy the following boundary conditions. At the origin $F(x; \mu)$ must vanish, for otherwise, because of the $\cos\theta$ factor, each of $f_\pm(0; \omega)$ will not have a unique value. In addition, as long as $\omega < \omega_c$, we require that $F(x; \mu)$ vanish for x tending to infinity. As we will see in the following, the latter boundary condition cannot be satisfied if $\mu = 2, 3, \ldots$, corresponding to the frequency of the applied field being in resonance with the dipole transitions from the $1s$ state. (For the frequency regime $\omega > \omega_c$ photoionization is possible and thus the appropriate boundary condition at infinity is

that $|f_\pm|$ remain bounded.)

As stated in the Introduction, in this article we are limiting our attention to the frequency regime $\omega < \omega_c$. Thus μ_+ ranges from 1 (static limit) to $2^{-1/2}$ (continuum edge) and μ_- ranges from 1 (static limit) to $+\infty$ (continuum edge). In the second article of this series we solve Eq. (7) for $\omega > \omega_c$. This corresponds to μ_+ decreasing from $2^{-1/2}$ to zero for increasing frequencies while μ_- decreases from $i\infty$ to 0 along the positive imaginary axis.

III. WAVEFUNCTION IN CLOSED FORM

A. Integral representation

In this section we obtain in closed form the solution of the differential equation of Eq. (10) which satisfies the required boundary conditions. Towards this end we define a new dependent variable $M(x;\mu)$ by the relation

$$F(x;\mu) = xe^{-x/2}M(x;\mu). \tag{13}$$

Substitution of Eq. (13) into Eq. (10) gives as the equation for M

$$xM'' + (4-x)M' + (\mu-2)M = A(\mu)e^{-(\mu-1)x/2}. \tag{14}$$

The homogeneous version of this differential equation is a special case of the confluent hypergeometric equation.

We now attempt to obtain the solution of Eq. (14), for values of μ in the range $\mu > 1$, in the form

$$M(x;\mu) = \int_\alpha^\beta dt\, u(t)e^{-xt} \quad (x>0), \tag{15}$$

where the function $u(t)$ and the real constants α and β are to be determined. This will be recognized as the method of integral transformations which is well known in the theory of *homogeneous*, second order, ordinary differential equations.[9] With this choice for M, Eq. (14) is equivalent to the equation

$$-u(t)t(t+1)e^{-xt}\Big|_{t=\alpha}^{\beta} + \int_\alpha^\beta dt\, e^{-xt}\left\{(\mu-2-4t)u + \frac{d}{dt}[(t^2+t)u]\right\} = A(\mu)e^{-(\mu-1)x/2}. \tag{16}$$

This equation is obtained by applying the differential operator $xd^2/dx^2 + (4-x)d/dx + \mu - 2$ to $M(x)$ of Eq. (15), and then integrating by parts. Equation (16) is satisfied if one can satisfy the following equations

$$\frac{d}{dt}[(t^2+t)u] + (\mu-2-4t)u = 0, \tag{17}$$

$$-u(\beta)\cdot\beta(\beta+1)e^{-\beta x} + u(\alpha)\cdot\alpha(\alpha+1)e^{-\alpha x} = A(\mu)e^{-(\mu-1)x/2}. \tag{18}$$

The first of these equations possesses a solution of the form

$$u(t) = Ct^{1-\mu}(1+t)^{1+\mu}, \tag{19}$$

where C is as yet an arbitrary constant. Equation (18) can be satisfied for three different choices of α and β: (i) $\alpha = 0, \beta = \frac{1}{2}(\mu-1)$; (ii) $\alpha = \frac{1}{2}(\mu-1), \beta = +\infty$; (iii) $\alpha = -1$, $\beta = \frac{1}{2}(\mu-1)$. Each choice of the pair α and β determines a possible solution of Eq. (14). However, a unique choice of α and β is established by requiring that the physical solution of Eq. (14) must satisfy the boundary conditions that $M(0)$ must be finite and $M(x)$ cannot grow as rapidly as $e^{x/2}$ for $x \to +\infty$. As shown in Appendix A, these requirements can be met only for the first of the above three possible choices of α and β. Furthermore, referring to Eq. (19) one finds that the term $u(\alpha)\alpha(\alpha+1)e^{-\alpha x}$ of Eq. (18) will vanish, when α is set equal to zero, only if $\mu < 2$. Finally, Eq. (18) will be satisfied if the heretofore arbitrary constant C of Eq. (19) is set equal to $C = -16A(\mu)(\mu-1)^{-(2-\mu)}(\mu+1)^{-(2+\mu)}$. Collecting the above results for α, β, and $u(t)$, and substituting in Eq. (15) yields the following expression for $M(x;\mu)$, which is valid for $1 < \mu < 2$ ($0 < \hbar\omega < |E_{1s} - E_{2p}|$),

$$M(x;\mu) = 4\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{\mu^2}{(\mu-1)^{2-\mu}(\mu+1)^{2+\mu}} \times \int_0^{(\mu-1)/2} dt\, e^{-xt}t^{1-\mu}(1+t)^{1+\mu}. \tag{20}$$

This result, in conjunction with Eqs. (9) and (13) gives $f_-(\vec{r};\omega)$ for $1 < \mu < 2$. The integral in Eq. (20) diverges at the lower limit for $\mu \geq 2$. Nevertheless, the expression in Eq. (20) can be continued analytically for all real, nonintegral (i.e., excluding exact resonances) values of μ (> 2), merely by integrating by parts a sufficient number of times. For example, integrating once by parts yields

$$M(x;\mu) = 4\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{\mu^2}{(\mu-1)^{2-\mu}(\mu+1)^{2+\mu}}\frac{1}{2-\mu} \times \left\{\tfrac{1}{8}(\mu+1)^{\mu+1}(\mu-1)^{2-\mu}e^{-(\mu-1)x/2} - \int_0^{(\mu-1)/2} dt\, e^{-xt}[1+\mu-x(1+t)]t^{2-\mu}(1+t)^{\mu}\right\} \tag{21}$$

thereby giving an expression for $f_-(r;\omega)$ of Eq. (9) which is valid for all values of μ in the range $1 < \mu < 3$ with the exception of $\mu = 2$.

For the range $0 < \mu < 1$, an analysis similar to the above yields

$$M(x;\mu) = 4\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{\mu^2}{(1-\mu)^{2-\mu}(1+\mu)^{2+\mu}} \times \int_0^{(1-\mu)/2} dt\, e^{xt}t^{1-\mu}(1-t)^{1+\mu}. \tag{22}$$

This form is relevant for the function $f_+(r;\omega)$ of Eq. (9), and it applies for *all* frequencies $\omega > 0$.

It is of interest to examine the solution of Eq. (14) for integral values of μ. For the case $\mu = 2$, corresponding to the $1s$-$2p$ resonance, Eq. (14) reduces to a *first* order equation for the function $M'(x)$ which can be solved directly by standard methods. One readily finds that

$$M(x;\mu=2) = M(0) + A(2)\int_0^x dt\, t^{-4}e^t \times \left[\tfrac{32}{27} - \tfrac{2}{3}e^{-3t/2}(t^3 + 2t^2 + \tfrac{8}{3}t + \tfrac{16}{9})\right], \tag{23}$$

where $M(0)$ is an arbitrary constant. Note first that the presence of this arbitrary constant is to be expected since any constant is a solution of the homogeneous version of Eq. (14) for $\mu = 2$. Secondly, note that the

quantity in square brackets is of order t^4 for $t \to 0$ so that there is no infrared divergence. However, for large x, the integral in Eq. (23) contributes a term which grows as $x^{-4}e^x$. The resulting expression for $F(x;\mu)$ will thus diverge as $x^{-3}e^{x/2}$ for large x. In short, for $\mu = 2$, it is impossible to satisfy the boundary condition that $F(x;\mu)$ vanish for large x. This behavior is actually heralded by the diverging simple-pole factor $(2-\mu)^{-1}$ appearing in Eq. (21). Furthermore, we have here the quantum analog of the behavior of a classical harmonic oscillator which has been driven for all times by an external electromagnetic field whose frequency equals the oscillator resonance frequency. For the classical oscillator the amplitude is infinitely large, whereas in the present case, the perturbed wavefunction, although finite for any finite value of r, fails to satisfy the large-r boundary condition, in fact growing exponentially for $r \to \infty$. Similar considerations apply for $\mu = 3, 4, \cdots$.

In the following we shall examine in detail the properties of $F(x;\mu)$, given by Eqs. (20) and (22). For obtaining numerical values of $F(x;\mu)$, the form of these equations is especially convenient, inasmuch as one can employ standard numerical techniques for evaluating the relevant integrals. We shall defer to Sec. III F, quoting and discussing numerical results for $F(x;\mu)$. We first derive a power series expansion for the integrals in Eqs. (20) and (22) as it displays several interesting properties.

B. Power series expansion

Writing the series expansion of e^{-xt} in powers of $-xt$, the integral in Eq. (20) is given by

$$\int_0^{(\mu-1)/2} dt\, e^{-xt} t^{1-\mu}(1+t)^{1+\mu} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!} \int_0^{(\mu-1)/2} dt\, t^{n+1-\mu}(1+t)^{1+\mu}.$$

The resulting integrals are given by[21]

$$\int_0^{(\mu-1)/2} dt\, t^{n+1-\mu}(1+t)^{1+\mu} = \left(\frac{\mu-1}{2}\right)^{n+2-\mu} \frac{1}{n+2-\mu} \times F(-1-\mu, n+2-\mu; n+3-\mu; -\tfrac{1}{2}(\mu-1)) = \tfrac{1}{16}(\mu-1)^{2-\mu}(\mu+1)^{2+\mu}\left(\frac{\mu-1}{2}\right)^n \frac{1}{n+2-\mu} \times F(n+4, 1; n+3-\mu; -\tfrac{1}{2}(\mu-1)), \quad (24)$$

where F denotes the hypergeometric function. Thus we obtain

$$M(x;\mu) = \frac{1}{4}\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0 a_0}{Z\hbar\omega}\right)\mu^2 \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\left(\frac{\mu-1}{2}\right)^n x^n \frac{1}{n+2-\mu} \times F(n+4, 1; n+3-\mu; -\tfrac{1}{2}(\mu-1)). \quad (25)$$

A similar treatment of the integral in Eq. (22) also yields Eq. (25). In fact, the series in Eq. (25) is defined for all nonintegral values of μ, even though it was obtained from Eq. (20) which applies only for $1 < \mu < 2$. Thus Eq. (25) constitutes an analytic continuation of Eqs. (20) and (22) which is valid for all nonintegral μ. Note that the coefficient of x^n has simple poles for $\mu = n+2, n+3, \cdots$. The pole at $\mu = n+2$ appears explicitly in Eq. (25), and the further poles for $\mu = n+3, n+4, \cdots$ originate in the hypergeometric function $F(n+4, 1; n+3-\mu; -\frac{1}{2}(\mu-1))$.

C. Asymptotic expansion

We now obtain the asymptotic expansion of $M(x;\mu)$ for large x. We consider first the expression for $M(x;\mu)$ of Eq. (20). As $x \to +\infty$ one can replace the factor $(1+t)^{1+\mu}$ of the integrand by its Taylor expansion in powers of t. Furthermore, one can extend the upper limit of integration to infinity, the resulting error being exponentially small. Thus one obtains

$$M(x;\mu) \underset{x\to+\infty}{\sim} 4\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0 a_0}{Z\hbar\omega}\right)\frac{\mu^2}{(\mu-1)^{2-\mu}(\mu+1)^{2+\mu}} \times \sum_{n=0}^{\infty} (-1-\mu)_n \frac{(-1)^n \Gamma(n+2-\mu)}{n!\, x^{n+2-\mu}}, \quad (26)$$

where $(b)_n = \Gamma(b+n)/\Gamma(b)$. This same result also obtains for Eq. (21) which gives $M(x;\mu)$ in the wider range $1 < \mu < 3$, $\mu = 2$ excluded. In short, Eq. (26) gives the asymptotic expansion of $M(x;\mu)$ in inverse powers of x for $x \to +\infty$ for all nonintegral values of μ with $\mu > 1$. For $= 2, 3, \cdots$, the expansion in Eq. (26) breaks down since various coefficients $\Gamma(n+2-\mu)$ diverge. This is consistent with our earlier result that $M(x;\mu)$ grows exponentially as $x \to +\infty$ for $\mu = 2, 3, \cdots$. Referring back to Eqs. (9) and (13) it is seen that for large r, the function $f_-(r;\omega)$ drops to zero, with the leading term being proportional to $r^{\mu_- - 1}\exp(-Zr/a_0\mu_-)$. Thus, the range parameter describing this portion of the first-order correction to the wavefunction is $a(\omega) = \mu_- a_0/Z = (1-\omega/\omega_c)^{-1/2} a_0/Z$ which is a continuous, monotonically increasing function of ω. Note that if ω equals one of the $1s$-np resonance frequencies, i.e., $\omega = \omega_n = \omega_c(1-1/n^2)$, $(n = 2, 3, \cdots)$, we have that $a(\omega) = na_0/Z$, which coincides with the range parameter of the eigenfunction of the unperturbed level np. That is, under the action of the electromagnetic field, the unperturbed $1s$ electron cloud with range parameter a_0/Z, is accompanied by a portion f_- with a p-like character and whose amplitude is proportional to the strength of the external field, but which is stretched out to a characteristic distance $(1-\omega/\omega_c)^{-1/2} a_0/Z$. Note that as ω is increased towards the ionization frequency ω_c, the characteristic range of f_- diverges, as is to be expected, in anticipation of the photoelectric process which commences with ω_c.

This picture of the behavior of f_- is complemented by the fact that as ω approaches any of the discrete resonance frequencies ω_n, f_- is closely approximated by $u_{n10}/(\omega - \omega_n)$, up to a constant of proportionality. This familiar result of perturbation theory will also be rederived later in this section by considering the limiting behavior of f_- as $\omega \to \omega_n$.

We should stress that we have obtained the frequency dependence of the range of f_- by a straightforward analysis of the integral representation of $M(x;\mu)$. By con-

trast, to obtain this result starting with the Rayleigh–Schrödinger perturbation series for $\delta\psi$ appears to be a hopeless task. This only serves to point out the enormous advantage of a direct attack on the differential equation for $\delta\psi$, rather than following the universal practice of relying on time-independent perturbation theory and its expansion of $\delta\psi$ in terms of the unperturbed hydrogenic eigenfunctions.

The strong sensitivity of f_- on ω stands in bold contrast to the behavior of f_+. For the function f_+, Eq. (22) is of relevance. Note that the exponential factor e^{xt} is largest at the upper limit of the integral. Thus in order to obtain the leading term of the asymptotic expansion of M for $x \to +\infty$, we need only replace the factor $t^{1-\mu}(1+t)^{1+\mu}$ of the integrand by its value at the upper limit $t = \frac{1}{2}(1-\mu)$. Hence we have

$$M(x;\mu) \underset{x\to+\infty}{\sim} \left(\frac{Z^3}{\pi a_0^3}\right)^{1/2} \left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{\omega_c}{\omega}\,\frac{e^{(1-\mu)x/2}}{x}, \tag{27}$$

and thus

$$f_+(\mathbf{r};\omega) \underset{r\to\infty}{\sim} \left(\frac{Z^3}{\pi a_0^3}\right)^{1/2} \left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{\omega_c}{\omega}\, e^{-Zr/a_0}\cos\theta. \tag{28}$$

That is, the range parameter for f_+ is frequency independent and equals a_0/Z.

D. Static limit

For the static limit, $\omega \to 0$, a closed-form expression for the first-order wavefunction $\delta\psi$ has been known for some years.[2] It is given by

$$\delta\psi(\mathbf{r}) = -\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0a_0}{2Z\hbar\omega_c}\right) e^{-Zr/a_0} \times\left[\frac{Zr}{a_0}+\frac{1}{2}\left(\frac{Zr}{a_0}\right)^2\right]\cos\theta. \tag{29}$$

As we now show, this result can be confirmed by carefully allowing ω to decrease to zero in Eqs. (20) and (22). Using Eqs. (8), (9), and (14) and making the change of variable in Eqs. (20) and (22), $t = \frac{1}{2}(1-\mu)(1-s)$, one easily finds

$$f_\pm(\mathbf{r};\omega) = \left(\frac{eF_0a_0}{Z\hbar\omega}\right)\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{Zr}{a_0}\right) e^{-Zr/a_0} H(r;\mu_\pm)\cos\theta, \tag{30}$$

where

$$H(r;\mu) = \frac{\mu}{1+\mu}\int_0^1 ds\,(1-s)^{1-\mu}\left[1+\left(\frac{1-\mu}{1+\mu}\right)s\right]^{1+\mu} \times\exp\left[-\frac{Zr}{a_0}\left(\frac{1}{\mu}-1\right)s\right]. \tag{31}$$

To first order in ω, $\mu_\pm = 1 \mp \omega/2\omega_c$ and

$$H(r;\mu) \simeq H(r;1) + (\mu-1)\frac{\partial}{\partial\mu}H(r;\mu)\Big|_{\mu=1} = \frac{1}{2}+\frac{1}{2}\left(1+\frac{rZ}{2a_0}\right)\cdot(\mu-1). \tag{32}$$

Substitution of these results in Eq. (5) yields the desired result, Eq. (29).

E. Resonances

We now present a brief discussion of the behavior of of $f_-(\mathbf{r};\omega)$ for values of ω in the vicinity of the $1s$–np resonance frequencies $\omega_n = \omega_c(1-1/n^2)$. As an example we consider frequencies ω where $\omega \simeq \omega_2 = \frac{3}{4}\omega_c$, or equivalently $\mu \approx 2$. Referring to Eq. (25), one notes that the part of $M(x;\mu)$ which is singular as $\mu \to 2$ is given by the $n=0$ term of the infinite series:

$$M^{\text{sing}}(x;\mu) = \frac{16}{81}\left(\frac{Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{eF_0a_0}{Z\hbar\omega}\right)\frac{1}{2-\mu}, \tag{33}$$

where we have used the fact that $F(4,1;1;-\frac{1}{2}) = \frac{16}{81}$. Using Eqs. (5), (9), (13), and (33) one finds directly that for $\omega \approx \frac{3}{4}\omega_c$ the singular part of the first-order wavefunction is given by

$$\delta\psi^{\text{sing}}(\mathbf{r},t) = \frac{64\sqrt{2}}{243}\left(\frac{eF_0a_0}{Z\hbar\omega_c}\right)\cdot\frac{u_{210}(\mathbf{r})}{(\omega/\omega_c)-\frac{3}{4}}\, e^{-i\omega t}. \tag{34}$$

where

$$u_{210}(\mathbf{r}) = \frac{1}{8}\left(\frac{2Z^3}{\pi a_0^3}\right)^{1/2}\left(\frac{Zr}{a_0}\right)e^{-Zr/2a_0}\cos\theta \tag{35}$$

is the unperturbed, normalized $2p$, $m=0$ hydrogenic eigenstate. The expression in Eq. (34), of course, coincides with the singular term of standard, first-order time-dependent perturbation theory. This result can be extended for angular frequencies ω in the vicinity of all the $1s$–np resonance frequencies. In fact, the standard perturbation series for $\delta\psi$ possesses the advantage that it displays the explicit behavior of $\delta\psi$ in the immediate vicinity of the resonance frequencies. Not too close to the resonance frequencies the contributions to $\delta\psi$ of all the unperturbed bound states and continuum states is important and the total contribution is given in closed form by the integral representations which we have derived for $f_\pm(\mathbf{r};\omega)$.

F. Numerical results

We close this section with a summary of our results for $\delta\psi(\mathbf{r},t) = e^{i\omega t}f_+ - e^{-i\omega t}f_-$ obtained by numerical integration of the function $M(x;\mu)$ which enters, via Eqs. (9) and (13). We re-express $\delta\psi$ in terms of its real and imaginary parts as

$$\delta\psi(\mathbf{r},t) = [\cos\omega t\,\phi_-(r;\omega) + i\sin\omega t\,\phi_+(r;\omega)]\cos\theta, \tag{36a}$$

where

$$\phi_\pm(r;\omega) = F(2Zr/a_0\mu_+) \pm F(2Zr/a_0\mu_-). \tag{36b}$$

In Figs. 1 and 2 are shown graphs of $\phi_-(r;\omega)$ as a function of Zr/a_0 for several frequencies, from the static limit through the $1s$–$2p$, $1s$–$3p$, and towards the $1s$–$4p$ resonances. We have chosen to graph ϕ_-, for it alone arises in the calculation of the frequency-dependent polarizability $\alpha(\omega)$ given in Sec. IV. The following general features should be noted. Denoting the $1s$–np resonance frequencies by $\omega_n = \omega_c(1-1/n^2)$, we note that ϕ_- has no nodes for $r>0$ when $0<\omega<\omega_2$, one node for $\omega_2<\omega<\omega_3$, two nodes for $\omega_3<\omega<\omega_4$, etc. These nodes move towards larger r as ω is increased. This behavior is consistent with the result we derived earlier that the range parameter of f_-, given by $a(\omega)$

$=(1-\omega/\omega_c)^{-1/2}a_0/Z$, is a monotonically increasing function of ω. In Fig. 1 we also show the contribution of the $2p$ and $3p$ terms of the perturbation series for ϕ_-. One might perhaps have argued that the contribution of these states to $\delta\psi$ should be very close to the exact solution. For large values of rZ/a_0 the approximate perturbation contribution is numerically close to the exact value of ϕ_-. However, this glosses over the fact that the former contribution cannot be characterized as decreasing exponentially with the range parameter $\alpha(\omega)$, which is the property of the exact solution. For smaller values of rZ/a_0 the approximate perturbation contribution differs very significantly from the exact solution. The large amplitude and change of sign of ϕ_- for two frequencies lying close to and straddling any resonance frequency is to be expected and does not warrant any special comment.

In the following section we employ our results for $\delta\psi$ to calculate $\alpha(\omega)$, the frequency-dependent polarizability.

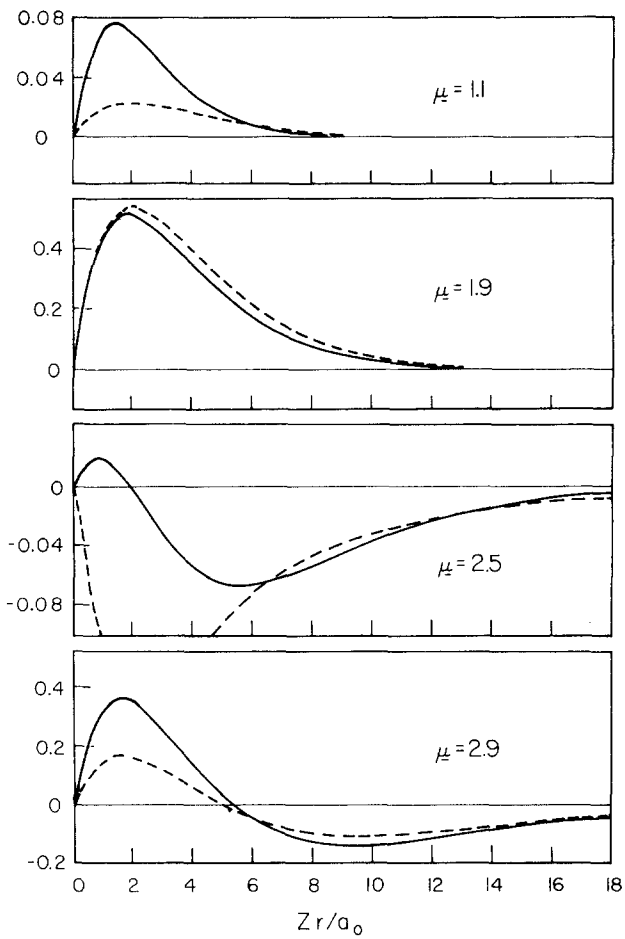


FIG. 1. The evolution of the exact, first-order correction to the wavefunction of the hydrogenic atom as the frequency of the external electromagnetic field sweeps from the static limit ($\mu_- = 1$) through the $1s$-$2p$ ($\mu_- = 2$), and towards the $1s$-$3p$ ($\mu_- = 3$), atomic resonances. The solid curve represents the radial function ϕ_-, defined by Eq. (36), divided by the constant $-8\pi(eF_0a_0/Z\hbar\omega_c)(Z^3/\pi a_0^3)^{1/2}$. The contribution of the $2p$ and $3p$ terms of the usual perturbation series is represented by the dashed curve.

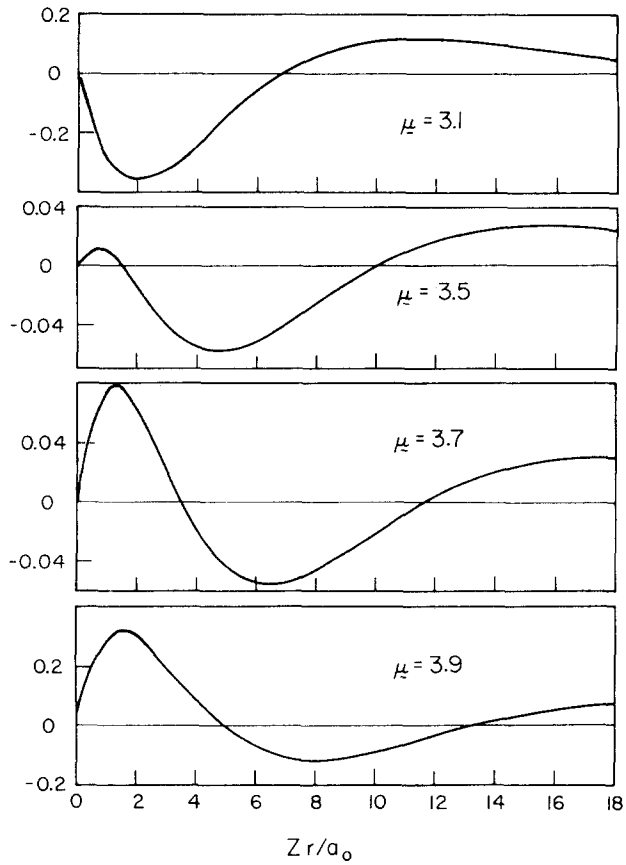


FIG. 2. The radial function $\phi_-(r;\omega)$, defined by Eq. (36), divided by the constant $-8\pi(eF_0a_0/Z\hbar\omega_c)(Z^3/\pi a_0^3)^{1/2}$, for frequencies between the $1s$-$3p$ ($\mu_- = 3$), and $1s$-$4p$ ($\mu_- = 4$), resonances.

IV. FREQUENCY-DEPENDENT POLARIZABILITY

The dipole moment induced in the atom by the electromagnetic field described by Eq. (1) is given by

$$\mathbf{p}(t;\omega) = \int d^3\mathbf{r}\,\delta\rho(\mathbf{r},t;\omega), \tag{37}$$

where $\delta\rho = -e(|\Psi|^2 - u_{100}^2)$ is the field induced change in the electron probability density. To first order in the amplitude F_0 of the external field

$$\delta\rho(\mathbf{r},t;\omega) = -2eu_{100}(r)\phi_-(r;\omega)\cos\theta\cos\omega t, \tag{38}$$

where we have used Eqs. (2), (5), (9), and (36), and the fact that F is a *real* function of its variables. Substituting Eq. (38) for $\delta\rho$ in Eq. (37), the induced dipole moment is given by

$$\mathbf{p}(t;\omega) = -\frac{8\pi}{3}e(\cos\omega t)\hat{\mathbf{z}}\int_0^\infty dr\,r^3u_{100}(r) \times\left[F\left(\frac{2Zr}{a_0\mu_+};\mu_+\right) - F\left(\frac{2Zr}{a_0\mu_-};\mu_-\right)\right]. \tag{39}$$

Note that the induced dipole moment is proportional to the electric field vector with no phase lag. This reflects the fact that, to first order in F_0, as long as ω is less than the ionization frequency ω_c (and does not equal one of the $1s$-np resonance frequencies ω_n), the electron does not absorb energy from the electromagnetic field.

It is convenient to write the polarizability, defined by $\mathbf{p}=\hat{z}\alpha(\omega)F_0\cos\omega t$, as the sum of the separate contributions of $F(x;\mu_+)$ and $F(x;\mu_-)$,

$$\alpha(\omega)=\alpha(\mu_+)+\alpha(\mu_-). \tag{40}$$

We first evaluate $\alpha(\mu_-)$ supposing, for the moment, that $1<\mu<2$. Employing Eqs. (13) and (20) and performing the integration over r one finds

$$\alpha(\mu)=32\left(\frac{a_0^3}{Z^4}\right)\frac{\omega_c}{\omega}\frac{\mu^6}{(\mu-1)^{2-\mu}(\mu+1)^{2+\mu}}\times\int_0^{(\mu-1)/2}dt\frac{t^{1-\mu}(1+t)^{1+\mu}}{[t+\frac{1}{2}(\mu+1)]^5}. \tag{41}$$

We now show that the integral in Eq. (41) can be expressed in terms of the hypergeometric function. Making the change of variables $t=(\mu-1)(\mu+1)^{-1}y[1-y(\mu-1)/(\mu+1)]^{-1}$, we have

$$\alpha(\mu)=2^{10}\left(\frac{a_0^3}{Z^4}\right)\frac{\mu^6}{(\mu+1)^9}\left(\frac{\omega_c}{\omega}\right)\times\int_0^1 dy\,y^{1-\mu}\left(1-\frac{\mu-1}{\mu+1}y\right)\Big/\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]^5. \tag{42}$$

Now with the rearrangement

$$1-\left(\frac{\mu-1}{\mu+1}\right)y=\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]-\frac{2(\mu-1)}{(\mu+1)^2}y,$$

the integral in Eq. (42) can be rewritten successively as

$$\int_0^1 dy\left\{y^{1-\mu}\left(1-\frac{\mu-1}{\mu+1}y\right)\Big/\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]^5\right\}$$

$$=\int_0^1 dy\,y^{1-\mu}\Big/\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]^4-\frac{2(\mu-1)}{(\mu+1)^2}\int_0^1 dy\,y^{2-\mu}\Big/\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]^5 \tag{43a}$$

$$=-\frac{1}{2}\frac{1}{\mu-1}\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2\right]^{-4}+\frac{\mu}{2(\mu-1)}\int_0^1 dy\,y^{1-\mu}\Big/\left[1-\left(\frac{\mu-1}{\mu+1}\right)^2y\right]^4. \tag{43b}$$

Equation (43b) is obtained from Eq. (43a) by transforming the second integral by an integration by parts. The hypergeometric function has the following well-known integral representation[22] (AS, p. 558, Eq. 15.3.1):

$$F(a,b;c;Z)=\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 dt\,t^{b-1}(1-t)^{c-b-1}(1-tZ)^{-a} \tag{44}$$

as long as $\mathrm{Re}c>\mathrm{Re}b>0$. Thus in accordance with our earlier restriction, $1<\mu<2$, it follows that the integral on the rhs of Eq. (43b) is given by $(2-\mu)^{-1}F(4,2-\mu;3-\mu;(\mu-1)^2/(\mu+1)^2)$ and the portion $\alpha(\mu_-)$ of the polarizability is given by

$$\alpha(\mu)=2\left(\frac{a_0^3}{Z^4}\right)\left(\frac{\omega_c}{\omega}\right)^2\left[\frac{256\mu^5}{(\mu+1)^8}\frac{1}{2-\mu}\times F\left(4,2-\mu;3-\mu;\left(\frac{\mu-1}{\mu+1}\right)^2\right)-1\right], \tag{45}$$

with μ replaced by μ_-. A similar analysis shows that Eq. (45), with μ replaced by μ_+, also gives $\alpha(\mu_+)$. Finally, the polarizability $\alpha(\omega)$ is given by Eq. (40).

Inspection of Eq. (45) shows that $\alpha(\mu)$ has simple poles for $\mu_-=2$, because of the factor $(2-\mu)^{-1}$, and for $\mu_-=3,4,\cdots$, because of the simple poles of F. For all other values of μ_- the quantity $\alpha(\mu_-)$ is analytic. Thus Eq. (45) constitutes the expression for $\alpha(\mu_-)$ for all nonintegral values of μ_-, even though it was derived using the expression of Eq. (20) which applies only for the limited range $1<\mu_-<2$. On the other hand, $\alpha(\mu_+)$ is analytic for all values of μ_+ in its range $0<\mu_+<1$, i.e., for all frequencies ω from the static limit to the infinite frequency limit. Thus the total polarizability $\alpha(\omega)=\alpha(\mu_+)+\alpha(\mu_-)$ has simple poles at all of the $1s-np$, $n=2,3,\cdots$, resonance frequencies, but is otherwise analytic.

The above results, Eqs. (40) and (45), for the polarizability were first derived by Podolsky,[3] although he expressed his answer as the sum of two infinite series, which in fact are the standard power series expansions of the hypergeometric function, involving the variables μ_- and μ_+.

The static limit of $\alpha(\omega)$ can be obtained by carefully evaluating Eq. (45) as $\omega\to 0$. The same result is most simply obtained by directly evaluating the dipole moment of Eq. (37) using the static limit wavefunction of Eq. (29). The final result is $\alpha(0)=\frac{9}{2}(a_0^3/Z^4)$. This result was first obtained in 1926 independently by Wentzel,[23] Waller,[24] and Epstein[25] in their treatment of the second-order Stark effect for atomic hydrogen.

As seen above, $\alpha(\omega)$ possesses simple poles for ω equal to any of the $1s-np$ resonance frequencies ω_n. In Appendix B we show that

$$\alpha(\omega)=\bar{\alpha}(\omega)+\alpha_s(\omega), \tag{46a}$$

where

$$\alpha_s(\omega)=-\tfrac{256}{3}\pi\frac{a_0^3}{Z^4}\left(\frac{\omega_c}{\omega}\right)^5\left(\frac{\mu_--1}{\mu_-+1}\right)^{2\mu_-}\cot\pi\mu_-, \tag{46b}$$

and $\bar{\alpha}(\omega)$ is real and finite for $0<\omega\leq\omega_c$. The explicit form for $\bar{\alpha}(\omega)$ is given in Eqs. (B6) and (B8). That is, Eq. (46) constitutes the decomposition of the polarizability as the sum of a regular function, $\bar{\alpha}(\omega)$, and a function which has a simple pole when $\omega=\omega_n$, or equivalently when $\mu_-=(1-\omega/\omega_c)^{-1/2}=n(=2,3,\cdots)$. The compact, closed form for the singular term $\alpha_s(\omega)$ should be contrasted with the perturbation theory result for $\alpha(\omega)$ which includes an infinite series of simple poles. This decomposition is particularly useful for values of ω just below ω_c, where the resonance frequencies ω_n become infinitely dense as $\omega\to\omega_c-0$. For very low frequencies, Eq. (46) is no longer useful since both $\bar{\alpha}(\omega)$ and $\alpha_s(\omega)$ possess spurious singularities in the static limit, even though $\alpha(\omega)$ is finite and well behaved for $0\leq\omega<\omega_2=\frac{3}{4}\omega_c$. We shall arbitrarily restrict usage of Eq. (46) to the interval $\frac{5}{9}\leq\omega/\omega_c<1$. The choice of lower limit $\omega/\omega_c=\frac{5}{9}$, is guided by the fact that for this value of ω we have $\alpha_s=0$ and $\bar{\alpha}=\alpha$.

In Table I we have listed values of $\bar{\alpha}(\omega)$ for selected

TABLE I. Total polarizability $\alpha(\omega)$ and partial polarizability $\bar{\alpha}(\omega)$ [see Eq. (46)], in units of a_0^3/Z^4, for $0 \le \omega/\omega_c < 1$.

ω/ω_c	$\alpha(\omega)$	ω/ω_c	$\bar{\alpha}(\omega)$
0	4.5	0.555 ···	8.729165
0.02	4.502660	0.56	7.939421
0.04	4.510661	0.57	6.316805
0.06	4.524065	0.58	4.886651
0.08	4.542979	0.59	3.625831
0.10	4.567553	0.60	2.514401
0.12	4.597987	0.61	1.534665
0.14	4.634534	0.62	0.671340
0.16	4.677505	0.63	−0.089003
0.18	4.727276	0.64	−0.758144
0.20	4.784300	0.65	−1.346435
0.22	4.849115	0.66	−1.862984
0.24	4.922358	0.67	−2.315818
0.26	5.004787	0.68	−2.712021
0.28	5.097301	0.69	−3.057854
0.30	5.200968	0.70	−3.358857
0.32	5.317065	0.72	−3.845451
0.34	5.447123	0.74	−4.204808
0.36	5.592992	0.76	−4.463004
0.38	5.756922	0.78	−4.640730
0.40	5.941675	0.80	−4.754458
0.42	6.150673	0.82	−4.817348
0.44	6.388211	0.84	−4.839944
0.46	6.659744	0.86	−4.830714
0.48	6.972315	0.88	−4.796474
0.50	7.335173	0.90	−4.742720
0.52	7.760721	0.92	−4.673887
0.54	8.265994	0.94	−4.593561
0.555 ···	8.729165	0.95	−4.550016
		0.98	−4.409462
		1.00	−4.309921

values of ω/ω_c in the interval $(\frac{5}{9}, 1)$. The values of $\bar{\alpha}(\omega)$ were obtained by calculating $\alpha(\omega)$, employing Eqs. (40), (45), and (46).[26] We have also listed values of α for the lower frequency regime $0 \le \omega/\omega_c \le \frac{5}{9}$. The limiting value $\bar{\alpha}(\omega_c) = -4.3099214 \cdots a_0^3/Z^4$ is derived in Appendix B. A graph of $\bar{\alpha}(\omega)$ versus ω is given in Fig. 3. A simple method for obtaining moderately accurate values of $\alpha(\omega)$, without recourse to extensive computer calculations, is to evaluate the rhs of Eq. (46) using the values of $\bar{\alpha}(\omega)$ listed in Table I. Specifically, by applying standard extrapolation techniques to the entries in Table I, one can obtain moderately accurate values of $\bar{\alpha}(\omega)$ for any angular frequency $\omega < \omega_c$.

Another advantage of the decomposition of Eq. (46) is that one can readily obtain the following result (see Appendix B) for $\alpha''(\omega)$, the imaginary part of $\alpha(\omega)$,

$$\alpha''(\omega) = 512 \frac{\pi}{3} \frac{a_0^3}{Z^4} \omega_c \sum_{n=2}^{\infty} \frac{n^7 (n-1)^{2n-5}}{(n+1)^{2n+5}} \delta(\omega - \omega_n). \tag{47}$$

This result expresses the fact that for $\omega < \omega_c$ the atom absorbs energy from the electromagnetic field only when the frequency of the field coincides with one of the $1s$–np resonance frequencies. The appearance of delta functions reflects the fact that in the present treatment we are excluding the possiblility of spectral line widths.

With the aid of Eq. (47), one can readily obtain the expressions for the oscillator strengths for the Lyman spectral series of a hydrogenic atom

$$g_n = \frac{256}{3} \frac{n^5 (n-1)^{2n-4}}{(n+1)^{2n+4}} \quad (n = 2, 3, \cdots). \tag{48}$$

It is traditional in the theory of optical and x-ray dispersion to simulate atomic scattering of electromagnetic radiation by an assembly of ersatz electric-dipole oscillators whose natural frequencies are identified with the atomic absorption frequencies.[27] The polarizability is related to the total scattering factor of the assembly of oscillators by the formula

$$\alpha(\omega) = -\frac{e^2}{m\omega^2} f(\omega), \tag{49}$$

where m is the electron mass. If there are g_n oscillators with natural frequency ω_n and if each oscillator is assumed to have vanishingly small damping, then $f(\omega)$ is given by[27]

$$f(\omega) = \sum_n \frac{g_n \omega^2}{\omega^2 - \omega_n^2 - i\omega^2 \epsilon}, \tag{50}$$

where ϵ is an arbitrarily small, real positive number. Using the identity $(x - i\epsilon)^{-1} = Px^{-1} + i\pi\delta(x)$, where P denotes the principal value and δ denotes the Dirac delta function, we have

$$\alpha''(\omega) = -\frac{\pi}{2} \frac{e^2}{m} \sum_n \frac{g_n}{\omega_n} \delta(\omega - \omega_n). \tag{51}$$

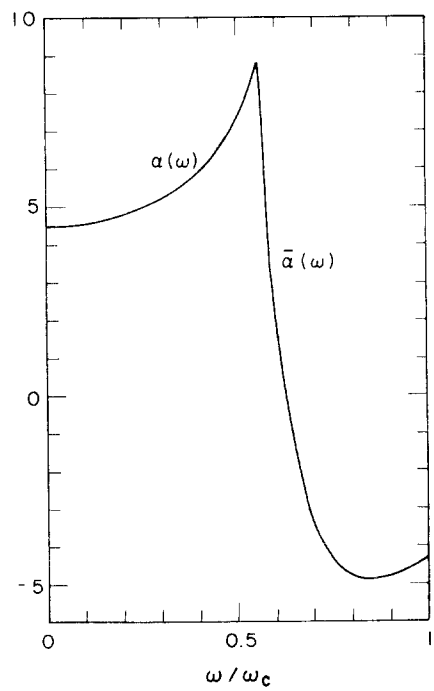


FIG. 3. The total polarizability $\alpha(\omega)$ and partial polarizability $\bar{\alpha}(\omega)$ in units of a_0^3/Z^4, for $0 \le \omega/\omega_c \le 1$. The quantity $\alpha(\omega)$ has simple poles for $\omega/\omega_c = 1 - n^{-2}$ $(n = 2, 3, \ldots)$. With the aid of Eq. (46a), $\alpha(\omega)$ is decomposed as the sum of a finite, well-behaved quantity $\bar{\alpha}(\omega)$, and a compact singular term $\alpha_s(\omega)$ given by Eq. (46b). Usage of Eq. (46) is arbitrarily restricted to the interval $\frac{5}{9} \le \omega/\omega_c < 1$. The choice of lower limit, $\omega/\omega_c = \frac{5}{9}$, is guided by the fact that for this value of ω one has $\alpha_s = 0$, and thus $\bar{\alpha} = \alpha$.

In order to be applicable to the hydrogenic atom, the summation sign in Eq. (51) must be interpreted to include a discrete sum, appropriate to the discrete $1s$–np resonance frequencies, as well as an integral to incorporate all dipole transitions from the $1s$ to continuum states. However, for our present discussion, where ω is restricted to below the ionization frequency ω_c, it is sufficient to suppose that the summation sign in Eq. (51) only includes a sum over discrete frequencies. Comparison of Eqs. (47) and (51) leads to the result Eq. (48) for the oscillator strengths g_n. Note carefully that this expression gives the oscillator strengths for $1s$–np transitions, i.e., for the Lyman spectral series of the hydrogenic atom. The conventional quantum–mechanical prescription for calculating g_n is to use the expression

$$g_n = \frac{2m\omega_n}{\hbar} |\langle n1|z|10\rangle|^2. \tag{52}$$

The values of the matrix elements $\langle n1|z|10\rangle$ and g_n can be found, for example, in the classic text of Bethe and Salpeter.[28]

V. DISCUSSION AND SUMMARY

In this article we have obtained an exact, closed–form solution for the first order correction to the wavefunction of a hydrogenic atom in the presence of an external, linearly polarized electromagnetic wave of angular frequency ω. The atom is assumed to be initially in the ground state and the external field is treated in the dipole approximation. This program has been achieved by employing the method of integral transformations to solve the inhomogeneous differential equation governing the first order wavefunction. Specifically, this wavefunction, given by Eqs. (2) and (5), is expressed in terms of two time–independent functions $f_\pm(\mathbf{r};\omega)$ which, in the dipole approximation, satisfy Eq. (7). It was shown that the radial portion of these two functions can be expressed, via Eqs. (9) and (13), in terms of the solution of an inhomogeneous version, Eq. (14), of the confluent hypergeometric equation.

There are two crucial aspects of Eq. (14) which assured a solution of the form of Eq. (15). The first is the fact that both of the linearly independent solutions of the homogeneous version of Eq. (14), i.e., the confluent hypergeometric equation, can be expressed in the form (15), with an exponential kernel (see p. 505 of Ref. 22). In fact, for the homogeneous solutions, the function $u(t)$ is also given by Eq. (19). Of course, these two solutions correspond to two different choices for the upper and lower limits of integration, α and β. Secondly, the inhomogeneous term of Eq. (14) is itself an exponential, originating from the term $\partial u_{100}/\partial z$ of Eq. (7). For the function $u(t)$, given by Eq. (19), the integral of Eq. (16) vanishes and the latter equation reduces to an algebraic equation, (18), requiring that the so–called bilinear concomitant be equal to the inhomogeneous term. For a different initial, unperturbed state of the hydrogenic atom, say the metastable u_{200} state, the replacement for the right–hand side of (14) will involve powers of x as well as an exponential function. As such, a solution of the form (15) no longer appears to be possible.

If one wishes to go beyond the dipole approximation, as is required for sufficiently high frequencies when the inequality $\omega/\omega_c \ll 2/Z\alpha$ is no longer satisfied, the mathematical difficulties are greatly increased. The relevant equation is Eq. (6). Now the presence of the factor $\exp(\mp i\omega y/c)$ on the right–hand side of Eq. (6) necessitates the series expansion of $f_\pm$ in terms of spherical harmonics. For each partial wave the radial function satisfies a different inhomogeneous differential equation.

The manifold advantages of a closed form, integral representation of the first order wavefunction have been apparent throughout this work. This form has enabled us to extract all essential analytic and numerical properties of the wavefunction. Especially noteworthy is the result that the first order wavefunction decays exponentially for large r with the frequency dependent range parameter $(a_0/Z)(1-\omega/\omega_c)^{-1/2}$. In addition, starting from the first order wavefunction, we have derived the frequency–dependent polarizability $\alpha(\omega)$, as well as a useful decomposition of α as the sum of a single term, embodying the poles of α, and a slowly–varying remainder term.

ACKNOWLEDGMENT

We are greatly indebted to I. Freund for numerous profitable discussions on all aspects of the work reported here.

APPENDIX A

For values of μ in the range $1<\mu<2$, the differential equation of Eq. (14) possesses a solution of the form

$$M(x;\mu) = \int_\alpha^\beta dt\, u(t) e^{-xt}, \tag{A1}$$

where

$$u(t) = ct^{1-\mu}(1+t)^{1+\mu}, \tag{A2}$$

for three possible choices of upper and lower limits α and β. These are: (i) $\alpha=0, \beta=\frac{1}{2}(\mu-1)$; (ii) $\alpha=\frac{1}{2}(\mu-1), \beta=+\infty$; (iii) $\alpha=-1, \beta=\frac{1}{2}(\mu-1)$. As shown in the text, the first of these choices leads to a solution which meets the requisite boundary conditions that $M(0)$ be finite and $M(x)$ cannot grow as rapidly as $e^{x/2}$ for $x\to+\infty$.

We show here that the remaining two choices of α and β are unacceptable since, although these lead to *bona fide* solutions of Eq. (14), these solutions fail to satisfy the above–mentioned boundary conditions. Consider first the solution

$$M(x;\mu) = c\int_{(\mu-1)/2}^\infty dt\, t^{1-\mu}(1+t)^{1+\mu}e^{-xt}. \tag{A3}$$

The boundary condition for large, positive x is met since the integral vanishes in this limit. However, the boundary condition at $x=0$ is not satisfied since the integral in Eq. (A3) diverges.

For the last choice of α and β the solution of Eq. (14) is of the form

$$M(x;\mu) = c\int_{-1}^{(\mu-1)/2} dt\, t^{1-\mu}(1+t)^{1+\mu}e^{-xt}. \tag{A4}$$

In this case $M(0;\mu)$ remains finite if μ lies within the interval $1<\mu<2$. However, for $x\to+\infty$ the integral di-

verges due to the exponential increase of the integrand for negative values of t.

APPENDIX B

In Sec. IV A we rewrote the frequency-dependent polarizability $\alpha(\omega)$ as a sum of two terms, [see Eq. (46)], where the first term $\bar{\alpha}(\omega)$ is real and regular throughout the frequency interval $0<\omega<\omega_c$, and the second term, $\alpha_s(\omega)$, possesses simple poles when ω equals one of the $1s$–np resonance frequencies ω_n. In this Appendix we derive the explicit form of $\bar{\alpha}(\omega)$ and show further that $\bar{\alpha}(\omega_c) = -4.30992137\cdots a_0^3/Z^4$. In addition, we derive Eq. (47).

The polarizability is given in Eqs. (40) and (45). The portion $\alpha(\mu_+)$ is well behaved for all frequencies $0<\omega<\omega_c$ and thus we shall leave its form intact. The singularities of $\alpha(\omega)$ are all simple poles, corresponding to the simple poles of the expression $\alpha(\mu)$ of Eq. (45) for $\mu_- = 2$, because of the factor $(2-\mu)^{-1}$, and for $\mu_- = 3,4,\cdots$, because of the poles of the hypergeometric function. Now the latter function can be rewritten using the relation (AS, p. 560, Eq. 15.3.12)

$$F(a,b;a+b-m;x) = \frac{\Gamma(m)\Gamma(a+b-m)}{\Gamma(a)\Gamma(b)}(1-x)^{-m}\sum_{n=0}^{m-1}\frac{(a-m)_n(b-m)_n(1-x)^n}{n!(1-m)_n}$$
$$-\frac{(-1)^m\Gamma(a+b-m)}{\Gamma(a-m)\Gamma(b-m)}\sum_{k=0}^{\infty}\frac{(a)_k(b)_k(1-x)^k}{k!(k+m)!}\left[\ln(1-x)-\psi(k+1)-\psi(k+m+1)+\psi(a+k)+\psi(b+k)\right], \quad \text{(B1)}$$

which applies if $|1-x|<1$, $|\arg(1-x)|<\pi$. Here m denotes any positive integer, $(a)_k = a(a+1)\cdots(a+k-1)$, $(a)_0 = 1$, and $\psi(y) = \Gamma'(y)/\Gamma(y)$ is the logarithmic derivative of the gamma function, or so-called psi function. It will be recalled that $\psi(y)$ possesses simple poles for $y = 0, -1, -2, \cdots$, and also that it obeys the reflection formula (AS, p. 259, Eq. 6.3.7)

$$\psi(1-y) = \psi(y) + \pi\cot\pi y \quad \text{(B2)}$$

and the recurrence formula (AS, p. 258, Eq. 6.3.5)

$$\psi(1+y) = \psi(y) + 1/y. \quad \text{(B3)}$$

In the case at hand, $a = 2-\mu$, $b = 4$, $m = 3$, and $x = (\mu-1)^2/(\mu+1)^2$. It is easily shown that for the parameters relevant to $\alpha(\mu_-)$, the first term on the right-hand side of Eq. (B1) equals $\frac{1}{192}(2-\mu)(\mu+1)^6\mu^{-3}[1+2\mu(1+\mu)^{-1}+8\mu^3(1+\mu)^{-3}]$. Examining Eq. (B1) it will be seen that the poles of F originate solely in the terms $\psi(a+k) = \psi(2+k-\mu)$. Furthermore, using Eqs. (B2) and (B3) we can write $\psi(2+k-\mu)$ as

$$\psi(2+k-\mu) = \cot\pi\mu + \psi(\mu-1) + \frac{1}{2-\mu} + \frac{1}{3-\mu} + \cdots + \frac{1}{k+1-\mu}, \quad \text{(B4)}$$

and it will be seen that the singular terms of F originate solely from the term $\pi\cot\pi\mu$ of $\psi(2+k-\mu)$. [The singular terms $(2-\mu)^{-1},\ldots,(k+1-\mu)^{-1}$ in Eq. (B4) are cancelled by the factor $(a)_k = (2-\mu)_k$ appearing in Eq. (B1).] These singularities of F are simple poles and they occur for $\mu = 3,4,\cdots$. Denoting by F_s that portion of F which is singular for $\mu = 3,4,\cdots$, we have

$$F_s\left(2-\mu,4;3-\mu;\left(\frac{\mu-1}{\mu+1}\right)^2\right) = \frac{1}{3!}(\mu-2)\mu(\mu^2-1)\pi\cot\pi\mu\sum_{k=0}^{\infty}\frac{(2-\mu)_k}{k!}\left(\frac{4\mu}{(\mu+1)^2}\right)^k$$
$$= \frac{1}{3!}(\mu-2)\mu(\mu^2-1)\pi(\cot\pi\mu)\left(\frac{\mu+1}{\mu-1}\right)^{2(2-\mu)}. \quad \text{(B5)}$$

The remaining, analytic, portion of F will be denoted by by $\bar{F}$, and it is given by

$$\bar{F}\left(2-\mu,4;3-\mu;\left(\frac{\mu-1}{\mu+1}\right)^2\right) = \frac{1}{192\mu^3}(2-\mu)(\mu+1)^6\left(1+\frac{2\mu}{1+\mu}+\frac{8\mu^3}{(1+\mu)^3}\right)$$
$$+\frac{1}{3!}(\mu-2)\mu(\mu^2-1)\Big\{[\gamma+\ln 4+\ln\mu-2\ln(\mu+1)]$$
$$\times\left(\frac{\mu+1}{\mu-1}\right)^{2(2-\mu)}+\frac{1}{1!}\frac{4\mu}{(\mu+1)^2}(2-\mu)\left(-\frac{1}{1}+\frac{1}{2-\mu}\right)$$
$$+\frac{1}{2!}\left(\frac{4\mu}{(\mu+1)^2}\right)^2(2-\mu)(3-\mu)$$
$$\times\left(-\frac{1}{1}-\frac{1}{2}+\frac{1}{2-\mu}+\frac{1}{3-\mu}\right)+\cdots\Big\}, \quad \text{(B6)}$$

where we have used the fact that $\psi(k+1) = -\gamma+1+\frac{1}{2}+\cdots+1/k$, and γ is Euler's constant. Employing Eq. (B5) we find that the singular part of $\alpha(\omega)$, denoted by $\alpha_s(\omega)$, is given by

$$\alpha_s(\omega) = -\tfrac{256}{3}\pi\frac{a_0^3}{Z^4}\left(\frac{\omega_c}{\omega}\right)^5\left(\frac{\mu_--1}{\mu_-+1}\right)^{2\mu_-}\cot\pi\mu_-. \quad \text{(B7)}$$

The well-behaved portion, $\bar{\alpha}(\omega)$, of $\alpha(\omega)$ is given by

$$\bar{\alpha}(\omega) = \alpha(\mu_+) + \frac{2a_0^3}{Z^4}\left(\frac{\omega_c}{\omega}\right)^2\left[\frac{256\mu_-^5}{(\mu_-+1)^8}\frac{1}{2-\mu_-}\right.$$
$$\left.\times\bar{F}\left(4,2-\mu_-;3-\mu_-;\left(\frac{\mu_--1}{\mu_-+1}\right)^2\right)-1\right]. \quad \text{(B8)}$$

Selected values of $\bar{\alpha}(\omega)$ are listed in Table I.

We now turn to the evaluation of $\bar{\alpha}(\omega)$ for the frequency $\omega = \omega_c$, which corresponds to $\mu_+ = 2^{-1/2}$ and $\mu_- \to +\infty$. The value of $\alpha(\mu_+ = 2^{-1/2})$ is readily established using Eq. (45). The value of the hypergeometric function $F(4,2-2^{-1/2};3-2^{-1/2};(2^{1/2}-1)^4) = 1.069957797\cdots$ is simply obtained using the usual power series expansion

$$F(a,b;c;x)=\sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{n!(c)_n}x^n, \tag{B9}$$

which proves to converge very rapidly for the requisite argument. The final result is

$$\alpha(\mu_+=2^{-1/2})=-0.961483353\cdots a_0^3/Z^4. \tag{B10}$$

Starting from Eq. (B1), and using the property (AS, 259, 6.3.18), $\psi(x)\to\ln x+O(1/x)$, the limiting value of $\bar{F}$ as $\mu_-\to\infty$ is found to be

$$\bar{F}\left(2-\mu,4;3-\mu;\left(\frac{\mu-1}{\mu+1}\right)^2\right)\underset{\mu\to\infty}{\longrightarrow}\frac{\mu^4}{3!}\left\{-\tfrac{11}{32}+(\gamma+\ln 4)e^{-4}+\left[\frac{4}{1!}-\frac{4^2}{2!}(1+\tfrac{1}{2})+\frac{4^3}{3!}(1+\tfrac{1}{2}+\tfrac{1}{3})-\cdots\right]+O\left(\frac{1}{\mu}\right)\right\}. \tag{B11}$$

Now the series in square brackets can be rewritten as follows:

$$\frac{4}{1!}-\frac{4^2}{2!}\left(1+\tfrac{1}{2}\right)+\frac{4^3}{3!}(1+\tfrac{1}{2}+\tfrac{1}{3})-\cdots=e^{-4}\sum_{n=0}^{\infty}\frac{4^n}{n\cdot n!}, \tag{B12}$$

as can be seen upon writing $e^{-4}=\sum_{n=0}^{\infty}(-1)^n4^n/n!$ and multiplying out the resulting right-hand side of this equation. Furthermore, the exponential integral function $Ei(x)$, defined as (AS, 230, 5.1.40)

$$Ei(x)=\gamma+\ln x+\int_0^x dt\,t^{-1}(e^t-1)\quad(x>0), \tag{B13}$$

has the power series expansion (AS, 229, 5.1.10)

$$Ei(x)=\gamma+\ln x+\sum_{n=1}^{\infty}\frac{x^n}{n\cdot n!}\quad(x>0). \tag{B14}$$

Collecting these results we arrive at the conclusion

$$\bar{F}\left(2-\mu,4;3-\mu;\left(\frac{\mu-1}{\mu+1}\right)^2\right)\underset{\mu\to\infty}{\longrightarrow}\frac{\mu^4}{3!}\left[-\tfrac{11}{32}+e^{-4}Ei(4)\right] \tag{B15}$$

and thus

$$\begin{aligned}\bar{\alpha}(\omega=\omega_c)&=\alpha(\mu_+=2^{-1/2})+\tfrac{2}{3}[41-128Ei(4)]a_0^3/Z^4\\&=-4.30992137\cdots a_0^3/Z^4.\end{aligned} \tag{B16}$$

We now proceed to derive Eq. (47). For this purpose we recall that $\alpha_s(\omega)$ is real and well behaved for all values of ω in the interval $(0,\omega_c)$ which are unequal to the $1s$-np resonance frequencies. To study the immediate vicinity of the singularities of α_s, we replace the frequency ω by $\omega(1-i\epsilon)$, where ϵ is a real, infinitesimal, positive number. Equivalently, μ_- is replaced by $\mu_- - i\epsilon$. It then follows that for values of μ_- in the immediate vicinity of $\mu_-=n$, the singular part of the quantity $\pi\cot\pi(\mu-i\epsilon)$ is given by $(\mu_- -n-i\epsilon)^{-1}=P(\mu_- -n)^{-1}+i\pi\delta(\mu_- -n)$, where P denotes the principal value, and δ denotes the Dirac delta function. That is, in the immediate vicinity of $\mu_-=n$, $\alpha_s(\omega-i\omega\epsilon)$ has an imaginary part in the limit $\epsilon\to 0+$ and it is proportional to $\delta(\mu_- -n)$. This argument can be made for any integer value of n. It follows, then, that for frequencies $0<\omega<\omega_c$,

$$\alpha''(\omega)=\tfrac{256}{3}\pi\frac{a_0^3}{Z^4}\sum_{n=2}^{\infty}\left(\frac{\omega_c}{\omega}\right)^5\left(\frac{n-1}{n+1}\right)^{2n}\delta(\mu_- -n), \tag{B17}$$

where α'' denotes the imaginary part of $\alpha(\omega)$. Now $\delta(\mu_- -n)=(2\omega_c/n^3)\delta(\omega-\omega_n)$, where $\omega_n=\omega_c(1-1/n^2)$, and thus one finds directly that $\alpha''(\omega)$ can be written in the form Eq. (47).

[1] A preliminary account of a portion of the results contained in this article appears in: M. Luban, B. Nudler, and I. Freund, Phys. Lett. A **47**, 447 (1974).

[2] See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed., Sec. 33. Credit for deriving the wavefunction in this case is given to M. Kotani, Ref. 6.

[3] B. Podolsky, Proc. Natl. Acad. Sci. U. S. **14**, 253 (1928).

[4] M. Kar[illegible]s and H. J. Kolker, J. Chem. Phys. **39**, 1493 (1963).

[5] M. Gav[illegible], Phys. Rev. **163**, 147 (1967).

[6] M. Kot[illegible] *Quantum Mechanics* (Yuwanami, Tokyo, 1951), Vol. I, p. 127.

[7] A. Dalgarno and J. T. Lewis, Proc. Roy. Soc. London **A233**, 70 (1955).

[8] C. Schwartz, Ann. Phys. (N.Y.) **6**, 156 (1959).

[9] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I, pp. 466–8; R. V. Churchill, *Operational Mathematics* (McGraw-Hill, New York, 1958), 2nd ed., Sec. 17; G. Doetsch, *Introduction to the Theory and Application of the Laplace Transformation* (Springer, New York, 1974), Sec. 38.

[10] L. Hostler and R. H. Pratt, Phys. Rev. Lett. **10**, 469 (1963).

[11] L. P. Rapoport and B. A. Zon, Phys. Lett. A **26**, 564 (1968); Zh. Eksp. Teor. Fiz. Pis'ma **7**, 70 (1968) [JETP Lett. **7**, 52 (1968)].

[12] B. A. Zon, N. L. Manakov, and L. P. Rapoport, Sov. Phys. JETP **28**, 480 (1969).

[13] L. P. Rapoport, B. A. Zon, and N. L. Manakov, Sov. Phys. JETP **29**, 220 (1969).

[14] S. Klarsfeld, Phys. Lett. A **30**, 382 (1969); Nuovo Cimento Lett. **1**, 682 (1969); **2**, 548 (1969); **3**, 395 (1970); Phys. Rev. A **6**, 506 (1972).

[15] M. Gavrila and A. Costescu, Phys. Rev. A **2**, 1752 (1970).

[16] M. Gavrila, Phys. Rev. A **6**, 1348, 1360 (1972).

[17] A. Costescu and M. Gavrila, Rev. Roum. Phys. **18**, 493 (1973).

[18] M. Gavrila, Rev. Roum. Phys. **19**, 473 (1974).

[19] M. Gavrila and M. N. Tugulea, Rev. Roum. Phys. **20**, 209 (1975).

[20] A. Sommerfeld and G. Schur, Ann. Phys. **4**, 409 (1930).

[21] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), 4th ed., p. 284, Eq. 3.194.

[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D. C., 1964). We shall use the code AS, followed by the relevant page and equation number to identify the source of the quoted formula.

[23] G. Wentzel, Z. Phys. **38**, 518 (1926).

[24] I. Waller, Z. Phys. **38**, 635 (1926).

[25] P. S. Epstein, Phys. Rev. **28**, 695 (1926).

[26] We remark that our calculated values of $\alpha(\omega)$ are in agreement with those published by Gavrila.[5] Strictly speaking, Gavrila has tabulated a quantity $M(\omega)$ which is proportional to the Kramers–Heisenberg matrix element describing the elastic scattering of photons by a hydrogen atom in the dipole approximation. The polarizability is related to Gavrila's function by the relation $\alpha(\omega)=-4(\omega_c/\omega)^2(a_0^3/Z^4)M(\omega)$.

[27] See, for example, R. W. James, *Optical Principles of the Diffraction of X-Rays* (Cornell U. P., Ithaca, New York, 1965), Chap. IV.

[28] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One and Two Electron Atoms* (Springer-Verlag, Berlin, 1957), p. 263.

Classes of solvable Volterra integral equations

Marshall Luban and Moshe Deutsch

Department of Physics, Bar-Ilan University, Ramat-Gan, Israel
(Received 21 March 1977)

We display several classes of Volterra equations which can be transformed to the so-called faltung form and are thus solvable in terms of Laplace transforms.

In this note we describe a method for obtaining the solution of several classes of Volterra integral equations. We begin by considering Volterra equations of the first kind, of the form

$$\int_0^x dy\, K((x^2-y^2)^{1/2})\phi(y)=g(x), \tag{1}$$

where the kernel K and the function g are given, and $\phi(x)$ is to be determined. An equivalent form of this equation is

$$\int_0^{\pi/2} d\theta\, f(x\sin\theta)\phi(x\cos\theta)=g(x), \tag{2}$$

where $f(u)=uK(u)$.

A special case of Eq. (2), namely $g(x)\equiv\pi/2$, is of central importance in correcting measured intensities in conventional small-angle x-ray scattering experiments employing long, narrow slits.[1-3] Prior to the present work the only known solution[2] of that equation was for the special case $f(u)=\exp(-u^2)$.

The key step in solving Eq. (1) consists of rewriting the equation so that it assumes the so-called faltung form,[4] which is solvable in terms of Laplace transforms. This is achieved by defining a triplet of functions K_2, g_2, ϕ_2, by the equations

$$K_2(u^2)=K(u), \tag{3}$$

$$g_2(u^2)=g(u), \tag{4}$$

$$\phi_2(u^2)=\phi(u)/u, \tag{5}$$

as well as a pair of variables $v=x^2$, $w=y^2$. Substituting in Eq. (1), one obtains

$$\int_0^v dw\, K_2(w)\phi_2(v-w)=2g_2(v), \tag{6}$$

which has the desired faltung form. Letting $\hat{K}_2$, $\hat{g}_2$, and $\hat{\phi}_2$ denote the Laplace transforms of K_2, g_2, and ϕ_2, respectively, Eq. (6) is then equivalent to the algebraic equation

$$\hat{K}_2(s)\hat{\phi}_2(s)=2\hat{g}_2(s). \tag{7}$$

In terms of the original functions we have

$$\hat{K}_2(s)=\int_0^\infty dv\, e^{-sv}k_2(v)$$
$$=2\int_0^\infty dx\, x\exp(-sx^2)K(x), \tag{8}$$

$$\hat{g}_2(s)=2\int_0^\infty dx\, x\exp(-sx^2)g(x), \tag{9}$$

and the solution of Eq. (1) is given by

$$\phi(x)=2xL_{x^2}^{-1}\{\hat{g}_2(s)/\hat{K}_2(s)\}$$
$$=x\,\frac{1}{\pi i}\int_{c-i\infty}^{c+i\infty} ds\exp(sx^2)\hat{g}_2(s)/\hat{K}_2(s). \tag{10}$$

The symbol $L_{x^2}^{-1}\{\cdots\}$ denotes the usual inverse Laplace transform[5] of the quantity in braces, but evaluated for the argument x^2.

The above method can also be used to solve Volterra equations of the third kind of the form

$$\int_0^x dy\, K((x^2-y^2)^{1/2})\phi(y)=g(x)+\tfrac{1}{2}\mu\,\frac{\phi(x)}{x}, \tag{11}$$

where μ is a constant. The final result is

$$\phi(x)=2xL_{x^2}^{-1}\{\hat{g}_2(s)/[\hat{K}_2(s)-\mu]\}. \tag{12}$$

We are, however, unaware of any physical realization of Eq. (11).

One can readily generalize the above method of solution so as to solve equations of the form

$$\int_0^x dy\, K[(x^n-y^n)^{1/n}]\phi(y)=g(x), \tag{13}$$

where n is any positive integer. The final result is

$$\phi(x)=nx^{n-1}L_{x^n}^{-1}\{\hat{g}_n(s)/\hat{K}_n(s)\}, \tag{14}$$

where

$$\hat{g}_n(s)=n\int_0^\infty dx\, x^{n-1}\exp(-sx^n)g(x), \tag{15}$$

and

$$\hat{K}_n(s)=n\int_0^\infty dx\, x^{n-1}\exp(-sx^n)K(x). \tag{16}$$

[1] A. Guinier and G. Fournet, *Small-Angle Scattering of X-Rays* (Wiley, New York, 1955), p. 118.
[2] O. Kratky, G. Porod, and L. Kahovec, Z. Elektrochem. **55**, 53 (1951).
[3] M. Deutsch and M. Luban, (J. Appl. Cryst.) to be published.
[4] See, for example, F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957), Sec. 1.9; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I, p. 972.
[5] See, for example, I. H. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972), pp. 174–5.

Structural properties of the self-conjugate SU(3) tensor operators

M. A. Lohe[a), b)] and L. C. Biedenharn[a)]

Department of Physics, Duke University, Durham, North Carolina 27706

J. D. Louck[c)]

Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 21 March 1977)

Denominator functions for the set of self-conjugate SU(3) tensor operators are explicitly obtained and shown to be uniquely related to SU(3)-invariant structural properties. This relationship becomes manifest through the appearance of zeroes of the denominator functions which thereby express the fundamental null space properties of SU(3) tensor operators. It is demonstrated that there exist characteristic denominator functions whose zeroes, in position and multiplicity, possess the interesting, and unexpected, property of forming SU(3) weight space patterns (in which the zeroes play the role of weights).

I. INTRODUCTION

The purpose of the present paper[1] is to illustrate in the simplest comprehensive case—the set of self-conjugate SU(3) tensor operators—precisely how the canonical operator construction[2] implies structural information equivalent to a unique [within (±) phases] definition of the operator. This structural information is SU(3)-invariant, and is contained in the SU(3) denominator function (the operator norm). The zeroes of this denominator function define the set of SU(3) vectors annihilated by the tensor operator in question; this set is called *the null space of the tensor operator*. It is a fundamental result for SU(3),[3] which we illustrate in this paper, that the canonical construction implies that the null spaces for the tensor operators in a given multiplicity set are simply ordered by inclusion, and this fact characterizes the operators uniquely.

To illustrate these results explicitly, and fully, would require that one construct the set of all canonical tensor operators, determine their invariant norms (denominator functions), and demonstrate directly the null space properties of this set of denominator functions. This we have done, and the results will be presented subsequently.[4] As one might expect, these general results are formidably complicated, and their very generality obscures the structure under discussion. For the set of self-conjugate operators, the canonical construction is greatly simplified, and the underlying structure can be much more readily understood. Since the self-conjugate operators do in fact realize the full multiplicity, their structure does indeed suffice to illustrate the general features of the problem. Accordingly we shall in this paper concentrate, for clarity, on this particular case alone.

The principal result which will be obtained in this paper is the generalization of the characteristic denominator function $G_q{}^t(\Delta,x)$ [cf. Sec. IV, Eq. (20)]. A special case of this function was defined earlier for *stretched* operator patterns[1] ($t=1$) and for *minimal* operator patterns ($t=\mathcal{M}$, where $\mathcal{M}$ = multiplicity). These functions, $G_q{}^t(\Delta,x)$, possess some very remarkable properties, which for greatest generality requires that the arguments (Δ) be in general position. The self-conjugate case [where $\Delta=(qqq)$ in special position] does suffice to illustrate one property which we feel is the most remarkable of all: *The zeroes of the characteristic denominator function $G_q{}^t(\Delta,x)$ are, in position and multiplicity, precisely given by SU(3) irrep weight space patterns.* For $G_q{}^t(\Delta,x)$, the associated weight space pattern is that of the irrep $[q-1,t-1,0]$.

This "weight space" property was, to us, totally unexpected, but it accords with all the results which we have obtained earlier.[1] For $t=1$, recall that we found a *triangle of zeroes* for G_q—this nicely correlates with the irrep $[q-1\ 0\ 0]$. For $t=\mathcal{M}$ we found an *inverted* triangle of zeroes, which correlates with the irrep $[q-1\ q-1\ 0]$.

The motivation of the present paper is to present this remarkable "weight space" result in the most accessible way.

The plan of the paper is as follows. In Sec. II we generate the (canonical) self-conjugate operators by coupling totally symmetric operators together with their conjugates; the denominator functions are then obtained in an explicit form from this construction. (It is this section which exploits the simplicity of construction mentioned above.) In Sec. III we recall the null space implications of the canonical construction, which proved that the null spaces of the canonical splitting are nested. In Sec. IV we apply these null space properties to the structure of the denominator functions, and in Sec. V we prove the property that the characteristic polynomials $G_q{}^t$ do indeed possess weight space patterns of zeroes, as asserted above.

a) Supported in part by the National Science Foundation under grant GP-14116.
b) Supported in part by an Overseas Scholarship from the Royal Commission for the Exhibition of 1851.
c) Work performed under the auspices of the USERDA.

II. A DETERMINANTAL FORM FOR THE DENOMINATOR FUNCTIONS OF SELF-CONJUGATE SU(3) TENSOR OPERATORS

The self-conjugate tensor operators in SU(3) belong to irreps labelled by $[p\ q\ r]=[2k\ k\ 0]$; here $p=2q=2k$ expresses the self-conjugate property and $r=0$ expresses the restriction to SU(3) [rather than U(3)]. We will confine our attention to operators in the *maximal multiplicity set*, that is, to operators having the shift labels $\Delta=(kkk)$, for which the multiplicity is $k+1$. The associated operator patterns, Γ_t, which canonically label the operators

$$\begin{pmatrix} \Gamma_t \\ 2k\ \ k\ \ 0 \end{pmatrix},$$

range over $t=1$: the stretched pattern

$$\begin{pmatrix} k \\ 2k\ \ 0 \\ 2k\ \ k\ \ 0 \end{pmatrix};$$

to $t=\mathcal{M}=k+1$: the minimal pattern

$$\begin{pmatrix} k \\ k\ \ k \\ 2k\ \ k\ \ 0 \end{pmatrix}.$$

It is useful, especially for self-conjugate operators, to employ a notation based on U(3), and permit *negative* integer labels in the irreps, in both the Gel'fand patterns and in the operator patterns. Accordingly, we will designate $[2k\ k\ 0]$ by $[k\ 0\ -k]$, with corresponding changes in the rest of the patterns.

Our starting point for constructing the desired set of self-conjugate operators is to consider the coupling of the multiplicity-free operator $\langle k\ 0\ 0\rangle$ together with its conjugate $\langle 0\ 0\ -k\rangle$. To simplify the construction, we need consider only the U3 : U2 projective operators, and accordingly use the coupling law:

$$\begin{bmatrix} \Gamma \\ k\ 0\ 0 \\ k\ 0 \\ k \end{bmatrix} \begin{bmatrix} \bar{\Gamma} \\ 0\ 0\ -k \\ 0\ 0 \\ 0 \end{bmatrix} = \sum_{t=1}^{k+1} \left\{ \begin{pmatrix} k\ 0\ -k \\ \Gamma_t \end{pmatrix} \begin{pmatrix} k \\ k\ 0 \\ k\ 0\ 0 \\ \Gamma \end{pmatrix} \begin{pmatrix} 0\ 0\ -k \\ \bar{\Gamma} \end{pmatrix} \right\} \begin{bmatrix} \Gamma_t \\ k\ 0\ -k \\ k\ 0 \\ k \end{bmatrix}. \tag{1}$$

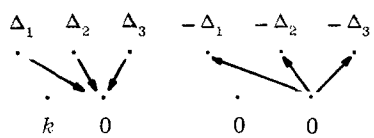
[In Eq. (1), recall that the symbol $[\cdots]$ denotes a U3 : U2 projective operator and the symbol $\{\cdots\}$ a Racah coefficient in SU(3).]

The procedure we will follow is to calculate the explicit forms of the operators on the left-hand side by using the pattern calculus.[1] To determine the canonical operators, given on the right-hand side, we will use a modification of the familiar Schmidt process. The Schmidt process is, of course, very far from canonical; one sees this from the fact that *arbitrary choices* are involved at each step, in selecting the sequence of vectors to be orthogonalized. But the canonical splitting of the multiplicity specifies precisely the proper sequence to be followed, beginning with Γ_1 (for which the Racah coefficient is known) and ending with $\Gamma_{\mathcal{M}}$ (for which the canonical result is also known). With this modification—specifying the canonical sequence—the Schmidt process becomes free of all arbitrary choices.

The form of the operators on the left-hand side of Eq. (1) has been explicitly given in Ref. 5 (Sec. 3D). For the right-hand side, we substitute the form:

$$\begin{bmatrix} \Gamma_t \\ k\ 0\ -k \\ k\ 0 \\ k \end{bmatrix} \begin{pmatrix} [m] \\ [m'] \end{pmatrix} = (\mathrm{NPCF})^{1/2} \rho_t(p_{22}) \Big/ D_2 \begin{pmatrix} [k\ 0] \\ [k\ 0] \end{pmatrix} (\,[m']\,) D \begin{pmatrix} \Gamma_t \\ k\ 0\ -k \end{pmatrix} (\,[m]\,), \tag{2}$$

where $[m']=[m_{12}m_{22}]$, $[m]=[m_{13}m_{23}m_{33}]$, NPCF stands for the numerator pattern calculus factor (evaluated by the pattern calculus rules on taking matrix elements), and ρ_t is a polynomial in p_{22} (and p_{i3}, but *not* p_{12}). The U(2) denominator D_2 is the same on both sides of Eq. (1) and cancels out. The same is true for NPCF, leaving the only contribution to the left-hand side of Eq. (1) to be that coming from the opposing arrows in the arrow patterns of the two operators. Using henceforth the notation $\Delta=(\Delta_1\ \Delta_2\ \Delta_3)$ to denote the Δ pattern of $\left(\begin{smallmatrix} \Gamma \\ k\ 0\ 0 \end{smallmatrix}\right)$, this contribution is evaluated from the diagrams



The result is $\left|\prod_{i=1}^{3}(p_{22}-p_{i3}+1)_{\Delta_i}\right|$, where $(x)_a = x(x+1)\cdots(x+a-1)$ denotes a rising factorial, and we recall also that $p_{ij} \equiv m_{ij}+j-i$.

Accounting for the phases[5] of the projective operator matrix elements, we thus obtain the following result from Eq. (1):

$$[p_{13}-p_{22}-1]_{\Delta_1}[p_{23}-p_{22}-1]_{\Delta_2}[p_{33}-p_{22}-1]_{\Delta_3}\Big/\left[D\begin{pmatrix}\Gamma\\ k\ 0\ 0\end{pmatrix}D\begin{pmatrix}\bar{\Gamma}\\ 0\ 0\ -k\end{pmatrix}\right]([m])$$

$$=\sum_{t=1}^{k+1}\left\{\begin{pmatrix}k\ 0\ -k\\ \Gamma_t\end{pmatrix}\begin{pmatrix}k\\ k\ 0\\ k\ 0\ 0\\ \Gamma\end{pmatrix}\begin{pmatrix}0\ 0\ -k\\ \bar{\Gamma}\end{pmatrix}\right\}([m])\rho_t(p_{22})\Big/D\begin{pmatrix}\Gamma_t\\ k\ 0\ -k\end{pmatrix}([m]), \tag{3}$$

where $[x]_a = x(x-1)\cdots(x-a+1)$ denotes a falling factorial ($(-x)_a = (-1)^a[x]_a$).

This is the basic equation from which we extract the Racah coefficients and, by orthonormality, the denominator functions; the key fact is that Eq. (3) holds identically as a polynomial in p_{22}.

It is convenient to use the variables $x_i = p_{j3} - p_{k3}$ (i,j,k cyclic), instead of p_{ij}; this is done by shifting $p_{ij} \to p_{ij} - \frac{1}{3}(p_{13}+p_{23}+p_{33})$, since Eq. (3) is invariant under constant shifts on p_{ij}, and Eq. (3) then becomes an identity in $y = p_{22} - \frac{1}{3}(p_{13}+p_{23}+p_{33})$.

The numerator of the left-hand side of Eq. (3) may now be expanded

$$[p_{13}-p_{22}-1]_{\Delta_1}[p_{23}-p_{22}-1]_{\Delta_2}[p_{33}-p_{22}-1]_{\Delta_3}$$

$$=[\tfrac{1}{3}(x_3-x_2)-y-1]_{\Delta_1}[\tfrac{1}{3}(x_1-x_3)-y-1]_{\Delta_2}[\tfrac{1}{3}(x_2-x_1)-y-1]_{\Delta_3} \equiv \sum_{i=1}^{k+1} y^{k-i+1}F_i(\Delta,x), \tag{4}$$

thereby defining the symmetric functions $F_i(\Delta,x)$. Let us also denote

$$V_i(\Delta,x) \equiv F_i(\Delta,x)\Big/\left[D\begin{pmatrix}\Gamma\\ k\ 0\ 0\end{pmatrix}D\begin{pmatrix}\bar{\Gamma}\\ 0\ 0\ -k\end{pmatrix}\right]([m]), \tag{5a}$$

$$R_t(\Gamma,x) \equiv \left\{\begin{pmatrix}k\ 0\ -k\\ \Gamma_t\end{pmatrix}\begin{pmatrix}k\\ k\ 0\\ k\ 0\ 0\\ \Gamma\end{pmatrix}\begin{pmatrix}0\ 0\ -k\\ \bar{\Gamma}\end{pmatrix}\right\}([m]), \tag{5b}$$

$$D(\Gamma_t,x) \equiv D\begin{pmatrix}\Gamma_t\\ k\ 0\ -k\end{pmatrix}([m]). \tag{5c}$$

Equation (3) becomes

$$\sum_{i=1}^{k+1} y^{k-i+1}V_i(\Delta,x) = \sum_{t=1}^{k+1} R_t(\Gamma,x)\frac{\rho_t(x,y)}{D(\Gamma_t,x)}. \tag{6}$$

We now assert that the Racah function $R_t(\Gamma,x)$ is obtained by carrying out the Schmidt process on the ordered set $\{V_i(\Delta;x)\}$, $i=1,\ldots,t$. It is seen from Eq. (6) that $R_t(\Gamma,x)/D(\Gamma_t,x)$ is determined by the set $\{V_i(\Delta,x)\}$, and the problem is to find the order which determines the orthogonalization. This ordering is determined by the power of y in $\rho_t(x,y)$; for $t=1$ we know that $\rho_1(x,y)$ is of degree k in y and that the coefficient of y^k is one (up to phase). (This follows from orthonormality of the Racah functions, and the known results for $t=1$.) Now, in Eq. (6) subtract from both sides the $t=1$ term $R_1(\Gamma,x)\,\rho_1(x,y)/D(\Gamma_1,x)$, leaving an identity in y of degree $k-1$. By equating coefficients of highest power in y, the $t=2$ term is separated off to give results for $t=2$, and the process can be repeated. Thus, the degree $k-t+1$ of y in $\rho_t(x,y)$ determines the relevant ordering of $\{V_i(\Delta,x)\}$ to be $i=1,\ldots,t$.

The Schmidt process now yields the expression for $R_t(\Gamma,x)/D(\Gamma_t,x)$, as determinants. To be explicit, we define

$$A_s \equiv \det\begin{bmatrix}(V_1,V_1) & \cdots & (V_1,V_s)\\ \vdots & & \vdots\\ (V_s,V_1) & \cdots & (V_s,V_s)\end{bmatrix}, \tag{7a}$$

where

$$(V_t, V_j) = \sum_{\substack{\Delta_1,\Delta_2,\Delta_3 \\ \Delta_1+\Delta_2+\Delta_3=k}} V_t(\Delta,x)V_j(\Delta,x)\,. \tag{7b}$$

Then

$$\frac{R_t(\Gamma,x)}{D(\Gamma_t,x)} = \det \begin{bmatrix} (V_1,V_1) & \cdots & (V_1,V_{t-1}) & V_1(\Delta,x) \\ \vdots & & & \vdots \\ (V_t,V_1) & \cdots & (V_t,V_{t-1}) & V_t(\Delta,x) \end{bmatrix} \Big/ A_t\,. \tag{7c}$$

Normalization of $R_t(\Gamma,x)$ yields:

$$D^2(\Gamma_t,x) = D^2\begin{pmatrix} \Gamma_t \\ k\ \ 0\ \ -k \end{pmatrix}([m]) = \frac{A_{t-1}}{A_t}\,. \tag{8}$$

Equation (8) establishes the main result of this section, and shows that the desired denominator functions are ratios of Gram determinants, whose elements are given by Eqs. (4), (5), and (7).

In Sec. IV, using results to be developed in Sec. III, following, we shall determine the polynomial structures involved in these determinantal forms.

It is of interest to note that Eq. (7c) gives explicit results for a large, and useful, class of SU(3) Racah functions.

III. NULL SPACES OF THE SELF-CONJUGATE OPERATORS

We have emphasized that the null spaces of operators in a multiplicity set are simply ordered, and characterize the operators uniquely. The null space is determined by the zeroes of the denominator function, but we show here how the precise knowledge of the null space can be applied to derive properties of the denominator functions. For example, particular linear factors can be deduced from lines of zeroes on the null space diagram, implying also the existence of polynomials which are required to have various patterns of zeroes.

The significance and properties of the null spaces of SU(3) tensor operators have been discussed previously in full generality,[3] and we apply the results here to the self-conjugate operators. These results can be depicted most concisely in terms of null space boundaries drawn on the Möbius plane (Ref. 3, Figs. 1 and 2); the corresponding boundaries for our case, i.e., the regions where $D(\Gamma_t,x) \neq 0$ and where $D(\Gamma_t,x) = 0$, are displayed in Fig. 1. The lexical region is $x_1, x_3 \geq 1$; the (hatched) region for which $D(\Gamma_t,x) > 0$ [when the operator with normalization $D(\Gamma_t,x)$ has nonzero matrix elements] is $x_1, x_3 \geq k - t + 2$, $x_2 \leq -2k + t - 3$, and otherwise in the lexical region $D(\Gamma_t,x) = 0$.

In order to illustrate the null space concept, let us examine the expressions for $D^2(\Gamma_t,x)$ in the cases which are known (Ref. 1), namely $t = 1$ (maximal null space) and $t = k + 1$ (minimal null space). These results are

$$D^2(\Gamma_1,x) = (k!)^2 \prod_{i=1}^{3} \prod_{\lambda=1}^{k} (x_i^2 - \lambda^2)/G_k(k,k,k;x)\,, \tag{9a}$$

$$D^2(\Gamma_{k+1},x) = [(k+1)/(2k+1)!]G_k(-1,-1,-1;-x)\,. \tag{9b}$$

The first denominator function ($t = 1$) consists partly of linear factors which result in lines of zeroes at x_1 (and x_3) $= 1, \ldots, k$ (plus symmetries), and partly of a triangle of zeroes from G_k *which reduce to first order the multiplicity of zeroes at the boundary of the null region.*

The second denominator function ($t = k + 1$) has first order zeroes in the lexical region only where $G_k(-1,-1,-1;-x)$ vanishes, namely in the inverted triangle $x_1, x_3 \geq 1$, $x_2 \geq -k - 1$. This is precisely the region where $D^2(\Gamma_{k+1},x)$ is required to be zero.

In the following sections we generalize these results to the full set of denominator functions for self-conjugate operators.

IV. POLYNOMIAL STRUCTURE OF THE DETERMINANTS AND IMPLICATIONS OF NULL SPACE

We propose now to examine separately the elements of the set of determinants A_t—which determine $D(\Gamma_t,x)$ through Eq. (8)—and show first by algebraic means how

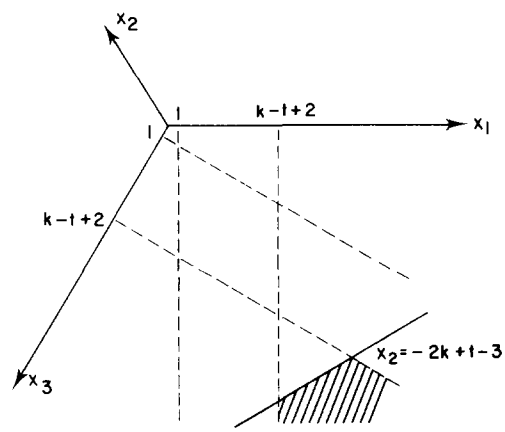


FIG. 1. Null space diagram for the self-conjugate operators. The lexical region is defined by $x_1, x_3 \geq 1$. The lattice points belonging to the hatched region defined by $x_1, x_3 \geq k - t + 2$, $x_2 \leq -2k + t - 3$ comprise the lexical region where the denominator function $D(\Gamma_t,x)$ is positive; at all other lattice points of the lexical region $D(\Gamma_t,x)$ is zero.

each element may be written as a polynomial divided by certain linear factors. Then, by considering the determinant as a whole, we show how more linear factors may be extracted using properties of the null space (discussed in Sec. III), so that the determinant finally appears as a polynomial $G_k^t(x)$ divided by linear factors.

The denominator functions appearing in $V_i(\Delta, x)$ [see Eq. (5)] have been given explicitly in Ref. 5, Eq. (4.18), and are as follows:

$$\left[D\begin{pmatrix} & \Gamma & \\ k & 0 & 0\end{pmatrix}D\begin{pmatrix} & \overline{\Gamma} & \\ 0 & 0 & -k\end{pmatrix}\right]^2([m]) = (1/k!^2)\prod_{i=1}^{3}(\Delta_i![x_i+\Delta_k]_{\Delta_j+\Delta_k+1})^2/x_i(x_i+\Delta_k-\Delta_j), \tag{10a}$$

i,j,k cyclic. (We have not carried out the explicit cancellations possible which would make the right-hand side manifestly a polynomial.) We need to express (V_i, V_j) as a polynomial over a lowest common denominator. This lowest common denominator is found from the denominator in Eq. (10a), and turns out to be

$$\prod_{i=1}^{3} x_i \prod_{\lambda=1}^{k-1}(x_i^2-\lambda^2)\prod_{\mu=1}^{k}(x_i^2-\mu^2). \tag{10b}$$

To see this, it is sufficient to observe that every factor of

$$\left[D\begin{pmatrix} & \Gamma & \\ k & 0 & 0\end{pmatrix}D\begin{pmatrix} & \overline{\Gamma} & \\ 0 & 0 & -k\end{pmatrix}\right]^2([m])$$

is contained in the expression (10b). (It is easiest to examine the factors containing only x_1 say, and then the others follow by symmetry.)

We can write

$$(V_i, V_j) = \sum_{\Delta} F_i(\Delta, x)F_j(\Delta, x)N(\Delta, x)\Big/\prod_{i=1}^{3} x_i \prod_{\lambda=1}^{k-1}(x_i^2-\lambda^2)\prod_{\mu=1}^{k}(x_i^2-\mu^2), \tag{11}$$

where the sum is over $\Delta_1, \Delta_2, \Delta_3$ with $\Delta_1+\Delta_2+\Delta_3 = k$. Here $N(\Delta, x)$ is defined by

$$N(\Delta, x) = k!^2\prod_{i=1}^{3}\left[x_i^2(x_i+\Delta_k-\Delta_j)\prod_{\lambda=1}^{k-1}(x_i^2-\lambda^2)\times\prod_{\mu=1}^{k}(x_i^2-\mu^2)(\Delta_i![x_i+\Delta_k]_{\Delta_j+\Delta_k+1})^{-2}\right], \tag{12}$$

i,j,k cyclic. It is important to observe that $N(\Delta, x)$ is a polynomial in x, but it is convenient to write $N(\Delta, x)$ in the form of Eq. (12). Next we show that the factors

$$\prod_{i=1}^{3} x_i \prod_{\lambda=1}^{k-1}(x_i^2-\lambda^2)$$

may be cancelled between numerator and denominator in Eq. (11). This will be done by showing that

$$\sum_{\Delta} F_i(\Delta, x)F_j(\Delta, x)N(\Delta, x) = 0 \quad \text{for } x_3 = \lambda, \text{ when } \lambda = 0, 1, \ldots, k-1. \tag{13}$$

This result implies that $x_3\prod_{\lambda=1}^{k-1}(x_3-\lambda)$ can be cancelled in Eq. (11), but it then follows by symmetry that $\prod_{i=1}^{3}x_i\prod_{\lambda=1}^{k-1}(x_i-\lambda)(x_i+\lambda)$ can also be cancelled. [The symmetry used here is that (V_i, V_j) is invariant under $(x)\to\delta_P P(x)$, where P denotes a permutation of the indices 1, 2, 3 (i.e., $P\in S_3$) and δ_P is the signature of P.] The proof of Eq. (13) is given in the Appendix.

With this cancellation, Eq. (11) is rewritten:

$$(V_i, V_j) = (\text{polynomial})\Big/\prod_{i=1}^{3}\prod_{\mu=1}^{k}(x_i^2-\mu^2).$$

The determinant therefore takes the form:

$$A_t = P_k^t(x)\Big/\left[\prod_{i=1}^{3}\prod_{\lambda=1}^{k}(x_i^2-\lambda^2)\right]^t \tag{14}$$

for some polynomial $P_k^t(x)$.

Equation (14) can be simplified further, using the properties of null space. First, from Eqs. (8) and (14), we have

$$D^{-2}(\Gamma_1, x) = A_1 = P_k^1(x)\Big/\prod_{i=1}^{3}\prod_{\lambda=1}^{k}(x_i^2-\lambda^2), \tag{15}$$

so that, by comparison with Eq. (9a), in Sec. III, we find that

$$k!^2 P_k^1(x) = G_k(k, k, k; x). \tag{16}$$

We use now the null space diagram for general t to show that the polynomial $P_k^t(x)$ [defined by Eq. (14)] contains linear factors; in fact,

$$P_k^t(x) = C_t\left[\prod_{i=1}^{3}\prod_{\mu=k-t+2}^{k}\prod_{\lambda=\mu}^{k}(x_i^2-\lambda^2)\right]G_k^t(x), \tag{17}$$

for some polynomial $G_k^t(x)$ and some constant C_t (depending only on t and k) yet to be fixed.

We will prove Eq. (17) by induction on t, and we begin by noting that Eq. (17) holds for $t = 1$. This follows from Eq. (16) if we define

$$G_k^1(x) = G_k(k, k, k; x). \tag{18}$$

[Let us remark that for $t = 1$ all linear factors are extracted from $P_k^t(x)$, but in general extra factors may appear, not from each determinantal element (V_i, V_j), but from the determinant as a whole.]

Suppose Eq. (17) holds for $t = r-1$, and let us show it to be true then for $t = r$. From Eqs. (8) and (14) and the induction hypothesis we have

$$D^{-2}(\Gamma_r, x) = P_k^r(x)\Big/C_{r-1}G_k^{r-1}(x)\left(\prod_{i=1}^{3}\prod_{\lambda=1}^{k-r+1}(x_i^2-\lambda^2)\right)\times\left(\prod_{i=1}^{3}\prod_{\lambda=k-r+2}^{k}(x_i^2-\lambda^2)\right)\left(\prod_{i=1}^{3}\prod_{\mu=k-r+3}^{k}\prod_{\lambda=\mu}^{k}(x_i^2-\lambda^2)\right). \tag{19}$$

From the null space diagram we have that $D^{-2}(\Gamma_r, x) > 0$ for $x_1, x_3 \geq k - r + 2$, $x_2 \leq -2k + r - 3$; in this region therefore there can be no lines of zeroes, implying that the factors $\prod_{i=1}^{3} \prod_{\mu=k-r+2}^{k} \prod_{\lambda=\mu}^{k} (x_i^2 - \lambda^2)$ in Eq. (19) do not actually appear in the denominator, and must therefore divide into $\rho_k^r(x)$. This shows that

$$G_k^r(x) = \rho_k^r(x)/C_r \prod_{i=1}^{3} \prod_{\mu=k-r+2}^{k} \prod_{\lambda=\mu}^{k} (x_i^2 - \lambda^2),$$

for some constant C_r, is a polynomial in x thus completing the induction proof of Eq. (17).

In summary to this point, we have shown that the function $G_k^t(x)$ defined by

$$G_k^t(x) = \frac{1}{C_t}\left[\prod_{i=1}^{3} \prod_{\mu=1}^{t} \prod_{\lambda=1}^{k-\mu+1} (x_i^2 - \lambda^2)\right] A_t, \tag{20}$$

which appears in the denominator function $D(\Gamma_t, x)$, is actually a polynomial in the (x_i). We shall see that, for general x, $G_k^t(x)$ is "irreducible," i.e., cannot be further factorized as the product of two polynomials.

This polynomial $G_k^t(x)$ is the natural generalization, to arbitrary t, of $G_k(kkk; x) = G_k^1(x)$ introduced in Refs. 1 and 3. It is of fundamental importance in the structure of the Wigner and Racah functions with which it is associated. As is the case for $G_q(\Delta, x)$, where Δ now refers to the Δ pattern of the operator $\langle pq0 \rangle$, the general $G_q^t(\Delta, x)$ possesses many remarkable properties which are discussed further in the next section, for the special case of the self-conjugate operators.

Let us note the form of the denominator function which we have established in this section:

$$D^{-2}(\Gamma_t, x) = G_k^t(x)/G_k^{t-1}(x) K_t \prod_{i=1}^{3} \prod_{\lambda=1}^{k-t+1} (x_i^2 - \lambda^2) \tag{21}$$

for some constant $K_t = C_t/C_{t-1}$, which is independent of the (x_i).

V. ZEROES OF THE CHARACTERISTIC POLYNOMIALS, $G_k^t(x)$

One of the most striking and unexpected properties of the set of polynomials $G_k^t(x)$ concerns their zeroes. We show now, from null space requirements, that $G_k^t(x)$ has a zero at the point x of a given region R^t of the null space diagram (see Fig. 2), of multiplicity

$$M^t(x) = \tfrac{1}{2}(k + 1 - |k - t + 1 - x_1| - |k + 1 + x_2| - |k - t + 1 - x_3|). \tag{22}$$

This result is to be compared with the formula for the multiplicity of the weight $(\Delta_1\ \Delta_2\ \Delta_3)$ in the representation $[p\ q\ 0]$ [Ref. 1, Eq. (2.17)], which can be written as follows [putting $\lambda_i = \frac{1}{2}(|q - \Delta_i| + q - \Delta_i)$]:

$$M = \tfrac{1}{2}(p + 2 - |\Delta_1 - q| - |\Delta_2 - q| - |\Delta_3 - q|). \tag{23}$$

If we substitute $p = k - 1$, $q = t - 1$, and $\Delta_1 = k - x_1$, $\Delta_2 = t - k - 2 - x_2$, $\Delta_3 = k - x_3$, the two formulas (22) and (23) are exactly the same. In other words, $G_k^t(x)$ has zeroes which in position and multiplicity fall into the familiar pattern of the weight space diagram of the representation $[k - 1, t - 1, 0]$. For example, in Fig. 3 we illustrate the case $G_3^2(x)$ which has an "octet" pattern. Note that the central point of the octet has a double zero.

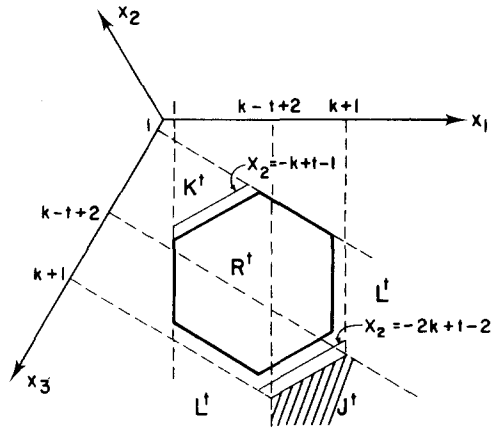


FIG. 2. Regions of the null space plane characterized by different properties of the denominator function $D(\Gamma_t, x)$. For discussing the properties of $D(\Gamma_t, x)$ the lexical region is conveniently divided into four disjoint regions identified as $\mathcal{J}^t$, R^t, L^t, and K^t corresponding to the following properties of $D(\Gamma_t, x)$: (a) On the lattice points of $\mathcal{J}^t$ $(x_1, x_3 \geq k - t + 2,\ x_2 \leq -2k - 3 + t)$, $D(\Gamma_t, x)$ is positive (it is zero on all other lattice points of the lexical region); (b) on the lattice points of R^t $(x_1, x_3 \leq k,\ -2k - 1 + t \leq x_2 \leq -k + t - 2)$, the polynomial form $G_k^t(x)$ occurring in $D(\Gamma_t, x)$ is zero; (c) on the lattice points of the two regions denoted by L^t $(x_1, x_3 \geq k + 1$, $\mathcal{J}^t$ excluded), the denominator function $D(\Gamma_t, x)$ has single lines (linear factors) of zeroes; (d) on the lattice points of K^t (lexical region where $x_2 \geq -k + t - 1$), the denominator function $D(\Gamma_t, x)$ has (three) intersecting lines of zeroes. Observe that the union $\mathcal{J}^t \cup K^t \cup L^t \cup R^t$ defines the lexical region except for the $t - 1$ points on the line $x_2 = -2k + t - 2$ enumerated by $x_1 = k, k - 1, \ldots, k - t + 2$, these points belonging to R^{t-1}.

Before beginning the proof let us first establish some preliminary properties of $G_k^t(x)$. From Eq. (20), by counting degrees in x on the right-hand side, we have that $G_k^t(x)$ is *at most* of degree $2t(k - t + 1)$ (shortly we show this to be the exact degree). Therefore, by putting $t = k + 1$, $G_k^{k+1}(x)$ is of degree 0, and hence is a constant which can be chosen to be 1:

$$G_k^{k+1}(x) = 1. \tag{24}$$

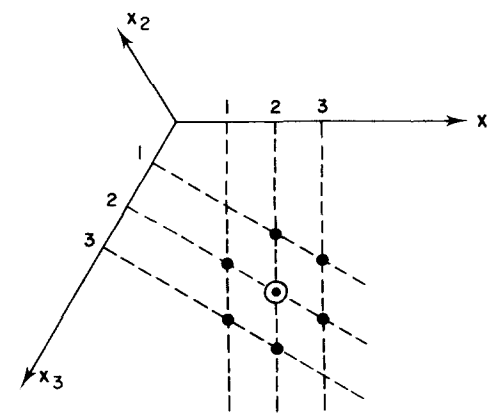


FIG. 3. Zeroes of $G_3^2(x)$. The eighth degree polynomial form $G_3^2(x)$ has zeroes (in the lexical region) which in position and multiplicity fall into the pattern of the weight space diagram of the octet.

Hence, from Eq. (21),

$$D^2(\Gamma_{k+1}, x) = K_{k+1} G_k^k(x),$$

so that, by comparison with Eq. (9b), we have

$$G_k^k(x) = G_k(-1, -1, -1; -x) \tag{25}$$

with $K_{k+1} = (k+1)/(2k+1)!$.

There are also symmetry properties of $G_k^t(x)$. Each element (V_i, V_j) of A_t is invariant under permutations P of S_3 (acting on x), and by Eq. (20) this symmetry extends to $G_k^t(x)$:

$$G_k^t(Px) = G_k^t(\delta_P x). \tag{26}$$

[There is actually also an additional symmetry in which $G_k^t(x) = G_k^t(-x)$.]

Now let us turn to the deduction of the zeroes of $G_k^t(x)$. From Eq. (21) for $D^2(\Gamma_t, x)$ we see that there are $k-t+1$ lines of zeroes in the null space arising from the linear factors. Let $L^t(x)$ denote the multiplicity of zeroes due to these lines, in the lexical region $x_1,\ x_3 \geq 1$. We require the following formula:

$$L^t(x) = 1 + \tfrac{1}{2}(|k-t+2-x_1| + |k-t+2-x_3| - |k-t+1-x_1| - |k-t+1-x_3|). \tag{27}$$

This is verified case by case for the following four regions (see Fig. 1):

(a) $x_1, x_3 \geq k-t+2$ for which $L^t(x) = 0$,

(b) $x_1 \geq k-t+2,\ x_3 \leq k-t+1$ for which $L^t(x) = 1$,

(c) $x_1 \leq k-t+1,\ x_3 \geq k-t+2$ for which $L^t(x) = 1$,

(d) $x_1, x_3 \leq k-t+1$ for which $L^t(x) = 2$.

Also let $M^t(x)$ denote the multiplicity of zeroes of $G_k^t(x)$ in the lexical region at the point (x_1, x_2, x_3). [This means we have $M^t(x) = 0$ at a point x where $G_k^t(x)$ is nonzero.]

It is necessary to divide the lexical region into the following sections, on each of which $D^2(\Gamma_t, x)$ has different properties (Fig. 2):

(a) Let $\mathcal{J}^t$ denote the region $x_1, x_3 \geq k-t+2,\ x_2 \leq -2k-3+t$ [in this region $D^2(\Gamma_t, x) > 0$].

(b) Let $\mathcal{R}^t$ denote the region $x_1, x_3 \leq k,\ -2k-1+t \leq x_2 \leq -k+t-2$ [this is the region of zeroes of $G_k^t(x)$].

(c) Let $\mathcal{K}^t$ denote the part of the lexical region for which $x_2 \geq -k+t-1$.

(d) Let $\mathcal{L}^t$ denote the region $x_1, x_3 \geq k+1$, but excluding $\mathcal{J}^t$ (i.e., the two regions covered by the single lines of the linear factors).

Note that $\mathcal{J}^{t-1} \subset \mathcal{J}^t$, $\mathcal{L}^t \subset \mathcal{L}^{t-1}$, and $\mathcal{K}^t \subset \mathcal{K}^{t-1}$. The lexical region can be expressed as the following unions:

$$\mathcal{J}^t \cup \mathcal{K}^t \cup \mathcal{L}^t \cup \mathcal{R}^t \cup \mathcal{R}^{t-1} = \mathcal{J}^{t+1} \cup \mathcal{L}^t \cup \mathcal{R}^t \cup \mathcal{K}^t. \tag{28}$$

The property of the null space to be used is

$$\begin{aligned} &D^2(\Gamma_t, x) = 0 \quad \text{for } x \in \mathcal{R}^t \cup \mathcal{R}^{t-1} \cup \mathcal{K}^t \cup \mathcal{L}^t, \\ &D^2(\Gamma_t, x) > 0 \quad \text{for } x \in \mathcal{J}^t. \end{aligned} \tag{29}$$

Firstly, we show by induction on t that

$$M^t(x) = 0 \quad \text{for } x \in \mathcal{J}^{t+1} \tag{30a}$$

and

$$M^t(x) \geq \tfrac{1}{2}(k+1-|k-t-x_1| - |k+1+x_2| - |k-t+1-x_3|) \quad \text{for } x \in \mathcal{R}^t. \tag{30b}$$

These equations are true for $t = k$, since $G_k^k(x)$ is completely known in Eq. (25). In particular, the zeroes are of first order, so that $M^k(x) \geq 1$ for $x \in \mathcal{R}^k$, and otherwise $G_k^k(x) \neq 0$.

Suppose Eqs. (30) hold for $t = r$, and we show them to be true for $t = r-1$. From Eq. (21)

$$D^2(\Gamma_r, x) = G_k^{r-1}(x) K_r \prod_{i=1}^{3} \prod_{\lambda=1}^{k-r+1} (x_i^2 - \lambda^2) / G_k^r(x). \tag{31}$$

It follows that at the point x, $D^2(\Gamma_r, x)$ has a zero of multiplicity $M^r(x)$ given by

$$M^r(x) = M^{r-1}(x) + L^r(x) - M^r(x). \tag{32}$$

For $x \in \mathcal{J}^r \subset \mathcal{J}^{r+1}$ we have $M^r(x) = 0$, $L^r(x) = 0$ and since $M^r(x) = 0$ [because $D^2(\Gamma_r, x) > 0$ for $x \in \mathcal{J}^r$] we obtain $M^{r-1}(x) = 0$. Hence Eq. (30a) is proved by induction on r.

If $x \in \mathcal{R}^r$ we must have $M^r(x) \geq 1$, since $D^2(\Gamma_r, x) = 0$. Using this result and relation (27) in Eq. (32), together with the induction hypothesis, we find that for $x \in \mathcal{R}^{r-1} \cap \mathcal{R}^r$ the following result must be valid:

$$\begin{aligned} M^{r-1}(x) &\geq 1 + M^r(x) - L^r(x) \\ &= 1 + \tfrac{1}{2}(k+1-|k-r+1-x_1| - |k+1+x_2| \\ &\quad - |k-r+1-x_3|) - 1 - \tfrac{1}{2}(|k-r+2-x_1| \\ &\quad - |k-r+1-x_1| + |k-r+2-x_3| \\ &\quad - |k-r+1-x_3|), \end{aligned}$$

that is,

$$M^{r-1}(x) \geq \tfrac{1}{2}(k+1-|k-r+2-x_1| - |k-r+2-x_3| - |k+1+x_2|), \tag{33}$$

as required for $x \in \mathcal{R}^{r-1} \cap \mathcal{R}^r$. The part of $\mathcal{R}^{r-1}$ not contained in $\mathcal{R}^r$ is the line section $x_2 = -2k+r-2,\ k-r+2 \leq x_1, x_3 \leq k$, which lies in $\mathcal{J}^{r+1}$. Since $L^r(x) = 0$, and $M^r(x) = 0$ at these points [from Eqs. (27) and (30a)], $M^r(x) \geq 1$ implies $M^{r-1}(x) \geq 1$ so that Eq. (33) extends to hold in all of $\mathcal{R}^{r-1}$. Hence Eq. (30b) also is proved.

We have shown now that, in $\mathcal{R}^t$, $G_k^t(x)$ has at least the zeroes corresponding to the weight space diagram of $[k-1\ t-1\ 0]$, and that $G_k^t(x)$ is nonzero on $\mathcal{J}^{t+1}$ [clearly $G_k^t(x)$ is in fact positive definite in this region].

Now consider the line $x_1 = k-t+1$. It follows from Eq. (30b) that

$$\sum_{x_3=1}^{k} M^t(x_1, x_2, x_3) \geq t(k-t+1)$$

for $x_1 = k-t+1$, so that on this line $G_k^t(x)$ has at least $t(k-t+1)$ zeroes in $\mathcal{R}^t$, and, by symmetry, $2t(k-t+1)$ zeroes on $x_1 = k-t+1$ in the full Möbius plane. Since $G_k^t(x)$ is a polynomial in x of degree at most $2t(k-t+1)$, we deduce that:

either $G_k^t(x)$ vanishes identically on $x_1 = k-t+1$, which

is impossible because from Eq. (30a), $G_k^t(x) \neq 0$ for $x \in \mathcal{J}^t$, which includes a section of $x_1 = k - t + 1$;

or $G_k^t(x)$ has exactly $2t(k-t+1)$ zeroes on $x_1 = k - t + 1$, and these zeroes occur only in R^t and the corresponding symmetrically placed regions of the Möbius plane. Hence $G_k^t(x)$ is exactly of degree $2t(k-t+1)$.

One of the lines symmetric to $x_1 = k - t + 1$ is $x_2 = -(k-t+1)$, a part of which lies in K^t (Fig. 2). On this line $G_k^t(x)$ has $2t(k-t+1)$ zeroes, all of which lie outside the lexical region, and so $G_k^t(x) \neq 0$ for the part inside the lexical region, i.e.,

$$\mathcal{M}^t(x) = 0 \quad \text{for } x \in K^t \quad \text{and} \quad x_2 = -(k-t+1). \tag{34}$$

Next we show

$$\mathcal{M}^t(x) = 0 \quad \text{for } x \in L^t. \tag{35}$$

This equation holds for $t = 1$, from the known properties of $G_k(x)$, by Eq. (18). We assume Eq. (35) is true for $t = r - 1$, and require then [from Eq. (32)]:

$$\mathcal{M}^{r-1}(x) + L^r(x) - \mathcal{M}^r(x) = M^r(x) \geq 0 \quad \text{for } x \in L^r.$$

For $x \in L^r \subset L^{r-1}$ we have $L^r(x) = 1$, and $\mathcal{M}^{r-1}(x) = 0$ from the induction hypothesis. Therefore $\mathcal{M}^r(x) \leq 0$, i.e., $\mathcal{M}^r(x) = 0$ for $x \in L^r$, thus proving Eq. (35).

We now show that

$$\mathcal{M}^t(x) \leq \tfrac{1}{2}(k+1-|k-t+1-x_1|-|k+1+x_2| - |k-t+1-x_3|) \quad \text{for } x \in R^t. \tag{36}$$

This holds for $t = 1$ [when $\mathcal{M}^t(x) = 1$], so we suppose it holds for $t = r - 1$. We require from Eq. (32) that

$$\mathcal{M}^{r-1}(x) + L^r(x) - \mathcal{M}^r(x) = M^r(x) \geq 1. \tag{37}$$

Now substitute for $L^r(x)$ from Eq. (27) and for $\mathcal{M}^{r-1}(x)$ from the induction hypothesis, thus obtaining

$$\mathcal{M}^r(x) \leq \tfrac{1}{2}(k+1-|k-t+1-x_1|-|k+1+x_2| - |k-t+1-x_3|) \quad \text{for } x \in R^{r-1} \cap R^r. \tag{38}$$

The part of R^r not contained in R^{r-1} is that section of K^{r-1} which lies on the line $x_2 = -(k-r+2)$ (see Fig. 2). From Eq. (34), putting $k = r - 1$, we obtain

$$\mathcal{M}^{r-1}(x) = 0 \quad \text{for } x \in K^{r-1} \quad \text{and} \quad x_2 = -(k-r+2).$$

Also, on this line section $L^r(x) = 2$, so that from Eq. (37) one obtains $\mathcal{M}^r(x) \leq 1$. Therefore, Eq. (38) extends to hold for all $x \in R^r$, and this completes the induction argument to prove Eq. (36).

Combining Eqs. (30b) and (36) now gives the exact expression for $\mathcal{M}^t(x)$ as was asserted in Eq. (22). We have also shown that $G_k^t(x) > 0$ in other parts of the lexical region, as is required through the positivity of an operator norm.

In summary, we have obtained the denominator functions, in the form of Eq. (21), in which the $G_k^t(x)$ are polynomials with the weight space pattern of zeroes [Eq. (22)]. It is of interest to take a simple case of this result, and examine the remarkable way in which these zeroes arrange themselves on the Möbius plane. Consider $k = 3$, for which the multiplicity is $\mathcal{M} = 4$. There are four denominator functions, and three polynomials:

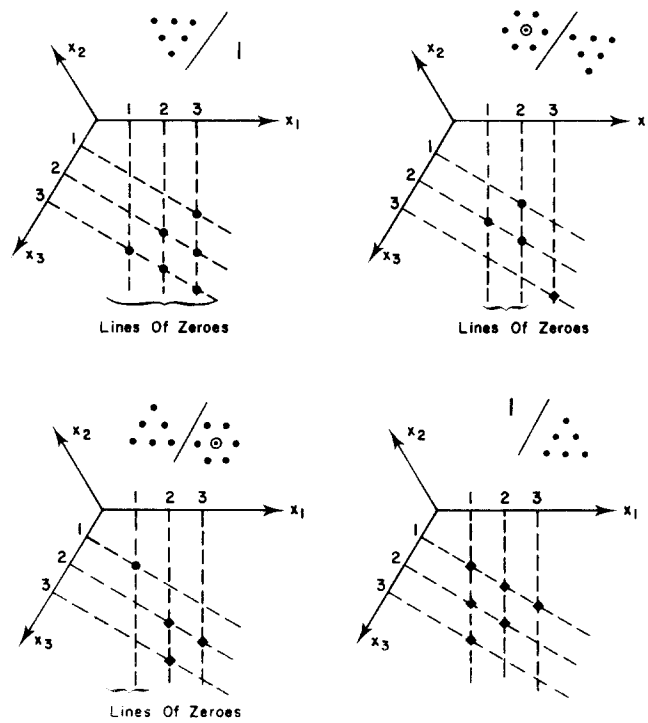


FIG. 4. The family of $\langle 630 \rangle$, $\Delta = (333)$ denominator functions, illustrating the weight space property. The ratios of the polynomials $G_3^1/(G_3^0 = 1)$, G_3^2/G_3^1, G_3^3/G_3^2 and $(G_3^4 = 1)/G_3^3$ possess first order zeros (circles) and poles (squares) characteristic of the denominator functions $D(\Gamma_1)$, $D(\Gamma_2)$, $D(\Gamma_3)$, and $D(\Gamma_4)$, respectively.

G_3^1 (the triangle of zeroes), G_3^2 (the octet of zeroes), and G_3^3 (the inverted triangle of zeroes). In Fig. 4 we have displayed the zeroes which arise only from the ratios of these polynomials — the diagram is completed by addition of suitable linear factors, according to Eq. (21), to account for the correct null space boundaries. Observe how the various multiplicities of zeroes combine to produce the proper *first order zeroes adjacent to these boundaries*. These features are completely general, even for arbitrarily high values of the operator multiplicity $\mathcal{M}$ when zeroes of arbitrarily high multiplicity appear in $G_k^t(x)$. This interplay of SU(3) symmetry patterns and null space boundaries we find to be of considerable interest and worthy of further study.

Remark: The question now arises as to whether the zeroes of $G_k^t(x)$, together with the symmetries, determine this polynomial uniquely. This property of uniqueness has been proved for $t = 1$ (Ref. 3), and for $t = k$ [via Eq. (25)], and provides a strong tool for handling these polynomials. Unfortunately, it is not true in the self-conjugate case for general t. This is shown by the following counterexample to the case $G_3^2(x)$:

$$(x_1^2 - 4)(x_2^2 - 4)(x_3^2 - 4)(x_1^2 + x_2^2 + x_3^2 - 26).$$

This latter polynomial, in linear combination with $G_3^2(x)$, has the same symmetries of S_3, the same (eighth) degree, and the octet of zeroes required of $G_3^2(x)$, but is not equal to $G_3^2(x)$.

This particular counterexample can be eliminated by considering $G_3^2(\Delta, x)$, that is, for the characteristic

polynomial in general position. However, our example shows that the uniqueness problem is of considerable subtlety.

APPENDIX

We prove here Eq. (13). First, note that the sum over Δ need be taken only for $\Delta_1 \geq \lambda$, since $(x_3 - \lambda)$ appears explicitly in $N(\Delta, x)$ for $\Delta_1 < \lambda$, so then $N(\Delta, x) = 0$ at $x_3 = \lambda > \Delta_1$. Now, for $\Delta_1 \geq \lambda$ it is necessary to cancel out explicitly in Eq. (12) the $(x_3 - \lambda)$ factors, before putting $x_3 = \lambda$. This is done using

$$[x_3]_k = [x_3]_\lambda [x_3 - \lambda - 1]_{k-\lambda-1}(x_3 - \lambda),$$

and

$$[x_3 + \Delta_2]_{\Delta_1+\Delta_2+1} = [x_3 + \Delta_2]_{\Delta_2+\lambda}[x_3 - \lambda - 1]_{\Delta_1-\lambda}(x_3 - \lambda).$$

Now put $x_1 = x$, $x_2 = -x - \lambda$, $x_3 = \lambda$. The part of $N(\Delta, x)$ which depends on Δ, and so remains under summation, is

$$(x + \Delta_3 - \Delta_2)(x + \lambda + \Delta_3 - \Delta_1)(\lambda + \Delta_2 - \Delta_1)/(\Delta_1!\,\Delta_2!\,\Delta_3!$$
$$\times [x + \lambda + \Delta_3]_{\Delta_1+\Delta_3+1}\ [x + \Delta_3]_{\Delta_2+\Delta_3+1}(\Delta_2 + \lambda)!\,(\Delta_1 - \lambda)!)^2.$$

Observe that this part of the summand is antisymmetric under the interchange $\Delta_1 - \lambda \rightarrow \Delta_2$. (To see this, one must use

$$[x + \lambda + \Delta_3]_{\Delta_1+\Delta_3+1}[x + \Delta_3]_{\Delta_2+\Delta_3+1}$$
$$= [x + \lambda + \Delta_3]_{\Delta_2+\lambda+\Delta_3+1}[x + \Delta_3]_{\Delta_1-\lambda+\Delta_3+1}).$$

Moreover, $F_i(\Delta, x)$ is symmetric under $\Delta_1 - \lambda \rightarrow \Delta_2$. This is because $F_i(\Delta, x)$ is defined from Eq. (4), the lhs of which is symmetric under $\Delta_1 - \lambda \rightarrow \Delta_2$, for $x_3 = \lambda$; this follows from the identity

$$[z + \lambda]_{\Delta_1}[z]_{\Delta_2} = [z + \lambda]_{\Delta_2+\lambda}[z]_{\Delta_1-\lambda},$$
$$\text{with } z = \tfrac{1}{3}x - \tfrac{1}{3}\lambda - \nu - 1.$$

We have shown now that the complete summand $N(\Delta, x)F_i(\Delta, x)F_j(\Delta, x)$ is antisymmetric under $\Delta_1 - \lambda \rightarrow \Delta_2$. Now in the sum $\sum_\Delta$ there will be for each index $(\Delta_1, \Delta_2, \Delta_3)$ a corresponding index $(\Delta_2 + \lambda, \Delta_1 - \lambda, \Delta_3)$. This is because

$$\lambda \leq \Delta_1 \leq k$$

implies

$$0 \leq \Delta_3 = k - \Delta_1 - \Delta_2 \leq k - \lambda - \Delta_2,$$

i.e.,

$$\Delta_2 \leq k - \lambda.$$

Therefore,

$$\lambda \leq \Delta_2 + \lambda \leq k, \quad 0 \leq \Delta_1 - \lambda \leq k - \lambda.$$

Now carry out the sum $\sum_\Delta N(\Delta, x)F_i(\Delta, x)F_j(\Delta, x)$ in pairs as indicated, in which the index $(\Delta_1, \Delta_2, \Delta_3)$ is summed with $(\Delta_2 + \lambda, \Delta_1 - \lambda, \Delta_3)$. The antisymmetry of the summand shows that each such pair must sum to zero. Summands for which $\Delta_1 = \Delta_2 + \lambda$ will be zero separately by the antisymmetry.

We have proved therefore

$$\sum_\Delta N(\Delta, x)F_i(\Delta, x)F_j(\Delta, x) = 0$$

for $x_3 = \lambda$, where $\lambda = 0, \ldots, k - 1$, as required.

[1] This paper is a continuation of the work of J. D. Louck, M. A. Lohe, and L. C. Biedenharn, J. Math. Phys. **16**, 2408–26 (1975).

[2] A survey of the relevant papers is in Ref. 1 above, and cited as Refs. 1–6 there.

[3] L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13**, 1985–2001 (1972).

[4] L. C. Biedenharn, M. A. Lohe, and J. D. Louck, in *Group Theoretical Methods in Physics*, edited by A. Janner, T. Janssen, and M. Boon (Springer-Verlag, Berlin, 1976), p. 395.

[5] L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13**, 1957–84 (1972).

A remark on the construction of quantum fields from Markov fields

YAO, Te Hai

Dublin Institute for Advanced Studies, School of Theoretical Physics, 10 Burlington Road, Dublin 4, Ireland
(Received 22 October 1976)

We give a proof of the existence of $\phi_0(g)$ as a continuous map from $\mathcal{S}(R^{d-1})$ to $\mathcal{L}(\tilde{H}^{k'}, \tilde{H}^{-k'})$, for k' sufficiently large, where $\tilde{H}^{k'}$ belongs to the scale associated with a certain Hilbert space $\tilde{H}$ and a certain self-adjoint operator (positive) $\tilde{H}$ in $\tilde{H}$, $\phi_0(g)$ being a sharp time field in the process of constructing a quantum field from an Euclidean field over $\mathcal{S}(R^d)$.

I. INTRODUCTION

Nelson[1] gives a method of explicitly constructing a quantum field satisfying all the Wightman axioms from a Euclidean field over $\mathcal{D}(R^d)$, satisfying the reflection principle, the ergodicity property, and the assumption that the field $\phi(f)$ is a continuous map from $\mathcal{D}(R^d)$ to $\mathcal{L}(K^k, K^l)$,[2] for some k and l, where K^k and K^l belong to the scale associated with a certain Hilbert space K and a certain self-adjoint operator (positive) K in K. However, his proof concerning the existence of a sharp time field $\phi_0(g)$ as a linear map from $\mathcal{S}(R^{d-1})$ to $\mathcal{L}(\tilde{H}^{k'}, \tilde{H}^{-k'})$ for k' sufficiently large, with $\tilde{H}^{k'}$ belonging to the scale associated with a subspace $\tilde{H}_-$ of K and a certain self-adjoint operator (positive) $\tilde{H}$ in $\tilde{H}$, is incorrect. For example, statements such as "Now any tempered distribution is a derivative of finite order of a function[3] with finite $\|\ \|_m$ norm," on p. 109, lines 25–7, and "To see this, let f_n be a sequence in $\mathcal{S}(R^d)$ with $\text{supp} f_n \subset Z^+$ which converges to f in $\|\ \|_m$ norm," on p. 110, lines 22–4, are false. Here we give a proof of a stronger result concerning $\phi_0(g)$ for the construction of a quantum field from a Euclidean field over $\mathcal{S}(R^d)$ satisfying the reflection principle, the ergodicity property, and the assumption that $\phi(f)$ is a continuous map from $\mathcal{S}(R^d)$ to $\mathcal{L}(K^k, K^l)$, where $\mathcal{S}(R^d)$ is assumed to be equipped with the usual topology. It then follows that the above fact also holds for the construction of a quantum field from a Euclidean field over $\mathcal{D}(R^d)$ satisfying the reflection principle, the ergodicity property, and the assumption that $\phi(f)$ is a continuous map from $\mathcal{D}(R^d)$ to $\mathcal{L}(K^k, K^l)$, provided that $\mathcal{D}(R^d)$ is equipped with the topology induced by the usual topology of $\mathcal{S}(R^d)$. We note that in the above statements the space $\mathcal{S}(R^{d-1})$ is again assumed to be equipped with the usual topology.

II. SHARP TIME FIELDS CONSTRUCTED FROM A EUCLIDEAN FIELD OVER $\mathcal{S}(R^d)$

We first note that a Euclidean field over $\mathcal{S}(R^d)$ is defined in exactly the same way as in Ref. 1 for Euclidean fields over $\mathcal{D}(R^d)$ except that $\mathcal{D}(R^d)$ is replaced by $\mathcal{S}(R^d)$, with $\mathcal{S}(R^d)$ equipped with the usual topology. We collect our results in two theorems.

Theorem I: There exists a sharp time field $\phi(g \otimes \delta)$ as an element of $\mathcal{L}(K^k, K^l)$ and it is a continuous map from $\mathcal{S}(R^{d-1})$ equipped with the usual topology to $\mathcal{L}(K^k, K^l)$, the space $\mathcal{L}(K^k, K^l)$ being defined in the same way as in Ref. 1, where $g \in \mathcal{S}(R^{d-1})$, δ is the delta distribution over $Z(R)^4$, for any Euclidean field ϕ over $\mathcal{S}(R^d)$ satisfying the properties mentioned in the Introduction.

Proof: We shall prove Theorem I along the following steps.

(1) We introduce the functions $\tilde{\nu}_a(\eta)$, $\eta \in R$, $a = 1, 2, \cdots$.

We first define

$$\tilde{s}(\eta) = e^{-\tilde{k}(\eta)},$$

$$\tilde{k}(\eta) = \left\{1 - \exp\left(-\frac{1}{(1-\eta^2)}\right)\right\} \times \exp\left\{-\frac{1}{\eta^2}\left[\exp\left(-\frac{1}{(1-\eta^2)}\right)\right]\right\}.$$

Then

$$\frac{d^l}{d\eta^l}\tilde{s}(\eta) = 0 \quad \text{for } \eta = 0, \quad \eta = 1, \quad l \geq 1.$$

$\tilde{s}(\eta)$ is monotonically decreasing as η increases from 0 to 1.

We then define

$$\tilde{s}_1(\eta) = \begin{cases} \tilde{s}(\eta) - \tilde{s}(1), & 1 \geq |\eta| \geq 0, \\ 0, & |\eta| > 1 \end{cases}$$

and

$$\tilde{\nu}_1(\eta) = \left[\frac{1}{\tilde{s}(0) - \tilde{s}(1)}\right] \tilde{s}_1(\eta).$$

The functions $\tilde{\nu}_a(\eta)$ are defined as

$$\tilde{\nu}_a(\eta) = \begin{cases} 1, & |\eta| < a - 1, \\ \tilde{\nu}_1(\eta - [a-1]), & a \geq \eta \geq a - 1, \\ \tilde{\nu}_1(\eta + [a-1]), & -a + 1 \geq \eta \geq -a, \\ 0, & |\eta| > a, \end{cases}$$

for $a = 2, 3, \cdots$.

(2) We introduce the topological spaces $(W(R^d), \tau_1)$ and $(Q(R^d), \tau_4)$.

We let

$$\gamma(l) = \max_{\eta \in R} \left| \frac{d^l}{d\eta^l} \tilde{\nu}_1(\eta) \right| = \max_{\eta \in R} \left| \frac{d^l}{d\eta^l} \tilde{\nu}_a(\eta) \right|, \quad l = 0, 1, 2, \cdots, \quad a = 1, 2, \cdots,$$

and let $\beta(l)$ be a sequence of numbers such that $1 \leq \beta(0)$

$\leq \beta(1) \leq \cdots$ and such that $\beta(l) \geq \gamma_i(l)$, $l = 0, 1, 2, \cdots$.

We introduce the space $W[-a, +a]$ of all functions f on R belonging to $\mathcal{S}(R)$ where $a = 1, 2, \cdots$, whose Fourier transform $\tilde{h}(\eta)$ is a finite linear combination of $\tilde{h}^a(\eta)$ belonging to $\mathcal{D}[-a, +a]$, the space of all functions belonging to $\mathcal{D}(R)$ with support contained in $[-a, +a]$ such that

$$\left| \frac{d^l}{d\eta^l} \tilde{h}^a(\eta) \right| \leq C(\tilde{h}_a)\beta(l), \quad l = 0, 1, 2, \cdots$$

and the products $\tilde{h}_\infty(\eta)\tilde{v}_j(\eta)$, where $\tilde{h}_\infty(\eta)$ is a bounded C^∞ function on R satisfying either $\tilde{h}_\infty(\eta) = e^{it\eta}$, $\infty > t > -\infty$ or

$$\left| \frac{d^l}{d\eta^l} \tilde{h}_\infty(\eta) \right| \leq C(\tilde{h}_\infty)\beta(l), \quad l = 0, 1, 2, \cdots,$$

where $C(\tilde{h}^a)$ and $C(\tilde{h}_\infty)$ are some constants depending on $\tilde{h}^a$ and $\tilde{h}_\infty$. We note that $\tilde{h}(\eta)$ belongs to $\mathcal{D}[-a, +a]$ and that either

$$\left| \frac{d^l}{d\eta^l} \{\tilde{h}_\infty(\eta)\tilde{v}_j(\eta)\} \right| \leq 2^l(1 + |t|)^l\beta(l)$$

for some t, or

$$\left| \frac{d^l}{d\eta^l} \{\tilde{h}_\infty(\eta)\tilde{v}_j(\eta)\} \right| \leq C(\tilde{h}_\infty)2^l[\beta(l)]^2.$$

We introduce, for each $h \in W[-a, +a]$, the norm

$$\|h\|_a = \int_{-a}^{a} d\eta \left(\sum_{l=0}^{\infty} \frac{\left| \frac{d^l}{d\eta^l} \tilde{h}(\eta) \right|^2}{[\beta(l)]^4 2^{2l}(l)!} \right) \frac{1}{\eta^2 + 1}.$$

We thus obtain the normed linear space $(W[-a, +a], \| \|_a)$. We let $\tau(a)$ be the topology in $W[-a, +a]$ defined by $\| \|_a$.

We let $W(R) = \cup_{a=1}^{\infty} W[-a, +a]$, and we define a topology τ_0 on the linear space $W(R)$ by the norm $\| \|_\infty$,

$$\|h\|_\infty = \int_{-\infty}^{\infty} d\eta \left(\sum_{l=0}^{\infty} \frac{\left| \frac{d^l}{d\eta^l} \tilde{h}(\eta) \right|^2}{[\beta(l)]^4 2^{2l}(l)!} \right) \frac{1}{\eta^2 + 1}$$

for $h \in W(R)$. We thus have the normed linear space $(W(R), \| \|_\infty)$ with the topology τ_0.

We now introduce a space $W(R^d)$ of all functions f on R^d whose Fourier transform $\tilde{f}$ is of the form $\tilde{f}(p) = \tilde{g}(p_1, \ldots, p_{d-1})\tilde{h}(p_d)$, where $p = (p_1, \ldots, p_d)$, $g \in \mathcal{S}(R^{d-1})$ and $h \in W(R)$, with $\tilde{}$ denoting Fourier transform. $W(R^d)$ becomes a linear topological space $(W(R^d), \tau_1)$ if addition $\dot{+}$ is defined as $(\widetilde{f_1 \dot{+} f_2})(p) = \{\tilde{g}_1(p_1, \ldots, p_{d-1}) + \tilde{g}_2(p_1, \ldots, p_{d-1})\}\{\tilde{h}_1(p_d) + \tilde{h}_2(p_d)\}$, multiplication by complex numbers is defined by $(\widetilde{cf})(p) = c\tilde{f}(p)$, and the topology τ_1 is defined as the coarsest topology on $W(R^d)$ for which the projections Π_1 and Π_2 from $(W(R^d), \tau_1)$ to $(\mathcal{S}(R^{d-1}), \tau_2^{(d-1)})$ and $(W(R), \tau_0)$ respectively, defined by

$$\Pi_1 f = g \in \mathcal{S}(R^{d-1}),$$

$$\Pi_2 f = h \in W(R)$$

for $\tilde{f}(p) = \tilde{g}(p_1, \ldots, p_{d-1})\tilde{h}(p_d)$ are respectively continuous, where $\tau_2^{(d-1)}$ is the usual topology on $\mathcal{S}(R^{d-1})$. $(W(R^d), \tau_1)$ is just the product of the linear topological spaces $(\mathcal{S}(R^{d-1}), \tau_2^{(d-1)})$ and $(W(R), \tau_0)$.

We now introduce the space $Q(R)$ of all distributions h on $Z(R)$ (equipped with the usual topology[4]) whose Fourier transform $\tilde{h}(\eta)$ is a finite linear combination of bounded C^∞ functions $\tilde{h}_\infty(\eta)$ satisfying either

$$\tilde{h}_\infty(\eta) = e^{it\eta}, \quad \infty > t > -\infty$$

or

$$\left| \frac{d^l}{d\eta_l} \tilde{h}_\infty(\eta) \right| \leq C(\tilde{h}_\infty)\beta(l), \quad l = 0, 1, 2, \cdots$$

and products $\tilde{h}_\infty(\eta)\tilde{v}_a(\eta)$, $a = 1, 2, \cdots$. We introduce the norm $\| \|_\infty$ on $Q(R)$ as follows:

$$\|h\|_\infty = \int_{-\infty}^{\infty} d\eta \left\{ \sum_{l=0}^{\infty} \frac{\left| \frac{d^l}{d\eta^l} \tilde{h}(\eta) \right|^2}{[\beta(l)]^4 2^{2l}(l)!} \right\} \frac{1}{\eta^2 + 1}$$

for $h \in Q(R)$, and we let τ_3 be the topology defined on $Q(R)$ by this norm. We note that $W(R)$ is a subspace of the normed linear space $(Q(R), \| \|_\infty)$ with the topology τ_3.

We finally introduce the space $Q(R^d)$, consisting of all distributions f on $Z(R^d)$ whose Fourier transform $\tilde{f}$ is of the form $\tilde{f}(p) = \tilde{g}(p_1, \ldots, p_{d-1})\tilde{h}(p_d)$, where $g \in \mathcal{S}(R^{d-1})$ and $h \in Q(R)$. $Q(R^d)$ becomes a linear topological space if addition $\dot{+}$ is defined by $(\widetilde{f_1 \dot{+} f_2})(p) = \{\tilde{g}_1(p_1, \ldots, p_{d-1}) + \tilde{g}_2(p_1, \ldots, p_{d-1})\} \times \{\tilde{h}_1(p_d) + \tilde{h}_2(p_d)\}$, multiplication by a complex number is defined by $(\widetilde{cf})(p) = c\tilde{f}(p)$ and a topology τ_4 is defined as the coarsest topology for which the projections Π_1^∞ and Π_2^∞ from $(Q(R^d), \tau_4)$ to $(\mathcal{S}(R^{d-1}), \tau_2^{(d-1)})$ and $(Q(R), \tau_3)$ are separately continuous, with

$$\Pi_1^\infty f = g \in \mathcal{S}(R^{d-1}),$$

$$\Pi_2^\infty f = h \in Q(R),$$

for $\tilde{f}(p) = \tilde{g}(p_1, \ldots, p_{d-1})\tilde{h}(p_d)$. Again the linear topological space $(Q(R^d), \tau_4)$ is just the product of the linear topological spaces $(\mathcal{S}(R^{d-1}), \tau_2^{(d-1)})$ and $(Q(R), \tau_3)$.

(3) *Lemma*: If $h_m \to h$ in $(W(R), \| \|_\infty)$, then h_m and h are in some $W[-a, +a]$.

Proof: We introduce the following scalar product $\langle , \rangle_a$ in $(W[-a, +a], \| \|_\infty)$,

$$\langle h_1, h_2 \rangle_a = \int_{-a}^{a} d\eta \left[\sum_{l=0}^{\infty} \frac{\left(\frac{d^l}{d\eta^l} \tilde{h}_1(\eta) \right)^* \left(\frac{d^l}{d\eta^l} \tilde{h}_2(\eta) \right)}{[\beta(l)]^4 2^{2l}(l)!} \right] \frac{1}{\eta^2 + 1}$$

for $h_1, h_2 \in W[-a, +a]$. Let $W^c(a)$ be the completion of $(W[-a, +a], \| \|_a)$. Then if $h(\eta) \in W^c(a)$, we have $h(\eta) = 0$ for $|\eta| > a$. Further we have $W^c(a)$ being a complete subspace of $(W^c(b), \| \|_b)$, for $b > a$. We introduce the normed linear space $(\cup_{a=1,2\ldots} W^c(a), \| \|_\infty)$.

We now let $h_m \to h$ in $(W(R), \| \|_\infty)$, and suppose for all n there is a h_m with $h_m \in X_n$ $[X_n = W^c(n)]$. Then one can construct a subsequence $t_i = h_{m(i)}$ and spaces $Y_i = W^c(n_i)$ with $t_i \in Y_{i+1}, t_i \notin Y_i$, $i = 1, 2, \cdots$, $\{n_i\}$ being the increasing sequence of positive integers. We let l_1 be a functional on $Y_1 \cup Z_1 \cup Z_2 \cup \cdots$ such that

$$l_1(s_1) = 0 \quad \text{for } s_1 \in Y_1,$$

$$l_1(\omega_1) = 1 \quad \text{for } \omega_1 \in Z_1,$$

where Z_i is the orthogonal complement of Y_i in $(Y_{i+1}, \| \|_{n_{i+1}})$.

We now let l_2 be a functional on $Y_2 \cup Z_2 \cup Z_3 \cup \cdots$ such that

$$l_2(s_2) = 0 \quad \text{for } s_2 \in Y_2,$$

$$l_2(\omega_2) = 2 - l_1(\omega_2) \quad \text{for } \omega_2 \in Z_2.$$

Similarly, we let l_n be a functional on $Y_n \cup Z_n \cup Z_{n+1} \cup \cdots$ such that

$$l_n(s_n) = 0 \quad \text{for } s_n \in Y_n,$$

$$l_n(\omega_n) = n - \sum_{k=1}^{n-1} l_k(\omega_n) \quad \text{for } \omega_n \in Z_n.$$

We now form the functional $l = \sum_{n=1}^{\infty} l_n$ on $Y_1 \cup Z_1 \cup Z_2 \cup \cdots$. We have

$$\begin{aligned} l(\omega_m) &= \sum_{n=1}^{\infty} l_n(\omega_m) \\ &= \sum_{n=1}^{m} l_n(\omega_m) \\ &= \left\{ m - \sum_{k=1}^{m-1} l_n(\omega_m) \right\} + \sum_{k=1}^{m-1} l_n(\omega_m) \\ &= m \quad \text{for } \omega_m \in Z_m, \end{aligned}$$

i.e., l maps Z_m into m. We have

$$\begin{aligned} Y_m &= Y_{m-1} \oplus Z_{m-1} \\ &= Z_{m-1} \oplus Z_{m-2} \oplus Y_{m-2} = \cdots \\ &= Z_{m-1} \oplus Z_{m-2} \oplus \cdots \oplus Z_0 \quad (Z_0 = Y_1). \end{aligned}$$

We extend l from $Z_0 \cup Z_1 \cup Z_2 \cup \cdots \cup Z_{m-1}$ to Y_m by linearity, i.e.,

$$l(s_m) = l(s_{m,0}) + l(s_{m,1}) + \cdots + l(s_{m,m-1}),$$

where

$$s_{m,i} \in Z_i \quad \text{and} \quad s_m = \sum_{i=0}^{m-1} s_{m,i}.$$

l is therefore defined on

$$\begin{aligned} \bigcup_{a=1,2\cdots} W^c(a) &= \bigcup_{m=1,2,\cdots} Y_m \\ &= \bigcup_{m=1,2,\cdots} (Z_0 \oplus Z_1 \oplus \cdots \oplus Z_{m-1}). \end{aligned}$$

We know that Z_i is a closed subset of $(Y_{i+1}, \| \|_{n_{i+1}})$. Further, since $Y_{i+1} = Z_i \oplus Y_i$, and $(Y_i, \| \|_{n_i})$ and $(Y_{i+1}, \| \|_{n_{i+1}})$ are complete, we conclude that $(Z_i, \| \|_{n_{i+1}})$ is complete and that Y_{i+1} is a closed subset of $(\cup_{a=1,2,\cdots} W^c(a), \| \|_\infty)$. Also, it can be shown that $Z_i \oplus Z_j$ is closed in $(\cup_{a=1,2,\cdots} W^c(a), \| \|_\infty)$.

Hence the inverse image of every closed set in R under the map l from $\cup_m Y_m = \cup_{a=1,2,\cdots} W^c(a)$ is a closed set in $(\cup_{a=1,2,\cdots} W^c(a), \| \|_\infty)$. l is therefore a continuous map from $(\cup_{a=1,2,\cdots} W^c(a), \| \|_\infty)$ to R. However, since $l(\omega_m) = m$ for $\omega_m \in Z_m$, the map l is unbounded. Hence the map l does not exist. Hence the assumption "suppose for all n, there is a h_m with $h_m \in X_n$" is wrong. Hence all h_m belong to some X_n. Hence h also belongs to that X_n since $(X_n, \| \|_n)$ is complete. Now h_m, for all m, and h, belong to $X_n \cap W(R)$. Hence h_m, for all m, and h, belong to $W[-a,+a]$ for some a.

(4) *Lemma*: If $\{f_n, f \in W(R^d)\}$, $n = 1, 2, \cdots$, and $f_n \to f$ in the τ_1 topology, then $f_n \to f$ in the usual topology $\tau_2^{(d)}$ defined by the norms $\| \|_{(d),\beta,r}$ (see below for complete definition of these norms).

Proof: We now let

$$\tilde{f}_n(p) = \tilde{g}_n(p_1, \ldots, p_{d-1}) \tilde{h}_n(p_d), \quad n = 1, 2, \cdots$$

and

$$\tilde{f}(p) = \tilde{g}(p_1, \ldots, p_{d-1}) \tilde{h}(p_d).$$

Since $f_n \to f$ in the τ_1 topology, we have $g_n \to g$ in the $\tau_2^{(d-1)}$ topology and $h_n \to h$ in the τ_0 topology. Hence all h_n and h are in $W[-a,+a]$, for some a. We have

$$\| f_n - f \|_{(d),\beta,r} = \sup_{x \in R^d} \left| (1 + |x|^2)^\beta D^{(r)} (f_n - f)(x) \right|$$

where $\| \|_{(d),\beta,r}$ are the norms defining the $\tau_2^{(d)}$ topology on $\mathcal{S}(R^d)$, with $\beta = 0, 1, 2, \cdots$, $r = \{r(1), \ldots, r(d)\}$, $r(i) = 0, 1, 2, \cdots$, $i = 1, 2, \ldots, d$,

$$D^{(r)} = \frac{\partial^{r(1)}}{\partial x_1^{r(1)}} \cdots \frac{\partial^{r(d)}}{\partial x_d^{r(d)}}, \quad x = (x_1, \ldots, x_d) \in R^d.$$

Hence

$$\begin{aligned} \| f_n - f \|_{(d),\beta,r} &= \sup_{x \in R^d} \left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dp_1 \cdots dp_{d-1} dp_d e^{ip_1 x_1} \cdots e^{ip_d x_d} \left(1 - \frac{\partial^2}{\partial p_1^2} - \cdots - \frac{\partial^2}{\partial p_{d-1}^2} - \frac{\partial^2}{\partial p_d^2} \right)^\beta [p_1^{r(1)} \cdots p_{d-1}^{r(d-1)} p_d^{r(d)} \right. \\ &\quad \left. \times \{ \tilde{f}_n(p) - \tilde{f}(p) \}] \right| \leq \sup_{x \in R^d} \left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dp_1 \cdots dp_{d-1} dp_d e^{ip_1 x_1} \cdots e^{ip_d x_d} \left(1 - \frac{\partial^2}{\partial p_1^2} - \cdots - \frac{\partial^2}{\partial p_{d-1}^2} - \frac{\partial^2}{\partial p_d^2} \right)^\beta \right. \\ &\quad \left. \times [p_1^{r(1)} \cdots p_{d-1}^{r(d-1)} p_d^{r(d)} \{ \tilde{g}_n(p_1 \cdots p_{d-1}) \tilde{h}_n(p_d) - \tilde{g}(p_1 \cdots p_{d-1}) \tilde{h}_n(p_d) \}] \right| \\ &\quad + \sup_{x \in R^d} \left| \int_{-\infty}^{\infty} dp_1 \cdots \int_{-\infty}^{\infty} dp_1 \cdots dp_{d-1} dp_d e^{ip_1 x_1} \cdots e^{ip_d x_d} \left(1 - \frac{\partial^2}{\partial p_1^2} - \cdots - \frac{\partial^2}{\partial p_{d-1}^2} - \frac{\partial^2}{\partial p_d^2} \right)^\beta \right. \\ &\quad \left. \times [p_1^{r(1)} \cdots p_{d-1}^{r(d-1)} p_d^{r(d)} \{ \tilde{g}(p_1 \cdots p_{d-1}) \tilde{h}_n(p_d) - \tilde{g}(p_1 \cdots p_{d-1}) \tilde{h}(p_d) \}] \right| \\ &\leq \sup_{x \in R^d} \left\{ \sum_{s=0}^{\beta} C_s^\beta \left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dp_1 \cdots dp_{d-1} e^{ip_1 x_1} \cdots e^{ip_{d-1} x_{d-1}} \left(1 - \sum_{i=1}^{d-1} \frac{\partial^2}{\partial p_i^2} \right)^s [p_1^{r(1)} \cdots p_{d-1}^{r(d-1)} \{ \tilde{g}_n(p_1 \cdots p_{d-1}) \right. \right. \\ &\quad \left. \left. - \tilde{g}(p_1 \cdots p_{d-1}) \}] \right| \left| \int_{-\infty}^{\infty} dp_d e^{ip_d x_d} \left(\frac{\partial^2}{\partial p_d^2} \right)^{\beta - s} [p_d^{r(d)} \tilde{h}_n(p_d)] \right| \right\} \end{aligned}$$

$$+\sup_{x\in R^d}\left\{\sum_{s=0}^{\beta} C_s^{\beta}\left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} dp_1\cdots dp_{d-1}e^{ip_1x_1}\cdots e^{ip_{d-1}x_{d-1}}\left(1-\sum_{i=1}^{d-1}\frac{\partial^2}{\partial p_i^2}\right)^s[p_1^{r(1)}\cdots p_{d-1}^{r(d-1)}\tilde{g}(p_1\cdots p_{d-1})]\right|\right.$$

$$\left.\times\left|\int_{-\infty}^{\infty} dp_d e^{ip_dx_d}\left(\frac{\partial^2}{\partial p_d^2}\right)^{\beta-s}[p_d^{r(d)}\{\tilde{h}_n(p_d)-\tilde{h}(p_d)\}]\right|\right\}$$

$$\leq \sup_{(x_1,\cdots,x_{d-1})\in R^{d-1}}\left\{\left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} dp_1\cdots dp_{d-1}e^{ip_1x_1}\cdots e^{ip_{d-1}x_{d-1}}\left(1-\sum_{i=1}^{d-1}\frac{\partial^2}{\partial p_i^2}\right)^s[p_1^{r(1)}\cdots p_{d-1}^{r(d-1)}\{\tilde{g}_n(p_1\cdots p_{d-1})\right.\right.$$

$$\left.\left.-\tilde{g}(p_1\cdots p_{d-1})\}]\right|\right\}\times\text{const}+\sup_{(x_1\cdots x_{d-1})\in R^{d-1}}\left\{\sum_{s=0}^{\beta} C_s^{\beta}\left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} dp_1\cdots dp_{d-1}e^{ip_1x_1}\cdots e^{ip_{d-1}x_{d-1}}\left(1-\sum_{i=1}^{d-1}\frac{\partial^2}{\partial p_i^2}\right)^s\right.\right.$$

$$\left.\left.\times[p_1^{r(1)}\cdots p_{d-1}^{r(d-1)}\tilde{g}(p_1\cdots p_{d-1})]\right|\right\}\int_{-a}^{a} dp_d\sum_{l=0}^{L} A_l\left|\frac{d^l}{dp_d^l}\{\tilde{h}_n(p_d)-\tilde{h}(p_d)\}\right|(p_d^2+1)^{\alpha},$$

since all h_n and h belong to $W[-a,+a]$, where C_s^{β}, L, α, and A_l are some positive integers.

Hence

$$\|f_n-f\|_{(d),\beta,r}\leq\left\{\sum_{(\beta',r')\in\text{finite set}} A(\beta',r')\|g_n-g\|_{(d-1),\beta',r'}\right\}\times\text{const}+\left\{\sum_{(\beta',r')\in\text{finite set}} A(\beta',r')\|g\|_{(d-1),\beta',r'}\right\}$$

$$\times\sum_{l=0}^{L} A_l\left\{\int_{-a}^{a} dp_d\left|\frac{d^l}{dp_d^l}\{\tilde{h}_n(p_d)-\tilde{h}(p_d)\}\right|^2\right\}^{1/2}\left\{\int_{-a}^{a} dp_d(p_d^2+1)^{2\alpha}\right\}^{1/2}\xrightarrow[n\to\infty]{}0,$$

where $A(\beta',r')$ are constants, since

$$\|g_n-g\|_{(d-1),\beta',r'}\xrightarrow[n\to\infty]{}0$$

for each fixed β',r', and since again

$$\sup_{p_d\in R}\left|\frac{d^l}{dp_d^l}\{\tilde{h}_n(p_d)-\tilde{h}(p_d)\}\right|\xrightarrow[n\to\infty]{}0,\quad l=0,1,2,\cdots.$$

Consequently $f_n\to f$ in the usual topology $\tau_2^{(d)}$.

(5) *Lemma*: $W(R^d)$ is dense in $(Q(R^d),\tau_4)$.

Proof: Since $(Q(R^d),\tau_4)$ is the product of $(S(R^{d-1}),\tau_2^{(d-1)})$ and $(Q(R),\tau_3)$, a neighborhood basis of an arbitrary $f=(g,h)$ of $(Q(R^d),\tau_4)$, $g\in S(R^{d-1})$, $h\in Q(R)$, is given by the family of all sets of the type B, $B=B_1\times B_2$, where B_1 is a neighborhood of g in $(S(R^{d-1}),\tau_2^{(d-1)})$ and B_2 is a neighborhood of h in $(Q(R),\tau_3)$. If $f=(g,h)$ with $h\notin W(R)$, then the sequence of elements $f_n=(g,h_n)$ with $\tilde{h}_n(p_d)=\tilde{h}_\infty(p_d)\tilde{\nu}_n(p_d)$ converges to f in the τ_4 topology, with $f_n\in W(R^d)$. This is because $\|h_n-h_\infty\|_\infty\xrightarrow[n\to\infty]{}0$. If $f=(g,h)$ with $h\in W(R)$, then of course $f\in W(R^d)$. Hence $W(R^d)$ is dense in $(Q(R^d),\tau_4)$.

(6) From (4) and the assumption on $\phi(f)$, we know that $\phi(f)$, for $f\in W(R^d)$, is a continuous map from $(W(R^d),\tau_1)$ to $L(K^k,K^l)$. Now let $f_n\to f$ in $(Q(R^d),\tau_4)$, with $f_n\in W(R^d)$. Then $\{f_n\}$ is a fundamental sequence in $(W(R^d),\tau_1)$. Hence $\phi(f_n)$ is also a fundamental sequence in $L(K^k,K^l)$. Consequently $\phi(f_n)$ approaches a limit $\phi(f)$ as $n\to\infty$. Hence since the space $L(K^k,K^l)$ is a regular space, the map ϕ can be extended to a continuous map from $(Q(R^d),\tau_4)$ to $L(K^k,K^l)$.[5] In particular, this gives a definition of $\phi(g\otimes\delta_t)$, for $g\in S(R^{d-1})$, $\delta_t(\eta)=\delta(\eta-t)$.

(7) Also, let $g_n\to g$ in the usual topology of $S(R^{d-1})$, then $g_n\otimes\delta\to g\otimes\delta$ in the τ_4 topology of $Q(R^d)$. Hence $\phi(g_n\otimes\delta)\to\phi(g\otimes\delta)$ in $L(K^k,K^l)$. Consequently

$$\phi:\ g\to\phi(g\otimes\delta)$$

is a continuous map from $S(R^{d-1})$ to $L(K^k,K^l)$.

Theorem I is therefore proven.

We remark at this point that

$$\phi(g\otimes h)=\int\phi(g\otimes\delta_t)h(t)dt$$

as elements of $L(K^k,K^l)$, for $h\in S(R)$, $t\in R$.

Theorem II. Let λ be a positive integer, and let $\tilde{H}$ be the Hilbert space $\tilde{H}=D(X_1^\lambda)\cap\cdots\cap D(X_{d-1}^\lambda)\cap H$ with the norm

$$\|u\|_{\tilde{H}}^2=\|(1+|X_1|^\lambda)u\|^2+\cdots+\|(1+|X_{d-1}|^\lambda)u\|^2.$$

Then the restriction $\tilde{H}$ of H to the domain $D(\tilde{H})$ of all μ in $\tilde{H}$ such that Hu is in $\tilde{H}$ is a self-adjoint positive operator in $\tilde{H}$. Let $\tilde{H}^\nu$ be the scale associated with the operator $\tilde{H}$ in $\tilde{H}$, then there exists a sharp time field $\phi_0(g)$ as an element of $L(\tilde{H}^{k'},\tilde{H}^{-k'})$, for g belonging to $S(R^{d-1})$ and some k' sufficiently large. Further, $\phi_0(g)$ is a continuous map from $S(R^{d-1})$, equipped with the usual topology, to $L(\tilde{H}^{k'},\tilde{H}^{-k'})$. In the above, the symbols $X_1,\ldots,X_{d-1}$, $D(X_1^\lambda),\ldots,D(X_{d-1}^\lambda)$, H, H, and $\|\ \|$ have the same meaning as in Ref. 1, except that the test function space of the field is $S(R^d)$ instead of $D(R^d)$.

Proof: Here we shall skip over the proof of the statement that $\tilde{H}$ is a positive self-adjoint operator in the Hilbert space $\tilde{H}$. We demonstrate the remaining parts of the Theorem in two steps.

(1) Let $S(R^d)_+ = \{g\otimes h \in S(R^d),\ g\in S(R^{d-1}),\ h\in S(R),\ \mathrm{supp}h\in[0,\infty)\}$. Then there exists a sequence $g\otimes h_n$ belonging to $S(R^d)_+$ such that $g\otimes h_n \to g\otimes\delta$ in the τ_4 topology.

Proof: We introduce the functions $h_n(\xi)$, for $n=1,2,\cdots$, as follows:

$$h_n(\xi)=\begin{cases} H_n e^{-1/\xi}e^{-1/(1-n\xi)}e^{-n\xi^2}, & \text{for } \dfrac{1}{n}\geqslant\xi\geqslant 0,\\ 0, \quad \xi<0,\ \xi>\dfrac{1}{n}. \end{cases}$$

Then $h_n(\xi)\in S(R)_+$, $S(R)_+=\{h\in S(R),\ \mathrm{supp}h\in[0,\infty)\}$ and $\tilde{h}_n(\eta)\in S(R)\in Z(R)'$, the adjoint space of $Z(R)$. In the above H_n is given by

$$H_n=\left(\int_0^{1/n} d\xi\ e^{-1/\xi}e^{-1/(1-n\xi)}e^{-n\xi^2}\right)^{-1},$$

so we have

$$\int_{-\infty}^{\infty} d\xi\, h_n(\xi)=1 \quad \text{for all } n.$$

$\tilde{h}_n(\eta)$ also belongs to $S(R)\subset Z(R)'$ (of course bounded and infinitely differentiable). Also we have

$$\frac{d^l}{d\eta^l}\tilde{h}_n(\eta)=\int_0^{1/n} d\xi\, e^{i\xi\eta}(i\xi)^l e^{-1/\xi}e^{-1/(1-n\xi)}e^{-n\xi^2}\Big/\int_0^{1/n} d\xi\, e^{-1/\xi}e^{-1/(1-n\xi)}e^{-n\xi^2},$$

and

$$\left|\frac{d^l}{d\eta^l}\tilde{h}_n(\eta)\right|\leqslant 1\leqslant\beta(l),\quad l=0,1,2,\cdots,\ n=1,2,\cdots.$$

Hence $h_n\in Q(R)$, for $h=1,2,\cdots$.

Now given any $b>0$, and $1>\gamma>0$, we choose $n\gg b$, i.e., $1\gg b/n$. Then for $|\eta|<b$, we have $|\xi\eta|\leqslant(1/n)|\eta|<b/n \ll 1$ for ξ in the interval $[0,1/n]$. Hence we have for $|\eta|<b$ and $n\gg b$, the inequality $1>|\tilde{h}_n(\eta)|>1-\gamma$. Further, we have

$$\left|\frac{d^l}{d\eta^l}\tilde{h}_n(\eta)\right|\leqslant\left(\frac{1}{n}\right)^l\ll(\gamma)^l,\quad l=1,2,\cdots \text{ for } n\gg 1/\gamma.$$

We now show that the sequence $h_n\to\delta$, the δ distribution in the τ_3 topology, i.e., we want to show that given any $\epsilon>0$, there exists an M such that for all $n\geqslant M$, we have $\|\tilde{\delta}-\tilde{h}_n\|_\infty<\epsilon$.

We proceed as follows.

Since

$$\left|\frac{d^l}{d\eta^l}\{\tilde{\delta}(\eta)-\tilde{h}_n(\eta)\}\right|\leqslant 2,\quad l=0,1,2,\cdots,\ n=1,2,\cdots,$$

we have

$$\sum_{\alpha=0}^{\infty}\left|\frac{d^l}{d\eta^l}\{\tilde{\delta}(\eta)-\tilde{h}_n(\eta)\}\right|^2\Big/[\beta(l)]^4 2^{2l}(l)!$$

$$\leqslant\sum_{\alpha=0}^{\infty}\frac{4}{[\beta(l)]^4 2^{2l}(l)!}\leqslant\text{const.}$$

Hence there exists an R, $\infty>R>0$, such that

$$\left[\int_{-\infty}^{-R}+\int_R^{\infty}\right]d\eta\left\{\sum_{l=0}^{\infty}\left(\left|\frac{d^l}{d\eta^l}\{\tilde{\delta}(\eta)-\tilde{h}_n(\eta)\}\right|^2\Big/[\beta(l)]^4 2^{2l}(l)!\right)\frac{1}{\eta^2+1}<\frac{\epsilon}{2}.$$

We choose $1>\gamma>0$ such that

$$\int_{-R}^{R}d\eta\left(\sum_{l=0}^{\infty}\frac{\gamma^2}{[\beta(l)]^4 2^{2l}(l)!}\right)\frac{1}{\eta^2+1}<\frac{\epsilon}{2},$$

i.e.,

$$\gamma^2<\frac{\epsilon}{2}\Big/\int_{-R}^{R}d\eta\left(\sum_{l=0}^{\infty}\frac{1}{[\beta(l)]^4 2^{2l}(l)!}\right)\frac{1}{\eta^2+1}.$$

Further, we choose M such that $M\gg 1/\gamma$. Then for $n\geqslant M$, we have

$$|\tilde{\delta}(\eta)-\tilde{h}_n(\eta)|<\gamma,\quad\left|\frac{d^l}{d\eta^l}[\tilde{\delta}(\eta)-\tilde{h}_n(\eta)]\right|\ll\gamma^l<\gamma,\quad l=1,2,\cdots,$$

so we have

$$\left|\frac{d^l}{d\eta^l}[\tilde{\delta}(\eta)-\tilde{h}_n(\eta)]\right|<\gamma,\quad l=0,1,2,\cdots,\ n\geqslant M.$$

Hence we have

$$\int_{-R}^{R}d\eta\left[\sum_{l=0}^{\infty}\left(\left|\frac{d^l}{d\eta^l}[\tilde{\delta}(\eta)-\tilde{h}_n(\eta)]\right|^2\Big/[\beta(l)]^4 2^{2l}(l)!\right)\right]\frac{1}{\eta^2+1}$$

$$\leqslant\int_{-R}^{R}d\eta\left(\sum_{l=0}^{\infty}\frac{\gamma^2}{[\beta(l)]^4 2^{2l}(l)!}\right)\frac{1}{\eta^2+1}<\frac{\epsilon}{2}.$$

Consequently, we have, for $n\geqslant M$

$$\|\tilde{\delta}-\tilde{h}_n\|_\infty=\int_{-\infty}^{\infty}d\eta\left[\sum_{l=0}^{\infty}\left(\left|\frac{d^l}{d\eta^l}[\tilde{\delta}(\eta)-\tilde{h}_n(\eta)]\right|^2\Big/[\beta(l)]^4 2^{2l}(l!)\right)\right]\frac{1}{\eta^2+1}<\epsilon.$$

Hence $h_n\xrightarrow[h\to\infty]{}\delta$ in the τ_3 topology, with δ and $h_n\in Q(R)$ and $h_n\in S_+(R)$, for all n. Part (1) is therefore proven.

(2) Following Nelson, we have for u,v belonging to $D(\tilde{H}^k)$ with k so large that $\phi(f)$ is continuous from K^{2k} to K^{-2k},

$$|\langle\check{v},\phi(g\otimes\delta)\hat{u}\rangle_{K^{-2k}}|$$

$$\leqslant\text{const}\,|||\,\phi(g\otimes\delta)\,|||_{K^{2k},K^{-2k}}\|v\|_{\tilde{H}^{k'}}\|u\|_{\tilde{H}^{k'}},$$

for sufficiently large $k'>0$. We denote the bilinear form

$$\phi(g\otimes\delta)(u,v)=\langle\check{v},\phi(g\otimes\delta)\hat{u}\rangle_{K^{-2k}}$$

by $\phi_0(g)$. $\phi_0(g)$ is consequently a bounded operator from $H^{k'}$ to $H^{-k'}$. Hence we have

$$|\langle\check{v},\phi(g\otimes\delta)\hat{u}\rangle_{K^{-2k}}|\leqslant\sum_{\{\gamma,\mu\}\in B}C_{\gamma,\mu}\|g\|_{\gamma,\mu}\|v\|_{H,k}\|u\|_{\tilde{H}^k}$$

where B is a finite set of $\{\gamma,\mu\}$, $C_{\gamma,\mu}$ are some constants, and where

$$\|g\|_{\gamma,\mu}=\sup_{x\in R^{d-1}}|(1+|x|^2)^\gamma D^{(\mu)}g(x)|,$$

$$\gamma = 0, 1, 2, \cdots,$$

$$\mu = \{\mu^{(1)}, \mu^{(2)}, \ldots \mu^{(d-1)}\},$$

$$\mu^{(i)} = 0, 1, 2, \cdots, \quad i = 1, 2, \ldots, d-1,$$

$$D^{(\mu)} = \frac{\partial^{\mu(1)}}{\partial x^{\mu(1)}} \frac{\partial^{\mu(2)}}{\partial x^{\mu(2)}} \cdots \frac{\partial^{\mu(d-1)}}{\partial x^{\mu(d-1)}}.$$

Hence the Theorem is proved using Ref. 6.

[1]E. Nelson, J. Funct. Anal. 12, 97 (1973).
[2]Here $\mathcal{L}(F, G)$ denotes the Banach space of all bounded linear operators from a Hilbert space F to another Hilbert space G.
[3]Presumably a function belonging to the completion of $\mathcal{S}(R^d)$ w.r.t. $\| \ \|_m$.
[4]I.M. Gelfand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.
[5]H. Schubert, *Topology* (Macdonald, London, 1968), p. 64, Theorem 2.
[6]L. Garding and J. L. Lions, Nuovo Cimento Suppl. 14, 9 (1959), Theorem 5.1.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*