A note on stochastic dynamics in the state space of a commutative C* algebra

Peter M. Alberti and Armin Uhlmann
Department of Physics and NTZ of Karl-Marx University Leipzig, DDR-701 Leipzig, German Democratic Republic

(Received 10 April 1980; accepted for publication 19 September 1980)

In this paper a functional characterization of stochastic evolutions within the state spaces of commutative C^* algebras with identity is derived. Consequences concerning the structure of those linear evolution equations (master equations) that give occassion to stochastic evolutions are discussed. In part, these results generalize facts which are well known from the finite-dimensional classical case. Examples are given and some important particularities of the W^* case are developed.

PACS numbers: 02.10.+w.

1. BASIC NOTIONS AND TOOLS

Let A denote a commutative C^* algebra with unity $\mathbf{1}$. Whereas the C^* norm of $x \in A$ will be marked by $\|x\|_\infty$, the functional norm of an element $\omega \in A^*$ of the topological dual A^* of A will be denoted by $\|\omega\|_1$. As usually, the state space of A, S_A in sign, is defined as the convex set $S_A = \{\omega \in A^*: \ \omega(x^*x) \geq 0, \ \omega(\mathbf{1}) = 1\}$.
Let $\mathrm{B}(A^*)$ denote the linear space of all bounded linear maps acting from the Banach space A^* into A^*. Then $\|\circ\|_1$ on A^* induced norm on elements of $\mathrm{B}(A^*)$ will be denoted by the same symbol $\|\circ\|_1$. Being equipped with this topology, $\mathrm{B}(A^*)$ becomes a Banach algebra. An element $\Phi \in \mathrm{B}(A^*)$ is said to be *stochastic* if $\Phi\omega(\mathbf{1}) = \omega(\mathbf{1})$ and Φ is positive, i.e., $\Phi\omega \in A_+^*$ whenever $\omega \in A_+^*$, where A_+^* means the positive cone in A^*. The convex set of all stochastic maps with respect to A will be denoted by $\mathrm{ST}(A)$. Let $\{\Phi_\lambda\}_{\lambda \in I}$ be a net of elements of $\mathrm{B}(A^*)$. Then, we say that the net is *weakly* converging towards $\Phi \in \mathrm{B}(A^*)$, $\Phi_\lambda \xrightarrow{w} \Phi$, if $\lim_\lambda (\Phi_\lambda(\omega))(x) = (\Phi\omega)(x)$ for every $\omega \in A^*$ and each element x of A. It is an important fact that $\mathrm{ST}(A)$ is *weakly compact*. Stochastic maps are exactly those linear transformations on A^* that throw states into states. This property makes them very useful for the *abstract* description of *dynamical evolutions* of systems (in our case *classical* systems, for only commutative C^* algebras will be under consideration throughout this paper).

In many applications we will meet commutative W^*-algebras. Then, by standard knowledge, we may identify the commutative W^* algebra A with $L^\infty(\Omega, \mu)$, for a suitable measure space Ω with measure μ. In this context, besides the whole set of states, there is the set of *normal* states deserving our interest. These states belong to the predual A_* of A. In the sense of the canonical identification from above, A_* can be identified with $L^1(\Omega, \mu)$. Thus, normal states correspond to *probability distributions* over certain measure spaces, and this is the frame in which problems of classical statistical mechanics usually will be dealt with. In this situation, the set of linear transformations that take normal states into normal ones will be referred to as $\mathrm{ST}_w(A)$.

Let f denote a *real-valued* function on the positive cone R_+^n of n-tuples of non-negative reals:

$$f: \ \mathrm{R}_+^n \ni (s_1, \ldots, s_n) \mapsto f(s_1, \ldots, s_n) \ .$$

We will refer to f as an *h-convex* function (of order n) if f is *finite*, *continuous*, *convex*, and *homogeneous of first degree* on R_+^n. We remark that homogeneity and convexity imply that h-convex functions are *subadditive* on R_n^+. By means of h-convex function f we define a functional S_f on n-tuples of positive linear forms of A by

$$S_f(\omega_1, \ldots, \omega_n) = \sup_{\{a_k\}} \sum_k f(\omega_1(a_k), \ldots, \omega_n(a_k)) , \quad (1.1)$$

where the *sup* runs through the set of decompositions $\{a_k\}$ of $\mathbf{1}$ into finitely many positive elements a_k of A (i.e., $\sum_k a_k = \mathbf{1}$). S_f in this situation will be spoken of as an *h-convex functional* (of order n), and S_f is called *positive* if f is non-negative on R_+^n.

In order to get a better idea of an h-convex functional S_f, we will take notice of

Proposition 1.1: Let $\omega_1(x) \ldots w_n(x) \in L^1(\Omega, \mu)$ correspond to positive normal functionals $\omega_1, \ldots, \omega_n$ of the W^* algebra $L^\infty(\Omega, \mu)$. Then, every positive h-convex functional S_f can be represented by

$$S_f(\omega_1, \ldots, \omega_n) = \int_\Omega f(\omega_1(x), \ldots, \omega_n(x))\, d\mu(x) \ . \quad (1.2)$$

A proof is given in the Appendix to this paper. The importance of h-convex functionals (of arbitrary order) is due to the following result (see Ref. 1):

Theorem 1.2: Let $\omega_1, \ldots, \omega_n, \ \sigma_1, \ldots, \sigma_n$ be states of the commutative C^* algebra A with identity. Then, there exists a stochastic map $\Phi \in \mathrm{ST}(A)$ performing the transformation

$$\omega_k = \Phi\sigma_k \ \forall \ k = 1, \ldots, n \ , \quad (1.3)$$

if and only if

$$S_f(\omega_1, \ldots, \omega_n) \leq S_f(\sigma_1, \ldots, \sigma_n) \quad (1.4)$$

for every h-convex functional S_f of order n over A_+^*. Moreover, the occurrence of (1.4) for all positive h-convex functionals S_f is sufficient to guarantee the existence of Φ obeying (1.3).

We close our preparations by introducing a relation $\gg$ between indexed sets of states.

Let $N = (\omega_i)_{i \in I}$, $N' = (\omega_i')_{i \in I}$ be two indexed sets (labeled by the same index set) of states on A. Then, we

 0022-2488/81/112345-07$01.00

define

Definition 1.3: $(\gg)$ $N \gg N'$ if, for any natural n and every choice $i_1, \ldots, i_n \in \mathrm{I}$, we have

$$S_f(\omega_{i_1}, \ldots, \omega_{i_n}) \leq S_f(\omega'_{i_1}, \ldots, \omega'_{i_n})$$

for every h-convex functional S_f of order n.

2. THE MAIN RESULT

We start our considerations in fixing the sense of what is called *stochastic dynamics*. Assume there is given a set of *stochastic maps* (T_{ts}), the members T_{ts} labeled by the pairs (t, s) of non-negative reals t, s with $t \geq s$.

Definition 2.1: (stochastic dynamics) (T_{ts}) is called stochastic dynamic if

$$T_{tt} = id \quad \forall t \geq 0, \tag{2.1}$$

$$T_{su} = T_{st} T_{tu} \quad \forall\ s \geq t \geq u \geq 0\ , \tag{2.2}$$

$$T_{st}(A^*) \text{ is dense in } A^*\ . \tag{2.3}$$

In case of a W^* algebra A, (T_{ts}) is said to be a *normal stochastic dynamic* if $T_{ts} \in \mathrm{ST}_w(A)$, and (2.3) is replaced with "$T_{ts}(A_*)$ is dense in A_*." A simple example of a stochastic dynamic is given by

Example 2.2: Let $A = l^\infty(\{1, \ldots, N\})$, and assume $M = (M_{ik})$ an $N \times N$ matrix with properties:

(i) $\sum_i M_{ik} = 0$, $\forall k$

(ii) $M_{ii} \leq 0$, $\forall_i$, $M_{ik} \geq 0$, $\forall\ i \neq k$.

Then, $\{T_{ts} = \exp M(t-s)\}$ is a stochastic dynamic within $A^* = l^1(\{1, \ldots, N\})$.

Let $\omega \in S_A$, and define a "trajectory" $(\omega_t)_{t \geq 0}$ within S_A by

$$\omega_t = T_{t0}\omega, \ \forall\ t \geq 0\ . \tag{2.4}$$

Then, $\omega = \omega_0$ will be called *initial state* of the trajectory $(\omega_t)_{t \geq 0}$ under this stochastic dynamic (T_{ts}). The total system $\{(\omega_t)_{t \geq 0}\}_{\omega \in S_A}$ of the trajectories generated by (T_{ts}) has the following properties

$$A_t^* = [(\omega_t)_{\omega \in S_A}] \text{ is dense in } A^* \text{ for } t \geq 0, \tag{2.5}$$

$$(\omega_t)_{\omega \in S_A} \gg (\omega_s)_{\omega \in S_A} \text{ whenever } t \geq s \geq 0\ , \tag{2.6}$$

where in (2.6) we referred to Theorem 1.2 and Definition 1.3 with $\mathrm{I} = S_A$, and $[\ ]$ in (2.5) means the operation of taking the linear hull. In the next step, let us ignore the presence of the generating dynamic (T_{ts}), and extract from (2.5) and (2.6) the following notion:

Definition 2.3: (c system) Let S_0 be a subset of states A. We call $\{(\omega_t)_{t \geq 0}\}_{\omega \in S_0}$ the c system (of trajectories) in one case that

$$A_t^* = [(\omega_t)_{\omega \in S_0}] \text{ is dense in } A^* \text{ for all } t \geq 0, \tag{2.7}$$

where $\omega_0 = \omega$ is supposed;

$$\{\omega_t\}_{\omega \in S_0} \gg \{\omega_s\}_{\omega \in S_0} \ \forall\ t \geq s \geq 0\ . \tag{2.8}$$

In the case that $A_t^* = A^*$ for all $t \geq 0$, we will speak of the c-system in question as a *proper* c system. The elements of S_0 will be referred to as *initial states* of the c system.

Remark 2.4: Let A be a commutative W^*-algebra. Then, $\{(\omega_t)_{t \geq 0}\}_{\omega \in S_0}$ is called a normal c system (resp. proper normal c system) if the trajectories belong to A_*, (2.8) is fulfilled, and (2.7) is replaced with the requirement that $A_{*t} = [(\omega_t)_{\omega \in S_0}]$ be dense in the Banach space A_* (resp. $A_{*t} = A_*\ \forall\ t \geq 0$).

We note that in the case of a normal c system the meaning of (2.8) becomes quite transparent due to the representation offered by Proposition 1.1. In the context of *normal* states, heuristically interesting geometrical and physically motivated interpretations of h-convex functionals are possible (we will not give them in this purely mathematical paper). As an example of a c-system we give

Example 2.5: [see Ref. 3, for instance] Let $A = L^\infty(\mathbb{R}^3)$. Then, $A_* = L^1(\mathbb{R}^3)$. Take for the set of initial states S_0 the probability distributions belonging to $C_0^\infty(\mathbb{R}^3)$ (space of all infinitely often differentiable functions with compact support in $\mathbb{R}^3$). Then, due to the fact that $C_0^\infty(\mathbb{R}^3)$ is dense in $L^1(\mathbb{R}^3)$, we find the linear hull of S_0 to be dense. Look at the heat equation $d_t V = \frac{1}{2}\Delta V$. Then, the Cauchy problem of this special master equation has a unique solution to the initial state $V \in S_0$, and $\{V_t\}_{t \geq 0}\}_{V \in S_0}$ forms a normal c system that is generated by a stochastic dynamic (the heat transformation) given in form of a stochastic integral operator

$$T_{ts}(y, x) = \left(\frac{1}{2\pi(t-s)}\right)^{3/2} \exp\left(-\frac{1}{2(t-s)}\,|y-x|^2\right),$$

with $t > s$, and $T_{tt} = id$ by definition.

We remark that Example 2.5 is not bounded to $\mathbb{R}^3$, and by extending Δ from $C_0^\infty(\mathbb{R}^3)$ in a suitable way a generalized heat equation solvable uniquely through the whole $L^1(\mathbb{R}^3)$ could be obtained.

The way we arrived at Definition 2.3 and the examples given suggest that one asks the following question:

Problem 2.6: Given a c system in the state space of a commutative C^* algebra A with identity, can one find a generating stochastic dynamics?

In this context, the stochastic dynamics (T_{ts}) is said to generate $\{(\omega_t)_{t \geq 0}\}_{\omega \in S_0}$ if $\omega_t = T_{t0}\omega$, $\omega \in S_0$. The decisive step toward an answer to the question is in proving the following

Lemma 2.7: Let $(\omega_i)_{i \in \mathrm{I}}$, $(\omega'_i)_{i \in \mathrm{I}} \subset S_A$. Then

$$(\omega_i)_{i \in \mathrm{I}} \gg (\omega'_i)_{i \in \mathrm{I}} \tag{2.9}$$

if and only if there is $T \in \mathrm{ST}(A)$ with

$$\omega_i = T\omega'_i \quad \forall\ i \in \mathrm{I}\ . \tag{2.10}$$

Proof: That (2.10) implies (2.9) is evident from Theorem 1.2 and the meaning of $\gg$ in 1.3.

Assume (2.9) to be fulfilled. Denote by $F(\mathrm{I})$ the set of all finite subsets of indices taken from I. In defining

$\Lambda \geq \Lambda'$ for Λ, $\Lambda' \in F(\mathrm{I})$ in case that $\Lambda \supset \Lambda'$, we may think of $\{F(\mathrm{I}), \geq\}$ as a directed set. Then, for $\Lambda \in F(\mathrm{I})$, we are assured of the existence of $T_\Lambda \in \mathrm{ST}(A)$ such that $\omega_i = T_\Lambda \omega_i'$, $\forall\, i \in \Lambda$, where we made use of Theorem 1.2. Since $\mathrm{ST}(A)$ is weakly compact, we find a converging subnet $(T_{\Lambda_\beta})_{\beta \in}$ of the net $(T_\Lambda)_{\Lambda \in F}$. Denote the limit by T. Assume $x \in A$, and $i \in \mathrm{I}$. Then, we find $\beta_0 \in \mathrm{K}$ such that $\Lambda_\beta \geq \{i\}$ whenever $\beta \geq \beta_0$, thus $(T\omega_i')(x) = \lim_\beta (T_{\Lambda_\beta}\omega_i')(x) = \lim_{\beta \geq \beta_0}(T_{\Lambda_\beta}\omega_i')(x) = (T_{\Lambda_{\beta_0}}\omega_i')(x) = \omega_i(x)$ by definition of T_Λ. The latter happens for every $x \in A$, so $T\omega_i' = \omega_i$ has to be required. Since i could range through the whole set I, we have arrived at the desired result.

Theorem 2.8: Every c system in the state space of a commutative C^* algebra with unit is generated by a uniquely determined stochastic dynamic. In case of a commutative W^* algebra and a normal c system a stochastic dynamic in $\mathrm{ST}_w(A)$ is uniquely given.

Proof: Let $\{(\omega_t)_{t \geq 0}\}_{\omega \in S_0}$ be the c system in question. Then, $(\omega_t)_{\omega \in S_0} \gg (\omega_s)_{\omega \in S_0}$ whenever $t \geq s \geq 0$, thus Lemma 2.7 applies to the time cuts $(\omega_t)_{\omega \in S_0}$ and $(\omega_s)_{\omega \in S_0}$, with $\mathrm{I} = S_0$.

That is we have $T \subset \mathrm{ST}(A)$ with $\omega_t = T\omega_s$ for any $\omega \in S_0$. Since A_s^* is dense in A^* and T is bounded, there is no other bounded linear map performing the transition from s-cut to t-cut. Hence we may define $T_{ts} = T$. This can be made for every pair with $t \geq s \geq 0$, and since $\omega_t = id\omega_t$, $T_{tt} = id$ has to hold. Let $t \geq u \geq s \geq 0$. Then, $\omega_t = T_{tu}T_{us}\omega_s$ and, by the same reasoning as above, we necessarily have $T_{tu}T_{us} = T_{ts}$, and $T_{ts}(A^*)$ dense in A_* by triviality follows. Finally, in case of a W^* algebra and a normal c system the assertion follows from the fact that A_* is a Banach space and the above constructed stochastic maps throw a dense set of A_* into A_*, so the restriction to A_* is in $\mathrm{ST}_w(A)$.

To make the correspondence between c-systems and stochastic dynamics complete, let us note that due to (2.3) *any* system of trajectories $\{\omega_t = T_{t0}\omega\}_{t \geq 0}$ with ω running through a set S_0 with $[S_0]$ being dense in A^* yields a c-system [(2.4) corresponds to $S_0 = S_A$].

Let $Z = \{(\omega_t)_{t \geq 0}\}_{\omega \in S_0}$ be a c-system in S_A. Z is said to be a *continuous* c system if any trajectory is continuously depending on t at any instant $t \geq 0$. Z is said to be *differentiable* if the time derivative $d_t\omega_t$ exists at any instant (at 0 the right derivative). Clearly, differentiable c systems will deserve our main interest, for the state solutions of many important master equation number among them (cf. Example 2.5).

We will justify the subsequent formulated regularity properties for continuous and differentiable c systems, respectively:

Proposition 2.9: Let Z be a continuous c system in S_A. Then, the Z generating stochastic dynamics (T_{ts}) is strongly continuous, i.e.,

$$T_{ts}\omega = \lim_{\substack{t' \to t \\ s' \to s \\ t' \geq s'}} T_{t's'}\omega, \quad \forall\, \omega \in A^*, \quad t \geq s.$$

Proof: It is plain to see that $\lim_{t' \to t} T_{t's}\omega = T_{ts}\omega$, $\forall\, \omega \in A^*$, $\forall\, t \geq s \geq 0$, for, the relation holds on a dense subset A_s^* of A^* and (T_{ts}) is uniformly bounded there. On the other hand, for $s \geq t \geq 0$, $s \geq t' \geq 0$ we have

$$\|T_{st}\omega_t - T_{st'}\omega_t\|_1 \leq \|T_{st'}\omega_t - T_{st'}\omega_{t'}\|_1 + \|T_{st'}(\omega_{t'} - \omega_t)\|_1, \tag{2.11}$$

and since $T_{st}\omega_t = T_{st'}\omega_{t'} = \omega_s$, by the uniform boundedness of (T_{st}) it follows from (2.11) that $\|T_{st}\omega_t - T_{st'}\omega_t\|_1 \leq \|\omega_{t'} - \omega_t\|_1$, i.e., $\lim_{t' \to t} T_{st'}\omega_t = T_{st}\omega_t$. Then, taking into account the denseness of A_t^* in A, and stressing again the uniform boundedness argument, we see that $\lim_{t' \to t} T_{st'}\omega = T_{st}\omega$ for every $\omega \in A^*$.

Finally, because $T_{s'u}T_{ut'} = T_{s't'}$ for every u with $s' \geq u \geq t'$, we get from the proven separate continuity, and uniform boundedness once more inserted, that with $s > u > t$

$$\lim_{\substack{s' \to s \\ t' \to t}} T_{s't'}\omega = \lim_{\substack{s' \to s \\ t' \to t}} T_{s'u}T_{ut'}\omega = T_{su}T_{ut}\omega = T_{st}\omega,$$

and for $s = t$ we may use (2.11) to make the argument complete. Q.E.D.

For a differentiable c system Z we have to state the following fact:

Proposition 2.10: Let Z be a differentiable c system in S_A. Then, there is a family $(L_t)_{t \geq 0}$ of linear operators, each of which is densely defined in A^*, such that Cauchy problem of the master equation

$$d_t\varphi = L_t\varphi, \tag{2.12}$$

has a solution for the dense set $[S_0]$ of initial elements, and the trajectories of the c-system Z are among the state solutions of (2.12). Moreover, a solution of (2.12) starting out from a state contained in $[S_0]$ evolves in S_A exclusively.

Proof: By Theorem 2.8 we are assured of a stochastic dynamic (T_{ts}) with

$$\omega_t = T_{ts}\omega_s, \quad \omega \in S_0. \tag{2.13}$$

Since Z is differentiable, we get from (2.13)

$$d_t\omega_t|_{t=s} = \lim_{t \to s} (t-s)^{-1}(T_{ts} - id)\omega_s, \quad \omega \in S_0. \tag{2.14}$$

On the dense linear set A_s^* we define an operator L_s acting into A^* by

$$\omega = \sum_i r_i \omega_s^i \mapsto L_s\omega = \sum_i r_i\, d_t\omega_t^i|_{t=s}. \tag{2.15}$$

Then, due to (2.13) and (2.14), we can be sure that L_s is linearly well defined on A_s. Let $\omega \in [S_0]$ be a state. Then, $\omega = \sum_i r_i \omega^i \in S_A$ with certain reals r_i and states $\omega^i \in S_0$. Since $T_{t0} \in \mathrm{ST}(A)$, we have

$$\omega_t = T_{t0}\omega \in S_A, \tag{2.16}$$

with $\omega_t = \sum_i r_i \omega_t^i$. Because of (2.14), however, we see $d_t\omega_t|_{t=s} = \sum_i r_i\, d_t\omega_t^i|_{t=s} = \sum_i r_i L_s \omega_s^i = L_s\omega_s$, i.e., $(\omega_t)_{t \geq 0}$ for $\omega \in [S_0] \cap S_A$ is a solution of the Cauchy prob-

lem for $d_t\varphi = L_t\varphi$ evolving totally in S_A [due to (2.16)]. That the problem has a solution for any $\varphi_0 \in [S_0]$ is seen in the same way. Q.E.D.

In other words, any differentiable c system can be interpreted as a subset of solutions of the Cauchy problem for a suitable master equation that is solvable on a dense set of initial conditions such that the equation admits the trajectory of an initial state to evolve in the state space.

Remark 2.11: In one case of a commutative W^* algebra A, all the derived results of this part remain true statements if we make the following replacements:

A^* replaced with A_*,

S_A replaced with normal states,

(proper) c-system replaced with (proper) normal c system,

$\mathrm{ST}(A)$ replaced with $\mathrm{ST}_w(A)$,

stochastic dynamic replaced with normal stochastic dynamic,

etc.

3. c SYSTEMS AND MASTER EQUATIONS

The aim of this part is to clarify the structure of those master equations on A^* that admit proper c systems as state solutions. We start with a class of master equations which are quite regular from our point of view.

Theorem 3.1: Let $\{L_t\}_{t\geq 0}$ be a family of bounded linear operators on A^*. Assume the following conditions to hold:

$$\|L_t\|_1 \leq C < \infty \text{ for } \forall\, t \geq 0\,, \tag{3.1}$$

$$L_t\omega \text{ depends continuously on } t \text{ for } \forall\, \omega \in A^*, \tag{3.2}$$

$\forall s \geq 0\ \exists\, \beta_s > 0$ such that

$$id + \beta_s\, L_t \in \mathrm{ST}(A) \text{ for } t \in [0, s]. \tag{3.3}$$

Then, the Cauchy problem for

$$d_t\varphi = L_t\varphi \tag{3.4}$$

is uniquely solvable through A^*. Moreover, the solutions starting out from states form a proper c-system.

Proof: Because of (3.1) and (3.2) the solution ω_t to $\omega \in A^*$ is uniquely given by

$$\omega_t = \omega + \int_0^t L_m\omega\, dm + \int_0^t L_m \int_0^m L_r\,\omega\, dr\, dm + \cdots \tag{3.5}$$
$$= T_{t0}\omega\,.$$

Fix $s \geq 0$, and define bounded linear maps by

$$T_{s's}\varphi = \varphi + \int_s^{s'} L_m\varphi\, dm + \int_s^{s'} L_m \int_s^m L_r\,\varphi\, dr\, dm + \cdots\,, \tag{3.6}$$

We are going to prove a formula for T_{ts} that explictly shows the stochasticity that we are looking for.

We put $\varphi_{s'} = T_{s's}\varphi$. Then, $\varphi_{s'}$ obeys the equation

$$\varphi_{s'} = \varphi + \int_s^{s'} L_r\varphi_r\, dr \quad \text{for } s \leq s' \leq t. \tag{3.7}$$

Suppose that $(t-s)C = \beta < 1$, and fix $\varphi \in A^*$. Since φ_r depends continuously on r and (3.1) and (3.2) hold, φ_r and $L_r\varphi_r$ are uniformly continuous on $[s, t]$. Let $\epsilon > 0$. Then, we find an integer N such that, with $t_i = (i/N)(t-s) + s$ for $i = 0, \ldots, N$, the following conditions are fulfilled

$$\|L_r\varphi_r - L_{t_i}\varphi_{t_i}\|_1 \leq \epsilon, \quad \|\varphi_r - \varphi_{t_i}\|_1 \leq \epsilon, \quad \forall\, r \in [t_{i-1}, t_i],$$
and (3.8)
$$(1 + e^{C(t-s)})\,\|\varphi\|_1\beta^N < \epsilon\,.$$

Then, from (3.7) it arises

$$\Big\| \int_s^u L_r\varphi_r\, dr - \frac{(t-s)}{N} \sum_{k=1}^{i} L_{t_k}\varphi_{t_k} \Big\|_1 \leq 2\epsilon(t-s), \tag{3.9}$$

whenever $u \in [t_{i-1}, t_i]$. Let us define $\varphi_r^0 = \varphi_r$, $r \in [s, t]$,

$$\varphi_r^1 = \varphi + \frac{(t-s)}{N} \sum_{k=1}^{i} L_{t_k}\varphi_{t_k} \quad \text{for } r \in [t_{i-1}, t_i], \tag{3.10}$$

and inductively

$$\varphi_r^n = \varphi + \frac{(t-s)}{N} \sum_{k=1}^{i} L_{t_k}\varphi_{t_k}^{n-1} \quad \text{for } r \in [t_{i-1}, t_i]. \tag{3.11}$$

One easily checks that (3.8)–(3.11) guarantee that

$$\|\varphi_r^n - \varphi_r^{n-1}\|_1 \leq 2\epsilon(t-s)\beta^n\,, \tag{3.12}$$

and in summing up over n running from 1 to N

$$\|\varphi_r^N - \varphi_r\|_1 \leq 2\epsilon(t-s)(1-\beta)^{-1}, \quad \forall\, r \in [s, t]. \tag{3.13}$$

In using (3.10) and (3.11) we will also obtain a representation of φ_t^N in the following form

$$\varphi_t^N = \prod_{k=N}^{1} \left(id + \frac{(t-s)}{N} L_{t_k}\right)\varphi + D_N, \tag{3.14}$$

with

$$D_N = \frac{(t-s)}{N^N} \Big\{ \sum_{i=1}^{N} L_{t_i} \sum_{j=1}^{i} L_{t_j} \cdots \sum L_{t_1}\varphi_{t_1} - L_{t_N} \cdots L_{t_1}\varphi \Big\}. \tag{3.15}$$

From (3.6) it arises that $\|\varphi_r\|_1 \leq e^{C(t-s)}\|\varphi\|_1$, $\forall\, r \in [s, t]$ hence we may estimate the norm of D_N given by (3.15) as

$$\|D_N\|_1 \leq \beta^N\,\|\varphi\|_1(e^{C(t-s)} + 1) < \epsilon\,, \tag{3.16}$$

where we made use of (3.8).

We define

$$T_{ts}^N = \left(id + \frac{(t-s)}{N} L_{t_N}\right) \cdots \left(id + \frac{(t-s)}{N} L_{t_1}\right),$$

and from (3.13) and (3.14) comes that

$$\| \varphi_t - T^N_{ts}\varphi \|_1 \leq \left(1 + \frac{2(t-s)}{(1-\beta)}\right)\epsilon\,, \tag{3.17}$$

and since $\epsilon > 0$ and $\varphi \in A^*$ were arbitrarily chosen, we may take as proven

$$T_{ts} = \text{st-}\lim_N T^N_{ts} \quad (\text{st- strong}), \tag{3.18}$$

with T^N_{ts} defined as above (where $t_i = (i/N)(t-s)+s$).

Because of (3.18) and the special structure of T^N_{ts} we see that $T^N_{ts} \in \mathrm{ST}(A)$ for $N > (t-s)\beta_t^{-1}$, so with t, s fixed, T_{ts} in the strong limit of a sequence of stochastic maps, so that $T_{ts} \in \mathrm{ST}(A)$ due to weak compactness of $\mathrm{ST}(A)$. Moreover, for $(t-s)C < 1/2$ we see

$$\| id - T^N_{ts} \|_1 \leq \frac{1(t-s)C}{1-(t-s)C} < 1, \quad \forall\, N\,,$$

hence $\|id - T_{ts}\|_1 < 1$, too. The latter means invertibility of T_{ts} in $\mathbf{B}(A^*)$. Since $T_{ts}T_{su} = T_{tu}$ for $t \geq s \geq u$ holds, T_{ts} is stochastic and bounded invertible for every pair t, s, i.e., the state solutions of (3.4) form a proper c system. Q.E.D.

We remark that Theorem 3.1 is a statement which closely relates to Example 2.2.

Let us close our considerations in proving that the class of equations described in Theorem 3.1 is primary in the set of all master equations that admit a proper c system as a solution

Theorem 3.2: Let $\{L_t\}_{t\geq 0}$ be a family of linear operators on A^* such that the Cauchy problem for

$$d_t\,\varphi = L_t\,\varphi$$

admits state solutions that form a proper c-system. Then, every L_t is bounded and there exists a sequence of families $\{(L^n_t)_{t\geq 0}\}_n$ of bounded linear operators, each of them being of the type described inTheorem 3.1 such that

$$L_t = \text{st-}\lim_n L^n_t, \quad \forall\, t \geq 0\,. \tag{3.19}$$

Proof: Let $\{(\omega_t)_{t\geq 0}\}_{\omega \in S_A}$ denote the proper c system known to be a solution of the master equation under discussion (the choice of $S_0 = S_A$ is not a restriction!). By Theorem 2.8 we are assured of the existence of a generating dynamic (T_{ts}) which, due to Proposition 2.9, is strongly continuous since $(\omega_t)_{t\geq 0}$ is a solution of a master equation. We define

$$L^n_t = n(T_{t+1/n\,t} - id)\,. \tag{3.20}$$

Inserting ω_t, we see from (3.20) and the assumptions

$$\lim_n L^n_t\,\omega_t = \lim_n \frac{\omega_{t+1/n} - \omega_t}{(1/n)} = d_t\,\omega_t = L_t\,\omega_t\,. \tag{3.21}$$

Since (3.21) is valid on $(\omega_t)_{\omega\in S_A}$, it is valid on $[(\omega_t)_{\omega\in S_A}] = A^*$, too, i.e., $\lim_n L^n_t\,\omega = L_t\,\omega$, $\forall\,\omega \in A^*$. The principle of uniform boundedness (Banach–Steinhaus theorem) then gives that L_t has to be bounded for every $t \geq 0$. Strong continuity of (T_{ts}) implies L^n_t to be strongly continuous. Finally, the special form of (3.20) makes all the other requirements of Theorem 3.1, (3.1)–(3.3) hold for $\{L^n_t\}_{t\geq 0}$. Q.E.D.

It should be clear from the proof that Theorem 3.2 can be modified in various aspects. So, for instance, we may replace "proper c system" if we require validity of (3.19) only on a dense subset (i.e., A^*_f). Also, in the case of a W^* algebra it is possible to formulate a "normal" variant of Theorem 3.2. Finally, we remark that the results of Proposition 2.10 and Theorem 3.2 which have been derived essentially on the basis of Theorem 2.8 give only one aspect of applications of Theorem 2.8. Another field for application and further concern is to ask for *stability* properties of c systems etc.

APPENDIX

Let A be a commutative W^* algebra we may identify with $L^\infty(\Omega, \mu)$ for a suitable measure space Ω with measure μ. Let $\omega_1, \ldots, \omega_n$ denote normal states on A. Then, there are functions $\omega_1(x), \ldots, \omega_n(x) \in L^1(\Omega, \mu)$ representing the states via the formula $\omega_i(a) = \int \omega_i(x)\,a(x)\,d\mu(x)$, with $a(x) \in L^\infty(\Omega, \mu)$ being a representative of $a \in A$ [more precisely, $a(x)$ is the representative of the class in $L^\infty(\Omega, \mu)$ which corresponds to A].

With this notion in mind, we are going to prove the result we touched on in Sec. 1.

Proposition: For any non-negative h-convex function f on $\mathbb{R}^n_+$ the corresponding h-convex functional is

$$S_f(\omega_1, \ldots, \omega_n) = \int_\Omega f(\omega_1(x), \ldots, \omega_n(x))\,d\mu(x)\,. \tag{A1}$$

Proof: Our first task will be to show that representation (A1) is valid in the case that $\omega_1(x), \ldots, \omega_n(x) \geq 0$ are simple functions in $L^1(\Omega, \mu)$. As usual, $\omega(x)$ is said to be a simple measurable function if $\omega = \sum_i r_i \chi_i$ with certain $r_i \in C^1$ and $\{\chi_i\}$ denoting the characteristic functions of a finite number of measurable sets that are pairwise disjunct.

Let us assume the simple functions ω_i to be represented by

$$\omega_i = \sum_l t_{il}\chi_l, \quad i = 1, \ldots, n\,. \tag{A2}$$

Let $\{Q_s\}_{s=1}^N$ be a finite orthogonal decomposition of $\mathbf{1}$ into orthoprojections Q_s. Let Q_s correspond to the characteristic function χ'_s of some measurable set G'_s.

Look at

$$f(\omega_1(Q_s), \ldots, \omega_n(Q_s))$$

$$= f\left(\sum_l t_{1l}\,\mu(G_l \cap G'_s), \ldots, \sum_l t_{nl}\,\mu(G_l \cap G'_s)\right). \tag{A3}$$

Employing subadditivity and homogeneity of f, (A3) turns into the inequality

$$f(\omega_1(Q_s), \ldots, \omega_n(Q_s)) \leq \sum_l \mu(G_l \cap G'_s)(f(t_{1l}, \ldots, t_{nl}). \tag{A4}$$

As usual, we adopt the convention $\infty \cdot 0 = 0$ which, due to

$f(0,\ldots,0)=0$, gives no contradiction in transition from (A3) to (A4). Because of $\cup_s G'_s=\Omega$ and since $G'_s\cap G'_t=\emptyset$ for $s\neq t$, from (A4) it follows that

$$\sum_s f(\omega_1(Q_s),\ldots,\omega_n(Q_s))\leq\sum_l \mu(G_l)f(t_{1l},\ldots,t_{nl}), \tag{A5}$$

and the right-hand side of (A5) equals $\int_\Omega f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x)$. We may suppose that $\{\chi_l\}$ corresponds to a finite orthogonal decomposition of 1 into orthoprojections [then, for $\mu(\Omega)=\infty$, one of t_{il}'s has to vanish], so we see from (A5)

$$\sup_{\{Q_s\}}\sum_s f(\omega_1(Q_s),\ldots,\omega_n(Q_s)) = \int_\Omega f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x), \tag{A6}$$

where the sup runs through all finite orthogonal decompositions of the unity. The left-hand side of (A6), however, equals $S_f(\omega_1,\ldots,\omega_n)$ by the results on h-convex functionals [see (5.9) in Ref. 1], i.e., we may take as proven

$$S_f(\sigma_1,\ldots,\sigma_n)=\int_\Omega f(\sigma_1(x),\ldots,\sigma_n(x))\,d\mu(x) \tag{A7}$$

for simple $\sigma_1,\ldots,\sigma_n\in L^1(\Omega,\mu)$.

In the next step, let $\omega_1(x),\ldots,\omega_n(x)$ correspond to normal states on $L^\infty(\Omega,\mu)$. Then, due to the continuity of f, $f(\omega_1(x),\ldots,\omega_n(x))$ is measurable, and positive by assumption. First, let us show the existence of an increasing sequence Ω_r of measurable sets with $\mu(\Omega_r)<\infty$, $\omega_i(x)\chi_r(x)\in L^\infty(\Omega,\mu)$, $\forall i$, and

$$\lim_r\int_{\Omega_r} f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x) = \int_\Omega f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x), \tag{A8}$$

where χ_r stands for the characteristic function of Ω_r. In fact, let us define $\Omega_r=\{x\in\Omega:\ (1/r)\leq h(x)\leq r\}$, with $h(x)=\sum_i\omega_i(x)$. Then, $\Omega_1\subset\Omega_2\subset\cdots$, and $\mu(\Omega_r)<\infty$ since $h(x)\in L^1(\Omega,\mu)$. Setting $\Omega'=\cup_r\Omega_r$, we see $\int_{\Omega'}f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x)=\int_\Omega f(\omega_1(x),\ldots,\omega_n(x))\times d\mu(x)$, for, from $x\notin\Omega'$ it follows that either $h(x)=0$ by homogeneity, or $h(x)=\infty$, which happens at most on a set of measure zero [$h\in L^1(\Omega,\mu)$!]. Thus, by Lebesgue's monotone convergence theorem we see equality (A5) to be true (f is positive!).

We are going to show that

$$\int_\Omega f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x)\leq S_f(\omega_1,\ldots,\omega_n). \tag{A9}$$

Since $\omega_i\chi_k\in L^\infty(\Omega,\mu)$, we find decreasing and increasing sequences $\{\omega^+_{iN}\}_N$, and $\{\omega^-_{iN}\}_N$ respectively, which consist of simple functions with support in Ω_k such that

$$\omega^+_{iN}\geq\omega_i\chi_k\geq\omega^-_{iN}\ \forall i,\quad \omega^\pm_{iN}=\sum_l^N t^\pm_{il}\chi^N_l\ \ (\text{with }\chi^N_l\neq 0), \tag{A10}$$

$$\lim_N\omega^+_{iN}=\omega_i\chi_k=\lim_N\omega^-_{iN}\ \text{in a uniform sense.}$$

Now, define on $L^1(\Omega,\mu)$ a contraction E^N by

$$E^N\sigma=\sum_l\frac{1}{\mu(G_l)}\int_\Omega\sigma(x)\chi^N_l(x)\,d\mu(x)\,\chi^N_l, \tag{A11}$$

where G_l is the measurable set for which χ^N_l is the characteristic function. Since E^N is a positive linear map with $\omega^\pm_{iN}$ being fixpoints, we get with (A10)

$$\omega^-_{iN}\leq E^N\omega_i\leq\omega^+_{iN},\ \text{where we used } E^N\omega_i=E^N\omega_i\chi_k, \tag{A12}$$

and from the second part of (A10) and from (A12) we see

$$\lim_N E^N\omega_i=\omega_i\chi_k\ \text{ uniformly }\ \forall i. \tag{A13}$$

Because the adjoint map of E^N, E^{N+}, is positive and $E^{N+}\mathbf{1}=\chi_k\leq 1$, from this, together with positivity of f and the original definition of S_f [cf. (1.1)], follows

$$S_f(E^N\omega_1,\ldots,E^N\omega_n)\leq S_f(\omega_1,\ldots,\omega_n). \tag{A14}$$

The $E^N\omega_i$ being simple functions makes (A7) to be applicable, thus from (A14) we are led to

$$\int_{\Omega_k} f(E^N\omega_1(x),\ldots,E^N\omega_n(x))\,d\mu(x)\leq S_f(\omega_1,\ldots,\omega_n), \tag{A15}$$

Applying (A13) and recalling $\mu(\Omega_k)<\infty$, (A15) yields

$$\int_{\Omega_k} f(\omega_1(x),\ldots,\omega_n(x))\,d\mu(x)\leq S_f(\omega_1,\ldots,\omega_n), \tag{A16}$$

from which inequality by means of (A8) the desired result (A9) can be seen.

Let us demonstrate the validity of the reverse of the inequality just proved. To this sake, by standard methods (see Ref. 4) we construct increasing sequences $\{s_{iN}\}_N$ of measurable simple functions with $0\leq s_{i1}\leq s_{i2}\leq\cdots\leq\omega_i$, and $\lim_N s_{iN}(x)=\omega_i(x)$ for all $\forall x\in\Omega$. We can choose the sequences in such a way that the convergence is uniform on any subset of Ω where ω_i's are bounded.

Especially, $\{s_{iN}\}$ tends uniformly to ω_i on Ω_k as defined above. One also easily recognizes that $\int_{\Omega'}\omega_i(x)\times d\mu(x)=1$. Now, by definition of S_f, we have

$$S_f(\sigma_1,\ldots,\sigma_n)=\sup_M S^M_f(\sigma_1,\ldots,\sigma_n), \tag{A17}$$

with

$$S^M_f(\sigma_1,\ldots,\sigma_n)=\sup_{\{a_k\}}\sum_k f(\sigma_1(a_k),\ldots,\sigma_n(a_k)),$$

with the sup running through all positive decompositions of 1 into, at most, M positive elements. It is plain to see that S^M_f is $\|\cdot\|_1$-continuous, so from $\|s_{iN}\chi_k-\omega_i\chi_k\|_1\xrightarrow{N}0$ follows that

$$S_f^M(\omega_1\chi_k, \ldots, \omega_n\chi_k) = \lim_N S_f^M(s_{1N}\chi_k, \ldots, s_{nN}\chi_k)$$

$$\leq \lim_N S_f(s_{1N}\chi_k, \ldots, s_{nN}\chi_k) \quad \text{(A18)}$$

$$= \lim_N \int_{\Omega_k} f(s_{1N}(x), \ldots, s_{nN}(x))\, d\mu(x)$$

$$= \int_{\Omega_k} f(\omega_1(x), \ldots, \omega_n(x))\, d\mu(x)$$

$$\leq \int_{\Omega} f(\omega_1(x), \ldots, \omega_n(x))\, d\mu(x)\,,$$

where in the last steps we made use of (A7) and the uniform convergence of $\{s_{iN}\}$ towards ω_i on Ω_k with finite measure, and positivity of f makes the conclusion of (A18) complete. Since the increasing sequences $\{\omega_i\chi_k\}_k$ converge pointwise to ω_i on Ω', from the property of Ω' (see above) and positivity of ω_i comes that $\|\omega_i\chi_k - \omega_i\|_1 \overset{k}{\to} 0$. Hence, from $\|\cdot\|_1$-continuity of S_f^M comes that (A18) can be turned into

$$S_f^M(\omega_1, \ldots, \omega_n) \leq \int_{\Omega} f(\omega_1(x), \ldots, \omega_n(x))\, d\mu(x)\,, \quad \text{(A19)}$$

and in applying (A17) to (A19) we have arrived at

$$S_f(\omega_1, \ldots, \omega_n) \leq \int_{\Omega} f(\omega_1(x), \ldots, \omega_n(x))\, d\mu(x)\,. \quad \text{(A20)}$$

Taking together (A20) with (A9), the desired result (A1) is obtained. We remark that the value ∞ is included in all considerations. Q.E.D.

[1] P. M. Alberti and A. Uhlmann, "Existence and Density Theorems for Stochastic Maps on Commutative C^* algebras," (submitted to Math. Nachr.).

[2] R. G. Douglas, *Banach Algebra Techniques in Operator Theory* (Academic, New York, 1972).

[3] T. Kato, *Perturbation Theory for Linear Operators* (Springer–Verlag, New York, 1966).

[4] W. Rudin, *Real and Complex Analysis* (McGraw–Hill, New York, 1970).

On the triangle anomaly number of SU(n) representations [a)]

J. Patera
Centre de Recherche de Mathématiques Appliqués, Université de Montréal, Montréal, Québec, Canada
R. T. Sharp
Department of Physics, McGill University, Montréal, Québec, Canada

(Received 6 March 1981; accepted for publication 27 March 1981)

A simple geometrical interpretation of the triangle anomaly number of SU(n) representations is demonstrated. The number is equal to the sum of cubes of projections of weights on the U(1) direction; U(1) is the group in the reduction SU(n)$\supset$U(1)$\times$SU($n-1$). Properties of the anomaly number are then consequences of properties of weight systems. Explicit formulas are derived for the anomaly numbers for reducible representations corresponding to general plethysms based on an arbitrary representation of SU(n). Generalizations to "anomaly numbers" corresponding to degrees other than three are discussed.

PACS numbers: 02.20.Km, 11.10.Np, 11.10.Gh

I. INTRODUCTION

Renormalizability of a unified nonabelian gauge theory based on a representation (λ) of a compact semisimple Lie group G requires[1] the vanishing of the triangular anomaly number $A_{(\lambda)}$. This turns out to be a quite restrictive condition when the group G is of the type[2] SU(n), $n\geqslant 3$ or contains it as one of the factors, $G = G'\times \mathrm{SU}(n)$, and the fermion representations are complex (nonself-contragredient).[3] Representations of simple groups other than SU(n), or of semisimple groups not involving SU(n), are known to be anomaly-free.[2] Examples of anomaly-free irreducible complex SU(n) representations are extremely rare.[2] It is therefore an imperative to study reducible complex representations of SU(n); see, for example, the papers of Ref. 4.

The purpose of this paper is to point out that the triangular anomaly number $A_{(\lambda)}$ of a representation (λ) has a simple geometrical interpretation in terms of the weights of the representation. For SU(n), it is, up to a normalization, the sum of cubes of the components of weights corresponding to the U(1) subgroup in the reduction SU(n)$\supset$U(1)$\times$SU($n-1$). For representations of simple groups other than SU(n), the triangle anomaly vanishes when it is defined as the sum of cubes of components of weights, where "component" refers to any direction in weight space. We exploit this definition of $A_{(\lambda)}$ to derive a number of new properties, and to simplify the proof of known ones.

The results of this paper are of three types. First, the new definition of the triangular anomaly number relating it directly to weights (quantum numbers) instead of eigenvalues of third-order Casimir operators[2,5] allows a straightforward physical interpretation depending only on the choice of quantum numbers assigned to that particular SU(n) weight space.[6] Second, the anomaly number is found here for a vast class of reducible representations. Finally, generalizations to similar characteristics of SU(n) representations such as "pentagon anomaly number" are described.

A remark concerning notations: There is a well known correspondence between Young tableaux and irreducible representations of SU(n). Here we specify both by integers $(\lambda_1,\lambda_2,\ldots,\lambda_{n-1})$ such that λ_i equals the number of *columns* of the Young tableau with i boxes. This is a notation often used for SU(n) representations but rarely for Young tableaux (in the usual notation for the latter λ_i would represent the number of *rows* with i boxes). For simplicity we sometimes denote the representations by their dimensions, or draw Young tableaux explicitly.

Section II contains the new definition of $A_{(\lambda)}$ and its properties. In Sec. III, we derive the formula for the triangle anomaly number for a general plethysm based on any representation of dimension N of any simple Lie group. A number of examples are shown here. In Sec. IV, we make some obvious generalizations of the triangle anomaly number to degrees $\neq 3$ and discuss their properties. The last section contains some additional comments.

II. THE ANOMALY NUMBER AND ITS PROPERTIES

The new definition of the anomaly number $A_{(\lambda)}$ of an SU(n) representation (λ) is given by the formula

$$A_{(\lambda)} = a\sum_{\mathbf{w}} w^3, \tag{2.1}$$

where $\mathbf{w}$ denotes a weight of the representation (λ), and w is the projection of the weight $\mathbf{w}$ on the direction in weight space corresponding to the U(1) subgroup of SU(n) which occurs in the reduction SU(n)$\supset$SU($n-1$)$\times$U(1). The normalization factor a is usually chosen so that $A_{(\lambda)} = 1$ when (λ) is the defining n-dimensional representation of SU(n). If the U(1) weight generator Y has the form

$$Y = \frac{1}{n}\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & -n+1 \end{pmatrix}, \tag{2.2}$$

then

$$a = -n^2[(n-1)(n-2)]^{-1}. \tag{2.3}$$

[a)]Work supported in part by the Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

 0022-2488/81/112352-05$01.00

The sum in (2.1) can be evaluated for a general irreducible representation (λ). The derivation parallels those for the second and fourth indices, described in Ref. 7; the resulting value of $A_{(\lambda)}$ agrees with the A_λ of Banks and Georgi,[5] and is twice the K_λ of Okubo.[2] A simpler proof of the equivalence makes use of Okubo's[2] identity

$$\mathrm{Tr}(\{t_a,t_b\}t_c) = d_{abc}K. \tag{2.4}$$

with $t_a = t_b = t_c = Y$, the left side of (2.4) becomes $2\Sigma w^3$; also

$$d_{YYY} = -4(n-1)(n-2)n^{-2}, \tag{2.5}$$

and the equivalence of the two anomaly numbers is established.

The formula for A_λ is

$$A_\lambda = \frac{2N_\lambda}{n}\left[\frac{p_3}{(n+1)(n+2)} + \frac{p_3 - p_1p_2}{(n-1)(n+1)} + \frac{p_3 - p_1^3 - 2p_1p_2}{(n-1)(n-2)}\right], \tag{2.6}$$

where $p_\alpha = p_\alpha(l)$ is defined by

$$\prod_{i=1}^{n}(1-zl_i)^{-1} = \sum_{\alpha=0}^{\infty} P_\alpha(l_1,\ldots,l_n)z^\alpha. \tag{2.7}$$

The l_j are defined in terms of the nonnegative integer labels λ_j by

$$l_j = \sum_{k=j}^{n-1}\lambda_k + n - j, \quad 1 \leqslant j \leqslant n-1, \quad l_n = 0. \tag{2.8}$$

Here we illustrate our definition on two simple examples, SU(3) representations of dimensions 3 and 6. Consider the two weight systems shown in Fig. 1. The sum

$$A_{(3)} = a\left[(-\tfrac{2}{3})^3 + 2(\tfrac{1}{3})^3\right] = -\tfrac{6}{27}a = 1$$

fixes the normalization as $a = -27/6$. Then for the sextet (6) one has

$$A_{(6)} = -\tfrac{27}{6}\left[(-\tfrac{4}{3})^3 + 2(-\tfrac{1}{3})^3 + 3(\tfrac{2}{3})^3\right] = 7,$$

which is the well known value of $A_{(6)}$.

From the definition (2.1) one finds immediately some properties of $A_{(\lambda)}$; it suffices to recall some properties of the weight systems of the representations involved.

Denoting by (λ) and $(\bar{\lambda})$ a pair of mutually contragredient representations of SU(n), one has

$$A_{(\lambda)} = -A_{(\bar{\lambda})}; \tag{2.9}$$

in particular, $A_{(\lambda)} = 0$, when $(\lambda) \equiv (\bar{\lambda})$ is a self-contragredient (real) representation.

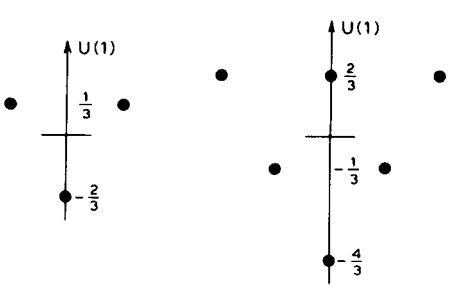


FIG. 1. Weights of the representations (3) and (6) of SU(3).

The anomaly number of a direct sum $(\lambda)\oplus(\lambda')$ is

$$A_{(\lambda)\oplus(\lambda')} = A_{(\lambda)} + A_{(\lambda')}. \tag{2.10}$$

For a product $(\lambda)\otimes(\lambda')$ one has

$$\begin{aligned} A_{(\lambda)\otimes(\lambda')} &= a\sum_{\mathbf{w},\mathbf{w}'}(w+w')^3 \\ &= a\sum_{\mathbf{w},\mathbf{w}'}(w^3 + w'^3 + 3w^2w' + 3ww'^2) \\ &= a\sum_{\mathbf{w},\mathbf{w}'}(w^3 + w'^3) = N_{(\lambda)}A_{(\lambda')} + A_{(\lambda)}N_{(\lambda')}, \end{aligned} \tag{2.11}$$

where $N_{(\lambda)}$ is the dimension of (λ). In (2.11) $\mathbf{w}$ and $\mathbf{w}'$ run through all the weights of (λ) and (λ'), respectively. We make use of the property $\Sigma_{\mathbf{w}} w = \Sigma_{\mathbf{w}'} w' = 0$ in deriving (2.11).

Consider now a group–subgroup pair SU(n)$\supset$SU(m), $n > m$, and the corresponding reduction of a given representation (λ) of SU(n) to the representation (λ') of SU(m). Typically (λ') is a direct sum of several irreducible components:

$$(\lambda)\supset(\lambda') = \oplus_i(\lambda'_i). \tag{2.12}$$

The definition (2.1) suggests immediately another property of anomaly numbers related to the reduction (2.12), namely,

$$A_{(\lambda)} = \rho A_{(\lambda')} = \rho\sum_i A_{(\lambda'_i)}, \tag{2.13}$$

where ρ is a constant depending only on the way the subgroup SU(m) is embedded into the group SU(n); it does not depend on the choice of the representation (λ). Hence it can be found once and for all from the equality of anomalies (2.13) for the simplest representation (λ), and then used for any other choice of (λ).

Let us illustrate the assertion by two examples.

1. SU(n)$\supset$SU($n-1$), the SU(n) representations (10···0) and (010···0) of dimension n and $\frac{1}{2}n(n-1)$, respectively. One has

$$(10\cdots0)_{\mathrm{SU}(n)} \supset (10\cdots0)_{\mathrm{SU}(n-1)} \oplus (0\cdots0)_{\mathrm{SU}(n-1)}$$

Clearly the anomalies on the left and right side add to 1; hence also $\rho = 1$. Similarly,

$$\begin{aligned} (010\cdots0)_{\mathrm{SU}(n)} &= (010\cdots0)_{\mathrm{SU}(n-1)} \oplus (10\cdots0)_{\mathrm{SU}(n-1)} \\ n-4 &= (n-1) - 4 + 1, \end{aligned}$$

where the second line indicates the corresponding values of A.

2. SU(6)$\supset$SU(3) so that the representation (10000) of SU(6) contains the six-dimensional representation (20) of SU(3). Thus, from

$$(10000)\supset(20)$$

one has

$$1 = A_{(10000)} = \rho A_{(20)} = 7\rho$$

so that

$$\rho = \tfrac{1}{7}$$

Then, for instance,

$$(00002)\supset(04)\oplus(20);$$

$$-9 = \tfrac{1}{7}(-70+7).$$

A proof of (2.13) is given at the end of the Appendix.

III. ANOMALY NUMBERS OF CERTAIN REDUCIBLE REPRESENTATIONS

The subject of our interest in this section is the tensor product of several copies of any given representation of SU(n). More precisely we consider the component (plethysm) of such a product with a definite permutation symmetry specified by a Young tableau. Our aim is to derive a formula for the anomaly number A_{YT} of an arbitrary plethysm and then to consider specific examples.

First let us fix an arbitrary representation of SU(n) of dimension $N < \infty$. In order to underline that it does not have to be irreducible, we denote it by the trivial Young tableau □. Its anomaly number is then $A_{\square}$. The content of the appendix is the proof of the relation

$$A_{\mathrm{YT}} = F_{\mathrm{YT}}(N)A_{\square}, \tag{3.1}$$

where the subscript YT stands for an arbitrary Young tableau, and the polynomial $F_{\mathrm{YT}}(N)$ depends on the representation □ only through its dimension N. Since we know $A_{\square}$ for any representation □, [cf. Eq. (2.6)] the determination of A_{YT} hinges on finding $F_{\mathrm{YT}}(N)$ as a function of N and the Young tableau YT.

We can determine $F_{\mathrm{YT}}(N)$ by making a very particular choice of the representation □, namely the defining representation of SU(N). Then any plethysm corresponding to a connected YT is irreducible [the Young tableau denotes an irreducible representation of SU(n)]. However, with this choice made we know the left-hand side of (3.1), and also $A_{\square} = 1$. Consequently, we know $F_{\mathrm{YT}}(N)$. Thus all one has to do is to identify the SU(n) representation labels in (2.6) with the corresponding Young tableau.

Let us now write down the final form of A_{YT} for the most interesting Young tableaux. Thus we fix a representation □ of SU(n), $N > 2$; its dimension is $N \geqslant n$. The lowest case is $A_{\square}$ given by (2.6). Every other A_{YT} is then expressed according to (3.1) as a multiple of $A_{\square}$:

$$A_{\square\square} = (N+4)A_{\square},$$

$$A_{\begin{smallmatrix}\square\\\square\end{smallmatrix}} = (N-4)A_{\square},$$

$$A_{\square\square\square} = \tfrac{1}{2}(N+3)(N+6)A_{\square},$$

$$A_{\begin{smallmatrix}\square\square\\\square\end{smallmatrix}} = (N-3)(N+3)A_{\square},$$

$$A_{\begin{smallmatrix}\square\\\square\\\square\end{smallmatrix}} = \tfrac{1}{2}(N-3)(N-6)A_{\square},$$

$$A_{\square\square\square\square} = \tfrac{1}{6}(N+3)(N+4)(N+8)A_{\square},$$

$$A_{\begin{smallmatrix}\square\square\square\\\square\end{smallmatrix}} = \tfrac{1}{2}(N+4)(N^2+N-8)A_{\square},$$

$$A_{\begin{smallmatrix}\square\square\\\square\square\end{smallmatrix}} = \tfrac{1}{2}(N-4)(N^2-N-8)A_{\square},$$

$$A_{\begin{smallmatrix}\square\\\square\\\square\\\square\end{smallmatrix}} = \tfrac{1}{6}(N-3)(N-4)(N-8)A_{\square},$$

$$A_{\begin{smallmatrix}\square\square\\\square\square\end{smallmatrix}} = \tfrac{1}{3}N(N-4)(N+4)A_{\square},$$

and also,

$$A_{\square\cdots\square\ (k)} = \frac{(N+2k)(N+k)!}{(N+2)!(k-1)!}A_{\square} = \frac{N+2k}{(k-1)!}(N+3)(N+4)\cdots(N+k)A_{\square},$$

$$A_{\begin{smallmatrix}\square\\\vdots\ (k)\\\square\end{smallmatrix}} = \frac{(N-2k)(N-3)!}{(N-k-1)!(k-1)!}A_{\square} = \frac{N-2k}{(k-1)!}(N-3)(N-4)\cdots(N-k)A_{\square}, \quad N \geqslant k,$$

$$A_{\square\cdots\square\ (k)} = \frac{(N+k-1)!(N+k+1)}{(N+2)!(k-1)!} \times [N^2+N(2k-5)-2k-2]A_{\square},$$

$$A_{\begin{smallmatrix}\square\square\\\vdots\ (k)\\\square\end{smallmatrix}} = \frac{(N-3)!(N-k-1)}{(N-k)!(k-1)!} \times [N^2-N(2k-5)-2k-2]A_{\square}, \quad N \geqslant k.$$

IV. GENERALIZATIONS

The summation in the definition (2.1) of $A_{(\lambda)}$ is over cubes of the projections w of the weights $\mathbf{w}$ of the representation (λ). Similarly one can consider the quantity

$$A^{(k)}_{(\lambda)} = a_k \sum_{\mathbf{w}} w^k, \quad k = 0,1,2,\cdots, \tag{4.1}$$

that is, the "k th anomaly number" of representation (λ) of a simple Lie group G. Here w is the orthogonal projection of the weight $\mathbf{w}$ on a suitably chosen one-dimensional subspace of the weight space, a convenient choice being, for instance, the direction of the highest weight of the defining representation of G. Clearly $A^{(3)}_{(\lambda)}$ is the triangle anomaly number of previous sections if $G = \mathrm{SU}(n)$ and the projection of $\mathbf{w}$ is done on the U(1) direction as before.

First however, let us introduce a different normalization convention. Instead of fixing the value of $A^{(k)}_{(\lambda)}$ for a chosen representation (λ), we fix the length of the basis vectors in weight space, i.e., the lengths of the simple roots. Such a normalization is independent of the value of k so that one can write

$$A^{(k)}_{(\lambda)} = \sum_{\mathbf{w}} w^k. \tag{4.2}$$

To be specific, we put $(\alpha,\alpha) = 2$, where (α,α) is the square of the length of the simple root α of the group G. For the groups of types O($2n+1$), Sp($2n$), G(2), and F(4) which have simple roots of two different lengths, α is the longer of the two.

From the properties of weight systems of representations of simple Lie groups one has the following properties of $A^{(k)}_{(\lambda)}$:

$$A^{(0)}_{(\lambda)} = \text{dimension of } (\lambda), \tag{4.3}$$

$$A^{(1)}_{(\lambda)} = 0, \tag{4.4}$$

$$A^{(2k+1)}_{(\lambda)} = 0 \text{ for } G \neq E_6, \text{ and } G \neq \mathrm{SU}(n),\ n > 2, \tag{4.5}$$

$$A^{(k)}_{(\lambda)\oplus(\lambda')} = A^{(k)}_{(\lambda)} + A^{(k)}_{(\lambda')}, \tag{4.6}$$

$$A^{(k)}_{(\lambda)\otimes(\lambda')} = \sum_{j=0}^{k} \binom{k}{j} A^{(k-j)}_{(\lambda)} A^{(j)}_{(\lambda')}, \tag{4.7}$$

$$A^{(k)}_{(\lambda)} = (-1)^k A^{(k)}_{(\bar{\lambda})}, \tag{4.8}$$

$A^{(2k)}_{(\lambda)} \neq 0$ for all nontrivial representations of any simple G.

$$A^{(2)}_{(\lambda)} = (1/l) I^{(2)}_{(\lambda)}, \tag{4.9}$$

$$A^{(4)}_{(\lambda)} = [3/l(l+2)] I^{(4)}_{(\lambda)}$$

for $G = \mathrm{SU}(2), E_6, E_7, E_8, F_4, G_2$, (4.10)

where l is the rank of the group and $I^{(m)}_{(\lambda)}$ denotes the mth degree index[7,8] of the representation (λ); the symbols E_6, E_7, E_8, F_4, G_2 stand for the exceptional simple Lie groups. The property (4.10) is a nontrivial consequence of symmetry properties[7] of weight systems of representations of the groups indicated. Extensive tables if $I^{(2)}$ and $I^{(4)}$ are found in Ref. 8, as well as a description of their properties.

Suppose that $G = G_1 \times G_2$, where G_1 and G_2 are simple. If (λ_1) and (λ_2) denote respectively the representations of G_1 and G_2, we use $(\lambda_1)(\lambda_2)$ for the corresponding representation of $G_1 \times G_2$. Then

$$A^{(k)}_{(\lambda_1)(\lambda_2)} = \sum_{j=0}^{k} \binom{k}{j} A^{(j)}_{(\lambda_1)} A^{(k-j)}_{(\lambda_2)}, \tag{4.11}$$

which resembles closely (4.7) except that here (λ_1) and (λ_2) are representations of different groups.

The relation (2.6) generalizes into

$$A^{(k)}_{(\lambda)} = \rho^{(k)} A^{(k)}_{(\lambda')}, \quad k = 2,3, \tag{4.12}$$

where (λ) and (λ') represent the group G and the subgroup $G' \subset G$ respectively. The constant $\rho^{(k)}$ is independent of (λ). Equality (4.12) holds also for $k = 4$ when G is one of the simple exceptional Lie groups and G' is any semisimple subgroup.[7] For even $k > 4$ equation (4.12) does not hold[7]; for odd values of $k > 4$ the equation has not been investigated.

V. COMMENTS

(1) We defined the triangle anomaly number of an SU(n) representation in terms of the sum of cubes of U(1) projections of weights. The sum of cubes of projections on any other direction is proportional to the sum of cubes of U(1) projections, the proportionality factor being less than or equal to unity in magnitude, and independent of the representation.

(2) The triangle anomaly number of any representation of any simple group other than SU(n) vanishes when defined as the sum of cubes of any component of weights. For groups other than O($4n+2$) and E_6 the statement is obvious, since their representations are all self-contragredient. For O($4n+2$), $n \geqslant 2$, the statement follows from the fact that the Weyl formula[7,9] for the character contains no third power of the dummy variables which carry the components of the weight as exponents. In the absence of an explicit character formula for E_6, the statement for E_6 is most easily verified by considering the $\mathrm{SU}(3) \times \mathrm{SU}(3) \times \mathrm{SU}(3)$ content of E_6 irreducible representations. Each subgroup representation $(a)(b)(c)$ appears together with $(c)(a)(b)$, $(b)(c)(a)$, $(\bar{a})(\bar{c})(\bar{b})$, $(\bar{c})(\bar{b})(\bar{a})$, $(\bar{b})(\bar{a})(\bar{c})$ (the set is smaller if some of its members coincide). Application of formula (4.11) shows that the triangle anomaly of this set of subgroup representations vanishes.

(3) All properties of the number $A^{(k)}_{(\lambda)}$, $k = 0,1,\cdots$, were deduced from properties of weight systems of finite-dimensional representations of semisimple Lie groups (algebras). Since the weight system of a representation (λ) does not change if a complex group is reduced to any of its real forms, the anomaly number $A^{(k)}_{(\lambda)}$ does not feel the difference between the complex group and any of its real forms.

(4) From (4.9) and (4.10) one can see that the anomaly numbers $A^{(k)}_{(\lambda)}$ and the indices $I^{(2k)}_{(\lambda)}$ are similar objects. Indeed, the former are sums of powers of certain projections of weights, while the latter are similar sums of powers of lengths of weights.[7,8] In particular, $A^{(0)}_{(\lambda)} = I^{(0)}_{(\lambda)}$ is the dimension of (λ). Many of their properties coincide and derivations of explicit expressions for them follow the same path,[7] the simplest of them being the derivation of Weyl's dimension formula for (λ). Next in order of increasing complexity are the expressions for $I^{(2)}_{(\lambda)}$, or $A^{(2)}_{(\lambda)}$ [see Eq. (4.9)]. Formulas for $A^{(3)}_{(\lambda)}$ are calculated from an equivalent definition in Refs. 2 and 5; the result is

$$A^{(3)}_{(\lambda)} = \frac{2N_{(\lambda)}}{n}\left(\frac{p_3(l)}{(n+1)(n+2)} + \frac{p_3(l) - p_1(l)p_2(l)}{(n-1)(n+1)} + \frac{p_3(l) - [p_1(l)]^3 - 2p_1(l)p_2(l)}{(n-1)(n-2)}\right), \tag{5.1}$$

where n labels the group SU(n), and $p_\alpha(l)$ is defined by

$$\prod_{i=1}^{n} (1 - zl_i)^{-1} = \sum_{\alpha=0}^{\infty} p_\alpha(l_1,\cdots,l_n) z^\alpha. \tag{5.2}$$

The representation labels l_j are given in terms of the nonnegative integer labels λ; by

$$l_j = \sum_{k=j}^{n-1} \lambda_k + n - j, \quad 1 \leqslant j \leqslant n-1,\ l_n = 0. \tag{5.3}$$

Formulas for $I^{(4)}_{(\lambda)}$ for all simple groups are again in Ref. 7. Because of (4.10), $A^{(4)}_{(\lambda)}$ is known for SU(2) and the five exceptional simple Lie groups. No other formulas are known for $A^{(k)}_{(\lambda)}$ or $I^{(2k)}_{(\lambda)}$.

(5) Recently a generalization of the group–subgroup relation, called subjoining, was found.[10] It requires consideration of formal linear combinations of representation with integer coefficients. If the coefficients are nonnegative the sum is the familiar direct sum. Here we would like to point out that one can extend the properties of $A^{(k)}$ in such a way that linear combinations of representations with negative coefficients have well-defined values of $A^{(k)}$. It suffices to put

$$A^{(k)}_{(\lambda)-(\lambda')} = A^{(k)}_{(\lambda)} - A^{(k)}_{(\lambda')}.$$

(6) The examples of Sec. III reveal symmetries between A_{YT}'s corresponding to two Young tableaux which differ by interchange of rows and columns. One is obtained from the other by changing certain signs. These symmetries arise from those of the coefficients $C^{(\nu)}_{(\mu)}$ in (A3). We have encountered a similar symmetry previously[11] while considering plethysms based on representations of SU(2) only.

ACKNOWLEDGMENTS

We are grateful to Dr. P. Frampton for bringing the problem to our attention and for stimulating discussions. One of us (J.P.) would like to thank the Aspen Center for Physics for its hospitality. We are grateful to the referee who suggested the simple proof in Sec. II of the equivalence of our anomaly number and the usual one.

APPENDIX

It is the purpose of this appendix to derive Eq. (3.1), which states that $A_{\rm YT}$ is a multiple of $A_{\square}$. The proportionality factor $F_{\rm YT}(N)$ depends on the Young tableau YT and on the dimension N of the representation $\square$, but not otherwise on $\square$, or indeed, on the group SU(n) of which $\square$ is a reducible or irreducible representation. More specifically, $F_{\rm YT}(N)$ is a polynomial of degree $p-1$ in N, where p is the number of boxes in the Young tableau YT.

In what follows $\square$ may denote the set of N weights $\mathbf{w}$ belonging to the representation $\square$. We denote by $\square^{\mu}$ the set of N weights $\mu\mathbf{w}$, i.e., the weights of $\square$, with the scale dilated by a factor μ. The triangle anomaly number of $\square^{\mu}$ is

$$A_{\square^{\mu}} = \mu^3 A_{\square}. \tag{A1}$$

Now consider the direct product

$$[\mu] = [\mu_1,\mu_2,\cdots,\mu_k] = \square^{\mu_1}\otimes\square^{\mu_2}\otimes\cdots\otimes\square^{\mu_k}, \tag{A2}$$

where $(\mu_1,\mu_2,\cdots,\mu_k)$ is a partition of p into nonnegative integers. The direct products corresponding to different partitions are clearly independent, as long as $N\geqslant p$.

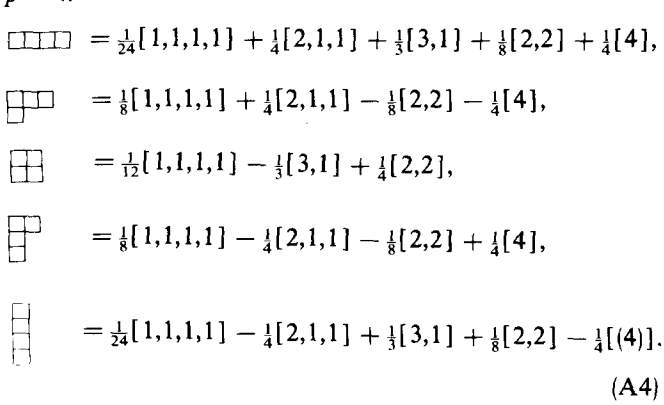
A Young tableau may be associated with a partition $(\nu)=(\nu_1,\nu_2,\cdots,\nu_k)$ of p. The presence of ν_i indicates a row of ν_i boxes. The Young tableaux are independent, as long as $N\geqslant p$.

Suppose for the time being that $N\geqslant p$. The Young tableaux with p boxes and the direct products $[\mu]$ with $\Sigma_i\mu_i = p$ span the same space, so either can be expressed as a linear combination of the other. We may write

$$\mathrm{YT}_{(\nu)} = \sum_{[\mu]} C^{(\nu)}_{[\mu]}[\mu]. \tag{A3}$$

The direct product $[1,1,\cdots,1]$ (p ones) is the only one which contains products of p different weights from $\square$. It appears with nonzero coefficient $C^{(\nu)}_{[1,1,\cdots,1]}$ on the right-hand side of every Young tableau.

For illustration we given the relations (A3) for the case $p=4$.



$$= \tfrac{1}{24}[1,1,1,1] + \tfrac{1}{4}[2,1,1] + \tfrac{1}{3}[3,1] + \tfrac{1}{8}[2,2] + \tfrac{1}{4}[4],$$

$$= \tfrac{1}{8}[1,1,1,1] + \tfrac{1}{4}[2,1,1] - \tfrac{1}{8}[2,2] - \tfrac{1}{4}[4],$$

$$= \tfrac{1}{12}[1,1,1,1] - \tfrac{1}{3}[3,1] + \tfrac{1}{4}[2,2],$$

$$= \tfrac{1}{8}[1,1,1,1] - \tfrac{1}{4}[2,1,1] - \tfrac{1}{8}[2,2] + \tfrac{1}{4}[4],$$

$$= \tfrac{1}{24}[1,1,1,1] - \tfrac{1}{4}[2,1,1] + \tfrac{1}{3}[3,1] + \tfrac{1}{8}[2,2] - \tfrac{1}{4}[(4)]. \tag{A4}$$

We now consider the situation $N\leqslant p$. Then there will be Young tableaux containing columns with more than N boxes. Equations (A3) remain valid provided Young tableaux with more than N boxes on the left-hand side are replaced by zero; the direct products $[\mu]$ are no longer independent. Nonvanishing Young tableaux are still given correctly by (A3). We emphasize that the coefficients $C^{(\nu)}_{[\mu]}$ depend only on the Young tableau and on the μ_i which label the direct product $[\mu]$.

Now Eq. (2.4) is easily generalized to give the triangle anomaly number of the direct product of k sets of weights.

$$A_{(\lambda_1)\otimes\cdots\otimes(\lambda_k)} = \left(\prod_{j=1}^{k} N_{(\lambda_j)}\right)\sum_{i=1}^{k} A_{(\lambda_i)}/N_{(\lambda_i)}, \tag{A5}$$

and hence for the direct product $[\mu_1,\mu_2,\cdots,\mu_k]$ we have, using (A1),

$$A_{[\mu]} = N^{k-1}\left(\sum_{i=1}^{k}\mu_i^3\right)A_{\square}. \tag{A6}$$

Since the largest value taken by k on the right-hand side of (A3) is p, it follows that $A_{\rm YT}$ is $A_{\square}$ times a polynomial of degree $p-1$ in N whose coefficients depend only on YT.

The best way to determine the polynomial $F_{\rm YT}(N)$ is from the formula (5.1) with $\square$ the defining representation of SU(n). Since the irreducible representations (λ) of SU(n) correspond to Young tableaux based on $\square$, the dependence of $A_{\rm YT}$ on N can be ascertained once and for all.

We complete this Appendix by giving a proof of Eq. (2.13).

Write $A_{(\lambda')} = A'_{\lambda}$; it is the sum of cubes of the projections of the weights of (λ) on some direction in weight space [specifically the U(1) direction of SU(m) weight space]. Write $A_{(10\cdots0)} = \rho A'_{(10\cdots0)}$. With the help of (3.1) we can write down $A'_{(\lambda)}$ for any irreducible representation (λ). We have

$$\begin{aligned} A'_{(\lambda)} &= A'_{\rm YT} = F_{\rm YT}(N)A'_{(10\cdots0)} \\ &= \rho^{-1}F_{\rm YT}(N)A_{(10\cdots0)} = \rho^{-1}A_{(\lambda)}. \end{aligned} \tag{A7}$$

YT in (A7) is the Young Tableau corresponding to the representation (λ).

[1] D. Gross and R. Jackiw, Phys. Rev. D **6**, 477 (1972).
[2] S. Okubo, Phys. Rev. D **16**, 3528 (1977).
[3] H. Georgi and S. Glashow, Phys. Rev. D **6**, 429 (1972).
[4] P. H. Frampton, Phys. Rev. Lett. **43**, 1912 (1979); P. H. Frampton and S. Nandi, Phys. Rev. Lett. **43**, 1460 (1979); P. H. Frampton, Phys. Lett. B **89**, 352 (1980); P. H. Frampton, Phys. Lett. B **88**, 299 (1979).
[5] J. Banks and H. georgi, Phys. Rev. D **14**, 1159 (1976).
[6] R. Slansky, "Group Theory for Unified Model Building," to be published in Phys. Rep. (1981).
[7] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976); Erratum **18**, 1519 (1977).
[8] W. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[9] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328,377 (1926).
[10] J. Patera and R. T. Sharp, J. Phys. A **13**, 397 (1980); M. Couture and R. T. Sharp, J. Phys. A **13**, 1925 (1980); J. Patera, R. T. Sharp, and R. Slansky, J. Math. Phys. **21**, 2335 (1980); R. V. Moody, "Lie Algebra Subjoining," C. R. Math. Rep. Acad. Sci. Canada **11**, 259 (1980); R. C. King, J. Phys. A **13**, 3563 (1980).
[11] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 261 (1981).

Wigner 9j symbols and the product group

C. Earp [a)]
Physics and Engineering Laboratory, Department of Scientific and Industrial Research, Wellington, New Zealand

(Received 23 July 1980; accepted for publication 19 September 1980)

The 9j symbols of a simply reducible Lie group G are related to the 3j symbols of the product group $G \times G$. It is shown that this relationship leads to certain identities satisfied by the 9j symbols. Two known identities involving 6j and 9j symbols are shown to be reduced forms of a new identity involving 9j symbols.

PACS numbers: 02.20.Qs

1. INTRODUCTION

The properties of Wigner's 6j symbols are now well understood[1]; in particular it is known that the 6j symbols satisfy two independent identities.[2] The usual method of deriving these identities is by considering different coupling schemes for three and four angular momentum vectors.

Wigner's 9j symbols have not been so thoroughly investigated, perhaps because of their relative complexity. Arima *et al.*[3] reviewed the known symmetry properties (previously derived by Jahn and Hope[4]), and derived an orthogonality relation and a number of identities. Their derivations were based on coupling schemes for four and five vectors. More recently, Jang[5] derived two further identities, involving both 6j and 9j symbols, from a consideration of the symmetry properties of the 15j symbols. This method, although indirect, is equivalent to the method of coupling schemes but is less tedious.

This paper deals with an alternative method, in which the coupling problem is transferred to a larger group. In the case of the three-dimensional rotation group SO(3), this larger group is the four-dimensional rotation group SO(4). It is interesting to note that recently[6] a new recurrence relation for the 3j symbols of SO(3) was derived from the properties of SO(4). Thus, although the properties of the larger group are usually determined from those of the smaller, the reverse procedure also has some merit.

2. THE PRODUCT GROUP

Let G be a simply reducible[7] Lie group, and denote by G^2 the Cartesian product group $G \times G$. It may be easily verified[8] that G^2 is also simply reducible, and therefore we may apply Wigner's[9] analysis of the properties of its coupling and recoupling coefficients. The notation of Ref. 9 will be used throughout, with a summation convention over greek and primed roman indices. An asterisk denotes complex conjugation.

The representations of G will be denoted by

$$\begin{bmatrix} j & R \\ \kappa & \lambda \end{bmatrix},$$

where R is an element of G. A basis for the representations of G^2 may be written in terms of tensor products of the representations of G:

$$\begin{bmatrix} (j,k) & (R,S) \\ (\kappa,\alpha) & (\lambda,\beta) \end{bmatrix} = \begin{bmatrix} j & R \\ \kappa & \lambda \end{bmatrix} \otimes \begin{bmatrix} k & S \\ \alpha & \beta \end{bmatrix}. \tag{2.1}$$

It follows that the 3j symbols of G^2 associated with the basis (2.1) are given, in terms of the 3j symbols of G, by

$$\begin{pmatrix} (j,m) & (k,n) & (l,p) \\ (\kappa,\alpha) & (\lambda,\beta) & (\mu,\gamma) \end{pmatrix} = \begin{pmatrix} j & k & l \\ \kappa & \lambda & \mu \end{pmatrix} \begin{pmatrix} m & n & p \\ \alpha & \beta & \gamma \end{pmatrix}. \tag{2.2}$$

Similarly, the 6j and 9j symbols are given by

$$\begin{Bmatrix} (a,d) & (b,e) & (c,f) \\ (g,j) & (h,k) & (i,l) \end{Bmatrix} = \begin{Bmatrix} a & b & c \\ g & h & i \end{Bmatrix} \begin{Bmatrix} d & e & f \\ j & k & l \end{Bmatrix}, \tag{2.3}$$

and

$$\begin{Bmatrix} (a,d) & (b,e) & (c,f) \\ (g,j) & (h,k) & (i,l) \\ (m,q) & (n,r) & (p,s) \end{Bmatrix} = \begin{Bmatrix} a & b & c \\ g & h & i \\ m & n & p \end{Bmatrix} \begin{Bmatrix} d & e & f \\ j & k & l \\ q & r & s \end{Bmatrix}. \tag{2.4}$$

A unitary transformation is now made to a different basis of G^2, defined by

$$\begin{bmatrix} (j,k) & (R,S) \\ (x,y) & (\phi,\lambda) \end{bmatrix} = \{[x][y]\}^{1/2} \begin{pmatrix} j & k & x \\ \alpha & \gamma & \phi \end{pmatrix} \begin{pmatrix} j & k & y \\ \beta & \delta & \lambda \end{pmatrix}^* \times \begin{bmatrix} (j,k) & (R,S) \\ (\alpha,\gamma) & (\beta,\delta) \end{bmatrix}. \tag{2.5}$$

This transformation is similar to the decomposition of the Kronecker product of representations of G. The 3j symbols associated with the new basis can be calculated by recoupling methods and the result is[10]

$$\begin{pmatrix} (j,m) & (k,n) & (l,p) \\ (q,\phi) & (r,\rho) & (s,\sigma) \end{pmatrix} = \{[q][r][s]\}^{1/2} \begin{pmatrix} q & r & s \\ \phi & \rho & \sigma \end{pmatrix}^* \times \begin{Bmatrix} j & m & q \\ k & n & r \\ l & p & s \end{Bmatrix}. \tag{2.6}$$

[a)]Present address: Applied Mathematics Division, Department of Scientific and Industrial Research, P. O. Box 1335, Wellington, New Zealand.

 0022-2488/81/112357-03$01.00

However, the $6j$ and $9j$ symbols are not affected by transformations such as (2.5), and so the recoupling coefficients of G^2 in the new basis are still given by (2.3) and (2.4).

3. A NEW IDENTITY

Equation (2.6) can be used to derive a number of properties for the $9j$ symbols of G; orthogonality relations for example. The purpose of this note is to demonstrate a new identity or sum rule for the $9j$ symbols.

Wigner's $9j$ symbols satisfy two identities invoving $3j$ symbols. For G^2, these are[9]

$$\begin{Bmatrix}(a,d) & (b,e) & (c,f)\\ (g,j) & (h,k) & (i,l)\\ (m,q) & (n,r) & (p,s)\end{Bmatrix}$$

$$= \begin{pmatrix}(a,d) & (b,e) & (c,f)\\ (r',\alpha) & (s',\beta) & (t',\gamma)\end{pmatrix}\begin{pmatrix}(g,j) & (h,k) & (i,l)\\ (u',\delta) & (v',\epsilon) & (w',\zeta)\end{pmatrix}$$

$$\times\begin{pmatrix}(m,q) & (n,r) & (p,s)\\ (x',\eta) & (y',\theta) & (z',\kappa)\end{pmatrix}\begin{pmatrix}(a,d) & (g,j) & (m,q)\\ (r',\alpha) & (u',\delta) & (x',\eta)\end{pmatrix}^*$$

$$\times\begin{pmatrix}(b,e) & (h,k) & (n,r)\\ (s',\beta) & (v',\epsilon) & (y',\theta)\end{pmatrix}^*\begin{pmatrix}(c,f) & (i,l) & (p,s)\\ (t',\gamma) & (w',\zeta) & (z',\kappa)\end{pmatrix}^*, \tag{3.1}$$

and[11]

$$\begin{pmatrix}(m,q) & (n,r) & (p,s)\\ (x,\eta) & (y,\theta) & (z,\kappa)\end{pmatrix}^*\begin{Bmatrix}(a,d) & (b,e) & (c,f)\\ (g,j) & (h,k) & (i,l)\\ (m,q) & (n,r) & (p,s)\end{Bmatrix}$$

$$= \begin{pmatrix}(a,d) & (b,e) & (c,f)\\ (r',\alpha) & (s',\beta) & (t',\gamma)\end{pmatrix}\begin{pmatrix}(g,j) & (h,k) & (i,l)\\ (u',\delta) & (v',\epsilon) & (w',\zeta)\end{pmatrix}$$

$$\times\begin{pmatrix}(a,d) & (g,j) & (m,q)\\ (r',\alpha) & (u',\delta) & (x,\eta)\end{pmatrix}^*\begin{pmatrix}(b,e) & (h,k) & (n,r)\\ (s',\beta) & (v',\epsilon) & (y,\theta)\end{pmatrix}^*$$

$$\times\begin{pmatrix}(c,f) & (i,l) & (p,s)\\ (t',\gamma) & (w',\zeta) & (z,\kappa)\end{pmatrix}^* \tag{3.2}$$

$$\begin{Bmatrix}a & b & c\\ g & h & i\\ m & n & p\end{Bmatrix}\begin{Bmatrix}d & e & f\\ j & k & l\\ q & r & s\end{Bmatrix}\begin{Bmatrix}m & n & p\\ q & r & s\\ x & y & z\end{Bmatrix} = [r'][s'][t'][u'][v'][w']$$

$$\times\begin{Bmatrix}r' & s' & t'\\ u' & v' & w'\\ x & y & z\end{Bmatrix}\begin{Bmatrix}a & b & c\\ d & e & f\\ r' & s' & t'\end{Bmatrix}\begin{Bmatrix}g & h & i\\ j & k & l\\ u' & v' & w'\end{Bmatrix}$$

$$\times\begin{Bmatrix}a & d & r'\\ g & j & u'\\ m & q & x'\end{Bmatrix}\begin{Bmatrix}b & e & s'\\ h & k & v'\\ n & r & y'\end{Bmatrix}\begin{Bmatrix}c & f & t'\\ i & l & w'\\ p & s & z'\end{Bmatrix}. \tag{3.4}$$

Use of (2.4) and (2.6) and (3.1) gives our result

$$\begin{Bmatrix}a & b & c\\ g & h & i\\ m & n & p\end{Bmatrix}\begin{Bmatrix}d & e & f\\ j & k & l\\ q & r & s\end{Bmatrix} = [r'][s'][t'][u'][v'][w'][x'][y'][z']$$

$$\times\begin{pmatrix}r' & s' & t'\\ \alpha & \beta & \gamma\end{pmatrix}^*\begin{pmatrix}u' & v' & w'\\ \delta & \epsilon & \zeta\end{pmatrix}^*\begin{pmatrix}x' & y' & z'\\ \eta & \theta & \kappa\end{pmatrix}^*$$

$$\times\begin{pmatrix}r' & u' & x'\\ \alpha & \delta & \eta\end{pmatrix}\begin{pmatrix}s' & v' & y'\\ \beta & \epsilon & \theta\end{pmatrix}\begin{pmatrix}t' & w' & z'\\ \gamma & \zeta & \kappa\end{pmatrix}$$

$$\times\begin{Bmatrix}a & d & r'\\ b & e & s'\\ c & f & t'\end{Bmatrix}\begin{Bmatrix}g & j & u'\\ h & k & v'\\ i & l & w'\end{Bmatrix}\begin{Bmatrix}m & q & x'\\ n & r & y'\\ p & s & z'\end{Bmatrix}$$

$$\times\begin{Bmatrix}a & d & r'\\ g & j & u'\\ m & q & x'\end{Bmatrix}\begin{Bmatrix}b & e & s'\\ h & k & v'\\ n & r & y'\end{Bmatrix}\begin{Bmatrix}c & f & t'\\ i & l & w'\\ p & s & z'\end{Bmatrix}$$

$$= [r'][s'][t'][u'][v'][w'][x'][y'][z']\begin{Bmatrix}r' & s' & t'\\ u' & v' & w'\\ x' & y' & z'\end{Bmatrix}$$

$$\times\begin{Bmatrix}a & b & c\\ d & e & f\\ r' & s' & t'\end{Bmatrix}\begin{Bmatrix}g & h & i\\ j & k & l\\ u' & v' & w'\end{Bmatrix}\begin{Bmatrix}m & n & p\\ q & r & s\\ x' & y' & z'\end{Bmatrix}$$

$$\times\begin{Bmatrix}a & d & r'\\ g & j & u'\\ m & q & x'\end{Bmatrix}\begin{Bmatrix}b & e & s'\\ h & k & v'\\ n & r & y'\end{Bmatrix}\begin{Bmatrix}c & f & t'\\ i & l & w'\\ p & s & z'\end{Bmatrix}. \tag{3.3}$$

Similarly, from (3.2) we find the alternative form

Some special cases of (3.4) are of interest. If x, y, and z refer to the unit representation, (3.4) reduces to

$$\begin{Bmatrix} a & b & c \\ g & h & i \\ m & n & p \end{Bmatrix} \begin{Bmatrix} j & k & l \\ d & e & f \\ m & n & p \end{Bmatrix} = [r'][s'][t'] \begin{Bmatrix} d & r' & a \\ g & m & j \end{Bmatrix} \begin{Bmatrix} e & s' & b \\ h & n & k \end{Bmatrix} \times \begin{Bmatrix} f & t' & c \\ i & p & l \end{Bmatrix} \begin{Bmatrix} a & b & c \\ d & e & f \\ r' & s' & t' \end{Bmatrix} \begin{Bmatrix} j & k & l \\ g & h & i \\ r' & s' & t' \end{Bmatrix} . \tag{3.5}$$

On the other hand, if a and d refer to the unit representation, the reduced form of (3.4) is

$$\begin{Bmatrix} g & n & p \\ b & i & h \end{Bmatrix} \begin{Bmatrix} j & r & s \\ e & l & k \end{Bmatrix} \begin{Bmatrix} j & r & s \\ g & n & p \\ x & y & z \end{Bmatrix} = [t'][v'][w'] \begin{Bmatrix} x & y & z \\ t' & v' & w' \end{Bmatrix} \times \begin{Bmatrix} g & h & i \\ j & k & l \\ x & w' & v' \end{Bmatrix} \begin{Bmatrix} b & e & t' \\ h & k & w' \\ n & r & y \end{Bmatrix} \begin{Bmatrix} e & b & t' \\ l & i & v' \\ s & p & z \end{Bmatrix} . \tag{3.6}$$

Equation (3.6) can also be derived from (2.3) by expressing the $6j$ symbol in terms of four $3j$ symbols.

The identities (3.5) and (3.6) were derived separately by Jang[5]; they are shown above to be particular cases of the same equation. The author has attempted to find a similar simple connection between (3.3) or (3.4) and the identities of Arima *et al.*,[3] without success, although it is possible to prove (3.4) by a lengthy calculation from the formulas of Ref. 3.

ACKNOWLEDGMENTS

The author wishes to thank Dr. G. Weir for his advice and comments during the preparation of the manuscript, and Dr. S. White for reading the manuscript.

[1]See, for example, *The Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965).
[2]L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Rev. Mod. Phys. **24**, 249 (1952).
[3]A. Arima, H. Horie, and Y. Tanabe, Prog. Theor. Phys. (Kyoto) **11**, 143 (1954).
[4]H. A. Jahn and J. Hope, Phys. Rev. **93**, 318 (1954).
[5]S. Jang, J. Math. Phys. **9**, 397 (1968).
[6]A. O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[7]E. P. Wigner, Am. J. Math. **63**, 57 (1941).
[8]For the case where G is SO(3), see L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961). Biedenharn's argument can be extended to the general case without difficulty.
[9]E. P. Wigner, "On the Matrices Which Reduce the Kronecker Products of Representations of Simply Reducible Groups," in Ref. 1, p. 87.
[10]For the case where G is SO(3), see A. Z. Dolginov, Sov. Phys. - JETP **3**, 746 (1956).
[11]The identity (3.2) is hinted at in Ref. 9; it arises during the derivation of Eq. (3.1) above, and also of the result (2.6). For the case where the group (here G^2) is SO(3), Eq. (3.2) was published by Jahn and Hope (Ref. 4).

Nonscalar extension of shift operator techniques for SU (3) in an O(3) basis. I. Theory

H. E. De Meyer[a] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent Krijgslaan 271-S9, B-9000 Gent, Belgium

J. W. B. Hughes
Department of Applied Mathematics, Queen Mary College, Mile End Road, London E1 4NS, United Kingdom

(Received 17 February 1981; accepted for publication 17 April 1981)

A set of relations is set up which connect quadratic products of the shift operators O_l^k $(k = 0,1,2)$, which are nonscalar with respect to the O(3) subgroup of SU (3). The usefulness of these relations is illustrated by the calculation of the eigenvalues of the scalar shift operator O_l^0 for various irreducible representations (p,q) of SU (3).

PACS numbers: 02.20.Qs

I. INTRODUCTION

The problem of l-degeneracies in the reduction of irreducible representations (p,q) of SU (3) with respect to its O(3) subgroup has been considered already two decades ago by various authors. The construction of bases, and particularly orthonormal bases, for SU (3) representations in the SU(3) $\supset$O(3) reduction has been the aim of numerous authors. In fact, most of the approaches which have been proposed can be viewed as searches for an additional Hermitian operator of which the eigenfunctions form a basis. It has been shown that only two such independent operators exist, and in the present paper we shall denote them by O_l^0 and Q_l^0. Many alternative notations have been introduced, and these are summarized by Partensky and Quesne,[1] who also study their interconnection. An excellent review of all common efforts and results obtained until 1975, more in particular on the missing label problem and the construction of bases, has been given by Moshinsky *et al.*[2]

In the present paper attention is drawn to a recursive method for calculating O_l^0 and Q_l^0 eigenvalues, which has been developed by Hughes.[3,4] The technique essentially relies on a set of relations among products of shift operators which behave as O(3) scalars. These shifts operators are constructed out of the three generators l_0, $l_\pm$ of O (3) and the five generators q_μ $(-2 \leqslant \mu \leqslant 2)$, which form a five-dimensional tensor representation of O(3). Later on, the O(3) shift operator construction has been generalized by Hughes and Yadegar[5] in such a way that a generally valid algorithm could be established, whereas many properties of shift operators were reconsidered. Although the shift operator technique has proven very successful for eigenvalue calculations,[4] it nevertheless should be criticized on the point that such calculations can become extremely tedious, especially when l-degeneracy occurs. In fact, whatever method is used an increase of l-degeneracy always involves new supplementary problems, and this is the main reason why a case of three-fold degeneracy has never been treated in a completely analytical way before.

Recently the shift operator technique has been applied by two of us[6] for solving the quadrupole phonon state labelling problem, which is related to the analysis of symmetric representations of O(5) into irreducibles of its O(3) subgroup. However, instead of setting up relations among scalar triple product operators, we have established relations among quadratic shift operator products which are no longer O(3) scalars. Motivated by the relevant advantages and simplifications which the latter type of equations induce in the O(5) case, we want to reinvestigate in the present paper the SU(3) state labelling problem.

In Sec. 2 we shall derive the so-called nonscalar relations and we shall discuss their conformity with certain relations among scalar triple product operators. Sections 3 and 4 are concerned with a complete analysis of the $(p,0)$ and $(p,1)$ representations, respectively, whereas in Sec. 5 we investigate the lowest angular momentum states.

2. NONSCALAR PRODUCT OPERATORS

In this paper we shall be concerned with the SU(3) shift operators $O_l^{\pm k}$ $(k = 0,1,2)$ which shift the eigenvalues of the O(3) Casimir operator L^2 by $\pm k$, in the form they have been originally defined by Hughes,[3] in terms of SU(3) generators. Also, we shall, without loss of generality, restrict our attention to SU(3) states which correspond to zero m, the eigenvalue of the diagonal O(3) generator l_0. In that case the shift operators can be written down as follows:

$$\begin{cases} O_l^0 = (\sqrt{6})l(l+1)q_0 - 3(q_{+1}l_- + q_{-1}l_+) - 3(q_{+2}l_-^2 + q_{-2}l_+^2), \\ O_l^{-1} = l(l-1)(q_{+1}l_- - q_{-1}l_+) - l(q_{+2}l_-^2 - q_{-2}l_+^2), \\ O_l^{-2} = (\sqrt{6})l^2(l-1)^2 q_0 - 2l(l-1)^2(q_{+1}l_- + q_{-1}l_+) + l(l-1)(q_{+2}l_-^2 + q_{-2}l_+^2), \\ O_l^{+1} = O_{-l-1}^{-1}, \\ O_l^{+2} = O_{-l-1}^{-2}, \end{cases} \tag{2.1}$$

[a]Research Associate N.F.W.O. (Belgium).

 0022-2488/81/112360-06$01.00

where $l(l+1)$ is the eigenvalue of L^2. For a detailed description of the SU(3) algebra of generators q_μ ($\mu = 0, \pm 1, \pm 2$), l_0 and $l_\pm$, the reader is referred to Ref. 3.

Previously, a set of very useful formulas, containing scalar shift operator products, the Casmirs I_2 and I_3, and a second Hermitian O(3) scalar operator Q_l^0 of fourth order in the group generators, has been set up by Hughes.[3] For later use, we recall here some of these formulas in slightly altered form:

$$(l+2)^2(O_l^0)^2 + 3\frac{(2l+3)(2l+5)}{(l+1)^2}O_{l+1}^{-1}O_l^{+1} + 3\frac{(l+3)}{(l+1)(l+2)^2}O_{l+2}^{-2}O_l^{+2} - 18(l+1)^2(l+2)^2(2l+3)^2[12I_2 - l(l+4)] = 0, \tag{2.2}$$

$$(2l+1)(O_l^0)^2 - 3\frac{l(l-2)}{(l+1)^2}O_{l+1}^{-1}O_l^{+1} + 3\frac{(l+1)(l+3)}{l^2}O_{l-1}^{+1}O_l^{-1} - 54l^2(l+1)^2(2l+1)(4I_2+1) = 0, \tag{2.3}$$

$$(l-1)^2(O_l^0)^2 + 3\frac{(2l-1)(2l-3)}{l^2}O_{l-1}^{+1}O_l^{-1} + 3\frac{(l-2)}{l(l-1)^2}O_{l-2}^{+2}O_l^{-2} - 18l^2(l-1)^2(2l-1)^2[12I_2 - (l+1)(l-3)] = 0. \tag{2.4}$$

Notice that Eq. (2.4) can be deduced from Eq. (2.2) if in the latter l is replaced formally by $-l-1$. All the relations (2.2)–(2.4) share the property that they contain only operators or operator products which leave the l-value of the states upon which they act invariant. One can, however, easily construct bilinear forms of shift operators which shift l by one, two, or even three units, and hence the question immediately arises as to whether there also exist relations connecting such product operators. By straightforward calculation it is readily verified that this question can be answered affirmatively. In fact, it can be shown that there are two independent relations among operators which lower l by one, namely,

$$(l+3)(2l+1)O_l^{-1}O_l^0 - (l+1)(2l+3)O_{l-1}^0O_l^{-1} - \frac{6l}{(l+1)^2}O_{l+1}^{-2}O_l^{+1} = 0, \tag{2.5}$$

$$(l-1)(2l-3)O_l^{-1}O_l^0 - (l-3)(2l-1)O_{l-1}^0O_l^{-1} - \frac{6l}{(l-1)^2}O_{l-2}^{+1}O_l^{-2} = 0, \tag{2.6}$$

one relation connecting operators which shift l by -2, i.e.,

$$(l-1)(l-2)O_l^{-2}O_l^0 - l(l+1)O_{l-2}^0O_l^{-2} - 6(2l-1)O_{l-1}^{-1}O_l^{-1} = 0, \tag{2.7}$$

and finally a last relation among the two quadratic operators which lower l by three units, i.e.,

$$O_{l-1}^{-2}O_l^{-1} - O_{l-2}^{-1}O_l^{-2} = 0. \tag{2.8}$$

The corresponding relations between operators that raise the l-value of the state upon which they act to change by, respectively, one, two, or three units are immediately deduced from Eqs. (2.5)–(2.8) by formally replacing l by $-(l+1)$ and by using $O_l^{+k} = O_{-(l+1)}^{-k}$. It is worthwhile mentioning that none of the relations (2.2)–(2.8) contain terms linear in the shift operators. This is essentially due to the fact that the commutator of any two generators q_μ is independent of the q's and can be expressed entirely in terms of $l_\pm$ and l_0.

There is also a certain connection between the newly derived nonscalar relations (2.5)–(2.8) and the relations between triple product operators previously given by Hughes.[3] Let us illustrate this by an example. If we multiply each term of Eq. (2.6) on the left with O_{l-1}^{+1}, and each term of Eq. (2.5) on the left also with O_{l-1}^{+1}, and if we then eliminate from the resulting two equations the term in $O_{l-1}^{+1}O_{l-1}^0O_l^{-1}$, the following relation among triple product operators is obtained:

$$-\frac{(l+1)(2l+3)}{l^2(l-1)^2}O_{l-1}^{+1}O_{l-2}^{+1}O_l^{-2} + \frac{(l-3)(2l-1)}{l^2(l+1)^2} \times O_{l-1}^{+1}O_{l+1}^{-2}O_l^{+1} = -\tfrac{4}{7}O_{l-1}^{+1}O_l^{-1}O_l^0.$$

The same relation is reproduced on eliminating the Casimir I_3 from Eqs. (40) and (42) of Ref. 3 and by making thereafter use of Eq. (36) of the same reference. Similarly, from the moment we can eliminate I_3 from the triple product operator relations of Hughes,[3] it is possible to indicate a suitable combination of the relations (2.5)–(2.8) which reproduces the same result. However, if use is made only of the nonscalar relations (2.5)–(2.8), the relations of Ref. 3 among triple product operators cannot be deduced, since there is no way by which the SU(3) invariant I_3 can be introduced. Also, it will become clear that for a complete classification of SU(3) irreducible representations, at least one relation containing I_3 and of the type considered by Hughes is needed.

In the following sections expressions will be derived for the eigenvalues of the O(3) scalar shift operator O_l^0, and this for certain SU(3) irreducible representations which as usual are labelled by the integers p and q satisfying $p \geqslant q \geqslant 0$ and related to I_2 and I_3 by the formulas

$$\langle I_2 \rangle = \tfrac{1}{9}(p^2 + q^2 - pq + 3p), \tag{2.9}$$

$$\langle I_3 \rangle = \tfrac{1}{162}(p-2q)(2p+3-q)(p+q+3), \tag{2.10}$$

where $\langle\ \rangle$ denotes the expectation value between states. Also, without loss of generality, we can assume hereafter and in a forthcoming paper that $p - 2q \geqslant 0$. Finally, the exact l-content of (p,q), which has been extensively studied in the past by various authors,[2] will be considered as a well established fact throughout this paper.

3. COMPLETE ANALYSIS OF THE $(p,0)$ REPRESENTATIONS

The $(p,0)$ representations have been studied already in detail by Hughes[4] by means of standard shift operator techniques. The relative ease with which the eigenvalue calculations could be performed rests essentially on the fact that no l degeneracy occurs in these representations. Nevertheless, it

is worthwhile to demonstrate how the presently derived nonscalar relations (2.5)–(2.8) bring about even more considerable simplifications into that type of calculation.

Indeed, let us start from the well-known property that l can take on only the nondegenerate values $p, p-2, p-4, \ldots$, 0 or 1, depending on whether p is even or odd. Since furthermore no confusion can arise, we suppress p and q in the basis vectors and since, moreover, we consider states corresponding to $m=0$, we can simply denote them by $|l\rangle$. Let us also denote the O_l^0 eigenvalue by λ_l. Now, if we select an acceptable l-value such that $|l\rangle$ exists and $|l-1\rangle$ is nonexistent and if we momentarily suppose and that λ_l is already known, we immediately derive from the action of Eq. (2.7) upon $|l\rangle$ that

$$\lambda_{l-2} = \frac{(l-1)(l-2)}{l(l+1)}\lambda_l. \tag{3.1}$$

Furthermore, the application of Eq. (2.8) upon $|l\rangle$ proves that $|l-3\rangle$ does not exist. Hence, with the only assumptions that $|p\rangle$ exists, that λ_p is known, and that $|p-1\rangle$ vanishes, recursion of (3.1) gives

$$\lambda_{p-2k} = \frac{(p-2k)(p-2k+1)}{p(p+1)}\lambda_p, \quad (k=0,1,2,\ldots), \tag{3.2}$$

whereas it also follows that $|p-2k-1\rangle (k=1,2,\ldots)$ are nonexistent states. Putting $l=p-2k$ in (3.2), this relation is rewritten as

$$\lambda_l = \frac{l(l+1)}{p(p+1)}\lambda_p, \ (p-l=\text{even}). \tag{3.3}$$

As a consequence, the problem is reduced to that of finding an expression for the eigenvalue λ_p belonging to the maximum angular momentum state $|p\rangle$. To that aim, we let Eq. (2.2) act upon $|p\rangle$ and by use of (2.9) we immediately can deduce that λ_p^2 takes on the value prescribed by the formula

$$\lambda_p^2 = 6p^2(p+1)^2(2p+3)^2. \tag{3.4}$$

In order to account for the correct sign of λ_p, we must use a triple product operator relation. Following Hughes[4] we arrive at

$$\lambda_p = (\sqrt{6})p(p+1)(2p+3). \tag{3.5}$$

It has to be noticed that the eigenvalues λ_l have been obtained without the necessity, as was the case with the standard shift operator technique, of having precise knowledge of matrix elements of the form $\langle l|O_{l-2}^{+2}O_l^{-2}|l\rangle$.

4. COMPLETE ANALYSIS OF THE $(p,1)$ REPRESENTATIONS

Assuming that we already know that no l-degeneracy occurs in $(p,1)$ representations, let us define c_l and d_l by

$$\begin{aligned} c_l &= \frac{1}{(l+1)^2}\langle l|O_{l+1}^{-1}O_l^{+1}|l\rangle \\ &= \frac{1}{(l+1)^2}\langle l+1|O_l^{+1}O_{l+1}^{-1}|l+1\rangle, \\ d_l &= \frac{1}{(l+1)^2(l+2)^2}\langle l|O_{l+2}^{-2}O_l^{+2}|l\rangle \\ &= \frac{1}{(l+1)^2(l+2)^2}(\langle l+2|O_l^{+2}O_{l+2}^{-2}|l+2\rangle). \end{aligned} \tag{4.1}$$

By suitable combination of the relations (2.2)–(2.4) and putting $l=p-k$, the following recursion relations can be derived:

$$\begin{aligned} c_{p-k-1} &= \frac{(p-k+1)(p-k+5)}{(p-k+2)^2}c_{p-k} + \frac{(2p-2k+1)}{(p-k+2)^2}d_{p-k} \\ &\quad - 24(p-k+1)^2(2p-2k+1)(2kp-p-k^3+3k+1), \\ d_{p-k-2} &= -4\frac{(2p-2k+1)}{(p-k+2)^2}c_{p-k} + \frac{(p-k)(p-k-4)}{(p-k+2)^2}d_{p-k} \\ &\quad + 48(p-k)(2p-2k+1)[2(p+1)^2+(p-k)(p-k+1)(p-k-5)], \end{aligned} \tag{4.2}$$

subject to the initial conditions $c_p = d_p = d_{p-1} = 0$. These relations are easily solved

$$\begin{aligned} \langle p-2k-1|O_{p-2k}^{-1}O_{p-2k-1}^{+1}|p-2k-1\rangle &= \langle p-2k|O_{p-2k-1}^{+1}O_{p-2k}^{-1}|p-2k\rangle \\ &= 24(p+1)(p-2k)^2(p-2k+1)(p-2k-1)(2p-2k+1), \end{aligned} \tag{4.3}$$

$$\begin{aligned} \langle p-2k|O_{p-2k+1}^{-1}O_{p-2k}^{+1}|p-2k\rangle &= \langle p-2k+1|O_{p-2k}^{+1}O_{p-2k+1}^{-1}|p-2k+1\rangle \\ &= 48k(p+1)(p-2k+1)^2(p-2k)(p-2k+2), \end{aligned} \tag{4.4}$$

$$\begin{aligned} \langle p-2k|O_{p-2k+2}^{-2}O_{p-2k}^{+2}|p-2k\rangle &= \langle p-2k+2|O_{p-2k}^{+2}O_{p-2k+2}^{-2}|p-2k+2\rangle \\ &= 48k(p-2k+1)^3(p-2k+2)^3(p-2k+3)(p-2k)(2p-2k+3), \end{aligned} \tag{4.5}$$

$$\begin{aligned} \langle p-2k-1|O_{p-2k+1}^{-2}O_{p-2k-1}^{+2}|p-2k-1\rangle &= \langle p-2k+1|O_{p-2k-1}^{+2}O_{p-2k+1}^{-2}|p-2k+1\rangle \\ &= 48k(p-2k+1)^3(p-2k)^3(p-2k+2)(p-2k-1)(2p-2k+1). \end{aligned} \tag{4.6}$$

Using (2.2) we then get

$$\lambda_{p-2k-1}^2 = 6(2p^3-8kp^2+3p^2+8k^2p-16kp-11p+20k^2+10k-6)^2, \tag{4.7}$$

$$\lambda_{p-2k}^2 = 6(2p^3-8kp^2+p^2+8k^2p-7p-4k^2+2k-6)^2, \tag{4.8}$$

where again λ_l is a short notation for the matrix element $\langle l|O_l^0|l\rangle$. At this point it becomes very tedious to obtain the sign of λ_{p-2k-1} and λ_{p-2k} with the usual techniques; therefore we try the nonscalar relations. If we let Eq. (2.5) act upon $|l\rangle$ we can prove that

$$(l+3)(2l+1)\lambda_l-(l+1)(2l+3)\lambda_{l-1}=\frac{6l}{(l+1)^2}\delta\left[\frac{c_l d_{l-1}}{c_{l-1}}\right]^{1/2}, \tag{4.9}$$

where δ is a sign factor not yet determined. Basic to the proof are the properties

$$|\langle l-k|O_l^{-k}|l\rangle|^2=\frac{1}{\alpha_{k,l-k}}\langle l|O_{l-k}^{+k}O_l^{-k}|l\rangle, \tag{4.10}$$

$$|\langle l+k|O_l^{+k}|l\rangle|^2=\alpha_{k,l}\langle l|O_{l+k}^{-k}O_l^{+k}|l\rangle, \tag{4.11}$$

with

$$\alpha_{k,l}=\frac{2l+1}{2l+2k+1}, \tag{4.12}$$

which are satisfied as long as $|l\rangle, |l\pm k\rangle$ are nondegenerate. If we replace l by $p-2k$, in (4.9) and then substitute herein the expressions for c_{p-2k}, d_{p-2k-1}, and c_{p-2k+1}, which are obtained from (4.3)–(4.6), we arrive at

$$(p-2k+3)(2p-4k+1)\lambda_{p-2k}-(p-2k+1)\times(2p-4k+3)\lambda_{p-2k-1}=24(\sqrt{6})\delta k(p-2k)^2(p-2k+2). \tag{4.13}$$

The only way to satisfy this equation with what we already know, from (4.7) and (4.8), is to define λ_{p-2k} and λ_{p-2k-1} as

$$\lambda_{p-2k-1}=\epsilon(\sqrt{6})(2p^3-8kp^2+3p^2+8k^2p-16kp-11p+20k^2+10k-6)\quad(q=1), \tag{4.14}$$

$$\lambda_{p-2k}=\epsilon(\sqrt{6})(2p^3-8kp^2+p^2+8k^2p-7p-4k^2+2k-6)\quad(q=1), \tag{4.15}$$

whereby ϵ equals $+1$ or -1, but has the same sign in (4.14) and (4.15). Of course ϵ can still be k-dependent, meaning that it could change sign for different pairs of consecutive eigenvalues. However, repeating the same calculations which have led us to (4.13), but choosing $l=p-2k-1$, we observe at the end that λ_{p-2k-1} and λ_{p-2k-2} should also be attributed the same ϵ, whereas λ_{p-2k-2} is found from (4.15) on replacing k by $k+1$. Hence, we can conclude that (4.14) and (4.15) are valid for all possible k values. It remains to determine what is the correct sign of ϵ. To that purpose we take recourse to a reasoning expounded already elsewhere[4] and based upon triple product relations, by which we can prove that the eigenvalue of the highest angular momentum state $|p\rangle$ is positive, just as was the case for $q=0$. Consequently ϵ has to be set equal to $+1$ in (4.14) and (4.15), as can be readily verified by putting $k=0$ in (4.15).

5. ANALYSIS OF THE LOW ANGULAR MOMENTUM STATES

As a next application of the extended shift operator technique, we study the situation where l is attributed a small numerical value, i.e., 0, 1,2, or 3. According to well-known properties concerning the exact l-content of representations, we have to distinguish between two cases:

(i) p even, q even: there is one $l=0$ state, one $l=3$ state, in general two $l=2$ states, and the $l=1$ state is always absent;

(ii) p odd, $q\neq 0$, $q\neq 1$ or p even, q odd: there is one $l=1$ state, one $l=2$ state, in general two $l=3$ states, but no $l=0$ state.

Notice that in case (ii) we had to exclude $q=0$ and $q=1$ when p is odd. For these two q values, however, we can immediately refer to the results of the previous sections. In the following, the orthonormal eigenvectors of O_l^0 will be denoted by $|l\rangle$ if l is nondegenerate, and by $|l,(i)\rangle$, $i=1,2,\ldots,n$, if l is n-degenerate. Also, we keep the notation λ_l or $\lambda_l^{(i)}$ for the corresponding eigenvalue(s).

Let us first investigate the situation described under (i). From the expression (2.1) of O_l^0 in terms of the generators it immediately follows that

$$\lambda_0=0,\quad (p \text{ even}, q \text{ even}). \tag{5.1}$$

Furthermore, the action of Eq. (2.4) upon both $l=2$ states $|2,(1)\rangle$ and $|2,(2)\rangle$ yields for both eigenvalues a similar condition, i.e.,

$$(\lambda_2^{(i)})^2=2^3 3^5[4\langle 2,(i)|I_2|2,(i)\rangle+1]\quad(i=1,2). \tag{5.2}$$

or with the use of (2.9),

$$(\lambda_2^{(i)})^2=2^3 3^3[4(p^2+q^2-pq+3p)+9]\quad(i=1,2). \tag{5.3}$$

Since both eigenvalues are necessarily different we can set

$$\lambda_2^{(i)}=(-1)^{i-1}6(\sqrt{6})[4(p^2+q^2-pq+3p)+9]^{1/2},\quad(i=1,2)\quad(p \text{ even}, q \text{ even}). \tag{5.4}$$

The result (5.4) agrees with that of Judd *et al.*[7] As a next step we replace l by $-(l+1)$ in Eq. (2.5), to obtain the relation

$$(l-2)(2l+1)O_l^{+1}O_l^0-l(2l-1)O_{l+1}^0O_l^{+1}+\frac{6(l+1)}{l^2}O_{l-1}^{+2}O_l^{-1}=0, \tag{5.5}$$

which when applied to the $l=2$ states yields

$$O_3^0O_2^{+1}|2,(i)\rangle=0, \tag{5.6}$$

immediately showing that

$$\lambda_3=0,\quad (p \text{ even}, q \text{ even}) \tag{5.7}$$

which again agrees with Judd's[7] result.

Although we have obtained all the required eigenvalues in closed form, we can as a supplement try to construct the eigenstates from appropriate shift operator actions. Therefore, we assume only that there is a minimum angular momentum state $|0\rangle$ with associated eigenvalue $\lambda_0=0$. With this knowledge, we let Eq. (2.3) act upon $|0\rangle$, which yields

$$O_l^{-1}O_0^{+1}|0\rangle=0, \tag{5.8}$$

meaning that since $|0\rangle$ is the minimum state, $|1\rangle$ is absent. As a consequence, we obtain, from the action of Eq. (2.2) upon $|0\rangle$, that

$$O_2^{-2}O_0^{+2}|0\rangle=3456\,I_2|0\rangle. \tag{5.9}$$

Now we define two $l=2$ states by means of shift operator

actions, namely one state $|2,(a)\rangle$ by

$$O_0^{+2}|0\rangle = |2,(a)\rangle, \tag{5.10}$$

and a second state $|2,(b)\rangle$ as

$$O_2^{-2}|2,(b)\rangle = 0. \tag{5.11}$$

Note that we reserve the notation $|2,(i)\rangle$, $i = 1,2$ for the orthonormal eigenstates of O_2^0, whereas the states $|2,(a)\rangle$ and $|2,(b)\rangle$ can be *a priori* any linear combinations of $|2,(1)\rangle$ and $|2,(2)\rangle$. Letting Eq. (2.4) first and then Eq. (2.3) act upon $|2,(a)\rangle$ and $|2,(b)\rangle$, we obtain

$$\begin{cases} (O_2^0)^2|2,(a)\rangle = 1944(4I_2+1)|2,(a)\rangle, \\ O_3^{-1}O_2^{+1}|2,(a)\rangle = 0, \\ (O_2^0)^2|2,(b)\rangle = 1944(4I_2+1)|2,(b)\rangle, \\ O_3^{-1}O_2^{+1}|2,(b)\rangle = 38880\, I_2|2,(b)\rangle. \end{cases} \tag{5.12}$$

This shows that only one $l = 3$ state exists. Furthermore, we see that $(O_2^0)^2$ is diagonal in the $\{|2,(a)\rangle, |2(b)\rangle\}$ basis, but O_2^2 is not. Indeed, let us define

$$\begin{cases} O_2^0|2,(a)\rangle = \alpha_{11}|2,(a)\rangle + \alpha_{12}|2,(b)\rangle, \\ O_2^0|2,(b)\rangle = \alpha_{21}|2,(a)\rangle + \alpha_{22}|2,(b)\rangle, \end{cases} \tag{5.13}$$

then $\alpha_{22} = -\alpha_{11}$ since $(O_2^0)^2$ is diagonal, and $\alpha_{12} = \alpha_{21}$. The calculation of α_{11} is straightforward but necessitates the introduction of the scalar operator Q_l^0. In particular, if we let Eq. (45) of Ref. 3 act upon $|2,(a)\rangle$ and $|2,(b)\rangle$ we find

$$\begin{cases} Q_2^0|2,(a)\rangle = 108(4I_2-7)|2,(a)\rangle, \\ Q_2^0|2,(b)\rangle = -108(4I_2+7)|2,(b)\rangle \quad (p \text{ even}, q \text{ even}) \end{cases} \tag{5.14}$$

Hence, Q_2^0 is diagonal in the basis $\{|2,(a)\rangle, |2,(b)\rangle\}$. From Eq. (40) of the same paper, we obtain with the help of (5.14) that $\alpha_{11} = -108(\sqrt{6})\langle I_3\rangle/\langle I_2\rangle$, where $\langle I_j\rangle$ $(j = 2,3)$ is a short notation for $\langle 2,(a)|I_j|2,(a)\rangle = \langle 2,(b)|I_j|2,(b)\rangle$. Finally, α_{12} is calculated from α_{11} and the eigenvalue of $(O_2^0)^2$, and it is readily verified that $\alpha_{12}^2 = -\alpha_{11}^2 - 1944(4\langle I_2\rangle + 1)$. The sign of α_{12} is arbitrary, since a change of that sign only corresponds to an interchange of $|2,(a)\rangle$ and $|2,(b)\rangle$. From the foregoing we may conclude that by only the assumption of the existence of a nondegenerate lowest angular-momentum state $|0\rangle$, we have been able to construct the other low angular momentum states explicitly using shift operator actions.

We now proceed with the analysis of case (ii). Here it appears that we cannot derive the O_1^0 eigenvalue directly. But, we learn immediately from the action of Eq. (2.6) upon $|2\rangle$ that the O_2^0 eigenvalue is proportional to the O_1^0 eigenvalue, namely,

$$\lambda_2 = -3\lambda_1, \quad \begin{pmatrix} p \text{ odd}, q \neq 0, q \neq 1 \\ p \text{ even}, q \text{ odd} \end{pmatrix} \tag{5.15}$$

To our knowledge, this property has never been mentioned explicitly before. Next, we let Eq. (2.7) act on each of the $l = 3$ states $|3,(1)\rangle$ and $|3,(2)\rangle$. Taking into account that $|\langle 1|O_2^{-1}|2\rangle| = (5\langle 1|O_2^{-1}O_1^{+1}|1\rangle/3)^{1/2}$ and that the state $|2\rangle$ is nondegenerate, we obtain

$$\{(\lambda_3^{(i)} - 6\lambda_1)\langle 1|O_3^{-2} - 15\sqrt{\tfrac{5}{3}}(\langle 1|O_2^{-1}O_1^{+1}|1\rangle)^{1/2} \times \langle 2|O_3^{-1}\}|3,(i)\rangle = 0 \quad (i = 1,2). \tag{5.16}$$

Similarly, if we let Eq. (2.6) act upon $|3,(i)\rangle (i = 1,2)$, and use $|\langle 2|O_1^{+1}|1\rangle| = (3\langle 1|O_2^{-1}O_1^{+1}|1\rangle/5)^{1/2}$, it follows that

$$\{4\lambda_3^{(i)}\langle 2|O_3^{-1} - 3\sqrt{\tfrac{3}{5}}(\langle 1|O_2^{-1}O_1^{+1}|1\rangle)^{1/2}\langle 1|O_3^{-2}\}|3,(i)\rangle = 0 \quad (i = 1,2). \tag{5.17}$$

Equations (5.16) and (5.17) are consistent if

$$4\lambda_3^{(i)}(\lambda_3^{(i)} - 6\lambda_1) - 45\langle 1|O_2^{-1}O_1^{+1}|1\rangle = 0 \quad (i = 1,2). \tag{5.18}$$

Next, applying Eq. (2.3) upon $|1\rangle$ gives

$$4\lambda_1^2 + \langle 1|O_2^{-1}O_1^{+1}|1\rangle - 864[4\langle I_2\rangle + 1] = 0. \tag{5.19}$$

Elimination of $\langle 1|O_2^{-1}O_1^{+1}|1\rangle$ from (5.18) and (5.19) yields

$$(\lambda_3^{(i)})^2 - 6\lambda_1\lambda_3^{(i)} - 45[216(4\langle I_2\rangle + 1) - \lambda_1^2] = 0 \quad (i = 1,2). \tag{5.20}$$

Hence, since $\lambda_3^{(1)}$ is necessarily different from $\lambda_3^{(2)}$, we can set

$$\lambda_3^{(i)} = 3\{\lambda_1 + 2(-1)^{i-1}[270(4\langle I_2\rangle + 1) - \lambda_1^2]^{1/2}\} \quad (i = 1,2), \tag{5.21}$$

Consequently, we have managed to express the eigenvalues λ_2 and $\lambda_3^{(i)}$ directly in terms of λ_1. In order to calculate the latter eigenvalue, we make use of Eq. (45) of Ref. 3, which after the replacement of l by $-(l+1)$ reads

$$Q_l^0 = \frac{1}{(l-1)^2}\left(\frac{O_{l-1}^{+1}O_l^{-1}}{l^2} + \frac{O_{l-2}^{+2}O_l^{-2}}{l^2(l-1)^2}\right) + 6l[-12(2l-1)I_2 + (l+1)(2l^2 - 10l - 9)]. \tag{5.22}$$

The action of (5.22) upon $|2\rangle$ produces

$$Q_2^0 = \tfrac{1}{4}O_1^{+1}O_2^{-1} - 108[4I_2 + 7]$$

which, due to (5.15) and (5.19), reduces to

$$Q_2^0 = 108(4I_2 - 5) - \tfrac{1}{6}\lambda_2^2. \tag{5.23}$$

Substituting this result in Eq. (40) with $l = 2$ of Ref. 3 we obtain

$$\lambda_2^3 - 2^3 3^4(3\langle I_2\rangle + 1)\lambda_2 + 2^5 3^7(\sqrt{6})\langle I_3\rangle = 0$$

or, with the help of (5.15),

$$\lambda_1^3 - 2^3 3^2(3\langle I_2\rangle + 1)\lambda_1 - 2^5 3^4(\sqrt{6})\langle I_3\rangle = 0. \tag{5.24}$$

This cubic equation yields three different solutions for λ_1, which after substitution of (2.9) and (2.10) into (5.24) read

$\lambda_1^{(1)} = -2(\sqrt{6})(p - 2q)$, or $\lambda_1^{(2)} = 2(\sqrt{6})(2p - q + 3)$,

or

$$\lambda_1^{(3)} = -2(\sqrt{6})(p + q + 3). \tag{5.25}$$

By numerical verification of a few (p,q)-representations it can be intuitively understood that $\lambda_1^{(1)}$ is the appropriate eigenvalue expression for the case that p is even and q is odd, $\lambda_1^{(2)}$ if p is odd and q even, and $\lambda_1^{(3)}$ if p and q are both odd.

6. DISCUSSION

A first and evident step in the development of shift operator techniques for the classification and analysis of irreducible representations of SU(3), is to discern the independent relationships between the O(3) scalar bilinear forms of the shift operators and the Casimirs L^2, l_0, and I_2. By this, how-

ever, one can only obtain relations in which O_l^0, the scalar shift operator, occurs quadratically. As a consequence, these relations alone are insufficient with regard to the O_l^0 eigenvalue and eigenstate determination. Also, the sign ambiguity of the eigenvalue corresponding to the highest angular momentum state, a fact that we observed at several occasions, is an immediate aspect of the quadratic occurrence of O_l^0. In order to transform the shift operator technique into a complete and adequate calculation method, relations among O(3) scalar triple product operators have been established and incorporated by one of us previously.[3,4] Although this opened the way to calculation successes, the fact that Q_l^0 and $[O_l^0, Q_l^0]$, where Q_l^0 is another O(3) scalar operator, enter these relations makes them somehow less handsome in practice. Therefore, we have envisaged at the beginning of this paper considering also nonscalar products of shift operators and analyzing the relationships between these objects. We found at the lowest level, namely, that of quadratic products, a set of formulas that permitted us to carry out, on a variety of illustrative examples, the eigenvalue calculation in a much shorter and sometimes almost immediate way. In fact, the most striking importance of the presently derived nonscalar relations is that O_l^0 occurs linearly. It should be remarked that none the less we could not forget completely about the triple product relations, and this is mainly due to the fact that the second independent SU(3) Casimir I_3 of third order in the group generators only enters equations which are at least of third order in the shift operators too.

Finally, the examples which we have quoted in the present paper have in common that either one of the SU(3) representation labels, either the angular momentum has been given a fixed numerical value. In a forthcoming paper we shall insist on the extreme usefulness of the relations between nonscalar products by also analyzing in the new context general (pq) representations.

ACKNOWLEDGMENT

The British Council and the National Science Foundation of Belgium are acknowledged by two of us (H.D.M. and G.V.B) for their financial support.

[1] A. Partensky and C. Quesne, J. Math. Phys. **20**, 2014 (1979).
[2] M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, Ann. Phys. **95**, 139 (1975), and references therein.
[3] J. W. B. Hughes, J. Phys. A **6**, 48 (1973).
[4] J. W. B. Hughes, J. Phys. A **6**, 281 (1973).
[5] J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[6] G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 2632 (1980).
[7] B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).

Nonscalar extension of shift operator techniques for SU (3) in an O(3) basis. II. Applications

H. E. De Meyer, [a] G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent Krijgslaan 271-S9, B-9000 Gent, Belgium
and
J. W. B. Hughes
Department of Applied Mathematics, Queen Mary College, Mile End Road, London E1 4NS, United Kingdom

(Received 17 February 1981; accepted for publication 17 April 1981)

In a preceding paper relations have been derived which connect nonscalar quadratic shift operator products. Here, the extreme usefulness of these relations is demonstrated by the example of the O_l^0-eigenvalue calculation for the cases $l = p - i$ ($i = 0,1,2,3$, and 4), where (p,q) is any SU (3) representation. For the first time a case of threefold l-degeneracy is completely solved in a purely analytical way.

PACS numbers: 02.20.Qs

I. INTRODUCTION

In a preceding paper[1] (to be referred to as I), it has been shown that the set of relations existing between nonscalar products of the SU (3) shift operators $O_l^{\pm k}(k = 0,1,2)$, constitutes a very useful complement to the standard shift operator calculus[2,3] especially with regard to the calculation of O_l^0 eigenvalues and the determination of mutually orthogonal eigenstates. Indeed, many of the results which, on use of the scalar product operator relations alone, were obtained only at the expense of laborious efforts, have been reproduced in I in a much easier and shorter way. Also, problems which were almost intractable with the usual shift operator techniques before, such as for example the complete analysis of the $(p,1)$ representations, were shown in I to become straightforwardly solvable on account of the nonscalar relations. Similar simplifications in eigenvalue calculations, caused by the inclusion of nonscalar product operators, have been also encountered recently by two of us while studying the group O(5).[4]

In order to raise the present extended shift operator technique to a level of applicability which can be compared to that of other calculation methods available,[5,6] we propose hereafter a full analysis of the highest angular momentum states of the general (p,q) representations. In the forthcoming sections our main concern will therefore be the determination in an analytical form of the O_l^0 eigenvalues if l equals consecutively $p, p-1, p-2, p-3$, and $p-4$. The last case is especially interesting since to our knowledge it has never been treated explicitly by any other technique before.

2. THE MAXIMUM ANGULAR MOMENTUM STATES

Since the labels p and q, although restricted in value by the inequalities $p \geqslant 2q \geqslant 0$, are kept arbitrary throughout, they can be suppressed in the states. As in I, we consider only states corresponding to $m = 0$. Hence, the O_l^0 eigenstates will be simply denoted by $|l,(i)\rangle$ $(i = 1,2,\ldots,n)$ if l is n-degenerate, and by $|l\rangle$ if l is nondegenerate. In order to conform with previous notations, the O_l^0 eigenvalues will be denoted by $\lambda_l^{(i)}$, i.e.,

$$O_l^0|l,(i)\rangle = \lambda_l^{(i)}|l,(i)\rangle. \tag{2.1}$$

The maximum angular momentum state $|p\rangle$ being nondegenerate, the square of λ_p is immediately found from the action of Eq. (I.2.2) on $|p\rangle$ and by using $I_2|l\rangle = \frac{1}{9}(p^2 + q^2 - pq + 3p)|l\rangle$. It follows that λ_p is given by

$$\lambda_p = \pm\sqrt{6}(p+1)(2p+3)(p-2q)$$

with either the + sign or the − sign. Unfortunately, none of Eqs. (I.2.2)–(I.2.8) can account for the correct sign. Therefore, we have to fall back upon triple product operator relations which contain the second invariant I_3. Following a reasoning expounded already in detail by one of us elsewhere,[3] we arrive at

$$\lambda_p = \sqrt{6}(p+1)(2p+3)(p-2q), \tag{2.2}$$

always on the assumption that $p - 2q \geqslant 0$.

3. EIGENVALUES OF THE STATES $|p-1\rangle$

For the derivation of the closed expression (2.2) for λ_p the nonscalar relations (I.2.5)–(I.2.8) were of no particular interest. Also, an analogous expression for λ_{p-1}, associated with the nondegenerate states $|p-1\rangle$, could be derived previously by using only relations connecting scalar products. However, the new relation (I.2.5) makes it possible to reproduce the same result in an immediate and nearly trivial way. Indeed, from the action of both sides of (I.2.5) upon $|p\rangle$ it follows that

$$(p+3)(2p+1)O_p^{-1}O_p^0|p\rangle = (p+1)(2p+3)O_{p-1}^0O_p^{-1}|p\rangle, \tag{3.1}$$

from which one easily deduces with the help of (2.2) that

$$\lambda_{p-1} = \sqrt{6}(p+3)(2p+1)(p-2q). \tag{3.2}$$

For the sake of completeness, it should be proved that the states $|p-1\rangle$ exist. In fact it is well known that this is not always true. But, this is easily verified, since by the action of Eq. (I.2.3) upon $|p\rangle$ one readily arrives at

[a] Research Associate N. F. W. O. (Belgium).

 0022-2488/81/112366-06$01.00

$$\langle p|O^{+1}_{p-1}O^{-1}_{p}|p\rangle = 24p^2(p+1)^2(2p+1)(p-q)q. \quad (3.3)$$

Hence, the state $|p-1\rangle$ exists only if $q\neq 0$ ($q=p$ does not occur due to the inequality $p-2q\geqslant 0$). Consequently, the expression (3.2) is valid only for $q\neq 0$, and also the case $q=0$ should therefore be treated separately if we continue to lower the l-value. This we have accomplished in I.

4. EIGENVALUES OF THE STATES $|p-2\rangle$

There are in general two ways of generating $|p-2\rangle$ states by the action of an l-lowering shift operator, namely $O^{-2}_{p}|p\rangle$ and $O^{-1}_{p-1}|p-1\rangle$. Hence we set

$$\begin{cases} O^{-2}_{p}|p\rangle = a^{(1)}_{p-2}|p-2,(1)\rangle + a^{(2)}_{p-2}|p-2,(2)\rangle, \\ O^{-1}_{p-1}|p-1\rangle = b^{(1)}_{p-2}|p-2,(1)\rangle + b^{(2)}_{p-2}|p-2,(2)\rangle, \end{cases} \quad (4.1)$$

where it is assumed that $|p-2,(1)\rangle$ and $|p-2,(2)\rangle$ are orthogonal to each other and normalized to unity if not equal to the zero vector. Note that we leave open the possibility that either $|p-2\rangle$ is nondegenerate or even does not exist.

Letting Eq. (I.2.4) act upon $|p\rangle$, it is straightforwardly deduced that

$$\langle p|O^{+2}_{p-2}O^{-2}_{p}|p\rangle = 24p^3(p-1)^2(2p+1)[p(p-2)(2p-1)+(p-2q)^2], \quad (4.2)$$

whereas the action of Eq. (I.2.3) upon $|p-1\rangle$ gives rise to

$$\langle p-1|O^{+1}_{p-2}O^{-1}_{p-1}|p-1\rangle = 24p(p+1)(p-1)^2[p(p-1)(p-2)-(p+1)(p-2q)^2]. \quad (4.3)$$

Now we take advantage of a nonscalar product relation by letting Eq. (I.2.5) act upon $|p-1\rangle$, which on account of the independence of $|p-2,(1)\rangle$ and $|p-2,(2)\rangle$, produces two relations which can be summarized as follows:

$$\sqrt{6}(p+2)(2p-1)(p+3)(2p+1)(p-2q)b^{(i)}_{p-2} - p(2p+1)\lambda^{(i)}_{p-2}b^{(i)}_{p-2} - \frac{6(p-1)}{p^2} \times [24p^2(p+1)^2(2p-1)(p-q)q]^{1/2}a^{(i)}_{p-2} = 0, \quad (i=1,2). \quad (4.4)$$

Here we also used the fact that

$$O^{+1}_{p-1}|p-1\rangle = \left[\frac{2p-1}{2p+1}\langle p-1|O^{-1}_{p}O^{+1}_{p-1}|p-1\rangle\right]^{1/2}|p\rangle \quad (4.5)$$

which, apart from an aribrary choice of phase, follows from a well-known general property,[2] and the fact that since $|p\rangle$ and $|p-1\rangle$ both exist and are nondegenerate (if $q\neq 0$), the matrix element in (4.5) can be inverted, i.e.,

$$\langle p-1|O^{-1}_{p}O^{+1}_{p-1}|p-1\rangle = \langle p|O^{+1}_{p-1}O^{-1}_{p}|p\rangle. \quad (4.6)$$

In a similar way, Eq. (I.2.7) acting upon $|p\rangle$ leads with the help of (3.3) and (4.1) to

$$\sqrt{6}(p-1)(p-2)(p+1)(2p+3)(p-2q)a^{(i)}_{p-2} - p(p+1)\lambda^{(i)}_{p-2}a^{(i)}_{p-2} - 6[24p^2(p+1)^2(2p+1)^2(2p-1)(p-q)q]^{1/2}b^{(i)}_{p-2} = 0, \quad (i=1,2). \quad (4.7)$$

It has to be emphasized that repeated occurrence of indices in (4.4) and (4.7) does not mean summation with respect to these indices. Equations (4.4) and (4.7) are four linear homogeneous equations with respect to $a^{(i)}_{p-2}$ and $b^{(i)}_{p-2}$ $(i=1,2)$. The determinantal condition which expresses that the solution differs from the trivial zero solution is a quadratic equation which has to be satisfied by $\lambda^{(1)}_{p-2}$ and $\lambda^{(2)}_{p-2}$, namely

$$\lambda^2 - 2\sqrt{6}(p+1)(2p+1)(p-2q)\lambda + 6(p-1)(2p-1)[(p^2-4)(2p^2+9p+9) - 4q(p-q)(2p^2+9p+1)] = 0. \quad (4.8)$$

Consequently, the eigenvalues of O^{0}_{p-2} can be brought into the following form:

$$\begin{cases} \lambda^{(1)}_{p-2} = \sqrt{6}\left[(p+1)(2p+1)(p-2q) + 6\sqrt{\Gamma}\right], \\ \lambda^{(2)}_{p-2} = \sqrt{6}\left[(p+1)(2p+1)(p-2q) - 6\sqrt{\Gamma}\right], \end{cases} \quad (4.9)$$

where

$$\Gamma = (p^2+p-1)^2 - 4p^2(p-q)q. \quad (4.10)$$

The substitution of the solution (4.9) into either Eqs. (4.4) or (4.7) leads to the following relations:

$$b^{(i)}_{p-2} = -\frac{(p^2+p-1)(p-2q)+(-1)^{i-1}p\sqrt{\Gamma}}{2p(2p+1)\sqrt{q(p-q)(2p-1)}}a^{(i)}_{p-2}, \quad (i=1,2). \quad (4.11)$$

It has to be noticed that in (4.11) infinities are excluded since neither q, $p-q$, nor p can be zero.

Finally, we want to investigate if there always exist two independent eigenstates of O^{0}_{p-2} in the case that $q\neq 0$. To that purpose, let us multiply each of the defining equations (4.1) on the left with its own Hermitian conjugate. On account of the orthonormality of eigenstates we obtain

$$\begin{cases} \sum_{i=1}^{2}|a^{(i)}_{p-2}|^2 = \frac{2p+1}{2p-3}\langle p|O^{+2}_{p-2}O^{-2}_{p}|p\rangle, \\ \sum_{i=1}^{2}|b^{(i)}_{p-2}|^2 = \frac{2p-1}{2p-2}\langle p-1|O^{+1}_{p-2}O^{-1}_{p-1}|p-1\rangle. \end{cases} \quad (4.12)$$

These equations, together with (4.2), (4.3), and (4.11), yield a unique solution for the unknowns $|a^{(i)}_{p-2}|$ and $|b^{(i)}_{p-2}|$, $(i=1,2)$. Solving the system explicitly for the $|a^{(i)}_{p-2}|$'s, namely,

$$|a^{(1)}_{p-2}|^2 = 24\frac{p^3(p-1)^2(2p+1)^2}{(2p-3)\sqrt{\Gamma}} \times\{-(p-2q)[2q(p-q)+(p-1)^2(p^2+p-1)] + [p(p-1)^2-2q(p-q)]\sqrt{\Gamma}\}, \quad (4.13)$$

$$|a^{(2)}_{p-2}|^2 = 24\frac{p^3(p-1)^2(2p+1)^2}{(2p-3)\sqrt{\Gamma}} \times\{(p-2q)[2q(p-q)+(p-1)^2(p^2+p-1)] + [p(p-1)^2-2q(p-q)]\sqrt{\Gamma}\}, \quad (4.14)$$

the $|b^{(i)}_{p-2}|$'s immediately follow from (4.11). Note that the right hand sides of (4.13) and (4.14) go over into each other on replacing $\sqrt{\Gamma}$ by $-\sqrt{\Gamma}$. Next, it can be easily verified that $|a^{(1)}_{p-2}|$, and hence also $|b^{(1)}_{p-2}|$, becomes equal to zero if

$q = 1$, whereas $|a^{(2)}_{p-2}|^2$ reduces to

$$|a^{(2)}_{p-2}|^2 = 24p^3(p-1)^3(p+1)(p-2)(2p+1)^2/(2p-3), \quad (q=1). \tag{4.15}$$

Since the eigenvector which we labeled as $|p-2,(1)\rangle$ vanishes identically for $q = 1$, the O^0_{p-2} eigenvalue is, in this case, given by the expression (4.9) for $\lambda^{(2)}_{p-2}$, which reduces to

$$\lambda_{p-2} = \sqrt{6}(2p^3 - 7p^2 + p - 8), \quad (q=1). \tag{4.16}$$

This expression agrees with formula (I.4.15) if in the latter k is replaced by one.

5. EIGENVALUES OF THE STATES $|p-3\rangle$

According to the results obtained in the previous section we shall from here on assume that $q \notin \{0,1\}$. There are three possible ways to define a $|p-3\rangle$ state by means of an l-lowering shift operator action upon higher angular momentum states, namely, by $O^{-2}_{p-1}|p-1\rangle$, $O^{-1}_{p-1}|p-2,(1)\rangle$, and $O^{-1}_{p-1}|p-2,(2)\rangle$. However, Eq. (I.2.8) applied on the $|p\rangle$ state immediately shows that at most two of these three actions can give rise to independent states. Hence, we set

$$\begin{cases} O^{-2}_{p-1}|p-1\rangle = a^{(1)}_{p-3}|p-3,(1)\rangle + a^{(2)}_{p-3}|p-3,(2)\rangle, \\ O^{-1}_{p-1}|p-2,(1)\rangle = b^{(1)}_{p-3}|p-3,(1)\rangle + b^{(2)}_{p-3}|p-3,(2)\rangle, \\ O^{-1}_{p-1}|p-2,(2)\rangle = c^{(1)}_{p-3}|p-3,(1)\rangle + c^{(2)}_{p-3}|p-3,(2)\rangle, \end{cases} \tag{5.1}$$

and the action of Eq. (I.2.8) upon $|p\rangle$ produces the following relations:

$$2\sqrt{6}(p+1)(2p+1)p[q(p-q)/(2p-1)]^{1/2}a^{(i)}_{p-3} = a^{(1)}_{p-2}b^{(i)}_{p-3} + a^{(2)}_{p-2}c^{(i)}_{p-3}, \quad (i=1,2), \tag{5.2}$$

where Eqs. (3.3) and (4.1) have been used explicitly. Next, if we act with Eq. (I.2.5) upon $|p-2,(i)\rangle$, $i = 1,2$, we obtain

$$\begin{aligned} &\sqrt{6}(p+1)(2p-3)\left[(p+1)(2p+1)(p-2q) + 6\sqrt{\Gamma}\right]b^{(i)}_{p-3} \\ &\quad - (p-1)(2p-1)\lambda^{(i)}_{p-3}b^{(i)}_{p-3} \\ &\quad - 6\frac{(p-2)}{(p-1)^2}\langle p-1|O^{+1}_{p-2}|p-2,(1)\rangle a^{(i)}_{p-3} = 0, \\ &\quad (i=1,2) \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} &\sqrt{6}(p+1)(2p-3)\left[(p+1)(2p+1)(p-2q) - 6\sqrt{\Gamma}\right]c^{(i)}_{p-3} \\ &\quad - (p-1)(2p-1)\lambda^{(i)}_{p-3}c^{(i)}_{p-3} \\ &\quad - 6\frac{(p-2)}{(p-1)^2}\langle p-1|O^{+1}_{p-2}|p-2,(2)\rangle a^{(i)}_{p-3} = 0, \\ &\quad (i=1,2). \end{aligned} \tag{5.4}$$

A first problem is to find general expressions for the matrix elements of O^{+1}_{p-2} which occur in (5.3) and (5.4). These can be obtained by inverting Eqs. (4.1), i.e.,

$$\begin{cases} \langle p-1|O^{+1}_{p-2}|p-2,(1)\rangle = \dfrac{\langle p-1|O^{+1}_{p-2}O^{-2}_{p}|p\rangle b^{(2)}_{p-2} - \langle p-1|O^{+1}_{p-2}O^{-1}_{p-1}|p-1\rangle a^{(2)}_{p-2}}{a^{(1)}_{p-2}b^{(2)}_{p-2} - a^{(2)}_{p-2}b^{(1)}_{p-2}}, \\ \langle p-1|O^{+1}_{p-2}|p-2,(2)\rangle = \dfrac{\langle p-1|O^{+1}_{p-2}O^{-1}_{p-1}|p-1\rangle a^{(1)}_{p-2} - \langle p-1|O^{+1}_{p-2}O^{-2}_{p}|p\rangle b^{(1)}_{p-2}}{a^{(1)}_{p-2}b^{(2)}_{p-2} - a^{(2)}_{p-2}b^{(1)}_{p-2}} \end{cases} \tag{5.5}$$

and by substituting herein the matrix element $\langle p-1|O^{+1}_{p-2}O^{-1}_{p-1}|p-1\rangle$ as given by (4.3) and the matrix element $\langle p-1|O^{+1}_{p-2}O^{-2}_{p}|p\rangle$ by its expression, which we can straightforwardly calculate from the action of Eq. (I.2.6) upon $|p\rangle$ and which reads

$$\langle p-1|O^{+1}_{p-2}O^{-2}_{p}|p\rangle = 48p^2(p-1)^2(p+1)(2p+1)(p-2q)[q(p-q)/(2p-1)]^{1/2}. \tag{5.6}$$

Expressing also the coefficients $b^{(i)}_{p-2)}(i = 1, 2)$ in terms of the $a^{(i)}_{p-2}$'s as dictated by (4.11), Eqs. (5.5) can be transformed into

$$\begin{cases} \langle p-1|O^{+1}_{p-2}|p-2,(1)\rangle = 24\dfrac{p^2(p-1)^2(p+1)(2p+1)[q(p-q)/(2p-1)]^{1/2}}{a^{(1)}_{p-2}\sqrt{\Gamma}} \\ \qquad \times\left[(p-2q)\sqrt{\Gamma} - (p^3 - 7p^2 + 7p - 2) - 4pq(p-q)\right], \\ \langle p-1|O^{+1}_{p-2}|p-2,(2)\rangle = 24\dfrac{p^2(p-1)^2(p+1)(2p+1)[q(p-q)/(2p-1)]^{1/2}}{a^{(2)}_{p-2}\sqrt{\Gamma}} \\ \qquad \times\left[(p-2q)\sqrt{\Gamma} + (p^3 - 7p^2 + 7p - 2) + 4pq(p-q)\right]. \end{cases} \tag{5.7}$$

As a next step, (5.7) is substituted into (5.3) and (5.4) and the $a^{(i)}_{p-3}$'s are expressed in terms of the $b^{(i)}_{p-3}$'s and $c^{(i)}_{p-3}$'s by means of Eq. (5.2). By this we arrive at

$$\begin{aligned} &\left[\sqrt{6}(p+1)^2(2p+1)(2p-3)(p-2q) + 6\sqrt{6}(p+1)(2p-3)\sqrt{\Gamma} - (p-1)(2p-1)\lambda^{(i)}_{p-3}\right]b^{(i)}_{p-3} - 12\sqrt{6}\frac{p(p-2)}{a^{(1)}_{p-2}\sqrt{\Gamma}} \\ &\quad \times\left[(p-2q)\sqrt{\Gamma} - (p^3 - 7p^2 + 7p - 2) - 4pq(p-q)\right]\left[a^{(1)}_{p-2}b^{(i)}_{p-3} + a^{(2)}_{p-2}c^{(i)}_{p-3}\right] = 0, \quad (i=1,2), \end{aligned} \tag{5.8}$$

$$\begin{aligned} &\left[\sqrt{6}(p+1)^2(2p+1)(2p-3)(p-2q) - 6\sqrt{6}(p+1)(2p-3)\sqrt{\Gamma} - (p-1)(2p-1)\lambda^{(i)}_{p-3}\right]c^{(i)}_{p-3} + 12\sqrt{6}\frac{p(p-2)}{a^{(2)}_{p-2}\sqrt{\Gamma}} \\ &\quad \times\left[(p-2q)\sqrt{\Gamma} + (p^3 - 7p^2 + 7p - 2) + 4pq(p-q)\right]\left[a^{(1)}_{p-2}b^{(i)}_{p-3} + a^{(2)}_{p-2}c^{(i)}_{p-3}\right] = 0, \quad (i=1,2). \end{aligned} \tag{5.9}$$

The system (5.8)–(5.9) of linear homogeneous equations with respect to the $b^{(i)}_{p-3}$'s and $c^{(i)}_{p-3}$'s yields a solution for these unknowns, which differs from the zero solution, if $\lambda^{(1)}_{p-3}$ and $\lambda^{(2)}_{p-3}$ are the roots of the following quadratic equation:

$$\lambda^2 - 2\sqrt{6}(p+3)(2p-1)(p-2q)\lambda + 6(p+1)(2p-3)[(p-2)(2p^3+15p^2+27p-54) - 4q(p-q)(2p^2+11p-3)] = 0. \tag{5.10}$$

Hence, we can make the choice

$$\begin{cases} \lambda^{(1)}_{p-3} = \sqrt{6}\left[(p+3)(2p-1)(p-2q) + 6\sqrt{\Lambda}\,\right], \\ \lambda^{(2)}_{p-3} = \sqrt{6}\left[(p+3)(2p-1)(p-2q) - 6\sqrt{\Lambda}\,\right], \end{cases} \tag{5.11}$$

where

$$\Lambda = 3(p-1)(p+1)(2p-3) + p^2(p-2q)^2. \tag{5.12}$$

It has to be noticed that in order to arrive at the closed expressions (5.11) for the O^0_{p-3} eigenvalues, the explicit forms of the $a^{(i)}_{p-2}$'s and $b^{(i)}_{p-2}$'s ($i = 1,2$) were not needed. Indeed, these coefficients, although they occur at many places in the intermediate steps, disappear in a natural way when the quadratic equation (5.10) is set up. If this property should turn out to be generally valid, it would mean that eigenvalue calculations can be carried out independently from the complete eigenstate construction. Hence, we believe at this point it is needless to evaluate explicitly the coefficients $a^{(i)}_{p-3}$, $b^{(i)}_{p-3}$ and $c^{(i)}_{p-3}$ ($i = 1,2$).

6. EIGENVALUES OF THE STATES $|p-4\rangle$

Our final purpose in the present paper is also to derive closed expressions for the O^0_{p-4} eigenvalues. This problem is particularly interesting, since for $q > 3$ one encounters for the first time a case of threefold l-degeneracy. Four l-lowering shift operator actions can account for the definition of $|p-4\rangle$ states, but Eq. (I.2.8) applied upon $|p-1\rangle$ shows that at most three of these states can be independent. Hence, we set

$$\begin{cases} O^{-2}_{p-2}|p-2,(1)\rangle \\ \quad = a^{(1)}_{p-4}|p-4,(1)\rangle + a^{(2)}_{p-4}|p-4,(2)\rangle + a^{(3)}_{p-4}|p-4,(3)\rangle, \\ O^{-2}_{p-2}|p-2,(2)\rangle \\ \quad = b^{(1)}_{p-4}|p-4,(1)\rangle + b^{(2)}_{p-4}|p-4,(2)\rangle + b^{(3)}_{p-4}|p-4,(3)\rangle, \\ O^{-1}_{p-3}|p-3,(1)\rangle \\ \quad = c^{(1)}_{p-4}|p-4,(1)\rangle + c^{(2)}_{p-4}|p-4,(2)\rangle + c^{(3)}_{p-4}|p-4,(3)\rangle, \\ O^{-1}_{p-3}|p-3,(2)\rangle \\ \quad = d^{(1)}_{p-4}|p-4,(1)\rangle + d^{(2)}_{p-4}|p-4,(2)\rangle + d^{(3)}_{p-4}|p-4,(3)\rangle. \end{cases} \tag{6.1}$$

Furthermore, Eq. (I.2.8) acting upon $|p-1\rangle$ gives a first set of relations among the $a^{(i)}_{p-4}$'s, $b^{(i)}_{p-4}$'s, $c^{(i)}_{p-4}$'s, and $d^{(i)}_{p-4}$'s if we take into account (4.1) and (5.1), i.e.,

$$b^{(1)}_{p-2}a^{(i)}_{p-4} + b^{(2)}_{p-2}b^{(i)}_{p-4} = a^{(1)}_{p-3}c^{(i)}_{p-4} + a^{(2)}_{p-3}d^{(i)}_{p-4}, \quad (i = 1,2,3). \tag{6.2}$$

Let us recall that so far we have no explicit knowledge of the coefficients $a^{(1)}_{p-2}$ and $a^{(2)}_{p-3}$.

A second set of relations can be formed as follows. If we let Eq. (I.2.7) act first upon $|p-2,(1)\rangle$ and thereafter on $|p-2,(2)\rangle$, then multiply the first equation with $a^{(1)}_{p-2}$ and the second with $a^{(2)}_{p-2}$, and if we finally take the sum of both, we can already take advantage of (6.2) to find, with the help of (5.2), that

$$(p-3)(p-4)[\lambda^{(1)}_{p-2}a^{(1)}_{p-2}a^{(i)}_{p-4} + \lambda^{(2)}_{p-2}a^{(2)}_{p-2}b^{(i)}_{p-4}] - (p-1)(p-2)\lambda^{(i)}_{p-4}[a^{(1)}_{p-2}a^{(i)}_{p-4} + a^{(2)}_{p-2}b^{(i)}_{p-4}] - 12\sqrt{6}(2p-5)(p+1)p(2p+1)[q(p-q)/(2p-1)]^{1/2}[b^{(1)}_{p-2}a^{(i)}_{p-4} + b^{(2)}_{p-2}b^{(i)}_{p-4}] = 0 \quad (i = 1,2,3). \tag{6.3}$$

Note that these relations involve only $a^{(i)}_{p-4}$ and $b^{(i)}_{p-4}$ ($i = 1,2,3$). Consequently, if we can manage to find a second set of relations in these unknowns, the eigenvalue problem will be reduced to that of solving a determinantal equation. The construction of such a set is far from trivial and goes as follows.

We let Eq. (I.2.5) act upon both $|p-3,(1)\rangle$ and $|p-3,(2)\rangle$,

$$\begin{cases} [p(2p-5)\lambda^{(1)}_{p-3}c^{(i)}_{p-4} - (p-2)(2p-3)\lambda^{(i)}_{p-4}c^{(i)}_{p-4}]|p-4,(i)\rangle - \dfrac{6(p-3)}{(p-2)^2}O^{-2}_{p-2}O^{+1}_{p-3}|p-3,(1)\rangle = 0, \\ [p(2p-5)\lambda^{(2)}_{p-3}d^{(i)}_{p-4} - (p-2)(2p-3)\lambda^{(i)}_{p-4}d^{(i)}_{p-4}]|p-4,(i)\rangle - \dfrac{6(p-3)}{(p-2)^2}O^{-2}_{p-2}O^{+1}_{p-3}|p-3,(2)\rangle = 0. \end{cases} \tag{6.4}$$

In order to analyze the action of the remaining shift operator product upon $|p-3,(1)\rangle$ and $|p-3,(2)\rangle$ the latter states are, on account of (5.1), rewritten as

$$\begin{cases} |\,p-3,(1)\rangle = \dfrac{c^{(2)}_{p-3}O^{-1}_{p-2}|\,p-2,(1)\rangle - b^{(2)}_{p-3}O^{-1}_{p-2}|\,p-2,(2)\rangle}{b^{(1)}_{p-2}c^{(2)}_{p-2} - b^{(2)}_{p-2}c^{(1)}_{p-2}}, \\ |\,p-3,(2)\rangle = \dfrac{-c^{(1)}_{p-3}O^{-1}_{p-2}|\,p-2,(1)\rangle + b^{(1)}_{p-3}O^{-1}_{p-2}|\,p-2,(2)\rangle}{b^{(1)}_{p-2}c^{(2)}_{p-2} - b^{(2)}_{p-2}c^{(1)}_{p-2}}. \end{cases} \tag{6.5}$$

After the substitution of (6.5) into (6.4), it is seen that we need to know how the triple product $O^{-2}_{p-2}O^{+1}_{p-3}O^{-1}_{p-2}$ affects the states $|\,p-2,(1)\rangle$ and $|\,p-2,(2)\rangle$. To this aim we split from this product the part $O^{+1}_{p-3}O^{-1}_{p-2}$, whose action upon a $|\,p-2,(i)\rangle$ state $(i=1,2)$ can be immediately expressed in terms of the action upon the same state of the product operator $O^{-1}_{p-1}O^{+1}_{p-2}$. Indeed, Eq. (I.2.3), when applied to $|\,p-2,(i)\rangle$ $(i=1,2)$, gives

$$\begin{aligned} 3\frac{(p-1)(p+1)}{(p-2)^2}O^{+1}_{p-3}O^{-1}_{p-2}|\,p-2,(i)\rangle &= 54(p-1)^2(p-2)^2(2p-3)(4I_2+1)|\,p-2,(i)\rangle \\ -(2p-3)(\lambda^{(i)}_{p-2})^2|\,p-2,(i)\rangle &+ \frac{3(p-2)(p-4)}{(p-1)^2}O^{-1}_{p-1}O^{+1}_{p-2}|\,p-2,(i)\rangle, \quad (i=1,2). \end{aligned} \tag{6.6}$$

As a next step we notice that the action of the shift operator O^{+1}_{p-2} upon the states $|\,p-2,(i)\rangle (i=1,2)$ has already been determined in (5.7), whereas the afterward action of $O^{-1}_{p-1}\rangle$ on $|\,p-1\rangle$ is directly contained in (4.1). Finally, O^{-2}_{p-2} $|\,p-2,(1)\rangle$ and $O^{-2}_{p-2}|\,p-2,(2)\rangle$ are immediately read off from (6.1). Resuming, the substitution of (6.5) and (6.6) into (6.4) and the consecutive use of (5.7) and (4.1) lead to the following relations which no longer contain shift operators:

$$\begin{aligned} &[\,p(2p-5)\lambda^{(1)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,]c^{(i)}_{p-4} - 6\frac{(p-3)}{(p-1)(p+1)}\frac{1}{b^{(1)}_{p-3}c^{(2)}_{p-3} - b^{(2)}_{p-3}c^{(1)}_{p-3}} \\ &\times\left[c^{(2)}_{p-3}\left(Aa^{(i)}_{p-4} + B\frac{b^{(2)}_{p-2}}{a^{(1)}_{p-2}}b^{(i)}_{p-4}\right) - b^{(2)}_{p-3}\left(B'\frac{b^{(1)}_{p-2}}{a^{(2)}_{p-2}}a^{(i)}_{p-4} + A'b^{(i)}_{p-4}\right)\right] = 0, \\ &[\,p(2p-5)\lambda^{(2)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,]d^{(i)}_{p-4} + 6\frac{(p-3)}{(p-1)(p+1)}\frac{1}{b^{(1)}_{p-3}c^{(2)}_{p-3} - b^{(2)}_{p-3}c^{(1)}_{p-3}} \\ &\times\left[c^{(1)}_{p-3}\left(Aa^{(i)}_{p-4} + B\frac{b^{(2)}_{p-2}}{a^{(1)}_{p-2}}b^{(i)}_{p-4}\right) - b^{(1)}_{p-3}\left(B'\frac{b^{(1)}_{p-2}}{a^{(2)}_{p-2}}a^{(i)}_{p-4} + A'b^{(i)}_{p-4}\right)\right] = 0 \quad (i=1,2,3). \end{aligned} \tag{6.7}$$

Herein A and B are functions of p and q alone, which after straightforward calculation can be brought into the form

$$\begin{cases} A = 6(p+1)(p-1)\{[2(p-2q)/\sqrt{\Gamma}\,][\,p^6 - 8p^5 + 18p^3 - 16p^2 + 14p - 6 - p^2(p^2+4p-2)(p-2q)^2] \\ \quad + p(p-4)(4p^3 - 7p^2 + 4p - 3) - (4p^3 + p^2 + 4p - 3)(p-2q)^2\}, \\ B = -24p^2(p+1)(p-2)(p-4)(2p+1)[q(p-q)/(2p-1)]^{1/2} \\ \quad \times\{(1/\sqrt{\Gamma}\,)[(p-1)(p-2)(2p-1) - p(p-2q)^2] - (p-2q)\}, \end{cases} \tag{6.8}$$

whereas similar expressions for A' and B' are obtained by the replacement of $\sqrt{\Gamma}$ by $-\sqrt{\Gamma}$ in, respectively, the expressions for A and B. In order to achieve the ultimate goal, all that remains to be done is the simultaneous elimination of $c^{(i)}_{p-4}$ and $d^{(i)}_{p-4}$ $(i=1,2,3)$ from Eqs. (6.7) and (6.2). By this, we end up with the following relations:

$$\begin{aligned} &[p(2p-5)\lambda^{(1)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,][p(2p-5)\lambda^{(2)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,][b^{(1)}_{p-2}a^{(i)}_{p-4} + b^{(2)}_{p-2}b^{(i)}_{p-4}\,] \\ &- \frac{6(p-3)[p(2p-5)\lambda^{(2)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,]}{(p^2-1)[b^{(1)}_{p-3}c^{(2)}_{p-3} - b^{(2)}_{p-3}c^{(1)}_{p-3}\,]} \\ &\cdot\left[a^{(1)}_{p-3}c^{(2)}_{p-3}\left(Aa^{(i)}_{p-4} + B\frac{b^{(2)}_{p-2}}{a^{(1)}_{p-2}}b^{(i)}_{p-4}\right) - a^{(1)}_{p-3}b^{(2)}_{p-3}\left(B'\frac{b^{(1)}_{p-2}}{a^{(2)}_{p-2}}a^{(i)}_{p-4} + A'b^{(i)}_{p-4}\right)\right] \\ &+ \frac{6(p-3)[p(2p-5)\lambda^{(1)}_{p-3} - (p-2)(2p-3)\lambda^{(i)}_{p-4}\,]}{(p^2-1)[b^{(1)}_{p-3}c^{(2)}_{p-3} - b^{(2)}_{p-3}c^{(1)}_{p-3}\,]} \\ &\cdot\left[a^{(2)}_{p-3}c^{(1)}_{p-3}\left(Aa^{(i)}_{p-4} + B\frac{b^{(2)}_{p-2}}{a^{(1)}_{p-2}}b^{(i)}_{p-4}\right) - a^{(2)}_{p-3}b^{(1)}_{p-3}\left(B'\frac{b^{(1)}_{p-2}}{a^{(2)}_{p-2}}a^{(i)}_{p-4} + A'b^{(i)}_{p-4}\right)\right] = 0 \quad (i=1,2,3). \end{aligned} \tag{6.9}$$

Together with (6.3) these relations form a system of six linear homogeneous equations in the six unknowns $a^{(i)}_{p-4}$, $b^{(i)}_{p-4}$ ($i = 1,2,3$). The condition that there exists a solution which differs from the trivial zero solution is a polynomial equation of third degree which has to be satisfied by each of the three eigenvalues $\lambda^{(1)}_{p-4}$, $\lambda^{(2)}_{p-4}$, and $\lambda^{(3)}_{p-4}$. At first sight it may be expected that this polynomial equation would also contain the coefficients $a^{(i)}_{p-2}$, $b^{(i)}_{p-2}$, $a^{(i)}_{p-3}$, $b^{(i)}_{p-3}$ and $c^{(i)}_{p-3}$ with $i = 1$ or $i = 2$, and this would necessitate the explicit determination of closed expressions for these coefficients. However, by doing the actual calculation, it turns out that they can be simply rearranged in combinations which on account of (4.11), (5.2), (5.3), and (5.4) allow their complete cancellation. If, finally, we fill in the expressions (4.9) and (5.11) for the eigenvalues $\lambda^{(i)}_{p-2}$ and $\lambda^{(i)}_{p-3}$ ($i = 1,2$), we arrive after very lengthy calculations at the following polynomial equation:

$$\begin{aligned}\lambda^3 &- 3\sqrt{6}(p+1)(2p-1)(p-2q)\lambda^2 \\ &+ 18[(4p^4+4p^3-51p^2+94p-63)(p-2q)^2 \\ &- 48(4p^3-13p^2+6p+9)]\lambda \\ &- 6\sqrt{6}[(4p^5+16p^4-107p^3-57p^2+459p-243) \\ &\times(p-2q)^2+144(-4p^4+13p^3+3p^2-27p+27)] \\ &\times(2p-5)(p-2q)=0. \qquad (6.10)\end{aligned}$$

It is, however, in general not possible to write down closed expressions such as (2.2), (3.2), (4.9), and (5.11) for the solutions $\lambda^{(1)}_{p-4}$, $\lambda^{(2)}_{p-4}$, $\lambda^{(3)}_{p-4}$ of Eq. (6.10). Consequently, for each given pair of values of (p,q), Eq. (6.10) must be solved numerically by means of some well-established method. There are, nevertheless, a few properties of Eq. (6.10) and its roots which can be traced without solving that equation explicitly.

So, for instance, it is clear that the sum of the three eigenvalues is $\sqrt{6}$ times a positive integer, whereas their product is positive if p is large enough compared to $2q$, and negative otherwise. These facts already indicate that for q values close to $[p/2]$ two eigenvalues are positive and one negative. On the other hand, if q is much smaller than $[p/2]$ either three positive roots or one positive root and two negative ones are expected. But, keeping q fixed in Eq. (6.10), each of its three roots tends towards $2p^3\sqrt{6}$ if p becomes very large. Hence, if $2q$ is small compared to p, three positive eigenvalues are most probable. All these properties concerning the sign of the eigenvalues can be checked on solving Eq. (6.10) numerically for various (p,q) representations, or by inspection of the extensive tables established by Judd *et al.*,[7] listing for $p \leqslant 12$ and $p - 2q \geqslant 0$ all the eigenvalues of a scalar operator K which is related to O^0_l by $K = O^0_l/2\sqrt{6}$.

Finally, it should be noticed that Eq. (6.10) admits a complete analytical solution if $p = 2q$, in which case the three eigenvalues read

$$\begin{cases}\lambda^{(1)}_{p-4} = 0, \\ \lambda^{(2)}_{p-4} = 12\sqrt{6}[4p^3-13p^2+6p+9]^{1/2}, \\ \lambda^{(3)}_{p-4} = -12\sqrt{6}[4p^3-13p^2+6p+9]^{1/2}.\end{cases} \qquad (6.11)$$

Since the particular representation (10,5) has been studied in detail by Partensky and Quesne,[6] their paper, which also contains the eigenvalues of an operator proportional to O^0_l, provides us with an immediate check of the validity of the expressions (6.11).

7. CONCLUSIONS

We have shown that with the recently derived relations among nonscalar shift operator products the shift operator calculus is raised into a powerful technique for the calculation of eigenvalues of the scalar shift operator O^0_l Indeed, where in the region of highest possible angular momentum values a case of double l-degeneracy became almost intractable in the usual shift operator approach, we have been able in the present paper to give even a complete analytical treatment of a situation whereby l is threefold degenerate.

ACKNOWLEDGMENTS

Two of us (H. De Meyer and G. Vanden Berghe) would like to express their gratitude to the British Council and the National Science Foundation of Belgium for their financial support, which made the present cooperation possible. Also they would like to thank all Staff members of the Dept. of Applied Math. at Queen Mary College (London) for their kind hospitality, which made a stay at Q. M. College a memorable one.

[1] H. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).
[2] J. W. B. Hughes, J. Phys. A **6**, 48 (1973).
[3] J. W. B. Hughes, J. Phys. A **6**, 281 (1973).
[4] G. Vanden Berghe and H. De Meyer, J. Math. Phys. **21**, 2632 (1980).
[5] M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, Ann. Phys. (NY) **95**, 139 (1975) and references quoted therein.
[6] A. Partensky and C. Quesne, J. Math. Phys. **20**, 2014 (1979).
[7] B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).

Lorentz and SU(2) ⊗ U(N) analysis of the Lie algebra sp(4N;r) [gl(4N;r)] for any integer $N \geqslant 1$ [a)]

Mitchell D. Adams [b),c)]
Department of Physics and Astronomy, Drake University, Des Moines, Iowa 50311

(Received 24 February 1981; accepted for publication 17 June 1981)

The generators of the fundamental representation of the Lie group GL(4N;r), integral $N \geqslant 1$, are constructed from Kronecker products of smaller matrices in such a way that their tensor character under the action of the (unique) full null-plane Lorentz subgroup is apparent. Commutation relations of these tensors are given in terms of symmetric and antisymmetric structure constants for the fundamental representation of U(N) used in their construction. Generators of the Sp(4N;r) subgroup are classified according to transformation character under a U(N) subgroup. Commutation relations of sp(4N;r) are given in terms of $SU(2)_{spin} \otimes U(N)$ multiplets.

PACS numbers: 02.20.Sv

I. INTRODUCTION

Since the advent of the dual resonance model,[1] and in particular the model known as the Nambu string,[2] some attention[3,4] has been focused upon the formal algebraic structure of physical systems composed of large numbers of boson operators. As is well known,[5] the Lie algebra of the maximal set of bilinears constructed from n bosons is the complex form of the real algebra sp(2n;r).

Now, for the applications of current interest, it is desirable to arrange the elements of sp(2n;r) into combinations which transform as tensors under a Lorentz subgroup. In the case of an orthogonal group, such an arrangement is simple; due, however, to the higher level of complexity of the root diagrams of large symplectic algebras, no such simple technique exists.

Recently, Staunton[6] has reported a Lorentz subgroup analysis of sp(8;r) based upon manual examination of a computer-generated set of commutation relations of 8×8 matrices constructed from 4×4 blocks of real Dirac matrices. The algorithm for sp(2^n;r) reported in Ref. 6, while systematic, is quite lengthy. Further, although it can be applied, case by case, to larger symplectic algebras, due to the necessity for manual examination, it rapidly becomes unwieldy.

We report here a compact analysis, based upon the formal properties of Kronecker products, which serves to classify the elements of any sp(4N;r), $N \geqslant 1$, according to their Lorentz tensor character under a Lorentz subgroup of sp(4N;r). In particular, we have selected the unique, null-plane Lorentz subgroup,[7] of interest for physical applications.[8,4] In fact, however, the analysis serves to classify the larger algebra gl(4N;r), of which sp(4N;r) is a subalgebra. Furthermore, since our Kronecker product construction singles out elements of the subalgebra u(N), the full set of commutation relations among the Lorentz tensors of gl(4N;r) are detailed in a form which explicitly involves the structure constants of U(N). Therefore, the results also facilitate a U(N) [in fact, an $SU(2)_{spin} \otimes U(N)$] analysis of the elements of the algebra, already classified according to their Lorentz character. These dual identifications may be of some future interest for the construction of relativistic, boson constituent models of composite particles exhibiting color or flavor symmetries.

In Sec. II, we present our Kronecker product construction, and exhibit the Lie algebra. In Sec. III, we present the Lorentz tensor classification of the elements of gl(4N;r), for any integer $N \geqslant 1$. The SU(2) ⊗ U(N) nature of the elements of sp(4N;r) are recorded in Sec. IV. Finally, as a particular case, the boson realization of sp(8;r) of Ref. 6 is classified in Sec. V.

II. AN ALGORITHM BASED UPON KRONECKER PRODUCTS

Consider the following set of four *real* Pauli-type matrices:

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.1}$$

The set $\{\sigma_i\}$, $i = 0, 1, 2, 3$ are then a realization of the algebra[5] gl(2;r). Let $\eta_{ij} = \text{diag}(1, 1, -1, 1)$ denote a metric. Then[9]

$$\sigma_i \sigma_j = a_{0ij}{}^{k} \sigma_k + \epsilon_{0ij}{}^{k} \sigma_k, \tag{2.2}$$

where

$$a_{0ij}{}^{k} = \eta_{ij}\eta_0^k + \eta_{i0}\eta_j^k + \eta_{j0}\eta_i^k + 2\eta_{i0}\eta_{j0}\eta_0^k. \tag{2.3}$$

Further, let the set of N^2 Hermitian $N \times N$ matrices $\{\lambda_I\}$, $I = 0, 1, \ldots, N^2 - 1$, generate[5] the *fundamental* representation of U(N), with $\lambda_0 = \frac{1}{2}I$, and with symmetric and antisymmetric structure constants given by

$$\{\lambda_I, \lambda_J\} = d_{IJK}\lambda_K, \tag{2.4a}$$

and

$$[\lambda_I, \lambda_J] = if_{IJK}\lambda_K. \tag{2.4b}$$

Consider next the set of N^2 real $N \times N$ matrices $\{G_I\}$,

[a)]Research partially supported by the Research Corporation under Grant C1232.
[b)]Partially supported by a grant from the Drake University Research Council.
[c)]Current address is Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106.

 0022-2488/81/112372-04$01.00

derived from the set $\{\lambda_I\}$ via $G_I = \lambda_I$, for symmetric λ_I, and $G_I = -i\lambda_I$, for antisymmetric λ_I. It follows[5] that the $\{G_I\}$ are a realization of the real Lie algebra gl(N;r), with symmetric and antisymmetric structure constants d'_{IJK} and f'_{IJK} differing from those of (2.4) by appropriate changes of sign.[10]

Finally, we define a set of $16N^2$ real matrices $\{M_{iIj}\}$ via the Kronecker product

$$M_{iIj} = \sigma_i \otimes G_I \otimes \sigma_j. \tag{2.5}$$

The real matrices M are linearly independent and therefore constitute[5] a realization of gl($4N$;r). Note that for the case $N = 1$, one realizes a familiar set of sixteen real 4×4 Dirac matrices.

Now the matrix product of any two of the matrices (2.5) is given by the multiplication rule for Kronecker products[11]:

$$\begin{aligned} M_{iIj}M_{kJl} &= (\sigma_i \otimes G_I \otimes \sigma_j)(\sigma_k \otimes G_J \otimes \sigma_l) \\ &= (\sigma_i\sigma_k) \otimes (G_IG_J) \otimes (\sigma_j\sigma_l), \end{aligned} \tag{2.6a}$$

and, of course

$$M_{kJl}M_{iIj} = (\sigma_k\sigma_i) \otimes (G_JG_I) \otimes (\sigma_l\sigma_j). \tag{2.6b}$$

Recall, however, that in all cases, $\sigma_k\sigma_i = \pm\sigma_i\sigma_k$, the positive sign obtaining whenever $k = i$ or at least one index is zero. It follows then, upon comparison of (2.6b) with (2.6a), that the commutation relations among the $\{M\}$ can be expressed in terms of the U(N) structure constants d_{IJK} and f_{IJK}, the symmetric constants being involved whenever the accumulated sign obtained upon interchanges of the σ matrices in (2.6b) is negative. Substitution of (2.2) into (2.6) yields, after straightforward algebra, the result[10]

$$\begin{aligned} [M_{iIj}, M_{kJl}] = {} & (a_{0ik}{}^m\epsilon_{0jl}{}^n + \epsilon_{0ik}{}^m a_{0jl}{}^n)d'_{IJK}M_{mKn} \\ & + (a_{0ik}{}^m a_{0jl}{}^n + \epsilon_{0ik}{}^m\epsilon_{0jl}{}^n)f'_{IJK}M_{mKn}. \end{aligned} \tag{2.7}$$

Now the Lie algebra of GL($4N$;r) is well known[5] and can be recorded directly, compactly, and systematically. Our particular, somewhat cumbersome result (2.7), however, serves to simplify to a great extent the classification of the elements of gl($4N$;r) according to their tensor character under particular subgroups of interest, as we shall presently demonstrate.

TABLE I. Generators of GL($4N$;r) classified according to their tensor character under the Lorentz group generated by $T^0_{\mu\nu}$. Indices S number symmetric matrices G_I, while A number those of the antisymmetric partition.

Symplectic	Nonsymplectic
$T^S_{23} = M_{2S1}$	$T^A_{23} = M_{2A1}$
$T^S_{31} = M_{2S3}$	$T^A_{31} = M_{2A3}$
$T^S_{12} = M_{0S2}$	$T^A_{12} = M_{0A2}$
$T^S_{10} = -M_{1S3}$	$T^A_{10} = -M_{1A3}$
$T^S_{20} = M_{1S1}$	$T^A_{20} = M_{1A1}$
$T^S_{30} = M_{3S0}$	$T^A_{30} = M_{3A0}$
$V^S_0 = -M_{2S0}$	$V^A_0 = -M_{2A0}$
$V^S_1 = -M_{3S3}$	$V^A_1 = -M_{3A3}$
$V^S_2 = M_{3S1}$	$V^A_2 = M_{3A1}$
$V^S_3 = -M_{1S0}$	$V^A_3 = -M_{1A0}$
$U^A_0 = -M_{1A2}$	$U^S_0 = -M_{1S2}$
$U^A_1 = M_{0A1}$	$U^S_1 = M_{0S1}$
$U^A_2 = M_{0A3}$	$U^S_2 = M_{0S3}$
$U^A_3 = -M_{2A2}$	$U^S_3 = -M_{2S2}$
$C^A = M_{0A0}$	$C^S = M_{0S0}$
$N^A = -M_{3A2}$	$N^S = -M_{3S2}$

III. LORENTZ TENSOR CLASSIFICATION

Recall that, for the case $N = 1$, our Kronecker product construction (2.5) yields a set of sixteen 4×4 real Dirac matrices. Such a set includes, of course, an antisymmetric tensor, $S_{\mu\nu}$, which generates the Lorentz group. It follows that a

TABLE II. Lie algebra gl($4N$;r) expressed in terms of the Lorentz tensor nature of the elements. S, T, U number symmetric G_I only, A, B, C antisymmetric G_I only.

Symplectic

$$\begin{aligned}
&[T^S_{\mu\nu}, T^T_{\alpha\beta}] = d_{STU}(g_{\mu\alpha}T^U_{\nu\beta} - g_{\nu\alpha}T^U_{\mu\beta} + g_{\nu\beta}T^U_{\mu\alpha} - g_{\mu\beta}T^U_{\nu\alpha}) \\
&\qquad + f_{STA}[(g_{\mu\alpha}g_{\nu\beta} - g_{\nu\alpha}g_{\mu\beta})C^A + \epsilon_{\mu\nu\alpha\beta}N^A] \\
&[T^S_{\mu\nu}, V^T_\alpha] = f_{STA}\epsilon_{\mu\nu\alpha}{}^\beta U^A_\beta + d_{STU}(g_{\mu\alpha}V^U_\nu - g_{\nu\alpha}V^U_\mu) \\
&[T^S_{\mu\nu}, U^A_\alpha] = f_{SAT}\epsilon_{\mu\nu\alpha}{}^\beta V^T_\beta + d_{SAB}(g_{\mu\alpha}U^B_\nu - g_{\nu\alpha}U^B_\mu) \\
&[T^S_{\mu\nu}, C^A] = f_{SAT}T^T_{\mu\nu} \qquad [T^S_{\mu\nu}, N^A] = f_{SAT}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^T_{\alpha\beta} \\
&[V^S_\mu, V^T_\nu] = f_{STA}g_{\mu\nu}C^A + d_{STU}T^U_{\mu\nu} \\
&[V^S_\mu, U^A_\nu] = -f_{SAT}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^T_{\alpha\beta} + d_{SAB}g_{\mu\nu}N^B \\
&[U^A_\mu, U^B_\nu] = -f_{ABC}g_{\mu\nu}C^C - d_{ABS}T^S_{\mu\nu} \\
&[C^A, C^B] = f_{ABC}C^C \qquad [C^A, N^B] = f_{ABC}N^C \\
&[N^A, N^B] = -f_{ABC}C^C \\
&[C^A, V^S_\mu] = f_{AST}V^T_\mu \qquad [C^A, U^B_\mu] = f_{ABC}U^C_\mu \\
&[N^A, V^S_\mu] = d_{ASB}U^B_\mu \qquad [N^A, U^B_\mu] = d_{ABS}V^S_\mu
\end{aligned}$$

Mixed

$$\begin{aligned}
&[T^S_{\mu\nu}, T^A_{\alpha\beta}] = d_{SAB}(g_{\mu\alpha}T^B_{\nu\beta} - g_{\nu\alpha}T^B_{\mu\beta} + g_{\nu\beta}T^B_{\mu\alpha} - g_{\mu\beta}T^B_{\nu\alpha}) \\
&\qquad - f_{SAT}[(g_{\mu\alpha}g_{\nu\beta} - g_{\nu\alpha}g_{\mu\beta})C^T + \epsilon_{\mu\nu\alpha\beta}N^T] \\
&[T^S_{\mu\nu}, V^A_\alpha] = -f_{SAT}\epsilon_{\mu\nu\alpha}{}^\beta U^T_\beta + d_{SAB}(g_{\mu\alpha}V^B_\nu - g_{\nu\alpha}V^B_\mu) \\
&[T^A_{\mu\nu}, V^S_\alpha] = -f_{AST}\epsilon_{\mu\nu\alpha}{}^\beta U^T_\beta + d_{ASB}(g_{\mu\alpha}V^B_\nu - g_{\nu\alpha}V^B_\mu) \\
&[T^S_{\mu\nu}, U^T_\alpha] = -f_{STA}\epsilon_{\mu\nu\alpha}{}^\beta V^A_\beta + d_{STU}(g_{\mu\alpha}U^U_\nu - g_{\nu\alpha}U^U_\mu) \\
&[T^A_{\mu\nu}, U^B_\alpha] = f_{ABC}\epsilon_{\mu\nu\alpha}{}^\beta V^C_\beta - d_{ABS}(g_{\mu\alpha}U^S_\nu - g_{\nu\alpha}U^S_\mu) \\
&[T^S_{\mu\nu}, C^T] = -f_{STA}T^A_{\mu\nu} \qquad [T^S_{\mu\nu}, N^T] = -f_{STA}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^A_{\alpha\beta} \\
&[T^A_{\mu\nu}, C^B] = f_{ABC}T^C_{\mu\nu} \qquad [T^A_{\mu\nu}, N^B] = f_{ABC}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^C_{\alpha\beta} \\
&[V^S_\mu, V^A_\nu] = -f_{SAT}g_{\mu\nu}C^T + d_{SAB}T^B_{\mu\nu} \\
&[V^S_\mu, U^T_\nu] = f_{STA}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^A_{\alpha\beta} + d_{STU}g_{\mu\nu}N^U \\
&[V^A_\mu, U^B_\nu] = -f_{ABC}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^C_{\alpha\beta} - d_{ABS}g_{\mu\nu}N^S \\
&[U^A_\mu, U^S_\nu] = -f_{AST}g_{\mu\nu}C^T + d_{ASB}T^B_{\mu\nu} \\
&[C^A, C^S] = f_{AST}C^T \qquad [C^A, N^S] = f_{AST}N^T \\
&[C^S, N^A] = f_{SAT}N^T \qquad [N^A, N^S] = -f_{AST}C^T \\
&[C^A, V^B_\mu] = f_{ABC}V^C_\mu \qquad [C^A, U^S_\mu] = f_{AST}U^T_\mu \\
&[C^S, V^T_\mu] = -f_{STA}V^A_\mu \qquad [C^S, U^A_\mu] = f_{SAT}U^T_\mu \\
&[N^A, V^B_\mu] = -d_{ABS}U^S_\mu \qquad [N^A, U^S_\mu] = -d_{ASB}V^B_\mu \\
&[N^S, V^T] = d_{STU}U^U_\mu \qquad [N^S, U^A_\mu] = -d_{SAB}V^B_\mu
\end{aligned}$$

Nonsymplectic

$$\begin{aligned}
&[T^A_{\mu\nu}, T^B_{\alpha\beta}] = -d_{ABS}(g_{\mu\alpha}T^S_{\nu\beta} - g_{\nu\alpha}T^S_{\mu\beta} + g_{\nu\beta}T^S_{\mu\alpha} - g_{\mu\beta}T^S_{\nu\alpha}) \\
&\qquad - f_{ABC}[(g_{\mu\alpha}g_{\nu\beta} - g_{\nu\alpha}g_{\mu\beta})C^C + \epsilon_{\mu\nu\alpha\beta}N^C] \\
&[T^A_{\mu\nu}, V^B_\alpha] = -f_{ABC}\epsilon_{\mu\nu\alpha}{}^\beta U^C_\beta - d_{ABS}(g_{\mu\alpha}V^S_\nu - g_{\nu\alpha}V^S_\mu) \\
&[T^A_{\mu\nu}, U^S_\alpha] = f_{AST}\epsilon_{\mu\nu\alpha}{}^\beta V^T_\beta + d_{ASB}(g_{\mu\alpha}U^B_\nu - g_{\nu\alpha}U^B_\mu) \\
&[T^A_{\mu\nu}, C^S] = f_{AST}T^T_{\mu\nu} \qquad [T^A_{\mu\nu}, N^S] = f_{AST}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^T_{\alpha\beta} \\
&[V^A_\mu, V^B_\nu] = -f_{ABC}g_{\mu\nu}C^C - d_{ABS}T^S_{\mu\nu} \\
&[V^A_\mu, U^S_\nu] = -f_{AST}\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}T^T_{\alpha\beta} + d_{ASB}g_{\mu\nu}N^B \\
&[U^S_\mu, U^T_\nu] = f_{STA}g_{\mu\nu}C^A + d_{STU}T^U_{\mu\nu} \\
&[C^S, C^T] = -f_{STA}C^A \qquad [C^S, N^T] = -f_{STA}N^A \\
&[N^S, N^T] = f_{STA}C^A \\
&[C^S, V^A_\mu] = f_{SAT}V^T_\mu \qquad [C^S, U^T_\mu] = -f_{STA}U^A_\mu \\
&[N^S, V^A_\mu] = d_{SAB}U^B_\mu \qquad [N^S, U^T_\mu] = -d_{STU}V^U_\mu
\end{aligned}$$

TABLE III. Elements of sp(4N;r) classified according to SU(2) ⊗ U(N). Superscripts denote tensor character under the SU(2) generated by S^1, S^2, and S^3. Subscripts denote transformation properties under the U(N) generated by $\{H, F_I^0\}$. Note, however, that $\Delta_{\pm A}$ is a single U(N) multiplet, as are the $\Omega^i{}_{\pm S}$.

$H = V_0^0 = -M_{200}$	$F_A^0 = C^A = M_{0A0}$	$F_S^0 = V_0^S = -M_{2S0}$,	$S \neq 0$
$S^1 = -T_{23}^0 = -M_{201}$	$F_A^1 = U_1^A = M_{0A1}$	$F_S^1 = -T_{23}^S = -M_{2S1}$,	$S \neq 0$
$S^2 = -T_{31}^0 = -M_{203}$	$F_A^2 = U_2^A = M_{0A3}$	$F_S^2 = -T_{31}^S = -M_{2S3}$,	$S \neq 0$
$S^3 = -T_{12}^0 = -M_{002}$	$F_A^3 = U_3^A = -M_{2A2}$	$F_S^3 = -T_{12}^S = -M_{0S2}$,	$S \neq 0$
	$\Delta_{+A} = U_0^A = -M_{1A2}$	$\Delta_{-A} = N^A = -M_{3A2}$	
	$\Omega^1{}_{+S} = T_{10}^S = -M_{1S3}$	$\Omega^1{}_{-S} = V_1^S = -M_{3S3}$	
	$\Omega^2{}_{+S} = T_{20}^S = M_{1S1}$	$\Omega^2{}_{-S} = V_2^S = M_{3S1}$	
	$\Omega^3{}_{+S} = T_{30}^S = M_{3S0}$	$\Omega^3{}_{-S} = V_3^S = -M_{1S0}$	

Lorentz subgroup of GL(4N;r) can be immediately identified, by one-to-one correspondence, from the set of sixteen $4N \times 4N$ Kronecker products of the form $\sigma_i \otimes G_0 \otimes \sigma_j$, since for any N, $G_0 = \frac{1}{2}I$. In particular, we choose to consider the null-plane[7] set

$$\begin{aligned} S_{23} &= M_{201}, \quad S_{10} = -M_{103}, \\ S_{31} &= M_{203}, \quad S_{20} = M_{101}, \\ S_{12} &= M_{002}, \quad S_{30} = M_{300}. \end{aligned} \tag{3.1}$$

The simplicity of our approach lies in the elemental observation that since, for all I,

$$\{G_0, G_I\} = G_I, \quad [G_0, G_I] = 0, \tag{3.2}$$

commutation of the $S_{\mu\nu}$ of (3.1) with any of the M_{iIj} will not alter the index I. In particular, we obtain, from (2.7),

$$[M_{i0j}, M_{kIl}] = (a_{0ik}{}^m \epsilon_{0jl}{}^n + \epsilon_{0ik}{}^m a_{0jl}{}^n) M_{mIn}. \tag{3.3}$$

Further, since in the set of 16 4×4 Dirac matrices there is found (neglecting parity) one antisymmetric tensor of second rank, two vectors, and two scalars, it follows that exactly the same number and types of Lorentz tensor quantities will be found in the $N > 1$ cases for *each* value of the index I. In particular, the tensor character of each M_{iIj} is exactly that of the 4×4 real Dirac matrix with the same indices i and j.

Let, for each value of the index I, antisymmetric Lorentz tensors of second rank be denoted $T_{\mu\nu}^I$, and, by analogy only since we are neglecting parity, vectors of polar (axial) type V_μ^I (U_μ^I), and scalars (pseudoscalars) C^I (N^I) respectively. Further, for later convenience, let the set of indices I be partitioned into two sets numbering the symmetric and antisymmetric elements of $\{G_I\}$, and let the indices S, T, U number those of the symmetric partition, and A, B, C those of the antisymmetric partition. Finally, let the null-plane Lorentz generators of (3.1) be denoted $S_{\mu\nu} \equiv T_{\mu\nu}^0$. Then the elements of gl(4N;r) are classified according to their Lorentz tensor properties in Table I. The commutation relations among these generators, again in terms of their Lorentz tensor character, are exhibited in Table II.

IV. sp(4N;r) CLASSIFICATION UNDER SU(2) ⊗ U(N)

From among the set $\{M\}$ of $16N^2$ real $4N \times 4N$ matrices we may identify a subset of $2N(4N+1)$ matrices $\hat{\Gamma}$ which have the $2N \times 2N$ block form

$$\hat{\Gamma} = \begin{bmatrix} Z & S \\ \hat{S} & -Z^T \end{bmatrix}, \tag{4.1}$$

TABLE IV. Lie algebra sp(4N;r) transcribed to reflect the SU(2) ⊗ U(N) classification of the elements ($i,j,k = 1,2,3$ only).

$[H, S^i] = 0$	$[H, F_I^0] = 0$	$[H, F_I^i] = 0$
$[S^i, S^j] = \epsilon_{ijk} S^k$	$[S^i, F_I^0] = 0$	$[S^i, F_I^j] = \epsilon_{ijk} F_I^k$
$[F_I^0, F_J^0] = f_{IJK} F_K^0$	$[F_I^0, F_J^i] = f_{IJK} F_K^i$	
$[F_I^i, F_J^j] = \delta_{ij} f_{IJK} F_K^0 + \epsilon_{ijk} d_{IJK} F_K^k$		
$[H, \Delta_{+A}] = \pm \Delta_{+A}$	$[S^i, \Delta_{+A}] = 0$	
$[F_A^0, \Delta_{+B}] = f_{ABC} \Delta_{+C}$	$[F_S^0, \Delta_{+A}] = \pm d_{SAB} \Delta_{\mp B}$	
$[F_A^i, \Delta_{+B}] = -d_{ABS} \Omega^i{}_{+S}$	$[F_S^i, \Delta_{+A}] = \pm f_{SAT} \Omega^i{}_{\mp T}$	
$[H, \Omega^i{}_{+S}] = \pm \Omega^i{}_{\mp S}$	$[S^i, \Omega^j{}_{+S}] = \epsilon_{ijk} \Omega^k{}_{+S}$	
$[F_A^0, \Omega^i{}_{+S}] = f_{AST} \Omega^i{}_{+T}$	$[F_S^0, \Omega^i{}_{+T}] = \pm d_{STU} \Omega^i{}_{\mp U}$	
$[F_A^i, \Omega^j{}_{+S}] = \delta_{ij} d_{ASB} \Delta_{\pm B} \mp \epsilon_{ijk} f_{AST} \Omega^k{}_{\mp T}$		
$[F_S^i, \Omega^j{}_{+T}] = \mp \delta_{ij} f_{STA} \Delta_{+A} + \epsilon_{ijk} d_{STU} \Omega^k{}_{\pm U}$		
$[\Delta_{+A}, \Delta_{+B}] = [\Delta_{-A}, \Delta_{-B}] = -f_{ABC} F_C^0$		
$[\Delta_{+A}, \Delta_{-B}] = -d_{ABS} F_S^0$		
$[\Omega^i{}_{+S}, \Delta_{+A}] = [\Omega^i{}_{-S}, \Delta_{-A}] = -d_{SAB} F_B^i$		
$[\Omega^i{}_{+S}, \Delta_{-A}] = f_{SAT} F_T^i$ $\quad [\Omega^i{}_{-S}, \Delta_{+A}] = -f_{SAT} F_T^i$		
$[\Omega^i{}_{+S}, \Omega^j{}_{+T}] = [\Omega^i{}_{-S}, \Omega^j{}_{-T}] = -\delta_{ij} f_{STA} F_A^0 - \epsilon_{ijk} d_{STU} F_U^k$		
$[\Omega^i{}_{+S}, \Omega^j{}_{-T}] = -\delta_{ij} d_{STU} F_U^0 + \epsilon_{ijk} f_{STA} F_A^k$		

where S and $\hat{S}$ are any two real *symmetric* matrices, Z is any $2N \times 2N$ matrix, and Z^T denotes the transpose of Z. The set $\{\hat{\Gamma}\}$ is a realization[5] of the Lie algebra sp(4N;r), and its elements are enumerated in Table I, in the column labeled symplectic. The Lorentz tensor classification of sp(4N;r) is therefore recorded in Table I, and its Lie algebra similarly classified among the entries of Table II.

Additionally, the matrices Γ have the important property

$$\hat{\Gamma} = 4M_{200}\hat{\Gamma}^T M_{200}, \tag{4.2}$$

which permits their use to construct boson realizations of sp(4N;r), for *any* integer N, as detailed below in Sec. V.

Included among the set $\{\hat{\Gamma}\}$ is the set of matrices $\{V_0^s, C^A\}$, which generate[12] a U(N) which commutes with the SU(2) generated by the space–space components of the Lorentz tensor $S_{\mu\nu} = T^0_{\mu\nu}$.

The symplectic matrices $\hat{\Gamma}$ are classified according to their transformation character under SU(2) ⊗ U(N) in Table III. The matrix V_0^0 is a scalar under both SU(2) and U(N) and is specially denoted H. Other entries in Table III are identified according to their scalar or vector character under SU(2) by their superscripts, with subscripts indicating their U(N) multiplet character. Commutation relations of the sp(4N;r) matrices according to their SU(2) ⊗ U(N) character are recorded in Table IV.

For sp(2n;r) with odd n, the construction and analysis above do not apply. However a construction using the subset of

$$M_{iI} \equiv \sigma_i \otimes G_I$$

That obey (4.2) with M_{20} instead of M_{200} yields a u(n) subalgebra $\{H, F_I\}$ and a multiplet $\Omega_{\pm s}$ with the same commutation relations as $\{H, {}^0F_I, \Omega^1_{\pm s}\}$.

V. BOSON REALIZATION AND SU(2)×U(2) CLASSIFICATION OF sp(8;r)

Consider the sets of n real variables q and associated canonical conjugates η, $\eta = -i\partial/\partial q$. The maximal set of bilinears constructed from these $2n$ boson operators form a realization of the (complex) Lie algebra of Sp(2n;r). For n even, the analysis of the preceding sections applies.

For $n = 2N$, form the column matrix

$$Q = \text{col}(q_1, q_2, ..., q_{2N}, \eta_1, \eta_2, ..., \eta_{2N}), \tag{5.1}$$

and denote the entries Q_a, $a = 1, 2, ..., 4N$. Then

$$[Q_a, Q_b] = -2iM_{200}, \tag{5.2}$$

where M_{200} is defined by (2.5).

Now Staunton[6] has used a result of Browne[13] to show that the self-adjoint[14] operators

$$\Gamma_{iIk} \equiv Q^T(M_{200}\hat{\Gamma}_{iIj})Q, \tag{5.3}$$

where the matrices $\hat{\Gamma}$ are those satisfying (4.2) and are listed in the symplectic column of Table I, satisfy for any N the Lie algebra of Table II (with, of course, a factor of the imaginary unit appearing on the right hand sides). The Lie algebra sp(8;r) i.e., $N = 2$, has been Lorentz classified in Ref. 6, and the special representational indentities obtaining for this boson realization have been detailed as well.

Our analysis permits these results to be extended to include a classification of these operators according to $\text{SU(2)}_{\text{spin}} \otimes \text{SU(2)}_{\text{isospin}}$, for example. Identification with the notation of Ref. 6. yields[15]

$$\begin{aligned} &V^0_\mu = A_\mu, \quad U^2_\mu = B_\mu, \quad V^3_\mu = C_\mu, \\ &V^1_\mu = D_\mu, \quad T^3_{\mu\nu} = E_{\mu\nu}, \quad T^1_{\mu\nu} = \hat{F}_{\mu\nu}, \\ &T^0_{\mu\nu} = S_{\mu\nu}, \quad N^2 = T_1, \quad C^2 = T_2, \end{aligned} \tag{5.4}$$

and the SU(2) ⊗ U(2) classification is that of Table III, with (i, j, k cyclic 1, 2, 3 only)

$$\begin{aligned} &H = A_0, \quad F^0_1 = D_0, \quad F^0_2 = T_2, \quad F^0_3 = C_0, \\ &S^i = -S_{jk}, \quad F^i_1 = -\hat{F}_{jk}, \quad F^i_2 = B_i, \quad F^i_3 = -E_{jk}, \\ &\Omega^i{}_{+0} = S_{i0}, \quad \Omega^i{}_{+1} = \hat{F}_{i0}, \Delta_{+2} = B_0, \quad \Omega^i{}_{+3} = E_{i0}, \\ &\Omega^i{}_{-0} = A_i, \quad \Omega^i{}_{-1} = D_i, \quad \Delta_{-2} = T_1, \quad \Omega^i{}_{-3} = C_i. \end{aligned} \tag{5.5}$$

The commutation relations follow easily upon substitution of the U(2) structure constants into Table IV.

ACKNOWLEDGMENTS

The author wishes to thank Dr. L. P. Staunton of Drake University for wisdom, encouragement, and valuable assistance with the manuscript. He would also like to thank Karen Fett for carefully typing the difficult manuscript.

[1]See, for example, P. H. Frampton, *Dual Resonance Models* (Benjamin, New York, 1974).
[2]Y. Nambu, Lectures at the Copenhagen Summer Symposium, 1970 (unpublished); for a recent review, see J. Scherk, Rev. Mod. Phys. **47**, 123 (1975).
[3]R. Marnelius, Nucl. Phys. B **104**, 477 (1976).
[4]L. P. Staunton, Phys. Rev. D **13**, 3269 (1976).
[5]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[6]L. P. Staunton, J. Math. Phys. **19**, 1471 (1978).
[7]P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).
[8]L. C. Biedenharn, M. Y. Han, and H. van Dam, Phys. Rev. D **8**, 1735 (1973).
[9]Our conventions are: $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$; $\hbar = c = 1$; $\epsilon_{0123} = +1$; $i, j, k, l, m, n = 0, 1, 2, 3$; $I, J, K, L = 0, 1, ..., N^2 - 1$; A, B, C denote indices of antisymmetric u(N) matrices, S, T, U those of symmetric u(N) matrices.
[10]With S, T, U denoting indices of symmetric matrices G_I and A, B, C those of antisymmetric matrices, $d'_{STU} = d_{STU}$, $d'_{SAB} = d_{SAB}$, $d'_{ABS} = -d_{ABS}$, $f'_{STA} = -f_{STA} f'_{SAT} = f_{SAT}$, $f'_{ABC} = f_{ABC}$.
[11]P. Halmos, *Finite-Dimensional Vector Spaces* (Springer, New York, 1974), p. 96.
[12]The entire U($2N$) subgroup is generated on a basis using, for complex numbers, a subset of a 2×2 matrix representation of quaternions.
[13]S. Browne, Nucl. Phys. B **79**, 70 (1974).
[14]The operators (5.3) are self-adjoint on a space of L_2 functions of q under the inner product $\int d^{2N}q$.
[15]The superscripts in (5.4) and subscript numbers in (5.5) reflect the choice of the $\{\lambda_I\}$ generating $\text{U(2)}_{\text{isospin}}$ to be typically numbered Pauli matrices times $\frac{1}{2}$.

Wigner coefficients for a semisimple Lie group and the matrix elements of the O(n) generators

M. D. Gould[a)]
Department of Mathematical Physics, University of Adelaide, G.P.O. Box 498, Adelaide, South Australia, 5001

(Received 29 May 1980; accepted for publication 22 August 1980)

A purely algebraic infinitesimal method for obtaining multiplicity-free Wigner coefficients is presented. The method is applied to obtain analytic expressions for the complete matrix elements of all O(n) generators. Moreover the structure of these matrix elements in terms of reduced matrix elements, Wigner coefficients, and reduced Wigner coefficients is made explicit. By comparison with the Wigner–Eckart theorem explicit analytic expressions are obtained for the fundamental Wigner coefficients of O(n). Finally the results are presented in a form which is directly analogous with the corresponding results for U(n).

PACS numbers: 02.20.Sv

I. INTRODUCTION

In the literature two distinct, yet intimately related, approaches to a general study of the classical groups have emerged. Firstly, there is the algebraic infinitesimal approach which exploits only the generators and their commutation relations. This approach has its origin in the pioneering researches of Casimir,[1,2] Van der Waerden,[2] and Racah.[3] Secondly, there is the integral approach as expounded in the classic works of Weyl.[4] The methods of Weyl have proved a powerful tool in group theoretic applications to physics and have been applied, in conjunction with Schwinger's boson calculus,[5] by various authors.[6]

From the point of view of applications to physics the principal problems to be solved are the complete determination of the states of an irreducible representation and the explicit determination of Wigner (or Clebsch–Gordan) coefficients. Probably the first major step in this direction was made in 1950 by Gel'fand and Zetlin[7] who constructed, with a full set of labels, a complete set of basis vectors for the irreducible representations of the orthogonal and unitary groups. The matrix elements of the group generators were also given initially by Gel'fand and Zetlin[7] and later by Baird and Biedenharn[8] who made an important contribution by revealing the structure of the matrix elements (i.e., a product of a reduced matrix element and a Wigner coefficient). The evaluation of all multiplicity-free Wigner coefficients of U(n) was subsequently given by Biedenharn, Louck, Baird, and Giovannini.[9] Although much work has been done on O(n) by several authors,[10–12] the complete program followed by Biedenharn, Louck *et al.* for U(n) has never been carried out for O(n) (although it seems evident[11] that it will extend to the orthogonal group with appropriate modifications).

The approach employed by Biedenharn, Louck *et al.* relies on the group theoretic methods of the Young tableaux and Schwinger's boson calculus. These methods are inherently integral in nature as are the techniques of Gel'fand *et al.*. In the 1960's an alternative infinitesimal approach to these problems was developed by Nagel and Moshinsky[13] for the unitary groups and extended to the orthogonal groups by Pang and Hecht[14] and Wong.[15] This approach relies on raising and lowering operators which are constructed from polynomials in the group generators. Gel'fand–Zetlin basis states may then be written as a product of lowering operators acting on the state of highest weight thus enabling, in principle, a complete determination of the group generator matrix elements. This approach has recently been given a more elegant treatment[16,17] by making use of polynomial identities satisfied by the infinitesimal generators of the group. Although such raising and lowering operators are useful for various applications we feel however that for obtaining the matrix elements of the group generators, and the more general and related problem of obtaining the multiplicity-free Wigner coefficients, this procedure is somewhat involved and undermines the simplicity apparent in the final results.

In recent work by the author[17,18] it was shown how one may obtain the complete matrix elements of the group generators algebraically using the concept of simultaneous shift operators which shift the representation labels of the group U(n) and each of its canonical[13] subgroups in a certain prescribed way. The principal tool in this approach is the construction of projection operators using polynomial identities satisfied by the infinitesimal generators of the group. The matrix elements of such projectors were shown to completely determine the fundamental Wigner coefficients of U(n). Moreover an expression for the reduced U(n):U($n-1$) Wigner coefficients (or isoscalar factors) was given as a U($n-1$)-trace of a polynomial in the U(n) generators. This operator is an element of the universal enveloping algebra of U(n) which commutes with all the U($n-1$) generators. Its eigenvalues determine the squares of the reduced Wigner coefficients. This reduces the problem of obtaining reduced Wigner coefficients to an algebraic problem of obtaining eigenvalues of a U($n-1$) Casimir invariant and provides a systematic method for their evaluation. The principal value of this approach is that it generalizes to arbitrary multiplicity-free tensor products and to more general groups. In particular it enables a treatment of the orthogonal group in exact analogy with the unitary group.

a)Present address: School of Physical Sciences, The Flinders University of South Australia, Bedford Park, South Australia, 5042.

 0022-2488/81/112376-13$01.00

In this paper we apply the techniques developed in Refs. 17 and 18 to obtain the matrix elements of all O(n) generators. As for the U(n) case the O(n):O($n-1$) reduced Wigner coefficients are obtained as an O($n-1$)-trace of a polynomial in the O(n) generators. Furthermore the general fundamental Wigner coefficients are obtained in exact analogy with the U(n) case.

We begin in Sec. II by outlining our procedure for an arbitrary semisimple Lie group. In Sec. III we obtain the reduced matrix elements and reduced Wigner coefficients of O(n). In Secs. IV–VII the structure of the matrix elements of all O(n) generators is determined.

II. MULTIPLICITY-FREE WIGNER COEFFICENTS

Since our approach is primarily based on the construction of projection operators using polynomial identities satisfied by the infinitesimal generators of the group we present a brief summary of results in this field.

Let G be a semisimple (compact) Lie group with Lie algebra L. Let H be a Cartan subalgebra of L, H^* the dual space to H, and $\Phi \subset H^*$ the set of roots with respect to the pair (L,H). Let $\Phi^+ \subset \Phi$ be a system of positive roots, δ the half sum of the positive roots, Λ the set of integral weights, and $\Lambda^+ \subset \Lambda$ the set of dominant integral weights. Let $(\ ,\ )$ denote the inner product induced on H^* by the Killing form. Finally let U denote the universal enveloping algebra of L and let Z be the center of U.

Now let $V(\lambda)$ be a finite dimensional irreducible module over U with highest weight $\lambda \in \Lambda^+$ and let π_λ be the representation afforded by $V(\lambda)$. Following Kostant[19] we consider the map

$$\partial: U \to \text{End } V(\lambda) \otimes U$$

defined for $x \in L$ by

$$\partial(x) = \pi_\lambda(x) \otimes 1 + 1 \otimes x,$$

which we extend to an algebra homomorphism to all of U. For example, if $x,y \in L$ then

$$\begin{aligned}\partial(xy) &= \partial(x)\partial(y) \\ &= \pi_\lambda(xy) \otimes 1 + \pi_\lambda(y) \otimes x + \pi_\lambda(x) \otimes y + 1 \otimes xy.\end{aligned}$$

If z is an element of the center Z of U we consider the operator $A(z)$ defined by

$$A(z) = -\tfrac{1}{2}[\partial(z) - \pi_\lambda(z) \otimes 1 - 1 \otimes z]$$

which may be viewed as a $d \times d$ matrix ($d = \dim V(\lambda)$) with entries from U. When acting on a finite dimensional irreducible module $V(\mu)$, $\mu \in \Lambda^+$, over U the matrix $A(z)$ may be written

$$A(z) = -\tfrac{1}{2}[\pi_\lambda \otimes \pi_\mu(z) - \pi_\lambda(z) \otimes 1 - 1 \otimes \pi_\mu(z)],$$

which is an operator on the tensor product space $V(\lambda) \otimes V(\mu)$.

Let us write the Clebsch–Gordan decomposition of $V(\lambda) \otimes V(\mu)$ into irreducible modules over U according to the standard convention

$$V(\lambda) \otimes V(\mu) = \bigoplus_\nu m(\nu) V(\nu), \tag{1}$$

where $m(\nu)$ is the multiplicity of π_ν in $\pi_\lambda \otimes \pi_\mu$ and $m(\nu)V(\nu)$ is shorthand notation for

$$V(\nu) \oplus V(\nu) \oplus \cdots \oplus V(\nu) \quad [m(\nu) \text{ times}].$$

We now note that on each space $V(\nu)$ occuring in (1) the operator $A(z)$ takes the constant value

$$\alpha_\nu(z) = -\tfrac{1}{2}(\chi_\nu(z) - \chi_\lambda(z) - \chi_\mu(z)), \tag{2}$$

where $\chi_\nu(z)$, $\nu \in H^*$, denotes the eigenvalue of the central element z on $V(\nu)$. From this it is an easy matter to deduce that acting on the space $V(\mu)$ the matrix $A(z)$ satisfies the polynomial identity

$$\prod_\nu (A(z) - \alpha_\nu(z)) = 0. \tag{3}$$

We remark here that if $\lambda_1,\ldots,\lambda_k$ are the distinct weights occuring in $V(\lambda)$ and if ν is a highest weight occurring in the decomposition (1) then ν is necessarily of the form $\nu = \mu + \lambda_i$ for some $i = 1,\ldots,k$. Moreover if the weight λ_i occurs with multiplicity $n(i)$ in $V(\lambda)$ then we necessarily have $m(\nu) \leqslant n(i)$. In particular if the weights occurring in $V(\lambda)$ have unit multiplicity then the tensor product space $V(\lambda) \otimes V(\mu)$ (for arbitrary $\mu \in \Lambda^+$) is multiplicity free.

In the special case where $z = C_L$ is the universal Casimir element the identities (3) reduce to the identities encountered recently by several authors for the various classical groups.[20–22] In this case the matrix A may be expressed in the form

$$A = -\sum_{r=1}^{n} \pi_\lambda(x^r) x_r, \tag{4}$$

where $\{x_1,\ldots,x_n\}$ ($n = \dim L$) is a basis for L and $\{x^1,\ldots,x^n\}$ is the corresponding dual basis with respect to the Killing form on L. Thus in this case the matrix A is a matrix with entries from L. Invariants of higher order than C_L may be constructed by taking traces of powers of the matrix A

$$I_m = \text{tr}(A^m).$$

The I_m are elements of the center Z which (at least for the simple Lie algebras) generate the center (although they are not of course all algebraically independent). Their eigenvalues are given by the explicit formula[23]

$$\chi_\mu(I_m) = \sum_{i=1}^{k} n(i)\alpha_i(\mu)^m \prod_{\alpha \in \Phi^+} \frac{(\mu + \lambda_i + \delta, \alpha)}{(\mu + \delta, \alpha)},$$

where $\lambda_1,\ldots,\lambda_k$ are the distinct weights occurring in $V(\lambda)$ with multiplicities $n(1),\ldots,n(k)$, respectively, and where α_i denotes the polynomial function

$$\begin{aligned}\alpha_i(\mu) &= -\tfrac{1}{2}[\chi_{\mu+\lambda_i}(C_L) - \chi_\mu(C_L) - \chi_\lambda(C_L)] \\ &= \tfrac{1}{2}(\lambda,\lambda + 2\delta) - \tfrac{1}{2}(\lambda_i, 2(\mu + \delta) + \lambda_i).\end{aligned}$$

It is easy to show that there exists an element z of the center Z such that the numbers $\chi_\nu(z)$, for each highest weight ν occuring in the decomposition (1), are all distinct. [In particular if $V(\lambda)$ is the fundamental vector representation of one of the classical Lie groups then the universal Casimir element will suffice.] By virtue of the polynomial identity (3) one may then construct a set of projection operators

$$P[\nu] = \prod_{\rho \neq \nu} \left(\frac{A - \alpha_\rho}{\alpha_\nu - \alpha_\rho} \right),$$

where $A = A(z)$. Since the matrix A takes the constant value $\alpha_\nu = \alpha_\nu(z)$ on the space $V(\nu)$ it follows that the projector $P[\nu]$

takes the constant value 1 on the space $V(\nu)$ and zero on the remaining $V(\rho), \rho \neq \nu$. Thus $P[\nu]$ projects $V(\lambda) \otimes V(\mu)$ onto the subspace $m(\nu)V(\nu)$. It follows that the matrix elements of the projector $P[\nu]$ between basis states in the space $V(\mu)$ are bilinear combinations of Clebsch–Gordan coefficients.

To be more explicit let $e_1^\lambda,\ldots,e_d^\lambda$ and $e_1^\mu,\ldots e_{d'}^\mu$, $d = \dim V(\lambda)$, $d' = \dim V(\mu)$, be orthonormal bases in the spaces $V(\lambda)$ and $V(\mu)$, respectively. In order to distinguish between the equivalent representations occurring in (1) we provide them with an additional index r; $V(\nu,r)$, $r = 1,\ldots,m(\nu)$. With this convention the decomposition (1) may be written

$$V(\lambda) \otimes V(\mu) = \bigoplus_{\nu} \bigoplus_{r=1}^{m(\nu)} V(\nu,r).$$

Let $e_1^{(\nu,r)},\ldots,e_{d(\nu)}^{(\nu,r)}$, $d(\nu) = \dim V(\nu)$, be an orthonormal basis of the space $V(\nu,r)$. The basis vectors $e_p^{\nu,r}$ form a basis for the product space $V(\lambda) \otimes V(\mu)$ which is related to the product basis $e_i^\lambda \otimes e_j^\mu$ by the change of basis transformation

$$e_p^{(\nu,r)} = \sum_{ij} \langle e_i^\lambda, e_j^\mu \mid e_p^{\nu,r} \rangle e_i^\lambda \otimes e_j^\mu .$$

The elements $\langle e_i^\lambda, e_j^\mu \mid e_p^{\nu,r} \rangle$ of this (unitary) basis transformation are Clebsch–Gordan coefficients.

Now let e_i^λ and e_j^λ be two arbitrary basis states in the space $V(\lambda)$. Since $P[\nu]$ is a polynomial in the matrix A it follows, since A is a matrix with entries from U, that the entries

$$P[\nu]^i{}_j = \langle e_i^\lambda | P[\nu] | e_j^\lambda \rangle$$

of the matrix $P[\nu]$ are well defined elements of U. On the other hand viewing $P[\nu]$ as an operator on the space $V(\lambda) \otimes V(\mu)$ we see that the matrix elements of $P[\nu]^i{}_j$ between basis states in the space $V(\mu)$ are given by

$$\langle e_k^\mu | P[\nu]^i{}_j | e_l^\mu \rangle = \langle e_k^\mu, e_i^\lambda | P[\nu] | e_j^\lambda, e_l^\mu \rangle,$$

where $|e_j^\lambda, e_l^\mu\rangle$ is bra-ket notation for the product state $e_j^\lambda \otimes e_l^\mu$. Introducing a complete set of states for the space $V(\lambda) \otimes V(\mu)$ we have, using the fact that $P[\nu]$ projects $V(\lambda) \otimes V(\mu)$ onto the subspaces $V(\nu)$,

$$\langle e_k^\mu | P[\nu]^i{}_j | e_l^\mu \rangle = \sum_{r=1}^{m(\nu)} \sum_{q=1}^{d(\nu)} \langle e_k^\mu, e_i^\lambda \mid e_q^{\nu,r} \rangle \langle e_q^{r,\nu} \mid e_j^\lambda, e_l^\mu \rangle . \quad (5)$$

In particular putting $i = j$, $k = l$ we obtain

$$\langle e_k^\mu | P[\nu]^i{}_i | e_k^\mu \rangle = \sum_{r=1}^{m(\nu)} \sum_{q=1}^{d(\nu)} | \langle e_k^\mu, e_i^\lambda \mid e_q^{\nu,r} \rangle |^2.$$

If, moreover, the module $V(\nu)$ occurs with multiplicity 1 then we may write (dropping the multiplicity label)

$$\langle e_k^\mu | P[\nu]^i{}_i | e_k^\mu \rangle = \sum_{q=1}^{d(\nu)} | \langle e_k^\mu, e_i^\lambda \mid e_q^\nu \rangle |^2. \quad (6)$$

In many cases the right-hand side reduces to a single term enabling an evaluation of certain Wigner coefficients (up to a phase) by an independent evaluation of the left-hand side. An immediate application of formula (6) is the matrix element of $P[\nu]^i{}_i$ between the maximal state e^μ in $V(\mu)$. Suppose we choose an orthonormal basis for the (reference) representation $V(\lambda)$ to be a weight basis. Recall from our previous remarks (see also Refs. 19 and 23) that the highest weight ν of $V(\nu)$ is given by $\nu = \mu + \lambda_i$ for some weight λ_i in $V(\lambda)$. Now let e_i^λ be a basis vector of wieght λ_i so that the tensor product state $e_i^\lambda \otimes e^\mu$ has weight $\nu = \mu + \lambda_i$. Then, according to (6), we have

$$\langle e^\mu | P[\nu]^i{}_i | e^\mu \rangle = \sum_{q=1}^{d(\nu)} | \langle e^\mu, e_i^\lambda \mid e_q^\nu \rangle |^2.$$

Since $e_i^\lambda \otimes e^\mu$ has weight ν which is the highest weight occurring in $V(\nu)$ it follows that the only surviving term in the above sum is

$$\langle e^\mu | P[\nu]^i{}_i | e^\mu \rangle = | \langle e^\mu, e_i^\lambda \mid e^\nu \rangle |^2.$$

This Clebsch–Gordan coefficient is important for the normalization of generalized raising and lowering operators for the group. This application is discussed more fully in Ref. 17. [Note that although a weight basis is required for $V(\lambda)$ we do not require a weight basis for the spaces $V(\mu)$ or $V(\nu)$ but only that the maximal weight vector belong to the basis. It is also not necessary that the weights occur in $V(\lambda)$ with multiplicity 1.]

More detailed information may be obtained by making use of a suitable chain of subgroups for G. In particular if the group G admits a canonical chain[13] of subgroups then all (multiplicity-free) Wigner coefficients may be obtained directly by this method. This procedure has been illustrated for the unitary group in Ref. 18. It is our aim here to generalize this technique to obtain the matrix elements of all O(n) generators. We remark however that in order to obtain the matrix elements of the group generators one need only consider tensor products of the form $V^* \otimes V(\mu)$, where V^* is the carrier space for the fundamental contragredient vector representation. Thus it suffices to apply only the vector identity (and its adjoint) for this problem.

III. REDUCED MATRIX ELEMENTS AND REDUCED WIGNER COEFFICIENTS FOR O(n)

Without loss of generality we may take as a set of generators for O(n) the operators $\alpha^i{}_j$ ($i, j = 1,\ldots,n$) which satisfy the relations

$$\alpha^i{}_j = -\alpha^j{}_i, \quad [\alpha^i{}_j, \alpha^k{}_l] = \delta^k_j \alpha^i_l - \delta^i_l \alpha^k_j - \delta^k_i \alpha^j_l + \delta^j_l \alpha^k_i,$$

and the Hermiticity property

$$(\alpha^i{}_j)^\dagger = \alpha^j{}_i.$$

This corresponds to the choice of O(n) metric $g_{ij} = \delta_{ij}$.

As for U(n) the generators of O(n) for naturally into an $n \times n$ matrix $\alpha = (\alpha^i{}_j)$ which is a special case of the matrix A appearing in Eq. (4) for the case where $\pi_\lambda = \pi^*$ is the fundamental contragredient vector representation. Associated with the matrix α is its adjoint $\bar{\alpha}$ with entries

$$\bar{\alpha}_j{}^i = -\alpha^i{}_j.$$

Polynomials in α and $\bar{\alpha}$ may be defined recursively according to

$$(\alpha^{m+1})^i{}_j = (\alpha^m)^i{}_k \alpha^k_j = \alpha^i_k (\alpha^m)^k{}_j,$$
$$(\bar{\alpha}^{m+1})_j{}^i = (\bar{\alpha}^m)_j{}^k \bar{\alpha}_k{}^i = \bar{\alpha}^k_j (\bar{\alpha}^m)_k{}^i.$$

The representations of O(n) may be labelled by the maximum eigenvalues of the commuting Hermitian operators

$$-i\alpha^{2r-1}_{2r}, \quad r = 1,\ldots,h,$$

where

$$h = \left[\frac{n}{2} \right] = \begin{cases} \frac{1}{2}n & n \text{ even}, \\ \frac{1}{2}(n-1) & n \text{ odd}. \end{cases}$$

The set of weights may be identified with the set of all tuples of the form $(\lambda_1,...,\lambda_h)$. We let Δ_r $(r = 1,...,h)$ denote the fundamental weights with 1 in position r and zeros elsewhere. We define weights Δ_r $(r = 1,...,n)$ by defining Δ_r for $r > h$ according to

$$\Delta_r = -\Delta_{n+1-r}. \tag{7}$$

For the case of odd $n = 2h + 1$ we have also the zero weight $\Delta_{h+1} = 0$ [which is consistent with Eq. (7)].

On a finite dimensional irreducible module $V(\lambda)$ with highest weight $\lambda = (\lambda_1,...,\lambda_h)$ the matrices α and $\bar{\alpha}$ satisfy the polynomial identities

$$\prod_{r=1}^{n} (\alpha - \alpha_r) = 0,$$

$$\prod_{r=1}^{n} (\bar{\alpha} - \bar{\alpha}_r) = 0,$$

where the roots α_r and $\bar{\alpha}_r$ are given by

$$\alpha_r = \bar{\alpha}_{n+1-r} = \lambda_r + n - 1 - r,$$

where we define labels λ_r for $r > h$ according to

$$\lambda_{n+1-r} = 1 - \lambda_r, \quad r = 1,...,h,$$

with

$$\lambda_{h+1} = 1, \quad \text{for} \quad n = 2h + 1. \tag{8}$$

As for the U(n) case one may construct projection operators $P[r]$ and $\bar{P}[r]$ by setting

$$P[r] = \prod_{l \neq r} \left(\frac{\alpha - \alpha_l}{\alpha_r - \alpha_l} \right),$$

$$\bar{P}[r] = \prod_{l \neq r} \left(\frac{\bar{\alpha} - \bar{\alpha}_l}{\bar{\alpha}_r - \bar{\alpha}_l} \right).$$

The matrix elements of these projectors in unitary representations of the group determine the fundamental Wigner coefficients of the group. Following Ref. 17 let $|^{\lambda}_{(\nu)}\rangle$ and $|^{\lambda}_{(\nu')}\rangle$ be two arbitrary Gel'fand basis states in the space $V(\lambda)$. As a special case of formula (5) (noting that the tensor product in this case is multiplicity free) we have

$$\left\langle {}^{\lambda}_{(\nu)} \middle| P[\nu]^i_j \middle| {}^{\lambda}_{(\nu')} \right\rangle = \sum_{(\mu)} \left\langle {}^{\lambda}_{(\nu)}; {}^{\overline{1\dot{0}}}_{i} \middle| {}^{\lambda-\Delta_r}_{(\mu)} \right\rangle \left\langle {}^{\lambda-\Delta_r}_{(\mu)} \middle| {}^{\overline{1\dot{0}}}_{j}; {}^{\lambda}_{(\nu')} \right\rangle, \tag{9}$$

where $|^{\overline{1\dot{0}}}_{i}\rangle$ constitutes an orthonormal basis for the contragredient vector representation (i.e., the row vectors with 1 in the ith position and zeros elsewhere) and where we define weights Δ_r for $r > h$ in accordance with (7). Similarly we have

$$\left\langle {}^{\lambda}_{(\nu)} \middle| \bar{P}[\nu]^i_j \middle| {}^{\lambda}_{(\nu')} \right\rangle = \sum_{(\mu)} \left\langle {}^{\lambda}_{(\nu)}; {}^{1\dot{0}}_{i} \middle| {}^{\lambda+\Delta_r}_{(\mu)} \right\rangle \left\langle {}^{\lambda+\Delta_r}_{(\mu)} \middle| {}^{1\dot{0}}_{j}; {}^{\lambda}_{(\nu')} \right\rangle, \tag{10}$$

where $|^{1\dot{0}}_{i}\rangle$ constitutes an orthonormal basis for the vector representation (i.e., the column vectors with 1 in position i and zeros elsewhere). We remark that the basis chosen here for the vector representation differs from that used in Ref. 17 since in the latter the basis states are weight states as distinct from the Gel'fand basis used here. As a special case of Eqs. (9) and (10) we have

$$\left\langle {}^{\lambda}_{(\nu)} \middle| P[r]^n_n \middle| {}^{\lambda}_{(\nu')} \right\rangle = \delta_{(\nu),(\nu')} \left| \left\langle {}^{\lambda}_{(\nu)}; {}^{\overline{1\dot{0}}}_{n} \middle| {}^{\lambda-\Delta_r}_{(\nu)} \right\rangle \right|^2,$$

$$\left\langle {}^{\lambda}_{(\nu)} \middle| P[r]_n{}^n \middle| {}^{\lambda}_{(\nu')} \right\rangle = \delta_{(\nu),(\nu')} \left| \left\langle {}^{\lambda}_{(\nu)}; {}^{1\dot{0}}_{n} \middle| {}^{\lambda+\Delta_r}_{(\nu)} \right\rangle \right|^2. \tag{11}$$

Thus we see that the operators $P[r]^n_n$ and $\bar{P}[r]_n{}^n$ are O$(n-1)$ invariants which determine squares of fundamental Wigner coefficients. Turning our attention to the group O$(n+1)$ we let β denote the O$(n+1)$ matrix whose (i,j) entry is the genrator α^i_j $(i,j = 1,...,n+1)$. Following our U(n) notation we write the characteristic identity satisfied by the matrix β in the form

$$\prod_{r=1}^{n+1} (\beta - \beta_r) = 0,$$

where the β_r take constant values on a finite-dimensional irreducible representation with highest weight λ given by $\beta_r = \lambda_r + n - r$. In a similar way we define the adjoint matrix $\bar{\beta}$ whose roots are given by $\bar{\beta}_r = \beta_{n+2-r}$. We denote the O$(n+1)$ projection operators by

$$Q[k] = \prod_{l \neq k} \left(\frac{\beta - \beta_l}{\beta_k - \beta_l} \right),$$

$$\bar{Q}[k] = \prod_{l \neq k} \left(\frac{\bar{\beta} - \bar{\beta}_l}{\bar{\beta}_k - \bar{\beta}_l} \right).$$

The operators $Q[k]^{n+1}{}_{n+1}$ and $\bar{Q}[k]_{n+1}{}^{n+1}$ are the O$(n+1)$ analogs of the operators $P[r]^n_n$ and $\bar{P}[r]_n{}^n$ whose matrix elements are squares of Wigner coefficients.

Following our U(n) notation we denote the O(n) invariants $Q[k]^{n+1}{}_{n+1}$ and $\bar{Q}[k]_{n+1}{}^{n+1}$ by C_k and $\bar{C}_k$ respectively. Applying the identities satisfied by the matrices α and β (see Ref. 24 and Appendix A) one may express the O(n) invariants C_k and $\bar{C}_k$ as a function of the β_k and α_r according to

$$C_k = \prod_{l \neq k} (\beta_k - \beta_l)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r - \eta_r),$$

$$C_{n+2-k} = \bar{C}_k = \prod_{l \neq k} (\bar{\beta}_k - \bar{\beta}_l)^{-1} \prod_{r=1}^{n} (\bar{\beta}_k - \bar{\alpha}_r - \eta_r), \tag{12}$$

where

$$\eta_r = \begin{cases} 1 & n = 2h, \\ 1 - \delta_{r,h+1} & n = 2h + 1. \end{cases} \tag{13}$$

In passing it is interesting to note, as in the U(n) case, that the spectral resolution of the matrix β implies

$$P(\beta)^{n+1}{}_{n+1} = \sum_{k=1}^{n+1} P(\beta_k) C_k,$$

which enables a systematic evaluation of O(n) invariants of the form $P(\beta)^{n+1}{}_{n+1}$, for arbitrary polynomials $p(x)$, by applying formula (12).

In an analogous way one may express the operators

$P[r]^n{}_n$ and $\bar{P}[r]_n{}^n$ as a function of the roots in the O(n) and O($n-1$) identities. This enables us to evaluate the Wigner coefficients (11) as required. However in order to determine the matrix elements of the group generators one must also obtain the corresponding reduced matrix elements.

Throughout the remainder of this paper let ψ denote the O(n) vector operator with components $\psi^i = \alpha^i{}_{n+1}$ ($i = 1,...,n$) with adjoint $\psi^\dagger$ whose components are given by $\psi^\dagger{}_i = \alpha^{n+1}{}_i$. Following Green and Bracken,[20] the vector operator ψ and its contragredient may be resolved into a sum of shift vectors

$$\psi = \sum_{r=1}^{n} \psi[r], \quad \psi^\dagger = \sum_{r=1}^{n} \psi^\dagger[r],$$

which alter the O(n) representation labels according to

$$\lambda_k \psi[r] = \psi[r](\lambda_k + \delta_{kr}),$$
$$\lambda_k \psi[n+1-r] = \psi[n+1-r](\lambda_k - \delta_{kr}),$$
$$\lambda_k \psi^\dagger[r] = \psi^\dagger[r](\lambda_k - \delta_{kr}),$$
$$\lambda_k \psi^\dagger[n+1-r] = \psi^\dagger[n+1-r](\lambda_k + \delta_{kr}),$$

for $k,r = 1,...h$ with

$$\lambda_k \psi^\dagger[h+1] = \psi^\dagger[h+1]\lambda_k, \quad \text{for } n = 2h+1.$$
$$\lambda_k \psi[h+1] = \psi[h+1]\lambda_k,$$

These shift vectors may be constructed by application of the projectors $P[r]$ and $\bar{P}[r]$ as follows

$$\psi[r] = P[r]\psi = \psi\bar{P}[r], \quad r = 1,...,n$$
$$\psi^\dagger[r] = \bar{P}[r]\psi^\dagger = \psi^\dagger P[r].$$

Since the matrix elements of the projectors $P[r]$ and $\bar{P}[r]$ are bilinear combinations of Clebsch–Gordan coefficients we obtain, from the Wigner–Eckart theorem, the result

$$\psi[r]\psi^\dagger[r] = \bar{M}_r P[r],$$
$$\psi^\dagger[r]\psi[r] = M_r \bar{P}[r], \tag{14}$$

where the $M_r(\bar{M}_r)$ are O(n) invariants whose eigenvalues determine the squares of the reduced matrix elements of $\psi(\psi^\dagger)$. Taking the trace of Eq. (14) the invariants M_r and $\bar{M}_r$ may be evaluated by applying the equation

$$\bar{M}_r = M_{n+1-r} = \frac{\psi[r]^i\psi^\dagger[r]_i}{\mathrm{tr}(P[r])}.$$

By this means we may express the M_r and $\bar{M}_r$ as a function of the β_k and α_r according to (see Ref. 24 and Appendix A)

$$\bar{M}_r = (-1)^n \prod_{k=1}^{n+1} (\bar{\beta}_k - \bar{\alpha}_r) \times \prod_{l \neq r} (\bar{\alpha}_r - \bar{\alpha}_l - \eta_l - \delta_{l,n+1-r})^{-1},$$
$$M_r = (-1)^n \prod_{k=1}^{n+1} (\beta_k - \alpha_r) \times \prod_{l \neq r} (\alpha_r - \alpha_l - \eta_l - \delta_{l,n+1-r})^{-1}, \tag{15}$$

where η_r is given by Eq. (13).

By taking the (n,n) entries of Eq. (14) we obtain

$$\psi[r]^n\psi^\dagger[r]_n = \bar{M}_r P[r]^n{}_n,$$
$$\psi^\dagger[r]_n\psi[r]^n = M_r \bar{P}[r]_n{}^n, \tag{16}$$

which, using formulas (12) and (15), enables one to determine the matrix elements of O(n) generators of the form $\alpha^{n+1}{}_n$. However in order to obtain the matrix elements of the remaining generators it is necessary to proceed down the subgroup chain in exact analogy with the U(n) case.[18] To this end we obtain a relationship between the O(n) and O($n+1$) projection operators which determines the O($n+1$):O(n) reduced Wigner operators.

Following our U(n) derivation we have the result (see also Appendix A)

$$\sum_{l,m=1}^{n} P[r]^i{}_l Q[k]^l{}_m P[r]^m{}_j = \alpha\begin{pmatrix} n+1 & n \\ k & r \end{pmatrix} P[r]^i{}_j, \tag{17}$$

where $\alpha\binom{n+1\ \ n}{k\ \ r}$ is an O(n) invariant whose eigenvalues determine the squares of the O($n+1$):O(n) reduced Wigner coefficients. Taking the O(n)-trace of Eq. (17) we see that $\alpha\binom{n+1\ \ n}{k\ \ r}$ may be evaluated algebraically from the equation

$$\alpha\begin{pmatrix} n+1 & n \\ k & r \end{pmatrix} = \frac{\mathrm{tr}(P[r]Q[k]P[r])}{\mathrm{tr}(P[r])} = \frac{\mathrm{tr}(P[r]Q[k])}{\mathrm{tr}(P[r])}, \tag{18}$$

where the second equality follows from the cyclic rule of traces and the projection property $P[r]^2 = P[r]$. One may express the $\alpha\binom{n+1\ \ n}{k\ \ r}$ as a function of the β_k and α_r according to the formula (see Appendix A)

$$\alpha\begin{pmatrix} n+1 & n \\ k & r \end{pmatrix} = C_k \bar{M}_r (\beta_k - \alpha_r - \eta_r)^{-1}(\beta_k - \alpha_r)^{-1} \times (\beta_k - \alpha_{n+1-r} - \eta_r)^{-1}(\beta_k - \alpha_{n+1-r} - 2\eta_r). \tag{19}$$

This equation is clearly an analog of the corresponding result for U(n) although it is a slightly more complicated expression. As seen in Appendix A this extra complication arises due to the dependence of the root α_{n+1-r} on the representation label λ_r.

IV. SIMULTANEOUS SHIFTS

The Lie group O(n) admits the canonical chain of subgroups

$$\mathrm{O}(n) \supset \mathrm{O}(n-1) \supset \cdots \supset \mathrm{O}(2),$$

where each group O(m) occurring in this chain has infinitesimal geneators $\alpha^i{}_j$ where i and j are restricted to the range $1,...,m$. Following our U(n) notation[18] we denote the O(m) matrix $(\alpha^i{}_j)$ $(i,j = 1,...,m)$ by α_m and we denote the characteristic roots of α_m by $\alpha_{r,m}$ $(r = 1,...,m)$. The $\alpha_{r,m}$ take constant values on a finite-dimensional irreducible module over O(m) with highest weight $(\lambda_{1,m},...,\lambda_{h,m})$ $\{h = [m/2]\}$ given by $\alpha_{r,m} = \bar{\alpha}_{m+1-r,m} = \lambda_{r,m} + m - 1 - r$, where we define labels $\lambda_{r,m}$ for $r > h$ in accordance with Eq. (8). We denote the corresponding O(m) projectors by $P\binom{m}{r}$ and $\bar{P}\binom{m}{r}$, respectively:

$$P\binom{m}{r} = \prod_{l \neq r} \left(\frac{\alpha_m - \alpha_{l,m}}{\alpha_{r,m} - \alpha_{l,m}} \right),$$

$$\bar{P}\begin{pmatrix} m \\ r \end{pmatrix} = \prod_{l \neq r} \left(\frac{\bar{\alpha}_m - \bar{\alpha}_{l,m}}{\bar{\alpha}_{r,m} - \bar{\alpha}_{l,m}} \right).$$

We denote the (m,m) entries of these projectors by $C_{r,m}$ and $\bar{C}_{r,m}$, respectively. From the previous section we know that these operators determine squares of Wigner coefficients whose eigenvalues are given by [see Eq. (12)]

$$C_{r,m} = \bar{C}_{m+1-r,m} = \prod_{k \neq r} (\alpha_{r,m} - \alpha_{k,m}) \prod_{l=1}^{m-1} (\alpha_{r,m} - \alpha_{l,m-1} - \eta_{l,m-1})^{-1}, \tag{20}$$

where $\eta_{l,m-1}$ is given by Eq. (13) (with n replaced by $m-1$); i.e.,

$$\eta_{l,m-1} = \begin{cases} 1 & \text{for } m-1 = 2h, \\ 1 - \delta_{l,h+1} & \text{for } m-1 = 2h+1. \end{cases} \tag{21}$$

We denote the O(m) vector operator $(\alpha^i{}_{m+1})$ $(i = 1,...,m)$ by $\psi(m)$ and denote its Hermitian conjugate by $\psi^\dagger(m)$. We denote the shift components of these operators by $\psi\binom{m}{r}$ and $\psi^\dagger\binom{m}{r}$, respectively. We then have, in accordance with Eq. (14),

$$\begin{aligned} &\psi\tbinom{m}{r}\psi^\dagger\tbinom{m}{r} = \bar{M}_{r,m} P\tbinom{m}{r}, \\ &\psi^\dagger\tbinom{m}{r}\psi\tbinom{m}{r} = M_{r,m} \bar{P}\tbinom{m}{r}, \end{aligned} \tag{22}$$

where the O(m) invariants $\bar{M}_{r,m}$ and $M_{r,m}$ (the squared reduced matrix elements) are given by

$$\begin{aligned} \bar{M}_{r,m} &= M_{m+1-r,m} \\ &= (-1)^m \prod_{k=1}^{m+1} (\alpha_{k,m+1} - \alpha_{r,m}) \\ &\quad \times \prod_{l \neq r} (\alpha_{r,m} - \alpha_{l,m} - \eta_{l,m} - \delta_{l,m+1-r})^{-1}. \end{aligned} \tag{23}$$

We therefore have, for the cases m even and m odd,

$$\begin{aligned} \bar{M}_{r,m} &= M_{m+1-r,m} \\ &= (-1)^m \prod_{k=1}^{m+1} (\alpha_{k,m+1} - \alpha_{r,m}) \\ &\quad \times \prod_{l \neq r} (\alpha_{r,m} - \alpha_{l,m} - 1 - \delta_{l,m+1-r})^{-1}, \end{aligned}$$

$m = 2h,$

$$\begin{aligned} \bar{M}_{r,m} &= M_{m+1-r,m} \\ &= (-1)^m \prod_{k=1}^{m+1} (\alpha_{k,m+1} - \alpha_{r,m}) \\ &\quad \times \prod_{l \neq r} (\alpha_{r,m} - \alpha_{l,m} - 1 + \delta_{l,h+1} - \delta_{l,m+1-r})^{-1}, \end{aligned}$$

$m = 2h+1.$

Finally, we write Eq. (17) in the form

$$P\begin{pmatrix} m \\ r \end{pmatrix} P\begin{pmatrix} m+1 \\ k \end{pmatrix} P\begin{pmatrix} m \\ r \end{pmatrix} = \alpha\begin{pmatrix} m+1 & m \\ k & r \end{pmatrix} P\begin{pmatrix} m \\ r \end{pmatrix}, \tag{24}$$

where the O(m) invariant $\alpha\binom{m+1\ \ m}{k\ \ r}$ [the squared O($m+1$):O(m) reduced Wigner coefficient] is given by

$$\alpha\begin{pmatrix} m+1 & m \\ k & r \end{pmatrix} = C_{k,m+1} \bar{M}_{r,m} (\alpha_{k,m+1} - \alpha_{r,m} - \eta_{r,m})^{-1} (\alpha_{k,m+1} - \alpha_{r,m})^{-1} (\alpha_{k,m+1} - \alpha_{m+1-r,m} - \eta_{r,m})^{-1} \times (\alpha_{k,m+1} - \alpha_{m+1-r,m} - 2\eta_{r,m}). \tag{25}$$

As in the case of U(n) one sees that an O(n) generator of the form $\alpha^{m+1}{}_l$ transforms as a contragredient vector operator with respect to the subgroups O(m),...,O(l) (for $2 < l \leqslant m$). We may then proceed to resolve $\alpha^{m+1}{}_l$ into its simultaneous shift components in direct analogy with the U(n) case. However in the case of the orthogonal groups a special derivation is evidently necessary for the special cases where $l = 1,2$. [This occurs primarily due to the fact that neither of the generators $\alpha^{m+1}{}_1$ or $\alpha^{m+1}{}_2$ $(m \geqslant 2)$ transform as components of a vector with respect to O(2).] We shall treat the case $l \geqslant 3$ first and consider the special case $l = 1,2$ later.

Firstly in the case $l = m$ we see that $\alpha^{m+1}{}_m$ can only be a component of a contragredient vector operator with respect to the subgroups O(m) and we obtain a decomposition into O(m) shift components according to

$$\alpha^{m+1}{}_m = \sum_{r=1}^{m} \psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix}_m.$$

Next we note that a generator of the form $\alpha^{m+1}{}_{m-1}$ $(m-1 > 2)$ transforms as a component of a contragredient vector with respect to the subgroups O(m) and O($m-1$). Resolving $\alpha^{m+1}{}_{m-1}$ into its O(m) shift components we obtain a primary resolution

$$\alpha^{m+1}{}_{m-1} = \sum_{r=1}^{m} \psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix}_{m-1},$$

where

$$\psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix}_i = \bar{P}\begin{pmatrix} m \\ r \end{pmatrix}_i{}^j \alpha^{m+1}{}_j = \alpha^{m+1}{}_j P\begin{pmatrix} m \\ r \end{pmatrix}_i{}^j.$$

Now each $\psi^\dagger\binom{m}{r}_{m-1}$ is also a component of a contragredient vector operator with respect to O($m-1$). Hence we may further decompose $\psi^\dagger\binom{m}{r}_{m-1}$ into its O($m-1$) shift components according to

$$\psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix}_{m-1} = \sum_{l=1}^{m-1} \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix}_{m-1},$$

where

$$\psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} = \bar{P}\begin{pmatrix} m-1 \\ l \end{pmatrix} \psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix} = \psi^\dagger\begin{pmatrix} m \\ r \end{pmatrix} P\begin{pmatrix} m-1 \\ l \end{pmatrix}.$$

Hence we obtain the resolution

$$\alpha^{m+1}{}_{m-1} = \sum_{r=1}^{m} \sum_{l=1}^{m-1} \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix}_{m-1},$$

where each shift component $\psi^\dagger\binom{m\ \ m-1}{r\ \ l}$ simultaneously alters the representation labels of O(m) and its subgroup O($m-1$) according to

$$\begin{aligned} \lambda_{k,m} \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} &= \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} \\ &\quad \times (\lambda_{k,m} - \delta_{kr} + \delta_{k,m+1-r}), \\ \lambda_{p,m-1} \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} &= \psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} \\ &\quad \times (\lambda_{p,m-1} - \delta_{pl} + \delta_{p,m-l}), \end{aligned}$$

$$k = 1,...,\left[\frac{m}{2}\right]; \quad p = 1,...,\left[\frac{m-1}{2}\right].$$

The simultaneous shift components $\psi^\dagger\binom{m\ \ m-1}{r\ \ \ l}$ are clearly given by

$$\psi^\dagger\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} = \psi^\dagger(m)P\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix} = \overline{P}\begin{pmatrix} m-1 & m \\ l & r \end{pmatrix}\psi^\dagger(m),$$

where $P\binom{m\ \ m-1}{r\ \ \ l}$ and $\overline{P}\binom{m-1\ \ m}{l\ \ \ r}$ are the operators defined by

$$P\begin{pmatrix} m & m-1 \\ r & l \end{pmatrix}^i_j = \sum_{q=1}^{m-1} P\begin{pmatrix} m \\ r \end{pmatrix}^i_q P\begin{pmatrix} m-1 \\ l \end{pmatrix}^q_j, \quad i = 1,...,m,$$

$$P\begin{pmatrix} m-1 & m \\ l & r \end{pmatrix}^i_j = \sum_{q=1}^{m-1} \overline{P}\begin{pmatrix} m-1 \\ l \end{pmatrix}^q_j \overline{P}\begin{pmatrix} m \\ r \end{pmatrix}^i_q, \quad j = 1,...,m-1.$$

More generally a generator $\alpha^{m+1}{}_l$ $(3 \leqslant l \leqslant m)$ may be resolved into shift components which simultaneously alter the representaion labels of the subgroups O(m),...,O(l). This decomposition may be written as

$$\alpha^{m+1}{}_l = \sum_{i(k)} \psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}_l,$$

where the summation symbol is shorthand notation for

$$\sum_{i(m)=1}^{m} \sum_{i(m-1)=1}^{m-1} \cdots \sum_{i(l)=1}^{l} .$$

Each shift component alters the representation labels of the subgroups O(m),...,O(l) according to

$$\lambda_{k,p}\psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} = \psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}(\lambda_{k,p} - \delta_{k,i(p)} + \delta_{k,p+1-i(p)}) \quad \text{for } p = l,...,m;\ k = 1,...,[p/2].$$

Clearly then if $i(r) \leqslant [r/2]$ then the above shift component decreases the O(r) representation label $\lambda_{i(r),r}$ by 1 unit. On the other hand if $i(r) \geqslant [r/2]$ then $\lambda_{r+1-i(r),r}$ is increased by 1 unit. In the special case where r is odd and $i(r) = [r/2] + 1$ then none of the O(r) representation labels are altered.

These shift components may be constructed by repeated application of the subgroup projectors as in the case of U(n). We denote the $m \times l$ matrix of operators with entries given by

$$\sum_{q=1}^{m-1} \sum_{r=1}^{m-2} \cdots \sum_{s=1}^{l} P\begin{pmatrix} m \\ i(m) \end{pmatrix}^i_q P\begin{pmatrix} m-1 \\ i(m-1) \end{pmatrix}^q_r \cdots P\begin{pmatrix} l \\ i(l) \end{pmatrix}^s_j, \quad i = 1,...,m;\ \ j = 1,...,l,$$

simply by $P\binom{m\ \cdots\ l}{i(m)\ \cdots\ i(l)}$. Similarly we define the matrix of operators $\overline{P}\binom{l\ \cdots\ m}{i(l)\ \cdots\ i(m)}$ constructed from the adjoint projectors. Clearly then, the simultaneous shift components of the generators $\psi^\dagger(m)_l = \alpha^{m+1}{}_l$ are given by

$$\psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} = \overline{P}\begin{pmatrix} l & \cdots & m \\ i(l) & \cdots & i(m) \end{pmatrix}\psi^\dagger(m) = \psi^\dagger(m)P\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}.$$

In the same way we define the matrix of operators

$$\overline{P}\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} \quad \text{and} \quad P\begin{pmatrix} l & \cdots & m \\ i(l) & \cdots & i(m) \end{pmatrix}$$

defined in the same way but with the order reversed. These operators project out the simultaneous shift components of the generators $\alpha^l{}_{m+1}$;

$$\alpha^l{}_{m+1} = \sum_{i(k)} \psi\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}^l,$$

$$\psi\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} = P\begin{pmatrix} l & \cdots & m \\ i(l) & \cdots & i(m) \end{pmatrix}\psi(m) = \psi(m)\overline{P}\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}.$$

We may now present a generalization of Eq. (16) for the multiple shift vectors $\psi\binom{m\ \cdots\ l}{i(m)\ \cdots\ i(l)}$. We have

$$\begin{aligned}\psi\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}^l \psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}_l &= P\begin{pmatrix} l & \cdots & m-1 & m \\ i(l) & \cdots & i(m-1) & i(m) \end{pmatrix}^l_i \psi(m)^i\psi^\dagger(m)_j \\ &\quad \times P\begin{pmatrix} m & m-1 & \cdots & l \\ i(m) & i(m-1) & \cdots & i(l) \end{pmatrix}^j_l \\ &= P\begin{pmatrix} l & \cdots & m-1 \\ i(l) & \cdots & i(m-1) \end{pmatrix}^l_i \psi\begin{pmatrix} m \\ i(m) \end{pmatrix}^i \psi^\dagger\begin{pmatrix} m \\ i(m) \end{pmatrix}_j \\ &\quad \times P\begin{pmatrix} m-1 & \cdots & l \\ i(m-1) & \cdots & i(l) \end{pmatrix}^j_l .\end{aligned}$$

Using Eq. (22) this in turn may be written

$$\overline{M}_{i(m),m} P\begin{pmatrix} l & \cdots & m-1 \\ i(l) & \cdots & i(m-1) \end{pmatrix} P\begin{pmatrix} m \\ i(m) \end{pmatrix} P\begin{pmatrix} m-1 & \cdots & l \\ i(m-1) & \cdots & i(l) \end{pmatrix}^l_l . \tag{26}$$

Hence by repeated application of Eqs. (24) and (25) we obtain

$$\begin{aligned}\psi\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}^l \psi^\dagger\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}_l &= \overline{M}_{i(l),l} C_{i(l),l} \prod_{r=l+1}^{m} \overline{M}_{i(r),r} C_{i(r),r} \\ &\quad \times \alpha_{i(r),r} - \alpha_{i(r-1),r-1} - \eta_{i(r-1),r-1})^{-1} \\ &\quad \times (\alpha_{i(r),r} - \alpha_{i(r-1),r-1})^{-1} \\ &\quad \times (\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - \eta_{i(r-1),r-1})^{-1} \\ &\quad \times (\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - 2\eta_{i(r-1),r-1}),\end{aligned} \tag{27}$$

which is the required generalization of Eq. (16).

We now consider the special case where $l = 1,2$. Although the generators $\alpha^{m+1}{}_1$ and $\alpha^{m+2}{}_2$ both transform as components of vectors with respect to the groups O(m),...,O(3) it is clear that neither of them transform as vectors with respect to the subgroup O(2). This leads us to consider the operators $\psi_\pm(m)$ defined by

$$\psi_\pm(m) = \frac{1}{\sqrt{2}}(\alpha^1{}_{m+1} \pm i\alpha^2{}_{m+1}).$$

(In the case $m = 2$ we see that these operators are the well-known ladder operators $L_\pm$ appearing in the theory of angular momenta.) These operators may be regarded as components of the O(m) vector operator

$$\begin{aligned}\psi(m)^1 &= \psi_+(m), \\ \psi(m)^2 &= \psi_-(m), \\ \psi(m)^i &= \alpha^i{}_{m+1}, \quad i = 3,...,m,\end{aligned}$$

which is obtainable from the O(m) vector operator $(\alpha^i{}_{m+1})$ $(i = 1,...,m)$ by considering the change of basis transformation

$$e_1 \to \frac{1}{\sqrt{2}}(e_1 - ie_2),$$
$$e_2 \to \frac{1}{\sqrt{2}}(e_1 + ie_2),$$
$$e_i \to e_i, \quad i = 3,\ldots,m.$$

The associated change of basis matrix is given by the $m \times m$ block matrix

$$M = M_2 \oplus I_{m-2},$$

where I_{m-2} is the $(m-2)\times(m-2)$ identity matrix and

$$M_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}.$$

Clearly the O(m) vector $\psi^i = (M^{-1})^i_j \alpha^j{}_{m+1}$ has components in their O(2) weight space forms. Similarly the entries of the O(m) matrix

$$a = M\alpha M^{-1}$$

consist of O(m) generators in their O(2) weight space forms. We may now proceed as we did before working with the matrix a instead of the matrix α. The modifications required are straightforward and the analysis proceeds in exactly the same way as before.

The result is that the operators $\psi_{\pm}(m)$ and their adjoints $\psi^{\dagger}_{\pm}(m) = (1/\sqrt{2})(\alpha_1^{m+1} \mp i\alpha_2^{m+1})$ may be resolved into simultaneous shift components

$$\psi_{\pm}(m) = \sum_{i(\underline{k})} \psi\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & 3(i) & \pm \end{pmatrix},$$
$$\psi^{\dagger}_{\pm}(m) = \sum_{i(\underline{k})} \psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & \pm \end{pmatrix},$$

where the simultaneous shift components simultaneously alter the representation labels of the subgroups O(m),···O(3) as before and also alter the representation labels of O(2) according to

$$\lambda_{1,2}\psi\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & \pm \end{pmatrix} = \psi\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & \pm \end{pmatrix}(\lambda_{1,2} \pm 1),$$
$$\lambda_{1,2}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & \pm \end{pmatrix} = \psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & \pm \end{pmatrix}(\lambda_{1,2} \mp 1).$$

Note that ψ_+ can only increase the representation label of O(2) while ψ_- can only decrease the representation label of O(2) (and similarly for $\psi^{\dagger}_{\mp}$).

In the case of O(2) all Wigner coefficients are unity and the result corresponding to Eq. (27) for the simultaneous shift components of the operators $\psi_{\pm}(m)$ is

$$\psi\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & + \end{pmatrix}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & + \end{pmatrix}$$
$$= \frac{M_{1,2}}{\bar{M}_{2,2}}\psi\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & - \end{pmatrix}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & - \end{pmatrix}$$
$$= \bar{M}_{1,2}\prod_{r=3}^{m} \bar{M}_{i(r),r} C_{i(r),r}$$
$$\times(\alpha_{i(r),r} - \alpha_{i(r-1),r-1} - \eta_{i(r-1),r-1})^{-1}$$
$$\times(\alpha_{i(r),r} - \alpha_{i(r-1),r-1})^{-1}$$
$$\times(\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - \eta_{i(r-1),r-1})^{-1}$$
$$\times(\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - 2\eta_{i(r-1),r-1}), \tag{28}$$

where $\bar{M}_{1,2}$ and $\bar{M}_{2,2}$ are given by Eq. (23).

Expressing the generators $\alpha^{m+1}{}_1$ and $\alpha^{m+1}{}_2$ in terms of $\psi^{\dagger}_{\pm}(m)$ we have

$$\alpha^{m+1}{}_1 = \frac{1}{\sqrt{2}}(\psi^{\dagger}_{+}(m) + \psi^{\dagger}_{-}(m)),$$
$$\alpha^{m+1}{}_2 = \frac{i}{\sqrt{2}}(\psi^{\dagger}_{+}(m) - \psi^{\dagger}_{-}(m)). \tag{29}$$

From this it follows, in our previous notation, that the generators $\alpha^{m+1}{}_1$ and $\alpha^{m+1}{}_2$ may be resolved into a sum of shift components

$$\alpha^{m+1}{}_i = \sum_{i(\underline{k})} \psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & i(2) \end{pmatrix}_i, \quad i = 1,2,$$

where the shift components alter the O(2) representation label according to

$$\lambda_{1,2}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 1 \end{pmatrix} = \psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 1 \end{pmatrix}(\lambda_{1,2} - 1),$$
$$\lambda_{1,2}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix} = \psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix}(\lambda_{1,2} + 1).$$

These shift components, in view of Eq. (29), are given by

$$\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 1 \end{pmatrix}_1 = \frac{1}{\sqrt{2}}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & + \end{pmatrix},$$
$$\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix}_1 = \frac{1}{\sqrt{2}}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & - \end{pmatrix},$$
$$\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 1 \end{pmatrix}_2 = \frac{i}{\sqrt{2}}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & + \end{pmatrix},$$
$$\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix}_2 = -\frac{i}{\sqrt{2}}\psi^{\dagger}\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & - \end{pmatrix}. \tag{30}$$

V. MATRIX ELEMENTS OF THE O(n) GENERATORS

Throughout this section we assume that we are working in a finite-dimensional irreducible module over O(n) which admits a Gel'fand–Zetlin basis. Our aim is to evaluate the matrix elements of the generators $\alpha^{m+1}{}_l$ ($l \leqslant m$).

For ease of notation we write an arbitrary Gel'fand–Zetlin basis state in the form $|\lambda_{j,k}\rangle$ ($k = 2,\ldots,n; j = 1,\ldots,[k/2]$) where the $\lambda_{j,k}$ satisfy the well-known Gel'fand betweenness conditions. In the special case $l = m \geqslant 3$ we see that the matrix elements of the generator $\alpha^{m+1}{}_m$ are given by

$$\alpha^{m+1}{}_m\,|\,\lambda_{j,k}\rangle = \sum_{r=1}^{m} \psi^{\dagger}\begin{pmatrix} m \\ r \end{pmatrix}_m |\,\lambda_{j,k}\rangle$$
$$= \sum_{r=1}^{m} N^m{}_r\,|\,\lambda_{j,k} - \Delta_{r,m}\rangle,$$

where $|\,\lambda_{j,k} - \Delta_{r,m}\rangle$ denotes the state obtained from $|\lambda_{j,k}\rangle$ by adding to the row $(\lambda_{1,m},\ldots,\lambda_{h,m})$ $\{h = [m/2]\}$, corresponding to the representation labels of O(m), the O(m) weight $-\Delta_{r,m}$. Here the weight $\Delta_{r,m}$ is defined for $r = 1,\ldots,m$ in accordance with Eq. (7).

The matrix elements $N^m{}_r(\lambda_{j,m+1};\lambda_{j,m};\lambda_{j,m-1})$ are given by (up to a phase)

$$|N^m{}_r| = \langle \lambda_{j,k} | \bar{M}_{r,m} C_{r,m} | \lambda_{j,k} \rangle^{1/2}$$

which may be evaluated using Eqs. (20) and (23). Hence for the case $m = 2h$ we obtain

$$|N^m{}_r| = \left(\frac{(-1)^m \Pi_{k=1}^{m+1}(\lambda_{k,m+1} - \lambda_{r,m} + r - p + 1)\Pi_{l=1}^{m-1}(\lambda_{r,m} - \lambda_{l,m-1} + l - r + \delta_{l,h})}{\Pi_{l \neq r}(\lambda_{r,m} - \lambda_{l,m} + l - r - 1 - \delta_{l,m+1-r})(\lambda_{r,m} - \lambda_{l,m} + l - r)} \right)^{1/2} \tag{31}$$

and in the case $m = 2h + 1$ we obtain

$$|N^m{}_r| = \left(\frac{(-1)^m \Pi_{k=1}^{m+1}(\lambda_{k,m+1} - \lambda_{r,m} + r - p + 1)\Pi_{l=1}^{m-1}(\lambda_{r,m} - \lambda_{l,m-1} + l - r)}{\Pi_{l \neq r}(\lambda_{r,m} - \lambda_{l,m} + l - r - 1 + \delta_{l,h+1} - \delta_{l,m+1-r})(\lambda_{r,m} - \lambda_{l,m} + l - r)} \right)^{1/2}, \tag{32}$$

where we have adopted convention (8) for defining labels $\lambda_{r,m}$ for $r > h = [m/2]$. The phases for the above matrix elements will be determined in the next section.

Returning to the general case we may resolve the generator $\alpha^{m+1}{}_l$ $(3 \leqslant l \leqslant m)$ into its simultaneous shift components to give

$$\begin{aligned}
&\alpha^{m+1}{}_l | \lambda_{j,k} \rangle \\
&\quad = \sum_{i(k)} \psi^\dagger \begin{pmatrix} m & & l \\ & \cdots & \\ i(m) & & i(l) \end{pmatrix}_l | \lambda_{j,k} \rangle \\
&\quad = \sum_{i(k)} N \begin{pmatrix} m & & l \\ & \cdots & \\ i(m) & & i(l) \end{pmatrix} | \lambda_{j,k} - \Delta_{i(m),m} - \cdots - \Delta_{i(l),l} \rangle .
\end{aligned}$$

The matrix elements in this case are given by (up to a phase)

$$\begin{aligned}
&\left| N \begin{pmatrix} m & & l \\ & \cdots & \\ i(m) & & i(l) \end{pmatrix} \right| \\
&\quad = \langle \lambda_{j,k} | \psi \begin{pmatrix} m & & l \\ & \cdots & \\ i(m) & & i(l) \end{pmatrix}^l \psi^\dagger \begin{pmatrix} m & & l \\ & \cdots & \\ i(m) & & i(l) \end{pmatrix}_l | \lambda_{j,k} \rangle^{1/2}
\end{aligned}$$

which, by virtue of Eq. (27), is equal to

$$\begin{aligned}
&\prod_{r=l}^{m} |N^r{}_{i(r)}| \prod_{r=l+1}^{m} [(\alpha_{i(r),r} - \alpha_{i(r-1),r-1} - \eta_{i(r-1),(r-1)})^{-1} \\
&\quad \times (\alpha_{i(r),r} - \alpha_{i(r-1),r-1})^{-1} \\
&\quad \times (\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - \eta_{i(r-1),(r-1)})^{-1} \\
&\quad \times (\alpha_{i(r),r} - \alpha_{r-i(r-1),r-1} - 2\eta_{i(r-1),r-1})]^{1/2},
\end{aligned} \tag{33}$$

where $|N^r{}_{i(r)}|$ are the phase-free matrix elements of the generator α_r^{r+1} which are given by Eqs. (31) and (32).

It is clear that $N\binom{m \cdots l}{i(m) \cdots i(l)}$ corresponds to the generator matrix element [suppressing the labels of the subgroups $O(m+2)$]

$$\left\langle \begin{matrix} \lambda_{j,m+1} \\ (\lambda') \end{matrix} \middle| \alpha^{m+1}{}_l \middle| \begin{matrix} \lambda_{j,m+1} \\ (\lambda) \end{matrix} \right\rangle, \quad l \geqslant 3$$

where (λ') is obtained from the pattern (λ) by adding to the first $(m-1)$ rows of the pattern (λ), corresponding to the representation labels of the respective subgroups $O(m),\ldots,O(l)$, the weights $-\Delta_{i(m)}, -\Delta_{i(m-1)}, \cdots -\Delta_{i(l)}$, respectively.

In the case of the generator α_2^{m+1} we have a resolution into simultaneous shift components which are defined by Eq. (30). In this case we may write

$$\begin{aligned}
&\alpha_2^{m+1} | \lambda_{j,k} \rangle \\
&\quad = \sum_{i(k)} \psi^\dagger \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & i(2) \end{pmatrix}_2 | \lambda_{j,k} \rangle \\
&\quad = \sum_{i(k)} N_2 \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & i(2) \end{pmatrix} | \lambda_{j,k} - \Delta_{i(m),m} - \cdots - \Delta_{i(2),2} \rangle ,
\end{aligned}$$

where we have adopted the same convention as before [keeping in mind that $\Delta_1 = 1 = -\Delta_2$ for O(2)]. We have found it convenient in this case to add a subscript 2 to the matrix element to indicate that we are considering the generator $\alpha^{m+1}{}_2$ rather than $\alpha^{m+1}{}_1$. (In the other cases this ambiguity does not arise).

Our phase-free matrix elements in this case are given by [see Eq. (30)]

$$\begin{aligned}
&\left| N_2 \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & 1 \end{pmatrix} \right| \\
&\quad = \frac{1}{\sqrt{2}} \left\langle \lambda_{j,k} \left| \psi \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & + \end{pmatrix} \psi^\dagger \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & + \end{pmatrix} \right| \lambda_{j,k} \right\rangle^{1/2}, \\
&\left| N_2 \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & 2 \end{pmatrix} \right| \\
&\quad = \frac{1}{\sqrt{2}} \left\langle \lambda_{j,k} \left| \psi \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & - \end{pmatrix} \psi^\dagger \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & - \end{pmatrix} \right| \lambda_{j,k} \right\rangle^{1/2}.
\end{aligned}$$

Using Eq. (28) we therefore have, in terms of the matrix elements of the generator $\alpha^{m+1}{}_3$ [see Eq. (33)],

$$\begin{aligned}
&\left| N_2 \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & 1 \end{pmatrix} \right| \\
&\quad = \frac{1}{\sqrt{2}} (\bar{M}_{1,2})^{1/2} \left| N \begin{pmatrix} m & & 3 \\ & \cdots & \\ i(m) & & i(3) \end{pmatrix} \right| [(\alpha_{i(3),3} - \alpha_{1,2} - 1)^{-1} \\
&\qquad \times (\alpha_{i(3),3} - \alpha_{1,2})^{-1} \\
&\qquad \times (\alpha_{i(3),3} - \alpha_{2,2} - 1)^{-1} (\alpha_{i(3),3} - \alpha_{2,2} - 2)]^{1/2} \\
&\quad = \left(\frac{\bar{M}_{1,2}}{\bar{M}_{2,2}} \right)^{1/2} \left| N_2 \begin{pmatrix} m & & 2 \\ & \cdots & \\ i(m) & & 2 \end{pmatrix} \right| .
\end{aligned}$$

The phase-free matrix elements $|N_1\binom{m \cdots 2}{i(m) \cdots i(2)}|$ of the generator $\alpha^{m+1}{}_1$ are given by the same expressions. From the commutation relation $\alpha^{m+1}{}_1 = [\alpha^{m+1}{}_2, \alpha^2{}_1]$ we see that the full matrix elements of the generators $\alpha^{m+1}{}_1$ and $\alpha^{m+1}{}_2$ are related by

$$N_1 \begin{pmatrix} m & & 3 & 2 \\ & \cdots & & \\ i(m) & & i(3) & 1 \end{pmatrix} = -iN_2 \begin{pmatrix} m & & 3 & 2 \\ & \cdots & & \\ i(m) & & i(3) & 1 \end{pmatrix},$$

and

$$N_1\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix} = +iN_2\begin{pmatrix} m & \cdots & 3 & 2 \\ i(m) & \cdots & i(3) & 2 \end{pmatrix}. \tag{34}$$

(This occurs since we are labeling our weights by the eigenvalues of $-i\alpha^1{}_2$; i.e., we are adopting the phase $-i$ for the generator $\alpha^2{}_1$).

VI. CHOICE OF PHASES

Unlike the U(n) case we cannot simply choose the phases of the generators $\alpha^{m+1}{}_m$ to be $(+1)$. This is because on unitary representations of the group the generators $\alpha^k{}_l$ are to be represented by anti-Hermitian matrices. Now the Cartan generators $\alpha^{2r}{}_{2r-1}$ have diagonal entries (in fact these are the only generators with diagonal matrix elements) and hence, by the anti-Hermitian property, these must be pure imaginary. Hence we adopt the phase $(-i)$ for the diagonal matrix elements. Although the remaining matrix elements of the Cartan generators can be chosen to be real we feel it is simpler to adopt the phase convention $(-i)$ for all matrix elements of the Cartan generators. In keeping with the unity of this choice of phase factor we adopt the phase $(-i)$ for the matrix elements of all generators of the form $\alpha^{m+1}{}_m$ (although one may equally well choose real phases for the matrix elements of the generators $\alpha^{2r+1}{}_{2r}$). It is easily checked that the choice $(-i)$ for the phases of the generators $\alpha^{m+1}{}_m$ is consistent with Hermiticity requirements.

The phases of the matrix elements of the remaining generators are now dictated by the Lie algebra commutation relations. It follows from these considerations that the phase-free matrix elements of the generator $\alpha^{m+1}{}_l$ ($l \geqslant 2$) are to be multiplied by a complex phase $(-i)^{m+1-l}$ and an additional real phase

$$S(i(m-1)-i(m))S(i(m-2)-i(m-1)) \cdots S(i(l)-i(l+1)),$$

where

$$S(x) = \text{sign}(x),$$

and

$$S(0) = 1$$

[c.f., U(n) case]. Thus for $m \geqslant l \geqslant 2$ we have

$$N\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} = (-i)^{m+1-l}\prod_{r=l}^{m-1} S(i(r)-i(r+1))\left|N\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix}\right|.$$

In the case of the generator $\alpha^{m+1}{}_1$ the matrix elements are obtained from those of $\alpha^{m+1}{}_2$ (whose phases are determined by the above conventions) by Eqs. (34).

As an example to clarify our notation and our choice of phase convention the matrix elements of the O(4) generators are written down in Appendix B. One sees that the absolute value of the matrix elements agrees with the results obtained in Ref. 12 but the phase conventions are different.

VII. ANALYSIS OF RESULTS

We have shown that the only nonvanishing matrix elements of the generator $\alpha^{m+1}{}_l$ are of the form [suppressing the labels of O($m+1$)]

$$\left\langle \begin{matrix} \lambda' \\ (\mu') \end{matrix} \middle| \alpha^{m+1}{}_l \middle| \begin{matrix} \lambda \\ (\mu) \end{matrix} \right\rangle,$$

where (μ') is obtained from the pattern (μ) by adding to the first $(m-l)$ rows of the pattern (μ), corresponding to the representation labels of the subgroups O(m),...,O(l), the weights $-\Delta_{i(m),m}, -\Delta_{i(m-1),m-1}, \cdots -\Delta_{i(l),l}$, respectively. The corresponding matrix element is given by formulas (31)–(34) (with the phase convention adopted in the previous section). From the Wigner–Eckart theorem it follows also that this matrix element may be written

$$\langle \lambda - \Delta_{i(m)} \| \psi(m) \| \lambda \rangle \left\langle \begin{matrix} \lambda \\ (\mu) \end{matrix}; \begin{matrix} \overline{10} \\ l \end{matrix} \middle| \begin{matrix} \lambda - \Delta_{i(m)} \\ (\mu') \end{matrix} \right\rangle, \tag{35}$$

where the first term is the O(m) reduced matrix element $(\bar{M}_{i(m),m})^{1/2}$. Comparing this with the formulas obtained for the matrix elements one may obtain all fundamental Wigner coefficients of O(n).

Also, by comparing Eq. (35) with Eq. (26), one sees that the general fundamental Wigner coefficient may be expressed algebraically, for the case $l > 2$, by

$$\left|\left\langle \begin{matrix} \lambda \\ (\mu) \end{matrix}; \begin{matrix} \overline{10} \\ l \end{matrix} \middle| \begin{matrix} \lambda - \Delta_{i(m)} \\ (\mu') \end{matrix} \right\rangle\right|^2 = \left\langle \begin{matrix} \lambda \\ (\mu) \end{matrix} \middle| P\begin{pmatrix} l & \cdots & m-1 & m & m-1 & \cdots & l \\ i(l) & \cdots & i(m-1) & i(m) & i(m-1) & \cdots & i(l) \end{pmatrix} \middle| \begin{matrix} l & \lambda \\ l & (\mu) \end{matrix} \right\rangle$$

in direct analogy with the U(n) case.

Finally, in the notation of Baird and Biedenharn,[8] let us denote the O(r) Wigner coefficient $(-i)(\bar{C}_{i(r),r})^{1/2}$ by $\begin{pmatrix} i(r):r \\ r-1 \end{pmatrix}$, the reduced matrix element $(M_{i(r),r})^{1/2}$ by $\begin{pmatrix} r+1 \\ i(r):r \end{pmatrix}$ and the corresponding reduced Wigner coefficient $\alpha\begin{pmatrix} r & r-1 \\ i(r) & i(r-1) \end{pmatrix}$ by $\begin{pmatrix} i(r) & r \\ i(r-1) & r-1 \end{pmatrix}$. Then the matrix elements may be expressed, for $l > 2$,

$$N\begin{pmatrix} m & \cdots & l \\ i(m) & \cdots & i(l) \end{pmatrix} = \begin{pmatrix} m+1 \\ i(m):m \end{pmatrix}\prod_{r=l+1}^{m} S(i(r-1)-i(r)) \times \begin{pmatrix} i(r):r \\ i(r-1):r-1 \end{pmatrix}\begin{pmatrix} i(l):l \\ l-1 \end{pmatrix}$$

in direct analogy with the U(n) case. Trivial modifications are required to extend this result to the case where $l = 1,2$.

APPENDIX A

Here we give a brief derivation of the results presented in Sec. III (see Ref. 24 for a more detailed account). We adopt the same notation as that used in Sec. III.

From the characteristic identity of the O($n+1$) matrix β we have $\beta Q[k] = \beta_k Q[k]$. Taking the $(i,n+1)$ entry of this matrix equation and rearranging we have

$$\alpha^i{}_{n+1} C_k = (\beta_k - \alpha)^i{}_j Q[k]^j{}_{n+1},$$

where C_k denotes the O(n) invariant $Q[k]^{n+1}{}_{n+1}$ and where we have summed over repeated index j from 1 to n. Inverting this equation we have

$$Q[k]^i{}_{n+1} = [(\beta_k - \alpha)^{-1}]^i{}_j \alpha^j{}_{n+1} C_k,$$

where $(\beta_k - \alpha)^{-1}$ denotes the matrix

$$(\beta_k - \alpha)^{-1} = \sum_{r=1}^{n} (\beta_k - \alpha_r)^{-1} P[r].$$

Resolving the O(n) vector $\psi^i = \alpha^i{}_{n+1}$ ($i = 1,...,n$) into its shift components we then have

$$Q[k]^i{}_{n+1} = \sum_{r=1}^{n} (\beta_k - \alpha_r)^{-1} \psi[r]^i C_k . \tag{A1}$$

When $r \leqslant h = [n/2]$ we may write

$$(\beta_k - \alpha_r)^{-1}\psi[r] = \psi[r](\beta_k - \alpha_r - 1)^{-1},$$

$$(\beta_k - \alpha_{n+1-r})^{-1}\psi[n+1-r]$$

$$= \psi[n+1-r](\beta_k - \alpha_{n+1-r} - 1)^{-1}.$$

However for $n = 2h + 1$ we have the zero shift component which satisfies

$$(\beta_k - \alpha_{h+1})^{-1}\psi[h+1] = \psi[h+1](\beta_k - \alpha_{h+1})^{-1}.$$

We combine these relations into the more compact form

$$(\beta_k - \alpha_r)^{-1}\psi[r] = \psi[r](\beta_k - \alpha_r - \eta_r)^{-1},$$

where $\eta_r = \eta_{n+1-r} = 1$ for $r \leqslant h$, with $\eta_{h+1} = 0$ for $n = 2h + 1$. Thus we may write Eq. (A1) in the form

$$Q[k]^i{}_{n+1} = \sum_{r=1}^{n} \psi[r]^i (\beta_k - \alpha_r - \eta_r)^{-1} C_k . \tag{A2}$$

Similarly we have

$$Q[k]^{n+1}{}_i = \sum_{r=1}^{n} (\beta_k - \alpha_r - \eta_r)^{-1} C_k \psi^\dagger[r]_i . \tag{A3}$$

Summing these equations over k from 1 to $n + 1$ one sees that the C_k satisfy

$$\sum_{k=1}^{n+1} (\beta_k - \alpha_r - \eta_r)^{-1} C_k = 0, \quad r = 1,...,n.$$

These equations together with the condition

$$\sum_{k=1}^{n+1} C_k = 1$$

uniquely determines the C_k. Using Lagrange-interpolation techniques (or else invert the equations using matrix methods) one sees that the unique solutions to these equations are given by

$$C_k = \prod_{l \neq k} (\beta_k - \beta_l)^{-1} \prod_{r=1}^{n} (\beta_k - \alpha_r - \eta_r). \tag{A4}$$

Similarly we have a corresponding expression for $\bar{C}_k$.

We now evaluate the O(n) invariants $\psi^\dagger[r]_i \psi[r]^i$. We first invert Eq. (A2) to get $\psi[r]^i$ in the form $P(\beta)^i{}_{n+1}$ for a suitable polynomial $p(x)$. We may then evaluate our invariants by a simple application of the O($n + 1$) characteristic identity. This leads us to consider the solutions γ_{lk} to the equations

$$\sum_{k=1}^{n+1} \gamma_{lk} C_k (\beta_k - \alpha_r - \eta_r)^{-1} = \delta_{rl}, \quad l = 1,...,n,$$

and

$$\sum_{k=1}^{n+1} \gamma_{lk} C_k = 0.$$

Then, for each $r = 1,...,n$, we have $(n + 1)$ equations in $(n + 1)$ unknowns γ_{rk} ($k = 1,...,n + 1$) which yield the unique solution

$$\gamma_{rk} = \gamma_r (\beta_k - \alpha_r - \eta_r)^{-1},$$

where

$$\gamma_r = (-1)^n \prod_{p=1}^{n+1} (\beta_p - \alpha_r - \eta_r) \prod_{l \neq r} (\alpha_r - \alpha_l)^{-1}.$$

Thus multiplying Eq. (A2) by γ_{rk} and summing over k we obtain

$$\psi^{[r]} = \sum_{k=1}^{n+1} Q[k]^i{}_{n+1} \gamma_r (\beta_k - \alpha_r - \eta_r)^{-1}. \tag{A5}$$

Multiplying on the left by $\alpha^{n+1}{}_i$ and summing on i from 1 to n we obtain, using $\alpha^{n+1}{}_i Q[k]^i{}_{n+1} = \beta_k C_k$,

$$\psi^\dagger[r]_i \psi[r]^i = \alpha^{n+1}{}_i \psi[r]^i = \gamma_r .$$

Using the O(n) commutation relations one may deduce the result (see Refs. 21 and 24)

$$Q[k]^i{}_j = Q[k]^i{}_{n+1} (C_k)^{-1} Q[k]^{n+1}{}_j. \tag{A6}$$

From this and Eq. (A4) we have the results

$$\psi[r]^i (\gamma_r)^{-1} \psi^\dagger[r]_j = P[r]^i{}_j ,$$

which may be rearranged to give

$$\psi[r]^i \psi^\dagger[r]_j = \bar{M}_r P[r]^i{}_j , \tag{A7}$$

which is Eq. (14). Finally applying the O(n) projector to both sides of Eq. (A6) we obtain, using Eqs. (A2) and (A3),

$$P[r]Q[k]P[r]^i{}_j = \psi[r]^i (\beta_k - \alpha_r - \eta_r)^{-2} C_k \psi^\dagger[r]_j .$$

From the form of C_k given by Eq. (A4) we then obtain, in view of Eq. (A7),

$$P[r]Q[k]P[r]$$

$$= C_k \bar{M}_r (\beta_k - \alpha_r - \eta_r)^{-1} (\beta_k - \alpha_r)^{-1}$$
$$\times (\beta_k - \alpha_{n+1-r} - \eta_r)^{-1} (\beta_k - \alpha_{n+1-r} - 2\eta_r)$$
$$\times P[r].$$

More details are given in Ref. 24.

APPENDIX B

We consider here the action of the O(4) generators on an arbitrary Gel'fand basis state

$$\begin{vmatrix} m_1 & m_2 \\ l & \\ m & \end{vmatrix}\rangle .$$

Before proceeding we note that acting on this state the O(4) roots $\alpha_{i,4}$ are given by $\alpha_{14} = m_1 + 2$, $\alpha_{24} = m_2 + 1$, $\alpha_{34} = 1 - m_2$, $\alpha_{44} = -m_1$ while the O(3) roots are given by $\alpha_{13} = l + 1$, $\alpha_{23} = 1$, $\alpha_{33} = -l$. Finally the O(2) roots are

$\alpha_{12} = m = -\alpha_{22}$.

The matrix elements of the O(3) generators are, according to our prescription, given by

$$\alpha^2{}_1 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle = -im \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle,$$

$$\alpha^3{}_2 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle = -\frac{i}{2}[(l-m)(l+m+1)]^{1/2} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m+1 & \end{matrix} \right\rangle - \frac{i}{2}[(l+m)(l-m+1)]^{1/2} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m-1 & \end{matrix} \right\rangle,$$

$$\alpha^3{}_1 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle = \tfrac{1}{2}[(l-m)(l+m+1)]^{1/2} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m+1 & \end{matrix} \right\rangle - \tfrac{1}{2}[(l+m)(l-m+1)]^{1/2} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m-1 & \end{matrix} \right\rangle.$$

The matrix elements of the generator $\alpha^4{}_3$, in view of our phase convention and Eq. (32), are given by

$$\alpha^4{}_3 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle = N^3{}_1 \left| \begin{matrix} m_1 & m_2 \\ l-1 & \\ m & \end{matrix} \right\rangle + N^3{}_2 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle + N^3{}_3 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m-1 & \end{matrix} \right\rangle,$$

where

$$N^3{}_1 = -i\left[\frac{(l+m)(l-m)(m_1-l+1)(l+m_2)(l-m_2)(l+m_1+1)}{l^2(2l+1)(2l-1)} \right]^{1/2},$$

$$N^3{}_2 = -i\left[\frac{mm_2(m_1+1)}{l(l+1)} \right],$$

$$N^3{}_3 = -i\left[\frac{(l+m+1)(l-m+1)(m_1-l)(l+m_1+2)(l+m_2+1)(l+1-m_2)}{(2l+1)(2l+3)(l+1)^2} \right]^{1/2}.$$

Similarly for the generator $\alpha^4{}_2$ we have

$$\alpha^4{}_2 \left| \begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix} \right\rangle = N_2\begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l-1 & \\ m-1 & \end{matrix} \right\rangle + N_2\begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m-1 & \end{matrix} \right\rangle + N_2\begin{pmatrix} 3 & 2 \\ 3 & 1 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l+1 & \\ m-1 & \end{matrix} \right\rangle + N_2\begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l-1 & \\ m+1 & \end{matrix} \right\rangle + N_2\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l & \\ m+1 & \end{matrix} \right\rangle + N_2\begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix} \left| \begin{matrix} m_1 & m_2 \\ l+1 & \\ m+1 & \end{matrix} \right\rangle.$$

The matrix elements in this case are given by

$$N_2\begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = -\frac{1}{2}\left[\frac{(l+m)(m_1-l+1)(l-m_2)(l+m_2)(l+m_1+1)(l+m-1)}{l^2(2l-1)(2l+1)} \right]^{1/2},$$

$$N_2\begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} = +\frac{1}{2}\left[\frac{(l+1-m)(l+m)m_2{}^2(m_1+1)^2}{l^2(l+1)^2} \right]^{1/2},$$

$$N_2\begin{pmatrix} 3 & 2 \\ 3 & 1 \end{pmatrix} = +\frac{1}{2}\left[\frac{(l-m+1)(l+m_1+2)(l+m_2+1)(l+1-m_2)(m_1-l)(l-m+2)}{(2l+3)(2l+1)(l+1)^2} \right]^{1/2},$$

$$N_2\begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix} = -\frac{1}{2}\left[\frac{(l-m)(m_1-l+1)(l-m_2)(l+m_2)(l+m_1+1)(l-m-1)}{l^2(2l-1)(2l+1)} \right]^{1/2},$$

$$N_2\begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} = -\frac{1}{2}\left[\frac{(l+1+m)(l-m)m_2{}^2(m_1+1)^2}{l^2(l+1)^2} \right]^{1/2},$$

$$N_2\begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix} = +\frac{1}{2}\left[\frac{(l+m+1)(l+m_1+2)(l+m_2+1)(l+1-m_2)(m_1-l)(l+m+2)}{(2l+3)(2l+1)(l+1)^2} \right]^{1/2}.$$

In a similar way we define the action of the generator $\alpha^4{}_1$ on the state

$$\left|\begin{matrix} m_1 & m_2 \\ l & \\ m & \end{matrix}\right\rangle.$$

The matrix elements in this case are related to those of the generator $\alpha^4{}_2$ by Eq. (34); i.e.,

$$N_1\begin{pmatrix} 3 & 2 \\ j & 1 \end{pmatrix} = -iN_2\begin{pmatrix} 3 & 2 \\ j & 1 \end{pmatrix}$$

$$N_1\begin{pmatrix} 3 & 2 \\ j & 2 \end{pmatrix} = iN_2\begin{pmatrix} 3 & 2 \\ j & 1 \end{pmatrix} \qquad j = 1,2,3.$$

One may check directly that the O(4) commutation relations are satisfied with these matrix elements.

ACKNOWLEDGMENTS

The author thanks Professor H. S. Green for his encouragement. The author is also grateful for the financial support of an A.R.G.C. research grant. Part of this work was done under the financial support of a Commonwealth postgraduate research award.

[1]H. Casimir, Proc. Koninkl. Akad. (Amsterdam) **34**, 844 (1931).
[2]H. Casimir and B. L. Van der Waerden, Math. Ann. **111**, 1 (1935).
[3]G. Racah, Rend. Atti. Accad. Naz. Lincei. **8**, 108 (1950).
[4]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen and Co., London, 1931); *The Classical Groups* (Princeton U. P., Princeton, N.J., 1939).
[5]J. Schwinger, "On Angular Momentum", preprinted in L.C. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[6]V. Bargmann, Commun Pure Appl. Math. **14**, 198 (1961). K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience, New York, 1953). M. Moshinsky, Nucl. Phys. **31**, 384 (1962). M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).
[7]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).
[8]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[9]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968); J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970); L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967); G. E. Baird and L. C. Biedenharn, J. Math. Phys. **12**, 1723 (1964); **5**, 1730 (1964); **6**, 1845 (1965).
[10]M. K. F. Wong, J. Math. Phys. **19**, 703 (1978); **17**, 1558 (1976); W. J. Holman III, Nuovo Cimento A **4**, 904 (1971); T. Maekawa, J. Math. Phys. **16**, 334 (1975). (For a more exhaustive list of references see above-mentioned papers.)
[11]M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971); M. K. F. Wong, J. Math. Phys. **10**, 1065 (1969); **12**, 1530 (1971).
[12]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (MacMillan, New York, 1963).
[13]J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965). Rev. Mex. Fis. **14**, 29 (1965).
[14]S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[15]M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[16]A. M. Bincer, J. Math. Phys. **18**, 1870 (1977); **19**, 1173 (1978); **19**, 1179 (1978).
[17]M. D. Gould, J. Math. Phys. **21**, 444 (1980); "General U(n) raising and lowering operators," Flinders University preprint (to appear in J. Math. Phys.).
[18]M. D. Gould, "On the matrix elements of the U(n) generators," Flinders University preprint (to appear in J. Math. Phys.).
[19]B. Kostant, J. Func. Anal. **20**, 257 (1975).
[20]H. S. Green, J. Math. Phys. **12**, 2106 (1971); A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).
[21]H. S. Green, J. Aust. Math. Soc. Ser. B, **19**, 129 (1972).
[22]S. O'kubo, J. Math. Phys. **16**, 528 (1975); **12**, 2382 (1977). K. C. Hannabuss, "Characteristic Equations for semisimple Lie Groups," preprint, Math. Inst. Oxford, 1972 (unpublished). D. M. O'Brien, A. Cant, and A. L. Carey, Ann. Inst. Henri Poincaré, Sec. A, 405 (1977).
[23]M. D. Gould, Ann. Inst. Henri Poincaré, Sec. A, 203 (1980).
[24]M. D. Gould, Ph.D. Thesis, 1979 (unpublished); J. Aust. Math. Soc. Ser. B **20**, 401 (1978).

Classical and quantal systems of imprimitivity

N. Giovannini
Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland

(Received 23 April 1981; accepted for publication 5 June 1981)

We discuss from a group-theoretical point of view, a simple framework in which the classical and the quantal state spaces appear in a unified way. The framework is characterized by the use of (possibly continuous) direct unions of Hilbert spaces. Symmetries are correspondingly represented by families of (anti-) unitary operators. We consider here the associated notions of projective representations as well as the corresponding observables. These observables, classical or quantal, are defined in terms of a slight generalization of the notion of a system of imprimitivity of Mackey.

PACS numbers: 02.20. + b, 03.20. + i, 03.65.Bz

INTRODUCTION

There exists a simple way for describing the classical and the quantal state spaces of an elementary physical system in a single mathematical framework. It is to consider (topological) direct unions of Hilbert spaces. Classical physics, where a state is usually given by a point in the phase space, corresponds then to the extreme case where each Hilbert space is of dimension one and the union is precisely taken over this phase space. Usual quantum physics, on the other hand, corresponds to the other extreme case, where the union is trivial and the space is thus a single Hilbert space. More generally and in the intermediate cases, such a framework is able to describe situations where parts of a system have a quantal behavior, whereas other aspects are of the classical type. Think for example of superselection sectors, or of a spin-spin interaction in a crystal (with fixed positions and momenta) or of a measurement procedure of a quantal system with a classical device.

On the other hand, the fact that we can compare different models (classical with quantal and, as we shall see, relativistic with nonrelativistic in particular) in a unified mathematical language is of course of great importance for the understanding of the results and of the difficulties encountered in each separate context.

In a previous work,[1] we have shown that, conversely, one can derive the classical and quantal state space, for a spinless particle, in a unified way, when considering these direct unions as representation spaces for the corresponding kinematical symmetry groups. More precisely we considered those representations for which there exist sufficiently faithful operators for each observable which corresponds to the system. This construction is based on the following briefly sketched points of view.

The state of a physical system is specified by the possible outcomes of the measurements that we can choose to perform on it. To these measurements correspond the observables, and to the outcomes correspond (subsets of) the spectra of these observables. Let us consider the space Γ of all spectra of the observables that correspond to some given system. A typical example of such a space Γ is the phase space, with the observables position and momentum. On this space Γ, the usual physical equivalence principles correspond to a natural action which forms the defining representation of a group. We have called the corresponding groups kinematical in order to emphasize that they are independent by definition of the dynamics and of the interactions, as the latter neither changes the representation nor the interpretation of the observables. It is only in a second step, when the state spaces have been specified, that it will be possible to consider the corresponding dynamics. This distinction is essential if we want to build up theories able to describe more than only free particles.

One of the main results of Ref. 1 was that both in the relativistic and nonrelativistic contexts there are two and only two solutions for the corresponding state spaces: the classical and quantal ones. The framework provides thus also by the way a new approach to quantization.

In order to generalize these models, in particular if we want to consider particles with an arbitrary spin or for the discussion of the nonrelativistic limit of the relativistic case we, however, need to consider a more general framework where arbitrary projective representations on such direct union spaces are considered. A first step in this direction was the solution of the corresponding cohomological problem (families of phase factor systems). This solution has been given in Ref. 2.

In the present paper we continue to analyze the mathematical basis of our approach, and we give the general solution for the state spaces and for the associated observables, the former in terms of group representations and the latter in terms of a generalization of the notion of systems of imprimitivity of Mackey.[3] We have also taken advantage of the present generalization for rewriting in a more direct algebraic way some of our previous results.[1]

The paper will be organized as follows. In the first part we briefly recall some axiomatic justifications of our approach and then discuss the structure of the state spaces in terms of carrier spaces of group representations. In the second part we give the general solution of the representation problem and in part three we discuss the associated observable representations. Finally in part four we illustrate our results at the hand of a simple representative example. The main applications of the theory will be discussed in a separate paper.[4]

 0022-2488/81/112389-08$01.00

1. K SPACES AND PROJECTIVE K REPRESENTATIONS

A K-space is defined as being a direct union over some (Borel) index set S of a family of (isomorphic) Hilbert spaces (that will be assumed here to be complex and separable)

$$K = \bigvee_{s\in S} \mathscr{H}_s. \tag{1.1}$$

The elements of K (respectively, the pure states) are thus given by some index $s_0 \in S$ and a corresponding vector (respectively, a ray) in $\mathscr{H}_{s_0}$.

The choice of (1.1) as a possible state space for a physical system can be justified by a pragmatic argument (see the Introduction), but also from an axiomatic point of view, which is interesting to briefly mention here (for details we refer to Ref. 5). On K there exists, namely, a natural so-called lattice of propositions denoted by $\mathscr{L}(K)$ and defined as follows. A *projection* in K is a family of projectors $\{P_s\}$ in the corresponding Hilbert spaces; in the set of all projections one can define a lattice structure using the lattice structure which is based on the operations of unions and intersections of closed subspaces in each single Hilbert space. More precisely the *inclusion*, *compatability*, *complementary*, and *intersection* in $\mathscr{L}(K)$ are defined in turn by

$$\begin{aligned}
&\text{(i)}\ \{P_s\} < \{Q_s\} \text{ iff } P_s < Q_s \quad \forall s \text{ iff } P_s Q_s = P_s \quad \forall s,\\
&\text{(ii)}\ \{P_s\} \leftrightarrow \{Q_s\} \text{ iff } P_s \leftrightarrow Q_s \quad \forall s \text{ iff } [P_s, Q_s] = 0 \quad \forall s,\\
&\text{(iii)}\ \{P_s\}' = \{P_s'\} = \{\mathbb{1}_{\mathscr{H}_s} - P_s\},\\
&\text{(iv)}\ \{P_s\} \wedge \{Q_s\} = \{P_s \wedge Q_s\} = \left\{\text{s-lim}_{n\to\infty} (P_s Q_s)^n\right\}.
\end{aligned} \tag{1.2}$$

The above algebraic structure allows application of the notion of morphism in a straightforward way and it follows from the representation theorem of Piron[5] that (up to small technical restrictions) any lattice of propositions corresponding to an elementary physical system is isomorphic to such a $\mathscr{L}(K)$. In other words, the space (1.1) is then indeed the most general state space for the corresponding system, and the projections (1.2) then correspond to the properties of the system. We refer to Ref. 5 for a more precise review of these aspects.

We can now define the notion of symmetry in this framework.

Definition 1.1: A symmetry in K is an automorphism of the corresponding lattice of propositions, i. e., an invertible map of $\mathscr{L}(K)$ onto itself that preserves its algebraic structure.

The essential result that we shall need in the sequel is the following generalization of the theorem of Wigner.

Theorem 1.2 (Ref.6): Every symmetry of a proposition system defined by a family $\{\mathscr{H}_s, s\in S\}$ is given by a permutation π of the index set S and a family of unitary or antiunitary operators

$$U_{\pi(s)}: \mathscr{H}_s \to \mathscr{H}_{\pi(s)}. \tag{1.3}$$

Moreover, each U_s is unique up to a phase.

It follows directly from this theorem that for each pair of elements g_1, g_2 in a group G of symmetries we have the following generalized projective representation relations:

$$U_s(g_1) U_{\pi_{g_1}(s)}(g_2) = \omega_s(g_1, g_2) U_s(g_1 g_2),$$

where $\omega_s(g_1, g_2) \in U(1)$ is a phase that may depend on g_1, g_2, and s.

As a consequence of the above theorem we are thus led to the following.

Definition 1.3 (Ref. 2): A *projective K-representation* of a group G in a space K is a set of automorphisms $U(g): K \to K$ that satisfies the following four conditions:

$$\text{(i)}\ U(e) = \mathbb{1}_K, \quad e \text{ the unit of } G, \tag{1.4}$$

$$\text{(ii)}\ U(g_1) U(g_2) = \Omega(g_1, g_2) U(g_1 g_2) \tag{1.5}$$

with Ω a family $\{\omega_s\}$ of c-numbers of length 1, $\Omega: G \times G \to (U(1))^S$, its action on an element Ψ of K being given by $(\Omega(g_1, g_2)\Psi)_s = \omega_s(g_1, g_2)(\Psi)_s$,

(iii) $U(g)$ can be decomposed as

$$U(g) = L(g) \cdot \zeta^\circ(g), \tag{1.6}$$

where $L(g) = \{L_s(g)\}$ is a family of (anti-) unitary operators acting in each $\mathscr{H}_s$, i. e., $L_s(g) \in \mathscr{U}(\mathscr{H}_s)$ and $\zeta^\circ(g)$ is a permutation operator acting (only) on the variable s:

$$(\zeta^\circ \Psi)_s = (\Psi)_{g^{-1}s}. \tag{1.7}$$

We denote gs, by a slight abuse of notation, as the image of s under the permutation induced by g on S.

(iv) $U(g)$ is continuous in g on $S \times \mathscr{H}$ (with for $\mathscr{H}$ the strong topology).

It follows from the above definition that L, as a mapping on $S \times G$, satisfies the following equations:

$$\text{(i)}\ L_s(g_1) L_{g_1^{-1}s}(g_2) = \omega_s(g_1, g_2) L_s(g_1 g_2),$$

$$\text{(ii)}\ L_s(e) = \mathbb{1}_{\mathscr{H}_s}, \quad \forall s \in S, \tag{1.8}$$

i. e., L satisfies (projective) normalized G–S–$\mathscr{U}(\mathscr{H})$-cocycle equations,[7] and the problem can thus be tackled using cocycle techniques.

Another important consequence of the above definition is that S is then a G-space in the sense of Mackey[8]: symmetry products being defined by composition we indeed can define a natural antihomomorphic map μ from G to the automorphisms of S, with $\mu(g)s = g^{-1}s$ and this map μ satisfies, as required,

$$\begin{aligned}
&(\mu(g_1)\mu(g_2))(s) = \mu(g_1)((\mu(g_2)s)),\\
&\mu(g_1)\mu(g_2) = \mu(g_2 g_1),\\
&\mu(e)s = s, \quad \forall s \in S.
\end{aligned} \tag{1.9}$$

In general S thus splits under the action of G into disjoint orbits but in the particular case where the action of G is transitive, S can be identified with a quotient space $S \simeq G/H$, with $H = \text{Stab}\, s_0$, the (closed) stabilizer of an (arbitrary) point $s_0 \in S$.

Definition 1.4: A projective K-representation is called irreducible if and only if

(i) the action of G on S is transitive,

(ii) the set $L_{s_0}(H)$, $s_0 \in S$ arbitrary, $H = \text{Stab} s_0$ is an irreducible projective representation of H in $\mathscr{H}_{s_0}$.

Obviously this definition does not depend on the choice of s_0 and it follows by restriction from Definition 1.3 that $L_{s_0}(H)$ is always a unitary/antiunitary representation of H in $\mathscr{H}_{s_0}$.

Defining the commutant as the set of commuting families $\{A_s\}, A_s \in \mathcal{L}(\mathcal{H}_s)$, it is easy to see that a representation is irreducible if and only if the commutant is trivial (the proof is as in Ref. 1).

Similarly one can define equivalent representations as representations intertwined by an automorphism of K, i. e.,

Definition 1.5: Two projective K-representations U and U' of G are called *equivalent* if and only if there exists a (Borel) isomorphism $\tau: S \to S'$ in the corresponding base spaces and a family of unitary operators $V = \{V_s, s \in S\}$, $V_s: \mathcal{H}_s \to \mathcal{H}_{\tau(s)}$ such that $\forall g \in G$

$$U'(g) = V \cdot U(g) \cdot V^{-1}. \tag{1.10}$$

The first problem is the one of the possible families of phase factors which may occur in (1.3). Analogous to the usual case it follows from the associativity that they are not arbitrary but satisfy the following generalized comultiplier equations:

$$\omega_s(g_1, g_2)\omega_s(g_1 g_2, g_3) = \omega_s(g_1, g_2 g_3)\omega^{\eta(g_1)}_{g_1^{-1}s}(g_2, g_3), \tag{1.11}$$

where $\eta(g)$ denotes the complex conjugation in the case where $U(g)$ is a family of antiunitary operators. In Ref. 2 we have investigated in detail the solutions, called K-comultipliers, of these equations and we have proven the following:

Theorem 1.6: Let U be an irreducible (unitary) projective K-representation. Then the K-multiplier Ω is equivalent to a K-multiplier Ω' given by

$$\omega'_s(g_1, g_2) = \omega_{s_0}(\nu(s, g_1),\ \nu(g_1^{-1}s, g_2)), \tag{1.12}$$

with $s_0 \in S$ arbitrary, $H = \mathrm{Stab}s_0$, and $\nu(s, g)$ is the (unique) element of H defined by

$$\nu(s, g) = k(s) \cdot g \cdot (k(g^{-1}s))^{-1}, \tag{1.13}$$

where $k(s)$ is, for each s, the (fixed) representative of the unique coset class in G/H satisfying the equation $k(s)^{-1} \cdot s_0 = s$. Conversely each equivalence class $[\omega] \in H^2(H, U(1))$ gives rise, via the same formula (1.12) with $\omega_{s_0} = \omega$, to one and only one equivalence class of solutions $[\Omega]$ of (1.11).

For more details on this cohomological problem we refer to Ref. 2. In words, the above theorem says that Ω is, up to equivalence, uniquely defined by its double restriction from $S \times G$ to $s_0 \times \mathrm{Stab}s_0$, for some arbitrary $s_0 \in S$, and that conversely each solution of this double restriction gives rise to a solution Ω. Moreover this correspondence preserves the equivalences.

Finally let us briefly show how the so-called lifting procedure[9] generalizes, i. e., how it is possible to construct a group G^{Ω} in such a way that the projective K-representations of G with K-multiplier Ω can be lifted to ordinary (vector) K-representations of this larger group G^{Ω}. We therefore consider G as given, $K = V_S \mathcal{H}_s$, and U some irreducible projective K-representation of G with Ω arbitrary but fixed. The group G^{Ω} is defined as follows. It has elements

$$(\Phi, g), \quad \Phi \in [U(1)]^S, \quad g \in G, \tag{1.14}$$

i. e., Φ is a family of phases, $\Phi = \{\varphi_s\}$, $s \in S$, $\varphi_s \in U(1)$. The product in G^{Ω} is given by

$$(\Phi_1, g_1)(\Phi_2, g_2) = (\Phi_1 \cdot \xi(g_1)\Phi_2 \cdot \Omega(g_1, g_2), g_1 g_2), \tag{1.15}$$

where the product in $(U(1))^S$ is defined term by term and the action ξ is given by

$$\xi(g)\Phi = U(g) \cdot \Phi \cdot U(g)^{-1} \tag{1.16}$$

or, equivalently, as can easily be computed, and in components, by

$$(\xi(g)\Phi)_s = \varphi^{\eta(g)}_{g^{-1}s}, \tag{1.17}$$

with $\eta(g)$ as in (1.11).

Using the K-comultiplier equations (1.11) it is easy to see that this product is associative. Together with a unit element $(\mathbb{1}_K, e)$ with e the unit of G, G^{Ω} forms a group and the projective K-representation U defines an ordinary representation V of G^{Ω} with

$$V(\Phi, g) \overset{\text{def}}{=} \Phi \cdot U(g). \tag{1.18}$$

Conversely, given any (vector) K-representation V of G^{Ω} which satisfies

$$V(\Phi, e) = \Phi, \tag{1.19}$$

one can construct a projective K-representation U of G with

$$U(g) = V(\mathbb{1}_K, g). \tag{1.20}$$

Obviously, G^{Ω} appears as an extension with the normal subgroup of phases in general not in the center, even if all $L_s(g)$ in (1.6) are unitary(this is in opposition to the usual case)

$$0 \to [U(1)]^S \to G^{\Omega} \to G \to 1, \tag{1.21}$$

the extension being characterized by the factor system Ω and the mapping ξ of G in the automorphisms of $(U(1))^S$

Similarly to the usual case,[10] if the mapping $S \to \Phi$, which can be considered as a cross section in the bundle of base S and fiber $U(1)$, is supposed to be continuous for all Φ and if G is continuous then G^{Ω} is continuous. Analogously if these cross sections are measurable and if G is a measurable group then G^{Ω} is measurable, too.

2. PROJECTIVE K REPRESENTATIONS

In this section we shall, in a first part, construct explicit examples of solutions for Eqs. (1.3)–(1.7), i. e., explicit examples of projective K-representations. In a second part we shall then show that this construction is exhaustive, i. e., we shall give the general solution of this problem.

Thus let G and S be given. As we are interested in elementary systems, we may suppose without loss of generality (see Definition 1.4) that the action of G on S is transitive, i. e., $S \simeq G/H$, $H = \mathrm{Stab}s_0$, $s_0 \in S$. Let ξ denote this action.

Furthermore, let D be some projective unitary representation of H, with multiplier ω, in some carrier space $\mathcal{H}(D)$. Then let K be the direct union over S of copies of this space.

$$K = V_s(\mathcal{H}(D))_s. \tag{2.1}$$

On this space K we define the following operators $\forall g \in G$:

$$U(g) = L(g) \cdot \xi(g), \tag{2.2}$$

where $L(g) = \{L_s(g)\}$ and

$$L_s(g) = D(\nu(s, g)), \tag{2.3}$$

with $\nu(s, g)$ as given in (1.13). Let us just remark here that these formulas are formally the same as in the nonprojective

case,[1] this in contradiction to the usual inducing procedure for projective representations in Hilbert spaces where additional phase factors need to be inserted.[9]

It follows straightforwardly from (2.3) that

$$L_s(g_1)L_{g_1{}^{-1}s}(g_2) = \omega(\nu(s, g_1), \nu(g_1^{-1}s, g_2))D(\nu(s, g_1g_2)), \tag{2.4}$$

where we have used the G–S–H-cocycle property[7] of $\nu(s, g)$,

$$\nu(s, g_1)\nu(g_1^{-1}s, g_2) = \nu(s, g_1g_2). \tag{2.5}$$

Comparing (2.4) with (1.8), using (2.3) again, and applying the last part of Theorem 1.6, it is now easy to see that (2.2) and (2.3) defines a projective K-representation in the space (2.1) and with K-multiplier Ω given by

$$(\Omega(g_1, g_2))_s = \omega(\nu(s, g_1), \nu(g_1^{-1}s, g_2)). \tag{2.6}$$

The above construction is of course nothing else than a generalization of the inducing procedure of Mackey.[8] The corresponding representation will therefore be called a K-induced representation. The important result is now the following converse.

Theorem 2.1: Let U be any irreducible projective K-representation with K-multiplier Ω, then U is equivalent to a K-induced representation.

Proof: It follows from the irreducibility that we may assume without loss of generality that $S \simeq G/H$. Let us choose a set of fixed right coset representatives $k(s) \in G$, $s \in S$. Then trivially with (1.13) each g can be written as

$$g = k(s)^{-1} \cdot \nu(s, g) \cdot k(g^{-1}s), \quad \forall s \in S$$

and correspondingly the unitary operators in the families $L = \{L_s(g)\}$, defined by the given representation U, can be written

$$L_s(g) = L_s(k(s)^{-1} \cdot \nu(s, g) \cdot k(g^{-1}s)). \tag{2.7}$$

We may expand this expression, using (1.8) (i) iteratively, and we find

$$\begin{aligned} L_s(g) = {} & \omega_s^{-1}(k(s)^{-1}, \nu(s, g)k(g^{-1}s)) \cdot \omega_{s_0}^{-1}(\nu(s, g), k(g^{-1}s)) \\ & \cdot \omega_{s_0}(k(s), k(s)^{-1}) \cdot L_{s_0}^{-1}(k(s)) \cdot L_{s_0}(\nu(s, g)) \cdot L_{s_0}(k(g^{-1}s)). \end{aligned} \tag{2.8}$$

Using Theorem 1.6 we may also assume without loss of generality that Ω is already in the form (1.12) and normalized (cf. Ref. 2). Using now (1.13) and (1.12) with $g_1 = k(s)^{-1}$ one easily finds that, $\forall g_2 \in G$

$$\omega_s(k(s)^{-1}, g_2) = \omega_{s_0}(k(s_0), g_2) = 1,$$

where we have used the normalization of Ω. The other phases in (2.8) can be seen to be trivial by similar straightforward calculations so that (2.8) reduces to

$$L_s(g) = (L_{s_0}(k(s)))^{-1} L_{s_0}(\nu(s, g)) L_{s_0}(k(g^{-1}s)). \tag{2.9}$$

Then using (1.10), with $S' = S$ and which then reads, in components,

$$(VU'(g)V^{-1})_s = V_s U_s(g) V_{g^{-1}s}, \tag{2.10}$$

one recognizes that, with $V = \{V_s\}$ and

$$V_s \overset{\text{def}}{=} L_{s_0}(k(s)), \tag{2.11}$$

the operators (2.9) can be brought back to the form (2.3), as the equivalent G–S–$\mathscr{U}(\mathscr{H})$-cocycle then reads

$$L'_s(g) = L_{s_0}(\nu(s, g)) \tag{2.12}$$

and this achieves the proof of the theorem.

This theorem thus implies that the problem of classifying all irreducible projective K-representations of a group G is reduced to the problem of the usual projective representations in Hilbert spaces of some closed subgroups H of G. In the next section we shall see which subgroups are relevant for our purposes.

3. SUPERSYSTEMS OF IMPRIMITIVITY

As in Ref. 1, and for a given physical system with a given kinematical symmetry group G, not any K-space carrying a representation of G will be considered as a possible state space for the system. We shall in addition require that, in K, the observables are faithfully represented, in a sense that we now want to make precise.

As mentioned in the Introduction, the state of the system is characterized by the possible outputs of the various measurements that we have chosen to eventually perform on it. We thus have *a priori* a given set of observables $\{A\}$. Let Γ_A denote the set of possible values for the observable A, i. e., the spectrum of A. We now demand that, for each (Borel) subset Δ of Γ_A there exists an operator P_Δ^A in K which selects the states having in actuality the property of having, for A, a value within Δ. It follows from the representation (1.2) of the properties in K, that P_Δ^A should be a projection in K, i. e., a family of projectors in the corresponding family of Hilbert spaces

$$P_\Delta^A = \{(P_\Delta^A)_s\}, \quad (P_\Delta^A)_s \in \mathscr{P}(\mathscr{H}_s). \tag{3.1}$$

Furthermore, it should satisfy the following measure theoretical properties:

(i) $P_\emptyset^A = 0_K$, $P_{\Gamma_A}^A = \mathbb{1}_K$,

(ii) $P_{\Delta_1 \cap \Delta_2}^A = P_{\Delta_1}^A P_{\Delta_2}^A$,

(iii) $P_{\bigcup_{i \in I} \Delta_i}^A = \sum_{i \in I} P_{\Delta_i}^A$ for $\Delta_i \perp \Delta_j$ if $i \neq j$ (i. e., disjoint) and I is countable. (3.2)

In (3.2) the sum and the products are meant as sum, respectively, products of the individual projectors in the families. On the other hand it is also by its action on $\Gamma = \{\Gamma_A\}$ that the group G is defined. We thus correspondingly require these mappings to transform covariantly, i. e.,

$$P_{g \cdot \Delta}^A = U(g) P_\Delta^A U(g)^{-1}. \tag{3.3}$$

In the particular case where S is a singleton [hence K is a (single) Hilbert space], the conditions (3.2) and (3.3) are nothing else than the conditions of systems of imprimivity of Mackey.[3] In analogy with this terminology we shall call in general such mappings *supersystems of imprimitivity* (s.s.o.i.) as in Ref. 1.

Example: Let $\Delta \in \mathscr{B}(S)$, the Borel sets of S, and χ_Δ the corresponding characteristic function. Then

$$P_\Delta = \{\chi_\Delta(s) \cdot \mathbb{1}_{\mathscr{H}_s}\} \tag{3.4}$$

obviously satisfies (3.1) and (3.2). Moreover, using (2.2) one

directly finds

$$U(g)\chi_\Delta U(g)^{-1} = \xi(g)\cdot\chi_\Delta\cdot\xi(g)^{-1} = \chi_{g\Delta}, \tag{3.5}$$

hence (3.4) also satisfies (3.3) and thus defines a supersystem of imprimitivity.

We thus see with this simple example that S always corresponds to one (or more) observable. On the other hand, we have seen that it can also be associated with an homogeneous space G/H of G. This is nothing but a particular case of a more general relationship between observables and homogeneous spaces that we shall also detail in the sequel.

Now let K and U be given. Because we want to consider elementary systems we may assume without loss of generality that U is irreducible. Let then P be some s.s.o.i. based on the Borel subsets of some G-space T. We may also assume that the corresponding action is transitive (or else P would define more than one observable). Hence T can be identified with some quotient space G/H_1, with $H_1 = \mathrm{Stab}t_0$, t_0 some (arbitrary) element of T.

On the other hand, if P defines an observable, T corresponds to some subset of Γ. The relationship between both is easy to recognize: $H_1 \cong H_\gamma$, where γ is some fixed value of the observable corresponding to P and H_γ is its stabilizer.

We thus have, $\forall F\in\mathscr{B}(T)$, projections P^F in K satisfying (3.2) and (3.3); hence, in each Hilbert space $\mathscr{H}_s$ we have

$$\begin{aligned}&\text{(i)}\ P_s^\emptyset = 0_{\mathscr{H}_s},\quad P_s^T = \mathbb{1}_{\mathscr{H}_s},\\ &\text{(ii)}\ P_s^{F_1\cap F_2} = P_s^{F_1}P_s^{F_2},\\ &\text{(iii)}\ (P^{\cup F_i})_s = \sum_i P_s^{F_i},\quad F_i|F_j \text{ for } i\neq j,\end{aligned} \tag{3.6}$$

as well as the covariance condition

$$(U(g)^{-1}P^FU(g))_s = P_s^{g^{-1}F}. \tag{3.7}$$

These last equations can be rewritten, using the decomposition of U in (1.6), as

$$L_s(g)^{-1}P_s^FL_s(g) = P_{g^{-1}s}^F, \tag{3.8}$$

where we have for simplicity omitted the imbedding maps of $\mathscr{H}_s$ in $\mathscr{H}_{gs}$ induced by $\xi(g)$.

Consider now the double restriction of (3.8) to some elements $s_0\in S$ and for the group elements $h\in H\subseteq G$ with $H = \mathrm{Stab}s_0$. We obtain

$$L_{s_0}(h)^{-1}P_{s_0}^FL_{s_0}(h) = P_{s_0}^{h^{-1}F}. \tag{3.9}$$

As (3.6) is true for each s, hence for s_0, we thus have, together with (3.9), that this restriction, i. e., the set $\{P_{s_0}^F, F\in\mathscr{B}(T)\}$ is, in $\mathscr{H}_{s_0}$, a usual system of imprimitivity of Mackey for H on T. In order to apply the theorem of imprimitivity we, however, need one more property, namely the transitivity of the action of the group. But this property is in general lost in the above restriction. In order to get out of this difficulty we have to analyze in some more detail the action of H on T.

Under the action of H, the space T splits in general into disjoints orbits $\{T_\mu, \mu\in M$, with M some index set$\}$. It is easy to verify that this set M can be identified with the set of $H{:}H_1$ double cosets: each t can indeed be written as $l(t)\cdot t_0$, t_0 fixed, $l(t)$ some coset representative of G/H_1, hence t and t' are in the same orbit if $l(t)\cdot t_0 = h\cdot l(t')\cdot t_0$ for some $h_1\in H_1$, i. e., iff

$$l(t) = h\cdot l(t')\cdot h_1,$$

i. e., iff $l(t)$ and $l(t')$ are in the same $H{:}H_1$ double coset.

Then let $t_\mu\in T_\mu$, some μ, and $H_1^\mu = \mathrm{Stab}t_\mu$, then one easily recognizes that $T_\mu \cong H/H\cap H_1^\mu$.

Proposition 3.1: The coset space $H_1^\mu/H\cap H_1^\mu$ can be identified with a (closed) subset S_1^μ of S, i. e., one can write

$$H_1^\mu = \bigcup_{s\in S_1^\mu}(H_1^\mu\cap H)\cdot k'(s), \tag{3.10}$$

where $k'(s)$ are some coset representatives of G/H.

Proof: Let s be arbitrary, t_μ fixed in T_μ, and $H_1^\mu = \mathrm{Stab}t_\mu$. If $k(s)t_\mu\notin T_\mu$ then no $h\in H$ can bring it back in T_μ because the orbits are disjoint. Hence the equation $h\cdot k(s)t_\mu = t_\mu$ has no solution h. If, however, $k(s)t_\mu\in T_\mu$ then $k(s)t_\mu = ht_\mu$ for some h, hence $(h^{-1}k(s))\in H_1^\mu$. We may choose $h^{-1}k(s)$ as new coset representative $k'(s)$ and obviously $(H_1^\mu\cap H)\cdot k'(s)\subseteq H_1^\mu$. Denoting by S_1^μ all $s\in G/H$, with this property we thus have

$$\bigcup_{s\in S_1^\mu}(H_1^\mu\cap H)k'(s)\subseteq H_1^\mu.$$

Conversely, each $h_1\in H_1^\mu$ can be written as $h\cdot k(s)$ and as $h\cdot k(s)t_\mu = t_\mu$ one has $s\in S_1^\mu$ hence finally the result asserted in (3.10).

Then let G_μ denote the stabilizer of the whole orbit T_μ. Obviously $G_\mu\supseteq H$ and we have the following.

Proposition 3.2: The coset space G_μ/H can be identified with a (closed) subset S_μ of S, i. e.,

$$G^\mu = \cup_{s_\mu} H\cdot k(s), \tag{3.11}$$

where $S_1^\mu\supseteq S_\mu$, $\forall_\mu\in M$.

Proof: $g\in G^\mu$ implies $h\cdot g\in G^\mu$. Conversely if $k(s)\notin G^\mu$ for some s then there exists at least one $t\in T_\mu$ such that $k(s)t\notin T_\mu$. But then $h\cdot k(s)t_\mu\notin T_\mu$, $\forall h\in H$ because the orbits are disjoint, hence there can be no element of G_μ in this coset. The condition $s\in S^\mu$ means $h\cdot k(s)\in G_\mu$, $\forall h\in H$, hence $h\cdot k(s)t_\mu\in T_\mu$. There thus exists an element $h'\in H$ such that $h\cdot k(s)\cdot t_\mu = h'\cdot t_\mu$ and $(h')^{-1}h\cdot k(s)\in H_1^\mu$; hence $s\in S_1^\mu$.

We may now come back to (3.8), the restriction at some s_0, for $H = \mathrm{Stab}s_0$ of the arbitrary but given s.s.o.i. P. In order to recover the transitivity property we restrict it further to the Borel subsets of T_μ:

$$P_{s_0}: F_\mu\to\mathscr{P}(\mathscr{H}_{s_0}),\quad F_\mu\in\mathscr{B}(T_\mu)\subseteq\mathscr{B}(T). \tag{3.12}$$

This restricted mapping still satisfies the covariance condition (3.9) and two of the measure theoretical conditions (3.6) (ii) and (3.6) (iii). It does, however, not necessarily satisfy the first one, as for $F_\mu = T_\mu$ the image of (3.12) is not necessarily the identity.

Proposition 3.3: Let s_0 be arbitrary but fixed and suppose F is some H-invariant Borel subset of T, then

$$P_{s_0}^F = \lambda\cdot\mathbb{1}_{\mathscr{H}_{s_0}} \text{ with } \lambda = 0 \text{ or } \lambda = 1.$$

Proof: It follows from (3.9) and the H-invariance that $P_{s_0}^F$ commutes with $L_{s_0}(h)$, $\forall h\in H$. But as U is irreducible it follows from Definition 1.4 (ii) and the lemma of Schur that $P_{s_0}^F = \lambda\mathbb{1}_{\mathscr{H}_{s_0}}$. That $\lambda = 0$ or 1 follows from the fact that $P_{s_0}^F$ is a projector [(3.6) (ii)].

In particular for the orbits T_μ one has thus

$$P^{T_\mu}_{s_0} = \lambda_\mu \mathbb{1}_{\mathscr{H}_{s_0}}, \quad \lambda_\mu \in \{0,1\}. \tag{3.13}$$

Proposition 3.4: There exist one and only one $\mu_0 \in M$ such that (for s_0 fixed) $\lambda_{\mu_0} = 1$.

Proof: Suppose this equation holds for at least two (disjoint) orbits T_μ and $T_{\mu'}$. Then from Proposition 3.3, as $T_\mu \cup T_{\mu'}$ is H-invariant and from (3.6) (iii) we find that $P^{T_\mu \cup T_{\mu'}}_{s_0} = (\lambda_\mu + \lambda_{\mu'})\mathbb{1}_{\mathscr{H}_{s_0}}$. As $\lambda_\mu + \lambda_{\mu'}$ is either 0 or 1 λ_μ and $\lambda_{\mu'}$ cannot be both equal to 1. On the other hand, if all $\lambda_\mu = 0$, then $P \equiv 0$ and it cannot thus satisfy the s.s.o.i. conditions. Hence there is exactly one orbit T_{μ_0} such that $\lambda_{\mu_0} = 1$.

Conversely, for a given orbit T_{μ_0} there may be, however, more $s \in S$ for which $P^{T_{\mu_0}}_s = \mathbb{1}_{\mathscr{H}_s}$.

Proposition 3.5: The set S_{μ_0} for which $P^{T_{\mu_0}}_s = \mathbb{1}_{\mathscr{H}_s}$ is the same as the set S_μ in Proposition 3.2, for $\mu = \mu_0$.

Proof: As follows from (2.3) and from Theorem 2.1, we may assume without loss of generality that $L_{s_0}(k(s)) = \mathbb{1}_{\mathscr{H}_{s_0}}$ so that, with (3.8), $P^F_s = P^{k(s)F}_{s_0}$, $\forall s \in S$, $F \in \mathscr{B}(T)$. Using now Proposition 3.5 we thus have

$$P^F_s = \mathbb{1}_{\mathscr{H}_s} \quad \text{iff} \quad k(s)F = T_{\mu_0}, \tag{3.14}$$

hence, with $F = T_{\mu_0}$, (3.14) reads

$$P^{T_{\mu_0}}_s = \mathbb{1}_{\mathscr{H}_s} \quad \text{iff} \quad k(s)T_{\mu_0} = T_{\mu_0},$$

i. e., $k(s) \in G_{\mu_0}$ as asserted.

It now follows from (3.13) and Proposition 3.4 that P_{s_0} restricted, as in (3.12) on the orbit T_{μ_0} determined by this Proposition 3.4, does satisfy all usual axioms of an s.o.i. for $H = \mathrm{Stab} s_0$. As the action is now transitive $T_{\mu_0} \cong H/H'$ where, as is easy to compute,

$$H' = H^{\mu_0}_1 \cap H. \tag{3.15}$$

Collecting all these results we now have proven the following.

Theorem 3.6: Let P be a (transitive) s.s.o.i. based on $T \cong G/H_1$, for the irreducible projective K-representation U of G, then P is given, up to equivalence, by

$$P^F_s = P^{k(s)F}_{s_0} = P^{(k(s)F)\cap T_{\mu_0}}_{s_0}, \tag{3.16}$$

where T_{μ_0} is the unique H-orbit in T for which $P^{T_{\mu_0}}_{s_0} = \mathbb{1}_{\mathscr{H}_{s_0}}$. Moreover the mapping (3.16) defines a usual system of imprimitivity on $\mathscr{H}_{s_0}$ and based on $H/H \cap H^{\mu_0}_1$, with $H^{\mu_0}_1 = \mathrm{Stab} t_{\mu_0}$, some $t_{\mu_0} \in T_{\mu_0}$.

The importance of the above theorem lies in the fact that P is thus completely determined by its restriction, at s_0, on H and on some orbit T_{μ_0}, and that this restriction forms a usual system of imprimitivity.

The main point is now to reverse the procedure and to show that each ordinary system of imprimitivity on $\mathscr{H}_{s_0}$ gives rise to a s.s.o.i. on K. Let us therefore suppose that we have given $s_0 \in S \cong G/H$, $H = \mathrm{Stab} s_0$ and $t_0 \in T \cong G/H_1$, $H_1 = \mathrm{Stab} t_0$, and $T_{\mu_0} = H \cdot t_0 \cong H/H \cap H_1$. Suppose furthermore that D is an irreducible representation of H in some Hilbert space $\mathscr{H}(D)$, which is induced from $H \cap H_1$, i. e., via Mackey's theorem of imprimitivity, that there exists an ordinary s.o.i. Q based on T_{μ_0},

$$Q: \mathscr{B}(T_{\mu_0}) \to \mathscr{P}(\mathscr{H}(D)) \tag{3.17}$$

with $\mathscr{P}(\mathscr{H}(D))$ the set of all projectors in $\mathscr{H}(D)$. One can now consider the K-space

$$K = \bigvee_{s \in G/H} (\mathscr{H}(D))_s \tag{3.18}$$

and, on K, consider (the irreducible) representation U defined in (2.2) and (2.3),

$$U_s(g) = D(\nu(s,g)) \cdot \xi_s(g). \tag{3.19}$$

In this framework we now define the following mapping:

$$P: \mathscr{B}(T) \to \mathscr{P}(K) \tag{3.20}$$

with

$$P^F_s \overset{\text{def}}{=} Q(k(s)F \cap T_{\mu_0}). \tag{3.21}$$

Theorem 3.7: The mapping constructed in (3.21) is a supersystem of imprimitivity for U in (3.19) on the space K given by (3.18).

Proof: We have to verify in turn the conditions (3.6) and (3.7) using (3.21) and the fact that Q satisfies the usual s.o.i. properties on T_{μ_0}.

(i) $P^T_s = Q(k(s)T \cap T_{\mu_0}) = Q(T_{\mu_0}) = \mathbb{1}_{\mathscr{H}_s}$ $\forall s$ and $P^\emptyset_s = Q(\emptyset \cap T_{\mu_0}) = Q(\emptyset) = 0$.

(ii) $P^{F_1 \cap F_2}_s = Q([k(s)F_1 \cap k(s)F_2] \cap T_{\mu_0}) = Q((k(s)F_1 \cap T_{\mu_0}) \cap (k(s)F_2 \cap T_{\mu_0})) = P^{F_1}_s \cdot P^{F_2}_s$ The product of two families of projectors being the family of the product, this verifies the condition (3.6) (ii).

(iii) Suppose $F_1 \downarrow F_2$, then $k(s)F_1 \downarrow k(s)F_2$, $\forall s$ and thus $(k(s)F_1 \cap T_{\mu_0}) \downarrow (k(s)F_2 \cap T_{\mu_0})$. We thus obtain that $P^{F_1 \cup F_2}_s = Q((k(s)F_1 \cup k(s)F_2) \cap T_{\mu_0}) = Q((k(s)F_1 \cap T_{\mu_0}) \cup Q(k(s)F_2 \cap T_{\mu_0}) = Q((k(s)F_1) \cap T_{\mu_0}) + Q(k(s)F_2 \cap T_{\mu_0}) = P^{F_1}_s + P^{F_2}_s$, hence (3.6) (iii) is also verified, the sum of two families being the family of the individual sums.

(iv) It still remains to verify (3.7): Using the expression (3.19) for U and the fact that Q is an ordinary s.o.i. for D, we find

$$\begin{aligned}(U(g)P^F U(g)^{-1})_s &= D(\nu(s,g)) P^F_{g^{-1}s} D^{-1}(\nu(s,g)) \\ &= D(\nu(s,g)) Q(k(g^{-1}s)F \cap T_{\mu_0}) D^{-1}(\nu(s,g)) \\ &= Q(\nu(s,g)[k(g^{-1}s)F \cap T_{\mu_0}]) \\ &= Q((\nu(s,g)k(g^{-1}s)F) \cap T_{\mu_0}),\end{aligned} \tag{3.22}$$

where we have used that, as $\nu(s,g) \in H$, it leaves T_{μ_0} invariant. Now using the definition (1.13) of the cocycle ν, i. e., $\nu(s,g) = k(s)g(k(g^{-1}s))^{-1}$ and inserting in (3.22) we find

$$\begin{aligned}(U(g)P^F U(g)^{-1})_s &= Q(k(s)gF \cap T_{\mu_0}) \\ &= P^{gF}_s, \quad \forall s \in S\end{aligned}$$

and this achieves the proof of the theorem.

Example: As each representation of H is trivially induced from itself, we have on $\mathscr{H}(D)$ with $H = H_1$ a usual s.o.i. on the 1-point set s_0 with

$$Q(s_0) = \mathbb{1}_{\mathscr{H}(D)}, \quad Q(\emptyset) = 0_{\mathscr{H}(D)}.$$

Using the definition (3.21) one correspondingly finds, as s.s.o.i. on K, $\forall E \in \mathscr{B}(G/H)$,

$$P^E_s = Q(s)E \cap s_0) = \begin{cases} 1 & \text{if } s_0 \in k(s)E \\ 0 & \text{otherwise} \end{cases},$$

hence

$$P_s^E = \chi_{k(s)E}(s_0) = \chi_E(k(s)^{-1}s_0) = \chi_E(s) \tag{3.23}$$

and we recover in this way the example given in (3.4).

It is useful to remark at this point that the subgroup H_1 is uniquely defined by the corresponding observable only up to conjugation. Different choices corresponding to different $H{:}H_1$ double coset elements can give in principle different (even possibly inequivalent) s.s.o.i. The choice of H_1 is, however, restricted by the following consideration. S always determines itself as above one or more observables so that H is the kinematical symmetry group corresponding to all the remaining observables. The subgroup H_1 is then, for each of these observables, uniquely determined up to conjugation by some $h \in H$. The $H{:}H_1$ double coset element is then fixed.

Reversing the above procedure we thus find all allowable K spaces if we proceed as follows. In a first step, for each subset of observables $\{B\} \subseteq \{A\}$ with corresponding kinematical symmetry subgroup H, $H \subseteq G$, we have to find those projective unitary representations which admit usual systems of imprimitivity for each observable in $\{B\}$ (and no other). For this, we can use the theorem of imprimitivity of Mackey and thus ask whether these representations are equivalent to representations induced by the corresponding subgroups or not. In a second step we construct U, the space K, and the s.s.o.i. as in (3.19), (3.18), and (3.21), and we find in this way all remaining observables.

The observables found in the first step are the usual quantal observables as they correspond to spectral measures in separable Hilbert spaces. The observables found in step 2 then play the role of superselection parameter or, in other words, are of the classical type, as they commute with all other ones and have a purely discrete point spectrum.

We shall see in the next section and in Ref. 4 how the above procedure applies in concrete physical examples and show then that the (quite mathematical) above framework is indeed the physically relevant one for our purposes.

4. EXAMPLE

In this last section we shall thus consider a simple representative physical example of application of the method found in the last section. As previously mentioned the main applications, i. e., the discussion of the classical and quantal, relativistic and nonrelativistic states spaces for elementary particles, will be treated in a separate paper.[4]

Consider therefore the two-dimensional phase space Ω with coordinates p and q. This space corresponds then to the space Γ discussed in Sec. 3, as it contains the possible values of the observables position and momentum. Let G be the group of translations in p and q. This group is determined by the physical postulates that there exists no absolute zero for the position nor for the momentum. We thus have $G \simeq \mathbb{R}^2$ with elements (w,a) and action

$$(w,a)(p,q) \overset{\text{def}}{=} (p - w, q - a). \tag{4.1}$$

Let us now determine the possible state spaces for this system.

Suppose first that both p and q are classical in the sense of Sec. 3. Then H is trivial and $\mathscr{H}_s$ is one-dimensional $\forall s = (p,q) \in \Omega$. We thus have

$$K = V_{(p,q)}(\mathbb{C})_{(p,q)} \tag{4.2}$$

and a state consists of a point (p,q) and a ray in the corresponding complex plane. This framework can thus be identified with the usual framework of classical mechanics in one space dimension. The observables are straightforwardly obtained as in (3.23) as given by

$$\begin{aligned} P_{\Delta p} f(p,q) &= \chi_{\Delta p}(p) f(p,q), \\ P_{\Delta q} f(p,q) &= \chi_{\Delta q}(q) f(p,q), \end{aligned} \tag{4.3}$$

with Δp and Δq Borel sets on the real line and $f(p,q)$ any state. The observables are thus given, as is usual, by the projections on the corresponding subsets of the phase space.

Suppose then that either p or q is quantal, say q. Then $H \cong \mathbb{R}$ and $K = V_{\mathbb{R}}(\mathscr{H}_p)$ with $\mathscr{H}_p$ carrying an irreducible projective representation of $\mathbb{R}$ which admits an observable position. But, as is easy to see, irreducibility implies that $\mathscr{H}_p$ is one-dimensional whereas, by the Mackey imprimitivity theorem, $\mathscr{H}_p$ admits an observable position if and only if it is isomorphic to $\mathscr{L}^2(\mathbb{R})$. Hence there can be no solution in this case.

Suppose finally that both p and q are quantal. Then $H = G \simeq \mathbb{R}^2$. The vector irreducible representations are one-dimensional and, as above, they cannot thus admit observables for the position nor for the momentum. There is, however, one single family of projective representations in

$$K = \mathscr{L}^2(\mathbb{R}) \tag{4.4}$$

with elements $f(x)$, and which is given by

$$\begin{aligned} U(a) f(x) &= f(x - a), \\ U(w) f(x) &= \exp(i\lambda x w) f(x), \quad \lambda \in \mathbb{R}. \end{aligned} \tag{4.5}$$

The observables p and q are given by the following spectral measures:

$$\begin{aligned} P_{\Delta q} f(x) &= \chi_{\Delta q}(x) f(x), \\ P_{\Delta p} \hat{f}(k) &= \chi_{\Delta p}(k) \hat{f}(k), \end{aligned} \tag{4.6}$$

with $\hat{f}(k)$ the usual Fourier transform of f. Equivalently, these observables are given by the operators

$$\begin{aligned} \hat{q} f(x) &= x f(x), \\ \hat{p} f(x) &= -i\lambda^{-1} \partial_x f(x) \end{aligned} \tag{4.7}$$

so that, with of course $\lambda = \hbar^{-1}$, we recover the usual framework of quantum physics in one space dimension.

We thus see by this simple example that our formalism provides a very direct method which allows us to derive the usual classical and quantal state spaces in a simple unified way. Moreover these two solutions are seen to be the only ones. We have not only built in this way a common language for both frameworks but we have done it in a way which is independent of the dynamics and which is only based on the realization of the properties of a physical system, in accordance with the points of view exposed in the Introduction.

[1] N. Giovannini and C. Piron, Helv. Phys. Acta **52**, 518 (1979).
[2] N. Giovannini, Lett. Math. Phys. **5**, 161 (1981).
[3] G. W. Mackey, Proc. Math. Acad. Sci. USA **35**, 537 (1949).

[4]N. Giovannini, J. Math. Phys. **22**, 2397 (1981).
[5]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, Mass., 1976).
[6]C. Piron, in *Fundamenti di Meccanica Quantistica*, International School of Physics Enrico Fermi, Vol. IL, edited by B. d'Espagnat (Academic, New York, 1971), p. 274.
[7]V. S. Varadarajan, *Geometry of Quantum Theory*, Vol. II (Van Nostrand Reinhold, New York, 1970), Chaps. 8 and 9.
[8]G. W. Mackey, *The Theory of Unitary Group Representations* (University of Chicago, Chicago, 1976).
[9]G. W. Mackey, Acta Math. **99**, 265 (1958).
[10]K. R. Parthasarathy, *Multipliers on Locally Compact Groups*, Springer Lect. Notes Math. **93** (Springer, Berlin, 1969).

State spaces for classical and quantal, relativistic and nonrelativistic elementary particles

N. Giovannini
Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland

(Received 23 April 1981; accepted for publication 5 June 1981)

We apply a recently developed mathematical formalism to the kinematical classification of elementary physical systems, more precisely to the derivation of the possible state spaces for single (massive) particles and to their corresponding group theoretical interpretation. We show in particular that in all considered cases, relativistic or not, we find in a unified way two solutions, a quantal one and a classical one.

PACS numbers: 02.20. + b, 03.20. + i, 03.65.Bz

INTRODUCTION

In a previous paper,[1] we have identified and analyzed the necessary mathematical basis of a group theoretical framework in which it is possible to treat simultaneously and in a common language classical and quantum physics. These two cases appear as two extreme cases of a more general formalism where some observables are represented by self-adjoint operators in Hilbert spaces (i.e., have a quantal behavior) whereas other observables are of the classical type as they commute with all other observables (and have a purely discrete point spectrum). Because of their physical interpretation we also call this latter type of observables superselection variables or parameters. They should not be confused with superselection rules and we do not use this latter terminology in order to emphasize that we do not assume (and in fact this is in general not true) that they correspond to conserved quantities. The latter can anyway only be asserted when the dynamics are known and our framework is precisely independent of the evolution and of the interactions of the physical system.

In Ref. 1 we have analyzed in general and with group theoretical arguments the structure of the possible state spaces for a given physical system starting from the following main considerations. The system itself is characterized by a set of observables corresponding to the measurements that we (eventually) perform on it. The usual physical equivalence postulates correspond then to the action of a group, called the kinematical symmetry group and defined by its action on the possible values for the observables. This is sufficient for the characterization of the properties of the system and this is, as it should be, independent of the dynamics. The dynamics appear in such a framework as a next separate step corresponding to the description of the changes in the properties, and not of the properties themselves. We shall briefly indicate at the end of the present paper how this occurs within our formalism.

This formalism is characterized by the fact that we consider as possible state spaces (topological) direct unions (or families) of Hilbert spaces. This is justified both from a pragmatic point of view (as it covers classical physics when all Hilbert spaces are one-dimensional, and quantum physics when the union is trivial) and from a more fundamental point of view as the result of an axiomatic approach.[1,2]

The state spaces, for an elementary physical system, are further specified by the following two main assumptions. Firstly, they should carry an (irreducible) representation of the above mentioned kinematical symmetry group. Secondly, they should admit a (sufficiently faithful) representation for each observable corresponding to the physical system. For the first condition we have appropriately generalized the idea of induced representations in Hilbert spaces[3] and, for the second, the idea of systems of imprimitivity.[4] In Ref. 1 we have just made mathematically more precise all these concepts.

In the present paper we shall apply the general mathematical results of Ref. 1 to a class of physical models, relativistic and nonrelativistic. We shall see in particular how, in all considered cases, two and only two solutions appear in a unified way: a classical one and a quantal one. This is quite interesting not only for aesthetic reasons but more essentially because the discussion of different models (classical or quantal, relativistic or not) in a common language, where the mathematical objects are the same and where the physical interpretation is the same, is of fundamental importance for the understanding of the results and of the difficulties encountered in each separate context.

For this last reason, and also in order to try to apply in classical physics the succcessful group-theoretical approach to elementary systems of quantum mechanics initiated by Wigner, an increasing number of papers have recently been published.[5] To our knowledge, however, it is the first time that such an approach is not only descriptive (i.e., explores the informative correspondences between different mathematical objects of different theories with the same physical interpretation) but really derives these theories in a unified mathematical language.

The present paper will be organized as follows. After a brief reminder of the framework and of the procedure determined in Ref. 1 for the construction of the state spaces, we recall the main characteristics of the case of spinless particles following the results already obtained in Ref. 6. We then consider the particles with (arbitrary) spin, relativistic and nonrelativistic, and we extend the nonrelativistic models, especially in view of the discussion of nonrelativistic limits. Finally, we briefly indicate some essential features concerning the description of the dynamics in our framework.

 0022-2488/81/112397-07$01.00

1. THE FRAMEWORK

Without entering into details (cf. Ref. 1), let us first very briefly recall what the general framework is, and correspondingly the method which can be extracted from the main results of Ref. 1.

As explained in the Introduction we define the state space of an elementary physical system as a space which carries an irreducible representation of the kinematical symmetry group and which admits operators for each of the corresponding observables.

As state spaces we consider direct unions, over some Borel set S, of complex Hilbert spaces

$$K = \bigvee_{s\in S} \mathscr{H}_s. \tag{1.1}$$

The elements of K are thus given by some index s_0 and an element in the corresponding $\mathscr{H}_{s_0}$. To the kinematical symmetry group G corresponds an (irreducible) projective K-representation U in the automorphisms of K, and this implies[1] that to each group element g corresponds a permutation π in the index set S and a family of (anti-) unitary operators $\{U_{\pi(s)}:\mathscr{H}_s\to\mathscr{H}_{\pi(s)}\}$. In Ref. 1 we have shown that U is always equivalent to a representation of the following form:

$$U_s(g) = D(\nu(s,g))\,\zeta(g), \tag{1.2}$$

where ζ is an action of G on S and where D is some usual projective representation, with some multiplier ω, in $\mathscr{H}_{s_0}$, $s_0\in S$ arbitrary but fixed, for $H = \mathrm{Stabs}_0 \subseteq G$ (the stabilizer of an arbitrary point $s_0\in S$). The cocycle ν in (1.2) is defined (writing gs for the image of s under the permutation induced by g on S) by

$$\nu(s,g) = k(s)g(.k(g^{-1}s))^{-1}, \tag{1.3}$$

with $k(s)$ some (fixed) coset representative of the unique coset class in G/H satisfying the equation $k(s)^{-1}s_0 = s$ for a given s: we indeed may assume that the action of G on S induced by π is transitive if U is irreducible. The representation (1.2) is projective and this means, in K, that

$$U_s(g_1)U_{g_1{}^{-1}s}(g_2) = \omega_s(g_1,g_2)U_s(g_1g_2), \tag{1.4}$$

where $\omega_s(g_1,g_2)$ is a phase factor which may depend on g_1,g_2 and on s. These phase factors are given, from (1.2), by

$$\omega_s(g_1,g_2) = \omega(\nu(s,g_1),\nu(g_1^{-1}s,g_2)) \tag{1.5}$$

and conversely this expression is, up to equivalence and as shown in Ref. 7, the most general solution of the corresponding cohomological problem.

Each observable A is represented in this formalism by mappings from the Borel subsets of the set Γ_A of its possible values (i.e., the spectrum) in the projections (i.e., in the families of projectors) in K,

$$P:\mathscr{B}(\Gamma_A)\to\mathscr{P}(K) \tag{1.6}$$

with two kinds of conditions: some measure theoretical properties as a map on Γ_A as well as a condition of covariance under the representation U (1.2) carried by the state space K (see Ref. 1 for details). In fact these conditions are nothing but a straightforward generalization of the conditions of systems of imprimitivity of Mackey.[4] If K is single Hilbert space, the mappings (1.6) are then the spectral measures of the corresponding self-adjoint operator, and if each Hilbert space is one-dimensional, these mappings correspond to characteristic functions in S. The most general solution for these generalized systems of imprimitivity was found in Ref. 1 to be equivalent to

$$P_s(\Delta) = Q(k(s)\Delta\cap T), \tag{1.7}$$

where $\Delta\in\mathscr{B}(\Gamma_A)$, $\Gamma_A \cong G/H_1$ for some subgroup $H_1\subseteq G$ and $T = H/H\cap H_1 \subseteq \Gamma_A$, and where Q is an usual system of imprimitivity in $\mathscr{H}_{s_0}$ for $H = \mathrm{Stabs}_0$ and for the representation D in (1.2).

The method that we could derive from these general results is then the following. One obtains all state spaces for a given set of observables (defining the physical system under consideration) with a given kinematical symmetry group G (defining the symmetry postulates) if one proceeds as follows.

For each subset of observables $\{B\}\subseteq\{A\}$ with corresponding kinematical symmetry subgroup $H\subseteq G$, find first those (usual projective unitary) representations D which admit usual systems of imprimitivity for each observable B and no other.

For each solution found in this way construct the space K as in (1.1) with $S = G/H$, the representation U as in (1.2) with ζ the canonical action of G on S, and finally with (1.7) for the observables.

The observables found in step 1 are the usual quantal ones as they correspond to self-adjoint operators in separable Hilbert spaces whereas the ones found in step 2 play the role of superselection parameters or, in other words, are of the classical type as they commute with all other ones and have a purely discrete point spectrum.

The calculations themselves are based on the same principles as the short example discussed at the end of Ref. 1, and are quite straightforward applications of standard group theoretical methods and of a simple generalization to projective representations of the arguments used and explained in more detail in Ref. 6. We shall therefore not enter into detailed calculations for the various physical models that we shall consider but we shall rather list the results and discuss these solutions from the physical point of view.

2. SPINLESS PARTICLES

As a first example, let us briefly consider the case of spinless particles (already considered in Ref. 6 to which we refer for more details).

A. The nonrelativistic spinless particles

In this case the physical system is characterized by the observables position $\mathbf{q}$, momentum $\mathbf{p}$, and time t. The set Γ (i.e., the collection of all possible outcomes of all observables) is thus isomorphic to $\mathbb{R}^7$. The kinematical symmetry group G follows from the following postulates. There exists no absolute zero for the position, no absolute zero for the momentum, no absolute zero for the time, and no privileged direction. This group G, that we have called the *Newton group*, is thus generated by the following operation on Γ:

$$
\begin{aligned}
&(\mathbf{q},\mathbf{p},t)\to(\mathbf{q}+\mathbf{a},\mathbf{p},t), \quad \mathbf{a}\in\mathbb{R}^3\\
&(\mathbf{q},\mathbf{p},t)\to(\mathbf{q},\mathbf{p}+\mathbf{w},t), \quad \mathbf{w}\in\mathbb{R}^3\\
&(\mathbf{q},\mathbf{p},t)\to(\mathbf{q},\mathbf{p},t+a^0), \quad a^0\in\mathbb{R}\\
&(\mathbf{q},\mathbf{p},t)\to(\alpha\mathbf{q},\alpha\mathbf{p},t), \quad \alpha\in\mathrm{SO}(3).
\end{aligned}
\tag{2.1}
$$

The group elements can thus be labeled by $(\mathbf{w},a^0,\mathbf{a},\alpha)$ and the group product is given by

$$
\begin{aligned}
&(\mathbf{w},a^0,\mathbf{a},\alpha)(\mathbf{w}',a^{0\prime},\mathbf{a}',\alpha')\\
&\quad(\mathbf{w}+\alpha\mathbf{w}',a^0+a^{0\prime},\mathbf{a}+\alpha\mathbf{a}',\alpha\alpha').
\end{aligned}
\tag{2.2}
$$

This group admits two and only two solutions as state spaces, i.e., two and only two spaces K as in (1.1) that carry an irreducible representation U of this group as well as supersystems of imprimitivity for each of the observables position, momentum, and time.

1. The classical particle

In this solution the superselection variable set S is given by Γ itself, hence the stabilizer of a point $s_0\in S$ is the rotation group and thus

$$
K = V_{\Gamma}\mathscr{H}_{\gamma}, \tag{2.3}
$$

where $\mathscr{H}_{\gamma}$ are all one-dimensional and carry the one-dimensional trivial representation of SO (3). The corresponding representation U of G is directly obtained from (1.2) and the defining representation (2.1) as given by

$$
(U(g)\psi)_{\gamma} = \psi_{g^{-1}\gamma}, \tag{2.4}
$$

where $\psi_{\gamma}\in\mathscr{H}_{\gamma}$ and $\psi=\{\psi_{\gamma}\}$ is any state in K, i.e., $\psi_{\gamma_0}\in U(1)$ for some $\gamma_0=\{\mathbf{p}_0,\mathbf{q}_0,t_0\}$ and 0 otherwise. Correspondingly, and for this state ψ, the observables are simply given by the following characteristic functions [as in (3.23) of Ref. 1]:

$$
\begin{aligned}
&P_{\Delta\mathbf{p}}\psi = \chi_{\Delta\mathbf{p}}(\mathbf{p}_0)\psi,\\
&P_{\Delta\mathbf{q}}\psi = \chi_{\Delta\mathbf{q}}(\mathbf{q}_0)\psi,\\
&P_{\Delta t}\psi = \chi_{\Delta t}(t_0)\psi,
\end{aligned}
\tag{2.5}
$$

with $\Delta\mathbf{p},\Delta\mathbf{q}$, and Δt Borel sets in $\mathbb{R}^3$, $\mathbb{R}^3$, and $\mathbb{R}$, respectively. The solution given by (2.3)–(2.5) can thus in an obvious way be identified with the usual framework of a classical single particle.

2. The quantal particle

In the only other solution, S is isomorphic to the time axis and can thus be identified with the coset space

$$
S\cong G/H_t, \quad H_t=\{(\mathbf{w},0,\mathbf{a},\alpha)\}. \tag{2.6}
$$

The state space K is thus the direct union over this S of Hilbert spaces that carry an irreducible representation of H_t in (2.6). These spaces are $\forall t$ isomorphic to $\mathscr{L}^2(\mathbb{R}^3)$ and thus

$$
K = \bigvee_{t\in\mathbb{R}}(\mathscr{L}^2(\mathbb{R}^3))_t. \tag{2.7}
$$

The representation U of the Newton group G is given by

$$
\begin{aligned}
&U(\mathbf{w})\psi_t(\mathbf{x}) = \exp(i\lambda\mathbf{w}\mathbf{x})\psi_t(\mathbf{x}),\\
&U(\mathbf{a})\psi_t(\mathbf{x}) = \psi_t(\mathbf{x}-\mathbf{a}),\\
&U(a^0)\psi_t(\mathbf{x}) = \psi_{t-a^0}(\mathbf{x}),\\
&U(\alpha)\psi_t(\mathbf{x}) = \psi_t(\alpha^{-1}\mathbf{x}),
\end{aligned}
\tag{2.8}
$$

where λ is an arbitrary nonzero real representation label. It directly follows from (2.8) that

$$
U(\mathbf{w})U(\mathbf{a}) = \exp(i\lambda\mathbf{w}\mathbf{a})U(\mathbf{a})U(\mathbf{w}) \tag{2.9}
$$

so that, identifying λ with the inverse of $\hbar$, we recover the usual Weyl commutation relations. The observables position momentum and time are respectively given by

$$
\begin{aligned}
&P_{\Delta\mathbf{q}}\psi_t(\mathbf{x}) = \chi_{\Delta\mathbf{q}}(\mathbf{x})\,\psi_t(\mathbf{x}),\\
&P_{\Delta\mathbf{p}}\hat{\psi}_t(\mathbf{k}) = \chi_{\Delta\mathbf{p}}(\mathbf{k})\,\hat{\psi}_t(\mathbf{k})\\
&P_{\Delta t}\psi_t(\mathbf{x}) = \chi_{\Delta t}(t)\,\psi_t(\mathbf{x}),
\end{aligned}
\tag{2.10}
$$

with $\hat{\psi}(\mathbf{k})$ the usual Fourier transform of $\psi(\mathbf{x})$. The first two observables in (2.10) can also be written in terms of the usual following operators:

$$
\begin{aligned}
&\hat{\mathbf{q}}\,\psi_t(\mathbf{x}) = \mathbf{x}\,\psi_t(\mathbf{x}),\\
&\hat{\mathbf{p}}\,\psi_t(\mathbf{x}) = -i\hbar\,\partial_{\mathbf{x}}\psi_t(\mathbf{x}).
\end{aligned}
\tag{2.11}
$$

This solution can thus be identified with the usual framework of quantum mechanics, up to the additional presence of time as an observable, due to the slightly more general state spaces K in (1.1), with respect to single Hilbert spaces. Such a generalized state space (1.1) is thus of direct use even for usual quantum physics, for problems, where time appears explicitly. Let us also remark here that this time observable has *a priori* nothing to do with an evolution parameter: it only characterizes the date, the measure of a clock, and not the changes of the system. In fact an evolution parameter, that labels the changes of the state, is not even an observable, as an observable corresponds to a property of the system, hence characterizes a state and not the changes of the properties.

Finally, let us also remark that the representations (2.8) is not characterized by a dynamical type relationship as was the case for the Galilei group approach where[8] the representations are such that

$$
E-(1/2m)\mathbf{p}^2 = \text{const.} \tag{2.12}
$$

This is important because a relation like (2.12) in fact already implies that the theory will strictly speaking only describe free particles.

B. The relativistic spinless particles

In the relativistic case, the observables are the 4-position q^{μ} and the 4-momentum p^{μ}. Hence $\Gamma=\{(q,p)\}\cong\mathbb{R}^8$. As symmetry postulates we assume that there exists no absolute zero for the space-time position, no absolute zero for the momentum (i.e., there exists no frame which is absolutely at rest), and no privileged Lorentz direction. Here again, the kinematical symmetry group is not a symmetry group for the dynamics even for the free particle, but it is a symmetry group for the description of the properties, i.e., of the state of a particle (in interaction or not). The group which is generated by these operations and which has been called the *Einstein group*[6] is thus generated by the following operations:

$$
\begin{aligned}
&(q,p)\to(q+a,p), \quad a\in\mathbb{R}^4\\
&(q,p)\to(q,p+w), \quad w\in\mathbb{R}^4\\
&(q,p)\to(\Lambda q,\Lambda p), \quad \Lambda\in\mathrm{SO}(3,1).
\end{aligned}
\tag{2.13}
$$

The group elements are thus labeled (w,a,Λ) and the group products reads

$$(w,a,\Lambda)(w',a',\Lambda') = (w + \Lambda w', a + \Lambda a', \Lambda\Lambda'). \quad (2.14)$$

Here also, and the distinction has now more important consequences, the time observable is purely kinematical and has thus *a priori* nothing to do with an evolution parameter.

This group admits also two and only two solutions for the state spaces.

1. The classical particle

As in the nonrelativistic case S is given by the phase space Γ itself with the Lorentz group as stabilizer of an arbitrary point and thus

$$K = V_\Gamma \mathscr{H}_\gamma, \quad (2.15)$$

where the $\mathscr{H}_\gamma$ are all one-dimensional and carry the one-dimensional trivial representation of the Lorentz group. The corresponding representation U of the Einstein group is directly given from the defining representation (2.13) and reads, with $\gamma = (q,p) \in \Gamma$ and $g = (w,a,\Lambda)$,

$$(U(g)\psi)_{q,p} = \psi_{\Lambda^{-1}(q-a),\Lambda^{-1}(p-w)}, \quad (2.16)$$

with $\psi_{q_0,p_0} \in U(1)$ for some (q_0,p_0) and zero otherwise. Each point of Γ thus defines exactly one state [one ray in $U(1)$] and this makes the identification with the classical phase space as state space, i.e., with the usual classical framework, complete. The observables position and momentum are correspondingly given as previously in (2.5) by characteristic functions [as again follows from (3.23) of Ref. 1]:

$$P_{\Delta q}\psi_\gamma = \chi_{\Delta q}(q_0)\psi_\gamma,$$
$$P_{\Delta p}\psi_\gamma = \chi_{\Delta p}(p_0)\psi_\gamma, \quad (2.17)$$

with Δq and Δp Borel sets in $\mathbb{R}^4$.

2. The quantal particle

In this solution S is a singleton, i.e., the direct union (1.1) is trivial and K is a single Hilbert space given by

$$K = \mathscr{L}^2(\mathbb{R}^4), \quad (2.18)$$

with the following representation of the Einstein group:

$$U(w)\psi(x) = \exp[i\hbar^{-1} g_{\mu\nu} w^\mu x^\nu]\psi(x),$$
$$U(a)\psi(x) = \psi(x-a),$$
$$U(\Lambda)\psi(x) = \psi(\Lambda^{-1}x), \quad (2.19)$$

where $g_{\mu\nu}$ has signature $(-,+,+,+)$ and where $\hbar^{-1}$ here also appears as a representation label. The representation (2.19) satisfies the following generalized Weyl commutation relation:

$$U(w)U(a) = \exp[i\hbar^{-1} g_{\mu\nu} w^\mu a^\nu] U(a)U(w). \quad (2.20)$$

The observables position and momentum are, respectively, given by the following spectral families:

$$P_{\Delta q}\psi(x) = \chi_{\Delta q}(x)\psi(x),$$
$$P_{\Delta p}\hat{\psi}(k) = \chi_{\Delta p}(k)\hat{\psi}(k), \quad (2.21)$$

where the characteristic functions are as in (2.17) and where

$$\hat{\psi}(k) = (2\pi\hbar)^{-2} \int e^{-i\hbar^{-1} g_{\mu\nu} k^\mu x^\nu} \psi(x) d^4x. \quad (2.22)$$

Equivalently these observables are given by the self-adjoint extensions of the operators

$$\hat{q}^\mu\psi(x) = x^\mu\psi(x),$$
$$\hat{p}^\mu\psi(x) = -i\hbar g^{\mu\nu}\partial_\nu\psi(x). \quad (2.23)$$

In this model, and in contradistinction to the usual case, the space-time position operators always exist and their (purely continuous) spectra are simply given by the space-time variables.

On the other side we can also remark here that the above representations are, analogously to (2.12), not characterized by an invariant of the form $p^2 = m^2$. We may thus here also hope that the corresponding dynamics will allow description of more than only free particles.

Finally, let us also remark that the states are not trajectories in space-time but, as elements in $\mathscr{L}^2(\mathbb{R}^4)$, really correspond to the events of Einstein. This is important because if the states are the trajectories, the dynamics only list these trajectories but nothing really changes, nothing really moves along these trajectories. Moreover it is difficult to speak of (quantum) probabilities in such a scheme (see also Refs. 2 and 9 on these aspects). In our formalism, however, the states are really capable of motion, hence of evolution, in space-time.

3. PARTICLES WITH SPIN

A. Nonrelativistic particles

The introduction of a new observable for the spin in the previously obtained nonrelativistic models, both classical and quantal, is very straightforward. This is so because in both cases the rotation group appears in the stabilizer of any point in the superselection parameter set S. We obtain for the classical particle

$$\psi_\gamma \in \mathbb{C}^{2\sigma+1}, \quad K = V_\Gamma(\mathbb{C}^{2\sigma+1})_\gamma \quad (3.1)$$

and the representation is the same as in (2.4) except for

$$(U(\alpha)\psi)_\gamma = (D^{(\sigma)}(\alpha)\psi)_{\alpha^{-1}\gamma}, \quad (3.2)$$

where $D^{(\sigma)}$ is an irreducible spin σ representation of the rotation group. This shows that the classical particle is compatible with an observable spin or, in other words, that it is possible to consider models where only the spin has a quantal behavior whereas the position and the momentum are both treated as classical observables.

The quantal solution can be generalized in the same way and we obtain in place in (2.7)

$$K = \int_{t\in\mathbb{R}} (\mathscr{L}^2(\mathbb{R}^3,\mathbb{C}^{2\sigma+1}))_t, \quad (3.3)$$

the representation being as in (2.8) except for

$$(U(\alpha)\psi(\mathbf{x}))_t = (D^{(\sigma)}(\alpha)\psi(\alpha^{-1}\mathbf{x}))_t. \quad (3.4)$$

These solutions (3.1) and (3.3) for the state spaces are, here also, the two only ones.

We can of course ask ourselves in which way it is possible to express the observable spin in a language completely similar to the one of the observables position and momentum, i.e., in terms of systems of imprimitivity (cf. Ref. 1). It turns out that this is possible if one observes that the corresponding measuring (Stern–Gerlach) apparatus in fact breaks the rotation invariance and if we consequently postu-

late that the restriction of the representation $U(G)$ to the subgroup $H_{\mathbf{n}}$ that leaves an arbitrary space direction $\mathbf{n}$ invariant admits a system of imprimitivity based on the spin spectrum [i.e., on a $(2\sigma+1)$ point discrete set]. We shall not detail more this aspect here, just mentioning the quite remarkable fact that the only corresponding systems of imprimitivity which are then *transitive*, hence elementary, are for the particles of spins 0 or $\frac{1}{2}$ (i.e., corresponding to the only massive particles which are considered as stable).

B. Relativistic particles

It is obvious that the just explained generalization does not apply to the relativistic case, as the relevant isotropy groups that appear both in the classical and in the quantal models contain the full Lorentz group whose nontrivial irreducible projective unitary representations are all ∞-dimensional, as is well known.

This difficulty can be avoided if one observes that in each inertial system the measure of the direction of the time axis commutes with all other observables. This is not in contradiction with the principle of relativity: the mixing of time and position observable values under changes of reference frames does not hinder that time has (trivially), in each reference frame, a role which is distinct to one of the space coordinates. Time is in fact measured with a clock, space position with a measuring rod. The measuring devices themselves reflect this essential difference. We can formulate these considerations as follows.

Postulate: In each Lorentz reference frame the *direction of the time axis* is a superselection parameter.

The direction of the time axis can be characterized by a unit vector $\hat{n}$ inside of the light cone, i.e., by a point in a unit upper half hyperboloid in space-time. It follows then from the above postulate that the classical and quantal states spaces can be decomposed as

$$K = V_{\hat{n}} K_{\hat{n}}, \tag{3.5}$$

the direct union being taken over all possible time directions, i.e., over the above mentioned upper half hyperboloid.

Each $K_{\hat{n}}$ should thus carry (as follows from Ref. 1) an irreducible representation of the stabilizer of the corresponding value of the superselection parameter. These stabilizers are given, up to isomorphism and for an arbitrary unit vector $\hat{n}$, by

$$H_{\hat{n}} \cong \{w, a, \Lambda_{\hat{n}_0}\}, \tag{3.6}$$

where $(w,a) \in \mathbb{R}^4 \times \mathbb{R}^4$ and $\Lambda_{\hat{n}_0}$ stands for the Lorentz transformations which stabilize the unit vector $\hat{n}_0 = (1,0,0,0)$, i.e., for the rotations.

Let us remark here that a quite different interpretation based on the Stern–Gerlach measuring device itself has previously been given[9,10] with, however, exactly the same mathematical consequences.

Under the above assumptions our formalism again leads in a unified way to the allowable state spaces and we again find a classical solution and a quantal one. For the classical particle we obtain

$$K_{\hat{n}} = V_{\Gamma}(\mathbb{C}^{2\sigma+1})_{\gamma} \tag{3.7}$$

and the representation of the Einstein group is given, for $\psi = \{\psi_{\hat{n},\gamma} | \psi_{\hat{n},\gamma} \in \mathbb{C}^{2\sigma+1}, \gamma = (q,p) \in \mathbb{R}^8\}$, by

$$\begin{aligned}
&(U(w)\psi)_{\hat{n},q,p} = \psi_{\hat{n},q,p-w},\\
&(U(a)\psi)_{\hat{n},q,p} = \psi_{\hat{n},q-a,p},\\
&(U(\Lambda)\psi)_{\hat{n},q,p} = D^{(\sigma)}(L(\hat{n})^{-1}\Lambda L(\Lambda^{-1}\hat{n}))\,\psi_{\Lambda^{-1}\hat{n},\Lambda^{-1}q,\Lambda^{-1}p},
\end{aligned} \tag{3.8}$$

where $L(\hat{n})$ is some coset representative which maps the vector $\hat{n}_0$ onto the vector $\hat{n}$ and the spinor rotation is thus nothing else than the well known Wigner rotation (with σ arbitrary).

Analogously, the quantal particle is given by the same (3.5) with now, in place of (3.7),

$$K_{\hat{n}} = \mathscr{H}_{\hat{n}} = (\mathscr{L}^2(\mathbb{R}^4, \mathbb{C}^{2\sigma+1}))_{\hat{n}}. \tag{3.9}$$

The representation U for (any) spin σ is given by

$$\begin{aligned}
&(U(w)\psi(x))_{\hat{n}} = (\exp[i\hbar^{-1} g_{\mu\nu} w^{\mu} x^{\nu}]\psi(x))_{\hat{n}},\\
&(U(a)\psi(x))_{\hat{n}} = (\psi(x-a))_{\hat{n}},\\
&(U(\Lambda)\psi(x))_{\hat{n}} = (D^{(\sigma)}(L(\hat{n})^{-1}\Lambda L(\Lambda^{-1}\hat{n}))\,\psi(\Lambda^{-1}x))_{\Lambda^{-1}\hat{n}}.
\end{aligned} \tag{3.10}$$

The observables p and q are as before, whereas the observable spin can be expressed in a straightforward way using the usual spin matrices (see Ref. 9 for details).

We shall mention some remarks concerning the physics of these models in Sec. 5. For the moment we want to present other examples of applications of our formalism.

4. GALILEAN BOOSTS AND NONRELATIVISTIC LIMITS

One of the (ultimate) goals of the present analysis is of course to understand better (and thus to solve) some of the well known difficulties of relativistic quantum dynamics and of field theories. With respect to this goal, it is of course an important advantages of our approach that, as we have seen, we have now really a unified mathematical formalism for the discussion and the interpretation of the physical quantities under consideration: as already emphasized, both the mathematical description and the interpretation of the physical objects is the same in all considered models, relativistic or not, classical or quantal. We can take advantage of this fact and compare effectively the relativistic quantum model with its classical and its nonrelativistic limits, and with the corresponding classical and nonrelativistic models, respectively. For that purpose we, however, obviously have to enlarge the observable space Γ in (2.1) with the observable energy E and, correspondingly to (2.13), with the transformations[11]

$$E \to E + w^0, \quad w_0 \in \mathbb{R} \tag{4.1}$$

and the Galilean boosts

$$\begin{aligned}
&\mathbf{q} \to \mathbf{q} - \mathbf{v}t,\\
&E \to E - \mathbf{p}\mathbf{v}, \quad \mathbf{v} \in \mathbb{R}^3.
\end{aligned} \tag{4.2}$$

We can apply in a straightforward way our formalism to this new example, and obtain again all possible state spaces for the corresponding physical system. Dropping again the detailed calculations, we find here also two and only two classes of solutions: a classical one with K the direct union over the 8-dimensional phase space $\{(t,\mathbf{q},E,\mathbf{p})\}$ of $(2\sigma+1)$-dimensional Hilbert spaces, and more interesting at this

point, a quantal solution: we find for the latter a single direct integral Hilbert space

$$K = \int^{\oplus} \mathscr{H}(\Delta)\, d^4x, \tag{4.3}$$

with $\mathscr{H}(\Delta)$ the carrier space of an arbitrary representation Δ of the homogeneous Galilei group, i.e., either an infinite dimensional representation on $\mathscr{L}^2$ of the two-dimensional sphere, or an usual spin representation[8]:

$$\Delta(\mathbf{v},\alpha) = D^{(\sigma)}(\alpha). \tag{4.4}$$

In this latter case, the representation is given in $\mathscr{L}^2(\mathbb{R}^4, \mathbb{C}^{2\sigma+1})$ by

$$\begin{aligned}
&(U(w)\psi)(x) = \exp[i\hbar^{-1}(\mathbf{w}\mathbf{x} - w^0 t)]\psi(x),\\
&(U(a)\psi)(x) = \psi(x-a),\\
&(U(\alpha)\psi)(x) = D^{(\sigma)}(\alpha)\,\psi(t,\alpha^{-1}\mathbf{x}),\\
&(U(\mathbf{v})\psi)(x) = \psi(t,\mathbf{x}+\mathbf{v}t).
\end{aligned} \tag{4.5}$$

Comparing this solution with the (spinless) relativistic model given in (2.18) and (2.19) one sees that the state spaces are (for $\sigma = 0$) essentially the same; only the groups and the group actions are different and this makes the comparison easily possible. In contradistinction with the relativistic case, in the case $\sigma \neq 0$, the spin enters here in the scheme in a more direct way. The observables position and momentum are, in this model, as before [see (2.11)] and for t and E, they are given by

$$\begin{aligned}
&\hat{t} = t\,\mathbb{1}_K,\\
&\hat{E} = i\hbar\,\partial_t.
\end{aligned} \tag{4.6}$$

The observables are thus also the same as in the relativistic case.

There is another possibility which is directly suggested by the corresponding relativistic problem: it is to consider, as in our previous results, the *direction of the time axis* as a superselection variable. This is even more natural if one realizes that there is in fact nothing typically relativistic in this assumption.

If we adopt this point of view, then, remarking that the variable t is by definition the time *coordinate* of a given event in space-time, we can identify the superselection variable set with the three dimensional vector space V of the pure Galilean boosts. We can correspondingly define in the space-time a three dimensional (horizontal) surface of coordinates $(1,\mathbf{n})$, where each $\mathbf{n}$ corresponds to a point in V. We then find for the state space

$$K = V_{\mathbf{n}\in\mathbb{R}^3} K_{\mathbf{n}} \tag{4.7}$$

where $K_{\mathbf{n}}$ carries an irreducible representation of the stabilizer of the vector $\mathbf{n}$. The action of the group on this space is easily found to be given by, for $g = (a,w,\mathbf{v},\alpha)$,

$$g\cdot\mathbf{n} = \alpha\mathbf{n} + \mathbf{v} \tag{4.8}$$

and the stabilizer $H_{\mathbf{n}}$ of the vector $\mathbf{n}$ is thus, up to isomorphism, given by

$$H_{\mathbf{n}} \cong \{(a,w,\mathbf{0},\alpha)\}. \tag{4.9}$$

In the more interesting quantal case, we find in the same way as before the following solution for the group representation: each $K_{\mathbf{n}}$ is isomorphic to $\mathscr{L}^2(\mathbb{R}^4,\mathbb{C}^{2\sigma+1})$ and the representation is given by

$$\begin{aligned}
&[U(w)\psi(x)]_{\mathbf{n}} = [\exp[i\hbar^{-1}(\mathbf{w}\mathbf{x} - w^0 t)]\psi(x)]_{\mathbf{n}},\\
&[U(a)\psi(x)]_{\mathbf{n}} = [\psi(x-a)]_{\mathbf{n}},\\
&[U(\mathbf{v})\psi(x)]_{\mathbf{n}} = [\psi(t,\mathbf{x}+\mathbf{v}t)]_{\mathbf{n}-\mathbf{v}},\\
&[U(\alpha)\,\psi(x)]_{\mathbf{n}} = [D^{(\sigma)}(\alpha)\,\psi(t,\alpha^{-1}\mathbf{x}]_{\alpha^{-1}\mathbf{n}}.
\end{aligned} \tag{4.10}$$

In other words, we have the same situation as before: the state spaces are essentially the same as in the relativistic case, only the groups and the group actions are different, whereas the hyperboloid of (3.5) is degenerated in a (horizontal) plane in space-time. Moreover, one sees in (4.10) that the Wigner rotations of (3.10) are now simply reduced to the rotation parts of the corresponding group elements. The observables are also the same as in the relativistic solution.

Let us finally remark here also that, if in place of (4.2) we had assumed that the energy was left unchanged by a Galilean boost, i.e., if we had considered the energy as a scalar quantity in place of considering it as the fourth component of a vector, we would have obtained in the quantal case a state space given by

$$K = V_{E,t}(\mathscr{L}^2(\mathbb{R}^3, \mathscr{H}(\Delta))_{E,t}, \tag{4.11}$$

with $\mathscr{H}(\Delta)$ as in (4.3) and the representation is, in the simplest spinless case, as in (2.8) with, in addition,

$$\begin{aligned}
&(U(w^0)\,\psi(\mathbf{x}))_{E,t} = (\psi(\mathbf{x}))_{E+w^0,t},\\
&(U(\mathbf{v})\,\psi(\mathbf{x}))_{E,t} = (\psi(\mathbf{x}+\mathbf{v}t))_{E,t}.
\end{aligned} \tag{4.12}$$

The observables $\mathbf{p}$ and $\mathbf{q}$ are thereby as in (2.11) and the observables E and t are as in the classical models. This last state space seems, however, less interesting when one considers the dynamics.

5. ON THE DYNAMICS

As previously emphasized, the framework of the present paper is purely kinematical, in the sense that the construction of the state spaces, the state spaces themselves, and the observables are defined independently of the dynamics. We have precisely exhibited and used those aspects of the concept of an elementary physical system which are independent of the interactions and of the evolution. We insist on this point, because it is one of our main motivations to construct a (group-theoretical) formalism in which it is possible to describe more than only free particles.

Although it is thus beyond the scope of this paper, we would like, for clarifying these points of view, to briefly sketch how the dynamics enter into the framework discussed in the present paper, from a very general point of view (see also for this discussion Ref. 12).

The evolution corresponds by definition to the changes of the states, hence it is specified (if reversible) by a two-parameter family of automorphisms of the state spaces. For specifying the dynamics we thus have first to introduce the parameter itself, an evolution parameter, that we call τ, the dynamical time. Such a parameter cannot be itself an observable as explained previously and it cannot thus enter into the dependence of the generator of the evolution. Hence the two-parameter family reduces to a one-parameter group. Togeth-

er with some differentiability conditions we are thus led, in the space K (1.1) of states, to the following generalized Schrödinger equations coupled with classical evolution equations in the superselection variable set S:

$$\dot{s} = X(s),$$
$$i\,\partial_\tau \psi_s = H_s \psi_s, \tag{5.1}$$

where the dot means differentiation with respect to τ, $\{H_s\}$ is a family of self-adjoint operators and X is a vector field on S. For illustration and corresponding to the particle given in (2.7) we may have (see Ref. 2)

$$t = 1,$$
$$i\partial_\tau \psi_t = i\partial_t \psi_t = H_t \psi_t, \tag{5.2}$$

where H_t is some usual Schrödinger Hamiltonian in each of the Hilbert spaces of the family (2.7). Correspondingly, in the here more interesting relativistic case, the evolution is characterized by a single Hamiltonian in the case (2.18) or by a family of Hamiltonians in the case with spin, the latter being coupled with a classical evolution equation for the parameter $\hat{n}$. Examples of this kind of relativistic Hamiltonians acting on the state spaces discussed in the present paper have been extensively studied in recent years (see, for examples, Refs. 9–13). In particular such Hamiltonians have been proposed for the interaction of single particles with external electromagnetic fields (e.g., for the spectrum of the Coulomb field problem[13c]) but also for the description of two-body interactions (in particular for the two-body H atom or for the positronium resonances [13d] with, in all considered cases, a very good agreement with the experimental data. As is well known, the no-go theorem of Currie[14] does not allow consideration of this last kind of problems in the usual framework.

[1]N. Giovannini, companion paper (I) J. Math. Phys. **22**, 2389 (1981).
[2]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, Mass, 1976).
[3]G. W. Mackey, *The Theory of Unitary Group Representations* (University of Chicago Press, Chicago, 1976).
[4]G. W. Mackey, Proc. Math. Acad. Sci. USA **35**, 537 (1949).
[5]L. Martinez Alsonso, J. Math. Phys. **20**, 219 (1979), and references therein.
[6]N. Giovannini and C. Piron, Helv. Phys. Acta **52**, 518 (1979).
[7]N. Giovannini, Lett. Math. Phys. **5**, 161 (1981).
[8]J. M. Levy Leblond, in *Group Theory and its Applications* Vol. 2, edited by E. Loebl (Academic, New York, 1971), p. 221.
[9]F. Reuse, Found. Phys. **9**, 865 (1979).
[10]L. P. Horwitz, C. Piron, and F. Reuse, Helv. Phys. Acta **48**, 546 (1975).
[11]C. Piron, "Particles, dynamics and covariance," preprint UGVA-DPT 1979/11-223 (to be published in the Proceedings of the Colloquium of Mathematical Physics at Trabzon and Istambul, 1979).
[12]C. Piron, "New quantum mechanics," preprint UGVA-DPT (1980/10-265 [to be published in *Essays in Honour of W. Yourgrau* (Plenum, New York)].
[13](a) L. P. Horwitz and C. Piron, Helv. Phys. Acta **46**, 316 (1973); (b) C. Piron and F. Reuse, *ibid.* **51**, 146 (1978); (c) F. Reuse, *ibid.* **51**, 157 (1978); (d) *ibid.* **53**, 416 and 552 (1980).
[14]D. G. Currie, J. Math. Phys. **4**, 1470 (1963).

Noncomplex representations and their relation to antiunitary symmetry

P. Kasperkovitz and G. Kahl

Institut für theoretische Physik, TU Wien, A-1040 Karlsplatz 13, Vienna, Austria
(Received 16 July 1980; accepted for publication 5 September 1980)

Equivalent eigenvalue problems in real, complex, and quaternionic Hilbert spaces are discussed. In all cases a compact symmetry group is assumed to exist, in the complex case also an antiunitary operator Θ commuting with both the self-adjoint operator and the elements of the symmetry group and satisfying $\Theta^2 = \pm 1$. It is shown that the extra degeneracies caused by the existence of Θ can also be obtained by consideration of an equivalent eigenvalue problem in a noncomplex Hilbert space.

PACS numbers: 02.20. + b, 03.65.Fd

1. INTRODUCTION

The purpose of this paper is to show that representations of groups in real and quaternionic Hilbert spaces can be as useful as the familiar complex representations and are by no means more difficult to handle. The study of these representations was initiated by the following example: In lattice dynamics one is given a real symmetric matrix (the matrix of force constants) the elements of which appear as coefficients in equations of motion for the basic quantities $\mathbf{p}_i$ (momenta) and $\mathbf{y}_i$ (displacements of lattice constituents). Although this set of linear differential equations of first order can be solved as it stands it is desireable to pass from the $\mathbf{p}$'s and $\mathbf{y}$'s to real linear combinations $\mathbf{r}_j$ and $\mathbf{z}_j$ (normal modes) and to consider a general solution as superposition (= real linear combination) of the $\mathbf{r}$'s and $\mathbf{z}$'s, which undergo harmonic motions. To find the proper linear combinations $\mathbf{r}_j$, $\mathbf{z}_j$ and the corresponding frequencies one has to diagonalize the real symmetric matrix, which amounts to finding a complete set of eigenvectors (= real column vectors) or the corresponding (real) orthogonal matrix. This program is greatly simplified by exploiting the results of representation theory since there exists a group of orthogonal matrices (reflecting the symmetry of the crystal) which commute with the matrix to be diagonalized. Assuming the eigenvalue problem to be already solved it becomes obvious that all maniupulations to simplify the eigenvalue problem and to explain the degeneracies of the eigenfrequencies should be performable in the real Hilbert space consisting of the real column vectors. But to apply the machinery of group theory as it is known, for instance, from electron-band theory a detour through the complex Hilbert space consisting of complex column vectors is necessary. The results obtained using complex representations of the symmetry group are then (i) a set of eigenvectors which are in general complex; and (ii) a relation between the minimal degeneracies of the eigenfrequencies and the dimensions of the complex irreducible matrix representation. To obtain real eigenvectors (the only ones admitting a direct physical interpretation) and a more satisfactory explanation of the degeneracies, complex conjugation is introduced as an antilinear operator commuting with the real symmetric matrix. If this extra symmetry is considered along with the one reflecting the crystal's symmetries one finds (i) that the eigenvectors can be chosen to be real (which was already clear from the beginning); and (ii) that the minimal degeneracies are doubled in some instances.[1] This two-step procedure originating from the classic paper of Maradudin and Vosko[2] is now standard in the literature on lattice dynamics[3] and is also found in the computer-adapted version of the theory.[4] But it is by no means the only way to the desired results. For, as has been shown elsewhere[5] for two specific examples and is described in full detail in this paper, the theory of group representations in real Hilbert spaces leads directly to real eigenvectors and the minimal degeneracies (being equal to the dimension of matrix representations irreducible over the reals), and all this is achieved by means of linear operators only.

The approach suggested here is not only more direct than the usual one but also more satisfactory regarding the foundation of the minimal degeneracies. The reason for this is that although Maradudin and Vosko denoted the operation of complex conjugation as "time reversal symmetry" on *formal* grounds only[6] this term was taken literally by subsequent workers in this field.[3,4] However, a short reflection shows the only meaning of this symmetry to be that a complex linear combination of the $\mathbf{p}$'s and/or $\mathbf{y}$'s satisfies the *same* equation of motion as the complex conjugate linear combination irrespective of whether the basic quantities $\mathbf{p}_i$, $\mathbf{y}_i$ are real-valued functions (classical mechanics) or operators or their expectation values (quantum mechanics). It is only in a related problem, namely the solution of the Schrödinger equation for the lattice constituents, that the term time reversal may be appropriate. But as far as the eigenvalue problem of the Hamilton operator is concerned this symmetry only means that the eigenvectors can be chosen to be real and that the degeneracies of the eigenvalues are again related to the real irreducible representations of the symmetry group. (This result can also be obtained restricting the complex Hilbert space to a *real* one; see Secs. 3A and 3C). It is only the part of the eigenvalue problem belonging to the subspace of 1-phonon-states, which are equivalent to diagonalizing the matrix of the force constants; more-phonon-states transform according to the symmetrized powers of Kronecker products of the (real) representations found there.

In the following we also discuss eigenvalue problems and group representations in quaternionic Hilbert spaces. As in the real case, the corresponding prob-

 0022-2488/81/112404-13$01.00

lem in a complex Hilbert space is characterized by the existence of an antiunitary operator θ commuting with the self-adjoint operator and all the elements of the symmetry group, but now satisfying $\theta^2=-1$ instead of $\theta^2=+1$ (real case). The mere fact that for $\theta^2=-1$ there exists an isomorphism between the operators in the complex Hilbert space and the operators in a quaternionic Hilbert space of half the dimension of the complex one entails a doubling of degeneracies for the complex eigenvalue problem and is hence a variant of Kramers' degeneracy theorem.[1]

The existence of isomorphic operator algebras in real, complex, and quaternionic Hilbert spaces, which is basic to our analysis, is not a new result. The cases, where one of the Hilbert spaces is real (cases 3 and 4 of Sec. 2) were treated by Stueckelberg and Guenin[7] and the relations between operators in complex and quaternionic Hilbert spaces (case 2 of Sec. 2) have been noticed by Chevally[8] and extensively discussed by Jauch[9] and collaborators. The main difference between their investigations and the following discussion is that they were interested in the foundations of quantum theory whereas we use this equivalence to simplify eigenvalue problems and to relate the degeneracies found there to representations of the symmetry group. A similar remark applies to group representations in a quaternionic Hilbert space. The reduction of a complex representation of the second kind to two identical quaternionic representations appearing here as one of the intermediate steps (see Secs. 5 and 6B) has been given by Finkelstein *et al.*[10] (using a related result of Frobenious and Schur[11]) but this result has not been used in any application. Apart from its different intention, the present paper extends the above cited results and the traditional approach in the following respects: (i) A sufficient condition is given to guarantee the existence of the operator isomorphism for infinite-dimensional Hilbert spaces ("densely defined") operator algebras (see Secs. 2C and 6A); (ii) the degeneracies are related to representations of the real group algebra in real and quaternionic Hilbert spaces; and (iii) in some of the remaining eigenvalue problems the number of real parameters is shown to be half the number appearing in the usual approach (this might be of interest for numerical calculations).

We tried to write this paper in a way that makes it readable to the nonspecialist. To this end, the general theory of noncomplex group representations was deferred to a companion paper[13] and only that part which is of relevance here was retained. The same purpose serves the detailed summary of the basic mathematical concepts (fields, Hilbert spaces, operators and their matrix representations) given in the next section. In Sec. 3 the equivalence of eigenvalue problems is stated and illustrated by well-known examples (real Hamiltonians, spin-orbit interactions). In Sec. 4 we show that the proper tool to simplify an eigenvalue problem in a real or quaternionic Hilbert space and to explain the degeneracies found there is the real group algebra. How to obtain the projection and shift operators belonging to this algebra from complex irreducible matrix representations is described in Sec. 5. The necessary proofs, all being of elementary algebraic nature, are given in Sec. 6 and our results are summarized in Sec. 7.

2. BASIC CONCEPTS AND NOTATION

A. Fields

The fields considered here are the *reals* $\mathbb{R}$, the *complex numbers* $\mathbb{C}$ and the *quaternions* $\mathbb{Q}$ (sometimes also called hypercomplex numbers). It is convenient to view every field $\mathbf{F}$ as an extension of $\mathbb{R}$. $\mathbf{F}$ is then the set of elements

$$f=\sum_{r=0}^{\rho-1} f_r a_r, \quad a_r\in\mathbb{R}. \tag{2.1}$$

ρ is called the *rank* of $\mathbf{F}$ over $\mathbb{R}$ and the sum and the product are defined by

$$\begin{aligned} f+f' &= \sum_{r=0}^{\rho-1} f_r(a_r+a_r'), \\ ff' &= \sum_{r,s=0}^{\rho-1} f_{rs}\sigma_{r,s}a_r a_s, \end{aligned} \tag{2.2}$$

where the product $rs=sr$ is one of the numbers $0,\ldots,\rho-1$ and $\sigma_{r,s}$ the r,s-element of a real matrix σ. The product rs and the matrix σ are specified below for the different fields. Equation (2.2) implies that the units f_r multiply according to $f_r f_s=f_{rs}\sigma_{r,s}$. If $\mathbf{F}=\mathbb{C}$ or $\mathbb{Q}$ we sometimes use special symbols for the elements, especially the units f_r.

$\mathbf{F}=\mathbb{R}$: $\rho^{\mathbb{R}}=1$; $f_0=1$;

$\sigma=1$; multiplication table $\begin{array}{c|c} & 0 \\ \hline 0 & \boxed{0} \end{array}$. (2.3)

$\mathbf{F}=\mathbb{C}$: $\rho^{\mathbb{C}}=2$; $f_0=1, f_1=i$;

general element $c=a+ib$ $(a,b\in\mathbb{R})$;

$\sigma^{\mathbb{C}}=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$; multiplication table $\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \boxed{0} & \boxed{1} \\ 1 & \boxed{1} & \boxed{0} \end{array}$. (2.4)

$\mathbf{F}=\mathbb{Q}$: $\rho^{\mathbb{Q}}=4$; $f_0=1, f_1=i, f_2=j, f_3=k$;

general element $q=a+ib+jc+kd$ $(a,b,c,d\in\mathbb{R})$

$$\sigma^{\mathbb{Q}}=\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & -1 & 1 & -1\\ 1 & -1 & -1 & 1\\ 1 & 1 & -1 & -1\end{pmatrix}; \text{ multiplication table } \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & \boxed{0} & \boxed{1} & \boxed{2} & \boxed{3} \\ 1 & \boxed{1} & \boxed{0} & \boxed{3} & \boxed{2} \\ 2 & \boxed{2} & \boxed{3} & \boxed{0} & \boxed{1} \\ 3 & \boxed{3} & \boxed{2} & \boxed{1} & \boxed{0} \end{array}. \tag{2.5}$$

$\mathbb{R}$ and $\mathbb{C}$ are commutative fields but $\mathbb{Q}$ is not ($qq' \neq q'q$ in general).

For each field $\mathbb{F}$ a mapping from $\mathbb{F}$ onto $\mathbb{F}$, called *conjugation*, is defined by

$$f^* = \sum_{r=0}^{\rho-1} f_r \sigma_{r,r} a_r \,. \tag{2.6}$$

The mapping $f \to f^*$ is an antiautomorphism since $(f+f')^* = f^* + f'^*$, $(ff')^* = f'^* f^*$. The modulus $|f|$ of an element $f \in \mathbb{F}$ is defined by

$$|f|^2 = ff^* = f^* f, \quad |f| \geq 0, \tag{2.7}$$

and the elements f with $|f| = 1$ are called *unimodular*.

Besides the antiautomorphism $f \to f^*$ we will also meet *automorphisms* $f \to f^\alpha$. These are mappings from $\mathbb{F}$ onto $\mathbb{F}$ such that $(f+f')^\alpha = f^\alpha + f'^\alpha$, $(ff')^\alpha = f^\alpha f'^\alpha$. For $\mathbb{R}$ there exists only the trivial automorphism $f \to f$. For $\mathbb{C}$ there is one nontrivial automorphism, namely the conjugation $f \to f^*$. It is called an *outer* automorphism since it is not possible to find a $g \in \mathbb{C}$ such that $f^* = gfg^{-1}$. For $\mathbb{Q}$ the situation is quite different: There exist infinitely many automorphisms and all of them are *inner* automorphisms of the form $f \to qfq^{-1} = ufu^{-1}$, $u = |q|^{-1} q$. If the automorphism $f \to f^\alpha$ is given the unimodular number u is determined up to a factor ± 1. The field $\mathbb{Q}$ contains subsets that can be identified with the fields $\mathbb{C}$ and $\mathbb{R}$. Whereas $\mathbb{R}$ is uniquely determined as a subfield of $\mathbb{Q}$ or $\mathbb{C}$, $\mathbb{Q}$ contains an infinity of subfields isomorphic to $\mathbb{C}$, namely the sets $\{a + ub : a, b \in \mathbb{R}, u \in \mathbb{Q}, -u = u^{-1}\}$. Since this fact is irrelevant for our purpose we shall always consider the chain $\mathbb{R} \subset \mathbb{C} \subset \mathbb{Q}$ with $\mathbb{C}$ fixed ($u = i$). We finally note that for $f_1, f_2, q \in \mathbb{Q}$,

$$\begin{aligned} &q = f_1 q f_1^{-1} = f_2 q f_2^{-1} \Leftrightarrow q \in \mathbb{R}, \\ &q = f_1 q f_1^{-1} \Leftrightarrow q \in \mathbb{C}. \end{aligned} \tag{2.8}$$

In the following frequent use will be made of the fact that the elements $f' \in \mathbb{F}'$ can be represented by matrices $F^{\mathbb{F}'}[f']$ with elements from a subfield $\mathbb{F}' \subseteq \mathbb{F}'$. The mapping $f' \to F^{\mathbb{F}'}[f']$ is an isomorphism since

$$\begin{aligned} &F^{\mathbb{F}'}[f_1' + f_2'] = F^{\mathbb{F}'}[f_1'] + F^{\mathbb{F}'}[f_2'], \\ &F^{\mathbb{F}'}[f_1' f_2'] = F^{\mathbb{F}'}[f_1'] \cdot F^{\mathbb{F}'}[f_2']; \\ &F^{\mathbb{F}'}[f'^*] = F^{\mathbb{F}'\dagger}[f'], \quad \text{where } (F^{\mathbb{F}'\dagger})_{rs} = (F^{\mathbb{F}'}_{sr})^*; \\ &F^{\mathbb{F}'}[f'] = 0\text{-matrix} \Leftrightarrow f' = 0. \end{aligned} \tag{2.9}$$

Apart from the trivial representation ($\mathbb{F} = \mathbb{F}'$)

$$F^{\mathbb{F}}[f] = f, \tag{2.10}$$

the following representations will be used ($\mathbb{F} \subset \mathbb{F}'$):

$$C^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix}, \tag{2.11}$$

$$R^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}, \tag{2.12}$$

$$R^{\mathbb{C}}[a + ib] = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \tag{2.13}$$

Equations (2.11)–(2.13) are by no means the only possibility to represent the elements of $\mathbb{F}'$ by matrices over a subfield $\mathbb{F}$. In Eq. (2.11) the basic quaternions i, j, k are identified with $i\sigma_1$, $i\sigma_2$, $i\sigma_3$ where the σ's are the Pauli matrices, but other conventions would work equally well (e.g., $-i\sigma_1$, $-i\sigma_2$, $-i\sigma_3$). For $\mathbb{F} = \mathbb{R}$, $\mathbb{F}' = \mathbb{Q}$ we shall also use the representation

$$\bar{R}^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix}. \tag{2.14}$$

The representations (2.12) and (2.14) are related by a remarkable property: If R' is a 4×4 matrix then

$$[R', \bar{R}^{\mathbb{Q}}(q)] = 0 \text{ for all } q \in \mathbb{Q} \Leftrightarrow R' = R^{\mathbb{Q}}(q') \text{ for some } q' \in \mathbb{Q}. \tag{2.15}$$

The statement remains valid if the symbols $R^{\mathbb{Q}}$ and $\bar{R}^{\mathbb{Q}}$ are interchanged.

B. Hilbert spaces

A Hilbert space $\mathcal{H}_{\mathbb{F}}$ is a *linear space over* $\mathbb{F}$, i.e., $v \in \mathcal{H}_{\mathbb{F}}$, $f \in \mathbb{F}$ implies $vf \in \mathcal{H}_{\mathbb{F}}$, $(v+v')f = vf + v'f$. We write the *number* f *to the right of the vector* v; this is a mere convention for $\mathbb{F} = \mathbb{R}$, $\mathbb{C}$ but essential for $\mathbb{F} = \mathbb{Q}$ since it is possible to define both left-$\mathbb{Q}$-Hilbert spaces (linear combinations $fv + f'v'$) and right-$\mathbb{Q}$-Hilbert spaces (linear combinations $vf + v'f'$). For $\mathcal{H}_{\mathbb{F}}$ a continuous bilinear mapping $\mathcal{H}_{\mathbb{F}} \times \mathcal{H}_{\mathbb{F}} \to \mathbb{F}$, the *scalar product*, is defined and satisfies

$$\begin{aligned} &\langle v, v' \rangle = \langle v', v \rangle^*, \quad \langle v, v' + v'' \rangle = \langle v, v' \rangle + \langle v, v'' \rangle, \\ &\langle vf, v'f' \rangle = f^* \langle v, v' \rangle f'. \end{aligned} \tag{2.16}$$

In the last equation care must be given to the order of the factors if $\mathbb{F} = \mathbb{Q}$. The *norm* $\|v\|$ of an element $v \in \mathcal{H}_{\mathbb{F}}$ is the real number defined by

$$\|v\|^2 = \langle v, v \rangle, \quad \|v\| \geq 0. \tag{2.17}$$

$\mathcal{H}_{\mathbb{F}}$ is *complete* with respect to the topology induced by the norm. We also assume $\mathcal{H}_{\mathbb{F}}$ to be *separable*, i.e., there exists a countable (or finite) set $\{v_n : n = 0, 1, \ldots\}$ of orthogonalized elements ($\langle v_n, v_N \rangle = \delta_{nN}$) such that for every $v \in \mathcal{H}_{\mathbb{F}}$, $v = \sum_n v_n \langle v_n, v \rangle$. In the following the term *basis* will always denote an orthonormalized basis if not explicitly stated otherwise.

We call the Hilbert space $\mathcal{H}_{\mathbb{F}}$ a *restriction* of the Hilbert space $\mathcal{H}_{\mathbb{F}'}$, if the following conditions are satisfied: (i) $\mathbb{F}$ is a proper subfield of $\mathbb{F}'$ ($\mathbb{F} \subset \mathbb{F}'$); (ii) $\mathcal{H}_{\mathbb{F}}$ is a subset of $\mathcal{H}_{\mathbb{F}'}$; (iii) $\mathcal{H}_{\mathbb{F}}$ is a Hilbert space over $\mathbb{F}$; (iv) $\dim \mathcal{H}_{\mathbb{F}} = \dim \mathcal{H}_{\mathbb{F}'}$. $\mathcal{H}_{\mathbb{F}}$ is a uniquely determined by the pair $\mathbb{F}$, $\{v_n\}$, where $\{v_n\}$ is a basis of $\mathcal{H}_{\mathbb{F}}$, (and as a consequence of its definition also of the Hilbert space $\mathcal{H}_{\mathbb{F}}$). If $\mathcal{H}_{\mathbb{F}}$ is a restriction of $\mathcal{H}_{\mathbb{F}'}$, the Hilbert space $\mathcal{H}_{\mathbb{F}'}$ is called an *extension* of $\mathcal{H}_{\mathbb{F}}$. It is obtained from $\mathcal{H}_{\mathbb{F}}$ by replacing the base field $\mathbb{F}$ by its extension $\mathbb{F}'$.

C. Operators

A mapping A from $\mathcal{H}_F$ into $\mathcal{H}_F$ is called an *operator on* $\mathcal{H}_F$. If A is a mapping from a subset $\mathfrak{D}(A)\subset\mathcal{H}_F$ into $\mathcal{H}_F$ A is called an *operator in* $\mathcal{H}_F$, and $\mathfrak{D}(A)$ the *domain* of A. We consider two classes of operators:

$$A \textit{ linear} \Leftrightarrow A(vf+v'f') = (Av)f + (Av')f', \tag{2.18}$$

$$A \textit{ antilinear} \Leftrightarrow A(vf+v'f') = (Av)f^* + (Av')f'^* . \tag{2.19}$$

An example of a linear operator is the *left-multiplication by real numbers* defined by

$$av = va \quad \text{for all } a\in\mathbb{R},\ v\in\mathcal{H}_F. \tag{2.20}$$

Obviously $aA = Aa$ for all operators A and the left-multiplier 1 is the 1-operator of $\mathcal{H}_F$.

An operator U on $\mathcal{H}_F$ is called *norm-preserving* if

$$\|Uv\| = \|v\| \quad \text{for all } v\in\mathcal{H}_F. \tag{2.21}$$

If U is also linear (2.21) is equivalent to

$$\langle Uv, U v'\rangle = \langle v, v'\rangle \quad \text{for all } v, v'\in\mathcal{H}_F. \tag{2.22}$$

Norm-preserving linear operators are called *orthogonal* if $F=\mathbb{R}$, *unitary* if $F=C$ and *hyperunitary* if $F=\mathbb{Q}$.

Defining the operator $A^\dagger$ *adjoint* to the operator A by

$$\langle v, Av'\rangle = \langle Av, v'\rangle \text{ for all } v\in\mathfrak{D}(A^\dagger), v'\in\mathfrak{D}(A), \tag{2.23}$$

linear norm-preserving operators may also be characterized by the relation

$$UU^\dagger = U^\dagger U = 1. \tag{2.24}$$

In $\mathcal{H}_C$ we also consider antilinear norm-preserving operators. Such an operator, usually denoted by the symbol θ, is called *antiunitary*; it satisfies Eqs. (2.19) and (2.21) or, equivalently,

$$\langle \theta v, \theta v'\rangle = \langle v', v\rangle \quad \text{for all } v, v'\in\mathcal{H}_C. \tag{2.25}$$

Of special interest are *self-adjoint linear operators with pure point spectrum*. For such an operator $H=H^\dagger$ and there exists a basis $\{v_n\}$ of $\mathcal{H}_F$ such that

$$Hv_n = v_n\epsilon_n,\ \epsilon_n\in\mathbb{R}. \tag{2.26}$$

The elements v_n are called *eigenvectors*, the numbers ϵ_n *eigenvalues*. The *spectrum* $\sigma(H)$ is the set of (different) eigenvalues of H. H admits a spectral decomposition

$$H = \sum_{\epsilon\in\sigma(H)} P_\epsilon\epsilon,\ P_\epsilon = P_\epsilon^\dagger = P_\epsilon^2. \tag{2.27}$$

The *degeneracy* of an eigenvalue ϵ of H is defined as

$$\deg(\epsilon, H) = \dim P_\epsilon\mathcal{H}_F. \tag{2.28}$$

If ϵ is not an eigenvalue of H we put $\deg(\epsilon, H)=0$. To solve the eigenvalue problem of H means to determine $\sigma(H)$ and a basis of eigenvectors or the set of *projectors* P_ϵ appearing in (2.27). What we are interested in is solving the eigenvalue problem for a given self-adjoint operator H in $\mathcal{H}_F$ exploiting the fact that there exists a group $\{U\}$ of norm-preserving operators commuting with H. We shall show that this can be achieved by considering an equivalent problem in a different Hilbert space ($\mathcal{H}_{F'}\neq\mathcal{H}_F$).

To make the last statement precise we introduce the concept of a *densely defined operator algebra* $A(\mathcal{H}_F)$. This is a set of linear operators including the left-multipliers $a\in\mathbb{R}$ which satisfies the following conditions: (i) $A(\mathcal{H}_F)$ is a linear space over $\mathbb{R}$; (ii) $A(\mathcal{H}_F)$ is closed under addition, multiplication, and adjunction; (iii) the domain of $A(\mathcal{H}_F)$, $\mathfrak{D}[A(\mathcal{H}_F)]$, defined as the intersection of all $\mathfrak{D}(A)$, $A\in A(\mathcal{H}_F)$ contains a basis of $\mathcal{H}_F$. The last condition may be dropped if $\mathcal{H}_F$ is finite-dimensional, since then all operators A can be defined on $\mathcal{H}_F$. We call two algebras $A(\mathcal{H}_F)$ and $A'(\mathcal{H}_{F'})$ *isomorphic operator algebras* if there exists a bijective mapping $I: A(\mathcal{H}_F)\to A'(\mathcal{H}_{F'})$ such that for all $A, B\in A(\mathcal{H}_F)$, $a\in\mathbb{R}$

$$I(A+B) = I(A) + I(B),\quad I(AB) = I(A)I(B)$$

$$I(A^\dagger) = I(A)^\dagger,\quad I(a) = a. \tag{2.29}$$

An isomorphism of operator algebras allows to solve the eigenvalue problem of an operator $H=H^\dagger\in A(\mathcal{H}_F)$ in an indirect way according to the following scheme: Given H, find $I(H)=H'$; solve the eigenvalue problem for H', i.e., find the spectral decomposition $H'=\sum_\epsilon P'_\epsilon\epsilon$; the spectral decomposition of H is $H=\sum_\epsilon P_\epsilon\epsilon$ where $P_\epsilon = I^{-1}(P'_\epsilon)$. If not only H but also a set $\{A\}$ of symmetry operators ($[A,H]=0$) is given in the beginning, the set $\{A'\}=\{I(A)\}$ may be used to simplify the solution of the eigenvalue problem of H'. In any case, for the method to be effective the mapping I and its inverse I^{-1} have to be known explicitly.

We now list four cases where it is possible to construct the isomorphism I. It is a common feature of all these examples that there exists a second operator algebra $C(\mathcal{H}_F)$ with $\mathfrak{D}[C(\mathcal{H}_F)]=\mathcal{H}_F$ such that

$$[A,C] = 0 \text{ for all } A\in A(\mathcal{H}_F),\ C\in C(\mathcal{H}_F). \tag{2.30}$$

Because of (2.30) $C(\mathcal{H}_F)$ is called the *commuting algebra* of $A(\mathcal{H}_F)$. In contrast to $A(\mathcal{H}_F)$, the algebra $C(\mathcal{H}_F)$ may contain antilinear operators; it is closed under adjunction if the adjoint $\theta^\dagger$ of the antilinear operator θ is defined via $\langle v, \theta^\dagger v'\rangle = \langle v', \theta v\rangle$.

Case 1:

$F=C$, $F'=\mathbb{R}$ and there exists an operator algebra $C(\mathcal{H}_C)=\{a+\theta b : a, b\in\mathbb{R}\}$ with

$$\theta \text{ antiunitary},\quad \theta^2 = 1. \tag{2.31}$$

In this case there exists a basis $\{w_m : m=0,1,\ldots\}$ of $\mathcal{H}_C$, $w_m\in\mathfrak{D}[A(\mathcal{H}_C)]$ such that

$$\theta w_m = w_m \tag{2.32}$$

and $\mathcal{H}_R$ is the restriction of $\mathcal{H}_C$ defined by one of the pairs $\mathbb{R}, \{w_m\}$.

Case 2:

$F=C$, $F'=\mathbb{Q}$ and there exists an operator algebra $C(\mathcal{H}_C)=\{a+\theta b : a, b\in\mathbb{R}\}$ with

$$\theta \text{ antiunitary},\quad \theta^2 = -1. \tag{2.33}$$

In this case there exists a basis $\{w_{mr} : m=0,1,\cdots; r=0,1\}$ of $\mathcal{H}_C$, $w_{mr}\in\mathfrak{D}[A(\mathcal{H}_C)]$ such that

$$\theta w_{m0} = w_{m1},\quad \theta w_{m1} = -w_{m0}, \tag{2.34}$$

implying that $\dim\mathcal{H}_C$ is even.

Case 3:

$\mathbf{F}=\mathbf{R}$, $\mathbf{F}'=\mathbf{C}$, and there exists an operator algebra $\mathbf{C}(\mathcal{H}_{\mathbf{R}})=\{\mathbf{F}_0 a+\mathbf{F}_1 b : a,b\in\mathbf{R}\}\cong\mathbf{C}$ with

$$\mathbf{F}_r \text{ linear}, \quad \mathbf{F}_r^\dagger=\sigma^{\mathbf{C}}_{r,r}\mathbf{F}_r,$$

$$\mathbf{F}_r\mathbf{F}_s=\sigma^{\mathbf{C}}_{r,s}\mathbf{F}_{rs}, \quad \sigma^{\mathbf{C}} \text{ and } rs \text{ see (2.4)}. \tag{2.35}$$

In this case there exists a basis $\{w_{mr} : m=0,1,\cdots; r=0,1\}$ of $\mathcal{H}_{\mathbf{R}}$, $w_{mr}\in\mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{R}})]$ such that

$$\mathbf{F}_r w_{ms}=\sigma^{\mathbf{C}}_{r,s} w_{m(rs)}, \tag{2.36}$$

implying that $\dim\mathcal{H}_{\mathbf{R}}$ is even.

Case 4:

$\mathbf{F}=\mathbf{R}$, $\mathbf{F}'=\mathbb{Q}$, and there exists an operator algebra $\mathbf{C}(\mathcal{H}_{\mathbf{R}})=\{\sum_{r=0}^{3}\mathbf{F}_r a_r : a_r\in\mathbf{R}\}\cong\mathbb{Q}$ with

$$\mathbf{F}_r \text{ linear}, \quad \mathbf{F}_r^\dagger=\sigma^{\mathbb{Q}}_{r,r}\mathbf{F}_r,$$

$$\mathbf{F}_r\mathbf{F}_s=\sigma^{\mathbb{Q}}_{r,s}\mathbf{F}_{rs}, \quad \sigma^{\mathbb{Q}}, \text{ and } rs \text{ see (2.5)}. \tag{2.37}$$

In this case there exists a basis $\{w_{mr} : m=0,1,\cdots; r=0,1,2,3\}$ of $\mathcal{H}_{\mathbf{R}}$, $w_{mr}\in\mathfrak{D}[\mathbf{A}(\mathcal{H})]$ such that

$$\mathbf{F}_r w_{ms}=\sigma^{\mathbb{Q}}_{r,s} w_{m(rs)}, \tag{2.38}$$

implying that $\dim\mathcal{H}_{\mathbf{R}}$ is a multiple of four.

D. Matrix representations

Isomorphisms between operator algebras are constructed via isomorphisms of matrix representations. A matrix A with elements $A_{nN}\in\mathbf{F}$ is called a *matrix representation* of the linear operator $\mathbf{A}$ in $\mathcal{H}_{\mathbf{F}}$ if there exists a basis $\{v_n\}\subset(\mathfrak{D}(\mathbf{A}))$ such that

$$\mathbf{A}v_n=\sum_N v_N A_{nN}, \quad A_{nN}=\langle v_n,\mathbf{A}v_N\rangle\in\mathbf{F}. \tag{2.39}$$

Given a matrix A and a basis $\{v_n\}$, (2.39) defines a linear operator $\mathbf{A}$. If a set $\{\mathbf{A}\}$ of operators is represented by a set $\{A\}$ of matrices we call $\{A\}$ a *matrix representation* of $\{\mathbf{A}\}$. The adjoint operator $\mathbf{A}^\dagger$ is represented by the *adjoint matrix* $A^\dagger$ with the elements

$$(A^\dagger)_{nN}=A^*_{Nn}. \tag{2.40}$$

The terms self-adjoint, norm-preserving, etc. are also used for matrices representing such operators.

In the first case listed in section 2C, Eqs. (2.31) and (2.32) ensure that the operators $\mathbf{A}\in\mathbf{A}(\mathcal{H}_{\mathbf{C}})$ have real matrix representatives. Hence the isomorphism $\{A\}\cong\{A'\}$ can be defined as follows:

$$\textit{Case 1}: A_{mM}=A'_{mM}. \tag{2.41}$$

In the other three cases the definition of the commuting algebra requires the operators $\mathbf{A}\in\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ to have matrix representatives which decompose into submatrices of size 2×2 (cases 2 and 3) or 4×4 (case 4) having a particular structure. This structure allows us to identify submatrices with elements from $\mathbf{F}$ with matrix elements (= numbers) from $\mathbf{F}'$ thereby establishing the isomorphism $\{A\}\cong\{A'\}$.

$$\textit{Case 2}: A_{mM}=C^{\mathbb{Q}}[A'_{mM}],\ C^{\mathbb{Q}} \text{ see (2.11)}. \tag{2.42}$$

$$\textit{Case 3}: A_{mM}=R^{\mathbf{C}}[A'_{mM}],\ R^{\mathbf{C}} \text{ see (2.13)}. \tag{2.43}$$

$$\textit{Case 4}: A_{mM}=\overline{R}^{\mathbb{Q}}[A'_{mM}],\ \overline{R}^{\mathbb{Q}} \text{ see (2.14)}. \tag{2.44}$$

Since in all four cases

$$B=A^\dagger\Leftrightarrow B'=A'^\dagger \tag{2.45}$$

a self-adjoint matrix H corresponds to a self-adjoint matrix H' and a norm-preserving matrix U to a norm-preserving matrix U'.

Now the isomorphism $I:\mathbf{A}(\mathcal{H}_{\mathbf{F}})\to\mathbf{A}'(\mathcal{H}_{\mathbf{F}'})$ can be defined as follows: First construct a basis $\{w_{mr}\}$ of $\mathcal{H}_{\mathbf{F}}$, transforming under $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ in the desired way. Such bases exist and can be constructed by means of a generalized Schmidt process. Next calculate the matrix representation $\{A\}$ of $\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ with respect to this basis and apply the correspondence rule given above to obtain the matrix representation $\{A'\}$. Finally use these matrices and an arbitrary basis $\{w'_{mr}\}$ of $\mathcal{H}_{\mathbf{F}'}$ to define $\mathbf{A}'(\mathcal{H}_{\mathbf{F}'})$. Proceeding in the reverse order one constructs a mapping $\bar{I}:\mathbf{A}'(\mathcal{H}_{\mathbf{F}'})\to\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ starting from a matrix representation $\{A'\}$ with respect to an arbitrary basis $\{\bar{w}'_m\}$. The proper correspondence rule gives the matrix representation $\{A\}$ and, augmented by an arbitrary basis $\{\bar{w}_{mr}\}$ of $\mathcal{H}_{\mathbf{F}}$, the operator algebra $\mathbf{A}(\mathcal{H}_{\mathbf{F}})$. The same basis $\{w_{mr}\}$ is then used to define the commuting algebra $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ by means of one of the equations (2.32), (2.34), (2.36), (2.38) and the (anti)linearity of the operators. I and $\bar{I}$ are inverse mappings if, and only if, the two bases involved in the definition of I and $\bar{I}$ coincide.

There are infinitely many equivalent ways to define the isomorphism I depending on the choice of the two bases $\{w_{mr}\}$ and $\{w'_{mr}\}$. This leads to the notion of equivalent matrix representations. Two sets of matrices with elements from $\mathbf{F}$, say $\{A\}$ and $\{\bar{A}\}$, are called $\mathbf{F}$-*equivalent* if there exists a norm-preserving matrix U with elements from $\mathbf{F}$ such that the matrices A and $\bar{A}$ are related by $UAU^\dagger=\bar{A}$. If $\mathbf{U}w_{MR}=\sum_{mr}w_{mr}U_{mr,MR}=\bar{w}_{MR}$ where both $\{w_{mr}\}$ and $\{\bar{w}_{mr}\}$ are bases transforming under $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ in the right way, then U has the same structure as the matrices A representing the operators $\mathbf{A}\in\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ [see (2.42)–(2.44)]. The concept of $\mathbf{F}$-equivalence allows one to understand the isomorphism of matrix algebras over different fields from another point of view.

$$\textit{Case 1}: A \text{ is } \mathbf{C}\text{-equivalent to a real } A'. \tag{2.46}$$

$$\textit{Case 2}: A \text{ is } \mathbb{Q}\text{-equivalent to } A'\oplus A'. \tag{2.47}$$

$$\textit{Case 3}: A \text{ is } \mathbf{C}\text{-equivalent to } A'\oplus A'^*. \tag{2.48}$$

$$\textit{Case 4}: A \text{ is } \mathbb{Q}\text{-equivalent to } A'\oplus A'\oplus A'\oplus A'. \tag{2.49}$$

Here $A'\oplus B''$ denotes the direct sum of the matrices A' and B'' and A'^* is the complex conjugate of A'.

Once the isomorphism I is fixed, the problem reduces to solving the eigenvalue problem for a self-adjoint operator $\mathbf{H}'\in\mathbf{A}(\mathcal{H}'_{\mathbf{F}})$. This is usually done by solving the eigenvalue equation $\mathbf{H}'v'=v'\epsilon$. The eigenvalue ϵ is also an eigenvalue of the operator $H\in\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ we are actually interested in but one might ask of what use the eigenvector $v'\in\mathcal{H}_{\mathbf{F}'}$ will be for the original problem. Now $\mathbf{P}'_\epsilon$ can be decomposed into a sum of projectors $\mathbf{P}'_k=|v'_k\rangle\langle v'_k|$ where the elements $\{v'_k\}$ form a basis of $\mathbf{P}'_\epsilon\mathcal{H}_{\mathbf{F}'}$. To every such projector $\mathbf{P}'=|v'\rangle\langle v'|$

there corresponds a projector $P = I^{-1}(P')$ in $\mathcal{K}_F$ projecting into a subspace of $P_\epsilon \mathcal{K}_F$. But the eigenvector v' contains even more information. For, knowing v', one is able to define a basis of $P\mathcal{K}$, i.e., a set $\{v_r\}$ of orthonormalized eigenvectors of H which belong to the eigenvalue ϵ and transform under the commuting algebra $C(\mathcal{K}_F)$ in the same way as the basis $\{w_{mr}\}$.

Case 1: $\langle w'_m, v' \rangle = a_m = \langle w_m, v \rangle \in \mathbf{R}$;

$$\theta v = v. \tag{2.50}$$

Case 2: $\langle w'_m, v' \rangle = q_m \in \mathbb{Q}$;

$r, s = 0, 1: \langle w_{mr}, v_s \rangle = C^{\mathbb{Q}}_{rs}(q_m) \in \mathbf{C}$,

$C^{\mathbb{Q}}$ see (2.11),

$$\theta v_0 = v_1, \quad \theta v_1 = -v_0. \tag{2.51}$$

Case 3: $\langle w'_m, v' \rangle = c_m \in \mathbb{C}$;

$r, s = 0, 1$: $\langle w_{mr}, v_s \rangle = R^{\mathbf{C}}_{rs}(c_m) \in \mathbf{R}$,

$\mathbf{R}^{\mathbf{C}}$ see (2.13),

$$\mathbf{F}_r v_s = v_{rs} \sigma^{\mathbf{C}}_{r,s}. \tag{2.52}$$

Case 4: $\langle w'_m, v' \rangle = q_m \in \mathbb{Q}$;

$r, s = 0, 1, 2, 3$: $\langle w_{mr}, v_s \rangle = \bar{R}^{\mathbb{Q}}_{rs}(q_m) \in \mathbf{R}$,

$R^{\mathbb{Q}}$ see (2.14),

$$\mathbf{F}_r v_s = v_{rs} \sigma^{\mathbb{Q}}_{r,s}. \tag{2.53}$$

Therefore constructing a basis $\{v_{mr}\}$ of eigenvectors of H especially adapted to the commuting algebra $C(\mathcal{K}_F)$ amounts to find a basis $\{v'_m\}$ of eigenvectors of H'. It is also a direct consequence of Eqs. (2.50)–(2.53) that the dimension of the subspace $P\mathcal{K}_F$ is a multiple of the dimension of the subspace $P'\mathcal{K}_{F'}$ if only P and P' are corresponding projectors (not necessarily into subspaces invariant under H and H', respectively).

$$I(P) = P' \tag{2.54}$$

Case 1: $\dim P\mathcal{K}_C = \dim P'\mathcal{K}_R$; (2.55)

Case 2: $\dim P\mathcal{K}_C = 2 \dim P'\mathcal{K}_Q$; (2.56)

Case 3: $\dim P\mathcal{K}_R = 2 \dim P'\mathcal{K}_C$; (2.57)

Case 4: $\dim P\mathcal{K}_R = 4 \dim P'\mathcal{K}_Q$. (2.58)

The determination of the eigenvalues of a real or complex self-adjoint matrix (case 1 or 3) is a standard problem if the matrix is finite-dimensional. It may be solved numerically (yielding at the same time a basis of eigenvectors) or by reduction to an algebraic equation.

A finite-dimensional quaternionic self-adjoint matrix raises no more problems since the isomorphism (2.42) allows to replace it by a complex matrix of double size thereby doubling the degeneracies of all eigenvalues. It is not necessary to exploit this fact if the minimal polynomial is used to find the eigenvalues since this polynomial has real coefficients only even if the matrix elements are quaternions.

We now summarize those results which are of primary interest for physical problems (cases 1 and 2) and give some examples where they apply.

3. EQUIVALENCE OF EIGENVALUE PROBLEMS

A. $\theta^2 = +1$ – real Hilbert space

(i) Let $\mathcal{K}_C$ be a complex Hilbert space and H a self-adjoint operator in $\mathcal{K}_C$ with a pure point spectrum $\sigma(H)$, and θ an antiunitary operator on $\mathcal{K}_C$ with $\theta^2 = +1$. If $[\theta, H] = 0$ then there exists a real Hilbert space $\mathcal{K}_R$ with $\dim\mathcal{K}_R = \dim\mathcal{K}_C$ and a self-adjoint operator H' in $\mathcal{K}_R$ such that the spectra of H' and H and the degeneracies of the common eigenvalues coincide, i.e.,

$$\sigma(H) = \sigma(H'), \ \deg(\epsilon, H) = \deg(\epsilon, H'). \tag{3.1}$$

Moreover, if there exists a group $\{U\}$ of unitary operators commuting with θ and H then there exists an isomorphic group $\{U'\}$ of orthogonal operators commuting with H':

$$U \text{ unitary}, \ [U, H] = 0; \ U' \text{ orthogonal}, \ [U', H] = 0, \quad \text{group}\{U\} \cong \text{group}\{U'\}. \tag{3.2}$$

(ii) Let H' be a self-adjoint operator in a real Hilbert space $\mathcal{H}_R$ and $\{U'\}$ a group of orthogonal operators commuting with H'. Then there exists a complex Hilbert space $\mathcal{K}_C$ with $\dim\mathcal{K}_C = \dim\mathcal{K}_R$ and operators H and U such that $H = H^\dagger$ and $U^{-1} = U^\dagger$ and Eqs. (3.1) and (3.2) are valid. Moreover, it is possible to define an antiunitary operator θ on $\mathcal{K}_C$ such that $\theta^2 = +1$ and $[\theta, H] = [\theta, U] = 0$.

B. $\theta^2 = -1$ – quaternionic Hilbert space

(i) Let $\mathcal{K}_C$ be a complex Hilbert space, H a self-adjoint operator in $\mathcal{K}_C$ with pure point spectrum $\sigma(H)$, and θ an antiunitary operator on $\mathcal{K}_C$ with $\theta^2 = -1$. If $[\theta, H] = 0$ then there exists a quaternionic Hilbert space $\mathcal{K}_Q$ with $\dim\mathcal{K}_Q = \frac{1}{2}\dim\mathcal{K}_C$ and a self-adjoint operator H' in $\mathcal{K}_Q$ such that the spectra and the degeneracies are related by

$$\sigma(H) = \sigma(H'), \quad \deg(\epsilon, H) = 2 \deg(\epsilon, H'). \tag{3.3}$$

Moreover, if there exists a group $\{U\}$ of unitary operators commuting with θ and H, then there exists an isomorphic group $\{U'\}$ of hyperunitary operators commuting with H':

$$U \text{ unitary}, \ [U, H] = 0; \ U' \text{ hyperunitary}, \ [U', H'] = 0, \quad \text{group}\{U\} \cong \text{group}\{U'\}. \tag{3.4}$$

(ii) Let H' be a self-adjoint operator in a quaternionic Hilbert space $\mathcal{K}_Q$ and $\{U'\}$ a group of hyperunitary operators commuting with H'. Then there exists a complex Hilbert space $\mathcal{K}_C$ with $\dim\mathcal{K}_C = 2 \dim\mathcal{K}_Q$ and operators H and U such that Eqs. (3.3) and (3.4) are valid. Moreover it is possible to define an antiunitary operator θ on $\mathcal{K}_C$ such that $\theta^2 = -1$ and $[\theta, H] = [\theta, U] = 0$.

C. Examples

(i) *Lattice vibrations* ($\theta^2 = +1$). In this example one starts with a finite real self-adjoint (=symmetric) matrix H, the matrix of force constants, and a group of orthogonal matrices U commuting with H. The group $\{U\}$ reflecting the symmetries of the interaction of the lattice constituents is a finite homomorphic image of a space group.[12] These real matrices can be considered

as representatives of operators on a real Hilbert space $\mathcal{H}_R$ or on a complex Hilbert space $\mathcal{H}_C$. In the second case the basis used to define the operators is assumed to be invariant under an antiunitary operator $\theta = K$. Since K transforms expansion coefficients (with respect to the invariant basis) into their complex conjugates $K^2 = 1$ and $[K, H] = [K, U] = 0$.

(ii) *Real Hamiltonians* ($\theta^2 = +1$). If the potential in the time-independent Schrödinger equation is a real-valued function the Hamiltonian H commutes with $\theta = K$ (complex conjugation of functions). Every real basis (e.g., the oscillator eigenfunctions) yields a real matrix representation of H. $\mathcal{H}_{R|}$ is the Hilbert space of real-valued square-integrable functions.

(iii) *Spin-orbit interaction – even number of electrons* ($\theta^2 = +1$). $H = \sum_{n=1}^{N} \sum_{r=1}^{3} A_n l_{nr} \sigma_{nr}$ where A_n is a real potential, l_{nr} the r-component of the orbital angular momentum operator of the nth electron, and σ_{nr} the corresponding Pauli-spin operator. The antiunitary operator commuting with H is $\theta = \Pi_{n=1}^{N} \sigma_{n2} K$. The original Hilbert space is the tensor space $\mathcal{H}_C = \mathcal{H}_{Space} \otimes \mathcal{H}_{Spin}$. If a real basis $\{\varphi_m\}$ is chosen for $\mathcal{H}_{Space}$ the action of H is determined by a set of operators on $\mathcal{H}_{Spin}$, each of the form $H_{mM} = i \sum_{n=1}^{N} \sum_{r=1}^{3} \sigma_{nr} a_{nr,mM}$, $a_{nr,mM} \in \mathbb{R}$. To see that these operators have real matrix representatives it is sufficient to consider the case $N = 2$ since the electrons can be collected into pairs. If $|\mu_1 \mu_2\rangle$ is a common eigenvector of σ_{13} and σ_{23}, $\sigma_{n3}|\mu_1 \mu_2\rangle = \mu_n |\mu_1 \mu_2\rangle$, and the basis $\{|s\rangle\}$ of $\mathcal{H}_{Spin}$ is defined by

$$\sqrt{2}\,|0\rangle = |+-\rangle - |-+\rangle\,, \quad \sqrt{2}\,|1\rangle = (-|++\rangle + |--\rangle)i\,,$$
$$\sqrt{2}\,|2\rangle = |++\rangle + |--\rangle\,, \quad \sqrt{2}\,|3\rangle = (-|+-\rangle - |-+\rangle)i\,, \tag{3.5}$$

then $\theta|s\rangle = |s\rangle$ and the corresponding matrix representation is

$$\left[2i \sum_{n=1}^{2} \sum_{r=1}^{3} s_{nr} a_{mr}\right]_{sS} = R^{Q}_{sS}(ia_{11} + ja_{12} + ka_{13}) + \bar{R}^{Q}_{sS}(ia_{21} + ja_{22} + ka_{23}),$$
$$R^{Q} \text{ see (2.12)}\,, \quad \bar{R}^{Q} \text{ see (2.14)}\,. \tag{3.6}$$

A short reflection shows that the operators $ij_{nr} = i(l_{nr} + s_{nr})$ also have real matrix representatives with respect to the basis $\{\varphi_m|s\rangle\}$ so that all rotations are represented by orthogonal matrices.

(iv) *Spin-orbit interaction — odd number of electrons* ($\theta^2 = -1$). The situation is the same as in (iii) except that N is now odd. If the electrons are again collected into pairs and the same basis $\{\varphi_m|s\rangle\}$ as in (iii) is used for $N-1$ electrons, the problem reduces to finding a matrix representative of $H_{ms,MS} = i\sum_{r=1}^{3} \sigma_r a_{r,msMS}$, $a_{r,msMS} \in \mathbb{R}$, which can be identified with a quaternion. This is provided by the representation

$$\left[i \sum_{r=1}^{3} \sigma_r a_r\right]_{sS} = C^{Q}_{sS}(ia_3 + ja_2 + ka_1), \quad C^{Q} \text{ [see(2.11)]}\,, \tag{3.7}$$

belonging to the basis $\{|s\rangle : s = 0, 1; \sigma_3|s\rangle = (-1)^s|s\rangle\}$.

This basis transforms under the antiunitary operator $\theta = -i\sigma_2 K$ in the way required by (2.33), i.e., $\theta|0\rangle = |1\rangle$, $\theta|1\rangle = -|0\rangle$. Both the infinitesimal generators and a finite rotation in this spin space are of the form (2.4); therefore rotations of the N-electron system can also be represented by hyperunitary matrices.

4. SIMPLIFICATION OF THE EIGENVALUE PROBLEM AND PREDICTION OF THE POSSIBLE DEGENERACIES

A. Complex Hilbert spaces

As has been shown in the preceeding section, the effect of a commuting antiunitary operator θ, with $\theta^2 = \pm 1$, can be taken into account by reformulating the eigenvalue problem in a noncomplex Hilbert space. If the unitary operators $U(x)$ representing the group $G = \{x\}$ in $\mathcal{H}_C$ commute not only with the self-adjoint operator H but also with θ then there exists a symmetry group $\{U'(x) : x \in G\}$ isomorphic to the original one which can be used to simplify the eigenvalue problem of the new self-adjoint operator H′ and to make predictions on the degeneracies of the eigenvalues of H′. Before doing so let us briefly recall how to proceed for a complex eigenvalue problem.

In the complex case it is convenient to extend the set of operators commuting with H from the symmetry group

$$G = \{U(x) : x \in G\} \tag{4.1}$$

to a set of bounded operators

$$B_C = G \cup A_C\,, \tag{4.2}$$

where A_C is a *representation of the complex group algebra in the complex Hilbert space* $\mathcal{H}_C$. The elements of A_C are related to the operators $U(x)$ by

$$A \in A_C : A = M_x a(x) U(x), \text{ i.e.,}$$
$$\langle v, Aw\rangle = M_x a(x) \langle v, U(x) w\rangle \text{ for all } v, w \in \mathcal{H}_C\,. \tag{4.3}$$

$$M_x f(x) = \frac{1}{|G|} \sum_{x} f(x) \text{ if } G \text{ is finite;}$$

$$M_x f(x) = \int_{\Xi} d\xi\, w(\xi) f(\xi) \text{ if } G \text{ is compact and continuous,}$$

Ξ a suitable parameter space, and w a positive weight function;

$$M_x f(x) = M_x f(x^{-1}) = M_x f(xyz) \text{ for all } y, z \in G;$$
$$f(x) = 1 \Rightarrow M_x f(x) = 1\,. \tag{4.4}$$
$$a \in L^2(G, \mathbb{C}), \text{ i.e., } a(x) \in \mathbb{C}, \quad M_x |a(x)|^2 < \infty\,. \tag{4.5}$$

The number $M_x f(x)$ in (4.4) is called the Haar integral of f. Every $B \in B_C$ can be decomposed as

$$B = \sum_{\lambda j J} E^{\lambda}_{jJ} C^{\lambda}_{jJ}(b), \quad C^{\lambda}_{jJ}(b) \in \mathbb{C}\,, \tag{4.6}$$

where

$\{\lambda\}$ = index set for $\mathbb{C}$-equivalence classes of matrix representations of G, subject to the convention $\lambda_1 = \lambda_+$, $\lambda_2 = \lambda_-$ if $C^{\lambda 1}(x)$ is $\mathbb{C}$-equivalent to $[C^{\lambda 2}(x)]^*$, (4.7)

$$j, J = 0, \ldots, d_{\mathbf{C}}^{\lambda} - 1 , \tag{4.8}$$

and the operators $\mathbf{E}_{jJ}^{\lambda}$ satisfy

$$E_{jJ}^{\lambda\dagger} = \mathbf{E}_{Jj}^{\lambda}, \quad \mathbf{E}_{jj'}^{\lambda}\mathbf{E}_{J'J}^{\Lambda} = \delta_{\lambda\Lambda}\delta_{j'J}\mathbf{E}_{jJ'}^{\lambda}; \quad [\mathbf{E}_{jJ}^{\lambda}, \mathbf{H}] = 0 . \tag{4.9}$$

It is just these relations which make the group algebra interesting for the solution of the eigenvalue problem of H. For the operators $\mathbf{E}_{jj}^{\lambda}$ are *projectors* which commute with H and therefore define subspaces of $\mathcal{H}_{\mathbf{C}}$ invariant under H,

$$\mathcal{H}_{\mathbf{C}} = \sum_{\lambda j} \oplus \mathcal{H}_{\mathbf{C},j}^{\lambda}, \quad \mathcal{H}_{\mathbf{C},j}^{\lambda} = \mathbf{E}_{jj}^{\lambda}\mathcal{H}_{\mathbf{C}} , \tag{4.10}$$

and the *shift operators* $\mathbf{E}_{jJ}^{\lambda}, j \neq J$, transform these subspaces among each other:

$$\mathbf{E}_{j0}^{\lambda}\mathcal{H}_{\mathbf{C},0}^{\lambda} = \mathcal{H}_{\mathbf{C},j}^{\lambda} . \tag{4.11}$$

Thus, given bases $\{v_0^{\lambda m}\}$ of the subspaces $\mathcal{H}_{\mathbf{C},0}^{\lambda}$,

$$m = 0, \ldots, m_{\mathbf{C}}^{\lambda} - 1; \; m_{\mathbf{C}}^{\lambda} = \dim \mathcal{H}_{\mathbf{C},j}^{\lambda} ; \tag{4.12}$$

it is possible to define a basis $\{v_j^{\lambda m}\}$ of $\mathcal{H}_{\mathbf{C}}^{\lambda}$ by

$$v_j^{\lambda m} = \mathbf{E}_{j0}^{\lambda} v_0^{\lambda m} . \tag{4.13}$$

$\{v_j^{\lambda m}\}$ is called a *symmetry-adapted basis* since

$$\mathbf{U}(x) v_J^{\lambda m} = \sum_j v_j^{\lambda m} C_{jJ}^{\lambda}(\mathbf{x}) ,$$

$\{C^{\lambda}(\mathbf{x}) : x \in G\}$ = complex irreducible matrix representation of G,

$$C^{\lambda}(\mathbf{x}) \text{ is unitary}, \quad \dim C^{\lambda}(\mathbf{x}) = d_{\mathbf{C}}^{\lambda} . \tag{4.14}$$

The matrix representation of H with respect to a symmetry-adapted basis decomposes into a direct sum of submatrices arranged into sequences of identical members

$$\langle v_j^{\lambda m}, \mathbf{H} v_J^{\Lambda M} \rangle = \delta_{\lambda\Lambda}\delta_{jJ} H_{mM}^{\lambda} ,$$
$$H_{mM}^{\lambda} = \langle v_0^{\lambda m}, \mathbf{H} v_0^{\lambda M} \rangle . \tag{4.15}$$

It is therefore only necessary to solve the eigenvalue problem for operators $\mathbf{H}^{\lambda}$ in complex Hilbert spaces $\mathcal{H}_{\mathbf{C}}^{\lambda} \cong \mathcal{H}_{\mathbf{C},0}^{\lambda}$ defined by matrices H^{λ} of dimension $m_{\mathbf{C}}^{\lambda}$ with elements given by (4.15).

$$\sigma(\mathbf{H}) = \bigcup_{\lambda} \sigma(\mathbf{H}^{\lambda}) . \tag{4.16}$$

Since the matrix representative of H contains $d_{\mathbf{C}}^{\lambda}$ copies of $\mathbf{H}^{\lambda}$

$$\deg(\epsilon, \mathbf{H}) = \sum_{\lambda} d_{\mathbf{C}}^{\lambda} \deg(\epsilon, \mathbf{H}^{\lambda}) . \tag{4.17}$$

B. Real Hilbert spaces

We assume the real eigenvalue problem to originate from a complex one; the notation H′ is used for the self-adjoint operator and $\mathbf{U}'(x)$ for the orthogonal operators forming the symmetry group G. Instead of $\mathbf{B}_{\mathbf{C}}$ we consider the bounded operators

$$\mathbf{B}_{\mathbf{R}} = G \cup \mathbf{A}_{\mathbf{R}} , \tag{4.18}$$

where $\mathbf{A}_{\mathbf{R}}$ is a *representation of the real group algebra in the real Hilbert space* $\mathcal{H}_{\mathbf{R}}$. The elements $\mathbf{A}' \in A_{\mathbf{R}}$ are again related to the operators $\mathbf{U}' \in \mathbf{G}$ by Eq. (4.3) but now with $v, w \in \mathcal{H}_{\mathbf{R}}$ and

$$a \in L^2(G, \mathbb{R}) , \text{ i.e., } a(x) \in \mathbb{R}, \quad M_x |a(x)|^2 < \infty . \tag{4.19}$$

The decomposition of an element $\mathbf{B}' \in \mathbf{B}_{\mathbf{R}}$ analogous to (4.6) now reads

$$\mathbf{B}' = \sum_{\lambda j J r} \mathbf{E}_{jJ}^{\lambda} \mathbf{F}_r^{\lambda} R_{jr,J0}^{\lambda}(\mathrm{b}), \; R_{jr,J0}^{\lambda}(\mathrm{b}) \in \mathbb{R} , \tag{4.20}$$

where

$\{\lambda\}$ = index set as in (4.7) with pairs $\lambda\pm$ replaced by single elements λ; (4.21)

$$j, J = 0, \ldots, d^{\lambda} - 1 , \tag{4.22}$$

$$r, R = 0, \ldots, r^{\lambda} - 1; \; r^{\lambda} = \text{rank of } \tilde{\mathbf{F}} . \tag{4.23}$$

Although not explicitly indicated by a subscript, the field $\tilde{\mathbf{F}}$ is uniquely determined by λ. According to whether $\tilde{\mathbf{F}} = \mathbb{R}$ or $\mathbb{C}$ or $\mathbb{Q}$, we distinguish *three types of representations*:

$$\begin{aligned} &\mathbb{R}\text{-type: } \tilde{\mathbf{F}} = \mathbb{R}, \quad r^{\lambda} = 1 , \\ &\mathbb{C}\text{-type: } \tilde{\mathbf{F}} = \mathbb{C}, \quad r^{\lambda} = 2 , \\ &\mathbb{Q}\text{-type: } \tilde{\mathbf{F}} = \mathbb{Q}, \quad r^{\lambda} = 4 . \end{aligned} \tag{4.24}$$

The operators $\mathbf{E}_{jJ}^{\lambda}$ appearing in (4.20) satisfy the same relations as stated in (4.9) for the complex case (H $\to$ H′); the operators $\mathbf{F}^{\lambda}$ satisfy

$$\begin{aligned} &\mathbf{F}_r^{\lambda}\mathbf{F}_R^{\Lambda} = \delta_{\lambda\Lambda}\mathbf{F}_{rR}^{\lambda}\sigma^{\tilde{\mathbf{F}}}{}_{r,R} , \; \mathbf{F}_r^{\lambda\dagger} = \mathbf{F}_r^{\lambda}\sigma^{\tilde{\mathbf{F}}}{}_{r,r} , \\ &\sigma^{\tilde{\mathbf{F}}} \text{ and } rR \text{ see (2.3) to (2.5)} ; \\ &[\mathbf{F}_r^{\lambda}, \mathbf{E}_{RS}^{\Lambda}] = 0; \; [\mathbf{F}_r^{\lambda}\mathbf{H}'] = 0 . \end{aligned} \tag{4.25}$$

Since the operators $\mathbf{E}_{jJ}^{\lambda}$ behave as in the complex case, the operators $\mathbf{E}_{jj}^{\lambda}$ are again *projectors* projecting into subspaces of $\mathcal{H}_{\mathbf{R}}$ invariant under H′,

$$\mathcal{H}_{\mathbf{R}} = \sum_{\lambda j} \oplus \mathcal{H}_{\mathbf{R},j}^{\lambda}, \quad \mathcal{H}_{\mathbf{R},j}^{\lambda} = \mathbf{E}_{jj}^{\lambda}\mathcal{H}_{\mathbf{R}} . \tag{4.26}$$

The set of *shift operators*, however, now contains the operators $\mathbf{F}_r^{\lambda}$ as well as the operators $\mathbf{E}_{jJ}, j \neq J$. Since $\mathbf{F}_r^{\lambda} v$ is orthogonal to $\mathbf{F}_R^{\lambda} v$ if $r \neq R$ and both elements of $\mathcal{H}_{\mathbf{R}}$ have the same norm it is possible to construct bases $\{v_{0r}^{\lambda m}\}$ of the subspaces $\mathcal{H}_{\mathbf{R},0}^{\lambda}$ such that

$$v_{0r}^{\lambda m} = \mathbf{F}_r^{\lambda} v_{00}^{\lambda m}, \quad m = 0, \ldots, m_{\mathbf{R}}^{\lambda} - 1, \tag{4.27}$$

and to extend these bases by

$$v_{jr}^{\lambda m} = \mathbf{E}_{j0}^{\lambda} v_{0r}^{\lambda m} \tag{4.28}$$

to a basis $\{v_{jr}^{\lambda m}\}$ of $\mathcal{H}_{\mathbf{R}}$ with

$$\dim \mathcal{H}_{\mathbf{R}} = \sum_{\lambda} d^{\lambda} r^{\lambda} m_{\mathbf{R}}^{\lambda} . \tag{4.29}$$

Here too $\{v_{jr}^{\lambda m}\}$ is called a *symmetry-adapted basis* since

$$\mathbf{U}'(x) v_{JR}^{\lambda m} = \sum_{jr} v_{jr}^{\lambda m} R_{jr,JR}^{\lambda}(\mathbf{x}) ,$$

$\{R^{\lambda}(\mathbf{x}) : x \in G\}$ = real irreducible matrix representation of G,

$R^\lambda(\mathbf{x})$ orthogonal, $\dim \mathbf{R}^\lambda(\mathbf{x}) = d^\lambda r^\lambda$. (4.30)

The matrices $R^\lambda(\mathbf{x})$ have a special structure depending on the type of the representation λ.

$\tilde{\mathbf{F}}$-type: $R^\lambda_{jr,JR}(\mathbf{x}) = R^{\tilde{\mathbf{F}}}_{rR}[\tilde{\mathbf{F}}_{jJ}(\mathbf{x})]$;

$R^{\tilde{\mathbf{F}}}$ see (2.10), (2.12), (2.13);

$\{\tilde{\mathbf{F}}^\lambda(\mathbf{x}): x \in G\} = \tilde{\mathbf{F}}$-valued irreducible matrix representation of G, $F^\lambda(\mathbf{x})$ norm-preserving . (4.31)

As in the complex case, the matrix representation of $\mathbf{H}'$ with respect to the symmetry-adapted basis $\{v^{\lambda m}_{jr}\}$ decomposes into a direct sum of submatrices of dimension $r^\lambda m^\lambda_{\mathbf{R}}$ grouped into chains of d^λ identical matrices,

$$\langle v^{\lambda m}_{jr}, \mathbf{H}' v^{\Lambda M}_{JR} \rangle = \delta_{\lambda\Lambda}\delta_{jJ} H^\lambda_{mr,MR},$$
$$H^\lambda_{mr,MR} = \langle v^{\lambda m}_{0r}, \mathbf{H}' v^{\lambda M}_{0r} \rangle . \quad (4.32)$$

Similarly to the matrix representatives $R^\lambda(\mathbf{x})$, the submatrices H^λ show a structure characteristic of the type of representation λ.

$\mathbb{R}$-type: $H^\lambda_{m0,M0} = R^{\mathbf{R}}[H^{\lambda(\mathbf{R})}_{mM}] = H^{\lambda(\mathbf{R})}_{mM} \in \mathbb{R}$,

$\mathbb{C}$-type: $H^\lambda_{mr,MR} = R^{\mathbf{C}}_{rR}[H^{\lambda(\mathbf{C})}_{mM}]$, $H^{\lambda(\mathbf{C})}_{mM} \in \mathbb{C}$, (4.33)

$\mathbb{Q}$-type: $H^\lambda_{mr,MR} = \bar{R}^{\mathbf{Q}}_{rR}[H^{\lambda(\mathbf{Q})}_{mM}]$, $H^{\lambda(\mathbf{Q})}_{mM} \in \mathbb{Q}$.

The matrix $H^{\lambda(\tilde{\mathbf{F}})}$ defines a self-adjoint operator $\mathbf{H}^\lambda$ in a Hilbert space $\mathcal{H}^\lambda_{\tilde{\mathbf{F}}}$ of dimension $m^\lambda_{\mathbf{R}}$. If the eigenvalue problem is solved for each of the operators $\mathbf{H}^\lambda$ it is solved for the operator $\mathbf{H}'$ since

$$\sigma(\mathbf{H}') = \bigcup_\lambda \sigma(\mathbf{H}^\lambda), \quad (4.34)$$

$$\deg(\epsilon, \mathbf{H}') = \sum_\lambda d^\lambda r^\lambda \deg(\epsilon, \mathbf{H}^\lambda), \quad (4.35)$$

and the eigenvectors of the operators $\mathbf{H}^\lambda$ can be used to define a basis $\{w^{\lambda m}_{0r}\}$ of $\mathcal{H}^\lambda_{\mathbf{R},0}$ [cf. Eqs. (2.51) and (2.52)] from which a symmetry adapted basis $\{w^{\lambda m}_{jr}\}$ of $\mathcal{H}_{\mathbf{R}}$ consisting of eigenvectors of $\mathbf{H}'$ is obtained by means of Eq. (4.28).

Moreover, if $\mathbf{H}'$ is the restriction of an operator $\mathbf{H}$ in a complex Hilbert space $\mathcal{H}_{\mathbf{C}}$ [Sec. 3B (i)] the basis $\{w^\lambda_{jr}\}$ is also a basis of $\mathcal{H}_{\mathbf{C}}$ and the elements w^λ_{jr} are eigenvectors of $\mathbf{H}$ invariant under the antiunitary operator θ. The spectrum of $\mathbf{H}$ and the degeneracies are given by (4.34) and (4.35) with $\mathbf{H}'$ replaced by $\mathbf{H}$, showing that the minimal degeneracies are given by the dimension of the real irreducible matrix representations instead of the complex ones [Eq. (4.17)]. The doubling of degeneracies of some of the eigenvalues caused by this change is usually derived by means of the antiunitary operator θ. In this approach the complex irreducible representations are partitioned into three classes and a comparison of the results, namely

$\mathbb{R}$-type: $d^\lambda = d^\lambda_{\mathbf{C}}$, $m^\lambda_{\mathbf{R}} = m^\lambda_{\mathbf{C}}$,

$\mathbb{C}$-type: $d^\lambda = d^{\lambda+}_{\mathbf{C}} = d^{\lambda-}_{\mathbf{C}}$, $m^\lambda_{\mathbf{R}} = m^{\lambda+}_{\mathbf{C}} = m^{\lambda-}_{\mathbf{C}}$, (4.36)

$\mathbb{Q}$-type: $d^\lambda = \frac{1}{2} d^\lambda_{\mathbf{C}}$, $m^\lambda_{\mathbf{R}} = \frac{1}{2} m^\lambda_{\mathbf{C}}$,

shows that the "kind" of a complex irreducible representation is uniquely related to the "type" of a real one. We therefore have the following equivalence of terms[1]:

$\mathbb{R}$-type = first kind,

$\mathbb{C}$-type = third kind, (4.37)

$\mathbb{Q}$-type = second kind.

This correspondence is discussed in more detail in the companion paper.[13]

It should be pointed out that the matrices to be diagonalized in our approach have the same dimension as the matrices considered in the traditional approach if they belong to the $\mathbb{R}$- or $\mathbb{C}$-type; moreover, in the first case they are real instead of complex matrices. The matrices $H^{\lambda(\mathbf{Q})}$ even have half the dimension of the corresponding complex matrices appearing in the usual approach.

If one does not want to calculate with quaternions, $H^{\lambda(\mathbf{Q})}$ may be replaced by a complex matrix of double size according to (2.11). This shows that the solution of the eigenvalue problem suggested here is by no means more difficult than the usual one. In some respects the method proposed here is even superior to the traditional one: (i) The number of real parameters characterizing the matrices to be diagonalized reduces by a factor $\frac{1}{2}$ for the $\mathbb{R}$- and $\mathbb{Q}$-type representations; (ii) the eigenvectors are real or, more precisely, θ-invariant; and (iii) the method is free of the burden of an unjustified and misleading interpretation (time reversal) if the antiunitary operator θ is introduced on only formal grounds (lattice dynamics, Landau theory).

C. Quaternionic Hilbert space

In the preceding sections we extended the symmetry group, here again denoted by $\mathbf{G} = \{\mathbf{U}(x): x \in G\}$, to an algebra $\mathbf{A}_{\mathbf{F}}$, to obtain operators more useful for the eigenvalue problem of $\mathbf{H}'$ than the operators $\mathbf{U}'(x)$. For the commutative fields $\mathbf{F} = \mathbb{R}$ or $\mathbb{C}$ we chose $\mathbf{F}' = \mathbf{F}$, which was possible since the left-action of an element $f \in \mathbf{F}$ on an element $v \in \mathcal{H}_{\mathbf{F}}$ could be defined by $fv = vf$. Consequently the left-multiplicators f commuted with every operator in $\mathcal{H}_{\mathbf{F}}$, especially with $\mathbf{H}'$, and therefore the linear combinations of the elements $\mathbf{U}'(x)$ with coefficients from $\mathbf{F}$, i.e., the elements of $\mathbf{A}_{\mathbf{F}}$ also commuted with $\mathbf{H}'$. This does not hold for $\mathbf{F} = \mathbb{Q}$ since the noncommutativity of $\mathbb{Q}$ entails in general $fv \neq vf$; $fv = vf$ for arbitrary $v \in \mathcal{H}_{\mathbf{Q}}$ is valid only if $f \in$. We therefore can extend $\mathbf{G}$ only to $\mathbf{A}_{\mathbf{R}}$ and have to exploit the properties of a *representation of the real group algebra in a quaternionic Hilbert space* $\mathcal{H}_{\mathbf{Q}}$.

Since the relations between the elements of $\mathbf{A}_{\mathbf{R}}$ are the same as in the real case we are again given operators $\mathbf{E}^\lambda_{jJ}$ and $\mathbf{F}^\lambda_r$ satisfying Eqs. (4.9) and (4.25). The difference to the real case appears in the construction of the *symmetry-adapted basis* for it is possible to find, starting from the subspaces $\mathbf{E}^\lambda_{00}\mathcal{H}_{\mathbf{Q}}$, a basis $\{v^{\lambda m}_j\}$ of $\mathcal{H}_{\mathbf{Q}}$ such that the following relations are valid:

$$\mathbf{E}^\Lambda_{J'J}, v^{\lambda m}_j = \delta_{\lambda\Lambda}\delta_{J'j} v^{\lambda m}_J ,$$
$$\mathbf{F}^\Lambda_r v^{\lambda m}_j = \delta_{\lambda\Lambda} v^{\lambda m} f_r, \; f_r \in \mathbf{F}(=\mathbb{Q}), \quad (4.38)$$
$$m = 0, \ldots, m^\lambda_{\mathbf{Q}} - 1, \quad (4.39)$$
$$\dim \mathcal{H}_{\mathbf{Q}} = \sum_\lambda d^\lambda m^\lambda_{\mathbf{Q}} . \quad (4.40)$$

Which numbers f_r appear on the right side of Eq. (4.38) depends on the type of the representation λ: For the R-type it is 1; for the C-type it is 1, i; and for the Q-type it is $1, i, j, k$.

The basis $\{v_j^{\lambda m}\}$ is symmetry adapted since

$$\mathbf{U}'(x)v_J^{\lambda m} = \sum_j v_j^{\lambda m} Q_{jJ}^{\lambda}(\mathbf{x}),$$

$\{Q^{\lambda}(x) : x \in G\}$ = quaternionic irreducible matrix representation of G, $Q^{\lambda}(\mathbf{x})$ hyperunitary,

$$\dim Q^{\lambda}(\mathbf{x}) = d^{\lambda}. \tag{4.41}$$

As in the case of a real or complex Hilbert space the matrix representation of $\mathbf{H}'$ belonging to the symmetry-adapted basis has a block form with chains of d^{λ} identical matrices arranged along the diagonal:

$$\langle v_j^{\lambda m}, \mathbf{H}' v_J^{\lambda M} \rangle = \delta_{\lambda\Lambda}\delta_{jJ} H_{mM}^{\lambda}, \quad H_{mM}^{\lambda} = \langle v_0^{\lambda m}, \mathbf{H}' v_0^{\lambda M} \rangle. \tag{4.42}$$

The elements H_{mM}^{λ} are quaternions but the transformation properties of the basis elements $v_j^{\lambda m}$ under the operators $\mathbf{F}_r^{\lambda}$ [Eq. (4.38)] and the commutativity of these operators with $\mathbf{H}'$ sometimes forces them to be complex or even real, entailing a further simplification of the eigenvalue problem of $\mathbf{H}'$:

$$\begin{aligned} &\mathbf{R}\text{-type: } H_{mM}^{\lambda} = H_{mM}^{\lambda(\mathbf{Q})} \in \mathbb{Q}, \\ &\mathbb{C}\text{-type: } H_{mM}^{\lambda} = H_{mM}^{\lambda(\mathbf{C})} \in \mathbf{C}, \\ &\mathbb{Q}\text{-type: } H_{mM}^{\lambda} = H_{mM}^{\lambda(\mathbf{R})} \in \mathbb{R}. \end{aligned} \tag{4.43}$$

The matrices $H^{\lambda(\mathbf{F})}$ define operators $\mathbf{H}^{\lambda}$ in Hilbert spaces $\mathcal{H}_{\hat{\mathbf{F}}}^{\lambda}$ of dimension $m_{\mathbf{Q}}^{\lambda}$. A solution of the eigenvalue problem of $\mathbf{H}'$ is obtained by successively solving the eigenvalue problems of these operators.

$$\sigma(\mathbf{H}') = \bigcup_{\lambda} \sigma(\mathbf{H}^{\lambda}), \tag{4.44}$$

$$\deg(\epsilon, \mathbf{H}') = \sum_{\lambda} d^{\lambda} \deg(\epsilon, \mathbf{H}^{\lambda}). \tag{4.45}$$

Each eigenvector of an operator $\mathbf{H}^{\lambda}$ defines d^{λ} elements of $\mathcal{H}_{\mathbf{Q}}$ which are linear combinations over $\hat{\mathbf{F}}$ of elements $v_j^{\lambda m}$ (j fixed) belonging to the subspaces $\mathbf{E}_{jj}^{\lambda}\mathcal{H}_{\mathbf{Q}}$ and are eigenvectors of $\mathbf{H}'$. Together they constitute a symmetry-adapted basis of $\mathcal{H}_{\mathbf{Q}}$.

If $\mathbf{H}'$ corresponds to an operator $\mathbf{H}$ in a complex Hilbert space $\mathcal{H}_{\mathbf{C}}$ [Sec. 3B (i)] then

$$\sigma(\mathbf{H}) = \bigcup_{\lambda} \sigma(\mathbf{H}^{\lambda}), \tag{4.46}$$

$$\deg(\epsilon, \mathbf{H}) = 2\sum_{\lambda} d^{\lambda} \deg(\epsilon, \mathbf{H}^{\lambda}). \tag{4.47}$$

Hence the minimal degeneracies are equal to $2d^{\lambda}$, which again agrees with the result derived with the aid of the antiunitary operator θ:

$$\begin{aligned} &\mathbf{R}\text{-type: } d^{\lambda} = d_{\mathbf{C}}^{\lambda}, \ m_{\mathbf{Q}}^{\lambda} = \tfrac{1}{2} m_{\mathbf{C}}^{\lambda}, \\ &\mathbf{C}\text{-type: } d^{\lambda} = d_{\mathbf{C}}^{\lambda^+} = d_{\mathbf{C}}^{\lambda^-}, \ m_{\mathbf{Q}}^{\lambda} = m_{\mathbf{C}}^{\lambda^+} = m_{\mathbf{C}}^{\lambda^-}, \\ &\mathbb{Q}\text{-type: } d^{\lambda} = \tfrac{1}{2} d_{\mathbf{C}}^{\lambda}, \ m_{\mathbf{Q}}^{\lambda} = m_{\mathbf{C}}^{\lambda}. \end{aligned} \tag{4.48}$$

A basis of $\mathcal{H}_{\mathbf{C}}$ consisting of eigenvectors of $\mathbf{H}$ is obtained from the symmetry-adapted eigenvectors of $\mathbf{H}'$ by means of Eq. (2.50). A comparison of this method with the traditional one gives a picture similar to the real case with the roles of the R- and the Q-type representations interchanged. The advantages of our method are: (i) Only half the number of real parameters is needed to characterize the matrices belonging to representations of the R- or the Q-type; and (ii) the eigenvectors transform in a simple way under the antiunitary operator θ [see Eq. (2.34)].

5. CALCULATION OF THE OPERATORS E_{jJ}^{λ} AND F_r^{λ} USING COMPLEX IRREDUCIBLE MATRIX REPRESENTATIONS

To be a true alternative to the usual method our program must include the rules according to which the projectors and shift operators are obtained from the norm-preserving operators given in the beginning. This amounts to knowing the functions appearing in the representations (4.3) of these operators. It is well known[14] that in the complex case

$$\mathbf{E}_{jJ}^{\lambda} = d_{\mathbf{C}}^{\lambda} M_x C_{jJ}^{\lambda *}(\mathbf{x})\, \mathbf{U}(x), \tag{5.1}$$

where $C_{jJ}(\mathbf{x})$ is the jJ element of a complex unitary irreducible matrix representation. The information needed in this case is therefore a complete set of these representations; this can be found in the literature for most of the groups which are of interest for applications in physics.

For the real group algebra a formula similar to (5.1) exists, viz.

$$\mathbf{E}_{jJ}^{\lambda} \mathbf{F}_r^{\lambda} = r^{\lambda} d^{\lambda} M_x R_{jr,J0}^{\lambda}(\mathbf{x}) \mathbf{U}'(x), \tag{5.2}$$

from which the desired operators are obtained by

$$\mathbf{E}_{jJ}^{\lambda} = \mathbf{E}_{jJ}^{\lambda} \mathbf{F}_0^{\lambda}, \quad F_r^{\lambda} = \sum_j \mathbf{E}_{jj}^{\lambda} \mathbf{F}_r. \tag{5.3}$$

The function $R_{jr,J0}^{\lambda}(\mathbf{x})$ in (5.2) is the $jr, J0$-element of a real orthogonal irreducible matrix representation of dimension $r^{\lambda} d^{\lambda}$; hence a complete set of this kind of representations has to be known to apply formula (5.2). In principle a complete set of representations R^{λ} can be constructed *ab initio*, starting with the multiplication law of the compact group G and proceeding in a manner analogous to that in the complex case [cf. Ref. 5 Chaps. 3,4]. Here, however, we adopt the practitioner's point of view and give only recipes for how to get real irreducible representations from complex ones, supposing that a complete set of these representations is already at hand.

First one has to find out what kind each representation C^{λ} is, since this determines how one proceeds. For complex irreducible representations of the *third kind* occurring in pairs the C-*type* real irreducible representation correspondong to the *pair* $C^{\lambda +}$, $C^{\lambda -}$ is simply

$$R_{jr,JR}^{\lambda}(\mathbf{x}) = R_{rR}^{\mathbf{C}}[C_{jJ}^{\lambda \dagger}(\mathbf{x})], \quad R^{\mathbf{C}} \text{ see (2.13)}. \tag{5.4}$$

More effort is needed for the two other kinds of representations for which unitary matrices Z^{λ} exist such that

$$Z^{\lambda} C^{\lambda}(\mathbf{x}) = C^{\lambda *}(\mathbf{x}) Z^{\lambda} \text{ for all } r \in G. \tag{5.5}$$

If, Eq. (5.5) is explicitly written down for a sufficient number of group elements (the generators of G) the matrix Z^λ can be determined from this set of linear equations and the unitary condition up to a phase factor. The two kinds of complex representations are distinguished by the product $Z^\lambda Z^{\lambda *}$: For representations of the *first kind* (R-*type*) $Z^\lambda Z^{\lambda *} = +E^\lambda$ (1-matrix) whereas $Z^\lambda Z^{\lambda *} = -E^\lambda$ for representations of the second kind.[1] If $Z^\lambda Z^{\lambda *} = +E^\lambda$,

$$R^\lambda(\mathbf{x}) = (Z^\lambda - \omega E^\lambda)^{-1} C^\lambda(\mathbf{x})(Z^\lambda - \omega E^\lambda), \quad \omega \in \mathbb{C}, \quad |\omega| = 1 \tag{5.6}$$

is the corresponding real irreducible representation provided that $Z^\lambda - \omega E^\lambda$ is nonsingular, i.e., ω is not an eigenvalue of Z^λ. This can be checked by means of the minimal or characteristic polynomial of Z^λ. Any other nonsingular matrix transforming $C^\lambda(\mathbf{x})$ into a real representation $\bar{R}^\lambda(\mathbf{x})$ would work equally well.

In the case of a representation of the *second kind* ($\mathbb{Q}$-*type*) we note that

$C^\lambda(\mathbf{x})$ is $\mathbb{C}$-equivalent to representations $\bar{C}^\lambda(\mathbf{x})$ having a special structure, viz., $\bar{C}^\lambda_{jr,JR}(\mathbf{x})$

$$= C^{\mathbf{Q}}_{rR}[\mathfrak{Q}^\lambda_{jJ}(\mathbf{x})], \; C^{\mathbf{Q}}, \text{ see (2.11)}. \tag{5.7}$$

The real irriducible representation $\{R^\lambda(\mathbf{x}) : x \in G\}$ corresponding to the complex representation $\{C^\lambda(\mathbf{x}) : x \in G\}$,

$$R^\lambda_{jr,JR}(\mathbf{x}) = R^{\mathbf{Q}}_{rR}[\mathfrak{Q}^\lambda_{jJ}(\mathbf{x})], \quad R^{\mathbf{Q}} \text{ see (2.12)}, \tag{5.8}$$

is then simply obtained from the matrices $\bar{C}^\lambda(\mathbf{x})$. This is all that has to be done if the original representation C^λ already has the structure (5.7) [e.g., the 2-dimensional representation of SU(2)]. In general, however, R^λ (and $\bar{C}^\lambda$) have to be determined from C^λ via a detour through the quaternionic representations Q^λ. To find Q^λ one first has to determine the matrix Z^λ satisfying (5.5) and

$$Z^\lambda Z^{\lambda *} = -E^\lambda . \tag{5.9}$$

Then the quaternionic matrix

$$P^\lambda_0 = \tfrac{1}{2}(E^\lambda + Z^\lambda j) = P^{\lambda\dagger}_0 = (P^\lambda_0)^2 \tag{5.10}$$

is a projection matrix of dimension $d^\lambda_{\mathbb{C}} = 2d^\lambda$ and there exist d^λ column vectors w^λ_{0j} with $2d^\lambda$ quaternionic components $(w^\lambda_{0j})_n$ such that

$$(P^\lambda_0 w^\lambda_{0j})_n = (w^\lambda_{0j})_n, \quad \langle w_{0j}, w_{0J} \rangle = \sum_n (w_{0j})^*_n (w_{0J})_n = \delta_{jJ} . \tag{5.11}$$

If these vectors are determined by application of P^λ_0 and a Schmidt process the problem is solved since

$$Q_{jJ}(\mathbf{x}) = \sum_n (w^\lambda_{0j})^*_n (C^\lambda(\mathbf{x}) w^\lambda_{0J})_n . \tag{5.12}$$

That the representation $\bar{C}^\lambda$ obtained from this Q^λ is $\mathbf{C}$-equivalent to C^λ is proved in Sec. 6B.

6. PROOFS

A. Construction of bases

A large number of arguments refer to bases transforming in a peculiar way under certain operators. The first thing we have to prove is the claimed existence of the bases adapted to the commuting algebra $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ [Eqs. (2.32), (2.34), (2.36), and (2.38)]. In the definition of the algebra $\mathbf{A}(\mathcal{H}_{\mathbf{F}})$ we supposed $[\mathbf{A}(\mathcal{H}_{\mathbf{F}})]$ to contain a basis of $\mathcal{H}_{\mathbf{F}}$, say $\{v_n\}$. Since $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ is defined on $\mathcal{H}_{\mathbf{F}}$ and $\mathbf{AC}v_n = \mathbf{CA}v_n$ for $\mathbf{A} \in \mathbf{A}(\mathcal{H}_{\mathbf{F}})$, $\mathbf{C} \in \mathbf{C}(\mathcal{H}_{\mathbf{F}})$, the set $\{\mathbf{C}v_n\}$ also belongs to $\mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{F}})]$. If $\mathbf{C}_1, \ldots$ are the generators of $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$ we can use the overcomplete set $\{v_n, \mathbf{C}_1 v_n, \ldots\} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{F}})]$ to construct a basis of $\mathcal{H}_{\mathbf{F}}$ belonging to $\mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{F}})]$ and adapted to $\mathbf{C}(\mathcal{H}_{\mathbf{F}})$.

In case 1 the generating element is θ. Suppose we have succeeded in constructing M orthonormalized vectors $w_0, \ldots, w_{M-1}$ with $\theta w_m = w_m$ and dropped all pairs $v_N, \theta v_N$ linearly dependent on $\{w_0, \ldots, w_{M-1}\}$ from the overcomplete set $\{v_n, \theta v_n\}$. Let $v, \theta v$ be the components of the first pair in this set orthogonal to $w_0 \ldots w_{M-1}$. Then there are three possibilities: (i) v and θv are orthogonal. In this case the vectors $v_+ = v + \theta v$, $v_- = [v - \theta v]i$ are orthogonal and invariant under θ. Normalization yields two new basis elements $w_M, w_{M+1} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{C}})]$; (ii) v and θv are linearly independent but not orthogonal. If $c \in \mathbb{C}$ is defined by $0 \neq \langle v, \theta v \rangle = c^2$ the vectors $v_+ = vc + (\theta v)c^*$, $v_- = [vc - (\theta v)c^*]i$ are orthogonal and invariant under θ. Normalization yields two basis elements $w_M, w_{M+1} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{C}})]$. (iii) v and θv are linearly dependent, say $\theta v = v\gamma$. Since θ is antiunitary $|\gamma| = 1$. The vectors $v_+ = v + \theta v = v(1+\gamma)$ and $v_- = v - (\theta v)i = v(1-\gamma)i$ are linearly dependent, invariant under θ, and cannot vanish simultaneously. Normalization yields a new basis element $w_M \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{C}})]$.

In case 2 we proceed in quite an analogous manner. Since $\theta^2 = -1$ and $\theta v = v\gamma$ implies $v = 0$ we are left with only two possibilities: (i) v and θv are orthogonal, $v_0 = v$ and $v = \theta v$ satisfy (2.34) and yield two new basis elements w_{2M}, $w_{2M+1} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{C}})]$. (ii) v and θv are linearly independent but not orthogonal. Defining c as in case 1, the vectors $v_0 = vc - (\theta v)c^*$, $v_1 = vc + (\theta v)c^*$ are orthogonal and of equal norm; they satisfy (2.34) and yield after normalization two new basis elements w_{2M}, $w_{2M+1} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{C}})]$.

Cases 3 and 4 can be treated simultaneously. Suppose as before that we have already obtained an orthonormalized set $w_{00}, \ldots, w_{(M-1)\rho} \in \mathfrak{D}[\mathbf{A}(\mathcal{H}_{\mathbf{R}})]$ transforming in the right way under the operators $\mathbf{F}_r$ and that all elements linearly dependent on $w_{00}, \ldots, w_{(M-1)\rho}$ have been rejected from the overcomplete set $\{v_{n0} = v_n, \ldots, v_{n\rho} = \mathbf{F}_\rho v_n\}$. Furthermore, let $v_0, \ldots, v_\rho$ be the components of the first ρ-tuple orthogonal to $w_{00}, \ldots, w_{(M-1)\rho}$. Then $v_r = \mathbf{F}_r v_0$ since the linear hull of $w_{00}, \ldots, w_{(M-1)\rho}$ is invariant under the operators $\mathbf{F}_r$. Moreover $\langle v_r, v_s \rangle = \delta_{r,s} \| v_0 \|^2$ because

$$\langle v_r, v_R \rangle = \langle \mathbf{F}^\dagger_r v_0 \sigma_{r,r}, \mathbf{F}_R v_0 \rangle = \langle v_0, \mathbf{F}_{rR} v_0 \rangle \sigma_{r,r} \sigma_{r,R} = \langle v_R, v_r \rangle = \langle v_0, \mathbf{F}_{rR} v_0 \rangle \sigma_{R,R} \sigma_{R,r} . \tag{6.1}$$

Inspection of the matrix σ [Eq. (2.4) or (2.5)] shows that $\langle v_0, \mathbf{F}_{rR} v_0 \rangle = \delta_{rR} \| v_0 \|^2$. It is therefore sufficient to normalize v_0 to obtain new basis elements $w_{M0}, \ldots, w_{M\rho} \in \mathfrak{D}[A(\mathcal{H}_{\mathbf{R}})]$.

The arguments just given also apply to the construction of a symmetry adapted basis of $\mathcal{H}_{\mathbf{R}}$. Since (4.25) in-

cludes (2.35) or (2.37)

$$\langle \mathbf{F}^{\lambda}_{\nu} v, \mathbf{F}^{\Lambda}_{R} w\rangle = \delta_{\lambda\Lambda}\delta_{\nu R}\langle v, w\rangle . \tag{6.2}$$

The properties of the operators $\mathbf{E}^{\lambda}_{jJ}$ [Eq. (4.9)] imply

$$\langle \mathbf{E}^{\lambda}_{j0} v, \mathbf{E}^{\Lambda}_{J0} w\rangle = \langle v, \mathbf{E}^{\lambda}_{0j}\mathbf{E}^{\Lambda}_{J0} w\rangle = \delta_{\lambda\Lambda}\delta_{jJ}\langle v, \mathbf{E}^{\lambda}_{00} w\rangle$$
$$= \delta_{\lambda\Lambda}\delta_{jJ}\langle \mathbf{E}^{\lambda}_{00} v, \mathbf{E}^{\lambda}_{00} w\rangle . \tag{6.3}$$

It is therefore sufficient to construct bases $\{w^{\lambda}_{0\nu}\}$ of the subspaces $\mathbf{E}^{\lambda}_{00}\mathcal{H}_{\mathbf{R}}$ satisfying $w^{\lambda}_{0\nu} = \mathbf{F}^{\lambda}_{\nu} w^{\lambda}_{00}$ and to extend these subbases by application of the shift operators $\mathbf{E}^{\lambda}_{j0}$ to a basis of $\mathcal{H}_{\mathbf{R}}$. Equation (6.3) holds true also for complex or quaternionic Hilbert spaces. Since in $\mathcal{H}_{\mathbf{C}}$ there are no operators $\mathbf{F}^{\lambda}_{\nu}$ except

$$\mathbf{F}^{\lambda}_{0} = \sum_{j} \mathbf{E}^{\lambda}_{jj} = \mathbf{E}^{\lambda} \tag{6.4}$$

and $\mathbf{F}^{\lambda}_{1} = \mathbf{E}^{\lambda} i$ they do not generate orthogonal vectors. The only thing we still have to show is how to obtain a basis of $\mathbf{E}^{\lambda}_{00}\mathcal{H}_{\mathbf{Q}}$ consisting of common eigenvectors of the operators $\mathbf{F}^{\lambda}_{\nu}$. But this can again be achieved by a generalized Schmidt process since

$$v = \sum_{R} \mathbf{F}^{\lambda}_{R}\, \bar{v} f_{R}\sigma_{R,R} \Rightarrow \mathbf{F}^{\lambda}_{\nu} v = v f_{\nu} , \tag{6.5}$$

and if $\bar{w}$ is orthogonal to v as defined in (6.5) then $\langle \mathbf{F}_{R}\bar{w}, v\rangle = 0$ and hence $\langle w, v\rangle = 0$ for w obtained from $\bar{w}$ according to (6.5).

B. The structure of matrix representatives

We first verify the structure of the matrix representation A with respect to a basis adapted to the commuting algebra [Eqs. (2.42)–(2.44) and the reality of A_{mM} in case 1]. In case 1

$$0 = (A\theta - \theta\mathbf{A})w_{N} = \sum_{N} w_{N}(A_{nN} - A^{*}_{nN}); \tag{6.6}$$

in case 2

$$0 = (\mathbf{A}\theta - \theta A)w_{M0} = \sum_{m} [w_{m0}(A_{m0,M1} - A^{*}_{m1,M0})$$
$$+ w_{m1}(A_{m1,M1} + A^{*}_{m0,M0})] ; \tag{6.7}$$

and in cases 3 and 4

$$\langle w_{mr}, \mathbf{A} w_{MR}\rangle = \langle \mathbf{F}_{\nu} w_{0}, \mathbf{A}\mathbf{F}_{R} w_{0}\rangle = \langle \mathbf{F}^{\dagger}_{R}\mathbf{F}_{\nu} w_{m0}\, \mathbf{A} w_{m0}\rangle$$
$$= \langle w_{m(rR)}, \mathbf{A} w_{m0}\rangle \sigma_{R,R}\sigma_{R\nu} . \tag{6.8}$$

Equations (6.6) and 6.7) refer to a complex Hilbert space whereas (6.8) applies to a real one. For a quaternionic Hilbert space the matrix elements with respect to a symmetry-adapted basis do not show a peculiar structure but are instead restricted to be complex or even real if the representation is of the $\mathbf{C}$- or $\mathbf{Q}$-type. This is a straight consequence of

$$\langle v, \mathrm{H} w\rangle = \langle v, \mathbf{F}^{\dagger}_{\nu}\mathrm{H}\mathbf{F}_{\nu} w\rangle = \langle \mathbf{F}_{\nu} v, \mathrm{H}\mathbf{F}_{\nu} w\rangle = \langle v f_{\nu}, \mathrm{H} w f_{\nu}\rangle$$
$$= f^{*}_{\nu}\langle v, \mathrm{H} w\rangle f_{\nu} , \tag{6.9}$$

and Eq. (2.8)

The second part of this section concerns the (possible) structure of the complex irreducible representations of the second kind[9,10] [cf. (5.7)]. We first note that if Y^{λ} is defined by

$$Y^{\lambda} = Z^{\lambda} j , \tag{6.10}$$

the properties of Z^{λ}, namely (5.5), $(Z^{\lambda})^{-1} = Z^{\lambda\dagger} = -Z^{\lambda *}$, and $Z^{\lambda}_{nN} \in \mathbf{C}$, imply

$$Y^{\lambda} = Y^{\lambda\dagger}, \quad (Y^{\lambda})^{2} = E^{\lambda} ; \tag{6.11}$$

$$-iY^{\lambda} i = -Y^{\lambda} ; \tag{6.12}$$

$$Y^{\lambda} C^{\lambda}(\mathbf{x}) = C^{\lambda}(\mathbf{x}) Y^{\lambda} \text{ for all } x \in G . \tag{6.13}$$

Equation (6.11) shows that Y^{λ} is self-adjoint with eigenvalues ± 1. The spectral decomposition is

$$Y^{\lambda} = P^{\lambda}_{0} - P^{\lambda}_{1}, \quad P^{\lambda}_{\nu} = \tfrac{1}{2}[E^{\lambda} + (-1)^{\nu} Y^{\lambda}] . \tag{6.14}$$

The degeneracy of both eigenvalues is equal since

$$\deg(+1, Y^{\lambda}) = \operatorname{trace} P^{\lambda}_{0} = \operatorname{trace}(-i)P^{\lambda}_{1} i$$
$$= \operatorname{trace} P^{\lambda}_{1} = \deg(-1, Y^{\lambda}) . \tag{6.15}$$

To verify (6.15) one has to use the definitions (6.14) and relation (6.12) and to observe that the trace of a self-adjoint quaternionic matrix is real.

Now since Y^{λ} commutes with $C^{\lambda}(\mathbf{x})$ the projection matrices P^{λ}_{ν} define two orthogonal subspaces of $l^{\lambda}(\mathbf{Q})$, the space of $\mathbf{Q}$-valued column vectors with scalar product

$$\langle v, w\rangle = \sum_{n} v^{*}_{n} w_{n} . \tag{6.16}$$

Both subspaces are invariant under $C^{\lambda}(\mathbf{x})$. If $\{w^{\lambda}_{0j}\}$ is a basis of $P^{\lambda}_{0} l^{\lambda}(\mathbf{Q})$ the elements

$$i w^{\lambda}_{0j} = w^{\lambda}_{1j} \tag{6.17}$$

form a basis of $P^{\lambda}_{1} l^{\lambda}(\mathbf{Q})$ because of (6.12). Since $C^{\lambda}(\mathbf{x})i = iC^{\lambda}(\mathbf{x})$, both bases transform according to the same quaternionic representation

$$[C^{\lambda}(\mathbf{x}) w^{\lambda}_{rj}]_{n} = \sum_{J} [w^{\lambda}_{rJ}]_{n} Q^{\lambda}_{Jj}(\mathbf{x}) . \tag{6.18}$$

The column vectors w^{λ}_{rj} may be collected into a hyperunitary matrix W^{λ} with elements

$$W^{\lambda}_{n,jr} = (w^{\lambda}_{jr})_{n} . \tag{6.19}$$

We now use the relation

$$\operatorname{trace}(A'B' + B'^{\dagger}A'^{\dagger}) = \operatorname{trace}(B'A' + A'^{\dagger}B'^{\dagger}) , \tag{6.20}$$

which can be proved using the isomorphism (2.24), the commutativity of the elements of $A = R^{\mathbf{Q}}[A']$ and $B = R^{\mathbf{Q}}[B']$, and the self-adjointness of the quaternionic matrices within the brackets.

$$\chi^{\lambda}(\mathbf{x}) + \chi^{\lambda *}(\mathbf{x}) = \operatorname{trace}[C^{\lambda}(\mathbf{x}) + C^{\lambda\dagger}(\mathbf{x})]$$
$$= \operatorname{trace} W^{\lambda\dagger}[C^{\lambda}(\mathbf{x}) + C^{\lambda\dagger}(\mathbf{x})]W^{\lambda} = 2\sum_{j}[Q^{\lambda}_{jj}(\mathbf{x}) + Q^{\lambda *}_{jj}(\mathbf{x})]$$
$$= 2\operatorname{trace} C^{\mathbf{Q}}[Q^{\lambda}(\mathbf{x})] = 2\operatorname{trace}\bar{C}^{\lambda}(\mathbf{x}) = 2\bar{\chi}^{\lambda}(\mathbf{x}) . \tag{6.21}$$

Since the character $\chi^{\lambda}(\mathbf{x})$ is real as a consequence of (5.5) this establishes the $\mathbf{C}$-equivalence of C^{λ} and $\bar{C}^{\lambda} = C^{\mathbf{Q}}(Q^{\lambda})$.

C. Equivalence of eigenvalue problems

It was stated in Sec. 2C that a self-adjoint operator H in a Hilbert space $\mathcal{H}_{\mathbf{F}}$ admits a spectral decomposition of the form (2.27) or, stated alternatively, that the col-

umns of a norm-preserving matrix U diagonalizing the matrix representative H of $\mathbf{H}$ define eigenvectors of $\mathbf{H}$ which constitute a basis of $\mathcal{H}_{\mathbf{F}}$. These are well-known facts for $\mathbf{F}=\mathbf{R}$ or $\mathbf{C}$. We sketch here a proof for $\mathbf{F}=\mathbf{Q}$ which is typical of the equivalence of eigenvalue problems. The first step is to find a matrix representative H' of the self-adjoint operator $\mathbf{H}'$ and to replace the quaternionic matrix H' by the real matrix $H=\overline{R}^{\mathbf{Q}}(H')$ which, as is easily verified, is symmetric because of the self-adjointness of H', the definition of q^* [Eqs. (2.6) and (2.5)], and the peculiar form of $\overline{R}^{\mathbf{Q}}$ [Eq. (2.14)]. Next suppose that v_0 is a normalized real column vector satisfying $Hv_0=v_0\epsilon$. If we introduce the matrices

$$F_r=\sum\oplus R^{\mathbf{Q}}(f_r) \tag{6.22}$$

then the column vectors $v_{0r}=F_rv_0$ are (i) all eigenvectors of H since (6.22) and (2.15) imply that $[H,F_r]=0$; and (ii) orthonormalized since their components can be divided into quadrupels, the correponding quadruples having the form of the columns of $\overline{R}^{\mathbf{Q}}$ [see (2.14)]. For the projection matrix $P^{(0)}$ reproducing the four vectors v_{0r} we have [cf. (2.53)]

$$P^0_{mr,MR}=\sum_s\langle w_{mr},v_{0s}\rangle\langle v_{0s},w_{MR}\rangle=\sum_s\overline{R}^{\mathbf{Q}}_{rs}(q_{0m})\overline{R}^{\mathbf{Q}}_{Rs}(q_{0M})$$
$$=\sum_s\overline{R}^{\mathbf{Q}}_{rs}(q_{0m})\overline{R}^{\mathbf{Q}}_{sR}(q^*_{0M})=\overline{R}^{\mathbf{Q}}_{rR}(P'_{mM}),\ P^0_{mM}{}'=q_{0m}q^*_{0M}\,. \tag{6.23}$$

In the next step one looks for a normalized eigenvector v_1 orthogonal to all four vectors v_{0r}, v_1 gives rise to four eigenvectors $v_{1r}=F_rv_1$ orthogonal to the vectors v_{0s}. Continuing in this manner one obtains two things: (i) A spectral decomposition of H that can be transformed into a spectral decomposition of H'; and (ii) a hyperunitary matrix U' which is the image of the orthogonal matrix U obtained by collecting the columns v_{0r}, $v_{1r},\ldots$ which diagonalizes the self-adjoint quaternionic matrix.

7. CONCLUSION

In this paper we discussed relations between eigenvalue problems in real, complex, and quaternionic Hilbert spaces. In all cases a compact symmetry group was supposed to exist and it was shown how to simplify the eigenvalue problem by means of a suitably-defined group algebra. For complex Hilbert spaces it was assumed that an antiunitary operator θ with $\theta^2=\pm1$ exists which commutes with the self-adjoint operator and with all unitary operators in the symmetry group. It was shown that in this case the extra degeneracies caused by θ can equally well be obtained by considering related eigenvalue problems in real ($\theta^2=+1$) or quaternionic ($\theta^2=-1$) Hilbert spaces. This approach, dispensing with antilinear operators, makes the underlying algebraic structure more transparent, especially the relation between the three kinds of matrix representations and the three fields $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{Q}$. It leads to a reduction of the real parameters needed in the calculation [see the remarks at the end of Secs. 4B and 4C] and is also more natural if the original problem is given in a noncomplex Hilbert space (lattice dynamics, Landau theory). To be a true alternative to the usual approach the complex irreducible matrix representations must be given in a peculiar form, namely the representations of the first kind in real form and those of the second kind must be composed of 2×2 matrices of the form (2.11). How to obtain these forms if they are not originally given is described in detail in Sec. 5.

[1] L. Jansen and M. Boon, *Theory of finite groups, Applications in Physics* (North-Holland, Amsterdam, 1967). See especially Chap. II, Sec. 5.9 and Chap. III, Sec. 5.1.

[2] A. A. Maradudin and S. H. Vosko, Rev. Mod. Phys. **40**, 1 (1968).

[3] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Solid State Physics Suppl. 3* (Academic, New York, 1971); W. Jones and N. H. March, *Theoretical Solid State Physics*, Vol. 1 (Wiley, New York, 1973).

[4] T. G. Worlton and J. L. Warren, Comp. Phys. Commun. **3**, 88 (1972); **4**, 382 (*E*) (1972); J. L. Warren, Phys. Rev. B **9**, 3603 (1974).

[5] R. Dirl and P. Kasperkovitz, *Anwendungen in der Atom-und Festkorperphysik Gruppentheorie* (Vieweg, Braunschweig, 1977).

[6] Ref. 2, pp. 7, 11, and 13.

[7] E. C. G. Stueckelberg, Helv. Phys. Acta **33**, 727 (1960); E. C. G. Stueckelberg and M. Guenin, Helv. Phys. Acta **34**, 621 (1961).

[8] C. Chevalley, *Theory of Lie Groups* (Princeton U. P., Princeton, N. J., 1946), pp. 16–24.

[9] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962); J. M. Jauch, "Projective Representations of the Poincaré Group in a Quaternionic Hilbert Space," in *Group Theory and its Applications*, Vol. I, edited by E. M. Loebl (Academic, New York, 1968).

[10] D, Finkelstein, J. M. Jauch, and D. Speiser, J. Math. Phys. **4**, 136 (1963).

[11] G. Frobenius and I. Schur, Sitzgsber. Preuss. Akad. B **II**, 186 (1906).

[12] P. Kasperkovitz, J. Phys. C **9**, 4069 (1976).

[13] P. Kasperkovitz, "Representation Theory of Compact Groups over Fields of Characteristic Zero," J. Math. Phys. **22**, 2417 (1981).

[14] M. A. Naimark, *Normed Rings* (Nordhoff, Groningen, 1964).

Representation theory of compact groups over fields of characteristic zero

P. Kasperkovitz

Institut für Theoretische Physik, TU Wien, Karlsplatz 13, Wien A-1040, Austria
(Received 16 July 1980; accepted for publication 5 September 1980)

The group algebras over the reals, the complex numbers, and the quaternions, are discussed and compared for a fixed compact group. Special emphasis is put on the properties of ring bases of the minimal two-sided ideals and their relation to the matrix representations of the group irreducible over the fields cited above.

PACS numbers: 02.20. + b, 03.65.Fd

1. INTRODUCTION

In the preceding paper[1] (hereafter referred to as I) it has been shown that the symmetry properties caused by the existence of a commuting antiunitary operator θ, $\theta^2 = \pm 1$, can be taken into account equally well reformulating the eigenvalue problem in a noncomplex Hilbert space. The existence of a symmetry group is conserved in this process, provided the unitary operators $\mathbf{U}(x)$, $x \in G$, all commute with θ. If one wants to use the new symmetry group $\{\mathbf{U}'(x): x \in G\}$ ($\cong\{\mathbf{U}(x): x \in G\}$) to simplify the eigenvalue problem of the new self-adjoint operator $\mathbf{H}'$ and to make predictions on the minimal degeneracies of its eigenvalues, one has to study the representations of G over the fields $\mathbf{R}$ (real numbers) and $\mathbf{Q}$ (quaternions).

Representations in real and quaternionic Hilbert spaces have been discussed some years ago for the noncompact Lorentz group[2,3] and for SU(2)[3]. There exists also a paper of Finkelstein[4] *et al.* on the reducibility of irreducible complex matrix representations of a compact group if the carrier space is extended to a quaternionic one (see also Emch[5]). But I am not aware of any published discussion of representations of compact groups in real Hilbert spaces exceeding the rudiments given by Frobenius and Schur[6] nor of detailed information about the structure of the noncomplex group algebras comparable to what is available in the complex case.[7,8] The present paper aims to fill this gap. Of course there are no surprising new results since everything is a direct consequence of the following three facts: (i) The group algebra is semisimple. (ii) The minimal two-sided ideals are finite-dimensional. (iii) Their structure is determined by Wedderburn's theorem. However the detailed and concrete discussion of the ring bases (leading to projection and shift operators in applications, see I) and of the corresponding matrix representations should give more insight into the underlying algebraic structure, especially into the relation of the three classes of representations to the three fields $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{Q}$.

The paper is organized as follows: In Sec. 2 the group algebra $\mathbf{A}_\mathbf{F}$ is defined, its properties are discussed in general terms and then specialized to the three cases $\mathbf{F} = \mathbf{R}$, $\mathbf{F} = \mathbf{C}$, $\mathbf{F} = \mathbf{Q}$. In Sec. 3 the group algebras $\mathbf{A}_\mathbf{C}$ and $\mathbf{A}_\mathbf{Q}$ are viewed as extensions of $\mathbf{A}_\mathbf{R}$ following the general pattern described by van der Waerden.[9] In Sec. 4 the uniqueness of the ring bases given in Sec. 2 is discussed. Section 5 deals with the matrix representations over the different fields and the necessary proofs are given in Sec. 6. Throughout this paper the notation introduced in I-Sec. 2 is used and G is assumed to be a fixed compact group.

2. GROUP ALGEBRAS

A. General properties

The group algebra $\mathbf{A}_\mathbf{F}$ over the field $\mathbf{F}$ (= $\mathbf{R}$ or $\mathbf{C}$ or $\mathbf{Q}$) is, roughly speaking, the set of linear combinations of group elements with coefficients in $\mathbf{F}$. For a precise definition one has to introduce $L^2(G,\mathbf{F})$, the Hilbert space of $\mathbf{F}$-valued square-integrable functions defined on G. In this space the scalar product is defined by

$$\langle a, b\rangle = M_x a^*(x) b(x), \tag{2.1}$$

where $x \in G$ and M_x denotes averaging over G [see I-(4.4)]. As a consequence of the topology of G the Hilbert space $L^2(G,\mathbf{F})$ is usually separable[7,8] and its dimension is the same for $\mathbf{F} = \mathbf{R}, \mathbf{C}, \mathbf{Q}$, since

$$L^2(G,\mathbf{F}) = \mathbf{F}L^2(G,\mathbf{R}) = L^2(G,\mathbf{R})\mathbf{F}, \text{ i.e.,}$$
$$a \in L^2(G,\mathbf{F}) \Leftrightarrow a = \sum_r f_r a_r, \quad a_r \in L^2(G,\mathbf{R}). \tag{2.2}$$

On $L^2(G,\mathbf{F})$ a group of norm-preserving linear operators, $\mathbf{G} = \{\mathbf{x}\}$, is introduced via

$$\mathbf{x} \in \mathbf{G}, \ v \in L^2(G,\mathbf{F}): \mathbf{x}v(y) = v(x^{-1}y). \tag{2.3}$$

The regular representation $x \to \mathbf{x}$ is an isomorphism, $G \cong \mathbf{G}$, and $\mathbf{x}^{-1} = \mathbf{x}^*$. Beside $\mathbf{G}$ a set of bounded operators $\mathbf{a}$ on $L^2(G,\mathbf{F})$ is introduced, each operator being in one-to-one correspondence with a function $a \in L^2(G,\mathbf{F})$. If the set $\{\mathbf{a}\}$ is denoted by $\mathbf{A}_\mathbf{F}$ and the convolution $a*b$ of two elements $a, b \in L^2(G,\mathbf{F})$ is defined by

$$a*b(y) = M_x a(x) b(x^{-1}y), \tag{2.4}$$

the operator $\mathbf{a}$ is given by

$$\mathbf{a} \in \mathbf{A}_\mathbf{F}, \ v \in L^2(G,\mathbf{F}): \mathbf{a}v(y) = a*v(y). \tag{2.5}$$

Comparing (2.5) with (2.3) it is seen that

$$\mathbf{a} = M_x a(x)\mathbf{x}, \quad a \in L^2(G,\mathbf{F}). \tag{2.6}$$

$\mathbf{A}$ is a linear space over $\mathbf{F}$ and, because of (2.2), also over $\mathbf{R}$ ($\subseteq \mathbf{F}$).

$$\mathbf{A}_\mathbf{F} = \mathbf{F}\mathbf{A}_\mathbf{R} = \mathbf{A}_\mathbf{R}\mathbf{F}. \tag{2.7}$$

$\mathbf{A}_\mathbf{F}$ is even a Hilbert space isomorphic to $L^2(G,\mathbf{F})$ if the

 0022-2488/81/112417-10$01.00

scalar product is defined by

$$\langle \mathbf{a}, \mathbf{b} \rangle = \langle a, b \rangle . \tag{2.8}$$

A is also a ring (closed under addition and multiplication) if the product $\mathbf{ab}$ is defined by the successive action of the operators $\mathbf{b}$ (first) and $\mathbf{a}$ (second). $\mathbf{ab} = \mathbf{c}$ if, and only if, $a*b = c$. The set $\mathbf{A}_F$ with this Hilbert ring structure is called the group algebra of G over $\mathbb{F}$. For finite groups $\mathbf{G} \subset \mathbf{A}_R$ ($\subseteq \mathbf{A}_F$); if G is continuous the operators $\mathbf{x} \in \mathbf{G}$ can be approximated (in the weak topology) by elements $\mathbf{x}_\epsilon \in \mathbf{A}_F$ and the elements $\mathbf{a} \in \mathbf{A}_F$ by finite linear combinations $\sum a(x)\mathbf{x}$, $a(x) \in \mathbb{F}$. To have at hand a compact notation we combine $\mathbf{G}$ and $\mathbf{A}_F$ into a single set of bounded operators

$$\mathbf{B}_F = \mathbf{A}_F \cup \mathbf{G} . \tag{2.9}$$

$\mathbf{G}$ is closed under adjunction ($\mathbf{x} \to \mathbf{x}^* = \mathbf{x}^{-1}$) and so is $\mathbf{A}_F$ (and hence $\mathbf{B}_F$) if $\mathbf{a}^*$ is the operator adjoint to $\mathbf{a}$. The function corresponding to $\mathbf{a}^*$ is a^+, where

$$a^+(x) = a^*(x^{-1}) . \tag{2.10}$$

The group algebra $\mathbf{A}_F$ decomposes into simple constituents $\mathbf{A}_F^\lambda$ called minimal two-sided ideals. This means that there exists a decomposition

$$\mathbf{A}_F = \sum_\lambda \mathbf{A}_F^\lambda , \quad \text{i.e.,} \quad \mathbf{a} = \sum_\lambda \mathbf{a}^\lambda \quad \text{for all } \mathbf{a} \in \mathbf{A}_F , \tag{2.11}$$

such that

$$\lambda \neq \Lambda \Rightarrow \mathbf{a}^\lambda \mathbf{b}^\Lambda = 0 \text{ and } \langle \mathbf{a}^\lambda, \mathbf{b}^\Lambda \rangle = 0 \text{ for all } \mathbf{a}, \mathbf{b} \in \mathbf{A}_F ; \tag{2.12}$$

$$\mathbf{B}\mathbf{A}^\lambda = \mathbf{A}^\lambda \mathbf{B} = \mathbf{A}_F^\lambda , \text{ i.e.,}$$

$$\mathbf{a}^\lambda \in \mathbf{A}^\lambda \Rightarrow \mathbf{b}\mathbf{a}^\lambda, \mathbf{a}^\lambda \mathbf{b} \in \mathbf{A}_F^\lambda \text{ for all } \mathbf{b} \in \mathbf{B}_F . \tag{2.13}$$

The decomposition (2.11) is unique but depends on which field $\mathbb{F}$ has been chosen and a precise notation would be $\lambda(\mathbb{F})$ instead of λ. The same argument applies to the element $\mathbf{e}^\lambda$ generating $\mathbf{A}_F^\lambda$ and playing the role of the number 1 there,

$$\mathbf{A}_F^\lambda = \mathbf{A}_F \mathbf{e}^\lambda = \mathbf{e}^\lambda \mathbf{A}_F , \quad \mathbf{e}^\lambda \mathbf{a}^\lambda = \mathbf{a}^\lambda \mathbf{e}^\lambda = \mathbf{a}^\lambda , \quad \mathbf{e}^\lambda = \mathbf{e}^{\lambda *} . \tag{2.14}$$

The ideals $\mathbf{A}_F^\lambda$ are both subrings of the ring $\mathbf{A}_F$ closed under addition, multiplication, and adjunction, and subspaces of the Hilbert space $\mathbf{A}_F$ invariant under the elements $\mathbf{b} \in \mathbf{B}_F$ acting as multipliers from the left or from the right. Although intimately connected, the subring and the subspace properties should be clearly distinguished. As a subring, $\mathbf{A}_F^\lambda$ is isomorphic to a full finite-dimensional matrix algebra over a field (Wedderburn's structure theorem). This means that there exist $\rho^\lambda (= \tilde{\rho})$ elements $\mathbf{f}_r^\lambda \in \mathbf{A}_F^\lambda$, $\mathbf{f}_0^\lambda = \mathbf{e}^\lambda$, such that

$$\mathbb{F}^\lambda = \Big\{ \sum_{r=0}^{\rho^\lambda - 1} \mathbf{f}_r a_r : a_r \in \mathbb{R} \Big\} \cong \Big\{ \sum_{r=0}^{\tilde{\rho}-1} f_r a_r : a_r \in \mathbb{R} \Big\} = \tilde{\mathbb{F}}$$

$$= \mathbb{R} \text{ or } \mathbb{C} \text{ or } \mathbb{Q}; \ \mathbf{R}\mathbf{e}^\lambda = \mathbf{e}^\lambda \mathbb{R} \subseteq \mathbf{F}_F^\lambda \subseteq \mathbf{A}_F^\lambda , \tag{2.15}$$

and a set of $(d^\lambda)^2$ elements $\mathbf{e}_{kK}^\lambda \in \mathbf{A}_F^\lambda$, $k, K = 0, \ldots, d^\lambda - 1$, such that

$$\mathbf{e}_{kK}^{\lambda *} = \mathbf{e}_{Kk}^\lambda , \tag{2.16}$$

$$\mathbf{e}_{kK}^\lambda \mathbf{e}_{k'K'}^\lambda = \delta_{Kk'} \mathbf{e}_{kK'}^\lambda . \tag{2.17}$$

Every $\mathbf{a}^\lambda \in \mathbf{A}_F^\lambda$ has a unique expansion

$$\mathbf{a} = \sum_{kK} \mathbf{e}_{kK}^\lambda \mathbf{f}_{kK}^\lambda (\mathbf{a}^\lambda) = \sum_{kK} \mathbf{f}_{kK}^\lambda (\mathbf{a}^\lambda) \mathbf{e}_{kK}^\lambda , \quad \mathbf{f}_{kK}^\lambda (\mathbf{a}^\lambda) \in \mathbf{F}_F^\lambda . \tag{2.18}$$

The elements $\mathbf{e}_{kK}^\lambda$ add up and multiply like matrices having a 1 in the kK-position and zeros elsewhere, and the elements $\mathbf{f} \in \mathbf{F}_F^\lambda$ behave like numbers $\tilde{f} \in \tilde{\mathbf{F}}$. Therefore the elements $\mathbf{a}^\lambda \in \mathbf{A}_F^\lambda$ have matrix representatives $\tilde{F}^\lambda(\mathbf{a}^\lambda)$ with elements

$$\tilde{F}_{kK}^\lambda (\mathbf{a}^\lambda) = \sum_{r=0}^{\tilde{\rho}-1} f_r R_{kK}^{\lambda(r)}(\mathbf{a}) \in \tilde{\mathbf{F}} , \quad R_{kK}^{\lambda(r)}(\mathbf{a}) \in \mathbb{R} . \tag{2.19}$$

The real numbers $R_{kK}^{\lambda(r)}(\mathbf{a})$ can be determined from

$$\mathbf{e}_{kk}^\lambda \mathbf{a}^\lambda \mathbf{e}_{KK}^\lambda = \mathbf{e}_{kK}^\lambda \mathbf{f}_{kK}^\lambda (\mathbf{a}^\lambda) = \mathbf{e}_{kK}^\lambda \sum_{r=0}^{\rho^\lambda - 1} \mathbf{f}_r^\lambda R_{kK}^{\lambda(r)}(\mathbf{a}) . \tag{2.20}$$

Because of (2.12) and (2.13) the matrix representation $\mathbf{a}^\lambda \to \tilde{F}(\mathbf{a}^\lambda)$ can be extended to a matrix representation $\mathbf{b} \to \tilde{F}^\lambda(\mathbf{b})$ of $\mathbf{B}_F$:

$$\tilde{F}^\lambda(\mathbf{b}) = \tilde{F}^\lambda(\mathbf{b}^\lambda), \quad \mathbf{b}^\lambda = \mathbf{b}\mathbf{e}^\lambda = \mathbf{e}^\lambda \mathbf{b} \in \mathbf{A}_F^\lambda . \tag{2.21}$$

Equations (2.18) and (2.20) show that $\mathbf{A}_F^\lambda$ is a linear space over $\mathbb{R}$ of dimension $(d^\lambda)^2 \rho^\lambda$. This dimension must not be confused with the dimension of the subspace over $\mathbb{F}$; in fact if $\mathbb{F} \neq \mathbb{R}$ these two dimensions are distinct.

As a subspace $\mathbf{A}_F^\lambda$ decomposes into isomorphic subspaces invariant under the operators $\mathbf{b} \in \mathbf{B}_F$ acting from the left (but not invariant if these operators are multiplied from the right!). This means that there exists a basis $\{\mathbf{v}_p^{\lambda q} : p = 0, \ldots, n^\lambda - 1; q = 0, \ldots, m^\lambda - 1\}$ such that

$$\mathbf{a}^\lambda \in \mathbf{A}_F^\lambda : \mathbf{a}^\lambda = \sum_{pq} \mathbf{v}_p^{\lambda q} \langle \mathbf{v}_p^{\lambda q}, \mathbf{a}^\lambda \rangle , \tag{2.22}$$

$$\mathbf{b} \in \mathbf{B}_F : \mathbf{b}\mathbf{v}_p^{\lambda q} = \sum_p \mathbf{v}_p^{\lambda q} F_{pP}^\lambda (\mathbf{b}), \quad F_{pP}^\lambda (\mathbf{b}) = \langle \mathbf{v}_p^{\lambda q}, \mathbf{b}\mathbf{v}_P^{\lambda q} \rangle \in \mathbb{F} . \tag{2.23}$$

Equation (2.23) shows that $\mathbf{A}_F^\lambda$ decomposes into $m^\lambda n^\lambda$-dimensional subspaces each carrying the matrix representation $\mathbf{b} \to F^\lambda(\mathbf{b})$. Since $\mathbf{A}_F^\lambda$ is a linear space of dimension $m^\lambda n^\lambda$ over $\mathbb{F}$ and $\mathbb{F}$ has rank ρ over $\mathbb{R}$

$$m^\lambda n^\lambda \rho = (d^\lambda)^2 \rho^\lambda . \tag{2.24}$$

The sets $\{\mathbf{e}_{kK}^\lambda, \mathbf{f}_r^\lambda\}$ and $\{\mathbf{v}_r^{\lambda q}\}$ can be chosen in such a way that the following relations between $\tilde{F}^\lambda$ and F^λ are valid:

$$\mathbb{F} \supseteq \tilde{\mathbb{F}} : p = k, \ P = K; \ F_{kK}^\lambda (\mathbf{b}) = \tilde{F}_{kK}^\lambda (\mathbf{b}) , \tag{2.25}$$

$$\mathbb{F} \subseteq \tilde{\mathbb{F}} : p \to kr, \ P \to KR; \ F_{kr,KR}^\lambda (\mathbf{b}) = F_{rR}^{\tilde{\mathbb{F}}} [\tilde{F}_{kK}^\lambda (\mathbf{b})] , \tag{2.26}$$

where the matrices F are [cf. Eqs. I-(2.10)–(2.14)]

$$F^{\mathbb{F}}[f] = f , \tag{2.27}$$

$$R^{\mathbb{C}}[a + ib] = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} , \tag{2.28}$$

$$R^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} , \tag{2.29}$$

$$\bar{R}^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix} , \tag{2.30}$$

$$C^{\mathbb{Q}}[a + ib + jc + kd] = \begin{pmatrix} a + ib & c + id \\ -c + id & a - ib \end{pmatrix} . \tag{2.31}$$

The matrix representations $\tilde{F}^\lambda$ and F^λ are both irreducible matrix representations. For F^λ this means that it is impossible to find a matrix representation $\bar{F}^\lambda$ $\mathbb{F}$-

equivalent to F^λ such that for all $\mathbf{b}\in\mathbf{B}_\mathbf{F}$ $\bar{F}^\lambda(\mathbf{b})$ decomposes into a direct sum of smaller matrices, say $\bar{F}^\lambda_1(\mathbf{b})\oplus\bar{F}^\lambda_2(\mathbf{b})$; an analogous assertion holds for $\tilde{F}^\lambda$.

It should be noted that the irreducibility of a matrix representation depends both on the set of operators to be represented and on the field from which the matrix elements are taken. The representation may become reducible, i.e., equivalent to a direct sum of smaller representations, if the set of operators is reduced or if the field is extended. Starting from a given matrix representation of $\mathbf{B}_\mathbf{F}$, say F^λ, it is possible to restrict the operators to $\mathbf{B}_\mathbf{R}$, in symbols $F^\lambda\downarrow\mathbf{B}_\mathbf{R}$, and consider all representations which are $\mathbf{Q}$-equivalent to $F^\lambda\downarrow\mathbf{B}_\mathbf{R}$. If the propositions (**R**), (**C**), (**Q**) are defined according to

$$(\mathbf{F}):F^\lambda\downarrow\mathbf{B}_\mathbf{R}\text{ is }\mathbf{Q}\text{-equivalent to a direct sum of irreducible }\mathbf{F}\text{-valued matrix representations of }\mathbf{B}_\mathbf{R}, \tag{2.32}$$

then it turns out that (**Q**) always holds true; in some cases even (**C**) or (**R**) are satisfied. Observing that (**R**) $\Rightarrow$(**C**)$\Rightarrow$(**Q**) this may be used to partition the set of irreducible matrix representations into three classes.

$$\begin{aligned}&\mathbf{R}\text{-type: }(\mathbf{R})\text{ is valid.}\\ &\mathbf{C}\text{-type: }(\mathbf{C})\text{ is valid but }(\mathbf{R})\text{ is not.}\\ &\mathbf{Q}\text{-type: }(\mathbf{Q})\text{ is valid but }(\mathbf{C})\text{ and }(\mathbf{R})\text{ are not.}\end{aligned} \tag{2.33}$$

Because of (2.25), (2.26), and

$$F^{\mathbf{F}'}[f']\text{ is }\mathbf{Q}\text{-equivalent to }f'\oplus\ldots\oplus f'(\rho'/\rho\text{ terms}), \tag{2.34}$$

$\tilde{F}^\lambda$ and F^λ are always of the same type and the partitioning into three classes may be extended to the ideals $\mathbf{A}^\lambda_\mathbf{F}$.

B. The complex algebra $\mathbf{A}_\mathbf{C}$

Group algebras over $\mathbf{F}=\mathbf{C}$ are extensively discussed in the mathematical and physical literature (see, e.g., Refs. 7 and 8). I assume the reader to be familiar with this theory and quote the results only to show how they fit into the general scheme and how they are related to the other cases discussed below ($\mathbf{F}=\mathbf{R},\mathbf{Q}$). First the properties common to all types of representations are stated.

$$F^\lambda_\mathbf{C}=\{\mathbf{e}^\lambda c : c\in\mathbf{C}\}\cong\tilde{\mathbf{F}}=\mathbf{C}=\mathbf{F}, \tag{2.35}$$

$$d^\lambda=n^\lambda=m^\lambda;\quad k=p=j,\quad K=P=J, \tag{2.36}$$

$$\mathbf{v}^{\lambda J}_j=(n^\lambda)^{-1/2}\mathbf{e}^\lambda_{jJ}, \tag{2.37}$$

$$F^\lambda_{jJ}(\mathbf{b})=\tilde{F}^\lambda_{jJ}(\mathbf{b})=C^\lambda_{jJ}(\mathbf{b})=(n^\lambda)^{-1}\langle\mathbf{e}^\lambda_{jJ},\mathbf{b}^\lambda\rangle\in\mathbf{C}. \tag{2.38}$$

Since always $\tilde{\mathbf{F}}=\mathbf{C}$ this field does not give any hint to the type of the representation. However, to recognize of what type a complex irreducible representation is, it is not necessary to go back to the general definition (2.32) since in this case $\mathbf{C}$-equivalence suffices to distinguish the three types.

$$\begin{aligned}&\mathbf{R}\text{-type: }C^\lambda\text{ is }\mathbf{C}\text{-equivalent to a real irreducible representation }R^\lambda.\\ &\mathbf{C}\text{-type: }C^\lambda\text{ and }C^{\lambda *}\text{ are }\mathbf{C}\text{-inequivalent.}\\ &\mathbf{Q}\text{-type: }C^\lambda\text{ and }C^{\lambda *}\text{ are }\mathbf{C}\text{-equivalent but }\mathbf{C}\text{-inequivalent to a real irreducible representation.}\end{aligned} \tag{2.39}$$

Because $\mathbf{C}$-equivalence implies $\mathbf{Q}$-equivalence the definitions (2.39) of the **R**- and the **C**-type imply the definitions (2.33) provided that these representations remain irreducible if $\mathbf{C}$-equivalence is extended to $\mathbf{Q}$-equivalence. This is actually the case as will be shown in Secs. 4 and 6 (see also Ref. 4). Taking for granted that the two definitions are equivalent for the **R**- and the **C**-type, the definitions (2.39) and (2.33) are also equivalent for the $\mathbf{Q}$-type since it is the only remaining possibility in both cases. In the literature the propositions in (2.39) are used to define what is called the kind of the complex irreducible matrix representation; hence we have the following equivalence of terms:

$$\begin{aligned}&\mathbf{R}\text{-type}=1\text{st kind},\\ &\mathbf{C}\text{-type}=3\text{rd kind},\\ &\mathbf{Q}\text{-type}=2\text{nd kind}.\end{aligned} \tag{2.40}$$

C. The real algebra $\mathbf{A}_\mathbf{R}$

The real group algebra $\mathbf{A}_\mathbf{R}$ has a slightly more complicated structure than the corresponding complex algebra $\mathbf{A}_\mathbf{C}$.

$$F^\lambda_\mathbf{R}=\Big\{\sum_{r=0}^{\rho^\lambda-1}\mathbf{f}^\lambda_r a_r : a_r\in\mathbf{R}\Big\}\cong\tilde{\mathbf{F}}\supseteq\mathbf{F}=\mathbf{R}, \tag{2.41}$$

$$\rho^\lambda d^\lambda=n^\lambda=\rho^\lambda m^\lambda;\quad k=j,\quad K=q=J,\quad p\to jr,\quad P\to JR, \tag{2.42}$$

$$\mathbf{v}^{\lambda J}_{jr}=(n^\lambda)^{-1/2}\mathbf{e}^\lambda_{jJ}\mathbf{f}_r, \tag{2.43}$$

$$\begin{aligned}F^\lambda_{jr,JR}(\mathbf{b})&=R^\lambda_{jr,JR}(\mathbf{b})=R^{\tilde{\mathbf{F}}}_{rR}[\tilde{F}^\lambda_{jJ}(\mathbf{b})]\\&=(n^\lambda)^{-1}\sigma^{\tilde{\mathbf{F}}}_{rR,R}\langle\mathbf{e}^\lambda_{jJ}\mathbf{f}^\lambda_r,\mathbf{b}^\lambda\rangle\in\mathbf{R},\end{aligned} \tag{2.44}$$

$$\tilde{F}^\lambda_{jJ}(\mathbf{b})=\sum_{r=0}^{\tilde{\rho}-1}\tilde{f}_r R^\lambda_{jr,J0}(\mathbf{b})\in\tilde{\mathbf{F}}. \tag{2.45}$$

For $\mathbf{A}_\mathbf{R}$ the type of the representation is determined by the division algebra (=field) $\mathbf{F}^\lambda_\mathbf{R}$,

$$\begin{aligned}&\mathbf{R}\text{-type: }\mathbf{F}^\lambda_\mathbf{R}\cong\tilde{\mathbf{F}}=\mathbf{R}=\mathbf{F};\quad\rho^\lambda=\tilde{\rho}=1;\\&\qquad\text{indices }r=R=0\text{ may be dropped;}\\&\qquad\tilde{F}^\lambda_{jJ}(\mathbf{b})=F^\lambda_{jJ}(\mathbf{b})=R^\lambda_{jJ}(\mathbf{b})\in\mathbf{R}.\end{aligned} \tag{2.46}$$

$$\begin{aligned}&\mathbf{C}\text{-type: }\mathbf{F}^\lambda_\mathbf{R}\cong\tilde{\mathbf{F}}=\mathbf{C}\supset\mathbf{R}=\mathbf{F};\quad\rho^\lambda=\tilde{\rho}=2;\\&\qquad\text{indices }r,R=0,1;\\&\qquad\tilde{F}^\lambda_{jJ}(\mathbf{b})=C^\lambda_{jJ}(\mathbf{b})=R^\lambda_{j0,J0}(\mathbf{b})+iR^\lambda_{j1,J0}(\mathbf{b})\in\mathbf{C}.\end{aligned} \tag{2.47}$$

$$\begin{aligned}&\mathbf{Q}\text{-type: }\mathbf{F}^\lambda_\mathbf{R}\cong\tilde{\mathbf{F}}=\mathbf{Q}\supset\mathbf{R}=\mathbf{F};\quad\rho^\lambda=\tilde{\rho}=4;\\&\qquad\text{indices }r,R=0,1,2,3;\\&\qquad\tilde{F}^\lambda_{jJ}(\mathbf{b})=Q^\lambda_{jJ}(\mathbf{b})=R^\lambda_{j0,J0}(\mathbf{b})+iR^\lambda_{j1,J0}(\mathbf{b})\\&\qquad\qquad+jR^\lambda_{j2,J0}(\mathbf{b})+kR^\lambda_{j3,J0}(\mathbf{b})\in\mathbf{Q}.\end{aligned} \tag{2.48}$$

D. The quaternionic algebra $\mathbf{A}_\mathbf{Q}$

Due to the noncommutativity of $\mathbf{Q}$ the quaternionic group algebra $\mathbf{A}_\mathbf{Q}$ has the richest structure. I do not consider the most general decoposition of the ideals $\mathbf{A}^\lambda_\mathbf{Q}$ in this paper; instead, fixing relations between some numbers $f\in\mathbf{Q}$ and the ring elements $\mathbf{e}^\lambda_{kK}$, the following discussion is restricted to matrix representations where the matrices $\tilde{F}^\lambda(\mathbf{b})$ reduce to direct sums of matrices $\tilde{F}^\mathbf{Q}[q]$ if $\mathbf{b}=\mathbf{e}q$, $q\in\mathbf{Q}$ [or to matrices $\tilde{F}^\mathbf{Q}[q]^*=C^\mathbf{Q}[q]^*$ in

case $\tilde{\mathbf{F}}=\mathbf{C}$]. A special feature of these representations is that for each type of representation not only the field $\tilde{\mathbf{F}}$ is of importance but also a second subfield of $\mathbf{Q}$ denoted by $\hat{\mathbf{F}}$,

$$\mathbf{F}_{\mathbf{Q}}^{\lambda}\cong\tilde{\mathbf{F}}\subseteq\mathbf{Q}=\mathbf{F},\quad\{\sum_{r=0}^{\hat{\rho}-1}\hat{f}_r a_r : a_r\in\mathbf{R}\}=\hat{\mathbf{F}}\subseteq\mathbf{Q},\quad\tilde{\rho}\hat{\rho}=4; \tag{2.49}$$

$$d^{\lambda}=n^{\lambda}=\hat{\rho}m^{\lambda};\quad k=p\to jr,\quad K=P\to JR,\quad q=J; \tag{2.50}$$

$$\hat{f}_r\mathbf{e}^{\lambda}_{js,JR}=\sigma^{\hat{\mathbf{F}}}_{r,s}\mathbf{e}^{\lambda}_{j(rs),JR},$$

$$\tilde{f}\,\mathbf{e}^{\lambda}_{jr,JR}=\mathbf{e}^{\lambda}_{jr,JR}(\hat{f}^{*}_{rR}\tilde{f}\hat{f}_{rR})\ \text{ for all }\tilde{f}\in\tilde{\mathbf{F}}; \tag{2.51}$$

$$\mathbf{v}^{\lambda J}_{jr}=(n^{\lambda})^{-1/2}\mathbf{e}^{\lambda}_{jr,Jr'},\quad r'\text{ see below}; \tag{2.52}$$

$$F^{\lambda}_{jr,JR}(\mathbf{b})=\tilde{F}^{\lambda}_{jr,JR}(\mathbf{b})\in\tilde{\mathbf{F}},$$

$$\sum_{R}\tilde{F}^{\lambda}_{jr,JR}(\mathbf{b})\sigma^{\hat{\mathbf{F}}}_{r',r'R}\hat{f}^{*}_{r'R}=(m^{\lambda})^{-1}\langle\mathbf{e}^{\lambda}_{jr',Jr'},\mathbf{b}^{\lambda}\rangle\in\mathbf{Q} \tag{2.53}$$

Similar to the real algebra, the type of an ideal $\mathbf{A}^{\lambda}_{\mathbf{Q}}$ is uniquely determined by one of the fields $\tilde{\mathbf{F}}$ or $\hat{\mathbf{F}}$.

$\mathbf{R}$-type: $\mathbf{F}^{\lambda}_{\mathbf{Q}}=\{\mathbf{e}^{\lambda}q : q\in\mathbf{Q}\}\cong\tilde{\mathbf{F}}=\mathbf{Q},\quad\hat{\mathbf{F}}=\mathbf{R},\quad\hat{\rho}=1;$

indices $r=R=r'=0$ may be dropped;

$$F^{\lambda}_{jJ}(\mathbf{b})=\tilde{F}^{\lambda}_{jJ}(\mathbf{b})=\mathfrak{Q}^{\lambda}_{jJ}(\mathbf{b})=(m^{\lambda})^{-1}\langle\mathbf{e}^{\lambda}_{jJ},\mathbf{b}^{\lambda}\rangle\in\mathbf{Q};$$

$$\mathfrak{Q}^{\lambda}_{jJ}(\mathbf{e}q)=\delta_{jJ}Q^{\mathbf{Q}}[q]=\delta_{jJ}q. \tag{2.54}$$

$\mathbf{C}$-type: $\mathbf{F}^{\lambda}_{\mathbf{Q}}=\{\mathbf{e}^{\lambda}a+\mathbf{i}^{\lambda}b : a,b\in\mathbf{R}\}\cong\tilde{\mathbf{F}}=\mathbf{C}\cong$

$$\hat{\mathbf{F}}=\{a-jb : a,b\in\mathbf{R},\ j\in\mathbf{Q}\},\quad\hat{\rho}=2;$$

indices $r,R=0,1$; $r'=0$ or 1 is the number appearing in the decomposition

$$\mathbf{e}^{\lambda}i=\sum_{jr}\mathbf{e}^{\lambda}_{jr,jr}\mathbf{i}^{\lambda}(-1)^{r+r'};$$

$$\mathbf{e}^{\lambda}_{jr,JR}=\mathbf{e}^{\lambda}_{jr,J0}j^{R};\quad i\mathbf{e}^{\lambda}_{jr,JR}=\mathbf{e}^{\lambda}_{jr,JR}i(-1)^{r+R};$$

$$F^{\lambda}_{jr,JR}(\mathbf{b})=\tilde{F}^{\lambda}_{jr,JR}(\mathbf{b})=\mathfrak{C}^{\lambda}_{jr,JR}(\mathbf{b})\in\mathbf{C},$$

$$\mathfrak{C}^{\lambda}_{jr,J0}(\mathbf{b})j^{r-r'}+\mathfrak{C}^{\lambda}_{jr,J1}(\mathbf{b})j^{1-r'}=(m^{\lambda})^{-1}\langle\mathbf{e}^{\lambda}_{jr,Jr'},\mathbf{b}^{\lambda}\rangle\in\mathbf{Q};$$

$$\mathfrak{C}^{\lambda}_{jr,JR}(\mathbf{e}q)=\delta_{jJ}C^{\mathbf{Q}}_{rR}[q]\text{ or }\delta_{jJ}C^{\mathbf{Q}}_{rR}[q]^{*}. \tag{2.55}$$

$\mathbf{Q}$-type: $\mathbf{F}^{\lambda}_{\mathbf{Q}}=\{\mathbf{e}^{\lambda}a : a\in\mathbf{R}\}\cong\tilde{\mathbf{F}}=\mathbf{R},\quad\hat{\mathbf{F}}=\mathbf{Q},\quad\hat{\rho}=4;$

indices $r,R=0,1,2,3$; $r'=0$;

$$f_r\mathbf{e}^{\lambda}_{js,JR}=\sigma^{\mathbf{Q}}_{r,s}\mathbf{e}^{\lambda}_{j(rs),JR};$$

$$F^{\lambda}_{jr,JR}(\mathbf{b})=\tilde{F}^{\lambda}_{jr,JR}(\mathbf{b})=\mathfrak{R}^{\lambda}_{jr,JR}(\mathbf{b})\in\mathbf{R},$$

$$\mathfrak{R}^{\lambda}_{jr,J0}(\mathbf{b})-i\mathfrak{R}^{\lambda}_{jr,J1}(\mathbf{b})-j\mathfrak{R}^{\lambda}_{jr,J2}(\mathbf{b})-k\mathfrak{R}^{\lambda}_{jr,J3}(\mathbf{b})$$

$$=(m^{\lambda})^{-1}\langle\mathbf{e}^{\lambda}_{jr,J0},\mathbf{b}^{\lambda}\rangle\in\mathbf{Q};$$

$$\mathfrak{R}^{\lambda}_{jr,JR}(\mathbf{e}q)=\delta_{jJ}R^{\mathbf{Q}}_{rR}[q]. \tag{2.56}$$

3. $\mathbf{A}_{\mathbf{C}}$ AND $\mathbf{A}_{\mathbf{Q}}$ AS EXTENSIONS OF $\mathbf{A}_{\mathbf{R}}$

It is instructive to view the complex and quaternionic group algebras as extensions of the real one. An extension $\mathbf{A}_{\mathbf{F}}\to\mathbf{A}_{\mathbf{F}'}$, $\mathbf{F}\subset\mathbf{F}'$, implies two things: (i) The Hilbert space $\mathbf{A}_{\mathbf{F}}$ [or $L^2(G,\mathbf{F})$] is extended to the Hilbert space $\mathbf{A}_{\mathbf{F}'}$ [or $L^2(G,\mathbf{F}')$]; and (ii) the set of left multipliers (=operators) $\{\mathbf{e}f=\sum_{\lambda}\mathbf{e}^{\lambda}f : \mathbf{e}^{\lambda}\in\mathbf{A}^{\lambda}_{\mathbf{F}},\ f\in\mathbf{F}\}\cong\mathbf{F}$ is extended to the set $\{\mathbf{e}f'=\sum_{\lambda}\mathbf{e}^{\lambda}f' : \mathbf{e}^{\lambda}\in\mathbf{A}^{\lambda}_{\mathbf{F}'},\ f'\in\mathbf{F}'\}\cong\mathbf{F}'$.

If one starts with the operators $\mathbf{b}\in\mathbf{B}_{\mathbf{R}}$ and considers the chain $L^2(G,\mathbf{R})\to L^2(G,\mathbf{C})\to L^2(G,\mathbf{Q})$ one obtains the following scheme:

$$\begin{aligned}
&\mathbf{R}\text{-type:}\quad R^{\lambda}\quad\to C^{\lambda}\quad\to Q^{\lambda}\\
&\qquad m^{\lambda}.m^{\lambda}\ =m^{\lambda}.m^{\lambda}\ =m^{\lambda}.m^{\lambda}\\
&\qquad R^{\lambda}\text{a.i.}\quad C^{\lambda}\sim R^{\lambda}\quad Q^{\lambda}\sim R^{\lambda}
\end{aligned}\tag{3.1}$$

$$\begin{aligned}
&\mathbf{C}\text{-type:}\quad R^{\lambda}\quad\to C^{\lambda*}\oplus C^{\lambda-}\quad\to Q^{\lambda}\oplus Q^{\lambda}\\
&\qquad 2m^{\lambda}.m^{\lambda}=m^{\lambda}.m^{\lambda}+m^{\lambda}.m^{\lambda}\ =m^{\lambda}.m^{\lambda}+m^{\lambda}.m^{\lambda}\\
&\qquad\qquad C^{\lambda*}\text{a.i.}\quad Q^{\lambda}\sim C^{\lambda*}\sim C^{\lambda-}\\
&\qquad\qquad C^{\lambda*}\not\sim C^{\lambda-}=(C^{\lambda*})^{*}
\end{aligned}\tag{3.2}$$

$$\begin{aligned}
&\mathbf{Q}\text{-type:}\quad R^{\lambda}\quad\to C^{\lambda}\oplus C^{\lambda}\quad\to Q^{\lambda}\oplus Q^{\lambda}\oplus Q^{\lambda}\oplus Q^{\lambda}\\
&\qquad 4m^{\lambda}.m^{\lambda}=2m^{\lambda}.m^{\lambda}+2m^{\lambda}.m^{\lambda}\\
&\qquad\qquad =m^{\lambda}.m^{\lambda}+m^{\lambda}.m^{\lambda}+m^{\lambda}.m^{\lambda}+m^{\lambda}.m^{\lambda}\\
&\qquad\qquad\qquad Q^{\lambda}\text{a.i.}
\end{aligned}\tag{3.3}$$

In (3.1)–(3.3) the product listed under a representation F^{λ} is $n^{\lambda}.m^{\lambda}$, where $n^{\lambda}=\dim F^{\lambda}$ and m^{λ} is the number of subspaces of $L^2(G,\mathbf{F})$ transforming according to F^{λ}; m^{λ} is also the dimension of the matrix algebra isomorphic to the *real* two-sided ideal. The symbol a.i. stands for absolutely irreducible, meaning that F^{λ} remains irreducible even if $\mathbf{F}$-equivalence is extended to $\mathbf{F}'$-equivalence ($\mathbf{F}\subset\mathbf{F}'$). As is seen from the scheme, m^{λ} is also the dimension of the a.i. representation contained in the representation F^{λ}. The equivalence symbol $\sim$ under F^{λ} denotes $\mathbf{F}$-equivalence and the first [second] arrow may be replaced by " is $\mathbf{C}$-[$\mathbf{Q}$-] equivalent to ". The result that complex irreducible group representations of type $\mathbf{R}$(=1st kind) or $\mathbf{C}$(=3rd kind) remain irreducible whereas representations of type-$\mathbf{Q}$(=2nd kind) split into two equivalent quaternionic group representations has already been obtained by Finkelstein[4] *et al.* It is based on the fact that every decomposition appearing in (3.2) and (3.3) may be achieved by solving the eigenvalue problem of a self-adjoint element of the group algebra not existing before the extension. In the first step this element must be $\pm\mathbf{i}^{\lambda}i$ ($\mathbf{C}$-type); in the second it can be chosen to be $\mathbf{j}^{\lambda}j$ ($\mathbf{Q}$-type) [see Sec. 6A].

For the $\mathbf{R}$- and the $\mathbf{C}$-type the decomposition of the matrix representation of $\mathbf{B}_{\mathbf{R}}$ is obtained choosing the elements $\mathbf{v}^{\lambda J}_{jr}$ given by (2.53), (2.54), and (2.55) as basis of the subspaces $\mathbf{A}^{\lambda}_{\mathbf{Q}}$ [or the corresponding functions $v^{\lambda J}_{jr}\in L^2(G,\mathbf{Q})$ as basis of $L^{\lambda}(G,\mathbf{Q})=\mathbf{e}^{\lambda}L^2(G,\mathbf{Q})$, $\mathbf{e}^{\lambda}\in\mathbf{A}_{\mathbf{R}}$]. To obtain also the $\mathbf{Q}$-type representations in the completely reduced form (3.3) one has to pass from the basis $\{\mathbf{v}^{\lambda J}_{jr}\}$ defined by Eqs. (2.53) and (2.56) to a basis $\{\mathbf{w}^{\lambda J}_{jr}\}$ according to

$$\mathbf{w}^{\lambda J}_{jR}=\sum_{r}\mathbf{v}^{\lambda J}_{jr}Q_{rR},\quad Q=(1/2)\begin{pmatrix}1&-i&-j&-k\\ i&1&k&-j\\ j&-k&1&i\\ k&j&-i&1\end{pmatrix}. \tag{3.4}$$

This shows that it is possible to find a basis $\{\mathbf{w}^{\lambda q}_{j}\}$ of the quaternionic Hilbert space $L^2(G,\mathbf{Q})$ such that the eigenvalue equations

$$\mathbf{e}^{\lambda}_{Jj}w^{\lambda q}_{j}=w^{\lambda q}_{J}\ \text{ for }\mathbf{e}^{\lambda}_{Jj}\in\mathbf{A}^{\lambda}_{\mathbf{R}},\quad\mathbf{f}^{\lambda}w^{\lambda q}_{j}=w^{\lambda q}_{j}\tilde{f}\ \text{ for }\mathbf{f}^{\lambda}\in\mathbf{F}^{\lambda}_{\mathbf{R}}, \tag{3.5}$$

are satisfied for all types of representations [cf. I-(4.38)]. The type of the representation shows up in the field $\{\tilde{f}\}=\tilde{\mathbf{F}}\subseteq\mathbf{Q}$ generated by the eigenvalues of the operators $\mathbf{f}^{\lambda}$.

The next step is to keep the Hilbert space $L^2(G,\mathbf{F})$ or the subspace $L^\lambda(G,\mathbf{F})=\mathbf{e}^\lambda L^2(G,\mathbf{F})$, $\mathbf{e}^\lambda\in\mathbf{A}_\mathbf{R}$, fixed and to extend the set of left-multipliers $\{\sum_\lambda \mathbf{e}^\lambda a : a\in\mathbf{R}\}\cong\mathbf{R}$ to the set $\{\sum_\lambda \mathbf{e}^\lambda f : f\in\mathbf{F}\}\cong\mathbf{F}$. For $\mathbf{F}=\mathbf{C}$ this does not change the structure of the decomposition into irreducible subspaces since in this case the Hilbert space is complex and

$$cb=bc \quad \text{for all } c\in\mathbf{C},\ \ b\in\mathbf{B}_\mathbf{R}. \tag{3.6}$$

Passing from $\mathbf{C}$ to $\mathbf{Q}$ by introducing the left-multiplier j, the situation becomes different because for a given $v\in L^2(G,\mathbf{Q})$ the elements v and jv may be linearly independent. This happens for the $\mathbf{C}$- and the $\mathbf{Q}$-type where a pair of complex irreducible representations of $\mathbf{B}_\mathbf{C}$ is absorbed into a single irreducible representation of $\mathbf{B}_\mathbf{Q}$. Therefore the full extension including both the extension of the Hilbert space, $L(G,\mathbf{R})\to L(G,\mathbf{C})\to L(G,\mathbf{Q})$, and that of the operator set, $\{\mathbf{B}_\mathbf{R}\}\to\{\mathbf{B}_\mathbf{R},i\}\to\{\mathbf{B}_\mathbf{R},i,j\}$, is given by the following scheme:

$$\begin{aligned}
&\mathbf{R}\text{-type:}\quad R^\lambda \quad\to C^\lambda \quad\to \mathfrak{Q}^\lambda\sim\mathfrak{Q}^\lambda\\
&\qquad m^\lambda.m^\lambda = m^\lambda.m^\lambda = m^\lambda.m^\lambda
\end{aligned}\tag{3.7}$$

$$\begin{aligned}
&\mathbf{C}\text{-type:}\quad R^\lambda \quad\to C^{\lambda*}\oplus C^{\lambda-} \quad\to \mathfrak{Q}^\lambda\sim\mathfrak{C}^\lambda\\
&\qquad 2m^\lambda.m^\lambda = m^\lambda.m^\lambda + m^\lambda.m^\lambda = 2m^\lambda.m^\lambda
\end{aligned}\tag{3.8}$$

$$\begin{aligned}
&\mathbf{Q}\text{-type:}\quad R^\lambda \quad\to C^\lambda\oplus C^\lambda \quad\to \mathfrak{Q}^\lambda\sim\mathfrak{R}^\lambda\\
&\qquad 4m^\lambda.m^\lambda = 2m^\lambda.m^\lambda + 2m^\lambda.m^\lambda = 4m^\lambda.m^\lambda.
\end{aligned}\tag{3.9}$$

The factors refer to the decomposition of the spaces $L^\lambda(G,\mathbf{F})$. The symbol $\sim$ denotes $\mathbf{Q}$-equivalence and the quaternionic, complex, and real representations listed at the end of the chains are just the representations given in Sec. 2.

4. UNIQUENESS

One may ask to what extent the properties (2.16) and (2.17), supplemented by (2.35)–(2.38) for $\mathbf{F}=\mathbf{C}$, by (2.41)–(2.45) for $\mathbf{F}=\mathbb{R}$, and by (2.49)–(2.53) for $\mathbf{F}=\mathbf{Q}$, fix the ring bases $\{\mathbf{e}^\lambda_{kK},\mathbf{f}^\lambda_r\}$ and the Hilbert space bases $\{\mathbf{v}^{\lambda q}_\rho\}$. To see this, suppose two ring bases $\{{}^{(1)}\mathbf{e}^\lambda_{kK},{}^{(1)}\mathbf{f}^\lambda_r\}$, $\{{}^{(2)}\mathbf{e}^\lambda_{kK},{}^{(2)}\mathbf{f}^\lambda_r\}$ satisfying the same set of equations to be given. It is then possible to find a pair k',k'' such that ${}^{(2)}\mathbf{e}^\lambda_{k''k''}\,{}^{(1)}\mathbf{e}^\lambda_{k'k'}\neq 0$ and to use these indices to define a norm-preserving element $\mathbf{u}^\lambda\in\mathbf{A}^\lambda_\mathbf{F}$,

$$\mathbf{u}^\lambda = b_{k'k''}\sum_k {}^{(2)}\mathbf{e}^\lambda_{kk''}\,{}^{(1)}\mathbf{e}^\lambda_{k'k},\quad b_{k'k''}\in\mathbf{R},\quad \mathbf{u}^\lambda\mathbf{u}^{\lambda*}=\mathbf{u}^{\lambda*}\mathbf{u}^\lambda=\mathbf{e}^\lambda. \tag{4.1}$$

For this $\mathbf{u}^\lambda$

$$\mathbf{u}^{\lambda*}\,{}^{(2)}\mathbf{e}^\lambda_{kK}\mathbf{u}^\lambda = {}^{(1)}\mathbf{e}^\lambda_{kK},\quad \mathbf{u}^{\lambda*}\,{}^{(2)}\mathbf{F}^\lambda_\mathbf{F}\mathbf{u}^\lambda = {}^{(1)}\mathbf{F}^\lambda_\mathbf{F},\quad \mathbf{u}^{\lambda*}\mathbf{e}^\lambda\mathbf{F}\mathbf{u}^\lambda=\mathbf{e}^\lambda\mathbf{F}. \tag{4.2}$$

In general the elements $\mathbf{u}^{\lambda*}\,{}^{(2)}\mathbf{f}^\lambda_r\mathbf{u}^\lambda$, $\mathbf{u}^{\lambda*}f_s\mathbf{u}^\lambda$ will differ from the elements ${}^{(1)}\mathbf{f}^\lambda_r$, f_s but are equally suited as basis elements for the fields ${}^{(1)}\mathbf{F}^\lambda_\mathbf{F}$, $\mathbf{F}$ since they satisfy the same relations. To be more precise,

$$\mathbf{u}^{\lambda*}\sum_r {}^{(2)}\mathbf{f}^\lambda_r a_r\ \mathbf{u}^\lambda = \alpha^{-1}\sum_r {}^{(1)}\mathbf{f}^\lambda_r a_r = \sum_r(\alpha^{-1}\,{}^{(1)}\mathbf{f}^\lambda_r)a_r, \tag{4.3}$$

where $\alpha^{-1}:{}^{(1)}\mathbf{F}^\lambda_\mathbf{F}\to{}^{(1)}\mathbf{F}^\lambda_\mathbf{F}$ is an automorphism of this field, and

$$\mathbf{F}\neq\mathbf{Q}:\mathbf{u}^{\lambda*}f\mathbf{u}^\lambda=f \text{ for all } f\in\mathbf{F}; \tag{4.4}$$

$$\mathbf{F}=\mathbf{Q}:\mathbf{u}^{\lambda*}\hat{f}\mathbf{u}^\lambda=\hat{f} \text{ for all } \hat{f}\in\hat{\mathbf{F}},\qquad \mathbf{u}^{\lambda*}\tilde{f}\mathbf{u}^\lambda=\tilde{\omega}\tilde{f} \text{ for all } \tilde{f}\in\tilde{\mathbf{F}}, \tag{4.5}$$

where $\tilde{\omega}$ is an automorphism of $\tilde{\mathbf{F}}$. Therefore, if a third ring basis $\{{}^{(3)}\mathbf{e}^\lambda_{kK},{}^{(3)}\mathbf{f}^\lambda_r\}$ is defined by

$${}^{(3)}\mathbf{e}^\lambda_{kK}=\mathbf{u}^{\lambda*}\,{}^{(2)}\mathbf{e}^\lambda_{kK}\mathbf{u}^\lambda,\quad {}^{(3)}\mathbf{f}^\lambda_r=\mathbf{u}^{\lambda*}\,{}^{(2)}\mathbf{f}^\lambda_r\mathbf{u}^\lambda, \tag{4.6}$$

the uniqueness of the matrix representations can be discussed in two steps, namely first (2)→(3) according to (4.4)–(4.6) and then (3)→(1) according to

$${}^{(1)}\mathbf{e}^\lambda_{kK}={}^{(3)}\mathbf{e}^\lambda_{kK},\quad {}^{(1)}\mathbf{f}^\lambda_r=\alpha\,{}^{(3)}\mathbf{f}^\lambda_r. \tag{4.7}$$

In the first step relations (4.6) immediately yield

$${}^{(2)}\tilde{F}^\lambda(\mathbf{b})={}^{(3)}\tilde{F}^\lambda(\mathbf{u}^{\lambda*}\mathbf{b}\mathbf{u}^\lambda)=\tilde{U}^{\lambda*}\,{}^{(3)}\tilde{F}^\lambda(\mathbf{b})\tilde{U}^\lambda,\quad \tilde{U}^\lambda={}^{(3)}\tilde{F}^\lambda(\mathbf{u}^\lambda), \tag{4.8}$$

showing that the matrix representation $\tilde{F}^\lambda$ is determined only up to $\tilde{\mathbf{F}}$-equivalence. The corresponding result for the matrix representation F^λ,

$$\begin{aligned}
&{}^{(2)}F^\lambda(\mathbf{b})={}^{(3)}F^\lambda(\mathbf{u}^{\lambda*}\mathbf{b}\mathbf{u}^\lambda)=U^{\lambda*}\,{}^{(3)}F^\lambda(\mathbf{b})U^\lambda,\\
&\mathbf{F}\supseteq\tilde{\mathbf{F}}:F^\lambda(\mathbf{b})=\tilde{F}^\lambda(\mathbf{b}),\quad U^\lambda=\tilde{U}^\lambda,\\
&\mathbf{F}\subseteq\tilde{\mathbf{F}}:F^\lambda(\mathbf{b})=F^{\tilde{\mathbf{F}}}[\tilde{F}^\lambda(\mathbf{b})],\quad U^\lambda=F^{\tilde{\mathbf{F}}}[\tilde{U}^\lambda],
\end{aligned}\tag{4.9}$$

is obtained if the Hilbert space bases $\{{}^{(2)}\mathbf{v}^{\lambda q}_\rho\}$, $\{{}^{(3)}\mathbf{v}^{\lambda q}_\rho\}$ are related to the ring bases $\{{}^{(2)}\mathbf{e}^\lambda_{kK},{}^{(2)}\mathbf{f}^\lambda_r\}$, $\{{}^{(3)}\mathbf{e}^\lambda_{kK},{}^{(3)}\mathbf{f}^\lambda_r\}$ as it is required in (2.37), (2.43), or (2.52). It should be noted that for $\mathbf{F}=\mathbf{Q}$ the automorphism $\tilde{\omega}$ appearing in (4.5) is the identical one for $\mathbf{R}$- and $\mathbf{Q}$-type representations; for the $\mathbf{C}$-type representations where $\tilde{\omega}c=c$ or $\tilde{\omega}c=c^*$ this automorphism is always compensated by the way the bases $\{{}^{(2)}\mathbf{v}^{\lambda R}_{jr}\}$, $\{{}^{(3)}\mathbf{v}^{\lambda R}_{jr}\}$ are chosen [cf. Eq. (2.55)].

To see what happens to the representations in the second step it is necessary to introduce an automorphism $\tilde{\alpha}:\tilde{\mathbf{F}}\to\tilde{\mathbf{F}}$ corresponding to the automorphism $\alpha:\mathbf{F}^\lambda_\mathbf{F}\to\mathbf{F}^\lambda_\mathbf{F}$,

$$\begin{aligned}
&\alpha\,{}^{(3)}\mathbf{f}_r=\sum_s {}^{(3)}\mathbf{f}_s O_{sr}={}^{(1)}\mathbf{f}_r,\\
&\alpha\sum_r {}^{(3)}\mathbf{f}_r a_r=\sum_r(\alpha\,{}^{(3)}\mathbf{f}_r)a_r=\sum_{rs}{}^{(3)}\mathbf{f}_r O_{rs}a_s,\\
&\tilde{\alpha}\sum_r f_r a_r=\sum_r(\tilde{\alpha}f_r)a_r=\sum_{rs}f_r O_{rs}a_s.
\end{aligned}\tag{4.10}$$

The special form of the orthogonal matrix O appearing in (4.10) depends on the field $\tilde{\mathbf{F}}$. The relation between the matrix representations ${}^{(3)}\tilde{F}^\lambda$, ${}^{(3)}F^\lambda$ and ${}^{(1)}\tilde{F}^\lambda$, ${}^{(1)}F^\lambda$ is then

$$\mathbf{F}\supseteq\tilde{\mathbf{F}}:\ {}^{(3)}F^\lambda(\mathbf{b})={}^{(3)}\tilde{F}^\lambda(\mathbf{b})=\tilde{\alpha}\,{}^{(1)}\tilde{F}^\lambda(\mathbf{b})=\tilde{\alpha}\,{}^{(1)}F^\lambda(\mathbf{b}), \tag{4.11}$$

$$\begin{aligned}
\mathbf{F}=\mathbf{R}:{}^{(3)}R^\lambda_{jr,jR}(\mathbf{b})&=\sum_{sS}O_{rs}\,{}^{(1)}R^\lambda_{js,jS}(\mathbf{b})O^*_{SR}\\
&=R^{\tilde{\mathbf{F}}}_{rR}[{}^{(3)}\tilde{F}^\lambda_{jJ}(\mathbf{b})]=R^{\tilde{\mathbf{F}}}_{rR}[\tilde{\alpha}\,{}^{(1)}\tilde{F}^\lambda_{jJ}(\mathbf{b})].
\end{aligned}\tag{4.12}$$

For $\mathbf{F}\neq\mathbf{C}$ all automorphisms of $\tilde{\mathbf{F}}$ can appear in this kind of equivalence of matrix representations. For $\mathbf{F}=\mathbf{C}$, however, only the identical one is admitted and two matrix representations belonging to different ring bases are always linked by a unitary transformation ($\mathbf{C}$-equivalence).

5. MATRIX ELEMENTS AS FUNCTIONS ON THE GROUP

A. Complex representations

In the complex group algebra $\mathbf{A_C}$ the elements $\mathbf{e}^\lambda_{jJ}$ are related to the irreducible matrix representations $C^\lambda(x) = C^\lambda(x)$ by

$$\mathbf{e}^\lambda_{jJ} = n^\lambda M_x C^{\lambda *}_{jJ}(x)\mathbf{x}, \quad n^\lambda = \dim C^\lambda(x). \tag{5.1}$$

Accordingly the properties of the ring bases $\{\mathbf{e}^\lambda_{jJ}\}$ [Eqs. (2.12), (2.16), (2.17), (2.37), and (2.38)] are reflected in an equivalent set of equations for the functions $C^\lambda_{jJ}: G \to \mathbb{C}$. This set is

$$C^\lambda_{jJ}(xy) = \sum_{j'} C^\lambda_{jj'}(x) C^\lambda_{j'J}(y), \tag{5.2}$$

$$C^\lambda_{jJ}(x^{-1}) = C^{\lambda *}_{Jj}(x), \tag{5.3}$$

$$C^\lambda_{jJ} * C^{\lambda'}_{j'J'} = \delta_{\lambda\lambda'}\delta_{Jj'} C^\lambda_{jJ'} (n^\lambda)^{-1}, \tag{5.4}$$

$$a \in L^2(G,\mathbb{C}) \Rightarrow a = \sum_{\lambda j J} C^\lambda_{jJ} \langle C^\lambda_{jJ} n^\lambda, a\rangle. \tag{5.5}$$

Equations (5.2) and (5.3) show that $\{C^\lambda(x) : x \in G\}$ is a unitary matrix representation of G. The convolution formula (5.4) implies the orthogonality of the C's which, combined with their completeness, gives the expansion formula (5.5). Equivalent sets $\{\mathbf{e}^\lambda_{jJ}\}$ and $\{\bar{\mathbf{e}}^\lambda_{jJ}\}$ correspond to equivalent matrix representations C^λ and $\bar{C}^\lambda$ related by

$$\bar{C}^\lambda(x) = U^{\lambda *} C^\lambda(x) U^\lambda, \quad U^\lambda \text{ unitary}. \tag{5.6}$$

B. Real representations

For the real group algebra $\mathbf{A_R}$ one finds

$$\mathbf{e}^\lambda_{jJ}\mathbf{f}^\lambda_r = n^\lambda M_x R^\lambda_{jr,J0}(x)\mathbf{x}, \quad n^\lambda = r^\lambda d^\lambda = \dim R^\lambda(x). \tag{5.7}$$

The functions corresponding to the elements $\mathbf{e}^\lambda_{jJ}$ and $\mathbf{f}^\lambda_r$ are obtained from (5.7) according to

$$\mathbf{e}^\lambda_{jJ} = \mathbf{e}^\lambda_{jJ}\mathbf{e}^\lambda = \mathbf{e}^\lambda_{jJ}\mathbf{f}^\lambda_0, \quad \mathbf{f}^\lambda_r = \mathbf{e}^\lambda \mathbf{f}^\lambda_r = \sum_j \mathbf{e}^\lambda_{jj}\mathbf{f}^\lambda_r. \tag{5.8}$$

Hence the properties of the ring basis $\{\mathbf{e}^\lambda_{jJ}, \mathbf{f}^\lambda_r\}$ are determined by the real irreducible matrix representation R^λ and vice versa. The elements of R satisfy

$$R^\lambda_{jr,JR}(x) = \sigma^{\mathbb{F}}_{rR,r} R^\lambda_{j(rR),J0}(x), \tag{5.9}$$

$$R^\lambda_{jr,JR}(xy) = \sum_{j'r'} R^\lambda_{jr,j'r'}(x) R^\lambda_{j'r',JR}(y), \tag{5.10}$$

$$R^\lambda_{jr,JR}(x^{-1}) = R^\lambda_{JR,jr}(x), \tag{5.11}$$

$$R^\lambda_{jr,J0} * R^\lambda_{j'r',J'0} = \delta_{\lambda\lambda'}\delta_{Jj'}\sigma^{\mathbb{F}}_{r,r'} R^\lambda_{j(rr'),J'0}, \tag{5.12}$$

$$a \in L^2(G,\mathbb{R}) \Rightarrow a = \sum_{\lambda j J r} R^\lambda_{jr,J0} \langle R^\lambda_{jr,J0} n^\lambda, a\rangle. \tag{5.13}$$

Equation (5.9) shows the real irreducible matrix representations of the group to have a structure which is typical for the type of the representation [cf. Eq. (2.44)]. Equations (5.10) and (5.11) state that $\{R^\lambda(x) : x \in G\}$ is an orthogonal matrix representation of G. As in the complex case, the convolution formula (5.12) implies the orthogonality of the functions $R^\lambda_{jr,J0}$, whereas the essential content of (5.13) is the completeness of these functions in $L^2(G,\mathbb{R})$. Equivalent matrix representations corresponding to equivalent ring bases are related by

$$\bar{R}^\lambda(x) = S^{\lambda *} U^{\lambda *} R^\lambda(x) U^\lambda S^\lambda; \quad U^\lambda \text{ orthogonal},$$
$$U^\lambda_{jr,JR} = \sigma^{\mathbb{F}}_{rR,r} U^\lambda_{j(rR),J0};$$
$$\mathbb{R}\text{- and } \mathbb{Q}\text{-type: } S^\lambda = E^\lambda (1\text{-matrix}), \tag{5.14}$$
$$\mathbb{C}\text{-type: } S^\lambda = E^\lambda \text{ or } S^\lambda_{jr,JR} = \delta_{jJ}\delta_{rR}(-1)^r.$$

C. Quaternionic representations

Due to the conventions made in Eqs. (2.51) for the elements $\mathbf{e}_{jr,JR}$ of the quaternionic group algebra $\mathbf{A_Q}$ the formula corresponding to (5.1) and (5.7) in the complex and in the real case, respectively, now reads

$$\mathbf{e}^\lambda_{jr,JR} = m^\lambda M_x \hat{f}_r f^\lambda_{jJ}(x) \hat{f}_R \mathbf{x}, \quad m^\lambda = \dim f^\lambda(x). \tag{5.15}$$

The numbers $\hat{f}_r$ are specified in (2.54)–(2.56) and the functions

$$f^\lambda_{jJ}(x) = \sum_{r=0}^{\rho-1} f_r R_{jr,J0}(x) \in \mathbb{F} \quad (\cong \mathbf{F}^\lambda_{\mathbf{R}}) \tag{5.16}$$

coincide with the matrix elements $\tilde{F}^\lambda_{jJ}(x)$ defined for $\mathbf{A_R}$ [cf. Eq. (2.45)]. They satisfy

$$f^\lambda_{jJ}(xy) = \sum_{j'} f^\lambda_{jj'}(x) f^\lambda_{j'J}(y), \tag{5.17}$$

$$f^\lambda_{jJ}(x^{-1}) = f^{\lambda *}_{Jj}(x), \tag{5.18}$$

$$f^\lambda_{jJ} * f^{\lambda'}_{j'J'} = \delta_{\lambda\lambda'}\delta_{Jj'} f^\lambda_{jJ'} (m^\lambda)^{-1}, \tag{5.19}$$

$$a \in L^2(G,\mathbb{Q}) \Rightarrow a = \sum_{\lambda j J r} \hat{f}_r f^\lambda_{jJ} \langle \hat{f}_r f^\lambda_{jJ} m^\lambda, a\rangle. \tag{5.20}$$

Equivalent matrix representations corresponding to equivalent ring bases are related by

$$\bar{f}^\lambda(x) = U^{\lambda *}[\omega f^\lambda(x)] U^\lambda, \quad U^\lambda \text{ norm-preserving}, \quad U^\lambda_{jJ} \in \mathbb{F};$$
$$\mathbb{R}\text{- and } \mathbb{Q}\text{-type: } \omega f = f, \quad \mathbb{C}\text{-type: } \omega c = c \text{ or } c^*. \tag{5.21}$$

6. PROOFS

A. Construction of bases

In this subsection proofs of the assertions made in Secs. 2 and 3 on the noncomplex representations are given or at least outlined. Complex representation theory is supposed to be known[7,8] and proofs being quite similar for all three fields are mostly omitted. Moreover, quaternionic representation theory is considered as an extension of the real one as was done in Sec. 3.

The first thing to be noted for the real group algebra is that $\mathbf{A_R}$ is semisimple. This is a straight consequence of the fact that $\mathbf{A_R}$ is a Hilbert ring.[10] The essential feature of the minimal two-sided ideals $\mathbf{A}^\lambda_{\mathbf{R}}$ is that they are simple (i.e., do not contain proper two-sided ideals) and finite-dimensional. The latter follows from the fact that the function $n^\lambda \chi^\lambda$ corresponding to the elements $\mathbf{e}^\lambda$ generating the ideals $\mathbf{A}^\lambda_{\mathbf{R}}$ are continuous.[11] With the simplicity and the finiteness of $\mathbf{A}^\lambda_{\mathbf{R}}$ guaranteed, Wedderburn's famous structure theorem[12] applies: $\mathbf{A}^\lambda_{\mathbf{R}}$ is isomorphic to a full matrix algebra over a division algebra (= field or skewfield) $\tilde{\mathbb{F}}$. The field $\tilde{\mathbb{F}}$ must be an extension field of $\mathbb{R}$ since $\mathbf{A}^\lambda_{\mathbf{R}}$ contains the

subset $\mathbf{e}^\lambda\mathbb{R}$ isomorphic to $\mathbf{R}$. But there are only three extensions of $\mathbf{R}$, namely $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{Q}$:[13] These three fields allow a natural partitioning of the two-sided ideals into three classes ("types"). For a given group all three types may occur, as the example of the double-point group D_3^* shows. Since $\mathbf{A}_\mathbb{R}^\lambda$ is isomorphic to a *full* matrix algebra over $\tilde{\mathbf{F}}$, the representation $\tilde{F}^\lambda$ associated with this isomorphism [Eqs. (2.18) and (2.19)] is irreducible.

It remains to discuss the Hilbert space properties of $\mathbf{A}_\mathbb{R}^\lambda$. That the subspaces $\mathbf{A}_\mathbb{R}^\lambda\mathbf{e}_{JJ}^\lambda$ are invariant under the left-action of $\mathbf{A}_\mathbb{R}$ follows from $\mathbf{A}_\mathbb{R}\mathbf{A}_\mathbb{R}^\lambda = \mathbf{A}_\mathbb{R}^\lambda$ [see Eq. (2.12)]. To prove that the elements $\mathbf{e}_{jJ}^\lambda\mathbf{f}_r^\lambda \in \mathbf{A}_\mathbb{R}^\lambda\mathbf{e}_{JJ}^\lambda$ are orthogonal and have the same norm one first puts $r=0$ and repeats the proof of the complex theory [Hint: use (2.16) and (2.17), and $\langle \mathbf{a}, \mathbf{b}\rangle = \langle \mathbf{a}^*, \mathbf{b}^*\rangle$]; then the factors $\mathbf{f}_r^\lambda$ satisfying

$$\mathbf{f}_r^{\lambda*} = \mathbf{f}_r^\lambda \sigma_{\tilde{\mathbf{F}},r}^{\tilde{\mathbf{F}}}, \quad \mathbf{f}_r^\lambda \mathbf{f}_R^\lambda = \mathbf{f}_{rR}^\lambda \sigma_{\tilde{\mathbf{F}},R}^{\tilde{\mathbf{F}}} \tag{6.1}$$

are introduced as operators acting on the elements $\mathbf{e}_{jJ}^\lambda$. The common normalization constant $n^\lambda = \dim R^\lambda$ is obtained from Eq. (5.7) relating $\mathbf{e}_{jJ}^\lambda\mathbf{f}_r^\lambda$ to the function $R_{jr,J0}^\lambda$. The proof of (5.7) is given in subsection 6C below. The peculiar structure of the matrix representations R^λ carried by the basis $\mathbf{e}_{jJ}^\lambda\mathbf{f}_r^\lambda$ is a direct consequence of (2.17) and (6.1). The representations R^λ are irreducible over $\mathbb{R}$: For a decomposition into a direct sum of smaller matrices would imply a decomposition of $\mathbf{F}_\mathbb{R}^\lambda$ into a direct sum of division algebras isomorphic to $\mathbf{F}_\mathbb{R}^\lambda$; but this is only possible if $\mathbb{R}$ is extended. The calculation of the matrix elements $R_{jr,JR}^\lambda(\mathbf{b})$ by means of the scalar product (2.44) follows from

$$\begin{aligned} n^\lambda R_{jr,JR}^\lambda(\mathbf{b}) &= \langle \mathbf{e}_{jJ}^\lambda, \mathbf{f}_r^\lambda, \mathbf{b}^\lambda \mathbf{e}_{JJ}^\lambda, \mathbf{f}_R^\lambda\rangle = \langle \mathbf{e}_{jJ}^{\lambda*}, \mathbf{f}_r^{\lambda*}, \mathbf{f}_R^{\lambda*}\mathbf{e}_{JJ}^{\lambda*}, \mathbf{b}^{\lambda*}\rangle \\ &= \langle \mathbf{e}_{JJ}^\lambda, \mathbf{e}_{J,j}^\lambda \mathbf{f}_R^{\lambda*}\mathbf{f}_r^{\lambda*}, \mathbf{b}^{\lambda*}\rangle = \langle \mathbf{e}_{jJ}^\lambda \mathbf{f}_{rR}^\lambda \mathbf{f}_R^\lambda \sigma_{\tilde{\mathbf{F}}_R,R}^{\tilde{\mathbf{F}}_R} \mathbf{f}_R^{\lambda*}, \mathbf{b}^\lambda\rangle . \end{aligned} \tag{6.2}$$

That the representation $\tilde{F}^\lambda$ can be obtained from the representation R^λ according to (2.45) follows from

$$\mathbf{b}^\lambda = \sum_{jJ} \mathbf{e}_{jJ}^\lambda \mathbf{f}_{jJ}^\lambda(\mathbf{b}) = \sum_{jJr} \mathbf{e}_{jJ}^\lambda \mathbf{f}_r^\lambda (n^\lambda)^{-1} \langle \mathbf{e}_{jJ}^\lambda \mathbf{f}_r^\lambda, \mathbf{b}^\lambda\rangle \tag{6.3}$$

[cf. Eqs. (2.18) and (6.2)] and the isomorphism $\mathbf{F}_\mathbb{R}^\lambda \simeq \tilde{\mathbb{F}}$.

Having clarified the structure of $\mathbf{A}_\mathbb{R}$, the complex algebra $\mathbf{A}_\mathbb{C}$ can be obtained by extending both the left-multipliers (=operators) and the right-multipliers (= elements of the base field of the vector space) from $\mathbb{R}$ to $\mathbb{C}$. Since $\mathbb{C}$ is commutative,

$$\mathbf{A}_\mathbb{C} = \mathbb{C}\mathbf{A}_\mathbb{R}\mathbb{C} = \mathbf{A}_\mathbb{R}\mathbb{C} . \tag{6.4}$$

The adjunction of $f_1 = i$ can generate a new self-adjoint element. Its spectral decomposition, if not trivial, yields a new decomposition of the element $\mathbf{e}^\lambda$ ($\epsilon\mathbf{A}_\mathbb{R}^\lambda$) into primitive idempotents. For the $\mathbb{C}$-type the new self-adjoint element is proportional to $\mathbf{f}_1^\lambda i$. The idempotents $\mathbf{g}_{rr}^\lambda$ obtained from its spectral decomposition satisfy

$$\begin{aligned} &\mathbf{g}_{rr}^\lambda = (1/2)[\mathbf{e}^\lambda + \mathbf{f}_1^\lambda(-1)^r i] = \mathbf{g}_{rr}^{\lambda*}, \quad \mathbf{g}_{rr}^\lambda \mathbf{g}_{RR}^\lambda = \delta_{rR}\mathbf{g}_{rr}^\lambda , \\ &\sum_r \mathbf{g}_{rr}^\lambda = \mathbf{e}^\lambda, \quad \mathbf{f}_1^\lambda \mathbf{g}_{rr} = \mathbf{g}_{rr}^\lambda (-1)^{r+1} i . \end{aligned} \tag{6.5}$$

The $\mathbf{g}$'s commute with all elements of $\mathbf{A}_\mathbb{R}^\lambda$, as does f_1^λ, and split this two-sided ideal of $\mathbf{A}_\mathbb{R}$ into a pair of minimal two-sided ideals of $\mathbf{A}_\mathbb{C}$, say $\mathbf{A}_\mathbb{C}^{\lambda\pm}$. If $\mathbf{A}_\mathbb{R}^\lambda$ is of $\mathbf{Q}$-type there is an infinity of new self-adjoint elements, namely the elements $\mathbf{u}^\lambda i$ for which $\mathbf{u}^{\lambda*} = -\mathbf{u}^\lambda = (\mathbf{u}^\lambda)^{-1}$. Since they do not commute, only one of them, say $\mathbf{f}_1^\lambda i$, can be used to decompose each $\mathbf{e}_{jj}^\lambda \in \mathbf{A}_\mathbb{R}^\lambda$ into two idempotents $\mathbf{e}_{jr,jr}^\lambda$ which are primitive in $\mathbf{A}_\mathbb{C}^\lambda$. The multiplication law (6.1) then implies

$$\mathbf{f}_2^\lambda \mathbf{g}_{00}^\lambda = \mathbf{g}_{11}^\lambda \mathbf{f}_2^\lambda ; \tag{6.6}$$

therefore the two-sided ideal $\mathbf{A}_\mathbb{R}^\lambda$ does not give rise to two new ideals but only changes its internal structure: The dimension of the matrix algebra isomorphic to $\mathbf{A}_\mathbb{C}^\lambda = \mathbf{A}_\mathbb{R}^\lambda\mathbf{C}$ is now twice the dimension of the matrix algebra isomorphic to $\mathbf{A}_\mathbb{R}^\lambda$ and the field of matrix elements is now $\mathbb{C}$ instead of $\mathbb{Q}$, since the subalgebra of $\mathbf{F}_\mathbb{R}^\lambda$ isomorphic to $\mathbb{C}$ has been "diagonalized." The adjunction of i does not give new self-adjoint elements if $\mathbf{A}_\mathbb{R}^\lambda$ is of $\mathbb{R}$-type. The only change in the structure of this ideal is that $\mathbf{F}_\mathbb{R}^\lambda \simeq \mathbb{R}$ becomes $\mathbf{F}_\mathbb{R}^\lambda\mathbb{C} = \mathbf{F}_\mathbb{C}^\lambda \simeq \mathbb{C}$.

The complex representations of $\mathbf{B}_\mathbb{R} = \mathbf{A}_\mathbb{R} \cup \mathbf{G}$ carried by the elements of $\mathbf{A}_\mathbb{C}^\lambda$ are easily computed from the real representations R^λ and Eqs. (6.5) and (6.6). The representations of the $\mathbb{R}$-type remain real if the ring basis is not changed. For the representations of $\mathbb{C}$-type the splitting of the two-sided ideal entails a splitting of the real representation R^λ into the pair $C^{\lambda\pm}$ of conjugate complex matrix representations [see (3.2)]. Since $\mathbf{e}^\lambda$, $\mathbf{f}_1^\lambda$ and $\mathbf{g}_{00}^\lambda, \mathbf{g}_{11}^\lambda$ are related by nonsingular complex transformations, R^λ is $\mathbf{C}$-equivalent to $C^{\lambda+} \oplus C^{\lambda-}$. If the elements $\mathbf{g}_{00}^\lambda \mathbf{e}_{JJ}^\lambda$, $-\mathbf{f}_1^\lambda \mathbf{g}_{00}^\lambda \mathbf{e}_{JJ}^\lambda$, J fixed, are used as the basis of an ideal of $\mathbf{Q}$-type, the real representation R^λ is transformed into $C^\lambda \oplus C^\lambda$, the complex representation C^λ being composed of submatrices of the form $C^\mathbf{Q}$ [see (2.31)].

The extension of $\mathbf{A}_\mathbf{C}$ to

$$\mathbf{A}_\mathbf{Q} = \mathbf{Q}\mathbf{A}_\mathbf{C}\mathbf{Q} \tag{6.7}$$

is accomplished by the adjunction of $f_2 = j$ both as a left-multiplier (=operator) and as a right-multiplier (=coefficient in a linear combination). For the $\mathbb{R}$-type the ring basis $\mathbf{e}_{jJ}^\lambda$ already obtained for the ideal $\mathbf{A}_\mathbb{R}^\lambda$ can be used also for $\mathbf{A}_\mathbf{Q}^\lambda$. Since the functions corresponding to these elements are then $\mathbb{R}$-valued,

$$q\mathbf{e}_{jJ}^\lambda = \mathbf{e}_{jJ}^\lambda q \quad \text{for all } q \in \mathbf{Q} . \tag{6.8}$$

Conversely, if (6.8) holds [as is required by convention (2.51)] the functions corresponding to the elements $\mathbf{e}_{jJ}^\lambda$ must be $\mathbb{R}$-valued and the $\mathbf{e}$'s form also a basis of the $\mathbb{R}$-type ideal $\mathbf{A}_\mathbb{R}^\lambda$. Obviously, $\mathbf{A}_\mathbf{Q}^\lambda$ is then isomorphic to a matrix algebra over $\mathbb{Q}$, the dimension being the same as in the real case. For $\mathbf{C}$-type ideals the situation is slightly more complicated. The relation

$$f_2 \mathbf{g}_{00}^\lambda = \mathbf{g}_{11}^\lambda f_2 , \tag{6.9}$$

meeting one of the conditions required in (2.51), shows that pairs of conjugate ideals, say $\mathbf{A}_\mathbb{C}^{\lambda+}$ and $\mathbf{A}_\mathbb{C}^{\lambda-}$, are linked by f_2, resulting in one ideal $\mathbf{A}_\mathbf{Q}^\lambda = \mathbf{Q}\mathbf{A}_\mathbb{R}^\lambda\mathbf{Q} = \mathbf{A}_\mathbb{R}^\lambda\mathbf{Q}$. The division algebra

$$\mathbf{F}_\mathbb{C}^{\lambda+} \oplus \mathbf{F}_\mathbb{C}^{\lambda-} = \mathbf{e}^{\lambda+}\mathbf{C} \oplus \mathbf{e}^{\lambda-}\mathbf{C} = \mathbf{e}^\lambda\mathbf{C} \tag{6.10}$$

remains isomorphic to $\mathbf{C}$, but the dimension of the matrix algebra is doubled. For ideals of $\mathbf{Q}$-type the adjunction of f_2 yields new self-adjoint elements such as

$f_2^\lambda f_2$. Proceeding as for $f_1^\lambda f_1$, the idempotents

$$h_{rr}^\lambda = (1/2)[e^\lambda + f_2^\lambda f_2(-1)^r] \tag{6.11}$$

appearing in the spectral decomposition of $f_2^\lambda f_2$ can be used to define a ring basis of A_Q^λ. The elements

$$e_{jr,\,JR}^\lambda = \hat{f}_r e_{jJ}^\lambda g_{00}^\lambda h_{00}^\lambda \hat{f}_R^* , \tag{6.12}$$

where e_{jJ}^λ is one of the basis elements of A_R^λ satisfy both Eqs. (2.16) and (2.17) and the convention (2.51). Moreover, using the decomposition

$$be^\lambda = b^\lambda = \sum_{jJr} e_{jJ}^\lambda f_r^\lambda R_{jr,\,J0}^\lambda(b) , \tag{6.13}$$

the definition (6.12), Eq. (6.1) and the corresponding multiplication law for the numbers $f_r \in Q$, and Eq. (5.9), one obtains

$$e_{jr,\,JR}^\lambda b e_{JR,\,JR}^\lambda = e_{jr,\,JR}^\lambda R_{jr,\,JR}^\lambda(b) . \tag{6.14}$$

Equations (6.12) and (6.14) show A_Q^λ to be isomorphic to a full matrix algebra over $\mathbb{R}$, the dimension being four times the dimension of the matrix algebra isomorphic to A_R^λ.

The elements $e_{jr,\,JR}^\lambda$ forming a ring basis of a two-sided ideal A_Q^λ are linearly dependent in the index R since (2.51) implies

$$e_{jr,\,JR}^\lambda = \hat{f}_r e_{j0,\,J0}^\lambda \hat{f}_R^* . \tag{6.15}$$

Treating the left-multipliers $\hat{f}_r$ as norm-preserving operators satisfying $\hat{f}_r^* = \hat{f}_r \sigma_{r,}^{\hat{F}}$ and $\hat{f}_r \hat{f}_R = \hat{f}_{rR} \sigma_{r,R}^{\hat{F}}$, the orthogonality of the elements (6.15) in r is easily established. Orthogonality in j is proved as for A_R or A_C. To prove orthogonality in J it has to be noted that Eq. (2.51) implies $e_{jJ}^\lambda \in A_R^\lambda$ if e_{jJ}^λ is defined by

$$e_{jJ}^\lambda = \sum_r e_{jr,\,Jr}^\lambda ; \tag{6.16}$$

this result is independent of the special definition (6.12). Therefore

$$\begin{aligned} \langle e_{j0,\,J0}^\lambda, e_{j0,\,J'0}^\lambda \rangle &= \langle e_{jJ}^\lambda, e_{j0,\,J'0}^\lambda \rangle = \langle e_{J'0,\,j0}^\lambda, e_{Jj}^\lambda \rangle \\ &= \langle e_{J'0,\,j0}^\lambda, e_{J0,\,j0}^\lambda \rangle = 0 \text{ for } J \neq J' , \end{aligned} \tag{6.17}$$

the second equality being true only because of $e_{jJ}^\lambda \in A_R$ [for $a, b \in A_Q$ $\langle a, b \rangle \neq \langle b^+, a^+ \rangle$ in general]. Analogous to the real and to the complex case, the common norm m^λ of the elements $e_{jr,\,JR}^\lambda$ follows from the representation (5.15).

The matrix representation of A_Q carried by the basis $\{e_{jr,\,Jr'}^\lambda : \lambda Jr' \text{ fixed}\}$ and denoted by F^λ is determined in several steps: First the division algebra F_Q^λ is determined. The result, listed in (2.54)–(2.56), follows either from convention (2.51) [R-type] or from the fact that F_Q^λ is the center of A_Q^λ [C- and Q-type]. Next r' is chosen such that

$$f_s^\lambda e_{jr,\,Jr'}^\lambda = e_{jr,\,Jr'}^\lambda f_s^\lambda = e_{jr,\,Jr'}^\lambda f_s \quad \text{for all } f_s^\lambda \in F^\lambda . \tag{6.18}$$

This is always possible since $f_s^\lambda f_s$ is self-adjoint, $e_{Jr',\,Jr'}^\lambda$ primitive, and $f_s^\lambda \in F_Q^\lambda$ and [due to (2.51)] $f_s \in \tilde{F}$ both commute with $e_{Jr',\,Jr'}^\lambda$. Equation (6.18) entails $\tilde{F}^\lambda = F^\lambda$ since every $f \in F_Q^\lambda$ is a real linear combination of the elements $f_s^\lambda \in F_Q^\lambda$. Finally

$$\begin{aligned} b^\lambda &= \sum_{jr\,JR} e_{jr,\,JR}^\lambda f_{jr,\,JR}^\lambda(b) = \sum_{jr\,JR_s} e_{jr,\,JR}^\lambda f_s R_{jr,\,JR}^{\lambda(s)}(b) \\ &= \sum_{jr\,JR_s} f_s^\lambda e_{jr,\,Jr'}^\lambda \hat{f}_{r'R}^* \sigma_{r',r'R}^{\hat{F}} R_{jr,\,JR}^{\lambda(s)}(b) \\ &= \sum_{jr\,JR} e_{jr,\,Jr'}^\lambda \sum_s f_s R_{jr,\,JR}^{(s)}(b) \hat{f}_{r'R}^* \sigma_{r',r'R}^{\hat{F}} \\ &= \sum_{jr\,JR} e_{jr,\,Jr'}^\lambda \tilde{F}_{jr,\,JR}^\lambda(b) \hat{f}_{r'R}^* \sigma_{r',r'R}^{\hat{F}} , \end{aligned} \tag{6.19}$$

from which (2.53) is obtained by taking the scalar product with $e_{jr,\,Jr'}^\lambda$. The matrix representations $\tilde{F}^\lambda(e^\lambda q)$, $q \in Q$, can be computed directly from (2.51), the results are listed in (2.54)–(2.56).

B. Uniqueness

To make the discussion in Sec. 4 complete, the definition and the properties of the element u^λ have to be described in more detail. First it is to be noted that

$$e^\lambda = \sum_k {}^{(1)}e_{kk}^\lambda = \sum_k {}^{(2)}e_{kk}^\lambda \tag{6.20}$$

because the element e^λ is uniquely determined by its properties. Equation (6.20) implies that for a given ${}^{(1)}e_{k'k'}^\lambda = e^\lambda\, {}^{(1)}e_{k'k'}^\lambda (\neq 0)$ there must be an element ${}^{(2)}e_{k''k''}^\lambda$ ${}^{(1)}e_{k'k'}^\lambda \neq 0$. Since ${}^{(1)}e_{k'k'}^\lambda$ is primitive

$$\begin{aligned} {}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda &= {}^{(1)}f_{k'k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda \\ &= a_{k'k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda , \quad a_{k'k''}^\lambda \in \mathbb{R} , \end{aligned} \tag{6.21}$$

the last equality following from the fact that both the lhs and ${}^{(1)}e_{k'k'}^\lambda$ are self-adjoint and hence ${}^{(1)}f_{k'k''}^\lambda = {}^{(1)}f_{k'k''}^{\lambda+}$ $= a_{k'k''}\, e^\lambda$. Exchange of ${}^{(1)}e_{k'k'}^\lambda$ and ${}^{(2)}e_{k''k''}^\lambda$ yields a real number $a_{k''k'}^\lambda$. But $a_{k''k'}^\lambda = a_{k'k''}^\lambda$ since

$$\begin{aligned} {}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda &= a_{k'k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda \\ &= a_{k''k'}^\lambda\, {}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda . \end{aligned} \tag{6.22}$$

Next observe that

$$[{}^{(2)}e_{k''k''}^\lambda\, {}^{(1)}e_{k'k'}^\lambda][{}^{(1)}e_{k'k'}^\lambda\, {}^{(2)}e_{k''k''}^\lambda] = a_{k'k''}^\lambda\, {}^{(2)}e_{k''k''}^\lambda \tag{6.23}$$

is positive definite. Hence $a_{k'k''}^\lambda > 0$ and it is possible to define a positive number $b_{k'k''}^\lambda$ by

$$b_{k'k''}^\lambda = (a_{k'k''}^\lambda)^{-1/2} > 0 . \tag{6.24}$$

The first of Eqs. (4.2) then follows directly from (4.1), (6.21), and (6.24). The second is verified by recognizing that $u^{\lambda+}{}^{(2)}F_F^\lambda u^\lambda$ is a field commuting with all elements ${}^{(1)}e_{kK}^\lambda$; hence it must be equal to ${}^{(1)}F_F^\lambda$. The last of Eqs. (4.2) is trivially satisfied for the commutative fields $F = \mathbb{R}$ or C. For $F = Q$ conventions (2.51) holding both for ${}^{(1)}e_{jr,\,JR}^\lambda$ and ${}^{(2)}e_{jr,\,JR}^\lambda$ and definition (4.1) imply $u^{\lambda*} Q u^\lambda = e^\lambda Q$ since

$$u^\lambda \hat{f}_s = \hat{f}_s u^\lambda , \quad u^\lambda \tilde{f} = (\hat{f}_{r'r''}^* \tilde{f} \hat{f}_{r',r''}) u^\lambda = (\tilde{\omega f}) u^\lambda , \tag{6.25}$$

where $j'r'(=k')$ and $j''r''(=k'')$ are the fixed indices appearing in the definition of u^λ.

Equation (4.8) relating ${}^{(2)}\tilde{F}^\lambda$ to ${}^{(3)}\tilde{F}^\lambda$ is obtained from

(4.6), (2.18)–(2.20), and

$$\mathbf{u}^{\lambda *}\mathbf{b}\mathbf{u}^{\lambda} = \sum_{kKr} [\mathbf{u}^{\lambda *}\,{}^{(2)}\mathbf{e}^{\lambda}_{kK}\mathbf{u}^{\lambda}][\mathbf{u}^{\lambda *}\,{}^{(2)}\mathbf{f}^{\lambda}_{r}\mathbf{u}^{\lambda}]\,{}^{(2)}R^{\lambda(r)}_{kK}(\mathbf{b})$$

$$= \sum_{kKr} {}^{(3)}\mathbf{e}^{\lambda}_{kK}\,{}^{(3)}\mathbf{f}^{\lambda}_{r}\,{}^{(3)}R^{\lambda(r)}_{kK}(\mathbf{u}^{\lambda *}\mathbf{b}\mathbf{u}^{\lambda})\,. \tag{6.26}$$

To verify Eq. (4.9) for $\mathbf{F} = \mathbf{R}$ one uses (2.25) and (2.26), and

$$\mathbf{b}\,{}^{(2)}\mathbf{e}^{\lambda}_{Jj},\,{}^{(2)}\mathbf{f}^{\lambda}_{R} = \sum_{jr} {}^{(2)}\mathbf{e}^{\lambda}_{jj},\,{}^{(2)}\mathbf{f}^{\lambda}_{r}\,{}^{(2)}R^{\lambda}_{jr,\,JR}(\mathbf{b})$$

$$= \mathbf{u}^{\lambda}[\mathbf{u}^{\lambda *}\mathbf{b}\mathbf{u}^{\lambda}]\,{}^{(3)}\mathbf{e}^{\lambda}_{Jj},\,{}^{(3)}\mathbf{f}^{\lambda}_{R}\mathbf{u}^{\lambda *}$$

$$= \sum_{jr} {}^{(2)}\mathbf{e}^{\lambda}_{jj},\,{}^{(2)}\mathbf{f}^{\lambda}_{r}\,{}^{(3)}R^{\lambda}_{jr,\,JR}(\mathbf{u}^{\lambda *}\mathbf{b}\mathbf{u}^{\lambda})\,. \tag{6.27}$$

The proofs for $\mathbf{F} = \mathbf{C}$ or $\mathbf{Q}$ are quite similar.

For $\mathbf{F} = \mathbb{h}$ the second part of the equivalence relation, i.e., step (3) → (1), follows from (2.19), (4.7), and (4.10),

$${}^{(3)}\mathbf{e}^{\lambda}_{jj}\mathbf{b}\,{}^{(3)}\mathbf{e}^{\lambda}_{JJ} = {}^{(3)}\mathbf{e}^{\lambda}_{jJ}\sum_{r} {}^{(3)}\mathbf{f}^{\lambda}_{r}\,{}^{(3)}R^{\lambda}_{jr,\,J0}(\mathbf{b}) = {}^{(1)}\mathbf{e}^{\lambda}_{jj}\mathbf{b}\,{}^{(1)}\mathbf{e}^{\lambda}_{JJ}$$

$$= {}^{(1)}\mathbf{e}^{\lambda}_{jJ}\sum_{s} {}^{(1)}\mathbf{f}^{\lambda}_{s}\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b})$$

$$= {}^{(3)}\mathbf{e}^{\lambda}_{jJ}\sum_{rs} {}^{(3)}\mathbf{f}^{\lambda}_{r}O_{rs}\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b})\,, \tag{6.28}$$

implying [cf. Eqs. (4.10)]

$${}^{(3)}\tilde{F}^{\lambda}_{jJ}(\mathbf{b}) = \tilde{\alpha}\,{}^{(1)}\tilde{F}^{\lambda}_{jJ}(\mathbf{b})\,, \tag{6.29}$$

and the relation

$$\mathbf{b}\,{}^{(3)}\mathbf{e}^{\lambda}_{Jj},\,{}^{(3)}\mathbf{f}^{\lambda}_{R} = \sum_{jr} {}^{(3)}\mathbf{e}^{\lambda}_{jj},\,{}^{(3)}\mathbf{f}^{\lambda}_{r}\,{}^{(3)}R^{\lambda}_{jr,\,JR}(\mathbf{b})$$

$$= \mathbf{b}\,{}^{(1)}\mathbf{e}^{\lambda}_{Jj},\sum_{S} {}^{(1)}\mathbf{f}^{\lambda}_{S}O^{*}_{SR}$$

$$= \sum_{jrsS} {}^{(3)}\mathbf{e}^{\lambda}_{jj},\,{}^{(3)}\mathbf{f}^{\lambda}_{r}O_{rs}R^{\lambda}_{js,\,JS}(\mathbf{b})O^{*}_{SR}\,. \tag{6.30}$$

That Eqs. (6.29) and (6.30) fit together, i.e., that

$${}^{(3)}R^{\lambda}_{jr,\,JR}(\mathbf{b}) = \sum_{sS} O_{rs}\,{}^{(1)}R^{\lambda}_{js,\,JS}(\mathbf{b})O^{*}_{SR} = \tilde{R}^{\mathbf{F}}_{rR}[{}^{(3)}\tilde{F}_{jJ}(\mathbf{b})]$$

$$= R^{\mathbf{F}}_{rR}[\tilde{\alpha}\,{}^{(1)}\tilde{F}_{jJ}(\mathbf{b})]\,. \tag{6.31}$$

can be verified using the peculiar form of the matrices O,

$$\mathbf{R}\text{-type:}\quad O = 1\,,$$

$$\mathbf{C}\text{-type:}\quad O = \begin{pmatrix} 1 & 0 \\ 0 & \pm 1 \end{pmatrix},$$

$$\mathbf{Q}\text{-type:}\quad O = R^{\mathbf{Q}}[u^{*}]\bar{R}^{\mathbf{Q}}[u^{*}] = \bar{R}^{\mathbf{Q}}[u^{*}]R^{\mathbf{Q}}[u^{*}]\,,\quad uu^{*} = 1\,. \tag{6.32}$$

The corresponding proof for $\mathbf{F} = \mathbf{C}$ is redundant since in this case (4.3), $\mathbf{u}^{\lambda}c = c\mathbf{u}^{\lambda}$, and $\mathbf{F}^{\lambda}_{\mathbf{C}} = \mathbf{e}^{\lambda}\mathbf{C}$, imply $\tilde{\alpha}(c\mathbf{e}^{\lambda}) = c\mathbf{e}^{\lambda}$. For $\mathbf{F} = \mathbf{Q}$, Eq. (4.9) follows from (4.7) and, for the $\mathbf{R}$-type, from

$${}^{(3)}\mathbf{f}_{s}\mathbf{e}^{\lambda} = \alpha^{-1}\,{}^{(1)}\mathbf{f}_{s} = [\tilde{\alpha}^{-1}\,{}^{(1)}f_{s}]\mathbf{e}^{\lambda} = [u\,{}^{(1)}f_{s}u^{*}]\mathbf{e}^{\lambda}$$

$$= \sum_{rt} {}^{(1)}f_{r}\bar{R}^{\mathbf{Q}}_{rt}[u]R^{\mathbf{Q}}_{ts}[u]\mathbf{e}^{\lambda} = \sum_{rt} {}^{(1)}f_{r}R^{\mathbf{Q}}_{rt}[u]\bar{R}^{\mathbf{Q}}_{ts}[u]\mathbf{e}^{\lambda}\,, \tag{6.33}$$

$${}^{(3)}\mathbf{e}^{\lambda}_{jj}\mathbf{b}\,{}^{(3)}\mathbf{e}^{\lambda}_{JJ} = {}^{(3)}\mathbf{e}^{\lambda}_{jJ}\sum_{r} {}^{(3)}f_{r}\,{}^{(3)}R^{\lambda}_{jr,\,J0}(\mathbf{b}) = {}^{(1)}\mathbf{e}^{\lambda}_{jj}\mathbf{b}\,{}^{(1)}\mathbf{e}^{\lambda}_{JJ}$$

$$= {}^{(1)}\mathbf{e}^{\lambda}_{jJ}\sum_{s} {}^{(1)}f_{s}\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b})\,, \tag{6.34}$$

$${}^{(3)}\mathcal{Q}^{\lambda}_{jJ}(\mathbf{b}) = \sum_{r} f_{r}\,{}^{(3)}R^{\lambda}_{jr,\,J0}(\mathbf{b}) = \sum_{rst} f_{r}\bar{R}^{\mathbf{Q}}_{st}[u]R^{\mathbf{Q}}_{tr}[u]\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b})$$

$$= \sum_{rst} f_{r}R^{\mathbf{Q}}_{rt}[u^{*}]\bar{R}^{\mathbf{Q}}_{ts}[u^{*}]\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b}) = \sum_{s} (u^{*}f_{s}u)\,{}^{(1)}R^{\lambda}_{js,\,J0}(\mathbf{b})$$

$$= u^{*}\,{}^{(1)}\mathcal{Q}^{\lambda}_{jJ}(\mathbf{b})u\,. \tag{6.35}$$

The simple proof for the $\mathbf{C}$-type is left to the reader. For the $\mathbf{Q}$-type $\tilde{\alpha}$ is the identical mapping and the step (3) → (1) is trivial.

C. Matrix elements

In this subsection the relation between the elements of the ring bases and the corresponding matrix representations is established for $\mathbf{F} = \mathbf{R}$ and $\mathbf{F} = \mathbf{Q}$. Let $E_{jJ,r} \in L^{2}(G, \mathbf{R})$ be the function corresponding to $\mathbf{e}^{\lambda}_{jJ}\mathbf{f}^{\lambda}_{r} \in \mathbf{A}_{\mathbf{R}}$, i.e.,

$$\mathbf{e}^{\lambda}_{jJ}\mathbf{f}^{\lambda}_{r} = M_{x}E^{\lambda}_{jJ,r}(x)\mathbf{x}\,. \tag{6.36}$$

The transformation properties under $\mathbf{y}^{-1}$, $y \in G$, are then given by

$$\mathbf{y}^{-1}\mathbf{e}^{\lambda}_{Jj},\mathbf{f}^{\lambda}_{R} = \sum_{jr} \mathbf{e}^{\lambda}_{jj},\mathbf{f}^{\lambda}_{r}R^{\lambda}_{jr,\,JR}(y^{-1})\,,$$

$$E^{\lambda}_{Jj',R}(yx) = \sum_{jr} E^{\lambda}_{jj',r}(x)R^{\lambda}_{jr,\,JR}(y^{-1})\,. \tag{6.37}$$

$\{R^{\lambda}(y) : y \in G\}$ has to be an orthogonal matrix representation [cf. Eqs. (5.10) and (5.11)] since $\mathrm{G} \cong G$ and $\mathbf{y}$ is a norm-preserving operator on $L^{2}(G, \mathbf{R})$. Putting $x = e$ one has

$$E^{\lambda}_{Jj',R}(y) = \sum_{jr} E^{\lambda}_{jj',r}(e)R^{\lambda}_{jR,\,jr}(y)\,. \tag{6.38}$$

Observing that $a^{*}b$ is a continuous function,[14] that

$$\langle \mathbf{a}^{*}, \mathbf{b}\rangle = a^{*}b(e)\,, \tag{6.39}$$

[cf. Eqs. (2.1), (2.4), and (2.10)], and that Eqs. (6.36), (2.15), and (2.17) imply $E^{\lambda}_{jJ,0}{}^{*}E^{\lambda}_{Jj',r} = E^{\lambda}_{jj',r}$, the numbers $E^{\lambda}_{jj',r}(e)$ can be related to a scalar product,

$$E^{\lambda}_{jj',r}(e) = E^{\lambda}_{jJ,0}{}^{*}E^{\lambda}_{Jj',r}(e) = \langle \mathbf{e}^{\lambda}_{Jj}\mathbf{f}^{\lambda}_{0}, \mathbf{e}^{\lambda}_{Jj'}\mathbf{f}^{\lambda}_{r}\rangle = N^{\lambda}\delta_{jj'}\delta_{r0}\,. \tag{6.40}$$

The normalization constant N^{λ} is determined from

$$n^{\lambda}N^{\lambda} = \sum_{jr} \langle \mathbf{e}^{\lambda}_{jJ}\mathbf{f}^{\lambda}_{r}, \mathbf{e}^{\lambda}_{jJ}\mathbf{f}^{\lambda}_{r}\rangle = \sum_{jr} M_{x}N^{\lambda}R^{\lambda}_{jr,\,J0}(x)N^{\lambda}R^{\lambda}_{jr,\,J0}(x)$$

$$= (N^{\lambda})^{2}M_{x}\sum_{jr} R^{\lambda}_{J0,\,jr}(x^{-1})R^{\lambda}_{jr,\,J0}(x) \tag{6.41}$$

to be equal to $n^{\lambda} = \dim R^{\lambda}(x)$. This completes the proof of (5.7). Equations (5.13) and (5.14) are only transcriptions of the corresponding properties of the elements $\mathbf{e}^{\lambda}_{jJ}\mathbf{f}^{\lambda}_{r}$. The equivalence relation (5.14) has already been derived in subsection 6B [see Eqs. (6.27), (6.31), and (6.32)].

Equation (5.15) relating the quaternionic ring elements to matrix representations $\{f^{\lambda}(x) : x \in G\}$ and Eqs. (5.17)–(5.20) are easily verified using Eqs. (5.1)–(5.13), if the elements $\mathbf{e}^{\lambda}_{jr,\,JR} \in \mathbf{A}_{\mathbf{Q}}$ are defined according to the follow-

ing conventions:

$\mathbf{R}$-type: $e^{\lambda}_{j0,\,J0} = e^{\lambda}_{jJ}$;

$\mathbf{C}$-type: $e^{\lambda}_{jr,\,JR} = j^{r} g^{\lambda}_{00} e^{\lambda}_{jJ} (-j)^{R}$, g^{λ}_{00} see (6.5);

$\mathbf{Q}$-type: $e^{\lambda}_{jr,\,JR} = f_r g^{\lambda}_{00} h^{\lambda}_{00} e^{\lambda}_{jJ} f_R$, g^{λ}_{00} see (6.5);

h^{λ}_{00} see (6.11);

all types: $\{e^{\lambda}_{jJ}\}$ = basis of $A^{\lambda}_{\mathbf{R}}$. (6.42)

This proof of Eqs. (5.15)–(5.20) for a special choice of elements $e^{\lambda}_{jr,\,JR}$ and representations f^{λ} can be extended to all equivalent sets using the equivalence relation (5.21) which has already been proved in subsection 6B using a different notation [Eqs. (6.26)–(6.29), $\tilde{F}^{\lambda} \to f^{\lambda}$].

Note added in proof: After completion of this paper I came across the paper "Quaternionic representations of compact metric groups" by S. Natarajan and K. Viswanath [J. Math. Phys. 8, 582 (1967)]. These authors derive orthogonality and completeness relations for matrix elements and characters of irreducible quaternionic representations and they point out the relation of the type of these representations to the fields $\mathbf{R}$, $\mathbf{C}$, $\mathbf{Q}$.

[1] P. Kasperkovitz and G. Kahl, J. Math. Phys. 22, 2402 (1981).
[2] E. C. G. Stueckelberg, Helv. Phys. Acta **33**, 727 (1960); E. C. G. Stueckelberg and M. Guenin, Helv. Phys. Acta **34**, 621 (1961).
[3] D. Finkelstein, J. M. Jauch, and D. Speiser, CERN 59-7, 59-9, 59-17; J. M. Jauch, "Projective representation of the Poincaré group in a quaternionic Hilbert space," in Vol. I of *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1968).
[4] D. Finkelstein, J. M. Jauch, and D. Speiser, J. Math. Phys. **4**, 136 (1963).
[5] G. Emch, Helv. Phys. Acta **36**, 739 (1963).
[6] G. Frobenius and I. Schur, Sitzungsber. Preuss. Akad. B **II**, 186 (1906).
[7] M. A. Naimark, *Normed Rings* (Noordhoff, Groningen, Holland, 1964).
[8] R. Dirl and P. Kasperkovitz, *Gruppentheorie, Anwendungen in der Atom- und Festkörperphysik* (Vieweg, Braunschweig, Germany, 1977).
[9] B. L. van der Waerden, *Algebra*, 2.Teil (Springer, Berlin, 1967), pp. 74–78.
[10] Ref. 7, p. 331.
[11] Ref. 7, p. 431.
[12] Ref. 9, p. 73.
[13] Ref. 9, p. 108.
[14] Ref. 7, p. 374.

Analytic evaluation of certain zeroth order coulombic hyperangular interaction integrals[a)]

Metin Demiralp and N. Abdülbaki Baykara

Applied Mathematics Department, Marmara Scientific and Industrial Research Institute, P. K. 141 Kadıköy–Istanbul, Turkey

(Received 16 July 1980; accepted for publication 19 September 1980)

As has been presented in a previous paper, the use of hyperspherical coordinates in research on a system of electrically charged particles carries mathematical complications into the evaluation of certain kinds of hyperangular interaction integrals. These integrals contain the hyperangular interaction potential and various powers of the inverse of the total angular momentum operator on the space of hyperharmonics, from which the zeroth order hyperharmonic is excluded. In this work the simplest one of these integrals has been taken into consideration. After some intermediate steps, it has been shown that it can be expressed in terms of elementary functions of the cosine of the angle (γ) between the hyperaxes of the potential term for an odd number of particles. In the case of an even number of particles, these integrals can be given in terms of a generalized hypergeometric function of the same argument (γ) and its derivatives.

PACS numbers: 02.30.Gp

I. INTRODUCTION

The use of the space-folding method in the hyperspherical coordinates respresentation of Schrödinger's equation for electrically charged particles brings us to a point where the evaluation of the system's energy value necessitates the determination of some complicated hyperangular integrals as presented in a previous paper.[1] Although the convergence of the perturbative series could not be proved, some results up to second order terms encouraged us to investigate the method presented there in details. But these investigations need the values of higher order terms; the evaluation of hyperangular interaction integrals plays an important role to this end. However, if we consider them generally they are not easy to compute. In a separate paper we present a numerical method for this purpose,[2] but an analytical expression for them seems to be a hopeless goal. Before making such a harsh judgment it is convenient to start with the simplest one and to try to find an analytical result, which is the purpose of this paper.

In the following sections, we shall define hyperangular interaction integrals generally, then choose the simplest one and reduce it to most convenient form by using several of its properties. Finally the use of hypergeometric function theory will take us to the desired result.

II. EXPANSION OF THE HYPERANGULAR INTERACTION INTEGRALS

Let us consider the following general integral

$$\chi_p\left[\begin{matrix} m_1, m_2, \cdots \\ \cdots m_p \end{matrix}\middle|\begin{matrix} u_1, \cdots \\ \cdots u_{p+1} \end{matrix}\right] \equiv \int_{S_\xi} \theta_0^*(\xi^T A_1 \xi)^{-1/2} \mathcal{L}^{m_1}(\xi^T A_2 \xi)^{-1/2} \mathcal{L}^{m_2} \cdots (\xi^T A_p \xi)^{-1/2} \times \mathcal{L}^{m_p}(\xi^T A_{p+1} \xi)^{-1/2} \theta_0 \, dS_\xi$$

ξ: $3N$-dimensional unit vector, (2.1)

where $m_1, \ldots, m_p$ represent some positive integers and the integral is evaluated on the hypersurface of the $3N$-dimensional hypersphere with unit radius. In addition the matrices A_j and the operator $\mathcal{L}$ are defined as below

$$A_j = u_j u_j^T \otimes I_3, \quad j = 1, 2, \ldots, p+1, \tag{2.2}$$

$$\mathcal{L}\varphi = -\sum_{j=1}^{\infty} \sum_{k=1}^{d_j} \frac{1}{j(j+3N-2)} \theta_j^{(k)} \int_S \theta_j^{(k)*} \varphi \, dS. \tag{2.3}$$

In the last expressions, u_j, $\theta_j^{(k)}$, and φ represent some unit $3N$-dimensional vector, a jth order hyperharmonic,[3] and a function in the space spanned by hyperharmonics, respectively.

Since we shall consider the simplest of the χ_p's hereafter our interest lies in the following integral:

$$\chi_1(m \mid u_1, u_2) \equiv \int_S \theta_0^* (\xi^T A_1 \xi)^{-1/2} \mathcal{L}^m (\xi^T A_2 \xi)^{-1/2} \theta_0 \, dS. \tag{2.4}$$

By using the addition theorem for hyperspherical harmonics[4] we can write the following equality for $\mathcal{L}^m$

$$\mathcal{L}^m \varphi = \frac{\Gamma(\alpha)}{2\pi^{\alpha+1}} \sum_{j=1}^{\infty} \frac{(-1)^m (j+\alpha)}{j^m (j+2\alpha)^m} \int_{S_\eta} C_j^\alpha(\xi^T \eta) \varphi(\eta) \, dS_\eta,$$

$\alpha = (3N-2)/2$, ξ, η: $3N$-dimensional unit vectors, (2.5)

where C_j^α denotes a Gegenbauer polynomial.[5]

By partial fractions we can obtain the following equation

$$\frac{1}{j^m (j+2\alpha)^m} = \sum_{k=0}^{m-1} \frac{\Gamma(2m-k-1)}{(m-k-1)!\,(m-1)!} \frac{1}{(2\alpha)^{2m-k-1}} \times \left(\frac{(-1)^{m-k-1}}{j^{k+1}} + \frac{(-1)^m}{(j+2\alpha)^{k+1}} \right). \tag{2.6}$$

The integral representations of the right-hand side terms give the following equality:

$$\frac{1}{j^m (j+2\alpha)^m} = \frac{(-1)^{m+1}}{(m-1)!} \sum_{k=1}^{m-1} \frac{\Gamma(2m-k-1)(-1)^k}{(m-k-1)!\,k!\,(2\alpha)^{2m-k-1}} \times \int_0^1 \frac{1-(-1)^k x^{2\alpha}}{x} x^j (-\ln x)_k^k \, dx. \tag{2.7}$$

[a)] This project has been supported by The Scientific and Technical Research Council of Turkey.

 0022-2488/81/112427-06$01.00

Since the Gegenbauer polynomials satisfy the following relation[6]:

$$\sum_{j=1}^{\infty}(j+\alpha)x^j C_j^{\alpha}(\xi^T\eta)=\alpha\left(\frac{1-x^2}{[1-2(\xi^T\eta)x+x^2]^{\alpha+1}}-1\right), \tag{2.8}$$

we can rewrite the Eq. (2.5) as follows:

$$\mathfrak{L}^m\varphi=-\frac{\Gamma(\alpha+1)}{(m-1)!\,2\pi^{\alpha+1}}\sum_{k=0}^{m-1}\frac{\Gamma(2m-k-1)(-1)^k}{k!(m-k-1)!(2\alpha)^{2m-k-1}}\times\int_{S_\eta}\int_0^1\frac{1-(-1)^k x^{2\alpha}}{x}\left(\frac{1-x^2}{[1-2(\xi^T\eta)x+x^2]^{\alpha+1}}-1\right)\times(-\ln x)^k\,dx\,\varphi(\eta)\,dS_\eta. \tag{2.9}$$

If we define the following function

$$\Lambda_k^{\alpha}(u_1,u_2)=\int_0^1\int_{S_\xi}\int_{S_\eta}(\xi^T A_1\xi)^{-1/2}(\eta^T A_2\eta)^{-1/2}\frac{1-(-1)^k x^{2\alpha}}{x}\times\left(\frac{1-x^2}{[1-2(\xi^T\eta)x+x^2]^{\alpha+1}}-1\right)\times(-\ln x)^k\,dS_\xi\,dS_\eta\,dx \tag{2.10}$$

by using the definition[7] of θ_0, $\theta_0=[\Gamma(\alpha+1)/2\pi^{\alpha+1}]^{1/2}$ we can write

$$\chi_1(m\,|u_1,u_2)=\left(\frac{\Gamma(\alpha+1)}{2\pi^{\alpha+1}}\right)^2\frac{1}{(m-1)!}\times\sum_{k=0}^{m-1}\frac{\Gamma(2m-k-1)(-1)^{k+1}}{k!(m-k-1)!(2m)^{2m-k-1}}\Lambda_k^{\alpha}(u_1,u_2). \tag{2.11}$$

If we expand Λ_k^{α} into a power series in $(\xi^T\eta)$, we can write

$$\Lambda_k^{\alpha}(u_1,u_2)=\sum_{j=0}^{\infty}\frac{([\alpha+1]/2)_j([\alpha+2]/2)_j}{(\frac{1}{2})_j\,j!}\beta_j^k\tau_j(u_1,u_2), \tag{2.12}$$

where $(a)_k$ denotes Pochhammer's symbol.[8] For the sake of simplicity the α dependence is not shown explicitly. β and τ in the last formula are defined as

$$\beta_j^k=\int_0^1\frac{1-(-1)^k x^{2\alpha}}{x}\left[\frac{1-x^2}{(1+x^2)^{\alpha+1}}\left(\frac{2x}{1+x^2}\right)^{2j}-\delta_{j,0}\right](-\ln x)^k\times dx, \tag{2.13}$$

$$\tau_j(u_1,u_2)=\int_{S_\xi}\int_{S_\eta}(\xi^T\eta)^{2j}(\xi^T A_1\xi)^{-1/2}(\eta^T A_2\eta)^{-1/2}dS_\xi\,dS_\eta, \tag{2.14}$$

ξ,η: $3N$-dimensional unit vectors. (2.15)

III. ANALYTICAL EVALUATION OF β INTEGRALS

The following transformation changes β_j^k into

$$\sqrt{1-u^2}=\frac{2x}{1+x^2},\quad x=\left(\frac{1-u}{1+u}\right)^{1/2}, \tag{3.1a, b}$$

$$\beta_j^k=\frac{1}{2^k}\int_0^1\frac{(1+u)^{\alpha}-(-1)^k(1-u)^{\alpha}}{(1+u)^{\alpha+1}(1-u)}\left\{\frac{u(1-u)^j(1+u)^{j+\alpha}}{2^{\alpha}}-\delta_{j0}\right\}\times\ln^k\left(\frac{1+u}{1-u}\right)du. \tag{3.2}$$

Since we can write

$$\ln^k\left(\frac{1+u}{1-u}\right)=\frac{d^k}{dt^k}\left\{\left(\frac{1+u}{1-u}\right)^t\right\}_{t=0}, \tag{3.3}$$

Eq. (3.2) implies

$$\beta_j^k=\frac{1}{2^k}\left\{\frac{d^k}{dt^k}[\sigma_1(j\,|t)-\sigma_2(j\,|t)-(-1)^k\sigma_1(j\,|t-\alpha)+(-1)^k\sigma_2(j\,|t-\alpha)]\right\}_{t=0}$$
$$=\frac{1}{2^k}\left\{\frac{d^k}{dt^k}[\sigma_1(j\,|t)-\sigma_2(j\,|t)-\sigma_1(j\,|-t-\alpha)+\sigma_2(j\,|-t-\alpha)]\right\}_{t=0}, \tag{3.4}$$

where

$$\sigma_1(j\,|t)=\frac{1}{2^{\alpha}}\int_0^1(1+u)^{j+t+\alpha+1}(1-u)^{j-t-1}u\,du$$
$$=\frac{1}{2^{\alpha}}\frac{1}{(j-t+1)(j-t)}\,{}_2F_1\left(\begin{matrix}-j-\alpha-t+1,2\\ j-t+2\end{matrix}\middle|-1\right)$$
$$=\frac{2^{j+t-1}}{(j-t+1)(j-t)}\,{}_2F_1\left(\begin{matrix}-j-\alpha-t+1,j-t\\ j-t+2\end{matrix}\middle|\frac{1}{2}\right), \tag{3.5}$$

$$\sigma_2(j\,|t)=\delta_{j0}\int_0^1(1+u)^{t-1}(1-u)^{-t-1}\,dt=-\frac{1}{2t}\,\delta_{j0}. \tag{3.6}$$

The use of some linear transformations for Gaussian hypergeometric functions (Ref. 9, pg. 47) and some simplifications for the special values of their parameters (Ref. 9, pg. 37) make it possible to write the following equations:

$$\sigma_1(j\,|t)=\frac{2^{-\alpha-2}}{(j+\alpha+t)(j+\alpha+t+1)}\,{}_2F_1\left(\begin{matrix}2j+\alpha+1,2\\ j+\alpha+t+2\end{matrix}\middle|\frac{1}{2}\right)+2^{2k}\frac{\Gamma(j-t)\Gamma(j+\alpha+t)}{\Gamma(2j+\alpha+1)}\left(t+\frac{\alpha}{2}\right), \tag{3.7}$$

$$\sigma_1(j\,|-t-\alpha)=\frac{2^{-\alpha-2}}{(j+\alpha+t)(j+\alpha+t+1)}\,{}_2F_1\left(\begin{matrix}2j+\alpha+1,2\\ j+\alpha+t+2\end{matrix}\middle|\frac{1}{2}\right).$$

However these last formulas together with Eqs. (3.4) and (3.6) bring us to the following result

$$\beta_j^k=\frac{1}{2^k}\left\{\frac{d^k}{dt^k}\left[\frac{2^{2j}\Gamma(j+\alpha+t)\Gamma(j-t)}{\Gamma(2j+\alpha+1)}\left(t+\frac{\alpha}{2}\right)+\left(\frac{1}{2t}+\frac{1}{2(t+\alpha)}\right)\delta_{j0}\right]\right\}_{t=0}. \tag{3.8}$$

IV. ANALYTICAL EVALUATION OF τ INTEGRALS

Let us consider the τ_j given by the Eq. (2.14). For its evaluation we can choose an orthonormal transformation on the frame of η that changes τ_j into the fol-

lowing form[1]:

$$\tau_j = \int_{S_\xi}\int_{S_\eta} \frac{[\eta_1(u_{21}^T\xi)+\eta_2(u_{22}^T\xi)+\eta_3(u_{23}^T\xi)+\sqrt{1-\sigma^2}\eta_4]^{2j}}{(\eta_1^2+\eta_2^2+\eta_3^2)^{1/2}(\xi^T A_1\xi)^{1/2}} \times dS_\eta\, dS_\xi\,, \tag{4.1}$$

where σ and u_{2m} are defined by

$$\sigma=(\xi^T A_2\xi)^{1/2}\,;\quad u_{2m}=u_2\otimes e_m,\quad m=1,2,3 \tag{4.2a, b}$$

$$e_1^T=[1,0,0],\quad e_2^T=[0,1,0],\quad e_3^T=[0,0,1]. \tag{4.2c, d, e}$$

The first four columns of the matrix representation for this orthonormal transformation are u_1, u_2, u_3, and q, respectively. The remaining columns are orthonormal to these four vectors. The definitions of q can be given as follows

$$q=(1-\sigma^2)^{-1/2}[\xi-A_2\xi]. \tag{4.3}$$

Another orthonormal transformation in the space of η_1, η_2, η_3 takes τ_j from Eq. (4.1) into the following form[1]:

$$\tau_j = \int_{S_\xi}\int_{S_\eta} (\sigma\eta_3+\sqrt{1-\sigma^2}\eta_4)^{2j}(\eta_1^2+\eta_2^2+\eta_3^2)^{-1/2}(\xi^T A_1\xi)^{-1/2} \times dS_\eta\, dS_\xi\,. \tag{4.4}$$

By passing to hyperspherical coordinates ($\eta_{3N}=\cos\theta_{3N-1}$, $\eta_{3N-1}=\sin\theta_{3N-1}\cos\theta_{3N-2}$, $\cdots$) and using some reduction formulas presented in the previous paper[1] we can write

$$\tau_j=\frac{4\pi^\alpha\,\Gamma(j+\frac12)}{\Gamma(j+\alpha+\frac12)}\int_0^1\int_S (\xi^T A_1\xi)^{-1/2}(1-\sigma^2t^2)^j\,dS\,dt\,. \tag{4.5}$$

Now let us consider the orthonormal transformation where the matrix representation has the columns u_2, $(1-\gamma^2)^{1/2}(u_1-\gamma u_2)$, and the vectors orthonormal to these two. Direct post-multiplication[10] of this orthonormal matrix by the three-dimensional unit matrix augments it to a $3N$-dimensional, orthonormal matrix. The use of this last transformation in Eq. (4.5) is followed by a rotation in the space of (ξ_1, ξ_2, ξ_3). Finally reduction formulas[1] give the following result for τ_j:

$$\tau_j(\gamma)=\frac{64\pi^{2\alpha}\,\Gamma(j+\frac12)}{\Gamma(j+\alpha+\frac12)\Gamma(\alpha-2)}\int_0^1\int_0^1\int_0^1\int_0^\pi \frac{(1-\mu_1^2\mu_2^2t^2)^j\mu_1^2(1-\mu_1^2)^{1/2}\mu_2^4(1-\mu_2^2)^{\alpha-3}}{\sqrt{\{\gamma^2\mu_1^2+2\gamma\sqrt{(1-\gamma^2)}\sqrt{(1-\mu_1^2)}\mu_1\cos\theta+(1-\gamma^2)(1-\mu_1^2)\}}}\sin\theta\,d\theta\,d\mu_1\,d\mu_2\,dt\,. \tag{4.6}$$

A careful investigation shows that

$$\int_0^\pi \sin\theta[\gamma^2\mu_1^2+2\gamma\mu_1\sqrt{1-\gamma^2}\sqrt{1-\mu_1^2}\cos\theta+(1-\gamma^2)(1-\mu_1^2)]^{-1/2}\,d\theta = \begin{cases}\dfrac{2}{\mu_1\gamma}\,, & \mu_1>\sqrt{1-\gamma^2}\,,\\[2ex] \dfrac{2}{\sqrt{(1-\gamma^2)}\sqrt{(1-\mu_1^2)}}\,, & \mu_1<\sqrt{1-\gamma^2}\,,\end{cases} \tag{4.7}$$

and

$$\int_0^1\int_0^1 (1-\mu_1^2\mu_2^2t^2)^j\mu_2^4(1-\mu_2^2)^{\alpha-3}\,d\mu_2\,dt=\frac{3\sqrt{\pi}}{8}\frac{\Gamma(\alpha-2)}{\Gamma(\alpha+\frac12)} \times {}_3F_2\left(\begin{matrix}-j,\frac12,\frac52\\ \alpha+\frac12,\frac32\end{matrix}\middle|\mu_1^2\right) = \frac{\sqrt{\pi}}{8}\frac{\Gamma(\alpha-2)}{\Gamma(\alpha+\frac12)}\frac{1}{\mu_1^2}\frac{d}{d\mu_1}\mu_1^3\,{}_2F_1\left(\begin{matrix}-j,\frac12\\ \alpha+\frac12\end{matrix}\middle|\mu_1^2\right). \tag{4.8}$$

Therefore

$$\tau_j(\gamma)=\frac{16\pi^{2\alpha+1/2}\,\Gamma(j+\frac12)}{\Gamma(\alpha+\frac12)\Gamma(j+\alpha+\frac12)}\left[(1-\gamma^2)\,{}_2F_1\left(\begin{matrix}-j,\frac12\\ \alpha+\frac12\end{matrix}\middle|1-\gamma^2\right) + \frac{1}{\gamma}\int_{\sqrt{(1-\gamma^2)}}^1 \frac{\sqrt{1-\mu^2}}{\mu}\frac{d}{d\mu}\mu^3\,{}_2F_1\left(\begin{matrix}-j,\frac12\\ \alpha+\frac12\end{matrix}\middle|\mu^2\right)d\mu\right]. \tag{4.9}$$

Integrating by parts gives the following result

$$\tau_j(\gamma)=\frac{16\pi^{2\alpha+1/2}\,\Gamma(j+\frac12)}{\Gamma(\alpha+\frac12)\Gamma(j+\alpha+\frac12)}\int_0^1 {}_2F_1\left(\begin{matrix}-j,\frac12\\ \alpha+\frac12\end{matrix}\middle|1-\gamma^2\mu^2\right)d\mu\,. \tag{4.10}$$

By expanding the hypergeometric function above in powers of $\gamma^2\mu^2$ and using the formula for hypergeometric functions with unit argument (Ref. 11, pg. 99–103) and then integrating term by term we arrive at the following formula

$$\tau_k(\gamma)=\frac{16\pi^{2\alpha+1/2}(k+\frac12)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma^2(k+\alpha+\frac12)}\,{}_3F_2\left(\begin{matrix}-k,\frac12,\frac12\\ 1-\alpha-k,\frac32\end{matrix}\middle|\gamma^2\right). \tag{4.11}$$

V. EVALUATION OF $\Lambda_k^\alpha(\gamma)$

By using Eq. (4.10) and (3.8) in the Eq. (2.12) and the integral representation for ${}_2F_1$ in the resulting equation, we can obtain the following formula after some algebra:

$$\Lambda_k^\alpha(\gamma)=\frac{2^{4-k}\pi^{2\alpha+1/2}}{\Gamma(\alpha)\Gamma(\alpha+\frac12)}\left\{\frac{d^k}{dt^k}\left[\frac{\Gamma(\alpha+t)\Gamma(-t)}{\Gamma(\alpha+1)}(t+\tfrac12\alpha)\Omega_\alpha(t,\gamma) + \left(\frac{1}{2t}+\frac{1}{2(t+\alpha)}\right)\frac{\Gamma(\alpha)\Gamma(\frac12)}{\Gamma(\alpha+\frac12)}\right]\right\}_{t=0}, \tag{5.1}$$

where $\Omega_\alpha(t,\gamma)$ is defined as

$$\Omega_\alpha(t,\gamma)=\int_0^1\int_0^1 \frac{(1-u)^{\alpha-1}}{\sqrt{u}}\,{}_2F_1\left(\begin{matrix}\alpha+t,-t\\ \alpha+\frac12\end{matrix}\middle|1-(1-\gamma^2\mu^2)u\right) \times du\,d\mu\,. \tag{5.2}$$

Linear transformation formula for the Gaussian hypergeometric functions makes it possible to write

$$\Omega_\alpha(t,\gamma)=\frac{\Gamma(\alpha+\frac12)\Gamma(\frac12)}{\Gamma(\alpha+\frac12+t)\Gamma(\frac12-t)}\Omega_\alpha^{(1)}(t,\gamma)+\frac{\Gamma(\alpha+\frac12)\Gamma(-\frac12)}{\Gamma(\alpha+t)\Gamma(-t)} \times\Omega_\alpha^{(2)}(t,\gamma)\,, \tag{5.3}$$

where $\Omega_\alpha^{(1)}$ and $\Omega_\alpha^{(2)}$ are

$$\Omega_\alpha^{(1)}(t,\gamma)=\int_0^1\int_0^1 \frac{(1-u)^{\alpha-1}}{\sqrt{u}}\,{}_2F_1\left(\begin{matrix}\alpha+t,-t\\ \frac12\end{matrix}\middle|(1-\gamma^2\mu^2)u\right) \times du\,d\mu\,, \tag{5.4}$$

$$\Omega_\alpha^{(2)}(t,\gamma)=\int_0^1\int_0^1 (1-u)^{\alpha-1}\sqrt{1-\gamma^2\mu^2}_2 F_1\left(\begin{matrix}\frac{1}{2}-t,\ \alpha+\frac{1}{2}+t\\ \frac{3}{2}\end{matrix}\middle|(1-\gamma^2\mu^2)u\right)du\,d\mu\,. \tag{5.5}$$

The integration of the Eq. (5.4) over u gives

$$\Omega_\alpha^{(1)}(t,\gamma)=\frac{\Gamma(\frac{1}{2})\Gamma(\alpha)}{\Gamma(\alpha+\frac{1}{2})}\int_0^1 {}_2F_1\left(\begin{matrix}\alpha+t,\ -t\\ \alpha+\frac{1}{2}\end{matrix}\middle|1-\gamma^2\mu^2\right)d\mu\,. \tag{5.6}$$

If we rewrite the hypergeometric function of $1-\gamma^2\mu^2$ in terms of hypergeometric functions of $\gamma^2\mu^2$ and integrate term by term and then change the resulting hypergeometric function of γ^2 into a hypergeometric function of $1-\gamma^2$ we obtain the following result

$$\Omega_\alpha^{(1)}(t,\gamma)=\frac{\sqrt{\pi}\Gamma(\alpha)}{\Gamma(\alpha-\frac{1}{2})(1+2t)(\alpha-\frac{1}{2}+t)}{}_2F_1\left(\begin{matrix}\alpha+t,\ -t\\ \alpha-\frac{1}{2}\end{matrix}\middle|1-\gamma^2\right)-\frac{\pi\Gamma(\alpha)}{(1+2t)(\alpha-\frac{1}{2}+t)\Gamma(\alpha+t)\Gamma(-t)}\frac{1}{\gamma}\,. \tag{5.7}$$

Term by term integration of Eq. (5.5) over u gives

$$\Omega_\alpha^{(2)}(t,\gamma)=\sum_{j=0}^{\infty}\frac{1}{\alpha}\frac{(\frac{1}{2}-t)_j(\alpha+\frac{1}{2}+t)_j}{(\frac{3}{2})_j(\alpha+1)_j}\rho_j(\gamma)\,, \tag{5.8}$$

where

$$\rho_j(\gamma)=\int_0^1 (1-\gamma^2\mu^2)^{j+1/2}\,d\mu={}_2F_1\left(\begin{matrix}-j-\frac{1}{2},\ \frac{1}{2}\\ \frac{3}{2}\end{matrix}\middle|\gamma^2\right). \tag{5.9}$$

The use of some linear transformations for ${}_2F_1$ results in

$$\rho_j(\gamma)=\frac{\Gamma(\frac{3}{2})\Gamma(j+\frac{3}{2})}{\Gamma(j+2)\gamma}-\frac{(1-\gamma^2)^{j+3/2}}{(2j+3)\gamma^{2j+4}}{}_2F_1\left(\begin{matrix}j+2,\ j+\frac{3}{2}\\ j+\frac{5}{2}\end{matrix}\middle|-\frac{1-\gamma^2}{\gamma^2}\right). \tag{5.10}$$

The last hypergeometric function can however be expressed in terms of elementary functions as follows

$${}_2F_1\left(\begin{matrix}j+2,\ j+\frac{3}{2}\\ j+\frac{5}{2}\end{matrix}\middle|-x\right)=\frac{(-1)^{j+1}(\frac{3}{2})_j}{(1)_j(\frac{1}{2})_j}\phi_{j+1}(x)\,, \tag{5.11}$$

$$\phi_{j+1}(x)=\frac{d^{j+1}}{dx^{j+1}}\frac{\arctan\sqrt{x}}{\sqrt{x}}\,. \tag{5.12}$$

As can be proved by induction $\phi_{j+1}(x)$ can be written explicitly as

$$\phi_{j+1}(x)=(-1)^{j+1}(\tfrac{1}{2})_{j+1}\frac{\arctan\sqrt{x}}{x^{j+3/2}}+\frac{(-1)^j}{2}\sum_{p=0}^{j}\frac{(\frac{3}{2})_j p!}{(\frac{3}{2})_p x^{j+1-p}(1+x)^{p+1}}\,. \tag{5.13}$$

This last equation implies, after some manipulation

$$\rho_j(\gamma)=\frac{(\frac{3}{2})_j}{2(2)_j}\left(\frac{\arcsin\gamma}{\gamma}+\sum_{p=0}^{j}\frac{p!}{(\frac{3}{2})_p}(1-\gamma^2)^{p+1/2}\right). \tag{5.14}$$

Therefore from Eq. (5.8) we conclude

$$\Omega_\alpha^{(2)}(\gamma)=\frac{1}{(1+2t)(\alpha+t-\frac{1}{2})}\left[1-{}_2F_1\left(\begin{matrix}-t-\frac{1}{2},\ \alpha+t-\frac{1}{2}\\ \alpha\end{matrix}\middle|1\right)\right]\frac{\arcsin\gamma}{\gamma}+\frac{1}{2\alpha}\sum_{p=0}^{\infty}\frac{(\frac{1}{2}-t)_p(\alpha+t+\frac{1}{2})_p}{(\alpha+1)_p(\frac{3}{2})_p(p+1)!}\,{}_3F_2\left(\begin{matrix}\frac{1}{2}-t+p,\ \alpha+\frac{1}{2}+t+p,\ 1\\ p+2,\ \alpha+p+1\end{matrix}\middle|1\right)(1-\gamma^2)^{p+1/2}\,. \tag{5.15}$$

By using transformation formulas [Ref. 11, pg. 111, Eq. (41)] we can write

$${}_3F_2\left(\begin{matrix}\frac{1}{2}-t+p,\ \alpha+t+p+\frac{1}{2},\ 1\\ p+2,\ \alpha+p+1\end{matrix}\middle|1\right)=\frac{(p+1)(\alpha+p)}{(\frac{1}{2}+t)(\alpha-\frac{1}{2}+t)}-\frac{(1)_p(p+1)(\alpha+1)_p\Gamma(\alpha+1)}{\Gamma(\frac{1}{2}-t)(\frac{1}{2}-t)_p(\frac{1}{2}+t)(\alpha-t-\frac{1}{2})(\alpha+t+\frac{1}{2})_p\Gamma(\alpha+t+\frac{1}{2})}\,. \tag{5.16}$$

However, this equation and the value of ${}_2F_1$ for unit argument make it possible to write the following result after expressing ${}_2F_1$ in terms of elementary functions:

$$\Omega_\alpha^{(2)}(\gamma)=\frac{1}{(1+2t)(\alpha+t-\frac{1}{2})}\left[\frac{\arcsin\gamma}{\gamma}-\frac{\pi\Gamma(\gamma)}{2\Gamma(\alpha+t+\frac{1}{2})\Gamma(\frac{1}{2}-t)}\frac{1}{\gamma}+\sqrt{1-\gamma^2}\,{}_3F_2\left(\begin{matrix}\frac{1}{2}-t,\ \alpha+t+\frac{1}{2},\ 1\\ \alpha,\ \frac{3}{2}\end{matrix}\middle|1-\gamma^2\right)\right]. \tag{5.17}$$

The use of values of $\Omega_\alpha^{(1)}$ and $\Omega_\alpha^{(2)}$ in the definition $\Lambda_k^\alpha(\gamma)$ leads us to the following result:

$$\Lambda_k^\alpha(\gamma)=\frac{2^{4+k}\pi^{2\alpha+1/2}}{\Gamma(\alpha)\Gamma(\alpha+\frac{1}{2})}\left\{\frac{d^k}{dt^k}\frac{\Gamma(\alpha+\frac{1}{2})\sqrt{\pi}\,(t+\frac{1}{2}\alpha)}{\Gamma(\alpha+1)(t+\frac{1}{2})(\alpha+t-\frac{1}{2})}\right.$$
$$\times\left[-\sqrt{1-\gamma^2}\,{}_3F_2\left(\begin{matrix}\frac{1}{2}-t,\ \alpha+\frac{1}{2}+t,\ 1\\ \alpha,\ \frac{3}{2}\end{matrix}\middle|1-\gamma^2\right)\right.$$
$$-\frac{\arcsin\gamma}{\gamma}-\frac{\pi\Gamma(\alpha)\Gamma(\alpha+t)\Gamma(-t)}{\Gamma(\alpha-\frac{1}{2})\Gamma(-\frac{1}{2})\Gamma(\frac{1}{2}-t)\Gamma(\alpha+\frac{1}{2}+t)}$$
$$\left.\left.\times{}_2F_1\left(\begin{matrix}\alpha+t,\ -t\\ \alpha-\frac{1}{2}\end{matrix}\middle|1-\gamma^2\right)\right]+\frac{\Gamma(\alpha)\sqrt{\pi}}{\Gamma(\alpha+\frac{1}{2})}\frac{1}{2}\left(\frac{1}{t}+\frac{1}{t+\alpha}\right)\right\}_{t=0}. \tag{5.18}$$

To check the last relation let us put $\alpha=2$, $\gamma=1$, and $k=0$. We obtain

$$\Lambda_0^2(1)=\frac{64\pi^4}{3}\left(\frac{23}{9}-\frac{\pi^2}{4}\right), \tag{5.19}$$

and this yields the result

$$\chi_1(1\,|\,u_1,u_2)=\frac{4}{3}-\frac{368}{27\pi^2} \tag{5.20}$$

which coincides with the result in the preceding paper.

VI. SPECIALIZATION OF $\Lambda_k^\alpha(\gamma)$ FOR AN ODD NUMBER OF PARTICLES

For an odd number of particles, α is an integer and this makes it possible to reduce the hypergeometric functions in Eq. (5.18) to more amenable forms. If we write $\alpha=n+1$, a careful investigation shows that

$$ {}_3F_2\left(\begin{matrix}\frac{1}{2}-t,\, n+\frac{3}{2}+t,\, 1\\ n+1,\, \frac{3}{2}\end{matrix}\middle| x\right)=\frac{(1)_n(\frac{3}{2}-n)_n}{(t+\frac{3}{2})_n(\frac{1}{2}-t-n)_n}\cdot\frac{1}{\cdot^n} \times\left[{}_2F_1\left(\begin{matrix}\frac{1}{2}-t-n,\, \frac{3}{2}+t\\ \frac{3}{2}-n\end{matrix}\middle| x\right) -\sum_{k=0}^{n-1}\frac{(\frac{1}{2}-t-n)_k(\frac{3}{2}+t)_k x^k}{(\frac{3}{2}-n)_k(1)_k}\right] \tag{6.1} $$

and

$$ {}_2F_1\left(\begin{matrix}\frac{1}{2}-t-n,\, \frac{3}{2}+t\\ \frac{3}{2}-n\end{matrix}\middle| x\right)=\frac{(1-x)^{-t}x^{n-1/2}}{(\frac{3}{2}-n)_n} \frac{d^n}{dx^n}\left[x^{1/2}(1-x)^{n+t}\,{}_2F_1\left(\begin{matrix}\frac{1}{2}-t,\, \frac{3}{2}+t\\ \frac{3}{2}\end{matrix}\middle| x\right)\right]. \tag{6.2} $$

However the hypergeometric function in the right-hand side of the last equation can be expressed in terms of elementary functions as follows:

$$ {}_2F_1\left(\begin{matrix}\frac{1}{2}-t,\, \frac{3}{2}+t\\ \frac{3}{2}\end{matrix}\middle| x\right)=\frac{1}{(2t+1)} \left(\frac{\sin(2t\arcsin\sqrt{x})}{\sqrt{x}}+\frac{\cos(2t\arcsin\sqrt{x})}{\sqrt{1-x}}\right). \tag{6.3} $$

These three equations and the Taylor expansion into powers of t makes it possible to evaluate the ${}_3F_2$ appearing in Eq. (5.18) in terms of elementary functions.

Another investigation of the hypergeometric functions shows that

$$ {}_2F_1\left(\begin{matrix}n+1+t,\, -t\\ n+\frac{1}{2}\end{matrix}\middle| x\right)=\frac{(-1)^n(\frac{1}{2})_n}{(n+t)_n(\frac{1}{2}+t)_n}\frac{1}{(1-x)^t}\frac{d^n}{dx^n} \times\left((1-x)^{n+t-1}\frac{\cos[(2t+1)\arcsin\sqrt{x}]}{\sqrt{1-x}}\right). \tag{6.4} $$

This completes the specialization of Λ_k^α for integer values of α due to the fact that all other functions of t appearing in Eq. (5.18) can be expressed in terms of elementary functions by using the properties of the Gamma function.

In the case of half-integer values of α, the Γ functions do not create great difficulties but the hypergeometric functions, especially the ${}_3F_2$ cause a problem. To show this, let us consider the simplest case $k=0, t=0$ in Eq. (5.18). By inserting $\alpha=n+\frac{3}{2}$ and employing the following property satisfied by Pochhammer's symbols

$$ (\tfrac{1}{2})_k(n+2)_k=\frac{(n+1)_k(\frac{3}{2})_k}{(2n+2)}+\frac{n+\frac{1}{2}}{n+1}(\tfrac{1}{2})_k(n+1)_k\,. \tag{6.5} $$

We can write

$$ {}_3F_2\left(\begin{matrix}\frac{1}{2},\, n+2,\, 1\\ \frac{3}{2},\, n+\frac{3}{2}\end{matrix}\middle| x\right)=\frac{1}{2(n+1)}{}_2F_1\left(\begin{matrix}n+1,\, 1\\ n+\frac{3}{2}\end{matrix}\middle| x\right)+\frac{(n+\frac{1}{2})}{(n+1)} {}_3F_2\left(\begin{matrix}\frac{1}{2},\, n+1,\, 1\\ \frac{3}{2},\, n+\frac{3}{2}\end{matrix}\middle| x\right), \tag{6.6} $$

and this implies

$$ {}_3F_2\left(\begin{matrix}\frac{1}{2},\, n+2,\, 1\\ \frac{3}{2},\, n+\frac{3}{2}\end{matrix}\middle| x\right)=\frac{1}{2}\sum_{k=0}^{n}\frac{(n+\frac{3}{2}-k)_k}{(n+1-k)_{k+1}}{}_2F_1\left(\begin{matrix}n+1-k,\, 1\\ n+\frac{3}{2}\end{matrix}\middle| x\right) +\frac{(\frac{1}{2})_{n+1}}{(2)_n}{}_3F_2\left(\begin{matrix}\frac{1}{2},\, 1,\, 1\\ \frac{3}{2},\, n+\frac{3}{2}\end{matrix}\middle| x\right). \tag{6.7} $$

The use of some properties of the generalized hypergeometric functions shows that [Ref. 11, pg. 111, Eq. (38)]

$$ {}_3F_2\left(\begin{matrix}\frac{1}{2},\, 1,\, 1\\ \frac{3}{2},\, n+\frac{3}{2}\end{matrix}\middle| x\right)=-\frac{1}{2}\sum_{k=0}^{n-1}\frac{(n+\frac{3}{2}-k)_k}{(n-k)_{n+1}}{}_2F_1\left(\begin{matrix}1,\, 1\\ n+\frac{3}{2}-k\end{matrix}\middle| x\right) +\frac{(\frac{3}{2})_n}{(1)_n}{}_3F_2\left(\begin{matrix}1,\, 1,\, \frac{1}{2}\\ \frac{3}{2},\, \frac{3}{2}\end{matrix}\middle| x\right). \tag{6.8} $$

All the Gaussian hypergeometric functions can be written in more elementary form, but the last hypergeometric function in the right-hand side of the Eq. (6.6) cannot be reduced any more. However, a rigorous investigation shows that it can be written in terms of the dilogarithm function[12] with a complex and complicated argument. Therefore it seems to be expressible at least in terms of tabulated functions.

However we must keep in mind that this is the simplest case, possibly for higher values of k the reduction of the ${}_3F_2$ in the Eq. (5.18) to elementary, or at least to tabulated functions will become more difficult. Therefore, we shall not discuss this point any further in this paper.

VII. CONCLUSIONS AND REMARKS

In this paper we have presented the evaluation of the simplest kind of hyperangular interaction integral as a function of γ, for a system of electrically charged particles. As we recall, γ is a scalar product of u_1 with u_2. However these vectors are eigenvectors of the potential matrices A_1, A_2. Therefore γ characterizes the angle between hyperaxes of the potentials. We have shown that they can be expressed in elementary terms.

Obviously in the case of such integrals of more complicated type, the number of γ's will increase and the evaluation of the τ type of integrals will become more difficult. Without obtaining some integral representation of these last ones, it seems to be very difficult to evaluate the Λ kind of integrals analytically. Our studies are continuing for this purpose.

An interesting point is that nature discriminates systems with an odd number of particles from systems with an even number. Indeed from a mathematical point of view for an odd number of particles the function Λ is expressible in terms of elementary functions such as $\arcsin x$ and algebraic functions. But in the other case more complicated functions such as the dilogarithm function appear.

Of course another interesting point is to investigate the Λ functions, for all complex values of γ, α, and even for k. But this is important only in the mathematical sense, it seems that there is no physical reason which necessitates the investigation of this point.

A careful look at the perturbative scheme presented in our previous publication shows that all of the different order perturbative terms will involve this kind of hyperangular interaction integrals. Therefore we have the possibility of evaluating some of these terms analytically. This encourages us to deal with the details of the more complicated cases.

[1]M. Demiralp, J. Chem. Phys. **72**, 3828–3836 (1980).
[2]N. A. Baykara and M. Demiralp, "Numerical evaluation of Coulombic hyperangular interaction integrals" (to be published).
[3]Upper index of $\theta_j^{(k)}$ represents the position of this hyperspherical harmonic in a conveniently choosen ordering (e.g., refer to M. Demiralp and E. Şuhubi, J. Math. Phys. **18**, 777–785 (1977).
[4]H. Hochstadt, *The Functions of Mathematical Physics* (Wiley-Interscience, New York, 1971).
[5]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), Vol. 1.
[6]A. Erdelyi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. II.
[7]θ_0 is a constant function of its arguments, this can be verified by looking at its definition which can be found in literature (e.g., see the reference given in the third article above).
[8]Pochhammer's symbol is defined as $(a)_k = \Gamma(a+k)/\Gamma(a)$.
[9]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).
[10]P. Lancaster, *Theory of Matrices* (Academic, New York, 1969).
[11]Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. I.
[12]M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Dover, New York, 1970), 9th printing, p. 1004.

Solution of Poisson's equation: Beyond Ewald-type methods

M. Weinert
Department of Physics and Astronomy and Materials Research Center, Northwestern University, Evanston, Illinois 60201

(Received 29 July 1980; accepted for publication 28 August 1980)

A general method for solving Poisson's equation without shape approximation for an arbitrary periodic charge distribution is presented. The method is based on the concept of multipole potentials and the boundary value problem for a sphere. In contrast to the usual Ewald-type methods, this method has only absolutely and uniformly convergent reciprocal space sums, and treats all components of the charge density equivalently. Applications to band structure calculations and lattice summations are also discussed.

PACS numbers: 02.30.Hq, 02.30.Mv, 71.25.Cx

I. INTRODUCTION

The solution of Poisson's equation for a periodic charge density is, in principle, straightforward. The practical aspects of the problem are of great importance in many areas of solid state physics, particularly in band structure calculations. In the past few years, different methods for solving Poisson's equation have been developed.[1] The basic approach has been to first consider the short-range Coulomb interactions arising from neutral charge distributions and then to account for the long-range interactions by lattice summations.

The problem of lattice sums was first considered by Madelung,[2] who obtained the self-potential of a lattice. A general method for treating these sums was developed by Ewald.[3] Since then, there have been a number of modifications and generalizations of this method.[4–8]

In this paper, we propose a new method using the concept of multipole potentials and the Dirichlet problem for a sphere in which all contributions are treated equivalently. Basing a method on these ideas was suggested by Hamann,[9,10] but the general formulation is new. The rest of the paper is organized as follows: The general considerations are given in Sec. II. In Sec. III, the Fourier representation of the charge density (and the potential) is derived, and then the properties of the Fourier expansion are discussed in Sec. IV. Extensions and applications are given in Secs. V and VI, and at the end are some mathematical appendices.

II. GENERAL CONSIDERATIONS

Consider a charge distribution $\rho(\mathbf{r})$ localized inside a sphere S. The potential at a point outside is given by the multipole expansion[11]

$$V(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} q_{lm} \frac{Y_{lm}(\hat{r})}{r^{l+1}}, \tag{1}$$

where the multipole moments q_{lm} are given by

$$q_{lm} = \int_S Y^*_{lm}(\hat{r}) r^l \rho(\mathbf{r})\, d^3r. \tag{2}$$

The actual form of the charge density is immaterial; the same potential (outside S) could be obtained from any one of an infinite number of charge distributions that have the correct multipole moments. This arbitrariness gives us the freedom to replace the charge density inside the sphere with another more convenient one.

If we have a periodic charge density, e.g., a crystal (the discussion will be in terms of a crystal), we can divide up the charge into that inside the spheres (centered on atomic sites) and that in the interstitial region (see Fig. 1):

$$\rho(\mathbf{r}) = \rho_I(\mathbf{r})\theta(\mathbf{r}\in I) + \sum_{i \text{ spheres}} \rho_i(\mathbf{r})\theta(\mathbf{r}\in S_i), \tag{3}$$

where $\rho_I(\mathbf{r})(\rho_i(\mathbf{r}))$ is the charge in the interstial (ith sphere) and where the unit step function θ enforces this division. (I refers to the interstitial and S_i refers to the ith sphere.)

In general, the charge density given by (3) will have a slowly convergent Fourier expansion because of the large oscillations near the nuclei, whereas the interstitial charge density is fairly smooth and can be continued into the spheres in such a way that it has a rapidly convergent Fourier expansion. We can make use of this observation and solve the problem of obtaining the Coulomb potential in two steps: 1) obtain the potential in the interstitial and then 2) solve the boundary value problem inside the sphere.

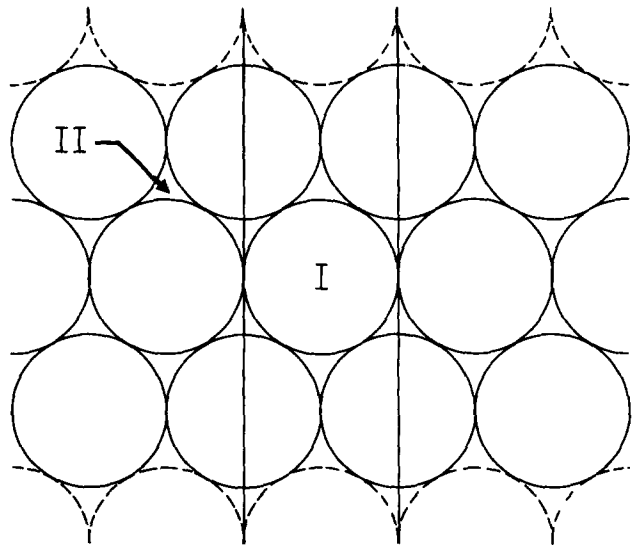


FIG. 1. Division of the unit cell into spheres (I) and into the interstitial (II). For a 2-D system, e.g., a film, the dotted spheres are replaced by vacuum and the unit cell is defined by the lines extending to infinity.

 0022-2488/81/112433-07$01.00

Since the potential in the interstitial depends only on the interstitial charge density and on the multipole moments of the spheres, we can replace the real charge density by a pseudocharge density

$$\rho(\mathbf{r}) \to \tilde{\rho}(\mathbf{r}) = \rho_I(\mathbf{r})\theta(\mathbf{r}\in I) + \sum_{\text{spheres}} \tilde{\rho}_i(\mathbf{r})\theta(\mathbf{r}\in S_i), \tag{4}$$

with the requirement that the pseudocharge density in the spheres have the same multipole moments as the original charge. This pseudocharge density will give the correct interstitial potential, but not the correct potential in the spheres. Let us assume that we can express (4) as a rapidly convergent Fourier series (we will come back to this point in the next section):

$$\tilde{\rho}(\mathbf{r}) = \sum_{\mathbf{K}} \tilde{\rho}(\mathbf{K}) e^{i\mathbf{K}\cdot\mathbf{r}}. \tag{5}$$

From this Fourier series, the interstitial potential is given by

$$V_I(\mathbf{r}) = \sum_{\mathbf{K}\neq 0} \frac{4\pi\tilde{\rho}(\mathbf{K})}{K^2} e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{6}$$

where $\mathbf{K}$ is a reciprocal lattice vector (dual space vector). The potential $V_I(\mathbf{r})$ is also correct on the sphere boundaries, hence the potential inside the sphere is a boundary value problem. The solution to this Green's function problem is[11] ($\mathbf{r}_i$ is the position relative to the ith nucleus)

$$V_i(\mathbf{r}_i) = \int_{S_i} \rho_i(\mathbf{r}')G(\mathbf{r}_i,\mathbf{r}')\,d^3r' - \frac{R_i^2}{4\pi}\oint_{S_i} V_I(\mathbf{R}_i')\frac{\partial G}{\partial n'}\,d\Omega', \tag{7}$$

where $\mathbf{R}_i$ is a point on the sphere and the Green's function G is given by

$$G(\mathbf{r}_i,\mathbf{r}') = 4\pi\sum_{lm}\frac{Y_{lm}^*(\hat{r}')Y_{lm}(\hat{r}_i)}{2l+1}\frac{r_<^l}{r_>^{l+1}}\left[1-\left(\frac{r_>}{R_i}\right)^{2l+1}\right], \tag{8}$$

where $r_>$ ($r_<$) is the greater (smaller) of r_i and r' and the normal derivative is

$$\frac{\partial G}{\partial n'} = \left.\frac{\partial G}{\partial r'}\right|_{r'=R_i} = -\frac{4\pi}{R_i^2}\sum_{lm}\left(\frac{r_i}{R_i}\right)^l Y_{lm}^*(\hat{r}')Y_{lm}(\hat{r}). \tag{9}$$

Note that since we are using the actual charge density in (7), we now have the correct potential everywhere. Equations (5)–(7) represent a complete solution to Possion's equation. We now turn to the critical step of obtaining the Fourier expansion of the pseudocharge density. Unless we can obtain a rapidly convergent expansion, this method will not be useful.

III. FOURIER REPRESENTATION OF THE PSEUDOCHARGE DENSITY

The pseudocharge density will have a dual representation: the interstitial charge is smooth and hence amenable to a Fourier expansion (valid in the interstitial), and the charge in the spheres will be in a spherical harmonic representation, which is the natural representation near an atomic site. In the spirit of the Ewald method (and using the linearity of Poisson's equation), we extend the interstitial Fourier representation over all space and subtract it off again inside the spheres, i.e., if the interstitial charge has the expansion

$$\rho_I(\mathbf{r}) = \sum_{\mathbf{K}} \rho_I(\mathbf{K})e^{i\mathbf{K}\cdot\mathbf{R}} \tag{10}$$

valid in the interstitial, then the charge density (3) is

$$\rho(\mathbf{r}) = \rho_I(\mathbf{r}) + \sum_{\text{spheres}} [\rho_i(\mathbf{r}) - \rho_I(\mathbf{r})]\,\theta(\mathbf{r}\in S_i), \tag{11}$$

where $\rho_I(\mathbf{r})$ is defined over all space by the Fourier representation (10) which converges to the actual charge density in the interstitial.

Let the multipole moments of the real charge density in the ith sphere be q_{lm}^i. The multipole moments of the plane-wave charge density (10) in the sphere located at ξ_i must be obtained in order to get the multipole moments of the total charge (11) in the spheres. First write (10) relative to ξ_i ($\mathbf{r}_i = \mathbf{r} - \xi_i$),

$$\rho_I(\mathbf{r}) = \sum_{\mathbf{K}} \rho_I(\mathbf{K})\, e^{i\mathbf{K}\cdot\xi_i}\, e^{i\mathbf{K}\cdot\mathbf{r}_i}. \tag{12}$$

Then substituting (12) into the definition of the multipole moments and using the well-known expansion for a plane-wave[11]

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{lm} 4\pi i^l j_l(Kr)Y_{lm}^*(\mathbf{K})Y_{lm}(\mathbf{r}) \tag{13}$$

(where the $j_l(Kr)$ are the spherical Bessel functions), then the multipole moments of the plane-wave charge in the ith sphere are given by

$$q_{lm}^{Ii} = \frac{\sqrt{4\pi}}{3}R_i^3\rho_I^{(\mathbf{K}=0)}\delta_{l0} + \sum_{K\neq 0} 4\pi i^l \rho_I(\mathbf{K})R_i^{l+3} \times \frac{j_{l+1}(KR_i)}{KR_i}e^{i\mathbf{K}\cdot\xi_i}Y_{lm}^*(\mathbf{K}), \tag{14}$$

where R_i is the sphere radius of the ith sphere. Finally, the multipole moments of the charge density (11) inside the ith sphere are given by

$$\tilde{q}_{lm}^i = q_{lm}^i - q_{lm}^{Ii}. \tag{15}$$

It is these multipole moments that we must consider in creating the pseudocharge density in the spheres.

Let us replace the real charge density inside the spheres by a pseudocharge density given in a spherical harmonic representation where the radial part is expanded in a power series:

$$\tilde{\rho}_i(\mathbf{r}_i) = \sum_{lm} Q_{lm}^i Y_{lm}(\hat{r}_i)\sum_\eta a_\eta r_i^{\nu_\eta}, \tag{16}$$

where a_η and ν_η are parameters for the power series and the Q_{lm}^i are constants such that the pseudocharge has the correct multiple moments. This requirement on the Q_{lm}^i is

$$\tilde{q}_{lm}^i = \sum_{l'm'} Q_{l'm'}^i \int_{S_i} Y_{lm}^*(\hat{r}_i)Y_{l'm'}(\hat{r}_i)\,d\Omega_i \int_0^{R_i}\sum_\eta a_\eta r_i^{\nu_\eta+l+2}\,dr_i$$

or

$$Q_{lm}^i = \tilde{q}_{lm}^i\left[\sum_\eta a_\eta \frac{R_i^{l+\nu_\eta+3}}{l+\nu_\eta+3}\right]^{-1}. \tag{17}$$

We need to obtain the Fourier expansion of the pseudo-

charge given by (16) inside the spheres and zero in the interstitial:

$$\tilde{\rho}_s(\mathbf{K}) = \frac{1}{\Omega}\int\left[\sum_{\text{spheres}}\tilde{\rho}_i(\mathbf{r}_i)\right]e^{-i\mathbf{K}\cdot\mathbf{r}}\,d^3r$$
$$= \frac{1}{\Omega}\sum_{\text{spheres}} e^{-i\mathbf{K}\cdot\boldsymbol{\xi}_i}\int_{S_i}\tilde{\rho}_i(\mathbf{r}_i)\,e^{-i\mathbf{K}\cdot\mathbf{r}_i}\,d^3r_i, \tag{18}$$

where Ω is the volume of the unit cell. Then using (16) in (18), we have

$$\tilde{\rho}_s(\mathbf{K}) = \frac{1}{\Omega}\sum_{lmi} Q^i_{lm}\,e^{-i\mathbf{K}\cdot\boldsymbol{\xi}_i}$$
$$\times\sum_{\eta} a_\eta\int_{S_i} Y_{lm}(\hat{r}_i)r^{\nu_\eta}e^{-i\mathbf{K}\cdot\mathbf{r}_i}\,d^3r_i.$$

After expanding the exponential as before, this becomes

$$\tilde{\rho}_s(\mathbf{K}) = \frac{1}{\Omega}\sum_{lmi} Q^i_{lm}\,e^{-i\mathbf{K}\cdot\boldsymbol{\xi}_i}\,4\pi(-i)^l Y_{lm}(\mathbf{K})$$
$$\times\sum_{\eta}\frac{a_\eta}{K^{\nu_\eta+3}}\int_0^{KR_i} t^{\nu_\eta+2}j_l(t)\,dt. \tag{19}$$

If we let $\nu_\eta = l+2\eta$ ($\eta = 0,1,2,\cdots$), then the integral in (19) can be expressed as a finite sum of spherical Bessel functions

$$\frac{1}{K^{l+2\eta+3}}\int_0^{KR_i} t^{l+2\eta+2}j_l(t)\,dt$$
$$= \frac{R_i^{l+2\eta+2}}{K}\sum_{\nu=0}^{\eta}\frac{(-1)^\nu 2^\nu\eta!}{(\eta-\nu)!}\frac{j_{l+1+\nu}^{(KR_i)}}{(KR_i)^\nu}. \tag{20}$$

If we define

$$A^i_l \equiv \sum_{\eta=0}^{n}\frac{a_\eta}{K^{l+2\eta+3}}\int_0^{KR_i} t^{l+2\eta+2}j_l(t)\,dt, \tag{21}$$

then using (20) in (21), we have

$$A^i_l = \sum_{\eta=0}^{n}\frac{a_\eta R_i^{l+2\eta+3}}{KR_i}\sum_{\nu=0}^{\eta}\frac{(-1)^\nu 2^\nu\eta!}{(\eta-\nu)!}\frac{j_{l+1+\nu}^{(KR_i)}}{(KR_i)^\nu}.$$

Since the sums are finite, we can rearrange them into sums of like powers of KR_i

$$A^i_l = R^{l+3}\sum_{\nu=0}^{n}(-1)^\nu 2^\nu\frac{j_{l+1+\nu}^{(KR_i)}}{(KR_i)^{\nu+1}}\sum_{\eta=\nu}^{n}\frac{a_\eta\eta!R_i^{2\eta}}{(\eta-\nu)!}. \tag{22}$$

For $\nu = n$, the sum over η reduces to $a_n n!R^{2n}$. For all other values of ν, let us require

$$\sum_{\eta=\nu}^{n}\frac{a_\eta\eta!R_i^{2\eta}}{(\eta-\nu)!} = 0, \quad \nu = 0,1,\ldots,n-1. \tag{23}$$

This set of equations has the solution (see Appendix B)

$$a_\eta = (-1)^{n-\eta}R_i^{2(n-\eta)}\binom{n}{\eta}a_n. \tag{24}$$

Let us consider what condition (23) means. The radial part of the pseudocharge density for a given l is proportional to [cf. (16)]

$$\tilde{\rho}_i(r_i)\sim\sum_{\eta=0}^{n}a_\eta r_i^{l+2\eta}$$
$$= (-1)^n a_n R_i^{2n}r_i^l\sum_{\eta=0}^{n}(-1)^\eta\binom{n}{\eta}\left(\frac{r_i}{R_i}\right)^{2\eta}.$$

The k th radial derivative evaluated at the sphere surface is

$$\left.\frac{\partial^k}{\partial r_i^k}\tilde{\rho}_i(r_i)\right|_{r_i=R_i}\sim\sum_{\eta=0}^{n}(-1)^\eta\frac{(2\eta+l)!}{(2\eta+l-k)!}\binom{n}{\eta}.$$

The factor $(2\eta+l)!/(2\eta+l-k)!$ is a polynomial in η of degree k, hence can write

$$\left.\frac{\partial^k}{\partial r_i^k}\tilde{\rho}_i(r_i)\right|_{r_i=R_i}\sim\sum_{p=0}^{k}\alpha_p\sum_{\eta=0}^{n}(-1)^\eta\eta^p\binom{n}{\eta},$$

where the α_p are the coefficients of the polynomial. Since this is zero for all $k<n-1$ (cf. Appendix A), the first $n-1$ derivatives of the pseudocharge density at the surface of the sphere are continuous. Hence, our choice of n is, in fact, a choice of how smooth a function we must Fourier analyze.

Now let us return to the evaluation of A^i_l. Using the values of a_η given in (24), the sum has the value

$$A^i_l = (-1)^n 2^n R_i^{l+3}\frac{j_{l+n+1}^{(KR_i)}}{(KR_i)^{n+1}}(a_n n!R^{2n}). \tag{25}$$

The pseudocharge multipole moments Q^i_{lm} (17) also depend on the a_η. Consider the sum

$$S^i_n = \sum_{\eta=0}^{n}a_\eta\frac{R_i^{2l+2\eta+3}}{2l+2\eta+3}.$$

We write this as

$$S^i_n = (-1)^n a_n R_i^{2l+2n+3}\sum_{\eta=0}^{n}\frac{(-1)^\eta\binom{n}{\eta}}{2l+2\eta+3},$$

or finally (see Appendix A)

$$S^i_n = (-1)^n a_n R_i^{2l+2n+3}2^n n!\frac{(2l+1)!!}{(2l+2n+3)!!}. \tag{26}$$

The Q^i_{lm} are then given by

$$Q^i_{lm} = \frac{(-1)^n\tilde{q}^i_{lm}}{2^n n!a_n R^{2l+2n+3}}\frac{(2l+2n+3)!!}{(2l+1)!!}. \tag{27}$$

Combining (27) and (25) with (19), we get the Fourier component

$$\tilde{\rho}_s(\mathbf{K}) = \frac{4\pi}{\Omega}\sum_{lmi}(-i)^l\left\{\frac{(-1)^n\tilde{q}^i_{lm}}{2^n n!a_n R_i^{2l+2n+3}}\frac{(2l+2n+3)!!}{(2l+1)!!}\right\}$$
$$\times\left\{(-1)^n 2^n n!a_n R_i^{l+2n+3}\frac{j_{l+n+1}^{(KR_i)}}{(KR_i)^{n+1}}\right\}e^{-i\mathbf{K}\cdot\boldsymbol{\xi}}Y_{lm}(\mathbf{K}),$$

of after simplifying,

$$\tilde{\rho}_s(\mathbf{K}) = \frac{4\pi}{\Omega}\sum_{lmi}\frac{(-i)^l(2l+2n+3)!!}{(2l+1)!!}$$
$$\times\frac{j_{l+n+1}^{(KR_i)}}{(KR_i)^{n+1}}\tilde{q}^i_{lm}e^{-i\mathbf{K}\cdot\boldsymbol{\xi}}Y_{lm}(\mathbf{K}). \tag{28}$$

For the case $\mathbf{K} = 0$, we have

$$\tilde{\rho}_s(\mathbf{K}=0) = \frac{\sqrt{4\pi}}{\Omega}\sum_i\tilde{q}^i_{00}. \tag{29}$$

The Fourier expansion for the total pseudocharge used to obtain the interstitial Coulomb potential is

$$\tilde{\rho}(\mathbf{r}) = \sum_{\mathbf{K}}[\rho_I(\mathbf{K})+\tilde{\rho}_s(\mathbf{K})]\,e^{i\mathbf{K}\cdot\mathbf{r}}. \tag{30}$$

A point to notice about the parameter n in (28) is that a different value of n can be picked for each l value and also for each sphere.

IV. PROPERTIES OF THE FOURIER EXPANSION

In this section we investigate some properties of the Fourier expansion of the (pseudo-) charge density. The first important point is that the Fourier series

$$\tilde{\rho}_s(\mathbf{r}) = \sum_{\mathbf{K}} \tilde{\rho}_s(\mathbf{K}) e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{31}$$

where the $\tilde{\rho}_s(\mathbf{K})$ are given in (28), is an absolutely and uniformly convergent series for $n > 0$ which converges to $\tilde{\rho}_i(\mathbf{r})$ in each sphere and to zero in the interstitial.[12] This property allows one to do the term by term integration needed to obtain the potential. It is interesting to note that the potential for the $l = 0$ term is also absolutely and uniformly convergent, whereas in the usual Ewald-type methods[4,5,7] this term is only conditionally convergent.

The step of extending the interstitial representation into the spheres requires some discussion. This step guarantees the continuity of the pseudocharge density everywhere for $n \geqslant 1$. In practice, there will only be a finite number of multipole moments considered, while the plane waves that are extended into the spheres contain multipole moments of all orders. This means that certain (higher) multipoles from the plane waves are included in the potential, but are neglected in the contributions from the actual charge in the spheres. At first glance this appears to be a disadvantage, but actually this approximation is quite good. First, the contributions from the higher multipole moments decrease rapidly (as $r^{-(l+1)}$, $r \geqslant R$) and hence are small in general. The second point is that the major contribution to the higher multipole moments comes from the region near the surface where, because of continuity, one expects that the charge density is still well represented by the Fourier expansion of the interstitial. In this way, we implicitly include even higher multipole moments to a good approximation.

Instead of extending the plane-wave representation into the spheres, one could match the interstitial charge density to the pseudocharge in the spheres. This method was rejected for two main reasons: 1) the matching coefficients depend on the $\rho_i(\mathbf{K})$ and 2) the pseudocharge density is manifestly discontinuous. The first reason is mainly aesthetic and computational: The particularly simple form of (28) would be ruined and more computing effort would be required. The second reason is the important one: Since only a finite number of spherical harmonics are included in the expansion of the charge density in the spheres, it is impossible to match a general plane wave, which contains all spherical harmonics, onto the pseudocharge density at the sphere boundary. Hence, higher l components will be discontinuous, thereby hurting the convergence of the Fourier series since a 3-D step function converges as K^{-2}.

Let us now consider the convergence properties of the Fourier series with respect to $\mathbf{K}$. The convergence rate as $|\mathbf{K}| \to \infty$ goes as

$$\lim_{|\mathbf{K}|\to\infty} \frac{(2l+2n+3)!!}{(2l+1)!!} \frac{j_{l+n+1}^{(KR_i)}}{(KR_i)^{n+1}} \sim \frac{(2l+2n+3)!!}{(2l+1)!!} \frac{1}{(KR_i)^{n+2}}.$$

This would seem to suggest that we should pick n as large as possible. Unfortunately, the situation is more complicated. For large n, $(2l+2n+3)!!$ is large and the value of KR_i for which the spherical Bessel function can be replaced by its asymptotic value increases. The form for the factor $(2l+2n+3)!!j_{l+n+1}(z)/((2l+1)!!z^{n+1})$ for $l = 0$ and $l = 4$ for different values of n is given in Fig. 2. The important features to notice are: i) for smaller n, there are larger oscillations and ii) for a given n, the largest contribution comes from z less than the first zero of the Bessel function. From considerations such as these, we have developed a criterion for determining the best n value for each l given the maximum value of KR_i: Choose the largest n such that the first zero of $j_{l+n+1}(z)$ is approximately equal to $(KR_i)_{\max}$. Table I gives these values for $0 \leqslant l \leqslant 8$ and $(KR)_{\max} \lesssim 21$. Again, it should be noted that each l component may have a different n, hence in practice one has larger n values for lower l. This means that the convergence is best for the lower l values, but these are, in general, the most important.

With these properties, the convergence of the Fourier representation of the pseudo-charge density, and hence the potential, can be monitored rather easily. We have also avoided the problems associated with conditionally convergent sums.

V. EXTENSIONS

As developed so far, the method is for a system with three-dimensional periodicity. This is, however, not a requirement. It we have a real system with two-dimensional periodicity, e.g., a film or surface, we can define a unit cell as in Fig. 1 where the dotted spheres are replaced by vacuum. From the theory of Fourier series,[12] we can artifically repeat the film in order to obtain a Fourier representation of the charge density that converges to the (pseudo-) charge in the film. In this way, the Fourier representation of the pseudocharge density of a 2-D periodic system is exactly the same as for the 3-D case [Eq. (28)], with the appropriate definition of the unit cell. The difference will be that now the potential is not given simply by (6), but will also include terms such as

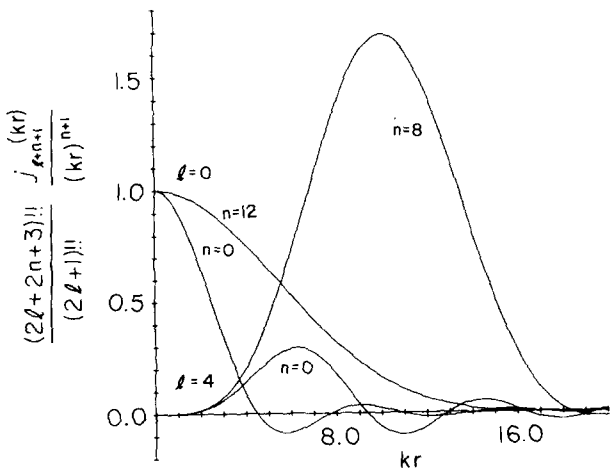


FIG. 2. Graph of the convergence factor of the pseudocharge density. The $l = 0$ terms ($n = 0, 12$) have the value unity at $kr = 0$, while the $l = 4$ terms ($n = 0, 8$) are zero there.

TABLE I. Optimum values of n for different values of l and $(KR)_{max}$ to obtain the best convergence possible of the pseudo-charge density and the interstitial potential.

ν	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
$z(j_\nu(z)=0)$	20.54	19.45	18.35	17.25	16.14	15.03	13.92	12.79	11.66	10.51	9.36	8.18	6.99	5.76	4.49
$l=0$	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	13	12	11	10	9	8	7	6	5	4	3	2	1	0	
2	12	11	10	9	8	7	6	5	4	3	2	1	0		
3	11	10	9	8	7	6	5	4	3	2	1	0			
4	10	9	8	7	6	5	4	3	2	1	0				
5	9	8	7	6	5	4	3	2	1	0					
6	8	7	6	5	4	3	2	1	0						
7	7	6	5	4	3	2	1	0							
8	6	5	4	3	2	1	0								

$$V_l'(\mathbf{r}) = \sum_{\mathbf{K}} \alpha(\mathbf{K})\tilde{\rho}(\mathbf{K}) e^{i\mathbf{K}\|\cdot\mathbf{r}\| \pm K\|z}. \tag{32}$$

In order to obtain the potential inside the spheres, we need to obtain this potential in a spherical harmonic representation. For a standard plane wave, we use expansion (13). For the case of the complex plane wave given in (32), we use the expansion (see Appendix C)

$$e^{i\mathbf{K}\|\mathbf{r}\| \pm K\|z} = \sum_{lm} (-1)^l (\mp 1)^{l-m} i^m (K\|r)^l \times \left[\frac{4\pi}{2l+1}\frac{1}{(l-m)!(l+m)!}\right]^{1/2} e^{-im\varphi_{\mathbf{K}}} Y_{lm}(\hat{r}). \tag{33}$$

Using this result, the rest proceeds straighforwardly.

Up to this point we have used only the translational symmetry. If we are considering a crystal, the symmetrized representations (symmetrized plane waves and lattice harmonics) are the natural ones to use. The extension of the method to these representations is simple and can often be written down by inspection.

VI. APPLICATIONS

The ultimate test of the method must be its usefulness. The method was programmed for the LAPW band structure method by Wimmer,[13] who compared the potential for a W(001) three-layer film calculated by this method and that using the old Ewald-type potential generator.[14] When only the $l=0$ term is included (which corresponds to the old method), the two potentials agrees to within a few mRyd everywhere. To obtain the full potential using this method, including exchange and correlation in the local density approximation and keeping multipoles up to $l=8$, the computing time was increased by not more than a few percent ($<5\%$). Hence, this is a reasonable method for doing band-structure calculations.

This method is also an efficient way to calculate lattice sums. As an example, let us calculate the Madelung potential for the NaCl structure. In this case, the Fourier coefficients depend only on the charges at each atomic site:

$$\tilde{\rho}_s(\mathbf{K}) = \frac{-(2n+3)!!}{\Omega}\frac{j_{n+1}(KR)}{(KR)^{n+1}}(1-e^{-i\mathbf{K}\cdot\boldsymbol{\eta}}),$$

where Cl^- is at the origin, Na^+ is at $\boldsymbol{\eta}$, and the sphere radii are equal. The coefficients depend on the vector property of $\mathbf{K}$ only through the phase factor. Expanding the plane-wave representation of the interstitial potential about the origin, the $l=0$ contribution is

$$V_{l=0}(\mathbf{r}) = \sqrt{4\pi}\left[\sum_{\mathbf{K}}{}' \frac{4\pi}{K^2}\tilde{\rho}_s(\mathbf{K}) j_0(KR)\right] Y_{00}(\hat{r}).$$

To obtain the Madelung constant, i.e., the self-potential of the lattice, we must subtract the potential due to the charge at the origin:

$$V'_{l=0}(\mathbf{r}) = \sqrt{4\pi}\left[\sum_{\mathbf{K}}{}' \frac{4\pi}{K^2}\tilde{\rho}_s(\mathbf{K}) j_0(KR) + \frac{1}{r}\right] Y_{00}(\mathbf{r}).$$

The potential inside the sphere is given by (7). Since there is no charge in the sphere, only the surface term contributes. Then at $\mathbf{r}=0$, the total potential is

$$V(\mathbf{r}=0) = \sum_{\mathbf{K}}{}' \frac{4\pi}{K^2}\tilde{\rho}_s(\mathbf{K}) j_0(KR) + \frac{1}{R}.$$

This is the Madelung constant. Table II presents results obtained for the NaCl structure for different n values. The number of terms included in the sum for each n value was determined by $(KR)_{max}$ using the values in Table I. The percent difference can be considered as limits on the results: For a given n, if the number of terms increase, then Δ is found to decrase. The convergence is quite good with increasing n, and supports the criterion for choosing n. The usefulness of doing lattice sums in this way has been demonstrated—extension to more general lattice sums can be handled in the same manner.

VII. SUMMARY

We have presented a general method for solving Poisson's equation in a periodic system. The method is based on the concept of multipole potentials and on the Dirichlet problem for a sphere. This method is not just an extension of the Ewald-type methods, but is a new alternative to them. Some advantages/properties of the method are:

1) there are reciprocal space sums only,
2) there are only absolutely and uniformly convergent sums (including the $l=0$ and $l=1$ terms),
3) general periodic charge densities are handled easily,
4) the convergence properties of the summations are easily monitored, and
5) all components are treated equivalently.

As in the Ewald method, we have one convergence parameter; ours is determined from a simple unique criterion. The

TABLE II. Madelung constant α for the NaCl structure calculated for different n values and compared to the exact result. Δ % is the absolute difference in percent and N_k is the number of distinct nonzero terms in the sum. Exact result from [C. Kittel, *Introduction to Solid State Physics*, 4th ed. (Wiley, New York, 1971), p. 118.]

n	α	N_k	Δ %
Exact	1.747565	...	...
14	1.747565	22	0.0
11	1.747569	15	0.0002
8	1.747517	10	0.003
5	1.748099	6	0.03
3	1.748214	4	0.04
1	1.735273	2	0.7

freedom of being able to choose different n values for each sphere and for each l to optimize the convergence is a freedom not available in the Ewald-type methods.

The application of this method to real problems has shown that the method is practical and can be used to advantage for treating those problems where the Ewald-type methods were used previously.

ACKNOWLEDGMENTS

I would like to thank E. Wimmer and H. Krakauer for helpful discussions and suggestions, A. J. Freeman for a critical reading of the manuscript, and D. R. Hamann for his suggestion as communicated to H. Krakauer. This work was supported by the National Science Foundation through the Northwestern University Materials Research Center (Grant No. DMR 76-80847).

APPENDIX A: SOME BINOMIAL SERIES IDENTITIES

1. Proof of $\sum_{n=0}^{m}(-1)^n(A+Bn)^p\binom{m}{n}=0, \quad p<m$

The binomial expansion is

$$(x+y)^m=\sum_{n=0}^{m}\binom{m}{n}x^n y^{m-n}, \tag{A1}$$

$$\binom{m}{n}=\frac{m!}{(m-n)!n!}.$$

Now consider the function

$$\left.\frac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})^m\right|_{x=0}.$$

If we do the differentiation by Leibniz's rule, then the k th term $(0\leqslant k\leqslant p)$ will be proportional to terms in x given by $(0\leqslant k'\leqslant k)$

$$\left[\frac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})^m\right]_{k,x=0}\sim\left.(1-e^{Bx})^{m-k'}\right|_{x=0}=0, \quad k'<m$$

If $p<m$, then all terms are zero, hence we can write

$$\left.\frac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})^m\right|_{x=0}=\begin{cases}\left.\dfrac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})\right|_{x=0}, & p\geqslant m\\ 0, & p<m\end{cases}. \tag{A2}$$

If we expand (A2) in a binomial series, then we have

$$\left.\frac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})^m\right|_{x=0}=\left.\frac{\partial^p}{\partial x^p}e^{Ax}\sum_{n=0}^{m}(-1)^n\binom{m}{n}e^{nBx}\right|_{x=0}=\sum_{n=0}^{m}(-1)^n\binom{m}{n}(A+Bn)^p. \tag{A3}$$

Then combining (A2) and (A3), we have

$$\sum_{n=0}^{m}(-1)^n\binom{m}{n}(A+Bn)^p=\begin{cases}\left.\dfrac{\partial^p}{\partial x^p}e^{Ax}(1-e^{Bx})^m\right|_{x=0}, & p\geqslant m\\ 0, & p<m\end{cases}. \tag{A4}$$

2. Proof of $\sum_{\nu=0}^{n}\frac{(-1)^\nu\binom{n}{\nu}}{2l+2\nu+3}=\frac{2^n n!(2l+1)!!}{(2l+2n+3)!!}=\frac{n!}{2}\frac{\Gamma(l+\frac{3}{2})}{\Gamma(l+n+\frac{5}{2})}$

Consider the integral

$$f=\int_0^1 t^{2(l+1)}(1-t^2)^n\,dt.$$

Expand the integral in a binomial series

$$f=\int_0^1 t^{2(l+1)}\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}t^{2\nu}\,dt=\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}\int_0^1 t^{2l+2\nu+2}\,dt.$$

Hence, f has the value of the sum we want:

$$\sum_{\nu=0}^{n}\frac{(-1)^\nu\binom{n}{\nu}}{2l+2\nu+3}=\int_0^1 t^{2(l+1)}(1-t^2)^n\,dt.$$

The integral is expressible as a beta function, hence we have

$$\sum_{\nu=0}^{n}\frac{(-1)^\nu\binom{n}{\nu}}{2l+2\nu+3}=\tfrac{1}{2}B\left(\frac{2l+3}{2},\frac{2n+2}{2}\right)=\frac{\frac{1}{2}\Gamma(l+\frac{3}{2})\Gamma(n+1)}{\Gamma(l+n+\frac{5}{2})}$$

or

$$\sum_{\nu=0}^{n}\frac{(-1)^\nu\binom{n}{\nu}}{2l+2\nu+3}=\frac{n!}{2}\frac{\Gamma(l+\frac{3}{2})}{\Gamma(l+n+\frac{5}{2})}. \tag{A5}$$

APPENDIX B: VERIFICATION OF EQ. (24)

Consider condition (23)

$$\sum_{\eta=\nu}^{n}\frac{a_\eta\eta!R^{2\eta}}{(\eta-\nu)!}=0.$$

Let us evaluate the series with the a_η given by (24)

$$a_\eta=(-1)^{n-\eta}R^{2(n-\eta)}\binom{n}{\eta}a_n.$$

Substituting into the sum, we have

$$\sum_{\eta=\nu}^{n}\frac{a_\eta\eta!R^{2\eta}}{(\eta-\nu)!}=\sum_{\eta=\nu}^{n}\left[(-1)^{n-\eta}R^{2(n-\eta)}\binom{n}{\eta}a_n\right]\frac{\eta!R^{2\eta}}{(\eta-\nu)!}=(-1)^nR^{2n}a_n\sum_{\eta=\nu}^{n}(-1)^\eta\binom{n}{\eta}\frac{\eta!}{(\eta-\nu)!}.$$

Now let $\eta' = \eta - \nu$; then

$$\sum_{\eta=\nu}^{n} \frac{a_\eta \eta! R^{2\eta}}{(\eta-\nu)!} = (-1)^{n-\nu} R^{2n} \frac{a_n n!}{(n-\nu)!} \sum_{\eta'=0}^{n-\nu} (-1)^{\eta'} \binom{n-\nu}{\eta'}.$$

Then from (A4), we have that this is zero for all $\nu < n$, hence

$$\sum_{\eta=\nu}^{n} \frac{a_\eta \eta! R^{2\eta}}{(\eta-\nu)!} = \begin{cases} 0, & \nu = 0,1,\ldots,n-1, \\ a_n n! R^{2n}, & \nu = n, \end{cases} \tag{A6}$$

which means that the a_η's given by (24) are solutions to (23).

APPENDIX C. EXPANSION OF A "COMPLEX" PLANE-WAVE

The general problem of determining the expansion in spherical harmonics of a plane wave of the form $\exp(i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel \pm k_z z)$ has been considered by Kambe[15] and also by Weinert.[16] The results of Kambe[15] differ from those of Ref. 16 by a phase because proper account was not taken of the restrictions associated with the cut in the complex plane of the associated Legendre functions $P_l^m(z)$.

To prove (33), i.e., the case $k_z = k_\parallel$, we do not follow the full development given in Ref. 16, but rather use a simple demonstration.[16] Consider the plane wave and make the normal power series expansion which is valid for all complex values:

$$e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel \mp k_\parallel z} = e^{k_\parallel r[i \sin\theta_r \cos(\varphi_r - \varphi_k) \mp \cos\theta_r]} = \sum_{l=0}^{\infty} \frac{(k_\parallel r)^l}{l!} (\mp 1)^l [\cos\theta_r \mp i \sin\theta_r \cos(\varphi_r - \varphi_k)]^l. \tag{A7}$$

From the theory of the associated Legendre functions, we have the expansion[17]

$$(\cos\theta \pm i \sin\theta \cos\varphi)^n = P_n(\cos\theta) + 2 \sum_{m=1}^{n} (\pm 1)^m e^{-(1/2)m\pi i} \frac{n!}{(n+m)!} \times \cos m\varphi \, P_n^m(\cos\theta) = \sum_{m=-n}^{n} (\pm 1)^m e^{-(1/2)m\pi i} \frac{n!}{(n+m)!} P_n^m(\cos\theta) e^{im\varphi}. \tag{A8}$$

(The associated Legendre functions $P_l^m(\cos\theta)$ have the standard phase, as for example in Ref. 18.) Using (A8) in (A7) with the identification

$$\varphi = \varphi_r - \varphi_k + \pi,$$

we have

$$e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel \mp k_\parallel z} = \sum_{l=0}^{\infty} \frac{(k_\parallel r)^l}{l!} (\mp 1)^l \sum_{m=-l}^{l} (\mp 1)^m (-i)^m \frac{l!}{(l+m)!} \times P_l^m(\cos\theta) e^{im(\varphi_r - \varphi_k)},$$

or writing this in terms of the normalized spherical harmonics

$$e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel \mp k_\parallel z} = \sum_{lm} (-1)^l (\pm 1)^{l-m} i^m \left[\frac{4\pi}{2l+1} \frac{1}{(l+m)!(l-m)!}\right]^{1/2} \times e^{-im\varphi_k} Y_{lm}(\hat{r}). \tag{A9}$$

We also quote the general result[16]

$$e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel \mp k_z z} = \sum_{lm} 4\pi i^l j_l(\kappa r) \left[\left(\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}\right)^{1/2} (\pm 1)^{l+m} i^m \times P_l^m\left(\frac{iK_z}{\kappa}\right) e^{-im\varphi_k}\right] Y_{lm}(\hat{r}), \tag{A10}$$

where

$$\kappa = \sqrt{k_\parallel^2 - k_z^2}$$

and $P_l^m(ik_z/\kappa)$ must be treated as a function of a complex variable.

Note added in proof: After submission of the manuscript, P. Herzig pointed out the similarity between the present method as used for calculating Madelung constants and the Bertaut method [E. F. Bertaut, J. Phys. Chem. Solids **39**, 97 (1978)]. The results of the present method as given in Sec. VI and Bertaut's Eq. (2.5) yield equivalent sums for the Madelung potential when the proper identifications are made. I would like to thank P. Herzig for bringing this reference to my attention and also for demonstrating the similarity between the two methods.

[1] P. D. DeCicco, Phys. Rev. **153**, 931 (1967); G. S. Painter, Phys. Rev. B **7**, 3520 (1973); N. Elyasher and D. D. Koelling, Phys. Rev. B **13**, 5362 (1976); A. Zunger and A. J. Freeman, Phys. Rev. B **15**, 4716 (1977); and references therin.
[2] E. Madelung, Z. Phys. **19**, 524 (1918).
[3] P. P. Ewald, Ann. Phys. **64**, 253 (1921).
[4] For a review of Ewald-type methods, see M. P. Tosi, Solid State Phys. **16**, 1 (1964).
[5] B. R. A. Nijboer and F. W. DeWette, Physica **23**, 309 (1957); Physica **24**, 422 (1958); F. W. DeWette and B. R. A. Nijboer, Physica **24**, 1105 (1958).
[6] F. G. Fumi and M. P. Tosi, Phys. Rev. **117**, 1466 (1960).
[7] W. E. Rudge, Phys. Rev. **181**, 1020 (1969).
[8] J. L. Birman, J. Phys. Chem. Solids **6**, 65 (1958).
[9] D. R. Hamann, private communication to H. Krakauer.
[10] D. R. Hamann also has a method working on these ideas, but the formulations are completely independent.
[11] J. D. Jackson, *Classical Electrodynamics*, 2nd Edition (Wiley, New York, 1975).
[12] G. P. Tolstov, *Fourier Series* (Dover, New York, 1976).
[13] E. Wimmer and H. Krakauer, unpublished.
[14] M. Posternak, H. Krakauer, A. J. Freeman, and D. D. Koelling, Phys. Rev. B **21**, 5601 (1980).
[15] K. Kambe, Z. Naturforsch, **23a**, 1280 (1968).
[16] M. Weinert, unpublished.
[17] E. W. Hobson, *Spherical and Ellipsoidal Harmonics* (Chelesa, New York, 1955), p. 97.
[18] M. Abramowitz and I. A. Stegun, editors, *Handbook of Mathematical Functions* (Dover, New York, 1972).

Linear equations invariant under arbitrary coordinate changes

G. H. Derrick
School of Physics, University of Sydney, Sydney N.S.W. 2006, Australia

(Received 4 June 1980; accepted for publication 12 September 1980)

Linear homogeneous equations of the type $i\alpha^\lambda \partial\psi/\partial x^\lambda = \beta\psi$ are considered, where α^λ and β are constant Hermitian $N \times N$ matrices and ψ is an N component column vector. The conditions that this equation is invariant under arbitrary changes of coordinates are shown to be $\alpha^\lambda C^\kappa_\iota + C^{\dagger\kappa}_\iota \alpha^\lambda = \alpha^\kappa \delta^\lambda_\iota - \alpha^\lambda \delta^\kappa_\iota$, $\beta C^\kappa_\iota + C^{\dagger\kappa}_\iota \beta = -\beta\delta^\kappa_\iota$, $\alpha^\lambda C^\kappa_\iota + \alpha^\kappa C^\lambda_\iota = 0$, where C^κ_ι is a set of 16 $N \times N$ matrices which specify the transformation law for the field components ψ. Some theorems are proved about solutions of these matrix relations and some explicit representations are given.

PACS numbers: 02.30.Jr

1. INTRODUCTION

Einstein's equations $R_{\iota\kappa} = 0$ for the gravitational field outside matter possess the remarkable property of invariance under arbitrary changes of coordinates. The term *invariance* is used here in its strongest sense: suppose we replace our coordinates x^λ by another set x'^λ (which are arbitrary functions of the unprimed set), and correspondingly transform the metric tensor $g_{\iota\kappa}$ and the Ricci tensor $R_{\iota\kappa}$ in the standard way, i.e.,

$$g'_{\iota\kappa} = g_{\lambda\mu}(\partial x^\lambda/\partial x'^\iota)(\partial x^\mu/\partial x'^\kappa),$$
$$R'_{\iota\kappa} = R_{\lambda\mu}(\partial x^\lambda/\partial x'^\iota)(\partial x^\mu/\partial x'^\kappa).$$

The property of invariance referred to stems from the fact that $R'_{\iota\kappa}$ is the *same* function of $g'_{\lambda\mu}$, $\partial g'_{\lambda\mu}/\partial x'^\rho$, $\partial^2 g'_{\lambda\mu}/\partial x'^\rho \partial x'^\nu$ as $R_{\iota\kappa}$ is of $g_{\lambda\mu}$, $\partial g_{\lambda\mu}/\partial x^\rho$, $\partial^2 g_{\lambda\mu}/\partial x^\rho \partial x^\nu$. Hence the transformed equations $R'_{\iota\kappa} = 0$ are in fact identical to the original set $R_{\iota\kappa} = 0$ except for the primes attached to the variables. These equations contain the dependent variables and their first and second derivatives but do not involve the independent variables, i.e., the coordinates, explicitly.

This strong property of *invariance* should be distinguished from the weaker concept of *covariance*. Consider, for example, the Laplace equation in two dimensions. In Cartesian coordinates x,y it takes the form

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = 0. \tag{1}$$

Introducing polar coordinates r,θ the equation transforms to

$$\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2} = 0. \tag{2}$$

Equations (1) and (2) may be written in the covariant form

$$\frac{1}{\gamma}\frac{\partial}{\partial x^\lambda}\left(\gamma g^{\lambda\mu}\frac{\partial\psi}{\partial x^\mu}\right) = 0, \tag{3}$$

where γ is the square root of the metric determinant $|g_{\iota\kappa}|$ with appropriate values for the metric tensor

$$g_{xx} = 1, \quad g_{yy} = 1, \quad g_{xy} = 0,$$
$$g_{rr} = 1, \quad g_{\theta\theta} = r^2, \quad g_{r\theta} = 0.$$

However the existence of a covariant representation, Eq. (3), certainly does not imply that Eqs. (1) and (2) are identical in form. Indeed Eq. (2) contains first derivatives and explicit dependence on coordinates, whereas Eq. (1) has neither of these features.

When matter is present Einstein's field equations become

$$R_{\iota\kappa} - \tfrac{1}{2}g_{\iota\kappa}R = -8\pi G T_{\iota\kappa}. \tag{4}$$

If the energy–momentum–stress source term $T_{\iota\kappa}$ can be written in terms of component scalar, vector, and tensor fields, then Eq. (4) together with the field equations for the particle fields again exhibits the property of coordinate invariance rather than mere covariance.

The above considerations suggest that it would be of interest to classify all possible sets of field equations which are invariant under the choice of coordinates. In general we shall have N dependent fields ψ^A, $(A = 1,2,...,N)$, whose dependence on the four coordinates x^λ is to be found from solution of a set of N differential equations

$$L^B(\psi^A, \partial\psi^A/\partial x^\iota, \partial^2\psi^A/\partial x^\iota \partial x^\kappa, ...) = 0. \tag{5}$$

We now seek a transformation law $\psi^A \to \psi'^A$ associated with the coordinate change $x^\lambda \to x'^\lambda$ and suitable functions L^B such that when Eq. (5) is satisfied it follows that

$$L^B(\psi'^A, \partial\psi'^A/\partial x'^\iota, \partial^2\psi'^A/\partial x'^\iota \partial x'^\kappa, ...) = 0. \tag{6}$$

Note that there is no prime on L^B in Eq. (6) since we demand that the same relations hold for the primed variables as for the unprimed.

As a first step in classification of coordinate invariant field equations, we shall consider in the present paper the simplest possible system, a set of linear, first order, homogeneous differential equations.

2. COORDINATE INVARIANT LINEAR EQUATIONS

Suppose we have a field ψ^A, $(A = 1,2,...,N)$, in general complex, whose dependence on the four coordinates x^λ, $(\lambda = 0,1,2,3)$, is governed by N coupled first order linear differential equations. For economy in writing we shall use matrix notation. Let ψ be the column vector formed from the N components ψ^A, and $\psi^\dagger$ the Hermitian conjugate vector, i.e., the row vector of complex conjugate fields ψ^{*A}. The most general linear first order homogeneous system may be written

 0022-2488/81/112440-07$01.00

$$i\alpha^\lambda \frac{\partial\psi}{\partial x^\lambda} = \beta\psi\,, \tag{7}$$

where α^λ, $(\lambda = 0,1,2,3)$, and β are $N\times N$ constant matrices. We shall assume that α^λ and β are Hermitian in order that the current $j^\lambda = \psi^\dagger\alpha^\lambda\psi$ be conserved, i.e., $\partial j^\lambda/\partial x^\lambda = 0$.

Suppose we now change coordinates from x^λ to x'^λ and try a transformation law $\psi\to\psi'$ of the form

$$\psi' = D\psi\,. \tag{8}$$

Here D is an $N\times N$ matrix which is a function of the coordinates and depends, in a way to be determined, on the functional relationship between x^λ and x'^λ. If ψ were made up of components of tensors and tensor densities then one would have $D = D(\xi)$, where ξ is the 4×4 matrix of the partial derivatives $\xi^\iota_\kappa = \partial x'^\iota/\partial x^\kappa$. In this case $D(\xi)$ would be an $N\times N$ matrix representation of the group of real nonsingular 4×4 matrices, i.e., $D(\xi_1)D(\xi_2) = D(\xi_1\xi_2)$. However, we shall *not* impose this representation condition *a priori*, but instead shall adopt a weaker hypothesis, relating to the form of D for infinitesimal coordinate changes, in Sec. 3. The question of integration to give finite transformations will then be take up in Sec. 5.

We now require that the primed field equations should take the same form as the unprimed

$$i\alpha^\lambda \frac{\partial\psi'}{\partial x'^\lambda} = \beta\psi'\,, \tag{9}$$

with the same constant matrices α^λ,β as in Eq. (7). The mathematical problem posed is to find matrices α^λ,β,D such that Eq. (7) implies Eq. (9) for arbitrary changes of coordinates.

Equation (7) may be derived from the variation principle $\delta\int\mathscr{L}d^4x = 0$ with

$$\mathscr{L} = \psi^\dagger\left(i\alpha^\lambda\frac{\partial\psi}{\partial x^\lambda} - \beta\psi\right) + \text{ complex conjugate}\,. \tag{10}$$

A sufficient condition for the invariance of Eq. (7) is that $\mathscr{L}$ should transform as a scalar density of weight 1;

$$\mathscr{L}' = \left|\frac{\partial x^\lambda}{\partial x'^\mu}\right|\mathscr{L}\,. \tag{11}$$

3. THE MATRIX RELATIONS

First let us consider the implications of demanding the invariance of Eq. (7) under infinitesimal coordinate changes. Such a transformation is specified by four arbitrary infinitesimal functions $v^\lambda(x^\mu)$:

$$x'^\lambda = x^\lambda + v^\lambda(x^\mu)\,,$$
$$\xi^\lambda_\mu = \frac{\partial x'^\lambda}{\partial x^\mu} = \delta^\lambda_\mu + v^\lambda_{,\mu}\,, \tag{12}$$

where a comma denotes partial differentiation. The transformation matrix D of Eq. (8) is assumed to possess an expansion about the neighborhood of the identical transformation of the form

$$D = I + v^\iota_{,\kappa}C^\kappa_\iota + \text{ second order terms}, \tag{13}$$

where I is the $N\times N$ unit matrix and C^κ_ι is a constant $N\times N$ matrix.

Let us now substitute into Eq. (11) the transformed coordinates and fields specified by Eqs. (8), (12), and (13). We obtain to first order of small quantities a linear combination of the first derivatives $v^\iota_{,\kappa}$ and second derivatives $v^\iota_{,\kappa\lambda}$. Since the functions v^λ are quite arbitrary the coefficients of these terms must vanish. The coefficient of $v^\iota_{,\kappa}$ is

$$i\psi^\dagger(\alpha^\lambda C^\kappa_\iota + C^{\dagger\kappa}_\iota\alpha^\lambda - \alpha^\kappa\delta^\lambda_\iota + \alpha^\lambda\delta^\kappa_\iota)\psi_{,\lambda}$$
$$+ \text{ complex conjugate} - 2\psi^\dagger(\beta C^\kappa_\iota + C^{\dagger\kappa}_\iota\beta + \beta\delta^\kappa_\iota)\psi\,.$$

The coefficient of $v^\iota_{,\kappa\lambda}$ is

$$\tfrac{1}{2}i\psi^\dagger(\alpha^\lambda C^\kappa_\iota + \alpha^\kappa C^\lambda_\iota - C^{\dagger\kappa}_\iota\alpha^\lambda - C^{\dagger\lambda}_\iota\alpha^\kappa)\psi\,.$$

Hence Eq. (11) will hold and Eq. (7) be invariant if the matrices $\alpha^\lambda,\beta,C^\kappa_\iota$ obey the following relations

$$\alpha^\lambda C^\kappa_\iota + \alpha^\kappa C^\lambda_\iota = C^{\dagger\kappa}_\iota\alpha^\lambda + C^{\dagger\lambda}_\iota\alpha^\kappa\,, \tag{14}$$
$$\alpha^\lambda C^\kappa_\iota + C^{\dagger\kappa}_\iota\alpha^\lambda = \alpha^\kappa\delta^\lambda_\iota - \alpha^\lambda\delta^\kappa_\iota\,, \tag{15}$$
$$\beta C^\kappa_\iota + C^{\dagger\kappa}_\iota\beta = -\beta\delta^\kappa_\iota\,. \tag{16}$$

Combining Eq. (14) with Eq. (15), we find that each side of Eq. (14) separately vanishes

$$\alpha^\lambda C^\kappa_\iota + \alpha^\kappa C^\lambda_\iota = 0\,. \tag{17}$$

Hence we conclude that Eqs. (15)–(17) are sufficient conditions that Eq. (7) be invariant under arbitrary infinitesimal coordinate transformations.

The interpretation of Eqs. (15) and (16) is interesting. If we consider how the current $\psi^\dagger\alpha^\lambda\psi$ transforms we find that Eq. (15) is simply the condition that this current be a vector density of weight 1. This is a desirable feature in view of the vanishing divergence of the current. Similarly Eq. (16) implies that $\psi^\dagger\beta\psi$ is a scalar density of weight 1.

4. SYMMETRIES OF THE MATRIX RELATIONS (15), (16), AND (17)

Suppose α^λ,β, and C^κ_ι satisfy Eqs. (15)–(17) and $\psi(x^\mu)$ is the corresponding solution of Eq. (7). Then the following are also solutions

(a) $S^\dagger\alpha^\lambda S,\quad S^\dagger\beta S,\quad S^{-1}C^\kappa_\iota S,\quad S^{-1}\psi(x^\mu)$,
where S is an arbitrary, nonsingular, complex $N\times N$ matrix.

(b) $\alpha^{*\lambda},\quad \beta^*,\quad C^{*\kappa}_\iota,\quad \psi^*(-x^\mu)$,
where * denotes complex conjugation (without taking the transpose).

(c) $T^\lambda_\mu\alpha^\mu,\quad T\beta,\quad T^\kappa_\lambda C^\lambda_\mu T^{-1\mu}_\iota,\quad \psi(TT^{-1\mu}_\nu x^\nu)$,
where T is any real number and the T^κ_λ are the elements of an arbitrary, real, nonsingular, 4×4 matrix.

(d) If we can find further matrices C^5_ι, C^κ_5, C^5_5 which satisfy the additional relations necessary to extend Eqs. (15)–(17) to the larger set

$$\alpha^l C^k_i + C^{\dagger k}_i\alpha^l = \alpha^k\delta^l_i - \alpha^l\delta^k_i\,,$$
$$\alpha^l C^k_i + \alpha^k C^l_i = 0\,, \tag{18}$$

where $\beta = \alpha^5$ and Latin indices range over the five values 0, 1, 2, 3, 5, then $\bar\alpha^l = T^l_m\alpha^m$, $\bar C^k_i = T^k_l C^l_m T^{-1m}_i$ again satisfy Eq. (18). Here T^l_m are the elements of an arbitrary 5×5 real nonsingular matrix.

The symmetries (a) should clearly be regarded as equivalence relations, with the transformed quantities belonging

to the same physical system—the field components have merely been replaced by linear combinations of themselves. The role of the symmetries (b), (c), and (d) is less clear. Perhaps (b) relates to charge conjugation or TCP invariance.

5. FINITE COORDINATE TRANSFORMATIONS

A. Integration of infinitesimal transformations

Let us restrict ourselves to transformations $x'^{\lambda} = f^{\lambda}(x^{\nu})$ which can be connected by a continuous path to the identical transformation, i.e., there exists a family of transformations $x'^{\lambda}_{t} = f^{\lambda}(x^{\nu},t)$ characterized by a parameter t, $0 \leqslant t \leqslant 1$, such that

$$f^{\lambda}(x^{\nu},0) = x^{\lambda},$$
$$f^{\lambda}(x^{\nu},1) = f^{\lambda}(x^{\nu}),$$
$$\left| \frac{\partial f^{\lambda}(x^{\nu},t)}{\partial x^{\mu}} \right| > 0 \quad \text{for } 0 \leqslant t \leqslant 1.$$

The Jacobian condition enables us to find the inverse transformation $x^{\nu} = F^{\nu}(x'^{\lambda}_{t},t)$. Assuming differentiability with respect to t we have

$$x'^{\iota}_{t+dt} = x'^{\iota}_{t} + dt\, \eta^{\iota}(x'^{\lambda}_{t},t), \tag{19}$$

where the function $\eta^{\iota}(y^{\lambda},t)$ is defined by

$$\eta^{\iota}(y^{\lambda},t) = \frac{\partial f^{\iota}}{\partial t}(x^{\nu},t)\Big|_{x^{\nu}=F^{\nu}(y^{\lambda},t)}.$$

Equation (19) gives the transformation at parameter value $t + dt$ in terms of an infinitesimal transformation applied to the coordinate system x'^{ι}_{t}. Hence we may apply the formalism of Sec. 3 and write a differential equation in t for the transformation matrix $D(t)$ along the path

$$D(t+dt) = [I + \eta^{\iota}_{,\kappa}(x'^{\lambda}_{t},t)C^{\kappa}_{\iota}]D(t)$$

or

$$\frac{\partial D(t)}{\partial t} = \eta^{\iota}_{,\kappa}(x'^{\lambda}_{t},t)C^{\kappa}_{\iota}D(t). \tag{20}$$

The initial conditions $D(0) = I$ then defines the transformation matrix $D(t)$ uniquely along the path, and in particular at the end point $t = 1$. In general the result will be path dependent, a point considered further in Sec. C.

B. The transformation matrix to second order

Let us expand the solution of Eqs. (19) and (20) as a power series in t. Retaining terms up to t^2 we obtain

$$x'^{\iota}_{t} = x^{\iota} + v^{\iota},$$
$$D(t) = I + v^{\iota}_{,\kappa}C^{\kappa}_{\iota} + \tfrac{1}{2}v^{\iota}_{,\kappa}v^{\lambda}_{,\mu}(C^{\kappa}_{\iota}C^{\mu}_{\lambda} - \delta^{\kappa}_{\lambda}C^{\mu}_{\iota}) + O(t^3), \tag{21}$$

where

$$v^{\iota} = t\eta^{\iota}(x^{\nu},0) + \tfrac{1}{2}t^2[\eta^{\iota}_{,t}(x^{\nu},0) + \eta^{\iota}_{,\lambda}(x^{\nu},0)\eta^{\lambda}(x^{\nu},0)].$$

If we now take t, and hence v^{ι} infinitesimal, we see that Eq. (21) provides the second order terms in Eq. (13). Note that to second order D is a function solely of $\partial x'^{\lambda}/\partial x^{\mu}$.

C. Path dependence of the transformation matrix D

If we evaluate the third order terms in Eq. (21) we find that it is no longer true that D is a function only of $\partial x'^{\lambda}/\partial x^{\mu}$. We obtain an additional term containing second derivatives which takes the form

$$\tfrac{1}{12}[\eta^{\iota}_{,\kappa t}(x^{\nu},0) + \eta^{\iota}_{,\kappa\rho}(x^{\nu},0)\eta^{\rho}(x^{\nu},0)]\eta^{\lambda}_{,\mu}(x^{\nu},0) \times [C^{\kappa}_{\iota}C^{\mu}_{\lambda} - C^{\mu}_{\lambda}C^{\kappa}_{\iota} - \delta^{\kappa}_{\lambda}C^{\mu}_{\iota} + \delta^{\mu}_{\iota}C^{\kappa}_{\lambda}]. \tag{22}$$

The presence of the initial t-derivative term $\eta^{\iota}_{,\kappa t}(x^{\nu},0)$ means that in general D depends on the path. A necessary condition for path independence is the vanishing of the second term, i.e.,

$$C^{\kappa}_{\iota}C^{\mu}_{\lambda} - C^{\mu}_{\lambda}C^{\kappa}_{\iota} = \delta^{\kappa}_{\lambda}C^{\mu}_{\iota} - \delta^{\mu}_{\iota}C^{\kappa}_{\lambda}. \tag{23}$$

These are the well-known commutation relations for the generators of the linear groups.[1] They are in fact also sufficient conditions that the representation condition hold, i.e., $D(\xi_1)D(\xi_2) = D(\xi_1\xi_2)$. The finite-dimensional matrix representations of C^{κ}_{ι} satisfying Eq. (23) correspond to tensors and tensor densities. Hence if we were to insist on Eq. (23) we would restrict the components of ψ to be such quantities. This seems unnecessarily restrictive and instead we shall adopt the following attitude: We ignore Eq. (23) and find what representations we can of $\alpha^{\lambda}, \beta, C^{\kappa}_{\iota}$ satisfying Eqs. (15)–(17). If Eq. (23) turns out to be satisfied then ψ is singly valued and thus an acceptable physical observable. If on the other hand Eq. (23) is violated then ψ is multivalued, since we can change its value by effecting a sequence of transformations leading from the identical transformation back to itself along a closed path. In this case ψ itself must be regarded as unobservable, and physical observables should be singly valued quantities constructed from ψ and $\psi^{\dagger}$, perhaps real quadratic forms of the type $\psi^{\dagger}M\psi$, where M is an Hermitian $N \times N$ matrix. For example the current $\psi^{\dagger}\alpha^{\lambda}\psi$ is an acceptable observable since it transforms as a vector density of weight 1 and is thus certainly singly valued. This approach is similar to that adopted in the quantum mechanics of half-integral spin systems. The wavefunction is there multivalued and unobservable and all measurable quantities correspond to singly valued real combinations of the wave function and its complex conjugate.

The problem of finding suitable observables of quadratic form reduces to the following. Supose we seek a matrix M^{a}, $(a = 1,2,...,n)$, such that $\psi^{\dagger}M^{a}\psi$ transforms as a tensor or tensor density with n components. Let K^{κ}_{ι} be the generators for the associated representation. K^{κ}_{ι} will be a set of 16 $n \times n$ matrices, matrix elements $K^{\kappa a}_{\iota b}$, $(\iota,\kappa = 0,1,2,3;\ a,b = 1,2,...,n)$, which satisfy the commutation relations, Eq. (23). For the infinitesimal transformation $x'^{\lambda} = x^{\lambda} + v^{\lambda}$ we require

$$(\psi^{\dagger}M^{a}\psi)' = (\delta^{a}_{b} + v^{\iota}_{,\kappa}K^{\kappa a}_{\iota b})\psi^{\dagger}M^{b}\psi. \tag{24}$$

Writing ψ' according to Eqs. (8) and (13) and equating coefficients of $v^{\iota}_{,\kappa}$ to zero yields

$$M^{a}C^{\kappa}_{\iota} + C^{\dagger\kappa}_{\iota}M^{a} = K^{\kappa a}_{\iota b}M^{b}. \tag{25}$$

In particular, if $\psi^{\dagger}M^{\lambda\cdots}_{\mu\cdots}\psi$ is a tensor density of weight W containing a number of contravariant indices λ and covar-

iant indices μ then Eq. (25) reduces to

$$M^{\lambda\cdots}_{\mu\cdots}C^{\kappa}_{\iota}+C^{\dagger\kappa}_{\iota}M^{\lambda\cdots}_{\mu\cdots}$$
$$=\sum(\delta^{\lambda}_{\iota}M^{\kappa\cdots}_{\mu\cdots}-\delta^{\kappa}_{\mu}M^{\lambda\cdots}_{\iota\cdots})-W\delta^{k}_{\iota}M^{\lambda\cdots}_{\mu\cdots}, \tag{26}$$

the summation being over all contravariant indices λ and all covariant indices μ.

6. REDUCIBILITY

A. Introduction

In order to classify the solutions of the matrix equations (15)–(17) one needs some definition of *reducible* and *irreducible* representations. The equivalence relations of relevance are those of type (a) in Sec. 4:

$$\alpha'^{\lambda}=S^{\dagger}\alpha^{\lambda}S,$$
$$\beta'=S^{\dagger}\beta S, \tag{27}$$
$$C'^{\kappa}_{\iota}=S^{-1}C^{\kappa}_{\iota}S,$$

where S is an arbitrary, nonsingular, complex $N\times N$ matrix. Two alternative definitions of reducibility suggest themselves, depending on whether we focus attention on the matrices α^{λ},β or on the matrices C^{κ}_{ι}.

B. $\alpha\beta$-reducibility

A solution $\alpha^{\lambda},\beta,C^{\kappa}_{\iota}$ of the matrix equations (15)–(17) will be called $\alpha\beta$-reducible if α^{λ},β can be transformed according to Eq. (27) to the direct sum form

$$\alpha'^{\lambda}=\begin{pmatrix}\alpha^{\lambda}_{11}&0\\0&\alpha^{\lambda}_{22}\end{pmatrix},$$
$$\beta'=\begin{pmatrix}\beta_{11}&0\\0&\beta_{22}\end{pmatrix}, \tag{28}$$

where $\alpha^{\lambda}_{11},\beta_{11}$ and $\alpha^{\lambda}_{22},\beta_{22}$ are respectively $N_1\times N_1$ and $N_2\times N_2$ matrices with $N_1+N_2=N$, $N_1\neq 0, N_2\neq 0$. Let us partition C'^{κ}_{ι} in the same way:

$$C'^{\kappa}_{\iota}=\begin{pmatrix}C^{\kappa}_{\iota 11}&C^{\kappa}_{\iota 12}\\C^{\kappa}_{\iota 21}&C^{\kappa}_{\iota 22}\end{pmatrix}.$$

Note that we do not require in this definition that the matrices $C^{\kappa}_{\iota 12}, C^{\kappa}_{\iota 21}$ vanish. This is because the reduced form, Eq. (28), implies that $\alpha^{\lambda}_{11},\beta_{11},C^{\kappa}_{\iota 11}$ and $\alpha^{\lambda}_{22},\beta_{22},C^{\kappa}_{\iota 22}$ are separately solutions of the matrix equations whether or not $C^{\kappa}_{\iota 12}$ and $C^{\kappa}_{\iota 21}$ are zero. Further the field equation (7) decomposes into two uncoupled equations of the same type, of dimensions N_1 and N_2. Hence it would seem most natural to set $C^{\kappa}_{\iota 12}$ and $C^{\kappa}_{\iota 21}$ both equal to zero and to thus effect complete decomposition into two disjoint sets of field equations.

C. C-reducibility

A solution $\alpha^{\lambda},\beta,C^{\kappa}_{\iota}$ of the matrix relations will be called C-reducible if the matrices C^{κ}_{ι} are reducible in the ordinary sense, i.e., C^{κ}_{ι} may be transformed to the form

$$C'^{\kappa}_{\iota}=\begin{pmatrix}C^{\kappa}_{\iota 11}&C^{\kappa}_{\iota 12}\\0&C^{\kappa}_{\iota 22}\end{pmatrix}, \tag{29}$$

where the N rows and columns are partitioned according to $N=N_1+N_2, N_1\neq 0, N_2\neq 0$. If we partition α'^{λ},β' in the corresponding way we find that $\alpha^{\lambda}_{11},\beta_{11}$, and $C^{\kappa}_{\iota 11}$ give an $N_1\times N_1$ representation of the matrix equations (15)–(17). However the field equations (17) will in general not decompose because of the coupling terms $\alpha^{\lambda}_{12},\beta_{12}$.

D. Analogs of Schur's lemma

We define a system to be irreducible if it is not reducible. The following results are proved in the Appendix.

Lemma 1.1: If $\alpha^{\lambda},\beta,C^{\kappa}_{\iota}$ is $\alpha\beta$-irreducible and a matrix P exists such that

$$\alpha^{\lambda}P=P^{\dagger}\alpha^{\lambda},\quad \beta P=P^{\dagger}\beta$$

then either P^{-1} exists or all the eigenvalues of P are zero.

Lemma 1.2: If the matrix P of Lemma 1.1 has any real eigenvalue then all the eigenvalues of P are real and equal. As a corollary, if P is also diagonalizable then it is a multiple of the unit matrix.

Lemma 2: If $\alpha^{\lambda},\beta,C^{\kappa}_{\iota}$ is C-irreducible then β^{-1} exists or $\beta=0$.

Note that Lemmas 1.1 and 1.2 are much weaker than the usual result that only a multiple of the unit matrix can commute with all members of an irreducible set.

7. FURTHER LEMMAS

The following results are also proved in the Appendix, and prove helpful in the construction of representations.

Lemma 3: If β^{-1} exists then Eq. (7) has only the trivial solution $\psi=0$.

Lemma 4.1: Suppose Hermitian matrices α_{ι} exist which satisfy the relations

$$\alpha^{\lambda}\alpha_{\iota}\alpha^{\kappa}+\alpha^{\kappa}\alpha_{\iota}\alpha^{\lambda}=\alpha^{\lambda}\delta^{\kappa}_{\iota}+\alpha^{\kappa}\delta^{\lambda}_{\iota},$$
$$\beta\alpha_{\iota}\alpha^{\kappa}+\alpha^{\kappa}\alpha_{\iota}\beta=\beta\delta^{\kappa}_{\iota}. \tag{30}$$

Then $C^{\kappa}_{\iota}=\alpha_{\iota}\alpha^{\kappa}-\delta^{\kappa}_{\iota}$ together with α^{λ},β gives a solution of the matrix equations. Equations (30) are reminiscent of the Duffin–Kemmer commutation relations,[2,3] but differ in that α_{ι} is not necessarily to be obtained from α^{κ} by lowering the superscript with the aid of a metric tensor.

Lemma 4.2: If in addition to Eq. (30) the matrices α_{ι} of Lemma 4.1 satisfy the further relations

$$\alpha_{\lambda}\alpha^{\iota}\alpha_{\kappa}+\alpha_{\kappa}\alpha^{\iota}\alpha_{\lambda}=\alpha_{\lambda}\delta^{\iota}_{\kappa}+\alpha_{\kappa}\delta^{\iota}_{\lambda}, \tag{31}$$

then the representation condition, Eq. (23), holds.

Lemma 5.1: Suppose a nonsingular Hermitian matrix B and matrices γ_{ι}, γ^{κ} exist such that

$$\gamma^{\dagger\kappa}=B\gamma^{\kappa}B^{-1},\quad \gamma^{\dagger}_{\iota}=B\gamma_{\iota}B^{-1},$$
$$\gamma^{\iota}\gamma^{\kappa}+\gamma^{\kappa}\gamma^{\iota}=0, \tag{32}$$
$$\gamma_{\iota}\gamma^{\kappa}+\gamma^{\kappa}\gamma_{\iota}=2\delta^{\kappa}_{\iota}.$$

Then $\alpha^{\kappa}=B\gamma^{\kappa}, C^{\kappa}_{\iota}=-\frac{1}{2}\gamma^{\kappa}\gamma_{\iota}, \beta=0$ gives a representation of Eq. (15)–(17) which is both $\alpha\beta$-reducible and C-reducible. $\beta=B$ also gives a solution, which is however of no interest on account of Lemma 3.

Lemma 5.2: If the matrices γ_ι of Lemma 5.1 satisfy the further relations

$$\gamma_\iota\gamma_\kappa + \gamma_\kappa\gamma_\iota = 0\,, \tag{33}$$

then the representation condition, Eq. (23), also holds.

Lemma 5.3: If the indices ι, κ in Eq. (32) range over the *five* values 0, 1, 2, 3, 5 instead of the usual four values 0, 1, 2, 3 then $\alpha^\kappa = B\gamma^\kappa$, $C_\iota^\kappa = -\frac{1}{2}\gamma^\kappa\gamma_\iota$, $\beta = \alpha^5$ gives a solution, again $\alpha\beta$-reducible and C-reducible. As in Lemma 5.2 the further condition, Eq. (33), will ensure the truth of Eq. (23).

Lemma 6: If Eqs. (15)–(17) hold then

$$\alpha^\nu P^{\kappa\mu}_{\iota\lambda} = P^{\dagger\kappa\mu}_{\iota\lambda}\alpha^\nu\,,$$
$$\beta P^{\kappa\mu}_{\iota\lambda} = P^{\dagger\kappa\mu}_{\iota\lambda}\beta\,, \tag{34}$$

where

$$iP^{\kappa\mu}_{\iota\lambda} = C^\kappa_\iota C^\mu_\lambda - C^\mu_\lambda C^\kappa_\iota - \delta^\kappa_\lambda C^\mu_\iota + \delta^\mu_\iota C^\kappa_\lambda\,. \tag{35}$$

Lemmas 1.1 and 1.2 are applicable to Eq. (34) and place severe limitations on the quantities $P^{\kappa\mu}_{\iota\lambda}$, which measure the extent to which the representation condition, Eq. (23), is violated. Further constraints on $P^{\kappa\mu}_{\iota\lambda}$ come from the Jacobi relation that the cyclic sum of the triple commutations $[[C^\kappa_\iota, C^\mu_\lambda], C^\pi_\nu]$ must vanish.

8. EXPLICIT REPRESENTATIONS

A. Method of construction

In this section we give all known nontrivial $\alpha\beta$-irreducible representations of Eqs. (15)–(17). They have dimensions 5, 6, 10, 10, 15, and have all been obtained with the aid of Lemmas 4.1 and 4.2.

Consider the Duffin–Kemmer algebra whose four generating elements β_λ satisfy the commutation relations

$$\beta_\lambda\beta_\iota\beta_\kappa + \beta_\kappa\beta_\iota\beta_\lambda = \beta_\lambda\delta_{\kappa\iota} + \beta_\kappa\delta_{\lambda\iota}\,. \tag{36}$$

This algebra has two nontrivial irreducible matrix representations, one in five dimensions and the other in ten dimensions.[4] We can clearly obtain a representation of Eq. (30) from one of Eq. (36) by taking $\alpha^\lambda = \alpha_\lambda = \beta_\lambda, \beta = 0$. Application of Lemma 4.1 then yields a 5×5 and a 10×10 representation of Eq (15)–(17).

If on the other hand we allow the indices in Eq. (36) to range over the five values 0,1,2,3,5 then the nontrivial irreducible representations for β_λ are of dimension 6,10,10,15. We now set $\alpha^\lambda = \alpha_\lambda = \beta_\lambda, \beta = (\text{const})\beta_5$ to obtain representations of these dimensions of Eqs. (15)–(17).

Note that in each case Eq. (31) also holds so that we have also satisfied the representation condition, Eq. (23). Hence the transformation matrix D is of the form $D(\xi)$ where $\xi^\iota_\kappa = \partial x'^\iota/\partial x^\kappa$, and the elements of ψ transform as the components of ordinary tensors and tensor densities.

All representations given are $\alpha\beta$-irreducible and C-reducible.

B. The 5×5 representation

This corresponds to the five-dimensional representation of the Duffin–Kemmer algebra in 4-space and thus has $\beta = 0$. The nonzero elements of α^λ are

$$\begin{aligned}
&(\alpha^0)_{15} = (\alpha^0)_{51} = 1\,,\\
&(\alpha^1)_{25} = (\alpha^1)_{52} = 1\,,\\
&(\alpha^2)_{35} = (\alpha^2)_{53} = 1\,,\\
&(\alpha^3)_{45} = (\alpha^3)_{54} = 1\,.
\end{aligned} \tag{37}$$

The first four field components $(\psi_1,\psi_2,\psi_3,\psi_4) \equiv (\chi^0,\chi^1,\chi^2,\chi^3)$ transform as the components of a contravariant vector density of weight 1, χ^λ, while the fifth component ψ_5 is a scalar. This corresponds to the transformation matrix (partitioned $4+1$)

$$D(\xi) = \begin{bmatrix} |\xi|^{-1}\xi & 0 \\ 0 & 1 \end{bmatrix}. \tag{38}$$

The field equation (7) takes the trivial form

$$\begin{aligned}
&\chi^\lambda_{,\lambda} = 0\,,\\
&\psi_{5,\lambda} = 0\,.
\end{aligned} \tag{39}$$

C. The 6×6 representation

This is obtained from the six-dimensional representation of the Duffin–Kemmer algebra in five-space. The nonzero elements of α^λ, β are

$$\begin{aligned}
&(\alpha^0)_{16} = (\alpha^0)_{61} = 1\,,\\
&(\alpha^1)_{26} = (\alpha^1)_{62} = 1\,,\\
&(\alpha^2)_{36} = (\alpha^2)_{63} = 1\,,\\
&(\alpha^3)_{46} = (\alpha^3)_{64} = 1\,,\\
&(\beta)_{56} = (\beta)_{65} = m\,,
\end{aligned} \tag{40}$$

where m is an arbitrary real constant. The first four field components transform as a contravariant vector density of weight 1, the fifth as a scalar density of weight 1 and the sixth as a scalar, corresponding to the transformation matrix (partitioned $4+1+1$)

$$D(\xi) = \begin{bmatrix} |\xi|^{-1}\xi & 0 & 0 \\ 0 & |\xi|^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{41}$$

The field equations (7) are again trivial

$$\begin{aligned}
&i\chi^\mu_{,\mu} = m\psi_5\,,\\
&\psi_{6,\mu} = 0\,,\\
&\psi_6 = 0\,.
\end{aligned} \tag{42}$$

Note that the four functions χ^μ are quite undetermined. The appearance of four arbitrary functions is of course to be expected in the solution of any field equations which permit arbitrary changes in four coordinates.

D. The 10×10 representation

First consider the representation with $\beta = 0$ obtained from the 4-space Duffin–Kemmer algebra. The independent nonzero elements of α^λ are

$$(\alpha^0)_{28} = (\alpha^0)_{39} = (\alpha^0)_{4,10} = 1\,,$$
$$-(\alpha^1)_{18} = (\alpha^1)_{37} = -(\alpha^1)_{46} = 1\,,$$
$$-(\alpha^2)_{19} = -(\alpha^2)_{27} = (\alpha^2)_{45} = 1\,, \tag{43}$$
$$-(\alpha^3)_{1,10} = (\alpha^3)_{26} = -(\alpha^3)_{35} = 1\,.$$

The first four field components $(\psi_1,\psi_2,\psi_3,\psi_4)\equiv(A_0,A_1,A_2,A_3)$ transform as a covariant vector A_λ, while the remaining six field elements $(\psi_6,\psi_7,\psi_8,\psi_9,\psi_{10})\equiv(F_{01},F_{02},F_{03},F_{23},F_{31},F_{12})$ behave as the components of an antisymmetric covariant tensor $F_{\lambda\mu}$. This defines the transformation matrix $D(\xi)$ so there is no need to write it explicitly.

In the case of the 10×10 representation obtained from the 5-space Duffin–Kemmer algebra one obtains the same matrices for α^λ as in Eq. (43) and the same transformation law for ψ. However β is no longer zero but has the following independent nonvanishing elements

$$(\beta)_{58} = (\beta)_{69} = (\beta)_{7,10} = m\,, \tag{44}$$

where m is an arbitrary constant. Thus the first case, $\beta = 0$, may be obtained by setting $m = 0$.

The field equations (7) are

$$F_{\iota\kappa,\lambda} + F_{\kappa\lambda,\iota} + F_{\lambda\iota,\kappa} = 0\,,$$
$$A_{\mu,\lambda} - A_{\lambda,\mu} = imF_{\lambda\mu}\,. \tag{45}$$

Once again the field equations possess redundancy associated with coordinate invariance. For $m\neq0$ the solution depends on four completely arbitrary functions, and for $m = 0$ on five.

E. The 15×15 representation

This representation comes from the 15-dimensional representation of the Duffin–Kemmer algebra in 5-space. Rather than write out the 15×15 matrices for α^λ, β it is less cumbersome to specify the Lagrangian density $\mathscr{L}$ of Eq. (10). The 15 components of ψ are comprised as follows:

$\mathscr{F}^{\lambda\mu}$, a contravariant antisymmetric tensor density of weight 1.

j^λ, a contravariant vector density of weight 1.

A_λ, a covariant vector.

A, a scalar.

In terms of these variables,

$$\psi^\dagger(i\alpha^\lambda\psi_{,\lambda} - \beta\psi)$$
$$= i(\mathscr{F}^{*\lambda\mu}A_{\lambda,\mu} + j^{*\lambda}A_{,\lambda} + A\overset{*}{\lambda}\mathscr{F}^{\lambda\mu}_{,\mu} + A^{*}j^{\lambda}_{,\lambda})$$
$$- m(j^{\lambda*}A_\lambda + A\overset{*}{\lambda}j^\lambda)\,, \tag{46}$$

where m is an arbitrary real constant. The matrix elements of α^λ, β may now be picked out by comparison of coefficients in Eq. (46).

The field equations are

$$A_{\lambda,\mu} - A_{\mu,\lambda} = 0\,,$$
$$iA_{,\lambda} = mA_\lambda\,,$$
$$i\mathscr{F}^{\lambda\mu}_{,\mu} = mj^\lambda\,, \tag{47}$$
$$j^\lambda_{,\lambda} = 0\,.$$

As before the field equations contain redundancy—$\mathscr{F}^{\lambda\mu}$ and A may be chosen quite arbitrarily.

9. DISCUSSION

Coordinate invariance implies that field equations cannot all be independent but must be weakened by Bianchi-type identities. In the present case there are four such identities which may be obtained in the standard way[5] from the invariance of $\int\mathscr{L}\,d^4x$. Alternatively they may be derived directly from Eq. (15)–(17). The identities are

$$(\psi^\dagger_{,\iota} + \psi^\dagger_{,\kappa}C^{\dagger\kappa}_\iota + \psi^\dagger C^{\dagger\kappa}_\iota\partial/\partial x^\kappa)(i\alpha^\lambda\psi_{,\lambda} - \beta\psi)$$
$$+ \text{complex conjugate} = 0\,. \tag{48}$$

Equation (48) holds for any ψ, regardless of whether it satisfies Eq. (7). We have $2N$ unknowns, the real and imaginary parts of the components of ψ. However because of the identities there are not $2N$ functionally independent field equations but only $2N-4$, leaving us with four degrees of freedom. Such degrees of freedom correspond to the fact that if ψ is a solution of Eq. (7) so is ψ' generated from ψ by a coordinate change $x^\lambda\to x'^\lambda(x^\mu)$, which brings in four arbitrary functions. In order to obtain a definite solution one needs to supplement Eq. (7) by four coordinate conditions, just as for the Einstein equations.[6]

The field equations found in Sec. 8 are unsatisfactory in that they contain *too much* redundancy. Imposing four coordinate conditions either yields an empty theory or else still leaves some functions arbitrary. It seems likely that to obtain a nontrivial system one must depart from the representation condition, Eq. (23). This condition was not imposed *a priori* but turned out to be satisfied in all solutions of Eq. (15)–(17) found to date. Its origin lies in the condition $\bar{\alpha}_\kappa = \alpha^\kappa$ which served to close the algebra generated by $\alpha_\kappa,\alpha^\lambda$ and thereby obtain a solution of Eq. (30) via Duffin–Kemmer theory. Equation (31) is then automatically satisfied leading to Eq. (23). One needs to investigate other ways to close the algebra of $\alpha_\kappa,\alpha^\lambda$. More generally we need to construct an algebra out of α^λ, C^κ_ι, $P^{\kappa\mu}_{\iota\lambda}$ without choosing $P^{\kappa\mu}_{\iota\lambda} = 0$.

Finally let us note that it is not within the spirit of the present paper to use a metric tensor to construct the field equations. Subject to first finding nontrivial representations of Eqs. (15)–(17), all physical fields including the metric tensor are to be sought via the prescription of Sec. 5C, Eqs. (25) and (26).

APPENDIX

Proof of Lemma 1: We are given $\alpha^lP = P^\dagger\alpha^l$, $l = 0,1,2,3,5$ with $\beta = \alpha^5$, where α^l is $\alpha\beta$-irreducible. Suppose P has r zero eigenvalues so that it may be transformed to the form

$$\begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix},$$

where P_1 is an $r\times r$ matrix all of whose eigenvalues are zero, and P_2 is a nonsingular $(N-r)\times(N-r)$ matrix. The rth power of P_1 is then zero. Let us partition α^l in the same way as P,

$$\alpha^l = \begin{pmatrix} \alpha^l_{11} & \alpha^l_{12} \\ \alpha^l_{21} & \alpha^l_{22} \end{pmatrix}.$$

Then the identity $\alpha^l(P)^r = (P^\dagger)^r\alpha^l$ yields $\alpha^l_{12}(P_2)^r = 0$ and hence $\alpha^l_{12} = 0$. Whence α^l decomposes into the direct sum of $r\times r$ and $(N-r)\times(N-r)$ matrices, contrary to the hypothesis of irreducibility. Hence either $r = N$ and all the eigenvalues of P are zero, or $r = 0$ and P^{-1} exists.

If we know that P has a real eigenvalue λ then $\alpha^l(P-\lambda) = (P-\lambda)^\dagger\alpha^l$ implies that all the eigenvalues of $P-\lambda$ are zero since $P-\lambda$ is certainly singular. In this case all the eigenvalues of P are λ. Finally, if P is also diagonalizable, then $P = \lambda I$.

Proof of Lemma 2: Equation (16) may be written

$$\beta C^\kappa_\iota = (-C^{\dagger\kappa}_\iota - \delta^\kappa_\iota)\beta\,. \tag{49}$$

If C^κ_ι is irreducible in the ordinary sense then so is $-C^{\dagger\kappa}_\iota - \delta^\kappa_\iota$. We can now apply Schur's Lemma[7] and deduce that either $\beta = 0$ or β^{-1} exists.

Proof of Lemma 3: Given that β^{-1} exists let us define $\gamma^\lambda = \beta^{-1}\alpha^\lambda$. Then Eqs. (15) and (17) become

$$\gamma^\lambda C^\kappa_\iota + \gamma^\kappa C^\lambda_\iota = 0\,, \tag{50}$$

$$\gamma^\lambda C^\kappa_\iota - C^\kappa_\iota\gamma^\lambda = \gamma^\kappa\delta^\lambda_\iota\,. \tag{51}$$

In particular

$$C^0_0\gamma^0 + \gamma^0 = \gamma^0 C^0_0 = 0\,.$$

Multiplying through on the left by γ^0 then gives $(\gamma^0)^2 = 0$. Similar manipulation yields $\gamma^0\gamma^1 + \gamma^0\gamma^1 = 0$ and in general

$$\gamma^\kappa\gamma^\lambda + \gamma^\lambda\gamma^\kappa = 0\,. \tag{52}$$

On multiplying Eq. (7) on the left by β^{-1} we obtain

$$i\gamma^\lambda\psi_{,\lambda} = \psi\,. \tag{53}$$

Let us now multiply Eq. (53) by $\gamma^0\gamma^1\gamma^2\gamma^3$. All terms on the left are zero leading to $\gamma^0\gamma^1\gamma^2\gamma^3\psi = 0$. Now multiply Eq. (53) by $\gamma^1\gamma^2\gamma^3$ to give $\gamma^1\gamma^2\gamma^3\psi = 0$, then by $\gamma^2\gamma^3$ to give $\gamma^2\gamma^3\psi = 0$ and so on. Finally one obtains $\psi = 0$.

Proof of Lemmas 4, 5, *and* 6: Lemmas 4 and 5 are best proved by direct substitution of the assumed forms for C^κ_ι into Eqs. (15)–(17), and Lemma 6 by direct evaluation. The algebra is straightforward but too lengthy to reproduce here.

[1] M. Hammermesh, *Group Theory and its Applications to Physical Problems* (Addison–Wesley, Reading, MA, 1962), p. 389.
[2] R. J. Duffin, Phys. Rev. **54**, 1114 (1938).
[3] N. Kemmer, Proc. R. Soc. London Ser. A **173**, 91 (1939).
[4] E. M. Corson, *Introduction to Tensors, Spinors, and Relativistic Wave-Equations* (Blackie, London, 1955), p. 184.
[5] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison–Wesley, Cambridge, MA, 1951), p. 293.
[6] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 161.
[7] H. Boerner, *Representations of Groups* (North–Holland, Amsterdam, 1963), p. 22.

Similarity solutions for the Ernst equations with electromagnetic fields

P. Kaliappan[a)]

Department of Physics, Autonomous Post-Graduate Centre, University of Madras, Tiruchirapalli-620 020, India

M. Lakshmanan

Department of Physics, Autonomous Post-Graduate Centre, University of Madras, Tiruchirapalli-620 020, India and Department of Mathematics, The University of Manchester Institute of Science and Technology, P. O. Box 88, Manchester M60 1QD, England

(Received 21 February 1980; accepted for publication 25 June 1980)

A class of exact similarity solutions is shown to exist for the Ernst equations with electromagnetic fields. We reduce the underlying coupled nonlinear partial differential equations for the potentials into a system of coupled ordinary nonlinear differential equations for the similarity variables. The reduced system is exactly solvable in terms of elementary functions. These solutions are bounded everywhere.

PACS numbers: 02.30.Jr

1. INTRODUCTION

It has been shown by Ernst[1] that any axially-symmetric stationary solution of Einstein's vacuum field equations may be described in terms of a single complex function ϵ, which is independent of the azimuth and satisfies the equation

$$(\mathrm{Re}\,\epsilon)\vec{\nabla}^2\epsilon = \vec{\nabla}\epsilon\cdot\vec{\nabla}\epsilon\,, \tag{1}$$

where

$$\vec{\nabla}^2\left(=\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{\partial^2}{\partial z^2}\right)$$

stands for the three Laplacian in axisymmetric coordinates. Equation (1) is also of interest because the SU(2) gauge field equations on E^4 can be reduced to a set of three Laplace equations,[2] which can equivalently be formulated[3] for static-axisymmetric fields in a form analogous to (1). Further, Ernst[4] had also proved that any axially-symmetric stationary solution of the coupled Einstein–Maxwell equations may also be described in terms of a pair of complex functions ϵ and Ψ, both of which are independent of the azimuth. They satisfy a set of coupled nonlinear partial differential equations of the form

$$(\mathrm{Re}\,\epsilon+|\Psi|^2)\vec{\nabla}^2\epsilon = (\vec{\nabla}\epsilon+2\Psi^*\vec{\nabla}\Psi)\cdot\vec{\nabla}\epsilon\,, \tag{2a}$$

$$(\mathrm{Re}\,\epsilon+|\Psi|^2)\vec{\nabla}^2\Psi = (\vec{\nabla}\epsilon+2\Psi^*\vec{\nabla}\Psi)\cdot\vec{\nabla}\Psi\,. \tag{2b}$$

For several purposes it is convenient to reexpress ϵ and Ψ in terms of the following quantities:

$$\epsilon=(\xi-1)/(\xi+1)\,,\quad \Psi=\eta/(\xi+1)\,. \tag{3}$$

Then Eq. (2) becomes

$$(\xi\xi^*-1)\vec{\nabla}^2\xi = 2\xi^*(\vec{\nabla}\xi)^2\,, \tag{4}$$

while the coupled set of equations (2) become

$$(\xi\xi^*+\eta\eta^*-1)\vec{\nabla}^2\xi = 2\vec{\nabla}\xi\cdot(\xi^*\vec{\nabla}\xi+\eta^*\vec{\nabla}\eta)\,, \tag{5a}$$

$$(\xi\xi^*+\eta\eta^*-1)\vec{\nabla}^2\eta = 2\vec{\nabla}\eta\cdot(\xi^*\vec{\nabla}\xi+\eta^*\vec{\nabla}\eta)\,. \tag{5b}$$

Extensive investigations on the structure and properties of Eqs. (4) and (5) have been carried out by many authors during the past few years.[5-8] From these investigations many interesting classes of exact solutions, such as the Kerr, Weyl, Tomimatsu–Sato classes, have been obtained and new techniques of generating further classes of solutions have emerged. Very recent investigations on the single potential equation (4) also show strong evidence about its complete integrability. Maison[9] had proved the existence of a Lax pair, while Harrison[10] and also Neugebauer[7] have found Bäcklund transformations for this equation, Belinskii and Zakharov[11] have applied the inverse scattering method to construct soliton solutions starting from a given base solution. It would be interesting to see how these inverse scattering techniques are extended to the coupled equations (5).

From a different point of view, these equations are also known to possess certain maximal symmetry groups.[8,12] The analysis of these symmetry groups, which correspond to invariance under infinitesimal Lie transformations, also suggests the existence of a similarity variable of the form

$$\zeta = z/\rho\,. \tag{6}$$

It is the purpose of this paper to show that systems (4) and (5) are exactly solvable in terms of the similarity variable (6) and that the solutions are expressible in terms of elementary functions of ζ, involving a minimal set of integration constants. Further these solutions are finite everywhere in the three-dimensional space. To our knowledge the type of solutions we report seem to be new and could be of some relevance. We might also mention that the existence of similarity systems having no movable critical points (which is the case for the equations under consideration here) seems to be an additional evidence for the complete integrability from the point of view of Ablowitz *et al.*'s[13] conjecture on the connection between nonlinear partial differential equations solvable by inverse scattering transform method and ordinary differential equations with no movable critical points. This was part of our original motiva-

[a)]On leave from N. G. M. College, Pollachi-642001, Tamilnadu, India.

 0022-2488/81/112447-05$01.00

tion for the search for similarity solutions to these systems.

The paper is organized as follows. In Sec. 2 of this paper, we reduce systems (4) and (5) to a set of coupled ordinary differential equations in terms of the similarity variable ζ and in Sec. 3 we solve this system explicitly. The solutions are expressed in terms of elementary functions involving a minimal set of integration constants and these are finite for all values of spatial variables and tend to constant values asymptotically. In the final section we give a brief discussion of the results and the analogy of present systems with Heisenberg's ferromagnetic spin system.

2. SIMILARITY FORM

Equation (4) is a special case of (5) and therefore we will discuss mostly the general case (5) and relegate the discussion on (4) to the end. After some manipulations, it is found that the following variable

$$r = \ln[\sqrt{1+\zeta^2} - \zeta] \equiv \ln[\sqrt{1+z^2/\rho^2} - z/\rho]\,, \tag{7a}$$

or

$$\zeta = -\sinh r \tag{7b}$$

is more convenient to work with, instead of ζ itself. Accordingly, Eq. (5) reduces to the set of coupled ordinary differential equations of the form

$$(\xi\xi^* + \eta\eta^* - 1)\frac{d^2\xi}{dr^2} = 2\xi^*\left(\frac{d\xi}{dr}\right)^2 + 2\eta^*\frac{d\xi}{dr}\cdot\frac{d\eta}{dr} \tag{8a}$$

and

$$(\xi\xi^* + \eta\eta^* - 1)\frac{d^2\eta}{dr^2} = 2\eta^*\left(\frac{d\eta}{dr}\right)^2 + 2\xi^*\frac{d\xi}{dr}\cdot\frac{d\eta}{dr}\,. \tag{8b}$$

We now make the following transformations (see also discussion in Sec. 4):

$$\xi = \kappa\exp\left[i\int\tau\,dr\right] \tag{9a}$$

and

$$\eta = p\exp\left[i\int q\,dr\right], \tag{9b}$$

where κ, τ, p, and q are real functions of r, and substitute in the set of equations (8). Equating real and imaginary parts, we obtain the following set of four coupled ordinary differential equations:

$$(\kappa^2+p^2-1)(\kappa'' - \kappa\tau^2) = 2\kappa(\kappa'^2 - \kappa^2\tau^2) + 2p(\kappa'p' - \kappa\tau pq)\,, \tag{10a}$$

$$(\kappa^2+p^2-1)(2\kappa'\tau + \kappa\tau') = 4\kappa^2\kappa'\tau + 2p(\kappa' pq + p'\kappa\tau) \tag{10b}$$

and

$$(\kappa^2+p^2-1)(p'' - pq^2) = 2p(p'^2 - p^2q^2) + 2\kappa(\kappa'p' - \kappa\tau pq)\,, \tag{11a}$$

$$(\kappa^2+p^2-1)(2p'q + pq') = 4p^2p'q + 2\kappa(p'\kappa\tau + \kappa' pq)\,, \tag{11b}$$

where prime stands for differentiation with respect to r. Now multiplying Eq. (10b) by κ and Eq. (11b) by p and adding we can show that

$$\frac{d(\kappa^2\tau + p^2q)}{(\kappa^2\tau + p^2q)} = \frac{2(2\kappa\kappa' + 2pp')}{(\kappa^2+p^2-1)}\,. \tag{12}$$

Integrating Eq. (12), we obtain

$$\kappa^2\tau + p^2q = (\kappa^2+p^2-1)^2C\,, \tag{13}$$

where C is a constant of integration. Similarly calculating $\{(10b)\times\kappa - (11b)\times p\}$, we find that

$$d\left\{\frac{\kappa^2\tau - p^2q}{(\kappa^2+p^2-1)}\right\} = C\cdot d(\kappa^2 - p^2) \tag{14}$$

and therefore

$$\kappa^2\tau - p^2q = (\kappa^2+p^2-1)\{(\kappa^2-p^2)C + D\}\,, \tag{15}$$

where D is the second integration constant. From (13) and (15), we obtain

$$\tau = (1/2\kappa^2)[(\kappa^2+p^2-1)\{(2\kappa^2-1)C + D\}] \tag{16}$$

and

$$q = (1/2p^2)[(\kappa^2+p^2-1)\{(2p^2-1)C - D\}]\,. \tag{17}$$

Now Eq. (10a) can be rewritten as

$$(\kappa^2+p^2-1)\kappa'' - 2\kappa\kappa'^2 - 2pp'\kappa' = (\kappa^2+p^2-1)\kappa\tau^2 - 2\kappa^3\tau^2 - 2\kappa\tau p^2q\,. \tag{18}$$

Substituting the expressions (16) and (17) in the right-hand side of Eq. (18), and after some manipulations, we obtain

$$(\kappa^2+p^2-1)\kappa'' - 2\kappa\kappa'^2 - 2pp'\kappa' = [(\kappa^2+p^2-1)^3/(4\kappa^3)]C^2[(1-D/C)^2 - 4\kappa^4]\,. \tag{19}$$

This can be reexpressed in the form

$$\frac{\kappa'}{(\kappa^2+p^2-1)}\,d\left\{\frac{\kappa'}{(\kappa^2+p^2-1)}\right\} = -(C-D)^2d\left(\frac{1}{8\kappa^2}\right) - C^2d\left(\frac{\kappa^2}{2}\right). \tag{20}$$

Integrating Eq. (20), we obtain

$$\frac{(\kappa')^2}{2(\kappa^2+p^2-1)^2} = -(C-D)^2\cdot\frac{1}{8\kappa^2} - C^2\frac{\kappa^2}{2} + E\,, \tag{21}$$

where E is the third integration constant. Proceeding in an analogous manner, Eq. (11a) can be integrated to obtain

$$\frac{(p')^2}{2(\kappa^2+p^2-1)^2} = -(C+D)^2\cdot\frac{1}{8p^2} - \frac{C^2p^2}{2} + F\,, \tag{22}$$

where F is the fourth integration constant.

We now make the following change of variable

$$\kappa^2 = x\,,\quad p^2 = y \tag{23}$$

and combine Eqs. (21) and (22) to get the following differential equation:

$$dy/[((4F^2 - (C+D)^2C^2)/(4C^4)) - (y - F/C^2)^2]^{1/2} = dx/[((4E^2 - (C-D)^2C^2)/(4C^4)) - (x - E/C^2)^2]^{1/2}\,. \tag{24}$$

On integrating, we obtain

$$y = l\left(x - \frac{E}{C^2}\right)(1-G^2)^{1/2} + \frac{Gl}{2C}\{8Ex - 4C^2x^2 - (C-D)^2\}^{1/2} + \frac{F}{C^2}\,, \tag{25a}$$

where

$$l=\{4F^2-(C+D)^2C^2\}^{1/2}/\{4E^2-(C-D)^2C^2\}^{1/2}. \quad (25b)$$

Here G is the fifth integration constant. Then Eq. (21) can be rewritten in the form

$$\frac{dx}{dr}=\left[\left(\frac{F}{C^2}-\frac{lE}{C^2}(1-G^2)^{1/2}-1\right)+x(1+l(1-G^2)^{1/2})+\frac{lG}{2C}\{8Ex-4C^2x^2-(C-D)^2\}^{1/2}\right][8Ex-4C^2x^2-(C-D)^2]^{1/2}, \quad (26)$$

which is the final similarity equation to be solved.

3. SOLUTIONS FOR THE ERNST POTENTIALS

To obtain the explicit solutions of Eq. (26), it is more advantageous to redefine the set of five integration constants $\{C,D,E,F,G\}$ in terms of a new set $\{x_1,x_2,\nu,A,\alpha\}$ in the following way. Let

$$x_1=\frac{E}{C^2}\left[1-\left\{1-\frac{C^2}{4E^2}(C-D)^2\right\}^{1/2}\right],$$

$$x_2=\frac{E}{C^2}\left[1+\left\{1-\frac{C^2}{4E^2}(C-D)^2\right\}^{1/2}\right],\quad x_1x_2=\frac{(C-D)^2}{4C^2}$$

$$(0\le x_1\le x_2) \quad (27)$$

so that x_1 and x_2 are the two roots of the quadratic equation

$$8Ex-4C^2x^2-(C-D)^2=0. \quad (28)$$

This enables us to make an Euler's transformation of Eq. (26) in the form

$$x=(x_1t^2+x_2)/(t^2+1) \quad (29a)$$

so that

$$\{8Ex-4C^2x^2-(C-D)^2\}^{1/2}=(2C(x_2-x_1)t)/(t^2+1). \quad (29b)$$

This, with the aid of three further redefined constants

$$\nu=C\left[\left(\frac{F}{C^2}-\frac{lE}{C^2}(1-G^2)^{1/2}-1\right)+x_1(1+l(1-G^2)^{1/2})\right], \quad (30a)$$

$$\alpha^2=\frac{l^2G^2C^2(x_2-x_1)^2}{\nu^2}, \quad (30b)$$

and

$$A^2=\alpha^2-\left[\left(\frac{F}{C^2}-\frac{lE}{C^2}(1-G^2)^{1/2}-1\right)+x_2(1+l(1-G^2)^{1/2})\right]\cdot\frac{C}{\nu}, \quad (30c)$$

enables us further to rewrite Eq. (26) as

$$dt/dr=-\nu[(t+\alpha)^2-A^2]. \quad (31)$$

On integrating, the solution becomes

$$t=-\alpha+A\coth(A\nu r+\tfrac{1}{2}\ln B), \quad (32)$$

where now B is the sixth integration constant. Using the original definition of the variable r, i.e. Eqs. (7) and (32) can be reexpressed in the form

$$t=\frac{(A+\alpha)+(A-\alpha)\cdot B[(1+z^2/\rho^2)^{1/2}-z/\rho]^{2\nu A}}{\{B[(1+z^2/\rho^2)^{1/2}-z/\rho]^{2\nu A}-1\}}. \quad (33)$$

Thus we finally obtain

$$\kappa=[(x_1t^2+x_2)/(t^2+1)]^{1/2}, \quad (34)$$

where t is defined as in Eq. (33). It essentially involves six constants A, B, ν, α, x_1, and x_2, pertaining to the underlying symmetries of the system under consideration. As we noted already these constants are directly related to the six explicit integrations we have carried out so far.

Finally, to obtain the Ernst potentials ξ and η, we need the expression $\exp[i\int\tau\,dr]$. This is easily obtained by using expression (16) for τ and the fact that

$$(x+y-1)=-\frac{dt}{dr}\cdot\frac{1}{C(t^2+1)}. \quad (35a)$$

and the various relations (27) and (29), so that

$$\int\tau\,dr=-(x_1x_2)^{1/2}\int\frac{dt}{(x_1t^2+x_2)}+\int\frac{dt}{t^2+1}. \quad (35b)$$

Thus

$$\exp\left[i\int\tau\,dr\right]=\exp\{+i[\tan^{-1}t-\tan^{-1}t(x_1/x_2)^{1/2}]\}+H$$

$$=\exp i\tan^{-1}\left[\frac{(\sqrt{x_2}-\sqrt{x_1})t}{(\sqrt{x_2}+\sqrt{x_1}\,t^2)}\right]\cdot e^{iH}, \quad (36)$$

where the seventh integration constant H could be identified with the NUT (Newman, Tamburino and Unti) parameter corresponding to the fact that if $\xi=\xi_0(r)$ is a solution so also is $\xi=\xi_0(r)\cdot e^{iH}$. Defining now

$$\varphi=\left[\frac{(\sqrt{x_2}-\sqrt{x_1})t}{\sqrt{x_2}+\sqrt{x_1}\,t^2}\right], \quad (37)$$

where t is explicitly given in (33), we obtain

$$\exp\left[i\int\tau\,dr\right]=(1+i\varphi)^{1/2}/(1-i\varphi)^{1/2}. \quad (38)$$

Combining (34) and (38) through the relation $\xi=\kappa\exp[i\int\tau\,dr]$, we finally obtain the first of the two Ernst potentials explicitly in terms of a minimal set of seven constants $\{x_1, x_2, A, B, \nu, \alpha, \text{and } H\}$.

As the second order coupled complex differential equation (8) involves eight integration constants, the remaining one enters in the definition of the second of the Ernst potentials $\eta=p\exp[i\int q\,dr]$, as a second NUT parameter when the integration in the exponential is carried out. Here the quantities p and q are determined through the relations (25) and (17), respectively, in terms of κ and τ. Thus the complete explicit expressions for ξ and η are obtained, which are easily seen to be finite everywhere.

The solutions of the single potential equation (4) can be obtained by putting $\xi=\eta$ and scaling by a factor of $\frac{1}{2}$ in the above results. In this limit the various constants reduce to

$$G=0,\quad E=F,\quad D=0,\quad l=1\quad (x=y) \quad (39a)$$

so that

$$x_1x_2=\tfrac{1}{4},\quad \nu=C((1/2x_2)-1),\quad A^2=2x_2,\quad \alpha=0. \quad (39b)$$

Correspondingly, we have

$$2\kappa^2 = 2(x_1 t^2 + x_2)/(t^2+1) = \frac{\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} + 1\}^2 + A^2\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} - 1\}^2}{\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} - 1\}^2 + A^2\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} + 1\}^2}, \tag{40a}$$

while

$$\exp\left[i\int \tau\, dr\right] = [\{A^2 + t^2 + i(A^2-1)t\}^{1/2}/\{A^2 + t^2 - i(A^2-1)t\}^{1/2}]\cdot e^{iH}. \tag{40b}$$

From (32), the form of t for the present case is

$$t = A\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} + 1\}/\{B[(1+z^2/\rho^2)^{1/2} - z/\rho]^{2\nu A} - 1\}. \tag{40c}$$

Thus the solution of the Ernst equation with single potential ξ, Eq. (4), is the suitable product of the right-hand sides of (40a) and (40b) involving four arbitrary constants A, B, ν and H.

To compare the solution (40) with the well-known Weyl, Kerr, Tomimatsu–Sato, Ernst, etc., classes of solutions,[5,6] one may rewrite the above expressions in terms of the prolate-spheroidal coordinates $(\hat{x}, \hat{y})$ so that

$$\rho = [(\hat{x}^2 - 1)(1 - \hat{y}^2)]^{1/2}, \quad z = \hat{x}\hat{y} \tag{41a}$$

and so

$$[(1+z^2/\rho^2)^{1/2} - z/\rho] = [(\hat{x}^2 + \hat{y}^2 - 1) - \hat{x}\hat{y}]/[(\hat{x}^2-1)(1-\hat{y}^2)]^{1/2}. \tag{41b}$$

It appears that the solutions (40) as well as the solutions (34)–(38) of the two potential case are of different structure and encompass a new larger class in view of the freedom of choice of the different integration constants occurring in the solutions.

4. DISCUSSION

We have shown that the Ernst equations are exactly solvable in terms of the similarity variable (7) and that these solutions are bounded for all values of the spatial variables. The similarity solutions assume importance because it is generally expected that such invariant solutions correspond to the asymptotic behavior of general solutions as in the case of many integrable systems.[13] From our own point of view, the crucial step in our analysis is the complex transformations (9). It was shown by one of us recently[14] that a similar transformation leads to the mapping of the evolution equation of a Heisenberg ferromagnetic spin system on a nonlinear Schrödinger equation and thereby it is completely integrable. This equivalence has been demonstrated as a gauge equivalence between the two systems recently by Zakharov and Takhtajan.[15] In the higher dimensional static situations of the ferromagnetic spin system, similar transformations lead to the exact solvability of the spherical and circular cases,[16] while a class of interesting point defect solutions are obtained for the planar and axial symmetric situations. As the Ernst equation with single potential has close analogy[17] with the ferromagnetic spin system, we were motivated to apply similar transformations to Eqs. (4) and (5), which lead to the explicit bounded solutions discussed above. Further analysis of the physical significance of these solutions is presently in progress.

ACKNOWLEDGMENTS

The authors would like to thank an anonymous referee for constructive suggestions. It is a great pleasure to acknowledge the marvellous hospitality extended to M. L. by Professor R. K. Bullough during his stay at U.M.I.S.T. He also benefited from discussions with Professors R. K. Bullough and A. Degasperis, Dr. R.K. Dodd, Dr. H.C. Morris, and Dr. R. Sasaki. M. L. would also like to acknowledge with gratitude the financial support of the Royal Society in the form of a Commonwealth and Nuffield Foundation Bursary during his stay in U. K. P. K. acknowledges with gratitude the Principal and management of N. G. M. College, Pollachi, for granting study leave and the University Grants Commission for financial support under the Faculty Improvement Program.

Note added in proof: After the original submission of this paper, an article by Fischer[18] appeared which considers similarity solutions for the Ernst equation with a single potential but with an imaginary rotation of the z axis. Such a system corresponds to the Einstein equation for cylindrical gravitational waves, as well as the Einstein–Maxwell equations for colliding plane gravitational and plane electromagnetic waves. For the system (4) considered here a solution in terms of the variable $\rho^2 + z^2$ is also presented by him. We might point out that our solutions for this system itself are of a different nature and our analysis is concerned with the more complicated coupled system (5) and so is of much wider scope.

[1] F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[2] C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[3] L. Witten, Phys. Rev. D **19**, 718 (1979).
[4] F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[5] See for e.g., R. M. Misra, D. B. Pandey, D. C. Srivastava, and S. N. Tripathy, Phys. Rev. D **7**, 1587 (1973); A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973); C. Reina and A. Treves, J. Math. Phys. **16**, 834 (1975); M. Yamazaki and S. Hori, J. Math. Phys. **18**, 2502 (1977); **19**, 1376 (1978) and references therein.
[6] F. J. Ernst, J. Math. Phys. **15**, 1409 (1974); **17**, 1097 (1976); **18**, 233 (1977); I. Hauser and F. J. Ernst, J. Math. Phys. **19**, 1316 (1978); Phys. Rev. D **20**, 362, 1783 (1979).
[7] C. Cosgrove, J. Phys. A **10**, 1481, 2093 (1977); A **11**, 2389 2405 (1978); G. Neugebauer, J. Phys. A **12**, L67 (1979); A **13**, L19 (1980); K. C. Das, J. Phys. A **13**, L23 (1980).
[8] W. Kinnersley, J. Math. Phys. **18**, 1529 (1977); W. Kinnersley and D. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926 2037 (1978); **20**, 2526, 2530 (1979).

[9]D. Maison, Phys. Rev. Lett. **41**, 521 (1978); J. Math. Phys. **20**, 871 (1979). See also H. C. Morris and R. K. Dodd, Phys. Lett. **75 A**, 20 (1980).
[10]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978); Phys. Rev. D **20**, 1695 (1980).
[11]V. A. Belinskii and V. E. Zakharov, Sov. Phys.-JETP **48**, 985 (1978); **50**, 1 (1979); see also A. Tomimatsu, Prog. Theor. Phys. **63**, 1054 (1980).
[12]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968); C. Hoenselaers, J. Math. Phys. **17**, 1264 (1976); Y. Tanabe, Prog. Theor. Phys. **57**, 840 (1977); E. Leibowitz and J. Meinhardt, J. Phys. A **11**, 1579 (1978); Phys. Lett. **66** A, 1 (1978).
[13]M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977); M. J. Ablowitz, A. Ramani and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978); J. Math. Phys. **21**, 715, 1006 (1980); P. Kaliappan and M. Lakshmanan, J. Phys. A **12**, L249 (1979); M. Lakshmanan and P. Kaliappan, Phys. Lett. **71 A**, 166 (1979); M. Lakshmanan, P. Kaliappan, and M. Daniel, Phys. Lett. **75 A**, 97 (1979).
[14]M. Lakshmanan, Phys. Lett. **61** A, 53 (1977); see also M. Lakshmanan, Th. W. Ruijgrok, and C. J. Thompson, Physica, **84** A, 577 (1976); M. Lakshmanan, J. Math. Phys. **20**, 1667 (1979).
[15]V. E. Zakharov and L. A. Takhtajan, Theor. Math. Phys. **38**, 17 (1979).
[16]M. Lakshmanan, M. Daniel, and P. Kaliappan, J. Phys. C **13**, 4743 (1980).
[17]M. Hirayama, H. C. Tze, I. Ishida, and T. Kawabe, Phys. Lett. **66** A, 352 (1978).
[18]E. Fischer, J. Phys. A **13**, L81 (1980).

Bounds on Green's functions of second-order differential equations

M. van den Berg
Institute for Theoretical Physics, State University of Groningen, P. O. Box 800, Groningen, The Netherlands

(Received 24 December; accepted for publication 20 March 1981)

We estimate the diagonal part of the Green's function for the equation $(-\Delta/2 + V(x) + \partial/\partial t)\psi(x,t) = 0, t > 0, x \in B$, where B is a finite region of the Euclidean space R^d with a regular boundary. In the special case $V(x) = 0$, $x \in B$, we also obtain bounds for the non-diagonal part of the Green's function which are uniform in t.

PACS numbers: 02.30.Jn

I. INTRODUCTION

Kac, in unpublished lecture notes summarized in Ref. 1, indicates how Wiener estimates of the Green's function for the one-particle diffusion equation can be used to derive results about the bulk properties of the free boson gas. The mathematical details were supplied in Lewis and Pulè.[2] The same strategy has been used in van den Berg and Lewis[3] to prove results (announced in Ref. 4) about the boson gas in an external potential. For this purpose we require Wiener estimates of the Green's function for the one-particle diffusion equation with an external potential. These estimates may be of use in other fields of application and the purpose of this paper is to provide their proofs.

We estimate the Green's function of the partial differential equation

$$(L + \partial/\partial t)\psi(x,t) = 0, \quad x \in B, t > 0, \tag{1}$$

where B is a finite region of the Euclidean space R^d with a regular boundary ∂B. We will restrict ourselves to Dirichlet boundary conditions: $\psi(x,t) = 0$ for $x \in \partial B$. L denotes the self-adjoint operator on the space $L^2(B)$ which is given on smooth functions by the differential operator $-\Delta/2 + V(x)$ with Dirichlet boundary conditions where $V(x)$ is a non-negative function (satisfying a Lipschitz condition almost everywhere in B). This operator has a discrete positive spectrum $E_1 < E_2 \leqslant E_3 \cdots$ and an orthonormal set of eigenfunctions $\{\phi_j(x)\}$ forming a basis in $L_2(B)$ (Davies[5]). Furthermore, the Green's function of (1) has the eigenfunction expansion

$$K(x,y;t) = \sum_{j=1}^{\infty} \exp(-tE_j)\phi_j(x)\phi_j(y). \tag{2}$$

Moreover, it has been shown by Rosenblatt[6] and Ray[7] that $K(x,y;t)$ can be written as

$$K(x,y;t) = \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \times \mathrm{E}\{\exp\left[-\int_0^t V(u(\tau))d\tau\right] : u(0) = x,\ u(t) = y;\ u(\tau) \in B\}, \tag{3}$$

where the quantity

$$\mathrm{E}\{\exp\left[-\int_0^t V(u(\tau))d\tau\right] : u(0) = x,\ u(t) = y;\ u(\tau) \in B\}$$

denotes the average value of

$$\exp\left[-\int_0^t V(u(\tau))d\tau\right]$$

for all paths $u(\cdot)$ of a Wiener process on R^d subject to $u(0) = x$ and $u(t) = y$. Furthermore Ray[7] proved that $K(x,y;t)$ is continuous (for all y) at a point x_0 of the boundary ∂B provided there exists a conical sector with vertex at x_0 entirely outside B. We will assume that this condition holds for all points on the boundary; we call such a boundary regular. Ray[7] has, in addition, results in the case in which B is an unbounded region of R^d and $V(x) \to \infty$ as $|x| \to \infty$; we will restrict ourselves, however, to the case in which B is a bounded region.

Our main result is that for t small

$$K(x,x;t) \sim e^{-tV(x)}/(2\pi t)^{d/2} \tag{4}$$

at all points x which are not too close to the boundary. Expression (4) can easily be understood from formula (3). For small times t the probability is small that $|x - u(\tau)|$ is large, so we may replace $u(\tau)$ by x and B by R^d, provided x is not too close to ∂B. This has been called the principle of not feeling the boundary.[8] Integrating both sides of (4) with respect to the volume we have, for t small,

$$\sum_{j=1}^{\infty} \exp(-tE_j) \sim \frac{1}{(2\pi t)^{d/2}} \int_{x \in B} e^{-tV(x)}\, dx. \tag{5}$$

This is allowed, since most points x are far from the boundary because the boundary is regular.

In Sec. 2 we will estimate the Green's function of the differential equation (1) for the special case $V(x) = 0, x \in B$. In Secs. 3 and 4 we will calculate bounds on the correction terms in (4) and (5).

2. A UNIFORM ESTIMATE WHEN V IS IDENTICALLY ZERO

Theorem 1: Let $K_0(x,y;t)$ be the Green's function of Eq. (1) with $V(x) = 0$, $x \in B$, and with Dirichlet boundary conditions for $\psi(x,t)$ at ∂B.
Then

$$\left| K_0(x,y;t) - \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \right| \leqslant \frac{2d}{(2\pi t)^{d/2}} \exp\left((4\sqrt{2} - 6)\frac{d_x^2}{dt}\right), \quad x \in B, y \in B, t > 0, \tag{6}$$

where d_x is the distance of x from ∂B.

Proof: Let $\square_x$ denote a hypercube with center x which lies entirely inside B and such that at least one corner vertex lies on ∂B. Then the length l_x of an edge of the cube is not less then $(2/d^{\frac{1}{2}})d_x$. From (3) we have

 0022-2488/81/112452-04$01.00

$$0 \leqslant K_0(x,y;t) = \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \times \mathrm{E}\{1{:}u(0)=x,\ u(t)=y;\ u(\tau)\in B\}$$
$$\leqslant \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}}\,\mathrm{E}\{1{:}u(0)=x,\ u(t)=y;\ u(\tau)\in R^d\} = \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}}. \tag{7}$$

We consider two cases:

(i) $|x-y| \geqslant \alpha l_x$, where $\alpha \in [0,\frac{1}{2})$. (We will choose α later). In this case we have by (7) the bound

$$\left| K_0(x,y;t) - \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \right| \leqslant \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \leqslant \frac{\exp(-\alpha^2 l_x^2/2t)}{(2\pi t)^{d/2}}. \tag{8}$$

(ii) $|x-y| \leqslant \alpha l_x$, where $\alpha \in [0,\frac{1}{2})$. It is obvious that $y \in \square_x$ since $\alpha \leqslant \frac{1}{2}$. Because of the inequality (7) we have only to derive a lower bound for $K_0(x,y;t)$

$$K_0(x,y;t) \geqslant \frac{\exp(-|x-y|^2/2t)}{(2\pi t)^{d/2}} \times \mathrm{E}\{1{:}u(0)=x,\ u(t)=y;\ u(\tau)\in \square_x\}. \tag{9}$$

Denote the right-hand side of (9) by $K_\square(x,y;t)$; $K_\square(x,y;t)$ is the Green's function for a cube with edges with lengths l_x and center x. We have the following explicit expression if we choose x as the origin and the rectangular coordinate frame parallel to the edges of the cube:

$$K_\square(0,y;t) = \prod_{i=1}^{d}\left(\frac{1}{l_x} \sum_{k=-\infty}^{+\infty} \exp\left[-\frac{\pi^2 t}{2l_x^2}(2k+1)^2 \right] \cos\frac{(2k+1)\pi y_i}{l_x} \right).$$

With the help of the Poisson formula[9] we obtain

$$K_\square(0,y;t) = \prod_{i=1}^{d} \frac{\exp(-y_i^2/2t)}{(2\pi t)^{\frac{1}{2}}} \times \left\{ 1 + 2\sum_{k=1}^{\infty} (-)^k \exp\left(-\frac{k^2 l_x^2}{2t} \right) \cosh\frac{k y_i l_x}{t} \right\}.$$

The terms of the alternating series in k are decreasing provided $|y_i| < l_x/2$. The term $k=1$ is negative so the sum is also negative but larger than or equal to -1, since the Green's function is non-negative. With the use of

$$\prod_{i=1}^{d}(1+a_i) \geqslant 1 + \sum_{i=1}^{d} a_i, \quad -1 \leqslant a_i \leqslant 0,\ d = 1,2\cdots$$

we have

$$K_\square(x,y;t) \geqslant \frac{\exp(-y^2/2t)}{(2\pi t)^{d/2}} \times \left[1 + 2\sum_{i=1}^{d}\sum_{k=1}^{\infty} (-)^k \exp\left(-\frac{k^2 l_x^2}{2t} \right) \cosh\frac{k y_i l_x}{t} \right]$$
$$\geqslant \frac{\exp(-y^2/2t)}{(2\pi t)^{d/2}} \left[1 - 2\sum_{i=1}^{d} \exp\left(-\frac{l_x^2}{2t} \right) \cosh\frac{y_i l_x}{t} \right]$$
$$\geqslant \frac{\exp(-y^2/2t)}{(2\pi t)^{d/2}} \left[1 - 2\sum_{i=1}^{d} \exp\left(-\frac{l_x^2}{2t} + \frac{|y_i| l_x}{t} \right) \right]$$
$$\geqslant \frac{\exp(-y^2/2t)}{(2\pi t)^{d/2}} \left[1 - 2d\exp\left(-\frac{l_x^2}{t}(\tfrac{1}{2}-\alpha) \right) \right], \tag{10}$$

since $|y| \leqslant \alpha l_x$. Now, we choose α to be the positive root of $\frac{1}{2} - \alpha = \alpha^2/2$. So $\alpha = -1 + \sqrt{2}$, which is also less than $\frac{1}{2}$. Combining the results (7), (8), and (9) we arrive at Theorem 1.

If $x = y$ we may choose $\alpha = 0$ and we have

$$\left| K_0(x,x;t) - \frac{1}{(2\pi t)^{d/2}} \right| \leqslant \frac{2d}{(2\pi t)^{d/2}} \exp\left(-\frac{2d_x^2}{dt} \right), \tag{11}$$

which is a stronger inequality than one would obtain from Theorem 1 by putting $x = y$. Notice that Theorem 1 is a stronger result than that of Arima[10] since it is uniform in t. On the other hand we have obtained it only in the case of Dirichlet boundary conditions.

3. THE MAIN ESTIMATE

Theorem 2: Let B be a finite region in R^d with a regular boundary ∂B and let $V(x)$ be a non-negative Borel-measurable function defined on B. Let $V(x)$ satisty a local Lipschitz condition of the form

$$|V(x) - V(x')| < M(x)\cdot|x-x'|^\alpha, \quad 0 < \alpha \leqslant 1$$

for almost all pairs x, x' in B; then

$$\left| \frac{e^{-tV(x)}}{(2\pi t)^{d/2}} - K(x,x;t) \right|$$
$$\leqslant \frac{e^{-tV(x)}}{(2\pi t)^{d/2}} \left\{ 2d\exp\left(-\frac{2d_x^2}{dt} \right) + \left| \int_0^t d\tau \int_{u\in B} du\,(V(u)-V(x)) \left(\frac{t}{2\pi\tau(t-\tau)} \right)^{d/2} \exp\left(-\frac{t|x-u|^2}{2\tau(t-\tau)} \right) \right| \right.$$
$$\left. + \left| \int_0^t \frac{d\tau}{t} \int_{u\in B} du \left(\frac{t}{2\pi\tau(t-\tau)} \right)^{d/2} \exp\left(t(V(x)-V(u)) - \frac{t|x-u|^2}{2\tau(t-\tau)} \right) - 1 \right| \right\}, \tag{12}$$

at all points x where the Lipschitz condition holds.

Proof: From the theorems of Ray[7] and Rosenblatt[6] it follows that

$$\frac{e^{-tV(x)}}{(2\pi t)^{d/2}} - K(x,x;t) = \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\left\{\mathbb{E}\left\{1 - \exp\left[-\int_0^t [V(u(\tau)) - V(x)]\,d\tau\right] : u(0) = u(t) = x; u(\tau)\in B\right\}\right.$$
$$\left. + 1 - \mathbb{E}\{1 : u(0) = u(t) = x; u(\tau)\in B\}\right\}$$
$$\leqslant \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\left\{\mathbb{E}\left\{\int_0^t [V(u(\tau)) - V(x)]\,d\tau : u(0) = u(t) = x; u(\tau)\in B\right\}\right.$$
$$\left. + 2d\exp\left(-\frac{2d_x^2}{dt}\right)\right\} = \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\left\{\int_0^t d\tau \int_{u\in B} du\,(V(u) - V(x))\cdot\left(\frac{t}{2\pi t(t-\tau)}\right)^{d/2}\exp\left(-\frac{t|x-u|^2}{2\tau(t-\tau)}\right)\right.$$
$$\left. + 2d\exp\left(-\frac{2d_x^2}{dt}\right)\right\} \leqslant \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\left\{\left|\int_0^t d\tau\int_{u\in B} du\,(V(u) - V(x))\right.\right.$$
$$\left.\left.\cdot\left(\frac{t}{2\pi t(t-\tau)}\right)^{d/2}\exp\left(-\frac{t|x-u|^2}{2\tau(t-\tau)}\right)\right| + 2d\exp\left(-\frac{2d_x^2}{dt}\right)\right\}, \tag{13}$$

by the inequality $1 - e^{-x} \leqslant x$, Theorem 1, and Fubini's theorem. Moreover

$$K(x,x;t) = \frac{1}{(2\pi t)^{d/2}}\mathbb{E}\left\{\exp\left[-\int_0^t V(u(\tau))\,d\tau\right] : u(0) = u(t) = x; u(\tau)\in B\right\}$$
$$\leqslant \frac{1}{(2\pi t)^{d/2}}\mathbb{E}\left\{\int_0^t \frac{d\tau}{t}\, e^{-tV(u(\tau))} : u(0) = u(t) = x; u(\tau)\in B\right\}$$
$$\leqslant \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\int_0^t \frac{d\tau}{t}\int_{u\in B}\exp\left(-t(V(u) - V(x)) - \frac{t|x-u|^2}{2\tau(t-\tau)}\right)\left(\frac{t}{2\pi\tau(t-\tau)}\right)^{d/2}, \tag{14}$$

by Jensen's inequality and Fubini's theorem. Combining (13) and (14) we have estimate (12).

4. AN ESTIMATE FOR THE PARTITION FUNCTION

In this section we estimate the correction term in (5). This can be done by simply integrating the inequalities (13) and (14) from which an estimate follows. However, due to an inequality of Ray[7] the result can be improved.

Theorem 3: Let $V(x)$ and B be as in Theorem 2, then

$$\left|\sum_{j=1}^{\infty}\exp(-tE_j) - \int_{x\in B} dx\,\frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\right| \leqslant \frac{1}{(2\pi t)^{d/2}}\int_{x\in B} dx\cdot e^{-tV(x)}\left\{2d\exp(-2d_x^2/dt)\right.$$
$$\left. + \left|\int_o^t d\tau\int_{u\in B} du\,(V(u) - V(x))\exp\left(-\frac{t|x-u|^2}{2\tau(t-\tau)}\right)\cdot\left(\frac{t}{2\pi\tau(t-\tau)}\right)^{d/2}\right|\right\}. \tag{15}$$

Proof: We first prove an upper bound (Ray[7]).

$$\sum_{j=1}^{\infty}\exp(-tE_j)$$
$$= \int_{x\in B} K(x,x;t)\,dx \leqslant \frac{1}{(2\pi t)^{d/2}}\int_{x\in B} dx\int_0^t \frac{d\tau}{t}\,\mathbb{E}\{e^{-tV(u(\tau))} : u(0) = u(t) = x; u(\tau)\in B\}$$
$$= \frac{1}{(2\pi t)^{d/2}}\int_{x\in B} dx\int_0^t \frac{d\tau}{t}\int_{u\in B} du\cdot\exp\left(-tV(u) - \frac{t|x-u|^2}{2\tau(t-\tau)}\right)\cdot\left(\frac{t}{(2\pi\tau)(t-\tau)}\right)^{d/2}$$
$$\leqslant \frac{1}{(2\pi t)^{d/2}}\int_{x\in R^d} dx\int_0^t \frac{d\tau}{t}\int_{u\in B} du\,\exp\left(-tV(u) - \frac{t|x-u|^2}{2\tau(t-\tau)}\right)\cdot\left(\frac{t}{(2\pi\tau)(t-\tau)}\right)^{d/2}$$
$$= \frac{1}{(2\pi t)^{d/2}}\int_{u\in B} e^{-tV(u)}du, \tag{16}$$

by Jensen's inequality. Moreover it follows from (13) that

$$\int_{x\in B} dx\left(K(x,x;t) - \frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\right) \geqslant -\int_{x\in B} dx\,\frac{e^{-tV(x)}}{(2\pi t)^{d/2}}\left\{\left|\int_0^t d\tau\int_{u\in B} du\,(V(u) - V(x))\left(\frac{t}{2\pi\tau(t-\tau)}\right)^{d/2}\right.\right.$$
$$\left.\left.\cdot\exp\left(-\frac{t|x-u|^2}{2\tau(t-\tau)}\right)\right| + 2d\exp\left(-\frac{2d_x^2}{dt}\right)\right\}. \tag{17}$$

Combining the bounds (16) and (17) we arrive at (15).

ACKNOWLEDGMENTS

The author wishes to thank Dr. J. T. Lewis for his helpful discussions. The work described in this paper is part of research program of the Foundation for Fundamental Research on Matter, which is financially supported by the Netherlands Organization for Pure Research.

[1]R. M. Ziff, G. E. Uhlenbeck, and M. Kac, Phys. Rep. **32**, 171 (1977).
[2]J. T. Lewis and J. V. Pulè, Commun. Math. Phys. **36**, 1 (1974).
[3]M. van den Berg and J. T. Lewis (to be published).
[4]M. van den Berg, Phys. Lett. A **78**, 88 (1980).
[5]E. B. Davies, J. London Math. Soc. (2) **7**, 483 (1973).
[6]M. Rosenblatt, Trans. Am. Math. Soc. **71**, 120 (1951).
[7]D. B. Ray, Trans. Am. Math. Soc. **82**, 299 (1954).
[8]M. Kac, *Proceedings of the Second Berkeley Symposium* (U. Cal. P., Berkeley, Calif., 1951), p. 189.
[9]E. D. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1952).
[10]R. Arima, J. Math. Kyoto Univ. **4**, 207 (1964).

Decay mode solution of the two-dimensional KdV equation and the generalized Bäcklund transformation

Akira Nakamura
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

(Received 13 January 1981; accepted for publication 1 May 1981)

We consider the generalization of the customary Bäcklund transform (≡BT) for the two-dimensional (2D) KdV equation, $(u_t + 6uu_x + u_{xxx})_x + 3\alpha^2 u_{yy} = 0$, with α being constant. A nonlinear superposition formula has been obtained and it is shown that the present generalized BT can produce multiple soliton–multiple decay mode solutions.

PACS numbers: 02.30.Jr, 02.30.Qy

I. INTRODUCTION

Recently "solitons" have become widely recognized as an important fundamental mode of certain nonlinear systems.[1] In these nonlinear systems, any initial disturbances evolve as superpositions of solitons and ripples (decay mode), where each soliton is stable and has infinite life time while ripples are unstable and decay as time passes.[2,3] So far, compared to the well-investigated soliton mode, this decay mode is not well understood except for a few studies.[4–7]

In this paper, we investigate this decay mode by taking the example of the two-dimensional KdV equation (or Kadomtsev–Petviashvili equations) written as[8,9]

$$(u_t + 6uu_x + u_{xxx})_x + 3\alpha^2 u_{yy} = 0. \tag{1.1}$$

Here α is an arbitrary constant. Throughout the paper, subscripts x, y, and t represent partial derivatives. We derive explicit multiple decay mode solutions for Eq. (1.1). Just as soliton solutions have the nonlinear superposition property, the present decay mode solutions also have the nonlinear superposition property. Moreover, solitons and the present decay mode also have the superposition property with respect to each other. This means that the general solution to Eq. (1.1) may be given by the superposed state of multiple solitons–multiple decay modes colliding with each other, preserving their identities. Physically, the only difference between solitons and present decay modes is that solitons are absolutely stable while decay modes vanish eventually as time passes.

Our present analysis is based on the bilinear Bäcklund Transform (hereafter abbreviated as BT) method, which is BT theory[10,11] performed in the transformed bilinear variables.[12–15] In Sec. 2, we derive a generalized bilinear BT and in Sec. 3, we study commutability and the superposition property of the obtained BT. In Sec. 4, by using the obtained superposition formula, we derive simple multiple ripple mode solutions and others, while in Sec. 5, a generalized BT in the original physical variables is presented. The brief results of the present work have been reported elsewhere.[16]

II. BILINEAR BT

In the following bilinear BT theory, we transform every quantity from the physical variable u to the bilinear variable f. The reason is that the form of solution becomes simple and transparent in the transformed bilinear variables.

By the dependent variable transform

$$u = (2\ln f)_{xx}, \tag{2.1}$$

Eq. (1.1) is transformed to the form[17]

$$\{(Ff\cdot f)/f^2\}_{xx} = 0,\quad F \equiv D_x D_t + D_x^4 + 3\alpha^2 D_y^2. \tag{2.2}$$

Here bilinear differential operators D_x, D_y, D_t are defined for arbitrary functions a, b as[12–14]

$$\begin{aligned} &D_x^l D_y^m D_t^n a(x,y,t)\cdot b(x,y,t) \\ &\quad \equiv (\partial_x - \partial_{x'})^l (\partial_y - \partial_{y'})^m (\partial_t - \partial_{t'})^n \\ &\qquad \times a(x,y,t) b(x',y',t')\big|_{x'=x, y'=y, t'=t}. \end{aligned} \tag{2.3}$$

We consider two different solutions u and u' of Eq. (1.1) and consider a transformation between u and u' (which is the BT to be investigated). We take a dependent variable transform also for u' as $u' = (2\ln f')_{xx}$. The simplest relation connecting u and u' is given by the direct subtraction of Eq. (1.1) for u and u' which, after integration twice with respect to x and setting the integration constants equal to zero, can be transformed into the form

$$P \equiv (Ff\cdot f)f'f' - ff(Ff'\cdot f') = 0. \tag{2.4}$$

Now, instead of the original variables u and u', we consider everything in the transformed variables f and f'. The relation between f and f' which satisfies Eq. (2.4) is a bilinear BT. We have found a general bilinear BT as follows:

$$\{D_x^2 + 2\lambda D_x + \alpha D_y + (\mu + \alpha\lambda_y(x + x_1))\}f\cdot f' = 0, \tag{2.5a}$$

$$\begin{aligned} &\{D_t + D_x^3 - 3\alpha D_x D_y - 6\alpha\lambda D_y - 3(\mu + \alpha\lambda_y(x + x_1))D_x \\ &\quad - 3\alpha(\nu + \mu_y(x + x_1) + \tfrac{1}{2}\alpha\lambda_{yy}(x + x_1)^2)\}f\cdot f' = 0, \end{aligned} \tag{2.5b}$$

where $\lambda = \lambda(y,t)$, $\mu = \mu(y,t)$, $\nu = \nu(y,t)$ are arbitrary functions of y and t, while x_1 is an arbitrary constant corresponding to the initial position of the wave packet. We call Eqs. (2.5a) and (2.5b), respectively the space part and time part of BT. In the case of $\lambda = \nu = 0$, $\mu = \text{const}$, Eqs. (2.5a) and (2.5b) reduce to the one obtained by Hirota and Satsuma.[15] In Appendix B, we show that Eqs. (2.5a) and (2.5b) satisfy equation (2.4).

Now we check the generation of solitons and decay mode by the present BT. When $f = 1$ ($u = 0$; trivial vacuum state), Eqs. (2.5a) and (2.5b) become pure linear equations for

 0022-2488/81/112456-07$01.00

f'. For the choice of the parameters

$$\lambda = -\alpha l/2, \mu = -k^2, \nu = -k^2 l, \tag{2.6}$$

where l, k are arbitrary constants, we have the solution

$$f' = c_1 e^{\theta} + c_2 e^{-\theta},$$
$$\theta \equiv k(x + ly + \omega t) + \theta_0, \quad \omega \equiv -3\alpha^2 l^2 - 4k^2, \tag{2.7}$$

where c_1, c_2, k, l, and θ_0 are arbitrary constants. This solution (with $c_1 = c_2 = 1$) corresponds to the well-known one-soliton solution, $u' = (2 \ln f')_{xx} = 2k^2 \operatorname{sech}^2\theta$. On the other hand, we have found that for the choice of parameters

$$\lambda = -\frac{y + y_1}{12\alpha t}, \quad \mu = -\lambda^2,$$
$$\nu = -\frac{2}{\alpha}\lambda\mu + \frac{1}{3}\lambda_y = -\frac{1}{3\alpha}\left(\frac{1}{12t} - 6\lambda^3\right), \tag{2.8}$$

there exists another type of solution to Eqs. (2.5a) and (2.5b) as

$$f' = c_3 \mathrm{Ai}(z_i) + c_4 \mathrm{Bi}(z_i),$$
$$z_i \equiv (x + x_i)(12t)^{-1/3} + \left(\frac{y + y_i}{\alpha}\right)^2 (12t)^{-4/3}, \tag{2.9}$$

where c_3, c_4 are arbitrary constants and Airy functions Ai and Bi are two linearly independent solutions of the ordinary differential equation

$$w_{zz}(z) - zw(z) = 0. \tag{2.10}$$

Solution (2.9) is the similarity type solution. Clearly, our present transform $u' = (2 \ln f')_{xx}$ indicates that the point $f' = 0$ gives $|u'| = \infty$. Since f' of Eq. (2.9) allows $f' = 0$ at some z_i values, it is actually a divergent solution. However, as will be shown later, at the higher-order stages of successive BT's, we can construct nondivergent solutions of this type. For the higher-order transformation, we should consider the commutability and superposition properties of the present BT.

III. COMMUTABILITY AND SUPERPOSITION FORMULAS OF PRESENT BT

First we consider the BT relation depicted in Fig. 1, where solid lines represent BT. Furthermore let us *assume* that $f_{12} = f_{21}$. Then the space part BT is written explicitly as

$$[D_x^2 + 2\lambda_1 D_x + \alpha D_y + \{\mu_1 + \alpha\lambda_{1y}(x + x_1)\}] f_0 \cdot f_1 = 0, \tag{3.1a}$$

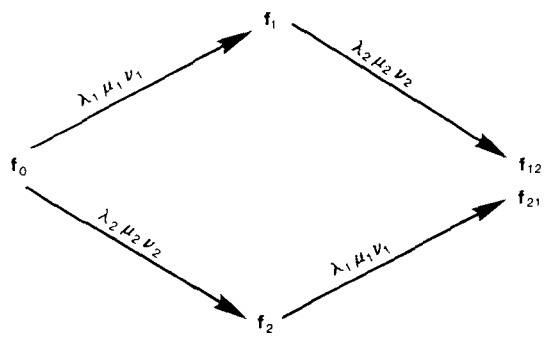


FIG. 1. Solid line represents the Bäcklund Transform (≡BT) with the respective parameters attached to the line. It can be proved that the present BT has the "commutability" property. Namely, operations of the BT first with parameters $\lambda_1\mu_1\nu_1$, and then with $\lambda_2\mu_2\nu_2$ result in the same function (f_{12}) as operations first with $\lambda_2\mu_2\nu_2$ and then with $\lambda_1\mu_1\nu_1$.

$$[D_x^2 + 2\lambda_1 D_x + \alpha D_y + \{\mu_1 + \alpha\lambda_{1y}(x + x_1)\}] f_2 \cdot f_{21} = 0, \tag{3.1b}$$

$$[D_x^2 + 2\lambda_2 D_x + \alpha D_y + \{\mu_2 + \alpha\lambda_{2y}(x + x_2)\}] f_0 \cdot f_2 = 0, \tag{3.1c}$$

$$[D_x^2 + 2\lambda_2 D_x + \alpha D_y + \{\mu_2 + \alpha\lambda_{2y}(x + x_2)\}] f_1 \cdot f_{12} = 0. \tag{3.1d}$$

Multiplying Eq. (3.1a) by $f_2 f_{21}$ and Eq. (3.1b) by $-f_0 f_1$; adding each other and using (A6), we obtain

$$D_x[(D_x f_0 \cdot f_{21}) \cdot f_2 f_1 + f_0 f_{21} \cdot (D_x f_2 \cdot f_1)] + (2\lambda_1 D_x + \alpha D_y) f_0 f_{21} \cdot f_2 f_1 = 0. \tag{3.2}$$

Similarly, from Eqs. (3.1c) and (3.1d) we obtain Eq. (3.2) with 1 and 2 interchanged in every subscript. By subtracting Eq. (3.2) from Eq. (3.2) with subscripts 1 and 2 interchanged, we have

$$D_x f_0 f_{12} \cdot \{(D_x + \lambda_1 - \lambda_2) f_2 \cdot f_1\} = 0, \tag{3.3}$$

which gives the superposition formula

$$f_0 f_{12} = c(y,t)(D_x - \lambda_1 + \lambda_2) f_1 \cdot f_2, \tag{3.4}$$

with $c(y,t)$ being an arbitrary function of y and t. Equation (3.4) generates a new solution f_{12} if we have three old (known) solutions: f_0, f_1, and f_2.

Although we have obtained an explicit superposition formula such as Eq. (3.4), the above argument was based on the assumption of commutability of BT, $f_{12} = f_{21}$. To be rigorous, this assumption must be proved. By a rather tedious calculation, we can prove that if $c(y,t)$ is chosen as $c(y,t) = c_0(t)^{1/2}$ with c_0 representing an arbitrary constant, commutability acutally holds ($f_{12} = f_{21}$), and the new function f_{12} constructed by Eq. (3.4) is actually a new solution (Fig. 2). We give the details of this calculation in Appendix C.

IV. GENERATION OF SOLITONS AND RIPPLONS

Now we consider generation of explicit solutions of physical interest. Since we have the relation $u = (2 \ln f)_{xx}$, multiplication of f by an arbitrary function of y and t does not affect u. Therefore, for convenience, in our superposition formula (3.4), we drop the unessential $(t)^{1/2}$ factor and consider

$$f_{12} \propto f_0^{-1}(D_x - \lambda_1 + \lambda_2) f_1 \cdot f_2. \tag{4.1}$$

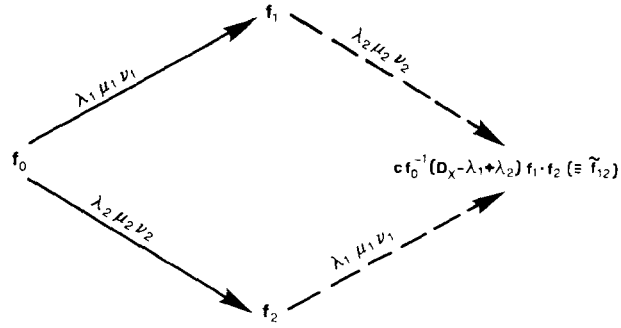


FIG. 2. Solid lines represent assumed BT and broken lines represent the relation which will be proved to be a BT.

In the following, we use the property of Airy functions:

$$D_x \mathrm{Bi}(z_i)\cdot\mathrm{Ai}(z_i) = (12t)^{-1/3}\{\mathrm{Bi}'(z_i)\mathrm{Ai}(z_i) - \mathrm{Bi}(z_i)\mathrm{Ai}'(z_i)\}$$
$$= (12t)^{-1/3}\pi^{-1}. \tag{4.2}$$

$$(x_{i'} - x_i)^{-1} D_x \mathrm{Ai}(z_{i'})\cdot\mathrm{Bi}(z_i)$$
$$= (12t)^{-1}\int^x dx\ \mathrm{Ai}(z_{i'})\mathrm{Bi}(z_i), \quad (\text{for } y_{i'} = y_i) \tag{4.3a}$$

$$(x_{i'} - x_i)^{-1} D_x \mathrm{Ai}(z_{i'})\cdot\mathrm{Ai}(z_i)$$
$$= (12t)^{-1}\int^x dx\ \mathrm{Ai}(z_{i'})\mathrm{Ai}(z_i), \quad (\text{for } y_{i'} = y_i) \tag{4.3b}$$

$$\lim_{x_{i'}\to x_i} (x_{i'} - x_i)^{-1} D_x \mathrm{Ai}(z_{i'})\cdot\mathrm{Ai}(z_i)$$
$$= \mathrm{Ai}_{xx}(z_i)\mathrm{Ai}(z_i) - \{\mathrm{Ai}_x(z_i)\}^2$$
$$= (12t)^{-1}\int^x dx\ \mathrm{Ai}^2(z_i), \quad (\text{for } y_{i'} = y_i). \tag{4.3c}$$

In Eqs. (4.3b) and (4.3c), the same relations hold for Bi instead of Ai. Here we note that integrals in equations (4.3a)–(4.3c) are indefinite integrals. Namely, the lower bound can be any constant including $+\infty$.

By taking the proper choice of c_3, c_4 in Eq. (2.9), we can have for example,

$$f_0 = 1, \quad f_1 = \rho_1 \mathrm{Ai}(z_{1'}) + (x_{1'} - x_1)\frac{\pi}{\rho_1}\mathrm{Bi}(z_{1'}),$$
$$f_2 = \frac{\rho_1 \mathrm{Ai}(z_1)}{x_{1'} - x_1}, \quad (y_{1'} = y_1), \tag{4.4}$$

where ρ_1 is arbitrary constant parameter. In this case Eq. (4.1) generates

$$f_{12} \propto \pi D_x \mathrm{Bi}(z_{1'})\cdot\mathrm{Ai}(z_1) + \rho_1^2\,(12t)^{-1}$$
$$\times \int^x dx\ \mathrm{Ai}(z_{1'})\mathrm{Ai}(z_1), \quad (y_{1'} = y_1). \tag{4.5}$$

In the limit $x_{1'} \to x_1$, this reduces to

$$f_{12} \propto 1 + \rho_1^2 (12t)^{-2/3}\int^x dx\ \mathrm{Ai}^2(z_1). \tag{4.6}$$

Similarly we can obtain

$$f_{12} \propto 1 + \rho_1^2 (12t)^{-2/3}\int^x dx\ \mathrm{Bi}^2(z_1). \tag{4.7}$$

These are two independent types of solutions different from the soliton mode. We call them "ripplons". The Ai^2 type is the same as the one obtained by Johnson and Thompson[18] and Reddekopp[19] and the Bi^2 type is new. Now it is worthy to look at the nature of the solution given by Eqs. (4.6) and (4.7) briefly. From $u = (2\ln f)_{xx}$, we see that both solutions have limit $u \to 0$ as $t \to \pm\infty$. From the definition of z_i in Eq. (2.9) we note that at any given time t, the points in the xy plane on the parabolic line $x + x_i + ((y + y_i)/\alpha)^2[1/(12t)] = \text{const}$ have equal value of u (namely constitute an equicontour of u). As time t increases in $0 < t < +\infty$, the curvature of this parabola becomes more and more flattened. Due to the equicontours, essentially one-dimensional information of the cross-section profile of $u(x,y,t)$ at $y = \text{const}$ (say $y = -y_i$) is sufficient to draw the complete profile of $u(x,y,t)$ over the whole xy plane. The graph of the cross-section profile for the Ai^2 type solution, Eq. (4.6), has already been reported in the study of cylindri-

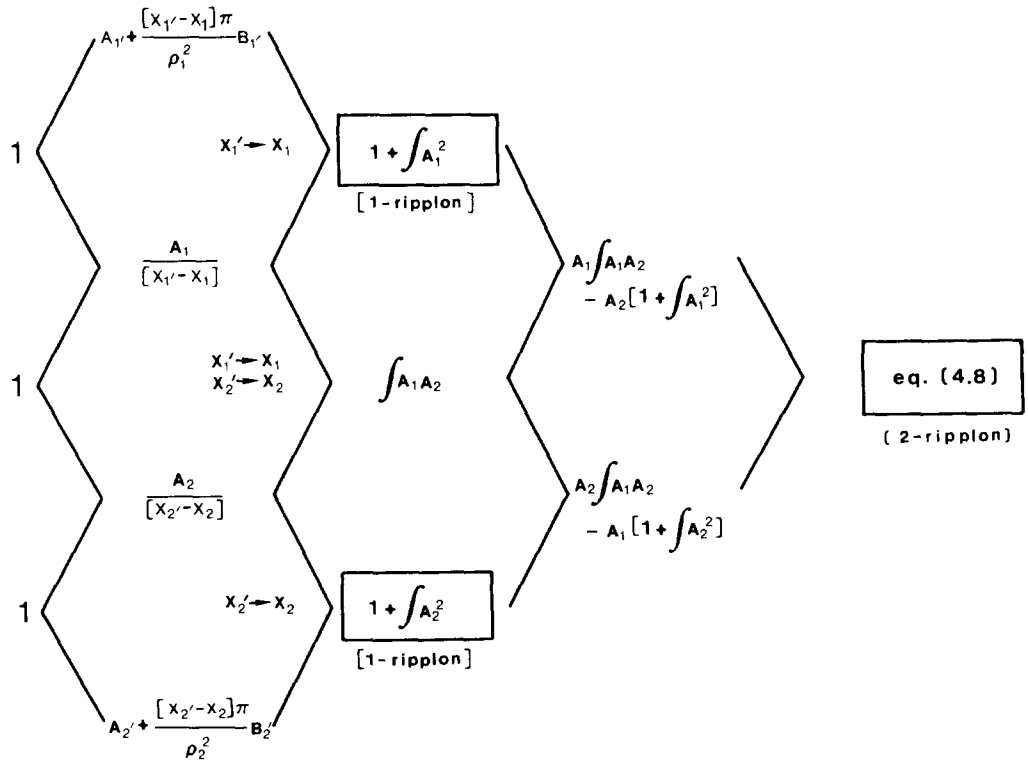


FIG. 3. Generation of two-ripplon solution by the BT for the simple case of $y_1 = y_2 (= y_{1'} = y_{2'})$. Special abbreviated notation is adopted as $A_n \equiv \rho_n \mathrm{Ai}(z_n)$, $B_n \equiv \rho_n \mathrm{Bi}(z_n)$, $\int \equiv (12t)^{-2/3}\int^x dx$.

cal KdV solutions[20] and recently for the present system[19] and is seen to be roughly one hump (either large or small) plus oscillating tails. Thus the whole form of this decay mode is like a horseshoe with oscillating tails both decaying in amplitude and flattening its parabolic curvature as time passes. Reddekopp[19] shows that in the $t\to\infty$ limit, the Ai^2 type ripplon tends to $\propto(1/t)\mathrm{Ai}(z_i)\mathrm{Ai}'(z_i)$ which coincides precisely with the solution of the linear initial-value problem with initial condition $u(x,y,0)\propto\delta(x)\delta(y)$, which gives a neat physical explanation for the present ripplon solution. In the higher orders of successive BT's, we can have a multiple superposition of ripplons and solitons. By the example shown in Fig. 3, we obtain two-ripplon solutions

$$f=\begin{vmatrix} 1+\rho_1^2(12t)^{-2/3}\int^x dx\ \mathrm{Ai}^2(z_1) & \rho_1\rho_2(12t)^{-2/3}\int^x dx\ \mathrm{Ai}(z_1)\mathrm{Ai}(z_2) \\ \rho_2\rho_1(12t)^{-2/3}\int^x dx\ \mathrm{Ai}(z_2)\mathrm{Ai}(z_1) & 1+\rho_2^2(12t)^{-2/3}\int^x dx\ \mathrm{Ai}^2(z_2) \end{vmatrix}. \tag{4.8}$$

where ρ_1,ρ_2 are arbitrary constants and for brevity we have considered the simple case of $y_1=y_2(=y_{1'}=y_{2'})$. Similarly, by the procedure shown in Fig. 4, we have the one soliton–one ripplon solution written as

$$\begin{aligned} &f=(e^{\theta}A_{1x}+e^{\theta}YA_1)^{-1}\left(D_x+\frac{y+y_1}{12\alpha t}-\frac{\alpha l}{2}\right)f_a\cdot f_b, \\ &f_a\equiv(12t)^{-2/3}A_1e^{\theta}(A_{1x}+YA_1)+\left\{1+(12t)^{-2/3}\int^x dx\ \mathrm{A}_1^2\right\}e^{\theta}\{Y^2-z_1(12t)^{-2/3}\}, \\ &f_b\equiv[\{z_1(12t)^{-2/3}+Y^2\}A_1+2YA_{1x}](1+e^{2\theta}/2k)+2k(A_{1x}+YA_1), \\ &Y\equiv(y+y_1)/12\alpha t-\alpha l/2-k,\quad A_1\equiv\rho_1\mathrm{Ai}(z_1), \end{aligned} \tag{4.9}$$

where θ is the same as in Eq. (2.7). In Eq. (4.9), in the limit $t\to\infty$ with x,y being finite, we have

$$Y\to-\alpha l/2-k\,(\equiv Y_\infty),\quad \mathrm{Ai}_x(z_1)/\mathrm{Ai}(z_1)\to 0,\quad f_a\to Y_\infty^2\,e^{\theta},\quad f_b\to Y_\infty\rho_1\mathrm{Ai}(z_1)\{Y_\infty(1+e^{2\theta}/2k)+2k\}$$

and the one soliton–one ripplon solution tends to the limit f_∞ as

$$f\to f_\infty\propto 1+e^{2(\theta+\delta)}/2k,\quad e^{\delta}\equiv(k+\alpha l/2)/(k-\alpha l/2), \tag{4.10}$$

which is seen to be equivalent to the one-soliton solution. This represents the limit where the decay mode has disappeared and only the soliton remains stable. As shown above, in each even stage of successive BT's, we have physical solutions. We can also have the solution given by Eq. (4.8) with

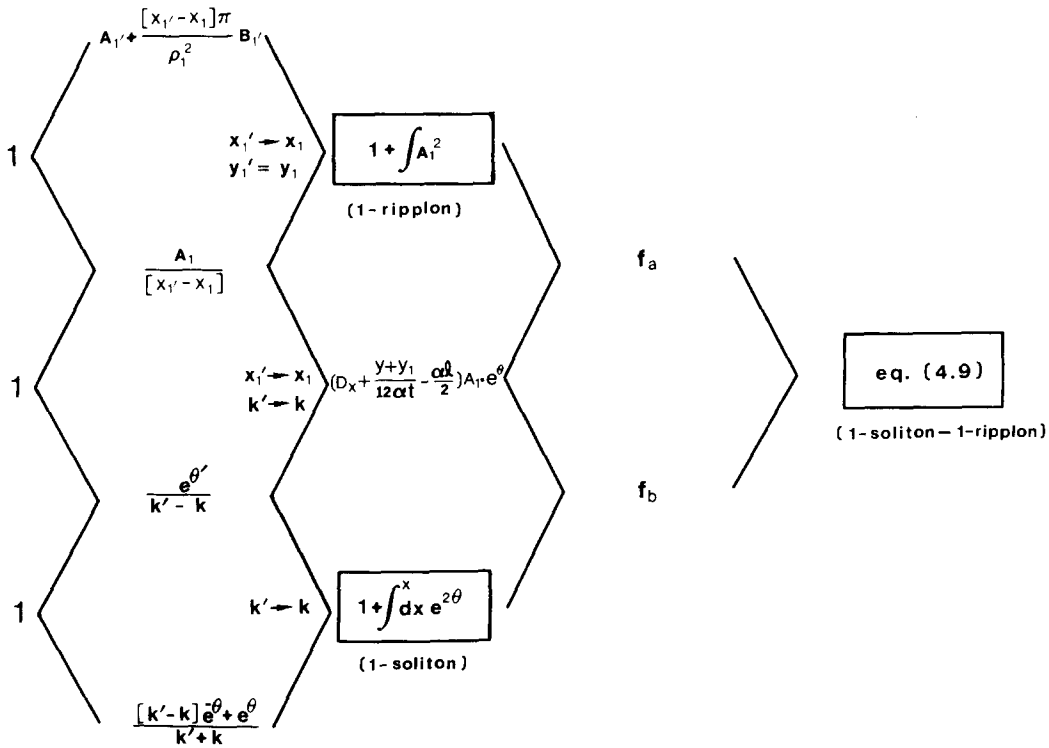


FIG. 4. Generation of one-soliton–one-ripplon solution by the BT. Special abbreviated notation is the same as in Fig. 3. θ' denotes θ given by (2.7) with k replaced by k'.

either both $\mathrm{Ai}(z_1)$ and $\mathrm{Ai}(z_2)$ replaced by $\mathrm{Bi}(z_1)$ and $\mathrm{Bi}(z_2)$ or only $\mathrm{Ai}(z_1)$ replaced by $\mathrm{Bi}(z_1)$ similarly to the case of cylindrical KdV equation.[21,22] By applying BT successively, *in principle* we can obtain either multiple soliton or multiple ripplon [of Ai type, Bi type, or Ai–Bi mixed type (in the case of two- and higher multiple ripplons)] or multiple soliton–multiple ripplon solutions.

V. BT IN PHYSICAL VARIABLES

We define potentials w, w' by $u \equiv -w_x$, $u' \equiv -w'_x$. From Eq. (2.1), this means $w = -(2\ln f)_x$, $w' = -(2\ln f')_x$. On the other hand, in bilinear BT equations (5a) and (5b), terms containing operators D_x, D_y, D_t: can be re-expressed by $\ln f$, $\ln f'$ as $(D_x f\cdot f')/ff' = (\ln f/f')_x$ and so on,[14] which in turn can be expressed by w, w'. In this way we obtain the BT in physical variables as

$$(w+w')_x - \tfrac{1}{2}(w-w')^2 + 2\lambda(w-w') + \alpha\partial_y \times \int dx\,(w-w') - 2\mu - 2\alpha\lambda_y(x+x_1) = 0, \tag{5.1a}$$

$$\partial_t \int dx\,(w-w') + (w-w')_{xx} - \frac{3}{2}(w-w')(w+w')_x + \tfrac{1}{4}(w-w')^3 - 3\alpha\left\{(w+w')_y - \left(\frac{w-w'}{2}\right)\partial_y \int dx\,(w-w')\right\} - 6\alpha\lambda\partial_y \int dx\,(w-w') - 3(\mu + \alpha\lambda_y(x+x_1))(w-w') + 6\alpha\left\{\nu + \mu_y(x+x_1) + \frac{\alpha}{2}\lambda_{yy}(x+x_1)^2\right\} = 0. \tag{5.1b}$$

VI. DISCUSSIONS

In the present paper, we have constructed generalized BT's for 2D KdV equations. We have newly revealed that precisely the same nonlinear superposition relation, expressed by Eq. (2.6), holds not only between the usual solitons but also between decay mode solutions (ripplons) and even between solitons and decay mode solutions (ripplons). This means that not only various solitons but also various ripplons and even solitons and ripplons can be superposed on each other freely, the last being an essentially new finding. Such unified simplicity may suggest the possibility of a unified picture such that just as any permanent profiles can be described completely as the nonlinear superposition of various solitons, any decaying profiles can be described completely as the nonlinear superposition of various ripplons. However, for the complete description, it seems necessary to have another generalization to the present simple ripplon solutions. For example, in the KdV limit or $\alpha \to 0$ limit, present simple ripplon solutions do not have the proper limit. To make a rough comparison in the $\alpha = 0$ (1D KdV) case, even the simplest, purely self-similar solution is known to be expressed by an infinite series expansion involving Ai and its integrals,[23] while in the present case the purely self-similar solution is not an infinite series as is seen from Eq. (2.9). This implies a certain additional complexity in the case of 1D KdV limit and/or the most general case.

In connection with the study of the cylindrical KdV equation, Freeman[24] and Johnson[25] reported the similarity reduction of the 2D KdV equation into the cylindrical KdV equation which may provide the implication that the N-soliton solutions of the cylindrical KdV equation can be the special class of solution of the 2D KdV equation (requirement for the similarity reduction needs $y_1 = y_2 = \cdots$; namely all the heads of horseshoe ripplons are aligned on one line parallel to the x axis). Our present work shows the generalization of superposition properties that horseshoe ripplons can be superposed not only in the limited aligned configuration but also in any arbitrary relative positions in xy plane and also with any usual solitons. Another advantage of the present BT approach is that both Ai^2 type and Bi^2 type solutions are obtained on an equal footing while in the inverse spectral transform method, Bi^2 type solutions are not obtainable because one cannot define the integral from infinity due to the divergence. Although the Bi^2 integral itself diverges at infinity, corresponding physical solutions in the u variable may provide a nondivergent useful solution if one is confined within a finite xy region. In the case of the cylindrical KdV, it was actually the Bi^2 type solution written as Eq. (4.7), which reproduced the results of experiment.[21]

APPENDIX A: BILINEAR OPERATOR IDENTITIES

The following identities hold for arbitrary functions a, b, c, d.

$$D_x ca\cdot a = c_x aa, \tag{A1}$$

$$(D_x D_t a\cdot a)bb - aa(D_x D_t b\cdot b) = 2D_x(D_t a\cdot b)\cdot ab = 2D_t(D_x a\cdot b)\cdot ab, \tag{A2}$$

$$(D_x^4 a\cdot a)bb - aa(D_x^4 b\cdot b) = 2D_x(D_x^3 a\cdot b)\cdot ab - 6D_x(D_x^2 a\cdot b)\cdot(D_x a\cdot b), \tag{A3}$$

$$D_x(D_x D_y a\cdot b)\cdot ab + D_x(D_y a\cdot b)\cdot(D_x a\cdot b) - D_y(D_x^2 a\cdot b)\cdot ab = 0, \tag{A4}$$

$$(D_x a\cdot b)cd - ab(D_x c\cdot d) = D_x ad\cdot cb, \tag{A5}$$

$$(D_x^2 a\cdot b)cd - ab(D_x^2 c\cdot d) = D_x\{(D_x a\cdot d)\cdot cb + ad\cdot(D_x c\cdot b)\}, \tag{A6}$$

$$(D_x a\cdot b)c - (D_x a\cdot c)b = -a(D_x b\cdot c), \tag{A7}$$

$$(D_x^2 a\cdot b)c - (D_x^2 a\cdot c)b = -2a_x(D_x b\cdot c) + a(D_x b\cdot c)_x, \tag{A8}$$

$$(D_x^3 a\cdot b)c - (D_x^3 a\cdot c)b = -3a_{xx}(D_x b\cdot c) + 3a_x(D_x b\cdot c)_x - \tfrac{1}{4}a\{(D_x^3 b\cdot c) + 3(D_x b\cdot c)_{xx}\}, \tag{A9}$$

$$(D_x^2 a\cdot b)_x c - (D_x^2 a\cdot c)_x b = -a_{xx}(D_x b\cdot c) - a_x(D_x b\cdot c)_x + \tfrac{1}{4}a\{(D_x^3 b\cdot c) + 3(D_x b\cdot c)_{xx}\}, \tag{A10}$$

$$D_x^3(D_x a\cdot c)\cdot ac - D_x(D_x^3 a\cdot c)\cdot ac + 3D_x(D_x^2 a\cdot c)\cdot(D_x a\cdot c) = 0, \tag{A11}$$

APPENDIX B: PROOF OF BILINEAR BT OF 2D KdV EQUATION

We can check that Eqs. (2.5a) and (2.5b) satisfy Eq. (2.4) as follows:

$$\tfrac{1}{2}P = D_x\{(D_t + D_x^3)f\cdot f'\}\cdot ff' - 3D_x(D_x^2 f\cdot f')\cdot(D_x f\cdot f') + 3\alpha^2 D_y(D_y f\cdot f')\cdot ff' \tag{B1}$$

$$\begin{aligned} &= D_x\left[\left\{\underline{(3\alpha D_x D_y)} + 6\alpha\lambda D_y + 3(\mu + \alpha\lambda_y(x+x_1))D_x + 3\alpha(\nu + \mu_y(x+x_1) + \frac{\alpha}{2}\lambda_{yy}(x+x_1)^2)\right\} f\cdot f'\right]\cdot ff' \\ &+ 3D_x[\{2\lambda D_x + \underline{\alpha D_y} + (\mu + \alpha\lambda_y(x+x_1))\}f\cdot f']\cdot(D_x f\cdot f') \\ &+ 3\alpha D_y[\{-\underline{D_x^2} - 2\lambda D_x - \underline{(\mu + \alpha\lambda_y(x+x_1))}\}f\cdot f']\cdot ff' = 0. \end{aligned} \tag{B2}$$

Equation (B1) is obtained by using equations (A2) and (A3). In (B2), terms with —— vanish from (A4) and terms with ---- from (A1). Remaining terms also vanish due to (A1) and (A2).

APPENDIX C: PROOF OF COMMUTABILITY OF PRESENT BT

We forget every assumption used so far and newly start from the Fig. 2 configuration where solid lines represent known or assumed BT's and broken lines represent the relation which we prove here to be a BT. First we prove the upper broken line of Fig. 2. The assumption that solid lines are BT's means that

$$(D_x^2 + 2\lambda_1 D_x + \alpha D_y + r_1)f_0\cdot f_1 = 0, \tag{C1}$$

$$(D_x^2 + 2\lambda_2 D_x + \alpha D_y + r_2)f_0\cdot f_2 = 0, \tag{C2}$$

$$(D_t + D_x^3 - 3\alpha D_x D_y - 6\alpha\lambda_1 D_y - 3r_1 D_x - 3\alpha s_1)f_0\cdot f_1 = 0, \tag{C3}$$

$$(D_t + D_x^3 - 3\alpha D_x D_y - 6\alpha\lambda_2 D_y - 3r_2 D_x - 3\alpha s_2)f_0\cdot f_2 = 0, \tag{C4}$$

$$r_i \equiv \mu_i + \alpha\lambda_{iy}(x+x_i),\ s_i \equiv \nu_i + \mu_{iy}(x+x_i) + \frac{\alpha}{2}\lambda_{iyy}(x+x_i)^2,\ s_{ix} = r_{iy}. \tag{C5}$$

Next, we define a new function $\tilde{f}_{12}$ in terms of f_0, f_1, f_2 as

$$f_0\tilde{f}_{12} \equiv c(y,t)(D_x - \lambda_1 + \lambda_2)f_1\cdot f_2. \tag{C6}$$

Notice here that $\tilde{f}_{12}$ is newly defined here and should not be confused with f_{12} or f_{21} in Fig. 1 (because we are now considering the Fig. 2 situation).

Then we prove that Q_1 defined by

$$Q_1 \equiv (D_x^2 + \lambda_2 D_x + \alpha D_y + r_2)f_1\cdot\tilde{f}_{12} \tag{C7}$$

vanishes. The proof is as follows: Multiplying (C1) by f_2, (C2) by $(-f_1)$, and adding the results, we have

$$\begin{aligned} 0 &= -2f_{0x}(D_x - \lambda_1 + \lambda_2)f_1\cdot f_2 + f_0\{(D_x - \lambda_1 + \lambda_2)f_1\cdot f_2\}_x + f_0\{-(\lambda_1 + \lambda_2)D_x - \alpha D_y + r_1 - r_2\}f_1\cdot f_2 \\ &= -(f_0/c)(D_x + \lambda_1 + \lambda_2)f_0\cdot\tilde{f}_{12} + f_0(-\alpha D_y + r_1 - \lambda_1^2 - r_2 + \lambda_2^2)f_1\cdot f_2. \end{aligned} \tag{C8}$$

$$(D_x + \lambda_1 + \lambda_2)f_0\cdot\tilde{f}_{12} = c(y,t)(-\alpha D_y + r_1 - \lambda_1^2 - r_2 + \lambda_2^2)f_1\cdot f_2. \tag{C9}$$

Then we see that

$$\begin{aligned} Q_1 f_0 f_2 &= \{(D_x^2 + 2\lambda_2 D_x + \alpha D_y + r_2)f_1\cdot\tilde{f}_{12}\}f_0 f_2 - f_1\tilde{f}_{12}(D_x^2 + 2\lambda_2 + \alpha D_y + r_2)f_0\cdot f_2, \\ &= D_x[\{(D_x - \lambda_1 + \lambda_2)f_1\cdot f_2\}\cdot f_0\tilde{f}_{12} + f_1 f_2\cdot\{(D_x + \lambda_1 + \lambda_2)f_0\cdot\tilde{f}_{12}\}] + \alpha D_y f_1 f_2\cdot f_0\tilde{f}_{12}, \\ &= -\alpha c D_x f_1 f_2\cdot(D_y f_1\cdot f_2) + \alpha c D_y f_1 f_2\cdot(D_x f_1\cdot f_2) - \alpha c_y f_1 f_2(D_x - \lambda_1 + \lambda_2)f_1\cdot f_2. \end{aligned} \tag{C10}$$

Thus for $c_y = 0$ or $c = c(t)$, Q_1 vanishes.

Next, consider the time part BT. We prove that Q_2, defined by

$$Q_2 \equiv (D_t + D_x^3 - 3\alpha D_x D_y - 6\alpha\lambda_2 D_y - 3r_2 D_x - 3\alpha s_2)f_1\cdot\tilde{f}_{12}, \tag{C11}$$

vanishes.

First, we multiply (C3), and (C4) respectively by f_2 and $(-f_1)$ differentiate (C1) and (C2) with respect to x and multiply the results, respectively, by $3f_2$ and $-3f_1$ and add all results; then we have, after using (C5),

$$\begin{aligned} &(3/2c)(D_x^2 - 2\alpha D_y)f_0\cdot\tilde{f}_{12} + \{D_t - 3\alpha(\lambda_1 + \lambda_2)D_y - 3(r_1 + r_2)D_x - \tfrac{1}{2}D_x^3 + 3\alpha(s_1 - s_2) + \frac{3}{2}(\lambda_1 - \lambda_2)D_x^2 \\ &+ 3\alpha(\lambda_1 - \lambda_2)_y\}f_1\cdot f_2 = 0. \end{aligned} \tag{C12}$$

Next, we multiply (C11) by $f_0 f_2$, (C4) by $f_1\tilde{f}_{12}$ and subtract the results to obtain

$$\begin{aligned} Q_2 f_0 f_2 &= \{D_t - 3\alpha\lambda_2 D_y - 3r_2 D_x + \tfrac{1}{4}D_x^3\}f_1 f_2\cdot f_0\tilde{f}_{12} + \tfrac{3}{4}D_x\{((D_x^2 - 2\alpha D_y)f_1\cdot f_2)\cdot f_0\tilde{f}_{12} \\ &+ 2(D_x f_1\cdot f_2)\cdot(D_x f_0\cdot\tilde{f}_{12}) + f_1 f_2\cdot(D_x^2 - 2\alpha D_y)f_0\cdot\tilde{f}_{12}\} - (3\alpha/2)D_y f_1 f_2\cdot\{(D_x + \lambda_1 + \lambda_2)f_0\cdot\tilde{f}_{12}\} - 3\alpha\lambda_{1y}f_1 f_2 f_0\tilde{f}_{12}. \end{aligned} \tag{C13}$$

Meanwhile, from $Ff_0 \cdot f_0 = 0$ and the assumption of a BT relation for $f_0 \longleftrightarrow f_1, f_2$ assures $Ff_1 \cdot f_1 = Ff_2 \cdot f_2 = 0$. Then we have

$$0 = (Ff_1 \cdot f_1) f_2 f_2 - f_1 f_1 (Ff_2 \cdot f_2) = 2D_x (D_t f_1 \cdot f_2) \cdot f_1 f_2 + 2D_x^3 (D_x f_1 \cdot f_2) \cdot f_1 f_2 + 6\alpha^2 D_y (D_y f_1 \cdot f_2) \cdot f_1 f_2. \tag{C14}$$

In (C13), we eliminate every $f_0 \tilde{f}_{12}$ by (C6), (C9), and (C12). Then using (C14), we obtain

$$Q_2 f_0 f_2 = \{3\alpha c(\mu_1 - \mu_2)_y + 6\alpha c\lambda_2(\lambda_2 - \lambda_1)_y + c(\lambda_1 - \lambda_2)_t + c_t(\lambda_1 - \lambda_2)\} f_1^2 f_2^2 - (c_t + 6\alpha c\lambda_{1,y}) f_1 f_2 (D_x f_1 \cdot f_2). \tag{C15}$$

For our present scheme, a set of parameters (λ,μ,ν) take two possible cases; constants given by Eq. (2.6) or certain functions of y, t given by Eq. (2.8). Thus in (C15), combinations of (λ_1,μ_1,ν_1) and (λ_2,μ_2,ν_2) have $2\times 2 = 4$ possibilities of the cases of Eqs. (2.6) and (2.8). We can check directly that in each of these four cases, the right-hand side of (C15) actually vanishes by taking c properly as $c =$ const or $c \propto t^{1/2}$.

Therefore, we have proved that the upper broken line of Fig. 2 is really a BT relation. Similarly, the lower broken line is also a BT. Thus we have proved both commutability, $f_{12} = f_{21}$, and that $\tilde{f}_{12}$ is connected by a BT to f_1 and f_2 and can be a new solution.

ACKNOWLEDGMENTS

Present work is partially supported by the Air Force Office of Scientific Research contract No. 78-3674B and Office of Naval Research contract No. N00014-74-C-0867.

[1]A. C. Scott, F. Y. F. Chu, and W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[2]R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[3]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Comm. Pure Appl. Math. **XXVII**, 97 (1974).
[4]M. J. Ablowitz and A. C. Newell, J. Math. Phys. **14**, 1277 (1973).
[5]M. J. Ablowitz and H. Segur, Stud. Appl. Math. **57**, 13 (1977).
[6]S. Oishi, J. Phys. Soc. Jpn. **47**, 1341 (1979); **48**, 639 (1980).
[7]V. E. Zakharov and A. B. Shabat, Func. Anal. Appl. **8**, 226 (1974).
[8]B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **192**, 753 (1970) [Soviet Phys. Dokl. **15**, 539 (1970)].
[9]M. Oikawa, J. Satsuma, and N. Yajima, J. Phys. Soc. Jpn. **37**, 511 (1974).
[10]G. L. Lamb, Jr., Rev. Mod. Phys. **43**, 99 (1971).
[11]H. D. Wahlquist and F. B. Estabrook, Phys. Rev. Lett. **31**, 1386 (1973).
[12]R. Hirota, *Bäcklund Transformations, the Inverse Scattering Method, Solitons and their Applications*, edited by R. M. Miura, Springer Lecture Notes in Mathematics, Vol. 515 (Springer, Berlin, 1976).
[13]R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).
[14]R. Hirota and J. Satsuma, Prog. Theor. Phys. Suppl. **59**, 64 (1976).
[15]R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. **45**, 1741 (1978).
[16]A. Nakamura, Phys. Rev. Lett. **46**, 751 (1981).
[17]J. Satsuma, J. Phys. Soc. Jpn. **40**, 286 (1976).
[18]R. S. Johnson and S. Thompson, Phys. Lett. **66**, 279 (1978).
[19]L. G. Reddekopp, Stud. Appl. Math. **63**, 185 (1980).
[20]F. Calogero and A. Degasperis, Lett. Nuovo Cimento **23**, 143 (1978).
[21]A. Nakamura and H. H. Chen, J. Phys. Soc. Jpn. **50**, 711 (1981).
[22]A. Nakamura, J. Phys. Soc. Jpn. **49**, 2380 (1980).
[23]M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977). (The results for the modified KdV equation pertain to the KdV equation via the Miura transform).
[24]N. C. Freeman, Adv. Appl. Mech. **20**, 1 (1980).
[25]R. S. Johnson, Phys. Lett. A **72**, 197 (1979).

Evolution equations associated with the discrete analog of the matrix Schrödinger spectral problem solvable by the inverse spectral transform

M. Bruschi, O. Ragnisco, and D. Levi[a)]
Istituto di Fisica dell'Universitá, Roma, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Italy

(Received 15 August 1980; accepted for publication 3 October 1980)

Through the generalized Wronskian technique we derive the whole class of nonlinear differential difference equations associated with the discrete analog of the matrix Schrödinger spectral problem. For such equations we briefly discuss soliton solutions, continuum limit, and Bäcklund transformations.

PACS numbers: 02.30.Ks

1. INTRODUCTION

In a previous paper, hereafter referred to as I,[1] the discrete analog of the matrix Schrödinger spectral problem (DMS) was investigated through the Riemann technique,[2] and the evolution equations for the nonabelian Toda-lattice were derived. The aim of the present paper is to construct the whole hierarchy of nonlinear differential difference equations (NDDE's) associated to such spectral problem which can be solved by the inverse spectral transform (IST), whose first member is just the nonabelian Toda-lattice.[3] For this purpose, it is convenient to look at DMS from a point of view closely analogous to that chosen by Calogero and Degasperis in order to investigate the continuous case,[4] and thus somewhat different from the approach followed in I. So, in Sec. 2, we briefly recast the direct and inverse spectral problem already treated in I; in Sec. 3 we derive the generalized Wronskian relations appropriate to DMS; in Sec. 4 we give the associated class of NDDE's elucidating the corresponding linear time evolution of the spectral data. Section 5 is devoted to the treatment of a special subclass of such NDDE's, which has an interesting continuum limit, while in Sec. 6 we give a preliminary sketch of Backlund transformations related to the spectral problem under scrutiny.

2. DIRECT AND INVERSE PROBLEM

A. Basic notations

Throughout this paper we shall use upper case characters for matrices, the only exception being given by the notation σ_k $(k = 1,...,N^2 - 1)$ which denotes the $N^2 - 1$ matrices that, together with the identity matrix $\sigma_0 = I$, constitute a basis for the space of $N \times N$ matrices. Moreover, in the following, greek indices will run over the values $0,...,N^2 - 1$, while latin indices will run over $1,...,N^2 - 1$; the summation convention for repeated indices is always understood. Upper case script characters will be used to indicate $2N \times N$ matrices.

We shall use the conventional notation for commutators (anticommutators) $[A,B] = AB - BA$ ($\{A,B\} = AB + BA$), and the Dirac notation for ordinary $\mathbb{C}^N$ vectors. We shall also introduce the operator η which, applied to a $2N \times N$ rectangular matrix, will transform it into a $2N \times 2N$ "block diagonal" matrix, according to the formula

$$\eta\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

Finally, matrix combinations of the type $A_1A_2A_3 + B_1B_2B_3$ will usually be expressed (somewhat loosely) through an inner product notation $(\mathscr{Y}_1,\eta[\mathscr{Y}_2]\mathscr{Y}_3)$, where

$$\mathscr{Y}_i = \begin{bmatrix} A_i \\ B_i \end{bmatrix} \quad \text{and} \quad (\mathscr{Y}_i,\mathscr{Y}_j) = A_iA_j + B_iB_j.$$

B. Direct problem

The discrete analog of the matrix Schrödinger spectral problem corresponds to the eigenvalue equation (see I)

$$\Phi(n-1) + B(n)\Phi(n) + A(n)\Phi(n+1) = \lambda\Phi(n), \quad (2.1)$$

where Φ,B,A are $N \times N$ matrices depending on the discrete variable n (running over all the integers) and possibly on a continuous variable, say t, and the "potentials" A,B satisfy the boundary conditions

$$\lim_{|n|\to\infty} A(n) - I = 0, \quad (2.2a)$$

$$\lim_{|n|\to\infty} B(n) = 0, \quad (2.2b)$$

with respect to some norm in the linear space of $N \times N$ matrices.

As remarked in I, under conditions (2.2) our spectral problem has a twice degenerate real continuous spectrum, which can be parametrized by setting $\lambda = z + z^{-1}$, z belonging to the unit circle of the complex plane. A fundamental matrix solution $\Psi(n,z)$ of (2.1) pertaining to the continuous spectrum can thus be characterized by the asymptotic behavior

$$\Psi(n,z) \underset{n\to+\infty}{\sim} z^{-n} + R(z)z^n, \quad (2.3a)$$

$$\Psi(n,z) \underset{n\to-\infty}{\sim} z^{-n}T(z), \quad (2.3b)$$

where $R(z)$ and $T(z)$ are the "reflection" and "transmission" matrix coefficients.[5]

As to the discrete spectrum, it has been shown in I that it consists of a finite number $\mathscr{N}$ of complex eigenvalues,

[a)]Present address: Department of Physics, Rockefeller University, New York.

 0022-2488/81/112463-09$01.00

which correspond to the poles z_j ($j = 1,...,\mathcal{N}$) of $T(z)$, all lying inside the unit circle. The corresponding normalizable vector eigenfunctions $|\psi_j(n)\rangle$ are characterized by the asymptotic behavior

$$|\psi_j(n)\rangle \underset{n\to+\infty}{\sim} z_j^n |c_j^{(+)}\rangle, \tag{2.4a}$$

$$|\psi_j(n)\rangle \underset{n\to-\infty}{\sim} z_j^{-n} |c_j^{(-)}\rangle, \tag{2.4b}$$

where

$$T_{z_j}^{-1} |c_j^{(-)}\rangle = 0. \tag{2.4c}$$

So, once given the "potentials" $A(n)$, $B(n)$ one can determine uniquely the matrices $R(z)$, $T(z)$, the eigenvalues z_j and the vectors $|c_j^{(\pm)}\rangle$.

C. Auxiliary problems and transformation properties

First of all, we associate to the eigenvalue equation (2.1) different asymptotic conditions for the solutions, by defining a new fundamental matrix solution $\hat{\Psi}(n,z)$ according to the formulas

$$\hat{\Psi}(n,z) \underset{n\to+\infty}{\sim} z^n \hat{T}(z), \tag{2.5a}$$

$$\hat{\Psi}(n,z) \underset{n\to-\infty}{\sim} z^n + z^{-n}\hat{R}(z). \tag{2.5b}$$

Moreover, we can introduce the adjoint eigenvalue equation

$$\bar{\Phi}(n+1) + \bar{\Phi}(n)B(n) + \bar{\Phi}(n-1)A(n-1) = \lambda\bar{\Phi}(n) \tag{2.6}$$

with the same potentials as Eq. (2.1), and define, as previously, two fundamental matrix solutions corresponding to different asymptotic conditions, i.e.,

$$\bar{\Psi}(n,z) \underset{n\to+\infty}{\sim} z^{-n} + z^n \bar{R}(z), \tag{2.7a}$$

$$\bar{\Psi}(n,z) \underset{n\to-\infty}{\sim} z^{-n}\bar{T}(z), \tag{2.7b}$$

$$\bar{\Psi}(n,z) \underset{n\to+\infty}{\sim} z^n \bar{T}(z), \tag{2.8a}$$

$$\bar{\Psi}(n,z) \underset{n\to-\infty}{\sim} z^n + z^{-n}\bar{R}(z). \tag{2.8b}$$

Let us now define the n-independent Wronskian matrix W through the position

$$W[\bar{\Phi}_1(n),\Phi_2(n)] = \bar{\Phi}_1(n)A(n)\Phi_2(n+1) - \bar{\Phi}_1(n+1)\Phi_2(n), \tag{2.9}$$

where $\bar{\Phi}_1$ is a solution of (2.6) and Φ_2 is a solution of (2.1).

It is easily seen that inserting into Eq. (2.9) the pairs $(\bar{\Psi},\Psi),(\bar{\Psi},\hat{\Psi}),(\hat{\bar{\Psi}},\Psi),(\hat{\bar{\Psi}},\hat{\Psi})$ in place of $(\bar{\Phi}_1,\Phi_2)$, one gets the following relationships between the four "reflection" and "transmission" coefficients previously introduced:

$$\bar{R}(z^{-1})R(z) + \bar{T}(z^{-1})T(z) = I, \tag{2.10a}$$

$$\bar{R}(z) = R(z), \tag{2.10b}$$

$$\hat{R}(z) = -T(z)R(z^{-1})T^{-1}(z^{-1}), \tag{2.11a}$$

$$\hat{T}(z) = \bar{T}(z) = (I - R(z)R(z^{-1}))T^{-1}(z^{-1}), \tag{2.11b}$$

$$\hat{\bar{R}}(z) = \hat{R}(z), \tag{2.12a}$$

$$\hat{\bar{T}}(z) = T(z). \tag{2.12b}$$

In particular, from (2.8b) it follows that the discrete eigenvalues of problem (2.6) associated to the asymptotic conditions (2.8) are the same as those pertaining to the problem (2.1) with the asymptotic conditions (2.3), while the corresponding vector eigenfunctions are characterized by the asymptotic behavior [6]

$$\langle\bar{\psi}_j(n)| \underset{n\to+\infty}{\sim} z_j^n \langle d_j^{(+)}|, \tag{2.13a}$$

$$\langle\bar{\psi}_j(n)| \underset{n\to-\infty}{\sim} z_j^{-n}\langle d_j^{(-)}|, \tag{2.13b}$$

where

$$\langle d_j^{(+)}|\hat{\bar{T}}^{-1}(z_j)| = \langle d_j^{(+)}|T^{-1}(z_j) = 0. \tag{2.13c}$$

D. Inverse problem

It was shown in I that the spectral data

$$S = \{R(z)(|z| = 1); z_j(|z_j| < 1), \rho_j |c_j^{(+)}\rangle\langle d_j^{(+)}|(j = 1,...,\mathcal{N})\}, \tag{2.14}$$

ρ_j being related to the singular behavior of $T(z)$ for $z \approx z_j$ through the formula

$$T(z) = -\rho_j z_j |c_j^{(-)}\rangle\langle d_j^{(+)}|(z - z_j)^{-1} + O(1) \tag{2.15}$$

allows us to recover uniquely the "potentials."

Actually, from S we can construct the kernel $H(n)$ of the discrete analog of the Gel'fand–Levitan–Marchenko equation,

$$K(n,l) + H(n+l) + \sum_{m=n+1}^{\infty} K(n,m)H(m+l) = 0, \tag{2.16a}$$

$$H(n) = \sum_{j=1}^{\mathcal{N}} z_j^n \rho_j |c_j^{(+)}\rangle\langle d_j^{(+)}| + (2\pi i)^{-1}\oint_{|z|=1} dz\, z^{n-1}R(z), \tag{2.16b}$$

and once (2.16a) is solved, we can find $A(n)$ and $B(n)$ through the formulas

$$A(n) - I = K(n,n+2) - K(n-1,n+1) + [K(n-1,n) - K(n,n+1)]K(n,n+1), \tag{2.17a}$$

$$B(n) = K(n,n+1) - K(n-1,n). \tag{2.17b}$$

3. GENERALIZED WRONSKIAN RELATIONS

A. Basic identities

In order to establish a one-to-one correspondence between the functional relations connecting two different pairs of "potentials" and those connecting the associated spectral parameters, it is convenient to define a generalized Wronskian.

To this aim, let us denote by $\bar{\Psi}'(n,z)$, respectively $\Psi(n,z)$, two fundamental matrix solutions of Eq. (2.6) [with "potentials" $A'(n),B'(n)$], respectively (2.1), with the asymptotic conditions (2.7), respectively (2.3), and associate to them the matrix bilinear form W', which we shall define as the gener-

alized Wronskian

$$W'[\bar{\Psi}'(n,z),\tilde{\Psi}(n,z)] = \bar{\Psi}'(n,z)A'(n)\tilde{\Psi}(n+1,z) - \bar{\Psi}'(n,1,z)\tilde{\Psi}(n,z), \tag{3.1}$$

where

$$\tilde{\Psi}(n,z) = F(n)\Psi(n,z) + G(n)\Psi(n+1,z), \tag{3.2}$$

$F(n)$ and $G(n)$ being so far two arbitrary n-dependent matrices.[7] A straightforward computation yields

$$\begin{aligned} &W'(n) - W'(n-1) \\ &= \lambda\{\bar{\Psi}'(n)[A'(n)G(n+1)A^{-1}(n+1) - G(n)]\Psi \\ &\quad \times\Psi(n+1) + \bar{\Psi}'(n)[F(n-1) - F(n)]\Psi(n)\} \\ &\quad + \bar{\Psi}'(n)[A'(n)(F(n+1) - G(n+1)A^{-1}(n+1)B(n+1)) \\ &\quad + B'(n)G(n) - F(n-1)A(n)]\Psi(n+1) \\ &\quad + \bar{\Psi}'(n)[B'(n)F(n) - F(n-1)B(n) \\ &\quad + G(n-1) - A'(n)G(n+1)A^{-1}(n+1)]\Psi(n). \end{aligned} \tag{3.3}$$

Equation (3.3) is the fundamental identity from which, by proper choices of the matrices $F(n)$, $G(n)$, any relevant relationship can be derived, as it will be clear in the following.

Let us first choose

$$F(n) = F(n\text{-independent}), \quad G(n) = 0. \tag{3.4}$$

Then, by summing up identity (3.4) over all integers and taking into account the asymptotic behaviors of the potentials and of the wavefunctions, Eqs. (2.2), (2.3), and (2.7), we immediately get, having set $\mu = z^{-1} - z$,

$$\begin{aligned} &-\mu(FR(z) - \bar{R}'(z)F) \\ &\quad = \sum_{n=-\infty}^{+\infty}\{\bar{\Psi}'(n)[A'(n)F - FA(n)\Psi(n+1) \\ &\quad + \bar{\Psi}'(n)[B'(n)F - FB(n)]\Psi(n)\}. \end{aligned} \tag{3.5}$$

If instead we set

$$F(n) = 0, \quad A'(n)G(n+1)A^{-1}(n+1) - G(n) = 0, \tag{3.6a}$$

which implies

$$G(n) = \Pi'(n)G\Pi^{-1}(n+1), \tag{3.6b}$$

where[8]

$$\Pi(n) = \prod_{j=n}^{\infty} A(j); \quad \Pi'(n) = \prod_{j=n}^{\infty} A'(j), \tag{3.6c}$$

we get

$$\begin{aligned} &\mu[zGR(z) - z^{-1}\bar{R}'(z)G] \\ &= \sum_{n=-\infty}^{+\infty}\{\bar{\Psi}'(n)[B'(n)\Pi'(n)G\Pi^{-1}(n+1) \\ &\quad - \Pi'(n)G\Pi^{-1}(n+1)B(n+1)]\Psi(n+1) \\ &\quad + \bar{\Psi}'(n)[\Pi'(n-1)G\Pi^{-1}(n) - \Pi'(n)G\Pi^{-1}(n+1)] \\ &\quad \times\Psi(n)\}. \end{aligned} \tag{3.7}$$

Finally, let us assume that $F(n)$ and $G(n)$ vanish (rapidly enough) as $|n|\to\infty$; in this case, we introduce two new matrices $P(n)$ and $Q(n)$ through the formulas

$$P(n) = A'(n)G(n+1)A^{-1}(n+1) - G(n), \tag{3.8a}$$

$$Q(n) = F(n-1) - F(n), \tag{3.8b}$$

which, once inverted, give

$$\begin{aligned} G(n) &= -\Pi'(n)\Big[\sum_{j=n}^{\infty}\Pi'(j)^{-1}P(j)\Pi(j+1)\Big]\Pi^{-1}(n+1) \\ &= = (\text{def.}) - \Sigma(n); \end{aligned} \tag{3.9a}$$

$$F(n) = \sum_{j=n+1}^{\infty} Q(j). \tag{3.9b}$$

It is perhaps worthwhile to emphasize here that, due to the arbitrarity of $F(n)$ and $G(n)$, also $P(n)$ and $Q(n)$ are essentially arbitrary matrices, the only constraint concerning their behavior as $|n|\to\infty$, where they must vanish as $G(n)$ and $F(n)$ or faster.

If we now define the operator Λ through the formula

$$\underline{\Lambda}\begin{bmatrix} P(n) \\ Q(n) \end{bmatrix} = \begin{bmatrix} P(n)B(n+1) + A'(n)[Q(n) + Q(n+1)] + B'(n)\Sigma(n) - \Sigma(n)B(n+1) + \sum_{j=n}^{\infty}[P(j)A(n) - A'(n)P(j)] \\ P(n) + B'(n)Q(n) - \Sigma(n) + \Sigma(n-1) + \sum_{j=n}^{\infty}[Q(j)B(n) - B'(n)Q(j)] \end{bmatrix} \tag{3.10}$$

we get from formula (3.3) the new identity, which holds for any entire function $\omega(\lambda)$,

$$\sum_{n=-\infty}^{+\infty}(\bar{\mathscr{V}}'(n),\eta\omega(\underline{\Lambda})[\mathscr{F}(n)]\mathscr{V}(n)) = \omega(\lambda)\sum_{n=-\infty}^{+\infty}(\bar{\mathscr{V}}'(n),\eta[\mathscr{F}(n)]\mathscr{V}(n)), \tag{3.11a}$$

where we have set

$$\mathscr{F}(n) = \begin{bmatrix} P(n) \\ Q(n) \end{bmatrix}, \tag{3.11b}$$

$$\bar{\mathscr{V}}'(n) = \begin{bmatrix} \bar{\Psi}'(n) \\ \bar{\Psi}'(n) \end{bmatrix}, \tag{3.11c}$$

$$\mathscr{V}(n) = \begin{bmatrix} \Psi(n) \\ \Psi(n+1) \end{bmatrix}, \tag{3.11d}$$

and nave used the notations introduced in Sec. 2A. We notice now that relation (3.5) as well as relation (3.7) can be written for each of the basis matrices σ_ν, and that Eq. (3.11a) holds for any choice of $P(n)$,$Q(n)$, provided they vanish at infinity: So we can, in particular, identify $\mathscr{F}(n)$ first with

$$\begin{bmatrix} A'(n)\sigma_\nu - \sigma_\nu A(n) \\ B'(n)\sigma_\nu - \sigma_\nu B(n) \end{bmatrix}$$

and secondly with

$$\begin{bmatrix} B'(n)\Pi'(n)\sigma_\nu \Pi^{-1}(n+1) - \Pi'(n)\sigma_\nu \Pi^{-1}(n+1)B(n+1) \\ \Pi'(n-1)\sigma_\nu \Pi^{-1}(n) - \Pi'(n)\sigma_\nu \Pi^{-1}(n+1) \end{bmatrix}.$$

Hence we can assert that two different pairs of potentials and the corresponding reflection coefficients are related by the equation

$$\mu\{f_\nu(\lambda)(\sigma_\nu R(z) - \bar{R}'(z)\sigma_\nu) + g_\nu(\lambda)(\sigma_\nu z R(z) - z^{-1}\bar{R}'(z)\sigma_\nu\} = \sum_{n=-\infty}^{+\infty} \left(\bar{\mathscr{V}}'(n), \left\{ \eta f_\nu(\underline{\Lambda}) \begin{bmatrix} A'(n)\sigma_\nu - \sigma_\nu A(n) \\ B'(n)\sigma_\nu - \sigma_\nu B(n) \end{bmatrix} \right.\right.$$
$$\left.\left. + \eta g_\nu(\underline{\Lambda}) \begin{bmatrix} B'(n)\Pi'(n)\sigma_\nu \Pi^{-1}(n+1) - \Pi'(n)\sigma_\nu \Pi^{-1}(n+1)B(n+1) \\ \Pi'(n-1)\sigma_\nu \Pi^{-1}(n) - \Pi'(n)\sigma_\nu \Pi^{-1}(n+1) \end{bmatrix} \right\} \mathscr{V}(n) \right). \tag{3.12}$$

B. The limit $(A'(n), B'(n)) \to (A(n), B(n))$

Let us now assume that the potentials, and consequently the wavefunctions and the spectral data, depend parametrically on a novel variable, say t (time) by setting

$$A(n) = A(n,t), \qquad A'(n) = A(n,t+\Delta t), \tag{3.13a}$$

$$B(n) = B(n,t), \qquad B'(n) = B(n,t+\Delta t), \tag{3.13b}$$

and thus

$$R(z) = R(z,t), \qquad R'(z) = R(z,t+\Delta t). \tag{3.14}$$

Then, we insert these positions into the equations derived in the previous subsection, and investigate the limit $\Delta t \to 0$. Let us begin by considering such limit for the operator $\underline{\Lambda}$, defined through formula (3.10): we get a new operator, say $\underline{L}$, which acts on elements like: $\begin{pmatrix} P(n) \\ Q(n) \end{pmatrix}$ as follows:

$$\underline{L} \begin{bmatrix} P(n) \\ Q(n) \end{bmatrix} = \begin{bmatrix} P(n)B(n+1) + A(n)(Q(n) + Q(n+1)) + B(n)S(n) - S(n)B(n+1) + \sum_{j=n}^{\infty} [P(j), A(n)] \\ B(n)Q(n) + P(n) - S(n) + S(n-1) + \sum_{j=n}^{\infty} [Q(j), B(n)] \end{bmatrix} \tag{3.15}$$

where $S(n)$ is the limit of $\Sigma(n)$ as $\Delta t \to 0$, i.e.,

$$S(n) = \Pi(n)\left[\sum_{j=n}^{\infty} \Pi^{-1}(j) P(j) \Pi(j+1) \right]. \tag{3.16}$$

Performing the same limit in Eq. (3.12), having set there $f_\nu = h(\Delta t)^{-1}\delta_{\nu,0}, g_\nu = 0$, and recalling Eq. (2.10b), we obtain

$$\mu h(\lambda) R_t(z,t) = \sum_{n=-\infty}^{+\infty} \left(\bar{\mathscr{V}}(n), \eta h(\underline{L}) \begin{bmatrix} A_t(n) \\ B_t(n) \end{bmatrix} \mathscr{V}(n) \right). \tag{3.17}$$

On the other hand, starting again from (3.12) and letting $\Delta t \to 0$ with no further position, we can write the identity

$$\mu[(f_\nu(\lambda) + \tfrac{1}{2}\lambda g_\nu(\lambda))\sigma_\nu, R(z) + \tfrac{1}{2}\mu g_\nu(\lambda)\{\sigma_\nu, R(z)\}$$
$$= \sum_{n=-\infty}^{+\infty} \left(\bar{\mathscr{V}}(n), \left\{ \eta f_\nu(\underline{L}) \begin{bmatrix} [A(n), \sigma_\nu] \\ [B(n), \sigma_\nu] \end{bmatrix} \right.\right.$$
$$\left.\left. + \eta g_\nu(L) \begin{bmatrix} B(n)\Pi(n)\sigma_\nu \Pi^{-1}(n+1) - \Pi(n)\sigma_\nu \Pi^{-1}(n+1)B(n+1) \\ \Pi(n-1)\sigma_\nu \Pi^{-1}(n) - \Pi(n)\sigma_\nu \Pi^{-1}(n+1) \end{bmatrix} \right\} \mathscr{V}(n) \right). \tag{3.18}$$

4. NONLINEAR EVOLUTION EQUATIONS SOLVABLE BY IST

A. NDDE's and evolution of the reflection coeffcient

The formulas derived at the end of the previous section allow us to define the class of NDDE's associated to DMSP which can be solved by IST.

Indeed, we can assert that if the potentials $(A(n), B(n))$ evolve in time according to the NDDE,

$$h(\underline{L})\begin{bmatrix}A_t(n)\\B_t(n)\end{bmatrix}+g_0(\underline{L})\begin{bmatrix}B(n)A(n)-A(n)B(n+1)\\A(n-1)-A(n)\end{bmatrix}+f_\kappa(\underline{L})\begin{bmatrix}[A(n),\sigma_\kappa]\\[B(n),\sigma_\kappa]\end{bmatrix}$$
$$+g_\kappa\begin{bmatrix}B(n)\Pi(n)\sigma_\kappa\Pi^{-1}(n+1)-\Pi(n)\sigma_\kappa\Pi^{-1}(n+1)B(n+1)\\\Pi(n-1)\sigma_\kappa\Pi^{-1}(n)-\Pi(n)\sigma_\kappa\Pi^{-1}(n+1)\end{bmatrix}=0, \tag{4.1}$$

the corresponding reflection coefficient, associated uniquely with the potentials for any value of t, evolves according to the linear evolution equation

$$h(\lambda)R_t(z)=-\mu g_0(\lambda)R(z)+(f_\kappa(\lambda)+\tfrac{1}{2}g_\kappa(\lambda))[\sigma_\kappa,R(z)]-\tfrac{1}{2}\mu g_\kappa(\lambda)\{\sigma_\kappa,R(z)\}, \tag{4.2}$$

where, we stress it again, h, f_κ, g_ν are arbitrary entire functions of their argument.

So, the solution of the Cauchy problem for any NDDE belonging to the class (4.1) can be achieved through three different steps, all involving just linear problems: (i) the solution of the direct spectral problem (2.1) at a given time t_0, in order to determine $R(z,t_0)$; (ii) the integration of Eq. (4.2), to get $R(z,t)$; (iii) the solution of the inverse problem to reconstruct the evoluted potentials (the existence of bound states will be taken into account later). We remark that Eq. (4.2) can be integrated in closed form, yielding

$$R(z,t)=U(z;t-t_0)R(z,t_0)U(z^{-1},t-t_0), \tag{4.3}$$

where

$$U(z;t)=\exp[tZ(z)], \tag{4.4a}$$
$$Z(z)=h^{-1}(\lambda)[f_\nu(\lambda)+zg_\nu(\lambda)]\sigma_\nu. \tag{4.4b}$$

B. Evolution of discrete spectrum parameters

In this subsection we shall prove that the discrete eigenvalues z_j do not evolve in time and that the degenerate matrix

$$C_j=\rho_j|c_j^{(+)}\rangle\langle d_j^{(+)}| \tag{4.5}$$

has the same time evolution as the reflection coefficient $R(z)$ [Eq. (4.3)] with the obvious substitution $z\to z_j$. To this aim, we need the time evolution of the transmission coefficients $T(z),\hat{T}(z)$, since $\langle d_j^{(+)}|,\rho_j$ are related to $T(z)$ [Eqs. (2.13c) and (2.15)] and $|c_j^{(+)}\rangle$ is related to $\hat{T}(z)$ by the formula

$$\hat{T}^{-1}(z_j)|c_j^{(+)}\rangle=0, \tag{4.6}$$

which follows by noting that $|\psi_j(n)\rangle=|\hat{\psi}_j(n)\rangle$ and taking into account the asymptotic behaviors (2.4a) and (2.5a).

The procedure to derive the time evolution of $T(z)[\hat{T}(z)]$ is closely analogous to that followed for the reflection coefficient.

The resulting evolution equations read

$$T_t(z)=[Z(z^{-1}),T(z)]+z^{-1}\tilde{H}(\lambda)T(z), \tag{4.7a}$$
$$\hat{T}_t(z)=[Z(z),\hat{T}(z)]-z\hat{T}(z)\tilde{G}(\lambda), \tag{4.7b}$$

where the time-dependent matrices $\tilde{H}(\lambda),\tilde{G}(\lambda)$ are defined as

$$\tilde{H}(\lambda)=D_1(\lambda)-D_2(\lambda)+zD_3(\lambda), \tag{4.7c}$$
$$\tilde{G}(\lambda)=D_1(\lambda)+D_2(\lambda)-z^{-1}D_3(\lambda), \tag{4.7d}$$

where

$$D_1(\lambda)=g_k(\lambda)\{\Pi(-\infty)\sigma_k\Pi^{-1})(-\infty)-\sigma_k\},$$
$$D_2(\lambda)=\sum_{j=-\infty}^{+\infty}\Pi(-\infty)\Pi^{-1}(j)[\tilde{g}_\nu(\underline{L})_{11}s_\nu(j)+\tilde{g}_\nu(\underline{L})_{12}u_\nu(j)+\tilde{f}_\nu(\underline{L})_{11}r_\nu(j)+\tilde{f}_\nu(\underline{L})_{12}v_\nu(j)]$$
$$\times\Pi(j+1)\Pi^{-1}(-\infty),$$
$$D_3(\lambda)=\sum_{j=-\infty}^{+\infty}[\tilde{g}_\nu(\underline{L})_{21}s_\nu(j)+\tilde{g}_\nu(\underline{L})_{22}u_\nu(j)+\tilde{f}_\nu(\underline{L})_{21}r_\nu(j)+\tilde{f}(\underline{L})_{22}v_\nu(j)],$$

having set

$$\tilde{f}_\nu(\underline{L})_{\kappa m}=\left(\frac{f_\nu(\underline{L})-f_\nu(\lambda)}{\underline{L}-\lambda}\right)_{\kappa m};\qquad \tilde{g}_\nu(\underline{L})_{\kappa m}=\left(\frac{g_\nu(\underline{L})-g_\nu(\lambda)}{\underline{L}-\lambda}\right)_{\kappa m}\quad(\kappa,m=1,2)$$

(where by the subscript indices κm we denote the corresponding matrix elements of the considered operator),

$$r_\nu(j)=B(j)\Pi(j)\sigma_\nu\Pi^{-1}(j+1)$$
$$-\Pi(j)\sigma_\nu\Pi^{-1}(j+1)B(j+1),$$
$$s_\nu(j)=[A(j),\sigma_\nu],$$
$$u_\nu(j)=[B(j),\sigma_\nu],$$
$$v_\nu(j)=\Pi(j+1)\sigma_\nu\Pi^{-1}(j)-\Pi(j)\sigma_\nu\Pi^{-1}(j+1).$$

Now performing the time-derivative of Eqs. (2.4c), (2.13c), and (4.6) and taking care of Eqs. (4.7a) and (4.7b), we get

$$|c_j^{(-)}\rangle_t=[Z(z_j^{-1})+z_j^{-1}\tilde{H}(\lambda_j)]|c_j^{(-)}\rangle+\gamma_j^{(-)}|c_j\rangle, \tag{4.8a}$$
$$\langle d_j^{(+)}|_t=-\langle d_j^{(+)}|Z(z_j^{-1})+\delta_j\langle d_j^{(+)}|, \tag{4.8b}$$
$$|c_j^{(+)}\rangle_t=Z(z_j)|c_j^{(+)}\rangle+\gamma_j^{(+)}|c_j^{(+)}\rangle, \tag{4.8c}$$

where $\gamma_j^{(\pm)},\delta_j$ are in principle arbitrary scalar time-dependent functions; however, as $|\psi_j(n)\rangle$ is determined up to a single arbitrary n-independent scalar function, it follows from formulas (2.4) that $\gamma_j^{(-)}=\gamma_j^{(+)}$. On the other hand, deriving formula (2.15) with respect to t, and comparing the resulting expression with Eq. (4.7a) one easily gets

$$(z_j)_t=0, \tag{4.9a}$$
$$(\rho_j)_t=-(\gamma_j^{(+)}+\delta_j)\rho_j. \tag{4.9b}$$

Of course, Eq. (4.9a) implies the isospectrality of the flow defined by the evolution equations (4.1). Finally, from for-

mulas (4.8b), (4.8c), and (4.9b) one gets

$$C_j(t) = U(z_j;t-t_0)C_j(t_0)U^{-1}(z_j^{-1};t-t_0). \tag{4.10}$$

C. Soliton solutions

As usually, soliton solutions are obtained from the Gel'fand–Levitan–Marchenko equation when the reflection coefficient vanishes identically on the unit circle, and therefore are expressed in terms of the bound state spectral data $\{z_j,C_j\}_{j=1,\dots,N}$, only the matrices C_j evolving in time according to Eq. (4.10). It is convenient to set

$$C_j(t) = \mu_j \exp[2\zeta_j\xi_j(t)]P_j(t), \tag{4.11}$$

where

$$z_j = \exp(-\zeta_j), \qquad \mu_j = 2\sinh\zeta_j,$$

$$P_j = \frac{|c_j^{(+)}\rangle\langle d_j^{(+)}|}{\langle d_j^{(+)}|c_j^{(+)}\rangle}. \tag{4.12}$$

It can be shown that, if we have N simple eigenvalues, the corresponding solution of Eq. (4.1) splits asymptotically into $\mathcal{N}$ separated solitions, whose "positions" and "polarizations" are given respectively by ξ_j and P_j. These parameters evolve in time according to the formulas

$$\xi_j(t) = \xi_j(t_0) + (2\zeta_j)^{-1} \times \ln\left\{\frac{\langle d_j^{(+)}|U^{-1}(z_j^{-1};t-t_0)U(z_j;t-t_0)|c_j^{(+)}(t_0)\rangle}{\langle d_j^{(+)}(t_0)|c_j^{(+)}(t_0)\rangle}\right\}, \tag{4.13}$$

$$P_j(t) = U(z_j;t-t_0)P_j(t_0)U^{-1}(z_j^{-1};t-t_0)\langle d_j^{(+)}(t_0)|c_j^{(+)}(t_0)\rangle \times[\langle d_j^{(+)}(t_0)|U^{-1}(z_j^{-1};t-t_0)U(z_j;t-t_0)|c_j^{(+)}(t_0)\rangle]^{-1} \tag{4.14}$$

As an example, we report explicitly the one-soliton solution, already derived in I:

$$A(n,t) = I + \sinh^2\zeta\,\mathrm{sech}^2\{\zeta[n+\tfrac{1}{2}-\xi(t)]\}P(t), \tag{4.15a}$$

$$B(n,t) = \sinh^2\zeta\,\mathrm{sech}\{\zeta[n-\tfrac{1}{2}-\xi(t)]\} \times\mathrm{sech}\{\zeta[n+\tfrac{1}{2}-\xi(t)]\}P(t). \tag{4.15b}$$

A detailed analysis of the time-behavior of such solution for a generic NDDE belonging to the class (4.1) will be performed in a separate paper.

5. AN INTERESTING SUBCLASS OF NDDE's

A. Definitions and soliton solutions

From the definition of the operator L Eq. (3.15) one can easily see that the initial condition $B(n,t_0) = 0$ for any NDDE of the class (4.1) implies $B(n,t) = 0$ iff the g_ν's are odd functions of their argument, while h_0 and the f_κ's are even functions. In this case, the system (4.1) reduces to a single equation for the potential $A(n,t)$, which reads

$$h(\mathcal{L})A_t(n) + g_0(\mathcal{L}[A(n-1)A(n) - A(n)A(n+1)] + f_\kappa\mathcal{L}[A(n),\sigma_\kappa] + g_\kappa(\mathcal{L})\mathcal{G}\sigma_\kappa, \tag{5.1}$$

where we have introduced the new operators

$$\mathcal{L}P(n) = A(n)(P(n)+P(n+1)) + S(n-1)A(n) - A(n)S(n+1) + \sum_{j=n}^{\infty}[P(j),A(n)], \tag{5.2a}$$

$$\mathcal{G}F = \Pi(n-1)F\Pi^{-1}(n+1) - \Pi(n)F\Pi^{-1}(n+2) - [F,A(n)] \tag{5.2b}$$

(F being any constant matrix).

To Eq. (4.4) corresponds the following linear evolution equation for the reflection coefficient:

$$h(\lambda^2)R_t(z) = -\lambda\mu g_0(\lambda^2)R(z) + f_\kappa(\lambda^2)[\sigma_\kappa,R(z)] + \tfrac{1}{2}\lambda g_\kappa(\lambda^2)(\lambda[\sigma_\kappa,R(z)] - \mu\{\sigma_\kappa,R(z)\}). \tag{5.3}$$

An analogous equation holds for the degenerate matrices $C_j(t)$, once replace z by z_j.

We notice that the requirement $B(n,t) = 0$ implies, through Eqs. (2.15), (2.16), and (2.17), that the reflection and transmission coefficients are even functions of z for any t. From this property it follows in particular that the discrete eigenvalues z_j occur in positive-negative pairs, each pair being associated to the same matrix C_j. The corresponding simplest soliton solution reads

$$A(n,t) = I + 2\sinh^2\zeta\cosh\zeta\,\mathrm{sech}\{\zeta[n+1-\xi(t)]\}\,\mathrm{sech}\{\zeta\{n-\xi(t)\}\}P(t). \tag{5.4}$$

B. Continuum limit

We introduce the parameter Δ to denote the spacing of the lattice and make the positions

$$x = n\Delta, \tag{5.5a}$$

$$z = \exp(i\kappa\Delta) \qquad \{z_j = \exp[(p_j\Delta)]\}, \tag{5.5b}$$

$$A(n) = \exp[-\Delta V(n)]\exp[+\Delta V(n+1)], \tag{5.5c}$$

where we have set

$$V(n) = \int_{n\Delta}^{\infty} dy\,Q(y) \tag{5.5d}$$

so that

$$V(n+1) = V(n) - \Delta Q(x) - \tfrac{1}{2}\Delta^2 Q_x(x) + o(\Delta^3). \tag{5.5e}$$

It is immediate to see that, with the above positions, as Δ goes to zero, the eigenvalue equation (2.1) with $B(n) = 0$ goes into the matrix Schrödinger equation[10]

$$-\Psi_{xx}(x) + Q(x)\Psi(x) = k^2\Psi(x). \tag{5.6}$$

In this subsection we shall show that, in the same limit, the whole class (5.1) goes into the class of NEE's associated with the spectral problem (5.6), already derived and investigated in CD. In particular, the first member of class (5.1) goes into the "boomeron" equation.[4,11] To perform the continuum limit it is convenient to start from the evolution equation in

the spectral space (5.13). From formula (5.5b), expanding the r.h.s. in power of Δ, it follows that

$$\mu = 2i\kappa\Delta + o(\Delta^3), \tag{5.7a}$$

$$\lambda = 2(1 - \tfrac{1}{2}(\kappa^2\Delta^2)) + o(\Delta^4). \tag{5.7b}$$

If we introduce the new arbitrary functions

$$\beta_\nu(-4\kappa^2) = g_\nu(\lambda^2), \tag{5.8a}$$

$$\alpha_l(-4\kappa^2) = \tilde{\alpha}_l(\lambda^2) = \Delta^{-1}[\tfrac{1}{2}\lambda^2 g_l(\lambda^2) + f_l(\lambda^2)], \tag{5.8b}$$

$$f_0(-4\kappa^2) = h(\lambda^2), \tag{5.8c}$$

and define the new time variable τ as follows:

$$\tau = \Delta^{-1}t, \tag{5.9}$$

it turns out that the reflection coefficient $R(z)$, which can be also considered as a function of the real variable k, evolves with respect to the new variable τ according to the formula

$$f_0(-4k^2)R_\tau(k) = 4ik\beta_0(-4\kappa^2)R(k) + 2ik\beta_l(-4k^2)\{\sigma_l, R(k)\} + \alpha_l(-4k^2)\{\sigma_l, R(k)\} + 0(\Delta^2), \tag{5.10}$$

which, in the limit $\Delta \to 0$, goes exactly into Eq. (4.2.3) of CD written for $\vec{\gamma} = 0$.

On the other hand, taking care of position (5.8b), the evolution equations (5.1) can be cast in the novel form

$$\begin{aligned} &h_0(\mathscr{L})A_t(n) + g_0(\mathscr{L})[A(n-1)A(n) - A(n)A(n+1)] \\ &\quad + \Delta\tilde{\alpha}_k(\mathscr{L})[A(n),\sigma_\kappa] + \tfrac{1}{2}g_k(\mathscr{L})[\sigma_\kappa A(n-1)A(n) - A(n)A(n+1)\sigma_\kappa] \\ &\quad + (2I - A(n))[A(n),\sigma_\kappa] + \left[A(n), \sum_{j=n}^{\infty}\right][A(j),\sigma_\kappa] + \Pi(n-1)\sigma_\kappa\Pi^{-1}(n+1) - \Pi(n)\sigma_\kappa\Pi^{-1}(n+2). \end{aligned} \tag{5.11}$$

Though, at a first glance, Eq. (5.11) may look more cumbersome that the original Eq. (5.1), it comes out to provide a more convenient starting point to evaluate the continuum limit.

First of all, we consider the simplest case, where the function g_ν and $\tilde{\alpha}_\kappa$ do not depend on $(\mathscr{L})$. After a tedious but straightforward calculation, taking care of the positions (5.5), one gets

$$A(n-1)A(n) - A(n)A(n+1) = 2\Delta^3 Q_x(x) + o(\Delta^4), \tag{5.12a}$$

$$[A(n),\sigma_\kappa] = \Delta^2[\sigma_\kappa, Q(x)] + o(\Delta^3), \tag{5.12b}$$

$$\begin{aligned} &\sigma_\kappa A(n-1)A(n) - A(n)A(n+1)\sigma_\kappa + (2I - A(n)) \\ &\quad \times [A(n),\sigma_\kappa] + \left[A(n), \sum_{j=n}^{\infty}[A(j),\sigma_\kappa] + \Pi(n-1)\sigma_\kappa\right. \\ &\quad \left.\times\Pi^{-1}(n+1) - \Pi(n)\sigma_\kappa\Pi^{-1}(n+2)\right. \\ &\quad = 2\Delta^3\{\{\sigma_\kappa, Q_x(x)\} + [Q(x),[V(x),\sigma_\kappa]]\} \end{aligned} \tag{5.12c}$$

and thus, dividing by Δ^3 and then letting $\Delta \to 0$,

$$\begin{aligned} Q_\tau(x;\tau) &= 2\beta_0 Q_x(x;\tau) + \alpha_\kappa[\sigma_\kappa, Q(x;\tau)] \\ &\quad + \beta_\kappa[\{\sigma_\kappa, Q_x(x;\tau)\} + [Q(x;\tau),[V(x;\tau),\sigma_\kappa]]] \end{aligned} \tag{5.13}$$

which is exactly the "boomeron" equation. To recover the higher NEE's associated to the matrix Schrödinger spectral problem, it is necessary to relate the discrete operator $\mathscr{L}$ defined by formula (5.2a) to the continuous operator $\underline{L}_c$ introduced in CD Eqs. (4.1.3) and (4.1.4). The correct connection is suggested by the analogous relation holding for discrete and continuous formulas in the spectral space: so we can assert that, due to the correspondence between application of $\underline{L}_c$ $(\mathscr{L})$ and multiplication by $-4\kappa^2$ (λ^2) in the continuous (discrete) case, the continuous operator $\underline{L}_c$ is the limit, as $\Delta \to 0$, of the discrete operator $(\mathscr{L} - 4I)\Delta^{-2}$. Indeed, if we apply this operator to each of the arguments involved in Eq. (5.11), which once expanded in powers of Δ, all exhibit the structure

$$\Delta^{-3}P(n) = F(x) + o(\Delta), \tag{5.14}$$

we obtain

$$\Delta^{-2}(\mathscr{L} - 4I)(\Delta^{-3}P(n)) = L_c F(x) + o(\Delta), \tag{5.15}$$

where

$$\begin{aligned} \underline{L}_c(x) &= F_{xx}(x) - 2\{Q(x), F(x)\} + \{Q_x(x), \int_x^\infty dx' F(x')\} \\ &\quad + \left[Q(x), \left[\int_x^\infty dx'\, Q(x'), \int_{x'}^\infty dx''\, F(x'')\right]\right], \end{aligned} \tag{5.16}$$

which is just the continuous operator obtained in CD. Hence, taking into account formulas (5.5), (5.12), and (5.5) it is a trivial task to realize that the class of NEED's (5.11) has the continuum limit

$$\begin{aligned} f_0(\underline{L}_c)Q_\tau(x;\tau) &= 2\beta_0(\underline{L}_c)Q_x(x;\tau) \\ &\quad + \alpha_\kappa(\underline{L}_c)[\sigma_\kappa, Q(x;\tau)] + \beta_k\underline{L}_c)G\sigma_\kappa \end{aligned} \tag{5.17}$$

which is the CD class of NEE's, the operator G being defined there through formula (4.1.4).

6. BÄCKLUND TRANSFORMATIONS

A. Basic formulas

First of all, we note that Eq. (3.12) implies that, if two pairs of potentials $(A'(n), B'(n)), (A(n), B(n))$ are related by the formula

$$\gamma_\nu(\underline{\Lambda})\begin{pmatrix} A'(n)\sigma_\nu - \sigma_\nu A(n) \\ B'(n)\sigma_\nu - \sigma_\nu B(n) \end{pmatrix} = \delta_\nu(\underline{\Lambda})\begin{pmatrix} B'(n)\Pi'(n)\sigma_\nu\Pi^{-1}(n+1) - \Pi'(n)\sigma_\nu\Pi^{-1}(n+1)B(n+1) \\ \Pi'(n-1)\sigma_\nu\Pi^{-1}(n) - \Pi'(n)\sigma_\nu\Pi^{-1}(n+1) \end{pmatrix}, \tag{6.1}$$

γ_ν and δ_ν being arbitrary entire functions, for the corresponding reflection coefficients $R(z), \bar{R}'(z')$ the following relationship holds:

$$\gamma_\nu(\lambda)[\sigma_\nu R(z) - \bar{R}'(z)\sigma_\nu] = \delta_\nu(\lambda)[z\sigma_\nu R(z) - z^{-1}\bar{R}'(z)\sigma_\nu]. \tag{6.2}$$

Denoting by $Y(z)$ the matrix $[\gamma_\nu(\lambda) - z\delta_\nu(\lambda)]\sigma_\nu$ the transformations (6.2) can be cast in the more compact form

$$Y(z)R(z) = \bar{R}'(z)Y(z^{-1}). \tag{6.3}$$

Among all the transformations belonging to the class (6.3), it is worthwhile to focus on those leading from a solution of a given NDDE of the class (4.1) to a new solution of the same NDDE, that is on the so-called "auto-Backlund" transformation, defined by the subclass of (6.3) which commute with the time-evolution (4.1). It is immediately seen that this subclass consists of matrices $Y(z)$ such that

$$[Z(z), Y(z)] = 0. \tag{6.4}$$

In the following we will restrict considerations to "auto-Bäcklund" transformations. Among such auto-Bäcklund transformations, we focus attention on those relating solutions with a different number of bound states (corresponding to a different number of solitons), which are given by the poles of the respective reflection coefficients $R(z), \bar{R}'(z) = R'(z)$ provided they can be analytically continued inside the unit circle.[12]

From Eq. (6.3) it follows that, if $Y(z^{-1})$ has some singular point inside the unit circle, $R'(z)$ supports a corresponding larger number of poles than $R(z)$.

Let us consider for simplicity a matrix $Y(z)$ of the form

$$Y(z) = [\gamma(\lambda) - z\delta(\lambda)]\sigma_0. \tag{6.5}$$

It implies

$$R'(z) = R(z)(\gamma(\lambda) - z\delta(\lambda))/(\gamma(\lambda) - z^{-1}\delta(\lambda)) \tag{6.6}$$

which shows that, generally speaking, $R'(z)$ has a pole not belonging to $R(z)$ whenever the analytic function of z, $\gamma(\lambda) - z^{-1}\delta(\lambda)$, has a zero. The simplest case occurs when $\gamma(\lambda) = \bar{\gamma}, \delta(\lambda) = \bar{\delta}$, so that only one new pole is generated, located at

$$\bar{z} = \bar{\delta}/\bar{\gamma} \tag{6.7}$$

and we assume of course $|\bar{\delta}/\bar{\gamma}| < 1$.

Formula (6.6) can then be rewritten in the form

$$R'(z) = R(z)(1 - z\bar{z})/(1 - \bar{z}^{-1}\bar{z}). \tag{6.8}$$

The corresponding pairs of potentials are related by the formula [see Eq. (6.1)]

$$\begin{aligned} A'(n) - A(n) &= \bar{z}[B'(n)\Pi'(n)\Pi^{-1}(n+1) \\ &\quad - (\Pi'(n)\Pi^{-1}(n+1)B(n+1)], \\ B'(n) - B(n) &= \bar{z}[\Pi'(n-1)\Pi^{-1}(n) - \Pi'(n)\Pi^{-1}(n+1)], \end{aligned} \tag{6.9}$$

where the pair $(A'(n), B'(n))$ supports one more soliton than the pair $(A(n), B(n))$.

For instance, if we set $(A(n), B(n)) = (I, 0)$ then $(A'(n), B'(n))$ is just the one-soliton solution.

Indeed, expressing $A'(n)$ in terms of the products $\Pi'(n)$ the system (6.9) gives immediately

$$\begin{aligned} &\Pi'(n+1) - \Pi'(n) = \bar{z}B'(n)\Pi'(n), \\ &B'(n) = z[\Pi'(n-1) - \Pi'(n)]. \end{aligned} \tag{6.10}$$

The resulting nonlinear functional equation for $\Pi'(n)$ can be solved by setting

$$\Pi'(n) = I + \alpha f(\beta n + \gamma)P, \tag{6.11}$$

P being some one-dimensional projector, which implies

$$\begin{aligned} f(x+\beta) - f(x) &= \bar{z}^2[f(x-\beta) - f(x)] \\ &\quad \times [1 + \alpha f(x)][1 + \alpha f(x+\beta)], \end{aligned} \tag{6.12}$$

where $x = \beta n + \gamma$, and the parameters $\bar{z}, \alpha, \beta$ are no longer independent.

Expanding (6.12) in powers of β, having set $\alpha = \beta + o(\beta^2), \bar{z}^2 = 1 - 2\beta + o(\beta^2)$, we get for small β the solution $f(x) = 1 - \tanh x$, which comes out to fulfill (6.12) identically (i.e. for any x and β) iff

$$\bar{z} = \exp(-\beta), \qquad \alpha = \tfrac{1}{2}[\exp(2\beta) - 1] \tag{6.13}$$

so that the one-soliton solution reads

$$\Pi'(n) = I + \tfrac{1}{2}[\exp(2\beta) - 1][1 - \tanh(\beta n + \gamma)]P. \tag{6.14}$$

This solution can be identified with that given by Eq. (5.4) with the positions: $\beta = \zeta, \gamma = \zeta(\tfrac{1}{2} - \xi)$.

As a final remark on this topic, we notice that, unlike the continuous case, the practical usefulness of Backlund transformations in order to find explicitly novel solutions of the same NDDE is rather questionable; actually, even in the simplest case we treated above, we had to solve a nonlinear functional equation, and we succeeded only through some *ad hoc* positions which can be hardly generalized to more complex cases. Of course, those computational difficulties do not affect at all the theoretical relevance of Backlund transformations.

7. CONCLUSIONS

The existence of a discrete counterpart also for the NEE's associated to the matrix Schrödinger spectral problem, as it was previously found[3,13] for the scalar Schrödinger and for the Zakharov–Shabat spectral problems, again raises the question whether the discretizing procedure based upon the underlying linear spectral problem, is the optimal one in order to solve completely integrable NEE's numerically.[14] It seems certainly worthwhile to compare the results which can be obtained this way to those arising from other more standard numerical techniques.

Besides this general problem, there remain to be investigated in more detail some other problems specifically related to the subject treated in the present paper: for instance, one can try to settle for our NDDE's a reduction technique analogous to that derived by Degasperis[15] for the continuous case, aiming at identifying subclasses of equations which can be relevant for applications. It would be also interesting to prove that our dynamical systems display an infinite sequence of conserved quantities, and to construct them explicitly.

Furthermore, the isospectrality of the flow (4.1) suggests the existence of a Lax pair for any equation of this class. Work is in progress in all these directions.

[1] M. Bruschi, D. Levi, S. V. Manakov, and O. Ragnisco, "The nonabelian Toda-lattice: Discrete analogue of the matrix Schrödinger spectral problem," J. Math. Phys. **21**, 2749 (1980).

[2]See for example: V. A. Zakharov and A. V. Mikhailov, Zh. Eksp. Teor. Fiz. **74**, 1953 (1978) [Soviet Phys. JETP **47**, 1017 (1978)].

[3]In the abelian case, the analogous problem was solved by R. K. Dodd, J. Phys. A, Math. Gen. **11**, 81 (1978); see also M. Bruschi, D. Levi, and O. Ragnisco, J. Phys. A, Math. Gen. (1980) (in press).

[4]F.Calogero and A. Degasperis, Nuovo Cimento B **39**, (1977); hereafter often referred to as CS.

[5]They are related to the elements of the monodromy matrix defined in I, Eqs. (2.5a) and (2.5b), as follows: $R(z) = b(z)a^{-1}(z)$, $T(z) = a^{-1}(z)$.

[6]The exact relationship between $|d_j^{(+)}\rangle$ and $|c_j^{(-)}\rangle$ is given in I, Eqs. (2.12) and (2.13).

[7]To simplify the formulas, we will always use in the following the shorthand notation $W'(n)$ in place of $W'(\bar{\Psi}'(n,z), \tilde{\Psi}'(n,z))$ and will also omit the argument z in the expression of the wave functions.

[8]Throughout this paper we understand the products to be ordered from the lower to the upper value of the running index. Note that the infinite products introduced in formula (3.6b) are well defined, due to the boundary condition (2.2a).

[9]We note that if $g_k = 0$, the evolution of the transmission coefficient is given by the similarity transformation

$$T(z,t) = U(z^{-1};t - t_0)T(z,t)U^{-1}(z^{-1};t - t_0).$$

[10]The discrete analog of the Schrödinger problem in the scalar case was extensively investigated by K. M. Case: K. M. Case and M. Kac, J. Math. Phys. **14**, 594 (1973); K. M. Case, **14**, 916 (1973); **15**, 143 (1974).

[11]F. Calogero and A. Degasperis, Lett. Nuovo Cimento **16**, 425 (1976).

[12][3]So far we never supposed this property; we now assume analiticity of $R(z)$ just for the sake of simplicity.

[13]M. J. Ablowitz and J. F. Ladik, J. Math. Phys. **16**, 598 (1975); **17**, 1011 (1976).

[14]M. J. Ablowitz and J. F. Ladik, Stud. Appl. Math., **55**, 213 (1976); **57**, 1 (1977).

[15]F. Calogero and A. Degasperis, "Reduction technique for matrix nonlinear evolution equations," J. Math. Phys. **22**, 23 (1981).

Some properties of Borel summable functions

G. Auberson and G. Mennessier
Département de Physique Mathématique,[a] U.S.T.L., 34060 Montpellier Cedex, France

(Received 15 October 1980; accepted for publication 16 January 1981)

Classes of Borel summable functions are defined and studied. Some properties of these functions, which are useful for high-order calculations pertaining to certain physical theories, are proved and discussed. They include reciprocal Watson-like theorems, sufficient conditions for membership in the classes, and asymptotic behaviors of the expansion coefficients after composition of functions.

PACS numbers: 02.30.Lt

1. INTRODUCTION

Recently, the combination of large order estimates with Borel summation was revealed as a powerful tool to push the use of perturbation expansions outside the weak coupling regime, both in quantum mechanics and quantum field theory.[1] To be mathematically satisfactory, such a treatment should include (i) a proof of the Borel summability and (ii) a careful justification of the rather intricate calculations needed to obtain large order behaviors. It also requires the manipulation of various Borel summable functions (especially to get an expansion of the asymptotic perturbation coefficients in inverse powers of the order). In practice, the steps (i) and (ii) are difficult and often (but not always[2,3]) by-passed (the Borel summability being, for instance, simply assumed if compatible with the large order estimates). As for the formal calculations on Borel summable functions, they are more easily justified (or even performed) when appeal is made to some general properties of these functions. The purpose of the present paper is to collect such properties, to give their proofs when necessary, and to comment on them. Some of these properties (at least in weaker versions) belong to the "know-how" of high-order practitioners. Others are new. They are presented here in the accurate form of theorems, some of which are mainly refinements of results imported from the mathematical literature. In order to state these theorems in a simple form we need to define suitable classes of Borel summable functions (Sec. 2). These classes are "optimal" in the sense that (i) they make the two Watson-like theorems of Sec. 3 fully reciprocal statements and (ii) they are invariant under the usual operations of algebra and analysis (Theorems 3 and 4 of Sec. 4). As for Theorem 5, which gives sufficient conditions for membership in these classes, it is of great use in proofs of Borel summability.[3] Theorems 6 and 7 have to deal with the composition of Borel summable functions, and have direct applications to high-order calculations.[4]

2. THE CLASSES $\mathcal{N}$ AND $\mathcal{W}$

As is well known, the class of Borel summable functions include those functions $f(z)$ which (a) are analytic within some sectorlike domain of the complex z plane (assumed to be centered at the origin) with sufficiently large opening angle; (b) admit an asymptotic power series

$$f(z) = \sum_{n=0}^{M-1} f_n z^n + R_M(z), \tag{1}$$

with remainders $R_M(z)$ of limited and uniform growth when $M \to \infty$. The "natural" analyticity domain which is referred to in (a) turns out to depend on two parameters λ and R ($0 \leqslant \lambda \leqslant \frac{1}{2}\pi$, $0 < R \leqslant \infty$). It will be defined as the kidney-shaped region (Fig. 1)

$$K(\lambda,R) = \bigcup_{|\theta| < \lambda} e^{i\theta} K(0,R), \tag{2}$$

where $K(0,R)$ is the circle $\mathrm{Re}(1/z) > 1/R$. As for the precise statement of condition (b), we shall give it in two different forms to which correspond two useful subclasses ("$\mathcal{N}$" and "$\mathcal{W}$") of Borel summable functions (and two variants of the forthcoming Watson-like theorem).

Definition 1: Given $R,\sigma > 0$ and $\lambda \geqslant 0$, we shall say that f belongs to $N-(\lambda,R,\sigma)$ if: (i) $f(z)$ is analytic in the domain $K(\lambda,R)$ as defined in Eq. (2), (ii) there exists a constant $A(\lambda,R,\sigma)$ such that $f(z)$ admits the asymptotic expansion (1), with

$$|R_M(z)| \leqslant A(\lambda,R,\sigma) M! |\sigma z|^M \quad \forall z \in K(\lambda,R) \text{ and } M = 0,1,\cdots \tag{3}$$

[$R_0(z)$ is identified with $f(z)$].

We next define the class $\mathcal{N}$ by

$$\mathcal{N}-(\lambda,R,\sigma) = \bigcap_{0 < R' < R, \sigma' > \sigma} N-(\lambda,R',\sigma'). \tag{4}$$

Definition 2: Given $\lambda,R,\sigma > 0$, $f \in W-(\lambda,R,\sigma)$ if (i) $f(z)$ is analytic in $K(\lambda,R)$ and (ii) $\exists A(\lambda,R,\sigma)$ such that $f(z)$ admits the asymptotic expansion (1) with

$$|R_M(z)| \leqslant A(\lambda,R,\sigma) M! |\sigma z|^M \times \begin{cases} 1 & \text{if } |\arg z| \leqslant \lambda \\ [\cos(|\arg z| - \lambda)]^{-M} & \text{if } \lambda \leqslant |\arg z| < \frac{1}{2}\pi + \lambda \end{cases} \tag{3'}$$

for all $z \in K(\lambda,R)$ and $M = 0,1,\cdots$.

The class $\mathcal{W}$ is defined by

$$\mathcal{W}-(\lambda,R,\sigma) = \bigcap_{0 < \lambda' < \lambda, 0 < R' < R, \sigma' > \sigma} W-(\lambda',R',\sigma') \tag{4'}$$

Remarks: (1) If $f \in N-$ or $W-(\lambda,R,\sigma)$, then

$$|f_n| \leqslant A(\sigma) n! \sigma^n, \quad n = 0,1,\cdots. \tag{5}$$

Indeed, $f_n = z^{-n}[R_n(z) - R_{n+1}(z)]$, and it suffices to make $z \to 0$ within $K(\lambda,R)$ on account of Eqs. (3) or (3').

(2) For $\lambda > 0$ it is obvious that

[a] Physique Mathématique et Théorique, Equipe de Recherche associated with C.N.R.S.

 0022-2488/81/112472-10$01.00

$$N(\text{resp.}\mathcal{N}) - (\lambda,R,\sigma) \subset W(\text{resp.}\mathcal{W}) - (\lambda,R,\sigma). \quad (6)$$

This inclusion relation is a strict one (as it will appear on the example following Theorem 2). On the other hand, for any μ such that $0 \leqslant \mu < \lambda$,

$$W(\text{resp.}\mathcal{W}) - (\lambda,R,\sigma) \subset N(\text{resp.}\mathcal{N}) - (\mu,R,\tau), \quad \text{where } \tau = \sigma/\sin(\lambda-\mu) \quad (7)$$

[this results from $K(\mu,R) \subset K(\lambda,R)$ and from the fact that $\sigma \leqslant \tau\cos(\varphi-\lambda)$ for $\lambda \leqslant \varphi \leqslant \frac{1}{2}\pi + \mu$].

(3) A function $f(z)$ which for some $b > 1$ enjoys properties (i) and (ii) of Definitions 1 (or 2) with $\Gamma(M+b)$ in place of $M!$, obviously belongs to $\mathcal{N}$ (or $\mathcal{W}$) $-(\lambda,R,\sigma)$.

(4) One can always choose the constants A in Eqs. (3) and (3') such that

$$A(\lambda',R',\sigma) \leqslant A(\lambda,R,\sigma), \quad \text{when } \lambda' < \lambda, R' < R. \quad (8)$$

3. BOREL TRANSFORM PROPERTIES

Let $\sum_{n=0}^{\infty} f_n z^n$ be a formal power series. Its *Borel transform* $\varphi(t)$ is defined by

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{f_n}{n!} t^n. \quad (9)$$

If this series converges in some disk, it uniquely defines $\varphi(t)$ as an analytic function.

On the other hand, let $\varphi(t)$ be an analytic function in a neighborhood of the positive real axis. Then its *inverse Borel transform* is defined by

$$f(z) = (1/z)\int_0^{\infty} dt e^{-t/z}\varphi(t). \quad (10)$$

If the Borel transform $\varphi(t)$ of a formal power series can be continued in a neighborhood of the positive real axis, and if the integral in Eq. (10) converges for some values of z, then $f(z)$ is called the *Borel sum* of the formal power series for those values of z. In favorable circumstances, the inverse Borel transform $f(z)$ turns out to be an analytic function whose asymptotic expansion at the origin concides with the formal power series. In this sense, the combination of formula (9) with analytic continuation and formula (10) realizes the reconstruction of the function $f(z)$ from its coefficients f_n, even in the case where the series $\sum_{n=0}^{\infty} f_n z^n$ diverges for all nonzero values of z.[5,6] Such a program clearly *can* make sense for functions in $\mathcal{N}$ or $\mathcal{W} - (\lambda,R,\sigma)$ since, according to Eq. (5), $\varphi(t)$ is analytic in the disk $|t| < 1/\sigma$. That it actually *does* is the main content of the following two theorems.

Theorem 1: (1) If a function $f(z)$ belongs to $\mathcal{N} - (\lambda,R,\sigma)$, then its Borel transform $\varphi(t)$ [defined through Eqs. (1) and (9)] has the following properties

(a) $\varphi(t)$ is analytic in the "corolla":

$$\begin{cases} C(\lambda,\sigma) = \bigcup_{|\theta| \leqslant \lambda} e^{i\theta} C(0,\sigma), \\ C(0,\sigma) = \{t \mid \text{dist}(t,R_+) < 1/\sigma\} \end{cases} \quad (11)$$

(see Fig. 2).

(b) For all $R' < R$ and $\sigma' > \sigma$,

$$|\varphi(t)| \leqslant B(\lambda,R',\sigma') e^{|t|/R'}, \quad \forall t \in C(\lambda,\sigma'). \quad (12)$$

(2) Conversely, a function $\varphi(t)$ which fulfills conditions (a) and (b) is the Borel transform of a function $f(z) \in \mathcal{N} - (\lambda,R,\sigma)$ uniquely defined by Eq. (10).

This improved version of Watson's theorem[6,7] is nothing but a reformulation of a theorem due to Nevanlinna[8] (hence our notation $\mathcal{N}$ for the corresponding class) and rediscovered by Sokal.[9] We shall not give its proof, which can be found in Ref. 9. It was the merit of Sokal to recognize that the analyticity domain of $f(z)$ is not required to enter the region $|\arg z| > \pi/2$ for a Watson-like theorem to be valid (thus allowing the value $\lambda = 0$ in its formulation).

The next theorem, which is an extension of Watson's theorem in another direction, is presumably new. It is more like the original Watson's formulation in the sense that the value $\lambda = 0$ is not allowed (hence our notation $\mathcal{W}$ for the corresponding class), but the restrictions on the behavior of the remainders $R_M(z)$, as given by Eq. (3'), are substantially weakened. We shall give its full proof (partly made of Laplace transform arguments in disguise), although the first part (1a') is essentially borrowed from Hardy.[6]

Theorem 2: (1) If a function $f(z)$ belongs to $\mathcal{W} - (\lambda,R,\sigma)$, then its Borel transform $\varphi(t)$ [defined through Eqs. (1) and (9)] has the following properties.

(a') $\varphi(t)$ is analytic in the "keyhole":

$$H(\lambda,\sigma) = \{t \mid |t| < 1/\sigma\} \cup \{t \mid |\arg t| < \lambda\} \quad (11')$$

(see Fig. 3).

(b') For all $\lambda' < \lambda$, $R' < R$, and $\sigma' > \sigma$:

$$|\varphi(t)| \leqslant B(\lambda',R',\sigma') e^{|t|/R'}, \quad \forall t \in H(\lambda',\sigma'). \quad (12')$$

(2) Conversely, a function $\varphi(t)$ which fulfills conditions (a') and (b') is the Borel transform of a function $f(z) \in \mathcal{W} - (\lambda,R,\sigma)$ uniquely defined by Eq. (10).

Proof: (1a') Let $f \in \mathcal{W} - (\lambda,R,\sigma)$. Then $f \in W - (\lambda',R',\sigma')$ for any $\lambda' < \lambda$, $R' < R$, $\sigma' > \sigma$. From Eq. (5) the Borel transform $\varphi(t)$ of $f(z)$ is analytic in the disk $|t| < 1/\sigma$. Consider now:

$$a(t) \equiv \frac{1}{2i\pi}\oint_{\partial K(\lambda',R')} dz \frac{1}{z} e^{t/z} f(z). \quad (13)$$

The function $f(z)$ being bounded on the boundary $\partial K(\lambda',R')$, this integral is absolutely convergent for $|\arg t| < \lambda'$ and defines a function (independent of λ',R') analytic in the sector $|\arg t| < \lambda$. We show that $a(t)$ and $\varphi(t)$ coincide on a segment of the positive real axis. Indeed, using Eq. (1) in (13), one has for $t > 0$,

$$|a(t) - \sum_{n=0}^{M-1} \frac{f_n}{n!} t^n| \leqslant \frac{1}{2\pi}\oint_{\partial K(\lambda',R')} \left|\frac{dz}{z}\right| e^{t\,\mathrm{Re}(1/z)} |R_M(z)|. \quad (14)$$

On the boundary of $K(\lambda',R')$,

$$\begin{cases} |z| = R' & \text{when } |\arg z| \leqslant \lambda', \\ |z| = R'\cos(|\arg z| - \lambda') & \text{when } \lambda' \leqslant |\arg z| \leqslant \frac{1}{2}\pi + \lambda', \end{cases}$$

so that Eq. (3') implies

$$|R_M(z)| \leqslant A(\lambda',R',\sigma') M! (\sigma' R')^M, \quad \forall z \in \partial K(\lambda',R'). \quad (15)$$

Then,

$$|a(t)-\sum_{n=0}^{M-1}\frac{f_n}{n!}t^n|\leqslant\frac{1}{2\pi}A(\lambda',R',\sigma')M!(\sigma'R')^M\oint_{\partial K(\lambda',R')}|\frac{dz}{z}|e^{t\,\mathrm{Re}(1/z)}=\frac{1}{2\pi}A(\lambda',R',\sigma')M!(\sigma'R')^M\left[\int_{-\lambda'}^{\lambda'}d\psi e^{(t/R')\cos\psi}\right.$$
$$\left.+2\int_0^\infty\frac{dv}{[(1/R')^2+v^2]^{1/2}}e^{(t/R')\cos\lambda'-t\sin\lambda' v}\right]\leqslant(1/\pi)A(\lambda',R',\sigma')M!(\sigma'R')^M e^{t/R'}(\lambda'+R'/t\sin\lambda'). \tag{16}$$

For any t, $0<t<1/\sigma'$, and M large enough, choose $R'=t/M$. In the limit $M\to\infty$, $A(\lambda',t/M,\sigma')\leqslant A(\lambda',R/2,\sigma')$ and the right-hand side of inequality (16) vanishes. Therefore, $a(t)$ coincides with the Borel transform $\varphi(t)$ and is analytic in $H(\lambda,\sigma)$.

(1b') Given λ', R', and σ' ($\lambda'<\lambda$, $R'<R$, $\sigma'>\sigma$), choose $\lambda''=(\lambda+\lambda')/2$. According to (1a'), we can write

$$\varphi(t)=\frac{1}{2i\pi}\oint_{\partial K(\lambda'',R')}dz\frac{1}{z}e^{t/z}f(z),\quad|\arg t|<\lambda''. \tag{17}$$

Then, Eq. (3') implies (for $t=|t|e^{i\theta}$)

$$|\varphi(t)|\leqslant\frac{1}{2\pi}A(\lambda'',R',\sigma')\oint_{\partial K(\lambda'',R')}|\frac{dz}{z}|e^{\mathrm{Re}(t/z)}=\frac{1}{2\pi}A(\lambda'',R',\sigma')\left[\int_{-\lambda''}^{\lambda''}d\psi e^{(|t|/R')\cos(\theta-\psi)}\right.$$
$$\left.+\left(\int_0^\infty\frac{dv}{[(1/R')^2+v^2]^{1/2}}\exp\left(\frac{|t|}{R'}\cos(\lambda''+\theta)-|t|\sin(\lambda''+\theta)v\right)+(\theta\to-\theta)\right)\right]$$
$$\leqslant\frac{1}{2\pi}A(\lambda'',R',\sigma')e^{|t|/R'}\left[2\lambda''+\frac{R'}{|t|}\left(\frac{1}{\sin(\lambda''+\theta)}+\frac{1}{\sin(\lambda''-\theta)}\right)\right]. \tag{18}$$

Hence, for $|\theta|<\lambda'$,

$$|\varphi(t)|\leqslant\frac{1}{\pi}A\left(\frac{\lambda+\lambda'}{2},R',\sigma'\right)\left[\lambda+\frac{R}{|t|\sin\frac{1}{2}(\lambda-\lambda')}\right]e^{|t|/R'}. \tag{19}$$

This bound has the form of (12') for $|t|>\epsilon$, and is immediately extended to the whole of $H(\lambda',\sigma')$.

(2i) Let $\varphi(t)$ be a function verifying conditions (a') and (b'). Since $\varphi(t)$ is analytic in the disk $|t|<1/\sigma$, it can be represented there as

$$\varphi(t)=\sum_{n=0}^\infty\frac{f_n}{n!}t^n, \tag{20}$$

with

$$|f_n|\leqslant A(\bar\sigma)n!\bar\sigma^n,\quad n=0,1,\cdots \tag{21}$$

for any $\bar\sigma>\sigma$. Defining the function $f(z)$ by Eq. (10), one easily checks, by using the bound (12') and rotating the integration path, that $f(z)$ is an analytic function in $K(\lambda,R)$.

(2ii) Defining now,

$$\Delta_M(t)=\varphi(t)-\sum_{n=0}^{M-1}\frac{f_n}{n!}t^n,\quad M=0,1,\cdots, \tag{22}$$

we obtain

$$R_M(z)\equiv f(z)-\sum_{n=0}^{M-1}f_n z^n$$
$$=\frac{1}{z}\int_0^\infty d\tau e^{i\theta}\exp\left(-\frac{1}{z}\tau e^{i\theta}\right)\Delta_M(\tau e^{i\theta}),(|\theta|<\lambda). \tag{23}$$

For any $\lambda'<\lambda$, $R'<R$, $\sigma'>\sigma$, and $z\in K(\lambda',R')$, let us choose λ_0 such that $\lambda'<\lambda_0<\lambda$, and

$$\theta=\begin{cases}\arg z & \text{if } |\arg z|\leqslant\lambda_0,\\ \pm\lambda_0 & \text{if } \lambda_0\leqslant\pm\arg z<\frac{1}{2}\pi+\lambda',\end{cases} \tag{24}$$

in Eq. (23). Then

$$|R_M(z)|\leqslant\frac{1}{|z|}\int_0^\infty d\tau e^{-\tau\cos\alpha/|z|}|\Delta_M(\tau e^{i\theta})|, \tag{25}$$

where

$$\alpha=\begin{cases}0 & \text{if } |\arg z|\leqslant\lambda_0,\\ |\arg z|-\lambda_0 & \text{if } \lambda_0\leqslant|\arg z|<\frac{1}{2}\pi+\lambda'.\end{cases} \tag{26}$$

To bound $|\Delta_M(\tau e^{i\theta})|$ we choose σ_0, σ_1, λ_1, and R_1 such that $\sigma<\sigma_1<\sigma_0<\sigma'$; $\lambda_0<\lambda_1<\lambda$, and $R'<R_1<R$. Then, using Eqs. (20), (21), and (12'), we obtain for $\tau\leqslant1/\sigma_0$,

$$|\Delta_M(\tau e^{i\theta})|\leqslant\sum_{n=M}^\infty\frac{|f_n|}{n!}\tau^n\leqslant A(\sigma_1)\sum_{n=M}^\infty(\sigma_1\tau)^n$$
$$\leqslant A(\sigma_1)\frac{\sigma_0}{\sigma_0-\sigma_1}(\sigma'\tau)^M, \tag{27}$$

and for $\tau\geqslant1/\sigma_0$,

$$|\Delta_M(\tau e^{i\theta})|\leqslant|\varphi(\tau e^{i\theta})|+\sum_{n=0}^{M-1}\frac{|f_n|}{n!}\tau^n$$
$$\leqslant B(\lambda_1,R_1,\sigma_1)e^{\tau/R_1}+A(\sigma_0)\sum_{n=0}^{M-1}(\sigma_0\tau)^n$$
$$\leqslant B(\lambda_1,R_1,\sigma_1)e^{\tau/R_1}+A(\sigma_0)M(\sigma_0\tau)^M$$
$$\leqslant B(\lambda_1,R_1,\sigma_1)e^{\tau/R_1}+\frac{A(\sigma_0)}{e\log(\sigma'/\sigma_0)}(\sigma'\tau)^M. \tag{28}$$

Inserting Eqs. (27) and (28) into Eq. (25) and noticing that $|z|\leqslant R'\cos\alpha$ when $z\in K(\lambda',R')$, we find

$$|R_M(z)|\leqslant\frac{D(\sigma')}{|z|}\int_0^\infty d\tau e^{-\tau\cos\alpha/|z|}(\sigma'\tau)^M+\frac{E(\lambda',R',\sigma')}{|z|}\int_{1/\sigma_0}^\infty d\tau e^{-\tau\cos\alpha/|z|+\tau/R_1}$$
$$=D(\sigma')M!\frac{|\sigma'z|^M}{(\cos\alpha)^{M+1}}+\frac{E(\lambda',R',\sigma')}{\cos\alpha-|z|/R_1}\exp\left(-\frac{\cos\alpha}{\sigma_0|z|}+\frac{1}{\sigma_0R_1}\right)$$
$$=\frac{D(\sigma')}{\cos\alpha}M!\left|\frac{\sigma'z}{\cos\alpha}\right|^M\left\{1+\left[\frac{E(\lambda',R',\sigma')}{1-R'/R_1}e^{1/\sigma_0R_1}\right]\left(\frac{\sigma_0}{\sigma'}\right)^M\frac{1}{M!}\left(\frac{\cos\alpha}{\sigma_0|z|}\right)^M e^{-\cos\alpha/\sigma_0|z|}\right\}$$

$$\leqslant \frac{D(\sigma')}{\sin(\lambda_0-\lambda')} M! \left|\frac{\sigma' z}{\cos\alpha}\right|^M \left[1 + F(\lambda',R',\sigma')\left(\frac{\sigma_0}{\sigma'}\right)^M \frac{M^M e^{-M}}{M!}\right]$$

$$\leqslant G(\lambda',R',\sigma') M! \left|\frac{\sigma' z}{\cos\alpha}\right|^M. \tag{29}$$

In view of Eqs. (3') and (26), we readily deduce from Eq. (29) that $f \in W - (\lambda',R',\sigma')$ for all $\lambda' < \lambda$, $R' < R$ and $\sigma' > \sigma$, i.e., $f \in \mathscr{W} - (\lambda,R,\sigma)$.

Q.E.D.

As an example, the function

$$\varphi(t) = \frac{1}{e^{i\lambda}/\sigma - t} + \frac{1}{e^{-i\lambda}/\sigma - t} \quad (0 < \lambda < \tfrac{1}{2}\pi), \tag{30}$$

is analytic in $H(\lambda,\sigma)$, in $C(\mu,\sigma/\sin(\lambda-\mu))$ for $\mu < \lambda$, and obeys the bounds (12') and (12) with $R = \infty$. It is the Borel transform of the function

$$f(z) = \frac{1}{z}\int_0^\infty dt e^{-t/z}\varphi(t) = \overset{\text{as.}}{\sum_{n=0}^{\infty}} 2\cos[n+1)\lambda\,]n!\sigma^{n+1}z^n \tag{31}$$

which can be analytically continued to the whole cut z plane $K(\frac{1}{2}\pi, \infty)$, but do *not* belong to $\mathscr{W} - (\frac{1}{2}\pi,\infty,\sigma)$. In fact, for $\frac{1}{2}\pi + \lambda < \arg z < \frac{1}{2}\pi$, $f(z)$ behaves essentially as $(1/z)\exp(-e^{i\lambda})/\sigma z)$ and diverges for $z \to 0$ in that sector, so that $f \in \mathscr{W} - (\lambda,\infty,\sigma)$ but $f \notin \mathscr{W} - (\nu,\infty,\sigma)$ for $\nu > \lambda$. Moreover, $f(z)$ does not belong to $\mathscr{N} - (\lambda,\infty,\sigma)$, although $f \in \mathscr{N} - [\mu,\infty,\sigma/\sin(\lambda-\mu)]$ for $\mu < \lambda$.

Theorem 2 makes clear the reason for having restricted λ to be strictly positive in the definition of the classes $\mathscr{W} - (\lambda,R,\sigma)$: in the limit $\lambda \to 0$ the piece of the analyticity domain $H(\lambda,\sigma)$ of $\varphi(t)$ that lies outside the disk $|t| < 1/\sigma$ shrinks to zero, and the analytic continuation of $\varphi(t)$ outside this disk (which is part of the practical Borel summation procedure) becomes meaningless. It is likely, however, that by a suitable redefinition of the classes $\mathscr{W} - (\lambda,R,\sigma)$, a remnant of Theorem 2 could hold even in the case $\lambda = 0$. In that case, the Borel transform $\varphi(t)$, still analytic in the disk $|t| < 1/\sigma$, would be required to admit a *quasianalytic* (and thus unique) continuation on the line $\mathrm{Im} t = 0$, $\mathrm{Re} t \geqslant 1/\sigma$. We shall not explore this interesting possibility here.

4. OTHER PROPERTIES

Theorem 3: The classes N or $\mathscr{N} - (\lambda,R,\sigma)$ and W or $\mathscr{W} - (\lambda,R,\sigma)$ are stable for the algebra structure of analytic functions in $K(\lambda,R)$.

Proof: If $f, g \in N(\lambda,R,\sigma)$, then clearly αf and $(f+g) \in N(\lambda,R,\sigma)$. Consider now the product $h(z) = f(z)g(z)$ and denote by R^h_M, R^f_M, and R^g_M the respective remainders. One has

$$h(z) = \sum_{n=0}^{M-1} h_n z^n + R^h_M(z), \quad M = 0,1,\ldots, \tag{32}$$

with

$$h_n = \sum_{m=0}^{n} f_m g_{n-m},$$

$$R^h_M(z) = \sum_{m=0}^{M-1} f_m z^m R^g_{M-m}(z) + R^f_M(z) g(z). \tag{33}$$

Hence, using Eqs. (3) and (5),

$$|R^h_M(z)| = A^f(\sigma) A^g(\lambda,R,\sigma)|\sigma z|^M \sum_{m=0}^{M-1} m!(M-m)! + A^f(\lambda,R,\sigma) A^g(\lambda,R,\sigma) M!|\sigma z|^M, \tag{34}$$

and noticing that $m!(M-m)! \leqslant (M-1)!$ for $1 \leqslant m \leqslant M-1$,

$$|R^h_M(z)| \leqslant 3A^f A^g(\lambda,R,\sigma) M!|\sigma z|^M, \tag{35}$$

which implies $h \in N - (\lambda,R,\sigma)$.

An analogous proof holds for the classes W or $\mathscr{W}$. Q.E.D.

We shall use Borel transformations together with Theorems 1 and 2 to establish the

Theorem 4: Let $f'(z)$ and $F(z)$ be respectively the derivative and a primitive of $f(z)$. Then,

(a) $f \in \mathscr{N}$ (resp. $\mathscr{W}$) $- (\lambda,R,\sigma)$ implies $f' \in \mathscr{N}$ (resp. $\mathscr{W}$) $- (\lambda,R,\sigma)$.

(b) $f \in \mathscr{N}$ (resp. $\mathscr{W},N,W$) $- (\lambda,R,\sigma)$ implies $F \in \mathscr{N}$ (resp. $\mathscr{W},N,W$) $- (\lambda,R,\sigma)$.

Proof:

(a) Let $f \in \mathscr{N}$ (resp. $\mathscr{W}$) $- (\lambda,R,\sigma)$. For any $\lambda' \leqslant \lambda, R' < R$, $\sigma' > \sigma$, choose λ'', R'', σ'' such that $\lambda' \leqslant \lambda'' < \lambda, R' < R'' < R$, $\sigma' > \sigma'' > \sigma$ (where the equality signs hold for the $\mathscr{N}$ case only). The Borel transform $\varphi(t)$ of the function $f(z)$ obeys the bound (12) [resp. (12')]. Then, using the Cauchy formula on a circle with center t and radius Σ^{-1},

$$\Sigma^{-1}(\lambda',\sigma') = \mathrm{Inf}|t_1 - t_2|, \tag{36}$$

$$t_1 \in C \text{ (resp. } H)(\lambda',\sigma'), \quad t_2 \in \partial C \text{ (resp. } \partial H)(\lambda'',\sigma''),$$

one derives the following bound for the derivatives of $\varphi(t)$:

$$|\varphi^{(p)}(t)| \leqslant \overline{B}(\lambda',R',\sigma') p! [\Sigma(\lambda',\sigma')]^p e^{|t|/R''}$$

$$\forall t \in C \text{ (resp. } H)(\lambda',\sigma'). \tag{37}$$

Now, due to the uniform convergence of the integrals, we are allowed to compute $f'(z)$ from $\varphi'(t)$ and $\varphi''(t)$ by first integrating Eq. (10) by parts:

$$f(z) = f_0 + \int_0^\infty dt e^{-t/z}\varphi'(t) = f_0 + z\int_0^\infty dx e^{-x}\varphi'(zx), \tag{38}$$

and next differentiating,

$$f'(z) = \int_0^\infty dx e^{-x}[\varphi'(zx) + zx\varphi''(zx)] = \frac{1}{z}\int_0^\infty dt e^{-t/z}[\varphi'(t) + t\varphi''(t)]. \tag{39}$$

This equation shows us that $f'(z)$ is the inverse Borel transform of the function

$$\phi(t) = \varphi'(t) + t\varphi''(t), \tag{40}$$

which according to Eq. (37), is bounded in C (resp. H) (λ',σ') by

$$|\phi(t)| \leqslant \overline{B}(\lambda',R',\sigma')\Sigma(\lambda',\sigma')[1 + 2|t|\Sigma(\lambda',\sigma')]e^{|t|/R''} \leqslant \overline{\overline{B}}(\lambda',R',\sigma')e^{|t|/R'}. \tag{41}$$

Then Theorem 1 (resp.2) immediately tells us that

$f' \in \mathcal{N}$ (resp. $\mathcal{W}$) $-(\lambda,R,\sigma)$.

(b) Let $f \in N$ or $W-(\lambda,R,\sigma)$. Then from Eq. (1),

$$F(z)=\sum_{n=0}^{M-1} F_n z^n + \mathcal{R}_M(z), \tag{42}$$

where $F_n = f_{n-1}/n$ $(n=1,2,\cdots)$, and

$$\mathcal{R}_M(z)=\int_0^z dz' R_{M-1}(z'). \tag{43}$$

Integrating on the straight line $z'=\rho z$ $(0\leqslant\rho\leqslant 1)$ and using the bounds (3) or (3') in Eq. (43), we obtain analogous bounds for $\mathcal{R}_M(z)$ and conclude that $F \in N$ or $W-(\lambda,R,\sigma)$ Q.E.D.

The next theorem gives necessary and sufficient conditions for a function $f(z)$ to be in the class $\mathcal{N}-(\lambda,R,\sigma)$ and sufficient conditions for f to be in the class $\mathcal{W}-(\lambda,R,\sigma)$. It will be recognized as a stronger version of a theorem already (and more or less implicitly) used in proofs of Borel summability.[3]

Theorem 5: Let $f(z)$ be a function analytic in $K(\lambda,R)$.

(a) f belongs to $\mathcal{N}-(\lambda,R,\sigma)$ if and only if its derivatives $f^{(n)}$ admit in $K(\lambda,R')$, for any $R'<R$ and $\sigma'>\sigma$, the bounds

$$|f^{(n)}(z)|\leqslant D(\lambda,R',\sigma')(n!)^2\left[\frac{\sigma'}{\cos\alpha_\lambda(z)-|z|/R'}\right]^n, \quad n=0,1,\cdots, \tag{44}$$

where

$$\alpha_\lambda(z)=\begin{cases}0 & \text{if } |\arg z|\leqslant\lambda,\\ |\arg z|-\lambda & \text{if } \lambda\leqslant|\arg z|<\tfrac{1}{2}\pi+\lambda.\end{cases}$$

(b) If for any $\lambda'<\lambda$, $R'<R$, and $\sigma'>\sigma$ the derivatives admit in $K(\lambda',R')$ the bounds

$$|f^{(n)}(z)|\leqslant D(\lambda',R',\sigma')(n!)^2\left[\frac{\sigma'}{\cos\alpha_{\lambda'}(z)(\cos\alpha_{\lambda'}(z)-|z|/R')}\right]^n, \quad n=0,1,\cdots, \tag{45}$$

then $f \in \mathcal{W}-(\lambda,R,\sigma)$.

Proof:

(a) (i) Let $f \in \mathcal{N}-(\lambda,R,\sigma)$ and $z \in K(\lambda,R')$.

Assume first $|\arg z|\leqslant\lambda$. Then, on account of Theorem 1,

$$f^{(n)}(z)=\int_0^\infty dx e^{-x}x^n\varphi^{(n)}(zx), \tag{46}$$

where, according to Eq. (37),

$$|\varphi^{(n)}(zx)|\leqslant\bar{B}(\lambda,R'',\sigma')n!(\sigma')^n e^{|zx|/R''} \tag{47}$$

for some R'' $(R'<R''<R)$ and any $\sigma'>\sigma$ [clearly $\Sigma(\lambda',\sigma')$ can be taken equal to σ' in this case]. Thus,

$$|f^{(n)}(z)|\leqslant\bar{B}(\lambda,R'',\sigma')n!\int_0^\infty dx x^n e^{-x(1-|z|/R'')}$$

$$=\frac{\bar{B}(\lambda,R'',\sigma')}{1-|z|/R''}(n!)^2\left(\frac{\sigma'}{1-|z|/R''}\right)^n$$

$$\leqslant\frac{\bar{B}(\lambda,R'',\sigma')}{1-R'/R''}(n!)^2\left(\frac{\sigma'}{1-|z|/R'}\right)^n. \tag{48}$$

If now $\lambda\leqslant\pm\arg z<\frac{1}{2}\pi+\lambda$, we use Eq. (46) integrated by parts,

$$f^{(n)}(z)=f_0\delta_{n0}+\int_o^\infty dx\, e^{-x}[nx^{n-1}\varphi^{(n)}(zx)+zx^n\varphi^{(n+1)}(zx)], \tag{49}$$

and rotate the integration path by putting

$$x=\rho\exp[-i(\arg z\mp\lambda)], \rho\geqslant 0.$$

Then the bound (47) is still valid on the new path, and for some σ'' $(\sigma<\sigma''<\sigma')$,

$$|f^{(n)}(z)|\leqslant|f_0|\delta_{n0}+\bar{B}(\lambda,R'',\sigma'')n!(\sigma'')^n\int_0^\infty d\rho\,[n\rho^{n-1}+(n+1)\sigma''|z|\rho^n]e^{-\rho\cos\alpha_\lambda+\rho|z|/R''}$$

$$\leqslant|f_0|\delta_{n0}+\bar{B}(\lambda,R'',\sigma'')(n!)^2(n+1)\left[\left(\frac{\sigma''}{\cos\alpha_\lambda-|z|/R''}\right)^n+|z|\left(\frac{\sigma''}{\cos\alpha_\lambda-|z|/R''}\right)^{n+1}\right]$$

$$\leqslant|f_0|\delta_{n0}+\bar{B}(\lambda,R'',\sigma'')(n!)^2\left[(n+1)\left(\frac{\sigma''}{\sigma'}\right)^{n+1}\left(\frac{\sigma'}{\sigma''}+\frac{\sigma'|z|}{\cos\alpha_\lambda-|z|/R''}\right)\left[\frac{\sigma'}{\cos\alpha_\lambda-|z|/R''}\right]^n\right]. \tag{50}$$

Hence, since $|z|\leqslant R'\cos\alpha_\lambda(z)$,

$$|f^{(n)}(z)|\leqslant|f_0|\delta_{n0}+\bar{B}(\lambda,R'',\sigma'')\frac{e\sigma'}{\log(\sigma'/\sigma'')}\left(\frac{1}{\sigma''}+\frac{R'}{1-R'/R''}\right)(n!)^2\left[\frac{\sigma'}{\cos\alpha_\lambda(z)-|z|/R'}\right]^n. \tag{51}$$

The bounds (44) follow from Eqs. (48) and (51).

(ii) Assume now $f(z)$ to be analytic in $K(\lambda,R)$ with its derivatives bounded according to Eq. (44). Let $\{z_i\}_{i=1,2,\cdots}$ be any sequence converging to 0 in $K(\lambda,R)$, and such that $|\arg z_i|\leqslant\frac{1}{2}\pi+\lambda-\epsilon$. Then Eq. (44) implies for any n,

$$|f^{(n)}(z_i)-f^{(n)}(z_j)|=\left|\int_{z_i}^{z_j} dz f^{(n+1)}(z)\right|\leqslant k_n|z_i-z_j|,$$

$$i,j=1,2,\cdots. \tag{52}$$

This means that $\{f^{(n)}(z_i)\}_{i=1,2,\cdots}$ is a Cauchy sequence, which ensures the existence of $\lim_{z\to 0}f^{(n)}(z)$, and allows us to define

$$f_n\equiv\frac{1}{n!}\lim_{z\to 0}f^{(n)}(z) \quad (z\in K(\lambda,R),\ |\arg z|\leqslant\tfrac{1}{2}\pi+\lambda-\epsilon) \tag{53}$$

and

$$R_M(z)\equiv f(z)-\sum_{n=0}^{M-1}f_n z^n. \tag{54}$$

Making then $z_0\to 0$ in the Taylor formula,

$$f(z)-\sum_{n=0}^{M-1}\frac{1}{n!}f^{(n)}(z_0)(z-z_0)^n$$

$$=\frac{1}{(M-1)!}\int_{z_0}^z dz'(z-z')^{M-1}f^{(M)}(z') \quad [z,z_0\in K(\lambda,R)], \tag{55}$$

we obtain for $M \geqslant 1$,

$$R_M(z) = \frac{1}{(M-1)!}\int_{L_z} dz'(z-z')^{M-1} f^{(M)}(z'), \tag{56}$$

where L_z is any path between the origin and z that is contained in $K(\lambda,R)$. Let $z \in K(\lambda,R')$. If $|\arg z| \leqslant \lambda$, we take for L_z the radial segment between 0 and z. Then, we can choose some R'' ($R' < R'' < R$) and use Eq. (44) in Eq. (56) to get for any $\sigma' > \sigma$,

$$\begin{aligned}|R_M(z)| &\leqslant D(\lambda,R'',\sigma')M!M\sigma'^M|z|^M \\ &\times \int_0^1 d\rho(1-\rho)^{M-1}(1-\rho|z|/R'')^{-M} \\ &= D(\lambda,R'',\sigma')M!|\sigma' z|^M M\int_0^1 du\frac{u^{M-1}}{(1-u|z|/R'')} \\ &\leqslant \frac{D(\lambda,R'',\sigma')}{(1-R'/R'')}M!|\sigma' z|^M. \end{aligned} \tag{57}$$

If $\lambda < \pm \arg z < \frac{1}{2}\pi + \lambda$, the previous choice of the path L_z would lead to a crude (and in fact useless) overestimate of $|R_M(z)|$. A better choice is obtained by considering the circle C through z and tangent to the line $\arg z' = \pm\lambda$ at the origin (notice that this corresponds to a straight line in the plane of the variable $1/z$, more natural from the Laplace transform point of view). We take for L_z that arc of C which joins 0 and z inside $K(\lambda,R')$. In the case $\lambda < \arg z < \frac{1}{2}\pi + \lambda$, this means

$$z' \in L_z \text{ iff } \mathrm{Im}(e^{i\lambda}/z') = \mathrm{Im}(e^{i\lambda}/z),\ \mathrm{Re}(e^{i\lambda}/z') \geqslant \mathrm{Re}(e^{i\lambda}/z). \tag{58}$$

Recasting now Eq. (56) in the form

$$\begin{aligned}R_M(z) = &-\frac{e^{i\lambda}z^{M-1}}{(M-1)!} \\ &\times \int_{z' \in L_z} d\left(\frac{e^{i\lambda}}{z'}\right)\left(\frac{e^{i\lambda}}{z'}\right)^{-(M+1)}\left(\frac{e^{i\lambda}}{z'}-\frac{e^{i\lambda}}{z}\right)^{M-1} f^{(M)}(z'),\end{aligned} \tag{59}$$

and using the following parametrization of L_z,

$$e^{i\lambda}/z' = i\,\mathrm{Im}(e^{i\lambda}/z) + (1/y)\mathrm{Re}(e^{i\lambda}/z), \quad 0 < y \leqslant 1, \tag{60}$$

we deduce from the bound (44),

$$\begin{aligned}|R_M(z)| &\leqslant D(\lambda,R'',\sigma'')M!M\sigma''^M|z|^M[\mathrm{Re}(e^{i\lambda}/z)]^M \\ &\times \int_0^1 dy\frac{|z'|}{y2}\left(\frac{1}{y}-1\right)^{M-1}\left[\frac{|z'|}{\cos\alpha_\lambda(z')-|z'|/R''}\right]^M.\end{aligned} \tag{61}$$

Namely, since $\mathrm{Re}(e^{i\lambda}/z) = \cos\alpha_\lambda(z)/|z|$ and $\cos\alpha_\lambda(z') = |z'|\cos\alpha_\lambda(z)/|z|y$,

$$\begin{aligned}|R_M(z)| &\leqslant D(\lambda,R'',\sigma'')M!M\sigma''^M|z|^{M-1} \\ &\times \int_0^1 dy\frac{|z'|}{y}(1-y)^{M-1}\left[1-y\frac{|z|}{R''\cos\alpha_\lambda(z)}\right]^{-M},\end{aligned} \tag{62}$$

and noticing that $|z'|/y = |z|\cos\alpha_\lambda(z')/\cos\alpha_\lambda(z) \leqslant |z|/\cos\alpha_\lambda(z)$,

$$\begin{aligned}|R_M(z)| &\leqslant D(\lambda,R'',\sigma'')M!(|\sigma''z|^M/\cos\alpha_\lambda(z)) \\ &\times M\int_0^1 du\frac{u^{M-1}}{(1-u|z|/R''\cos\alpha_\lambda(z))} \\ &\leqslant \frac{D(\lambda,R'',\sigma'')}{(1-R'/R'')}M!\frac{|\sigma''z|^M}{\cos\alpha_\lambda(z)}.\end{aligned} \tag{63}$$

To remove the unwanted factor $1/\cos\alpha_\lambda(z)$, it suffices to reinsert this bound in the rhs of

$$R_M(z) = f_M z^M + R_{M+1}(z), \tag{64}$$

together with $|f_n| \leqslant D(\lambda,R',\sigma')M!\sigma'^M$, which follows from the Definition (53). Then,

$$\begin{aligned}|R_M(z)| &\leqslant \{D(\lambda,R',\sigma') \\ &+ \frac{D(\lambda,R'',\sigma'')}{1-R'/R''}\sigma'\left[(M+1)\left(\frac{\sigma''}{\sigma'}\right)^{M+1}\right]\frac{|z|}{\cos\alpha_\lambda}\Big\}M!|\sigma' z|^M \\ &\leqslant \left[D(\lambda,R',\sigma') + \frac{D(\lambda,R'',\sigma'')}{1-R'/R''}\frac{\sigma' e}{\log(\sigma'/\sigma'')}R'\right]M!|\sigma' z|^M.\end{aligned} \tag{65}$$

Equations (57) and (65) reproduce the bounds (3), and we conclude that $f \in \mathscr{N} - (\lambda,R,\sigma)$.

(b) Minor and obvious changes in the part (ii) of the proof above suffice to establish that conditions (45) imply the membership of f in $\mathscr{W} - (\lambda,R,\sigma)$.

Q.E.D.

We have stated Eqs. (45) only as sufficient conditions for $f \in \mathscr{W} - (\lambda,R,\sigma)$, and actually they are *not* necessary conditions. This can be seen as follows. Starting from the function $\varphi(t)$ defined in Eq. (30) [which is the Borel transform of a function $f(z) \in \mathscr{W} - (\lambda,\infty,\sigma)$] and using the representation (46), it is not difficult to derive the large n behavior of $f^{(n)}(z)$ by applying the saddle-point method. The result shows that

$$\lim_{z\to 0}|f^{(n)}(z)| \underset{n\to\infty}{\simeq} 2(n!)^2\sigma^{n+1}, \tag{66}$$

in agreement with Eq. (31), *but*

$$\sup_{\substack{z\in K(\lambda,\infty) \\ \arg z = \varphi}}|f^{(n)}(z)| \underset{n\to\infty}{\simeq} 2(n!)^2\left[\frac{2\sigma}{1+\sin(\lambda-|\varphi|)}\right]^{n+1}, \tag{67}$$

which violates inequalities (45) indeed. The supremum in Eq. (67) is attained for $|z| = \cos(\lambda - |\varphi|)/2\sigma n$. Moreover, Eqs. (66) and (67) tell us that there are no "simple" necessary and sufficient conditions of the form (45) for $f \in \mathscr{W} - (\lambda,R,\sigma)$ (i.e., conditions expressed solely by means of bounds uniform in $|z|$ when $R = \infty$). In fact, using a Cauchy representation for $\varphi(t)$ and Eq. (67), it is possible to determine completely the analyticity domain of $\varphi(t)$ implied by conditions (45), which is of course strictly larger than $H(\lambda,\sigma)$ but strictly smaller that $C(\lambda,\sigma)$. Its shape is quite complicated, but the sets $\{|z| = 1/\sigma, \lambda + \frac{1}{2}\pi \leqslant |\arg t| \leqslant \pi\}$ and $\{|\arg t| = \lambda, |t| \geqslant 2/\sigma\}$ are still parts of its boundary.

The following theorem shows that, under rather general conditions, the composition of functions preserves the Borel summability.

Theorem 6: Let $f(z)$ be a function in $\mathscr{N}(\text{resp. } \mathscr{W}) - (\lambda_1,R_1,\sigma_1)$ with $f_0 = 0$ and $g(z)$ a function in $\mathscr{N}(\text{resp. } \mathscr{W}) - (\lambda_2,R_2,\sigma_2)$. If there are $\lambda \leqslant \lambda_1$ (resp. $\lambda < \lambda_1$) and $R < R_1$ such that

$$f(K(\lambda,R)) \subset K(\lambda_2',R_2'), \tag{68}$$

with $\lambda_2' = \lambda_2$ (resp. $\lambda_2' < \lambda_2$) and $R_2' < R_2$, then there exists a σ such that $h \equiv g \circ f$ belongs to $\mathscr{N}(\text{resp. } \mathscr{W}) - (\lambda,R,\sigma)$.

Proof: Consider the $\mathcal{N}$-case. First of all, $K(\lambda,R)\subset K(\lambda_1,R_1)$, so that the function $h(z)$ is defined and analytic in $K(\lambda,R)$ on account of (68). We can write for all $z\in K(\lambda,R)$,

$$h(z)=g_0+\sum_{p=1}^{M-1}g_p[f(z)]^p+R_M^g(f(z)). \tag{69}$$

Now, according to Theorem 3, $f^p\in N(\lambda,R,\sigma_1')$ for any p and $\sigma_1'>\sigma_1$. Hence,

$$[f(z)]^p=\sum_{n=p}^{M-1}f_n^p z^n+R_M^p(z), \tag{70}$$

where

$$f_n^p=\sum_{m_1+\cdots+m_p=n}f_{m_1}f_{m_2}\cdots f_{m_p} \tag{71}$$

and where the M th order remainder $R_M^p(z)$ is bounded by

$$|R_M^p(z)|\leqslant\tfrac{1}{2}(3^p-1)(M-p+1)!\times[A^f(\lambda,R,\sigma_1')]^p|\sigma_1'z|^M,\quad 1\leqslant p\leqslant M-1, \tag{72}$$

as shown in Appendix A. Then,

$$h(z)=\sum_{n=0}^{M-1}h_n z^n+R_M^h(z), \tag{73}$$

with

$$\begin{cases}h_0=g_0,\\ h_n=\sum_{p=1}^{n}g_p f_n^p,\quad n\geqslant 1,\end{cases} \tag{74}$$

and

$$R_M^h(z)=\sum_{p=1}^{M-1}g_p R_M^p(z)+R_M^g(f(z)). \tag{75}$$

To bound $R_M^h(z)$, we use Eq. (72) together with

$$|g_p|\leqslant A^g(\lambda_2,R_2',\sigma_2')p!(\sigma_2')^p \tag{76}$$

and

$$|R_M^g(f(z))|\leqslant A^g(\lambda_2,R_2',\sigma_2')M!(\sigma_2')^M|f(z)|^M, \tag{77}$$

which are valid for any $\sigma_2'>\sigma_2$ since the point $f(z)$ is contained in $K(\lambda_2,R_2')$. Thus,

$$|R_M^h(z)|\leqslant A^g|\sigma_1'z|^M\sum_{p=1}^{M-1}p!(M-p+1)!\tfrac{1}{2}(3A^f\sigma_2')^p+A^gM!|A^f\sigma_1'\sigma_2'z|^M, \tag{78}$$

where we have used $|f(z)|=|R_1^f(z)|\leqslant A^f|\sigma_1'z|$. Finally, introducing $\mu=\max(1,3A^f\sigma_2')$ and noticing that $\sum_{p=1}^{M-1}p!(M-p+1)!\leqslant 3M!$ [see Eq. (A5)], we obtain

$$|R_M^h(z)|\leqslant A^gM!|\sigma_1'z|^M\left[\tfrac{3}{2}\mu^{M-1}+(A^f\sigma_2')^M\right]\leqslant\tfrac{11}{6}A^g(\lambda_2,R_2',\sigma_2')M!|\sigma'z|^M, \tag{79}$$

where

$$\sigma'=\max[\sigma_1',3A^f(\lambda,R,\sigma_1')\sigma_1'\sigma_2']. \tag{80}$$

It follows from Eqs. (79) and (80) that $h\in\mathcal{N}-(\lambda,R,\sigma)$, with

$$\sigma=\inf_{\sigma_1'>\sigma_1}\max[\sigma_1',3A^f(\lambda,R,\sigma_1')\sigma_1'\sigma_2]. \tag{81}$$

A similar proof holds in the $\mathcal{W}$ case, with

$$\sigma=\inf_{\sigma_1'>\sigma_1}\max\left[\sigma_1',3A^f(\lambda,R,\sigma_1')\sigma_1'\sigma_2,\frac{A^f(\lambda,R,\sigma_1')\sigma_1'\sigma_2}{\sin(\lambda_2-\lambda_2')}\right] \tag{82}$$

Q.E.D.

Remark: In particular, Theorem 6 always applies when the function $g(z)$ is analytic in a disk $|z|<\rho$: it suffices to choose $\lambda\leqslant\lambda_1$ and $R(<R_1)$ so small that $A^f(\lambda,R,\sigma_1')\sigma_1'R<\rho$ for some $\sigma_1'>\sigma_1$. Moreover, $\sigma\leqslant\sigma_1$ in that case, since σ_2 can be taken arbitrarily small.

In our last theorem, we extend the range of validity of a formula first obtained by Figerou[4] (but previously known and used in its "low order" forms). The virtue of this formula is to give, under suitable conditions, a simple expression for the expansion coefficients h_n of the function $h(z)$ in Theorem 6, approximate up to a given order r in $1/n$. An application can be found in Ref. 4.

Theorem 7: Let f and g be as in Theorem 6. If moreover, the sequence $\{f_n\}$ of the expansion coefficients of $f(z)$ is absolutely bounded by a positive sequence $\{\bar f_n\}$ such that:

(i) $\{\log\bar f_n\}$ is a convex sequence,

(ii) $\bar f_n\geqslant n\bar\sigma\bar f_{n-1}$ for some $\bar\sigma>0$ $(n=2,3,\cdots)$,

(iii) for some integer $r\geqslant 0$, the expansion coefficients of $g(z)$ satisfy

$$|g_n|\leqslant(G/k^n)\bar f_{n-r-1}\text{ for some } G>0\ (n=r+2,r+3,\cdots), \tag{83}$$

where

$$k=2\bar f_1+\bar f_2/\bar\sigma\quad(\neq 0), \tag{84}$$

then the expansion coefficients h_n of $h=g\circ f$ are given in the large n limit by

$$h_n=\sum_{l=0}^{r}e_l f_{n-l}+\bar f_{n-r}O\left(\frac{1}{n}\right), \tag{85}$$

where the e_l's are defined by (in the sense of formal power series)

$$\sum_{l=0}^{\infty}e_l x^l=\frac{d}{du}g(u)\Big|_{u=\sum_n^{\infty}f_n x^n} \tag{86}$$

Proof: According to Eqs. (71) and (74), the coefficients h_n are given in terms of the coefficients f_n and g_n by the expression

$$h_n=\sum_{p=1}^{n}g_p\sum_{\substack{m_1+\cdots+m_p=n\\ 1\leqslant m_i\leqslant n-p+1}}f_{m_1}f_{m_2}\cdots f_{m_p}. \tag{87}$$

We split up this sum into three parts:

$$h_n=h_n^{(1)}+h_n^{(2)}+h_n^{(3)}, \tag{88}$$

$$h_n^{(1)}=g_1f_n+\sum_{p=2}^{r+1}g_p\sum_{\substack{m_1+\cdots m_p=n\\ 1\leqslant m_i\leqslant n-p+1\\ \text{at least one } m_i\geqslant n-r}}f_{m_1}\cdots f_{m_p}, \tag{89}$$

$$h_n^{(2)}=\sum_{p=2}^{r+1}g_p\sum_{\substack{m_1+\cdots+m_p=n\\ 1\leqslant m_i\leqslant n-r-1}}f_{m_1}\cdots f_{m_p}, \tag{90}$$

$$h_n^{(3)}=\sum_{p=r+2}^{n}g_p\sum_{\substack{m_1+\cdots+m_p=n\\ 1\leqslant m_i\leqslant n-p+1}}f_{m_1}\cdots f_{m_p}. \tag{91}$$

Introducing the sums

$$S(n,p,q)=\sum_{\substack{m_1+\cdots+m_p=n\\ 1\leqslant m_i\leqslant n-q}}\bar{f}_{m_1}\bar{f}_{m_2}\cdots\bar{f}_{m_p}, \tag{92}$$

and using for them the bounds (B2) established in Appendix B, we are able to show that $h^{(2)}$ is of order $\bar{f}_{n-r-1}$. Indeed,

$$|h_n^{(2)}|\leqslant\sum_{p=2}^{r+1}|g_p|S(n,p,r+1)\leqslant\left[\sum_{p=2}^{r+1}|g_p|k(p,r+1)\right]\bar{f}_{n-r-1}. \tag{93}$$

Likewise, from Eqs. (B2), (B3), and (83),

$$\begin{aligned}|h_n^{(3)}|&\leqslant\sum_{p=r+2}^{n}|g_p|S(n,p,p-1)\leqslant\frac{G}{k}\sum_{p=r+2}^{n}\bar{f}_{p-r-1}\bar{f}_{n-p+1}\\&=(G/k)S(n-r,2,1)\leqslant G\bar{f}_{n-r-1}.\end{aligned} \tag{94}$$

Hence, using again property (ii), we get in the large n limit,

$$\begin{aligned}&|h_n-h_n^{(1)}|\\&\leqslant H(r)\bar{f}_{n-r-1}\leqslant(H(r)/(n-r)\bar{\sigma})\bar{f}_{n-r}=\bar{f}_{n-r}O(1/n)\end{aligned} \tag{95}$$

Considering now $h_n^{(1)}$, we see that for $n>2r$, only one m_i can be $\geqslant n-r$ in Eq. (89), so that

$$\begin{aligned}h_n^{(1)}&=g_1f_n+\sum_{p=2}^{r+1}g_pp\sum_{l=p-1}^{r}f_{n-l}\sum_{\substack{m_1+\cdots+m_{p-1}=l\\ 1\leqslant m_i\leqslant l-(p-2)}}f_{m_1}\cdots f_{m_{p-1}}\\&=g_1f_n+\sum_{l=1}^{r}f_{n-l}\sum_{p=2}^{l+1}pg_p\sum_{\substack{m_1+\cdots+m_{p-1}=l\\ 1\leqslant m_i\leqslant l-p+2}}f_{m_1}\cdots f_{m_{p-1}}.\end{aligned} \tag{96}$$

Finally, we obtain by putting together Eqs. (95) and (96),

$$\begin{aligned}h_n&=g_1f_n+\sum_{l=1}^{r}f_{n-l}\left[\sum_{q=1}^{l}(q+1)g_{q+1}\sum_{\substack{m_1+\cdots+m_q=l\\ 1\leqslant m_i\leqslant l-q+1}}f_{m_1}\cdots f_{m_q}\right]\\&\quad+\bar{f}_{n-r}O\left(\frac{1}{n}\right).\end{aligned} \tag{97}$$

On the other hand,

$$\begin{aligned}\left|\frac{d}{du}g(u)\right|_{u=\sum_{n=1}^{\infty}f_nx^n}&=\sum_{q=0}^{\infty}(q+1)g_{q+1}\left[\sum_{n=1}^{\infty}f_nx^n\right]^q\\&=g_1+\sum_{l=1}^{\infty}x^l\left[\sum_{q=1}^{\infty}(q+1)g_{q+1}\sum_{\substack{m_1+\cdots+m_q=l\\ 1\leqslant m_i\leqslant l-q+1}}f_{m_1}\cdots f_{m_q}\right].\end{aligned} \tag{98}$$

Therefore,

$$\begin{aligned}&e_0=g_1\\&e_l=\sum_{q=1}^{l}(q+1)g_{q+1}\sum_{\substack{m_1+\cdots+m_q=l\\ 1\leqslant m_i\leqslant l-q+1}}f_{m_1}\cdots f_{m_q},\quad l=1,2,\cdots,\end{aligned} \tag{99}$$

and Eq. (97) reduces to the announced formula (85).

Q.E.D.

Remarks: (1) In actual applications of Theorem 7, some work may be required to determine an appropriate majorant sequence $\{\bar{f}_n\}$ (if it exists at all!). First, in order that Eq. (84) be a sensible formula, it is necessary that the remainder $\bar{f}_{n-r}O(1/n)$ become really negligible with respect to f_{n-r} when $n\to\infty$ (at least for a subsequence of the integers n). This imposes the asymptotic behaviors of $\{f_n\}$ and $\{\bar{f}_n\}$ to be "essentially" the same, and fixes an upper bound to $\bar{\sigma}$. Then the sequence $\{\bar{f}_n\}$ can be extrapolated down to $\bar{f}_1$. This has to be done with the aim of minimizing $k=2\bar{f}_1+\bar{f}_2/\bar{\sigma}$, but under the constraints $\bar{f}_n>|f_n|$ and the conditions (i) and (ii). A rescaling of the $\bar{f}_n$'s [which does not spoil (i) and (ii)] may be needed in this step. Finally, one is in a position to check condition (iii).

(2) We wish to emphasize the usefulness and nontriviality of formula (85). Assuming for definiteness that $\lim_{n\to\infty}\inf|f_n|/\bar{f}_n>0$, one obtains for $r=0$, the remarkably simple result

$$h_n=g_1f_n[1+O(1/n)], \tag{100}$$

and more generally,

$$h_n=\sum_{l=0}^{r}e_lf_{n-l}[1+O(1/n^{r+1})]. \tag{101}$$

It is interesting to compare this formula with the following one:

$$h_n=\sum_{l=0}^{n-1}e_lf_{n-l}(1-l/n), \tag{102}$$

which is *exact* (and easy to derive) but useless when $n\to\infty$. One sees that, at the order $(1/n)^r$, acute cancellations occur in Eq. (102) between the corrections l/n for $l\leqslant r$ and all the terms for which $r+1\leqslant l\leqslant n-1$.

(3) Condition (iii) is certainly satisfied if the function $g(z)$ is analytic in a disk $|z|\leqslant\rho$. In that case indeed, $|g_n|\leqslant\gamma/\rho^n$, and condition (ii) implies

$$\begin{aligned}&\bar{f}_{n-r-1}\geqslant\bar{\sigma}^{n-r-2}(n-r-1)!f_1\\&\geqslant(k/\rho)^n\gamma/G\geqslant(k^n/G)|g_n|,\quad\forall n\geqslant r+2,\end{aligned} \tag{103}$$

for G large enough.

To conclude, let us give, as an example, a family of sequences $\{f_n\}$ for which suitable majorant sequences $\{\bar{f}_n\}$ do exist:

$$\begin{aligned}f_n&=c\Gamma(n+b)a^ne^{n^\alpha}\cos(n\varphi)\\&\quad\times\left[1+\frac{d_1}{n}+\frac{d_2}{n^2}+\cdots+\frac{d_r}{n^r}+O\left(\frac{1}{n^{r+1}}\right)\right],\end{aligned} \tag{104}$$

with $a>0$, $0<\varphi\leqslant\pi$ and $0<\alpha<1$ [$\alpha<1$ in order that $f(z)$ belong to some $\mathscr{W}-(.,.,a)$]. Since $|f_n|\leqslant C\Gamma(n+b)a^ne^{n^\alpha}$, we can try the choice

$$\bar{f}_n=C\Gamma(n+b)a^ne^{n^\alpha}\quad\text{for } n>N. \tag{105}$$

Then it is readily seen that conditions (i) and (ii) are satisfied indeed (with $\bar{\sigma}=a$), when n is larger than some N, depending only on b and α. Of course, the optimal extrapolation of $\{\bar{f}_n\}$ below N and the determination of k will depend on the exact values of f_n for $n\leqslant N$. If now the condition (iii) turns out to be satisfied, Eqs. (85) and (99) give the asymptotic expression

$$\begin{aligned}h_n&=g_1c\Gamma(n+b)a^ne^{n^\alpha}\\&\times\left[\cos(n\varphi)+\frac{\delta_1(n)}{n}+\frac{\delta_2(n)}{n^2}+\cdots+\frac{\delta_r(n)}{n^r}+O\left(\frac{1}{n^{r+1}}\right)\right],\end{aligned} \tag{106}$$

and allow us to compute explicitly $\delta_1(n),\ldots,\delta_r(n)$ from $d_1,\ldots,d_r,f_1,\ldots,f_r$, and $g_1,\ldots,g_{r+1}$.

ACKNOWLEDGMENTS

We are indebted to G. Mahoux for his help and interest in an early stage of this work and to V. Figerou for fruitful discussions.

APPENDIX A

Let $f \in N - (\lambda,R,\sigma)$ with $f_0 = 0$. Denoting by $R^{\,p}_{M}(z)$ the M th-order remainder of $[f(z)]^p$, we show that for $z \in K(\lambda,R)$,

$$|R^{\,p}_{M}(z)| \leqslant \tfrac{1}{2}(3^p - 1)\,(M - p + 1)![A(\lambda,R,\sigma)]^p|\sigma z|^M, \quad 1 \leqslant p \leqslant M - 1. \tag{A1}$$

The proof is based on the formula

$$R^{\,p}_{M}(z) = \sum_{m=1}^{M-p} f_m z^m R^{\,p-1}_{M-m}(z) + R^{\,1}_{M-p+1}(z)[R^{\,1}_{1}(z)]^{p-1}, \tag{A2}$$

which follows directly from the decomposition.

$$\begin{aligned}[f(z)]^p &= f(z)[f(z)]^{p-1} \\ &= \sum_{m=1}^{M-p} f_m z^m \left[\sum_{q=p-1}^{M-m-1} f_q^{p-1} z^q + R^{\,p-1}_{M-m}(z)\right] \\ &\quad + R^{\,1}_{M-p+1}(z)[f(z)]^{p-1},\end{aligned} \tag{A3}$$

since $f(z) = R^{\,1}_{1}(z)$.

We proceed by induction. The bound (A1) is true by definition for $p = 1$. Assuming it to be true for $R^{\,p-1}_{M}(z)$, we obtain from Eq. (A2),

$$\begin{aligned}&|R^{\,p}_{M}(z)| \\ &\leqslant \sum_{m=1}^{M-p} Am!|\sigma z|^m \frac{3^{p-1} - 1}{2} A^{p-1}(M - m - p + 2)!|\sigma z|^{M-m} \\ &\quad + A(M - p + 1)!|\sigma z|^{M-p+1} A^{p-1}|\sigma z|^{p-1} = A^p|\sigma z|^M \\ &\quad \times \left[\frac{3^{p-1} - 1}{2} \sum_{m=1}^{M-p} m!(M - m - p + 2)! + (M - p + 1)!\right].\end{aligned} \tag{A4}$$

Then, noticing that $m!(P - m)! \leqslant 2(P - 2)!$ for $2 \leqslant m \leqslant P - 2$, we have

$$\sum_{m=1}^{M-p} m!(M - m - p + 2)! \leqslant 3(M - p + 1)!, \tag{A5}$$

which inserted into Eq. (A4) completes the proof of (A1).

APPENDIX B

Let $\{\bar{f}_n\}_{n=1,2,\ldots}$ be a positive sequence enjoying properties (i) and (ii) of Theorem 7. We show that for all integers n, p, q such that $2 \leqslant p \leqslant n$ and $p - 1 \leqslant q \leqslant n(p-1)/p$, the sums

$$S(n,p,q) \equiv \sum_{\substack{m_1 + \cdots + m_p = n \\ 1 \leqslant m_i \leqslant n - q}} \bar{f}_{m_1} \cdots \bar{f}_{m_p} \tag{B1}$$

admit the bounds

$$S(n,p,q) \leqslant k(p,q)\bar{f}_{n-q}, \tag{B2}$$

where

$$k(p,p-1) = k^{p-1}, \quad k = 2\bar{f}_1 + \bar{f}_2/\bar{\sigma}. \tag{B3}$$

The proof is by induction over p. For $p = 2$, one has

$$S(n,2,q) = \sum_{m=q}^{n-q} \bar{f}_m \bar{f}_{n-m} = 2\bar{f}_q \bar{f}_{n-q} + \sum_{m=q+1}^{n-q-1} \bar{f}_m \bar{f}_{n-m}. \tag{B4}$$

The convexity property (i) implies

$$\bar{f}_m \bar{f}_{n-m} \leqslant \bar{f}_{q+1}\bar{f}_{n-q+1} \quad \text{for } q + 1 \leqslant m \leqslant n - q - 1, \tag{B5}$$

so that

$$\begin{aligned}&\sum_{m=q+1}^{n-q-1} \bar{f}_m \bar{f}_{n-m} \\ &\leqslant (n - 2q - 1)\bar{f}_{q+1}\bar{f}_{n-q-1} \quad \left(\text{for } q \leqslant \frac{n}{2} - 1\right) \\ &\leqslant (n - q)\bar{f}_{q+1}\bar{f}_{n-q-1} \\ &\leqslant \frac{\bar{f}_{q+1}}{\bar{\sigma}} \bar{f}_{n-q}\end{aligned} \tag{B6}$$

on account of property (ii).

Thus,

$$S(n,2,q) \leqslant (2\bar{f}_q + \bar{f}_{q+1}/\bar{\sigma})\bar{f}_{n-q}, \tag{B7}$$

which is of the form (B2)–(B3) indeed, since $k(2,1) = 2\bar{f}_1 + \bar{f}_2/\bar{\sigma} = k$. Let us now assume Eqs. (B2) and (B3) to hold for some $p - 1 \geqslant 2$, and all q $[p - 2 \leqslant q \leqslant n(p-2)/(p-1)]$. In the expression (B1) for $S(n,p,q)$, we can single out the sum over m_p. This allows us to express $S(n,p,q)$ in terms of $S(\cdots,p-1,\cdots)$:

$$\begin{aligned}S(n,p,q) &= \sum_{s=1}^{q-p+1} \bar{f}_s S(n - s, p - 1, q - s) \\ &\quad + \sum_{s=q-p+2}^{n-q} \bar{f}_s S(n - s, p - 1, p - 2)\end{aligned} \tag{B8}$$

(the first term in the rhs is absent if $q = p - 1$).

Hence,

$$\begin{aligned}S(n,p,q) &\leqslant \sum_{s=1}^{q-p+1} \bar{f}_s k(p - 1, q - s)\bar{f}_{n-q} \\ &\quad + \sum_{s=q-p+2}^{n-1} \bar{f}_s k(p - 1, p - 2)\bar{f}_{n-s-p+2} \\ &= \left[\sum_{s=1}^{q-p+1} \bar{f}_s k(p - 1, q - s)\right]\bar{f}_{n-q} \\ &\quad + k^{p-2} S(n - p + 2, 2, q - p + 2) \\ &\leqslant \left[\sum_{s=1}^{q-p+1} \bar{f}_s k(p - 1, q - s) + k^{p-2}(2, q - p + 2)\right]\bar{f}_{n-q} \\ &= k(p,q)\bar{f}_{n-q},\end{aligned} \tag{B9}$$

with

$$k(p,p-1) = k^{p-2}k(2,1) = k^{p-1},$$

and the proof is complete.

Notice that the constants $k(p,q)$ still depend on *some* of the $\bar{f}_m$'s. The important point is that the number of those $\bar{f}_m$'s is bounded independently of n. In fact, it is easily drawn from Eqs. (B3) and (B9), that $k(p,q)$ depends only on $\bar{f}_1, \ldots, \bar{f}_{q-p+3}$.

[1] C. M. Bender and T. T. Wu, Phys. Rev. Lett. **27**, 461 (1971); Phys. Rev. D **7**, 1620 (1973); T. Banks, C. M. Bender, and T. T. Wu, Phys. Rev. D **8**, 3346 (1973); D **8**, 3366 (1973); L. N. Lipatov, Pis'ma Zh. Eksp. Teor. Fiz. **24**, 179 (1976) [JETP Lett. **24**, 157 (1976)]; **25**, 116 (1977) [**25**, 104 (1977)]; Zh. Eksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)]; E. Brezin, J. C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **15**, 1544 (1977); D **15**, 1558 (1977); G. Parisi, Phys. Lett. B **66**, 167 (1977); J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977); See also J. Zinn-Justin and G. Parisi

in *Cargese Lectures* 1977 (Plenum, New York, 1979).
[2]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970); B. Simon, Adv. in Math. **7**, 240 (1971); J. Avron, I. Herbst, and B. Simon, Phys. Lett. A **62**, 214 (1977); P. Renouard; Ann. Inst. Henri Poincaré, A **27**, 237 (1977); **31**, 235 (1979); S. Graffi and V. Grecchi, Commun. Math. Phys. **62**, 83 (1978); I. Herbst and B. Simon, Phys. Rev. Lett. **41**, 67 (1978).
[3]J. P. Eckmann, J. Magnen, and R. Seneor, Commun. Math. Phys. **39**, 251 (1975); J. Magnen and R. Seneor, Commun. Math. Phys. **56**, 237 (1977); J. P. Eckmann and H. Epstein, Commun. Math. Phys. **68**, 245 (1979).
[4]R. Figerou, "3^{d} Cycle Thesis," Montpellier, 1980, Montpellier Preprint PM/80/6.
[5]E. Borel, *Lecons sur les séries divergentes* (Gauthier-Villars, Paris, 1928).
[6]G. H. Hardy, *Divergent Series* (Oxford U.P., London, 1949).
[7]G. N. Watson, Philos. Trans. R. Soc. London, Ser. A **211**, 279 (1912).
[8]F. Nevanlinna, Ann. Acad. Sci. Fenn. Ser. A **12**, No.3 (1918–19); Jahresber. Fortschr. Math. **46**, 1463 (1916–18).
[9]A. D. Sokal, J. Math. Phys. **21**, 261 (1980).

The Coulomb unitarity relation and some series of products of three Legendre functions

H. van Haeringen
Department of Mathematics, University of Technology, Delft, The Netherlands
L. P. Kok
Institute for Theoretical Physics, P. O. Box 800, University of Groningen, The Netherlands

(Received 29 April 1981; accepted for publication 12 June 1981)

We obtain from the off-shell Coulomb unitarity relation a closed expression for $\Sigma_{l=0}^{\infty}(2l+1)P_l(x)\times Q_l^{i\gamma}(y)\,Q_l^{-i\gamma}(z)$, and we consider some related series of products of Legendre functions.

PACS numbers: 02.30.Lt, 02.30.Gp, 03.65.Nk

In this paper we shall consider the Coulomb unitarity relation[1–4] and derive from this relation a closed expression for an infinite series of products of three Legendre functions, P_l, $Q_l^{i\gamma}$, and $Q_l^{-i\gamma}$ [see Eq. (12)]. By taking the limit $\gamma\to 0$ we obtain agreement with an expression[5] for the corresponding series, which exists in the literature. However, our expression has a much simpler form, which means that we have obtained a substantial reduction of the expression given in.[5] After the derivation of our main result, Eq. (12), we shall briefly consider some related series of products of Legendre functions [see Eqs. (14)–(25)].

The unitarity relation, or generalized optical theorem, or Low equation, in quantum-mechanical scattering theory establishes a simple relation between the imaginary part of the off-shell T matrix and its half-off-shell elements.[1,2] Suppressing the energy, $E=k^2+i\eta$, $\eta\downarrow 0$, we have

$$\langle \mathbf{p}|T-T^{\dagger}|\mathbf{p}'\rangle = -i\pi k\int \langle \mathbf{p}|T|\mathbf{k}\rangle\langle \mathbf{k}|T^{\dagger}|\mathbf{p}'\rangle\,d\hat{k}, \tag{1}$$

where the integration is over the unit sphere. Equation (1) is valid when the potential associated with T has a short range. However, for the Coulomb potential V_c Eq. (1) has to be modified because the half-shell limit of the off-shell Coulomb T matrix T_c does not exist. Instead we have[4]

$$\begin{aligned}\langle \mathbf{p}|T_c-T_c^{\dagger}|\mathbf{p}'\rangle &= -i\pi k\int \langle \mathbf{p}|T_c|\mathbf{k}\infty\rangle\langle \mathbf{k}\infty|T_c^{\dagger}|\mathbf{p}'\rangle\,d\hat{k}\\ &= -i\pi k\int \langle \mathbf{p}|V_c|\mathbf{k}+\rangle_c\,{}_c\langle \mathbf{k}+|V_c|\mathbf{p}'\rangle\,d\hat{k},\end{aligned} \tag{2}$$

where $|\mathbf{k}\infty\rangle$ is the so-called Coulombian asymptotic state and $|\mathbf{k}+\rangle_c$ is the Coulomb scattering state with energy $(k+i\epsilon)^2$, $\epsilon\downarrow 0$. The left-hand side of Eq. (2) is known in closed form (Ref. 4). We rewrite the right-hand side by inserting

$$\langle \mathbf{p}|V_c|\mathbf{k}+\rangle_c = \sum_{l=0}^{\infty}(4\pi)^{-1}(2l+1)P_l(\hat{p}\cdot\hat{k})\langle p|V_{cl}|kl+\rangle_c, \tag{3}$$

and using the orthogonality relation

$$\int P_l(\hat{p}\cdot\hat{k})P_{l'}(\hat{p}'\cdot\hat{k})\,d\hat{k} = 4\pi(2l+1)^{-1}P_l(\hat{p}\cdot\hat{p}')\delta_{ll'}. \tag{4}$$

In Eq. (3), $|kl+\rangle_c$ is the partial-wave Coulomb scattering state. Denoting $(p^2+k^2)/(2pk)$ by y and assuming $p>k$, we have[4]

$$\langle p|V_{cl}|kl+\rangle_c = 2\gamma(\pi p)^{-1}e^{(1/2)\pi\gamma}Q_l^{i\gamma}(y), \tag{5}$$

where γ is Sommerfeld's parameter, which is real ($k>0$). It is important to note that $Q_l^{i\gamma}(y)$ is not real-analytic: For the complex conjugate of both members of Eq. (5) we obtain

$$\langle p|V_c|kl+\rangle_c^{*} = 2\gamma(\pi p)^{-1}e^{-(3/2)\pi\gamma}Q_l^{-i\gamma}(y). \tag{6}$$

In the above indicated way we obtain from Eqs. (2)–(6),

$$\begin{aligned}&\sum_{l=0}^{\infty}(2l+1)P_l(x)Q_l^{i\gamma}(y)Q_l^{-i\gamma}(z)\\ &\quad-[\Gamma(1+i\gamma)\Gamma(1-i\gamma)/2i\gamma(\alpha_+\alpha_-)^{1/2}]\\ &\quad\times(Y^{i\gamma}-Y^{-i\gamma})\\ &\quad=-\pi\sin(\gamma\ln Y)/(\alpha_+\alpha_-)^{1/2}\sinh\pi\gamma.\end{aligned} \tag{7}$$

Here $x=\hat{p}\cdot\hat{p}'$, $z=(p'^2+k^2)/(2p'k)$, $p'>k$,

$$\alpha_{\pm} = yz-x\pm(y^2-1)^{1/2}(z^2-1)^{1/2}, \tag{8}$$

$$Y=(\alpha_+^{1/2}-\alpha_-^{1/2})/(\alpha_+^{1/2}+\alpha_-^{1/2}). \tag{9}$$

For convenience we introduce the quantity W,

$$W=W(x,y,z)=x^2+y^2+z^2-2xyz-1. \tag{10}$$

Then we have $\alpha_+\alpha_- = W\geqslant 0$,

$$Y^2=(yz-x-W^{1/2})/(yz-x+W^{1/2}), \tag{11}$$

so that Eq. (7) can be rewritten as

$$\begin{aligned}&\sum_{l=0}^{\infty}(2l+1)P_l(x)Q_l^{i\gamma}(y)Q_l^{-i\gamma}(z)\\ &\quad=-\pi\sin(\tfrac{1}{2}\gamma\ln Y^2)/W^{1/2}\sinh\pi\gamma.\end{aligned} \tag{12}$$

By analytic continuation it follows that Eq. (12) is valid for complex x, y, z, and γ. The series in Eq. (12) is convergent if $\mathrm{Re}\,x>0$, $\mathrm{Re}\,y>0$, $\mathrm{Re}\,z>0$, and

$$|x+(x^2-1)^{1/2}|<|y+(y^2-1)^{1/2}|\cdot|z+(z^2-1)^{1/2}|. \tag{13}$$

When $\mathrm{Re}\,x<0$, one should replace x by $-x$ in Eq. (13), and similarly for y and z. It may be noted that

$$P_l(-y)=(-1)^lP_l(y),$$

$$Q_l^{i\gamma}(-z)=(-1)^{l+1}Q_l^{i\gamma}(z).$$

Now we are going to consider the more general expression

$$\begin{aligned}&F_{mn}(x_1,\ldots,x_m;z_1,\ldots,z_n)\\ &\quad=\sum_{l=0}^{\infty}(2l+1)P_l(x_1)\cdots P_l(x_m)Q_l(z_1)\cdots Q_l(z_n)\end{aligned} \tag{14}$$

 0022-2488/81/112482-02$01.00

(cf. Ref. 5) for $n, m = 0,1,2,3$, $x_i \in \mathbb{C}$, $y_i \in \mathbb{C}\backslash[-1,1]$. When $\mathrm{Re}\, x_i > 0$, $\mathrm{Re}\, z_j > 0$, this series is convergent if

$$\prod_{i=1}^{m} |x_i + (x_i^2 - 1)^{1/2}| < \prod_{j=1}^{n} |z_j + (z_j^2 - 1)^{1/2}|. \tag{15}$$

Let us first consider F_{12}. By taking the limit for $\gamma \to 0$ in Eq. (12) we obtain

$$F_{12}(x;y,z) = \sum_{l=0}^{\infty} (2l+1) P_l(x) Q_l(y) Q_l(z)$$
$$= \tfrac{1}{2} W^{-1/2} \ln \frac{yz - x + W^{1/2}}{yz - x - W^{1/2}}. \tag{16}$$

It is interesting to note that

$$F_{12}(x;y,z) = \frac{1}{4\pi} \int (y - \hat{p}\cdot\hat{q})^{-1} (z - \hat{p}'\cdot\hat{q})^{-1}\, d\hat{q},$$

where $\hat{p}\cdot\hat{p}' = x$ and the integration is over the unit sphere. In Ref. 5 an expression has been given for $F_{12}(x;z,z)$. Our result given by Eq. (16) means a considerable reduction of that expression. Indeed, in the notation of Ref. 5 we have $d = z^2 - 1 + t^2$, $2t^2 = 1 - y$, and Eq. (16) gives

$$F_{12}(y;z,z) = \frac{1}{4td^{1/2}} \ln\left(\frac{t + d^{1/2}}{t - d^{1/2}}\right)^2, \tag{17}$$

whereas in Ref. 5 the following result is given,

$$F_{12}(y;z,z) = \frac{1}{4td^{1/2}}$$
$$\times \ln\left[\frac{1 + t[z + 2 + 2t^2/(z-1)]d^{-1/2} + 2t^2/(z-1)}{1 + t[z - 2 + 2t^2/(z+1)]d^{-1/2} - 2t^2/(z+1)}\right].$$

To demonstrate the equivalence of this result and that in Eq. (17) is *not* completely trivial. It can be done by dividing out the common factor $(1 + tzd^{-1/2})(z^2 - 1)^{-1}$ from the numerator and the denominator of the fraction which forms the argument of the logarithm. By this procedure Eq. (17) is retrieved.

We shall briefly consider some other interesting particular cases of the general function F_{mn}. By taking $x = 1$ in Eq. (16) we obtain the well-known result

$$F_{02}(y,z) = \tfrac{1}{2}(y - z)^{-1} \ln\left(\frac{y-1}{y+1}\frac{z+1}{z-1}\right). \tag{18}$$

Other well-known formulas are[6]

$$F_{21}(x,y;z) = W^{-1/2}, \tag{19}$$
$$F_{11}(y;z) = (z - y)^{-1}, \tag{20}$$
$$F_{01}(z) = (z - 1)^{-1}. \tag{21}$$

Eq. (20) is called Heine's formula.

When $n = 0$ we shall restrict x, y, and z in F_{m0} to the interval $[-1,1]$. According to Ref. 6, p. 307 we have

$$F_{30}(x,y,z) = \begin{cases} 0 & \text{if } W > 0 \\ 2\pi^{-1}(-W)^{-1/2} & \text{if } W < 0 \end{cases}. \tag{22}$$

Furthermore, we have [cf. Eq. (4)]

$$F_{20}(x,y) = 2\delta(y - x), \tag{23}$$
$$F_{10}(x) = 2\delta(1 - x), \tag{24}$$

where δ is Dirac's delta distribution.

Finally we shall briefly consider F_{03}. In virtue of Eq. (16) we have

$$F_{03}(p,y,z) = \sum_{l=0}^{\infty} (2l+1) Q_l(p) Q_l(y) Q_l(z)$$
$$= \frac{1}{2}\int_{-1}^{1} \frac{dx}{p - x} \sum_{l=0}^{\infty} (2l+1) P_l(p) Q_l(y) Q_l(z)$$
$$= \frac{1}{4}\int_{-1}^{1} \frac{W^{-1/2}}{p - x} \ln \frac{yz - x + W^{1/2}}{yz - x - W^{1/2}}\, dx.$$

Putting $a = (y^2 - 1)^{1/2}(z^2 - 1)^{1/2}$, $v = \mathrm{arcosh}((yz - x)/a)$, $v_{\pm} = \mathrm{arcosh}((yz \pm 1)/a)$ we get $W^{1/2} = a \sinh v$ and

$$F_{03}(p,y,z) = (1/2)a \int_{v_-}^{v_+} \frac{v\, dv}{\cosh v - (yz - p)/a}. \tag{25}$$

According to formula 2.478.7 of Ref. 7 we have

$$\int \frac{x\, dx}{\cosh 2x - \cos 2t}$$
$$= \frac{1}{2\sin 2t}[L(u + t) - L(u - t) - 2L(t)], \tag{26}$$

where $u = \arctan(\tanh x \cot t)$ and L is Lobachevski's function, defined by

$$L(x) = -\int_0^x \ln(\cos t)\, dt. \tag{27}$$

This implies that F_{03} cannot be expressed in terms of elementary functions.

By using the series representation

$$L(x) = -x \ln 2 + (1/2) \sum_{n=1}^{\infty} (-1)^n n^{-2} \sin 2nx, \tag{28}$$

the right member of Eq. (26) can be rewritten as

$$\frac{1}{4\sin 2t} \sum_{n=0}^{\infty} (-1)^n n^{-2} \sin 2nt \cos^2 nu. \tag{29}$$

We point out that on p. 377 of Ref. 6, Eq. (56.8.1), a closed formula is given for the series

$$\sum_{l=0}^{\infty} (2l+1) P_l(x) P_l^m(y) P_l^{-m}(z), \tag{30}$$

where $m \in \mathbb{N}$ and $x,y,z \in [-1,1]$.

ACKNOWLEDGMENT

This work is part of the project "The Coulomb Potential in Quantum Mechanics and Related Topics."

[1] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Chap. 7.2.2.

[2] K. M. Watson and J. Nuttall, *Topics in Several Particle Dynamics* (Holden-Day, San Francisco, 1967), Chap. 2.4.

[3] J. Nuttall and R. W. Stagat, Phys. Rev. A **3**, 1355 (1971).

[4] H. van Haeringen, "Long Range Potentials in Quantum Mechanics," Report 150 (University of Groningen, 1979).

[5] J. M. Greben, Ph.D. Thesis (University of Groningen, 1974), Prop. 3 (in Dutch).

[6] E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, N.J., 1975).

[7] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Corrected and Enlarged Edition* (Academic, New York, 1980). Unfortunately this work contains a number of misprints and errors. Some of these (i.e., a few hundred) have been collected in H. van Haeringen and L. P. Kok, Corrigenda Report (Delft-Groningen, 1981).

An expansion of the moments of powers of a Fermi distribution

H. Krivine and J. Treiner
Division de Physique Théorique, [a] *Institut de Physique Nucléaire, 91406 Orsay Cedex, France*

(Received 24 March 1981; accepted for publication 22 May 1981)

We derive an expansion of the moments of the νth power of a Fermi distribution for any real value of ν in powers of the ratio of the surface thickness to the radius of the distribution. For $\nu = 1$ we recover the Sommerfeld lemma.

PACS numbers: 02.30.Mv

In nuclear physics, the nuclear density is often described by a Fermi distribution

$$\rho(r) \sim \frac{1}{1+\exp[(r-R)/a]}, \tag{1}$$

where R is the half-density radius and a the surface thickness of the nucleus. The various moments of the density can be expanded in powers of $x = a/R$, using the Sommerfeld lemma,[1] and volume, surface, curvature, and higher-order terms can then be extracted. However, in some recent developments of the energy density formalism,[2] one has to deal with the moments of noninteger powers of the density, i.e., with the following integrals:

$$I_{\nu,\mu} = \int_0^\infty \frac{r^\mu\, dr}{\{1+\exp[(r-R)/a]\}^\nu} = R^{\mu+1}\int_0^\infty \frac{r^\mu\, dr}{\{1+\exp[(r-1)/x]\}^\nu} \tag{2}$$

for any positive value of ν and any positive integer value of μ. We present here a generalization of the Sommerfeld expansion, which stands as follows:

$$I_{\nu,\mu} = \frac{R^{\mu+1}}{\mu+1}[1 + xP_{\nu,\mu}(x) + \omega_{\nu\mu}(x)], \tag{3}$$

where (i) $P_{\nu,\mu}$ is a polynomial of order μ and (ii) the function $\omega_{\nu\mu}(x)$ goes to zero faster than any power of x.

We split the last integral of Eq. (2) in the following two contributions:

$$I_1 = \int_0^1 \frac{r^\mu\, dr}{\{1+\exp[(r-1)/x]\}^\nu} \tag{4a}$$

and

$$I_2 = \int_1^\infty \frac{r^\mu\, dr}{\{1+\exp[(r-1)/x]\}^\nu}. \tag{4b}$$

Extracting the zeroth order term of Eq. (4a)

$$I_2 = \frac{1}{\mu+1} + \int_0^1 \left(\frac{r^\mu}{\{1+\exp[(r-1)/x]\}^\nu} - r^\mu\right) dr \tag{5}$$

and introducing the change of variable

$$r = 1 - ux, \tag{6}$$

we obtain

$$I_1 = \frac{1}{\mu+1} + x\int_0^{1/x} (1-ux)^\mu \left[\frac{1}{(1+e^{-u})^\nu} - 1\right] du \tag{7}$$

or, equivalently,

$$I_1 = \frac{1}{\mu+1} + x\int_0^\infty (1-ux)^\mu \left[\frac{1}{(1+e^{-u})^\nu} - 1\right] du - x\int_{1/x}^\infty (1-ux)^\mu \left[\frac{1}{(1+e^{-u})^\nu} - 1\right] du. \tag{8}$$

We can thus write

$$I_1 = \frac{1}{\mu+1} + \sum_{k=1}^{\mu+1} \gamma_k^{(1)} x^k - xJ_{\nu\mu} \tag{9}$$

with

$$\gamma_k^{(1)} = (-)^{k-1}\binom{\mu}{k-1}\int_0^\infty u^{k-1}\left[\frac{1}{(1+e^{-u})^\nu} - 1\right] du \tag{10a}$$

and

$$J_{\nu\mu} = \int_{1/x}^\infty (ux-1)^\mu \left[\frac{1}{(1+e^{-u})^\nu} - 1\right] du. \tag{10b}$$

In Eq. (10a), $\binom{n}{k}$ stands for the binomial coefficient.

An equivalent calculation for I_2, Eq. (4b) with the change of variable

$$t = 1 + \mu x \tag{11}$$

leads to

$$I_2 = x\int_0^\infty \frac{(1+ux)^\mu}{(1+e^u)^\nu}\, du = \sum_{k=1}^{\mu+1} \gamma_k^{(2)} x^k \tag{12}$$

with

$$\gamma_k^{(2)} = \binom{\mu}{k-1}\int_0^\infty \frac{u^{k-1}}{(1+e^u)^\nu}\, du. \tag{13}$$

We thus have proven the first part of the theorem, with $P_{\nu,\mu}(x)$ given by

$$P_{\nu,\mu}(x) = (\mu+1)\sum_{k=0}^{\mu}\binom{n}{k}\eta_\nu^{(k)} x^k, \tag{14a}$$

where

$$\eta_\nu^{(k)} = (-)^k \int_0^\infty u^k \left[\frac{1+(-)^k e^{-u\nu}}{(1+e^{-u})^\nu} - 1\right] du \tag{14b}$$

is μ-independent.

Let us note that in the special case when $k = 0$ and ν rational the integral in (14b) can be performed analytically. It is now straightforward to obtain an upper bound for the integral $J_{\nu\mu}$ which is related to $\omega_{\nu\mu}$ by

$$\omega_{\nu\mu} = (\mu+1)xJ_{\nu\mu}. \tag{14c}$$

[a] Laboratoire associé au C. N. R. S.

 0022-2488/81/112484-02$01.00

Starting from the inequality

$$\left|1-\frac{1}{(1+\epsilon)^{\nu}}\right|\leqslant\nu\epsilon \tag{15}$$

which holds when $\epsilon\geqslant 0$, we get

$$|J_{\nu\mu}|\leqslant\nu x^{\mu}\int_{1/x}^{\infty}u^{\mu}e^{-u}\,du. \tag{16}$$

Defining

$$K_{\mu}=\int_{1/x}^{\infty}u^{\mu}e^{-u}\,du, \tag{17}$$

which satisfies the recursion relation,

$$K_{\mu}=(1/x)^{\mu}e^{-1/x}+\mu K_{\mu-1} \tag{18}$$

we obtain

$$K_{\mu}=e^{1/x}[(1/x)^{\mu}+\mu(1/x)^{\mu-1}+\mu(\mu-1)(1/x)^{\mu-2}+\cdots+\mu!(1/x)+\mu!]. \tag{19}$$

So the inequality (16) reads

$$|J_{\nu\mu}|\leqslant\nu e^{-1/x}\frac{1-(\mu x)^{\mu+1}}{1-\mu x}. \tag{20}$$

If one assumes now $\mu x<1$, Eq. (14c) leads to

$$|\omega_{\nu\mu}|\leqslant(\mu+1)\nu\frac{xe^{-1/x}}{1-\mu x}, \tag{21}$$

which proves the second part of the theorem.

Table I gives some numerical values of the coefficients defined by Eq. (14b) for $\nu=\frac{1}{6},\frac{1}{3}$, and $\frac{2}{3}$. Other values may be fixed by using the following recursion relations:

$$I_{\nu,\mu}=\frac{\nu}{a(\mu+1)}(I_{\nu,\mu+1}-I_{\nu+1,\mu+1}), \tag{22a}$$

$$\frac{dI_{\nu,\mu}}{dR}=\frac{\nu}{a}(I_{\nu,\mu}-I_{\nu+1,\mu}), \tag{22b}$$

$$\mu I_{\nu,\mu-1}=\frac{dI_{\nu,\mu}}{dR}. \tag{22c}$$

Equation (22a) is obtained by integrating by parts Eq. (2), Eq. (22b) by differentiating Eq. (2) with respect to R. Equation (22c) is a combination of Eqs. (22a) and (22b).

TABLE I. Values of $\eta_{\nu}^{(k)}$ [see Eq. (14b)] for $k=0,\ldots,4$ and for $\nu=\frac{1}{6},\frac{1}{3}$, and $\frac{2}{3}$.

k \ ν	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$
0	5.75491	2.55482	0.741019
1	36.0412	9.13381	2.62896
2	431.493	53.1261	5.28519
3	7776.45	487.333	33.7307
4	186620.	5822.22	165.466

If $\nu=1$, it is clear from Eq. (14b) that

$$\eta_1^{(k)}=0 \quad \text{if } k \text{ is even} \tag{23a}$$

and

$$\eta_1^{(k)}=2\int_0^{\infty}\frac{e^{-u}}{1+e^{-u}}u^{k-1}\,du \quad \text{if } k \text{ is odd.} \tag{23b}$$

By expanding $e^{-u}/(1+e^{-u})$ in powers of e^{-u} and using the properties of the Riemann function ζ, one obtains

$$\eta_1^{(2p-1)}=2(2p-1)!(1-2^{1-2p})\zeta(2p), \tag{24}$$

recovering the result of the Sommerfeld lemma.

Finally, let us stress that, even in the domain where $\omega_{\nu\mu}(x)$ is small, the various contributions of order 1, 2, ..., μ in $P_{\nu,\mu}(x)$ are not necessarily decreasing. Particularly for small values of ν, it could be misleading to truncate the polynomial to its lowest order terms.

ACKNOWLEDGMENT

We thank Professor J. L. Krivine for a helpful discussion.

[1] S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Dover, New York, 1967), 2nd ed. (1st ed., 1939).

[2] J. Treiner, H. Krivine, O. Bohigas, and J. Martorell, Nucl. Phys. A (in press).

Irreducible representations of the complex rotation group in terms of the axis and angles of rotations

K. N. Joshi
Department of Mathematics, Garhwal University, Srinagar (Garhwal), U. P., India

B. S. Rajput
Department of Physics, Garhwal University, Srinagar (Garhwal), U. P., India

(Received 28 April 1980; accepted for publication 12 September 1980)

The expression for matrix elements corresponding to finite complex rotations have been derived in terms of the complex Euler angles of the complex rotations. Using this expression the values of matrix elements for special cases of eigenvalues of complex angular momentum operators have been computed.

PACS numbers: 02.20.Qs

1. INTRODUCTION

In our earlier papers[1-3] we developed a compact operator formulation to reformulate the Gel'fand–Naimark[4] theory of representations of SL(2,C) group by combining the generators of homogeneous Lorentz group. Diagonal elements of the operators thus obtained, have been interpreted as the components $\hat{Z}_1$, $\hat{Z}_2$, and $\hat{Z}_3$ of complex angular momentum operator $\hat{Z}$ in a complex space. Representations of complex angular momentum operator $\hat{Z}$ have been derived in the basis of eigenvectors of third component of ordinary angular momentum operator and it has been shown that the space R in which the generators of homogeneous Lorentz group operate may be analyzed into a linear sum of invariant subspaces R_l in each of which an irreducible representation of weight 'l' of the group of rotations is obtained. It has also been shown that the eigenvectors of the generators of complex angular momentum operator form the complete set in the subspace in which only the weight 'l' participates in the representations while in the subspaces in which weights '$l-1$' and '$l+1$' participates separately, they do not form the complete set.

Using Moses[5] parameterization of rotation group in terms of well-known Euler's theorem, we have obtained the parameterization of the group of complex angular momentum operators[3] (group of complex rotations) in terms of Euler angles in complex planes as the natural generalization of the method developed by Bars and Gursey[6] and Huszar.[7] The matrix elements corresponding to finite complex rotations are given in terms of the complex Euler angles of the complex rotation. We have obtained the complex rotations by means of a complex vector ϕ, where the direction of ϕ is the direction of the axis of complex rotation and $|\phi|$ is the angle of complex rotation.

Our representation is useful in showing how wave functions transform in a rotated frame of reference when the axis of complex rotation is prescribed. One could also obtain the transformation properties of the wave function in terms of the Euler angle parameterization but the expressions are extremely cumbersome for a general axis of complex rotation. In the present representation, on the other hand, the expressions are quite simple. The matrix elements for finite complex rotation for the cases of complex angular momentum eigenvalues 1, 2, and 3 in complex space corresponding to the values of the spin $\frac{1}{2}$, 1, $\frac{3}{2}$ in real space, have been computed.

2. IRREDUCIBLE REPRESENTATIONS OF THE COMPLEX ROTATION GROUP

In our previous papers,[1,2] we combined the generators of proper, orthochronous, homogeneous Lorentz group to define the components of complex angular momentum operators in $(\hat{J},\hat{K})$ space in the following form

$$\hat{Z}_j = \tfrac{1}{2}(\hat{J}_j + i\hat{K}_j), \tag{1}$$

where $\hat{J}_j$ and $\hat{K}_j$ are respectively the generators of rotations about, and boosts along the jth axis and satisfy the commutation rules

$$[\hat{Z}_j,\hat{Z}_k] = i\epsilon_{jkl}\hat{Z}_l. \tag{2}$$

(ϵ_{jkl} is the Levi–Civita three index symbol). It has already been shown[1,2] that the complex angular momentum operators $\hat{Z}$ exhibits the properties similar to those obtained by many authors[8-10] for ordinary angular momentum operators. It has also been shown that the space 'R' in which the generators of homogeneous Lorentz group operate, may be analyzed into a linear sum of invariant subspaces R_l, in each of which an irreducible representation of weight 'l' of the group of ordinary rotations is obtained in the canonical basis $|l,m\rangle$. In cases when only the weight 'l' participates in the representation, the complex angular momentum operators act on the canonical basis $|l,m\rangle$ in the following way

$$\begin{aligned}
&\hat{Z}_3|l,m\rangle = A_1 m|l,m\rangle,\\
&(\hat{Z}_2 - i\hat{Z}_1)|l,m\rangle = A_1[(l-m)(l+m+1)]^{1/2}|l,m+1\rangle,\\
&(\hat{Z}_2 + i\hat{Z}_1)|l,m\rangle = A_1[(l+m)(l-m+1)]^{1/2}|l,m-1\rangle,
\end{aligned} \tag{3}$$

where

$$\begin{aligned}
&A_1 = \left(1 + \frac{l_0 l_1}{l(l+1)}\right);\\
&m = -l, -l+1,\cdots, l-1, l;\\
&l = l_0, l_0+1,\cdots;
\end{aligned} \tag{4}$$

and l_1 is some arbitrary complex number. Let us change the canonical basis $|l,m\rangle$ to $|L,M\rangle$, such that

 0022-2488/81/112486-03$01.00

$$M = A_1 m \tag{5a}$$

$$L = \frac{-1 + [1 + 4A_1^2 l(l+1) + 4A_1 M - 4M]^{1/2}}{2}, \tag{5b}$$

for

$$m = -l, \quad L = A_1 l, \tag{6a}$$

and for a given value of L, variable M takes on the values

$$M = -L, -L + A_1, -L + 2A_1, \cdots, L - A_1, L. \tag{6b}$$

The set of the kets $|L,M\rangle$ then satisfy the following orthonormality and completeness relations

$$\sum_M |L,M\rangle \langle L,M| = I,$$
$$\langle L,M'|L,M\rangle = \delta_{M,M'}. \tag{7}$$

In the basis $|L,M\rangle$, the operators $\hat{Z}_1$, $\hat{Z}_2$, and $\hat{Z}_3$ act in the following way

$$\hat{Z}_3|L,M\rangle = M|L,M\rangle,$$
$$(\hat{Z}_2 - i\hat{Z}_1)|L,M\rangle = [(L-M)(L+M+A_1)]^{1/2}|L,M+1\rangle, \tag{8}$$
$$(\hat{Z}_2 + i\hat{Z}_1)|L,M\rangle = [(L+M)(L-M+A_1)]^{1/2}|L,M-1\rangle.$$

As such, the matrix elements of the operators $\hat{Z}_j$ are given by

$$\langle L,M'|\hat{Z}_3|L,M\rangle = M\delta_{M,M'},$$
$$\langle L,M'|\hat{Z}_2 - i\hat{Z}_1|L,M\rangle = [(L-M)(L+M+A_1)]^{1/2}\delta_{M',M+1}, \tag{9}$$
$$\langle L,M'|\hat{Z}_2 + i\hat{Z}_1|L,M\rangle = [(L+M)(L-M+A_1)]^{1/2}\delta_{M',M-1}.$$

The kets in the basis of Refs. 9 and 10 may be denoted by $|L,M)$ and satisfy the completeness and orthonormality conditions (7). In the basis $|L,M)$ the operators act in the following way

$$\hat{Z}_3|L,M) = M|L,M),$$
$$(\hat{Z}_1 + i\hat{Z}_2)|L,M) = [(L-M)(L+M+A_1)]^{1/2}|L,M+1), \tag{10}$$
$$(\hat{Z}_1 - i\hat{Z}_2)|L,M) = [(L+M)(L-M+A_1)]^{1/2}|L,M-1).$$

Comparison of Eqs. (8) and (10) gives the following relation between the kets in two bases:

$$|L,M) = i^{L+M}|L,M\rangle,$$

and the matrix elements of an operator $\hat{A}$ in the two bases are related by

$$(L,M'|\hat{A}|L,M) = i^{M-M'}\langle L,M'|\hat{A}|L,M\rangle.$$

Let us introduce a third basis in which $\boldsymbol{\phi}\cdot\mathbf{Z}$ is diagonal for any complex vector $\boldsymbol{\phi}$. Let the eigenvalues of $(\boldsymbol{\phi}\cdot\mathbf{Z})/\phi$ be denoted by λ, where λ can take on the following $(2L + A_1)/A_1$ values

$$-L, -L + A_1, \cdots, L - A_1, L.$$

It may easily be verified that the eigenkets $|L,\lambda\rangle$ of $(\boldsymbol{\phi}\cdot\mathbf{Z})/\phi$ satisfy the same orthogonality and completeness relations as those for kets $|L,M\rangle$.

We may also write

$$\boldsymbol{\phi}\cdot\mathbf{Z}|L,\lambda\rangle = \phi\lambda|L,\lambda\rangle. \tag{11}$$

Kets $|L,\lambda\rangle$ and $|L,M\rangle$ are related by transformation matrix $\langle L,\lambda|L,M\rangle$ given by

$$\langle L,\lambda|L,M\rangle = \left(\frac{1}{2}\right)^M \left(-\frac{\phi_2 - i\phi_1}{\phi}\right)^{M-\lambda} \left(1 + \frac{\phi_3}{\phi}\right)^{\lambda} \times \left(\frac{(L-M)!(L+M)!}{(L-\lambda)!(L+\lambda)!}\right)^{1/2} S(L,M,\lambda,\phi_3/\phi), \tag{12}$$

where function $S(L,M,M',x)$ is defined as follows[11];

$$S(L,M,M',x) = (L-M)!(L+M)!2^{M'-L} \times \sum_{q=0}^{L-M'} \frac{(x+1)^q (x-1)^{L-M'-q}}{q!(L-M+q)!(L-M'-q)!(L+M+M')!}. \tag{13}$$

Matrix elements of $\exp[i(\boldsymbol{\phi}\cdot\mathbf{Z})]$ may then be defined in the following form

$$\langle L,M'|\exp(i\boldsymbol{\phi}\cdot\mathbf{Z})|L,M\rangle = \left(\frac{(L-M)!(L+M)!}{(L-M')!(L+M')!}\right)^{1/2} \left(\sin\frac{\phi}{2}\right)^{M-M'} \times \left(\frac{\phi_1 + i\phi_2}{\phi}\right)^{M-M'} \left(\cos\frac{\phi}{2} + i\frac{\phi_3}{\phi}\sin\frac{\phi}{2}\right)^{M+M'} \times S(L,M,M',x), \quad \text{for} M \geqslant |M'|,$$
$$= (-1)^{2L+M'+M} \left(\frac{(L-M)!(L+M)!}{(L-M')!(L+M')!}\right)^{1/2} \times \sin\left(\frac{\phi}{2}\right)^{M'-M} \left(\frac{\phi_1 - i\phi_2}{\phi}\right)^{M'-M} \times \left(\cos\frac{\phi}{2} - i\frac{\phi_3}{\phi}\sin\frac{\phi}{2}\right)^{-(M'+M)} \times S(L,-M,-M',x), \quad \text{for } M \leqslant -|M'|,$$

and

$$\langle L,M'|\exp(i\boldsymbol{\phi}\cdot\mathbf{Z})|L,M\rangle = \left(\frac{(L-M')!(L+M')!}{(L-M)!(L+M)!}\right)^{1/2} \left(\sin\frac{\phi}{2}\right)^{M-M'} \times \left(\frac{\phi_1 + i\phi_2}{\phi}\right)^{M-M'} \left(\cos\frac{\phi}{2} - i\frac{\phi_3}{\phi}\sin\frac{\phi}{2}\right)^{-(M+M')} \times S(L,-M',-M,x), \quad \text{for } M' \leqslant -|M|. \tag{14}$$

where M is given by (5), $L = A_1 l$ for $m = -l$, and

$$x = (1 - \phi_3^2/\phi^2)\cos\phi + \phi_3^2/\phi^2. \tag{15}$$

Using the expressions (14) and (15) the matrix elements of $\exp[i(\boldsymbol{\phi}\cdot\mathbf{Z})]$ may be calculated for different finite values of spin. For instance, let us compute the matrix elements

$$\langle L,M'|\exp[i(\boldsymbol{\phi}\cdot\mathbf{Z})]|L,M\rangle = (M',M)$$

for the three cases of complex angular momentum eigenvalues 1, 2, and 3 in complex space corresponding to the values of the spin $\frac{1}{2}$, 1, and $\frac{3}{2}$ in real space (see Appendix), in terms of u, v, and w given by

$$u = \sin\frac{\phi}{2},$$

$$v = \frac{\phi_1 + i\phi_2}{\phi},$$

$$w = \cos\frac{\phi}{2} + i\frac{\phi_3}{\phi}\sin\frac{\phi}{2}.$$

The matrix elements of $\exp[i(\boldsymbol{\phi},\mathbf{Z})]$ are calculated for the finite rotation for the cases spin $\frac{1}{2}$, 1, and $\frac{3}{2}$. For the finite representation l_1 is integral or half-integral with l_0, the lowest weight participating in the representation and $|l_1| > l_0$ so that $l_1 = l + 1$ and $l = l_0$.

APPENDIX

Matrix elements of $\exp[i(\boldsymbol{\phi}\cdot\mathbf{Z})]$.

Case I: For spin $\frac{1}{2}$ we have $l = l_0 = \frac{1}{2}$, $l_1 = \frac{3}{2}$, $A_1 = 2$, $L = 1$, and hence we obtain:

$$(-1,-1) = w^{*2},$$
$$(-1,0) = 2^{1/2}uvw^*,$$
$$(-1,1) = u^2v^2,$$
$$(0,-1) = -2^{1/2}uv^*w^*,$$
$$(0,0) = -u^2vv^* + ww^*,$$
$$(0,1) = 2^{1/2}uvw,$$
$$(1,-1) = u^2w^{*2},$$
$$(1,0) = -2^{1/2}uv^*w,$$
$$(1,1) = w^2.$$

Case II: For spin 1 we have $l = l_0 = 1$, $l_1 = 2$, $A_1 = 2$, $L = 2$, and the following values of matrix elements are obtained:

$$(-2,-2) = w^{*4},$$
$$(-2,0) = (6)^{1/2}u^2v^2w^{*2},$$
$$(-2,-2) = u^4v^4,$$
$$(0,-2) = (6)^{1/2}u^2v^{*2}w^{*2},$$
$$(0,0) = u^2vv^* + 2u^2vv^*ww^* + ww^*,$$
$$(0,2) = (6)^{1/2}u^2v^2w^2,$$
$$(2,0) = (6)^{1/2}u^2v^{*2}w^2,$$
$$(2,-2) = u^4v^{*4},$$
$$(2,2) = w^4,$$
$$(2,1) = -\tfrac{1}{2}uv^*(w^*)^{-3},$$
$$(1,2) = 2uvw^3,$$
$$(-2,1) = u^3v^3w^*,$$
$$(2,-1) = -\tfrac{1}{2}(uv^*)^3(w^*)^{-1},$$
$$(1,-2) = -2v^3v^{*3}w^3,$$
$$(-1,2) = \tfrac{1}{2}u^3v^3(w^*)^{-1},$$
$$(-1,-2) = -2uv^*w^{*3},$$
$$(-2,-1) = 2uvw^{*3}.$$

Case III: For spin $\frac{3}{2}$ we have $l = l_0 = \frac{3}{2}$, $l_1 = \frac{5}{2}$, $A_1 = 2$, $L = 3$, and the following values of matrix elements are obtained:

$$(-3,-3) = (w^*)^6,$$
$$(3,-3) = (uv^*)^6,$$
$$(-3,3) = u^6v^6,$$
$$(3,3) = w^6,$$
$$(3,1) = (\tfrac{1}{15})^{1/2}u^2v^2w^4,$$
$$(-1,3) = (15)^{1/2}u^4v^4w^4,$$
$$(3,-1) = (\tfrac{1}{15})^{1/2}u^4v^{*4}(w^*)^{-2},$$
$$(-3,1) = (15)^{1/2}(uv)^4w^{*2},$$
$$(1,-3) = (15)^{1/2}(uv^*)^4(w^*)^2,$$
$$(-1,-3) = (15)^{1/2}u^2v^{*2}w^{*4},$$
$$(-3,-1) = (15)^{1/2}u^2v^2w^{*4},$$
$$(1,3) = (\tfrac{1}{15})^{1/2}u^2v^2w^4.$$

[1] K. N. Joshi and B. S. Rajput, Prog. Theor. Phys. (to appear).
[2] K. N. Joshi and B. S. Rajput, J. Math. Phys. **21**, 1579 (1980).
[3] K. N. Joshi and B. S. Rajput, Phys. Lett. A (to appear).
[4] I. M. Gel'fand and M. A. Naimark, Izv. Acad. Nauk SSSR Ser. Mat. **11**, 1411 (1947).
[5] H. E. Moses, Nuovo Cimento **40**, 1120 (1965).
[6] I. Bars and F. Gursey, J. Math. Phys. **13**, 131 (1972).
[7] M. Huszar, J. Math Phys. **14**, 1620 (1973).
[8] V. Bargmann, Ann. Math. **48**, 568 (1947).
[9] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 15 (translated by J. J. Griffin).
[10] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., , Princeton, NJ, 1957).
[11] G. Szego, *Orthogonal Polynomials*, Vol. XXIII Rev. Ed. (Am. Math. Soc. Colloquium Publications, 1959).

Interpolation theory and refinement of nested Hilbert spaces

J.-P. Antoine
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
W. Karwowski[a)]
Centre de Physique Théorique, CNRS, Luminy, F-13288 Marseille, Cedex 2, France

(Received 20 January 1981; accepted for publication 22 May 1981)

Using the method of quadratic interpolation between Hilbert spaces, we show that a nested Hilbert space with countable index set admits arbitrarily many proper refinements by nested Hilbert spaces.

PACS numbers: 02.30.Sa, 02.30.Tb

I. INTRODUCTION

Let us begin with an example, namely, the familiar weighted L^2-spaces $H_\alpha \equiv L^2(\mathbb{R}, (1+x^2)^\alpha dx) \equiv L^2{}_\alpha$, $\alpha \in \mathbb{R}$. For every pair $\alpha > \beta$, one has $H_\alpha \subset H_\beta$, and, for each α, $H_{\bar\alpha} \equiv H_{-\alpha}$ is the (anti)dual of H_α with respect to the inner product $\langle f, g\rangle = \int_{-\infty}^{\infty} \bar f(x) g(x)\, dx$. Accordingly these spaces constitute a *continuous scale of Hilbert spaces* $H_{\bar I} = \{H_\alpha\}_{\alpha\in\mathbb{R}}$ in the sense of Palais,[1] Krein and Petunin,[2] or Berezanskii.[3] The continuous scale $H_{\bar I}$ contains the discrete scale $H_I = \{H_\alpha\}_{\alpha\in\mathbb{Z}}$ and one or the other is used in applications, depending on the problem at hand.

In fact, given the discrete scale H_I, the continuous scale $H_{\bar I}$ can be reconstructed uniquely ("functorially") by the method of *quadratic interpolation*, discussed by Palais[1] and Lions and Magenes.[4] Our aim in this paper is to extend this construction to the more general case of an involutive lattice of Hilbert spaces.

The proper framework for this problem is that of partial inner product (PIP) spaces, developed by Grossmann and one of us in a series of four papers,[5] to be quoted below as I–IV. Indeed both scales H_I, $H_{\bar I}$ are PIP-spaces on the vector space $H_{-\infty} = \cup_{\alpha\in\mathbb{R}} H_\alpha$, with compatibility $f \# g \Leftrightarrow \exists\ \alpha \in \mathbb{Z}$ (resp. $\mathbb{R}$) such that $f \in H_\alpha$, $g \in H_{\bar\alpha}$ and partial inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} \bar f(x) g(x)\, dx \quad \text{for } f \# g.$$

The interesting point is that $H_{\bar I}$ is a (proper) refinement of H_I (see III), in the sense that the complete involutive lattice generated by H_I is an involutive sublattice of the one generated by $H_{\bar I}$. Thus quadratic interpolation may be viewed as a particular method for refining the PIP-space H_I. It is well known (see III, IV) that the problem of refining an arbitrary PIP-space has in general no solution. But we will show here that such a solution (and in fact infinitely many solutions) always exists for the case of a countable lattice of Hilbert spaces [PIP-space of type (H)]. The method of the proof is a straightforward combination of quadratic interpolation and the spectral theorem for self-adjoint operators.

As said above, our framework will consist, as in our previous work,[6] of a vector space V together with a countable, covering, involutive lattice of Hilbert subspaces of V, $\mathcal{F} = \{F_n,\ n \in I\}$, that is:

(a) $\mathcal{F}$ covers V: $V = \Sigma_{n\in I} F_n$.

(b) $\mathcal{F}$ is a lattice with respect to intersection and vector sum; given $n\ m \in I$, $F_{n\wedge m} \equiv F_n \cap F_m$ and $F_{n\vee m} \equiv F_n + F_m$ also belong to $\mathcal{F}$.

(c) $\mathcal{F}$ carries an involution $n \longleftrightarrow \bar n$ which interchanges the lattice operations: $F_{\overline{n\wedge m}} = F_{\bar n \vee \bar m}$.

(d) For each $n \in I$, F_n is a Hilbert space (that is, with a *fixed* inner product, not only a Hilbertian topology) and $F_{\bar n} = (F_n)$, the antidual of F_n.

As an immediate consequence, we have:

(i) For each $n \in I$, the norm topology on F_n coincides with the Mackey topology $\tau(F_n, F_{\bar n})$.

(ii) Whenever $F_n \subseteq F_m$, the injection $E_{mn} : F_n \to F_m$ is continuous and has dense range.

(iii) $V^\# \equiv \cap_{n\in I} F_n$ is dense in every F_n.

As in Ref. 6, we will assume that each E_{mn} has norm at most one, i.e., $F_n \subseteq F_m$ implies $\|f\|_m \leqslant \|f\|_n$ for all $f \in F_n$.

It is clear (see III, IV, and Ref. 6) that the family $\mathcal{F}$ defines on V a structure of PIP-space, with compatibility

$$f \# g \Leftrightarrow \exists\ n \in I \quad \text{such that } f \in F_n,\ g \in F_{\bar n} \tag{1.1}$$

and partial inner product

$$\langle f \mid g\rangle = \overline{f(g)}, \quad \text{for } f \# g, \tag{1.2}$$

where $f(g)$ is the value of the antilinear functional f evaluated on the element g. More precisely, we obtain in this way a nondegenerate PIP-space of type (H) with countable index set I.

We assume further the form $\langle\cdot|\cdot\rangle$ to be positive, i.e., $\langle f \mid f\rangle > 0$ for every nonzero, self-compatible $f \in V$. We can then assume, without loss of generality,[6] the existence of a central, self-dual Hilbert space $H_0 \equiv F_0 = F_{\bar 0}$, which is the completion of $V^\#$ in the norm $\|f\| = \langle f \mid f\rangle^{1/2}$, and consists exactly of all self-compatible vectors. Then our PIP-space is simply a nested Hilbert space (NHS), in the sense of Grossmann.[7]

We will review first some results from the theory of interpolation between Hilbert spaces, following Refs. 1–4. Then we proceed to the proof of our main result, namely that any countable NHS admits arbitrarily many proper refinements, which are themselves NHS. We treat successively a scale of Hilbert spaces (that is, a chain built on the powers of a single positive self-adjoint operator), then a general chain (in both cases, we denote assaying subspaces by H_n), and finally a general lattice, as described above. From the concluding remarks, we would like to point out here a different

[a)] Permanent address : Institute of Theoretical Physics, University of Wrocław, Poland.

 0022-2488/81/112489-08$01.00

method of construction of an NHS, starting from a suitable family of positive self-adjoint operators on a Hilbert space, or, equivalently, a family of closed quadratic forms.

Our method has many applications for chains or even lattices of Hilbert spaces which are quite common in many fields of mathematical physics. To mention a few examples:

(i) In scattering theory, chains of Hilbert spaces built on the powers of $(1 + |\mathbf{x}|^2)$ or $(1 + |\mathbf{p}|^2)$ are an essential tool. We will discuss such spaces in Sec. V below.

(ii) Nelson's formulation of Euclidean field theory[8] uses the scale built from the Hamiltonian, or alternatively from the Casimir operator of the translation group. In fact, the Euclidean field is defined precisely as an operator (in the sense of II) in the PIP-space corresponding to the scale in question. Examples of such fields are the free Euclidean field[9] and elements of the Borchers class of the latter.[10]

(iii) The second scale mentioned in (ii) is in fact a particular case of a standard construction in group representation theory,[11] namely the C^∞-vectors of a continuous unitary representation of a Lie group in a Hilbert space. Again the proper framework is the scale corresponding to the so-called Nelson operator, i.e., the closure of the second order Casimir operator of the Lie algebra in the given representation.

(iv) A beautiful application of the concept of (finite) scale may be found in the recent papers of Grossmann, *et al.*[12] on quantum mechanical systems with local, many-center Hamiltonians. Here the defining operator is just $(K_0 - \lambda)$, where K_0 is the free Hamiltonian and λ is a real point in the resolvent set of K_0. This type of analysis applies to many situations of condensed matter physics (crystals, polymers, monomolecular layers,···).

In each of those examples the scale is used either for defining very singular operators, or for getting precise estimates on some operators of interest. In either case, refining the scale can only increase the effectiveness of the procedure. However, we will see below that chains are not quite enough, one is led very quickly to consider genuine lattices of Hilbert spaces as well.

Presumably, most of our results can be extended (with even more nonuniqueness) to the case of a countable PIP-space of type (B). Every F_n is then assumed to be a reflexive Banach space, and the general theory of interpolation has to be used (see, for instance, Ref. 13). However, we will restrict ourselves here to the simpler case of Hilbert spaces, which seems quite sufficient for most applications in mathematical physics (see, however, the comments after Lemma 3.2.).

II. RESULTS FROM INTERPOLATION THEORY

For future reference, we will collect here some well-known results about interpolation between Hilbert spaces. Further details may be found in Refs. 1, 2, and 4.

Let X_1, X_0 be two Hilbert spaces, with inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle$, respectively, such that $X_1 \subset X_0$, the injection $X_1 \to X_0$ is continuous with norm not greater than 1 and has dense range. Then there exists a *unique* self-adjoint operator A in X_0, with $A > 1$, such that $X_1 = Q(A)$, the form domain of A, and

$$\langle f,f\rangle_1 = \langle f, Af\rangle = \langle A^{1/2}f, A^{1/2}f\rangle, \quad f\in X_1 = Q(A) = D(A^{1/2}).$$

Then, for $0\leqslant t\leqslant 1$, one defines the space $X_t \equiv Q_t(X_1,X_0)$ as the form domain of the self-adjoint operator A^t, equipped with the inner product

$$\langle f,f\rangle_t = \langle f, A^t f\rangle = \langle A^{t/2}f, A^{t/2}f\rangle, \qquad f\in Q(A^t)\equiv D(A^{t/2}).$$

The spaces X_t $(0\leqslant t\leqslant 1)$ have the following properties:

(i) X_t is a Hilbert space, and for $0\leqslant t\leqslant s\leqslant 1$, one has $X_1 \subseteq X_s \subseteq X_t \subseteq X_0$, where all injections are continuous and have dense range.

(ii) The definition of X_t *is intrinsic* in the following sense: Replacing the inner products on X_1 and X_0 by equivalent ones results in a different operator A, but the *same* vector space with an equivalent topology.

(iii) *Interpolation property*: Let $X_1 \subseteq X_0$, $Y_1 \subseteq Y_0$ two such pairs, $\{X_t\}$, $\{Y_t\}$ being the corresponding interpolating scales. Let T be a linear bounded operator from X_0 into Y_0 and also from X_1 into Y_1. Then T is linear and bounded from X_t into Y_t for every t.

(iv) *Duality property*: Let $X_1^\times, X_0^\times$ be the (anti)duals of X_1, X_0. Then $[Q_t(X_1,X_0)]^\times = Q_{1-t}(X_0^\times, X_1^\times)$.

As an example we take again $X_1 = L^2(\mathbb{R},(1 + x^2)\,dx)$ and $X_0 = L^2(\mathbb{R},dx)$. Then $X_t = L^2(\mathbb{R},(1 + x^2)^t\,dx)$. More generally, with $X_n = L^2(\mathbb{R},(1 + x^2)^n\,dx)$, $n\in\mathbb{N}$, we have $Q_t(X_n,X_{-n}) = X_{n(2t-1)}$. In particular, $Q_{1/2}(X_n,X_{-n}) = X_0$ for every $n\in\mathbb{N}$.

This method of constructing the continuous scale $\{X_t\}_{0\leqslant t\leqslant 1}$ from the pair $\{X_1,X_0\}$ is called *quadratic interpolation* in Palais.[1]

III. REFINEMENT OF A SCALE OF HILBERT SPACES

Let us begin with the simplest situation, namely the familiar triplet of Hilbert spaces (we write $\bar{1} = -1$).

$$H_1 \subset H_0 \subset H_{\bar{1}}, \tag{3.1}$$

where $H_{\bar{1}} = H_1^\times$ is the antidual of H_1. This scale defines a PIP-space H_I on $H_{\bar{1}}$, with index set $I = \{1,0,\bar{1}\}$, complete lattice $\mathcal{F}$ identical with (3.1) and the natural partial inner product defined by Eq. (1.2). By quadratic interpolation we get a continuous scale of Hilbert spaces between H_1 and $H_{\bar{1}}$:

$$H_1 \subset \cdots \subset H_\alpha \subset \cdots \subset H_0 \subset \cdots \subset H_{\bar{\alpha}} \subset \cdots \subset H_{\bar{1}}, \tag{3.2}$$

where $H_\alpha = Q_\alpha(H_1,H_0)$ $(0\leqslant\alpha\leqslant 1)$ and $H_{\bar{\alpha}} = Q_{1-\alpha}(H_0,H_{\bar{1}}) = (H_\alpha)^\times$. The new scale (3.2) defines another PIP-space $H_{\tilde{I}}$ on $H_{\bar{1}}$, with index set $\tilde{I} = [1,-1]$ and the same partial inner product as H_I. Obviously, $H_{\tilde{I}}$ is a proper refinement of the original PIP-space H_I. The corresponding complete lattice, $\tilde{\mathcal{F}}$ is indexed by the "nonstandard" interval obtained by adding to each point α infinitesimally close points α_- (for $1\geqslant\alpha > -1$) and α_+ (for $1 > \alpha\geqslant -1$). This construction was described in detail in III.[5] Two facts must be emphasized:

(i) The complete lattice $\tilde{\mathcal{F}}$ is still totally ordered, it is a chain;

(ii) None of the additional spaces $H_{\alpha+}$, $H_{\alpha-}$ is a Hilbert space; they are, respectively, (nontrivial) inductive and projective limits of infinitely many Hilbert spaces.

Remarks 3.1: (a) The same construction works if $H_1, H_{\tilde{1}}$ are reflexive Banach spaces, dual of each other,[2] but not for more general spaces. For instance, Girardeau[14] has shown that no normed space exists between Schwartz spaces $\mathscr{S}$ and $\mathscr{S}'$, such that the interpolation property (iii) of Sec. II holds for every operator T.

(b) On the example discussed so far the intrinsic character of the spaces H_α stated in Sec. II.(ii) can be seen very explicitly. For instance, the multiplication operators by $(1+x^2)$ and $(1+|x|+x^2)$ have the same form domain H_1 and their form graph norms $\langle f, Af\rangle$ are different, but equivalent. Thus they both define the same intermediate spaces H_α, with different but equivalent norms.

Now we can go further.

Lemma 3.2: The PIP-space $H_{\tilde{I}}$ admits infinitely many proper refinements of type (H).

Proof: The statement follows from a straightforward application of the spectral theorem to the self-adjoint operator A which interpolates between H_1 and H_0. One has indeed, for each $0\leqslant\alpha\leqslant 1$,

$$H_\alpha \equiv Q(A^\alpha) = \left\{ f\in H_0 \,\middle|\, \int_1^\infty s^\alpha d(f,E(s)f) < \infty \right\}$$

with inner product

$$\langle f,g\rangle_\alpha = \langle f, A^\alpha g\rangle, \quad f,g\in H_\alpha.$$

Now let φ be any continuous, positive function on $[1,\infty)$ such that $\varphi(t)$ is unbounded for $t\to\infty$, but increases more slowly than any power t^α $(0<\alpha\leqslant 1)$. An example is $\varphi(t)=\log t\,(t\geqslant 1)$. Then $\varphi(A)$ is a well-defined self-adjoint operator, with form domain

$$Q(\varphi(A)) = \left\{ f\in H_0 \,\middle|\, \int_1^\infty (1+\varphi(s))d(f,E(s)f) < \infty \right\}.$$

With the corresponding inner product

$$\langle f,g\rangle_\varphi = \langle f,g\rangle + \langle f,\varphi(A)g\rangle$$

$Q(\varphi(A))$ becomes a Hilbert space H_φ. For every α, $0<\alpha\leqslant 1$, one has, with proper inclusions and continuous embeddings,

$$H_\alpha \subset H_\varphi \subset H_0. \tag{3.3}$$

Taking duals, one gets thus a new chain indexed by $\tilde{\tilde{I}} = \tilde{I}\cup\{\varphi,\bar{\varphi}\}$, where $H_{\bar{\varphi}}$ is the antidual of H_φ. One has, in fact, with proper inclusions

$$H_\alpha \subset H_{0+} \equiv \bigcup_{\alpha>0} H_\alpha \subset H_\varphi \subset H_0 \subset H_{\bar{\varphi}} \subset H_{0-} \equiv \bigcap_{\alpha<0} H_\alpha \subset H_{\bar{\alpha}}.$$

Indeed, H_φ, $H_{\bar{\varphi}}$ being Hilbert spaces, they cannot coincide with $H_{0\pm}$, which are not. Hence $H_{\tilde{I}}$ is a proper refinement of $H_{\tilde{I}}$. One can go further by interpolating between H_φ and H_0 and then iterating the construction. In the same fashion, additional spaces can be inserted between any fixed H_{α_0} and all the H_β, $\beta>\alpha_0$. Clearly this process can be repeated indefinitely with different functions φ, which proves the assertion.□

It is interesting to observe that, in questions related to the limiting absorption principle of scattering theory, Agmon and Hörmander[15] have also used (general) interpolation theory for refining the scale of weighted L^2-spaces $\{L^2_s(\mathbb{R}^n)\}$. More precisely they consider a space $\mathscr{B}$ such that

$$\bigcup_{s>1/2} L^2_s \equiv L^2_{1/2+} \subset \mathscr{B} \subset L^2_{1/2}.$$

$\mathscr{B}$ is the inverse Fourier transform of the so-called Besov space $B^{1/2}_{2,1}$, which is a nonreflexive Banach space.[13] The elementary Hilbert space method just described, of course, cannot yield such a space (although it does yield $B^{1/2}_{2,2}\equiv L^2_{1/2}$), but it allows one to construct many Hilbert spaces sitting between $L^2_{1/2+}$ and $L^2_{1/2}$, which could possibly be used for the same purposes as $\mathscr{B}$.

Let us turn now to an infinite discrete chain H_I [1-3]

$$\cdots\subset H_2\subset H_1\subset H_0\subset H_{\tilde{1}}\subset H_{\tilde{2}}\subset\cdots. \tag{3.4}$$

As before, there exists for each $j=1,2,\ldots$ a self-adjoint operator $A_j>1$ on H_0 such that $H_j = Q(A_j)$ and $\langle f,f\rangle_j = \langle f, A_j f\rangle$. Then the chain (3.4) is called a *scale* if

$$A_j = (A_1)^j, \quad j=1,2,\cdots. \tag{3.5}$$

Remark: If we write $A_1 = 1+\hat{A}_1$, with $\hat{A}_1>0$, then standard estimates show that $(A_1)^j = (1+\hat{A}_1)^j$ and $1+(\hat{A}_1)^j$ define equivalent norms on H_j, i.e., $\langle f,f\rangle_j = \langle f, A_j f\rangle$ is equivalent to the form graph norm $\langle f,f\rangle + \langle f,(\hat{A}_1)^j f\rangle$ of $(\hat{A}_1)^j$.

We assume now H_I to be a discrete scale ($I\subseteq\mathbb{Z}$). Then quadratic interpolation yields a unique continuous scale of Hilbert spaces $H_{\tilde{I}}$, with $\tilde{I}=\mathbb{R}$, which is obviously a proper refinement of H_I. Again the corresponding index set $\tilde{F}$ is the "nonstandard" real line described in III,[5] it is still totally ordered, and none of the additional spaces is a Hilbert space. Then we proceed exactly as before. For a suitable function φ, we consider the self-adjoint operator $\varphi(A_1)$ and the scale

$$H_{\varphi,k} \equiv Q([\varphi(A_1)]^k), \quad k\in\mathbb{J}^+ \quad (\mathbb{J}=\mathbb{R} \text{ or } \mathbb{Z})$$

with norms

$$\langle f,f\rangle_{\varphi,k} = \int [1+\varphi(s)]^k d(f,E(s)f).$$

Here k is positive and may take either integer values ($\mathbb{J}=\mathbb{Z}$) or arbitrary positive values ($\mathbb{J}=\mathbb{R}$). This yields a new chain, indexed by $\tilde{I}\equiv I\cup\mathbb{J}^+$, with proper inclusions, valid for every $0<\alpha<1$ and every $k>0$:

$$\cdots\subset H_1\subset H_\alpha\subset H_{\varphi,k}\subset H_0\subset H_{\frac{1}{\varphi,k}}\subset H_{\bar{\alpha}}\subset H_{\tilde{1}}\subset\cdots.$$

The same procedure may be repeated for each interval $[n,n-1]$ or, more generally, $[\alpha,\beta]$ with $\alpha>\beta$. It can also be iterated or applied with a different function φ. Thus we have proved the following general result:

Proposition 3.3: Every scale of Hilbert spaces possesses infinitely many proper refinements which are themselves chains of Hilbert spaces.□

The standard example is, of course, again obtained with $A_1 =$ multiplication by $(1+x^2)$ on $H_0 = L^2(\mathbb{R})$, and $\varphi(s) = \log s\,(s\geqslant 1)$.

It follows from this result that no chain of Hilbert spaces can be maximal; it can always be refined into another such chain. Yet maximal compatibility relations do exist, as a consequence of Zorn's lemma (see III). What Proposition 3.3 tells us, then, is that the corresponding PIP-spaces have of necessity a more complicated structure than a chain of

Hilbert spaces. This result falls in line with the general philosophy behind PIP-spaces: The choice of a given (indexed) PIP-space is dictated by the problem at hand, but never uniquely; there always remains the freedom of refining a given structure in order to answer more precise questions.

IV. REFINEMENT OF GENERAL CHAINS OF HILBERT SPACES

We start again with a finite chain:

$$H_2 \subset H_1 \subset H_0 \subset H_{\bar{1}} \subset H_{\bar{2}}. \tag{4.1}$$

For $j = 1,2$ there exists a unique self-adjoint operator $A_j > 1$ in H_0 such that $H_j = Q(A_j)$, but we assume $A_2 \neq (A_1)^2$, i.e., the chain (4.1) is *not* a scale. First we can easily characterize such chains.

Lemma 4.1:

$H_2 \subset H_1$ iff $A_1^{1/2} A_2^{-1/2}$ is bounded.

Proof: (a) Let $A_1^{1/2} A_2^{-1/2}$ be bounded, Since $A_2 > 1$, $A_2^{-1/2}$is bounded and $H_2 = D(A_2^{1/2}) = A_2^{-1/2} H_0$. Since $A_1^{1/2} A_2^{-1/2} f$ is well defined for every $f \in H_0$, it follows that $D(A_2^{1/2}) \subset D(A_1^{1/2}) = H_1$.

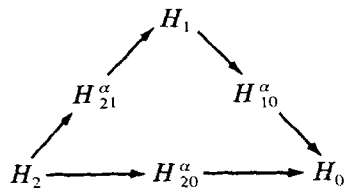
(b) As for the converse, notice that $A_1^{1/2}$ and $A_2^{-1/2}$ are closed operators, $A_1^{-1/2}$ is bounded, and hence $A_1^{1/2} A_2^{-1/2}$ is closed. Now, if $H_2 \subset H_1$, $A_1^{1/2} A_2^{-1/2}$ is defined on all of H_0 and therefore it is a bounded operator.□

Notice that the embedding $H_2 \to H_1$ is continuous, under the same conditions, if H_j, $j = 1,2$, is equipped, as usual, with the norm $\langle f, f \rangle_j = \langle f, A_j f \rangle$.

Let now B be the interpolating operator between H_2 and H_1. B has the form domain H_2, is self-adjoint in H_1 (but not necessarily in H_0), and is larger than 1. It verifies the relations

$$\langle f, f \rangle_2 = \langle f, Bf \rangle_1 = \langle B^{1/2} f, B^{1/2} f \rangle_1 \quad (f \in H_1).$$

By the uniqueness of the interpolation operator, one has $A_2 = B^{1/2*} A_1 B^{1/2}$, where $B^{1/2*}$ is the adjoint of $B^{1/2}$ in H_0. Thus we can interpolate between H_2 and H_1 with B, between H_1 and H_0 with A_1, and also directly between H_2 and H_0 with A_2. Schematically



where

$$\left.\begin{aligned} H_{20}^{\alpha} &= Q(A_2^{\alpha}) \quad \text{in } H_0 \\ H_{10}^{\alpha} &= Q(A_1^{\alpha}) \quad \text{in } H_0 \\ H_{21}^{\alpha} &= Q(B^{\alpha}) \quad \text{in } H_1 \end{aligned}\right\} \quad (0 \leqslant \alpha \leqslant 1).$$

So the question arises, how do the two chains $\{H_{21}^{\alpha}, H_{10}^{\beta}\}$ and $\{H_{20}^{\alpha}\}$, both between H_2 and H_0, compare to each other?

We have to distinguish two cases:

(1) $H_1 = H_{20}^{\gamma} \equiv Q(A_2^{\gamma})$ *for some* γ $(0 < \gamma < 1)$. Equality here means equality as vector spaces with equivalent norms (comparable norms would be sufficient, of course). By property (ii) of Sec. II, we can replace the norm of H_1 by that of $Q(A_2^{\gamma})$ without changing any of the interpolation spaces $H_{21}^{\alpha}, H_{10}^{\beta}$. This amounts to writing $A_1 = A_2^{\gamma}, B = A_2^{1-\gamma}$. In that case, there is only one continuous interpolating scale between H_2 and H_0 as follows from the interpolation property (iii) of Sec. II applied to the identity operator.

(2) *There is no* γ *such that* $H_1 = Q(A_2^{\gamma})$. In that case, the three scales $\{H_{21}^{\alpha}\}$, $\{H_{10}^{\alpha}\}$, $\{H_{20}^{\gamma}\}$ generate, by intersection and vector sum, a genuine lattice, i.e., *not* a chain. First there are some inclusion relations.

Lemma 4.2: With the notations as above, the following inclusions hold, where all the injections are continuous and have dense range:

(i) For any pair α, β, $H_{21}^{\alpha} \subseteq H_{10}^{\beta}$.

(ii) For every fixed α, $0 \leqslant \alpha \leqslant 1$: $H_{21}^{\alpha} \subseteq H_{20}^{\alpha} \subseteq H_{10}^{\alpha}$.

Proof: Part (i) is obvious. The inclusions (ii) follow from repeated application of the interpolation property to the identity operator. Indeed let (X_1, X_0) and (Y_1, Y_0) be two pairs of Hilbert spaces as defined in Sec. II, with X_0 and Y_0 subspaces of the same vector space. Then one has, for every $0 \leqslant t \leqslant 1$, and with continuous injection:

$$Q_t(X_1 \cap Y_1, X_0 \cap Y_0) \subseteq X_t \cap Y_t, \tag{4.2}$$

here, as usual, $E \cap F$ is the intersection of E and F, equipped with the projective topology. The two inclusions in (ii) follow by taking, respectively, the pairs $(H_2, H_1), (H_2, H_0)$ and $(H_2, H_0), (H_1, H_0)$.□

Let us now build the lattice, taking Lemma 4.2 into account. The first "generation" consists of all the spaces $H_{20}^{\alpha}, H_{21}^{\beta}, H_{10}^{\gamma}$. At the second generation, new spaces appear: $H_{20}^{\alpha} \cap H_{21}^{\beta}$ and $H_{20}^{\alpha} + H_{21}^{\beta}$ $(\alpha > \beta)$, $H_{20}^{\alpha} \cap H_{10}^{\gamma}$ and $H_{20}^{\alpha} + H_{10}^{\gamma}$ $(\alpha < \gamma)$. The same happens at the third generation, e.g., for $\alpha > \beta > \gamma$: $H_{10}^{\alpha} \cap H_{20}^{\beta} \cap H_{21}^{\gamma}$ and $H_{10}^{\alpha} + H_{20}^{\beta} + H_{21}^{\gamma}$ or $(H_{10}^{\alpha} + H_{20}^{\beta}) \cap H_{21}^{\gamma} = H_{10}^{\alpha} \cap (H_{20}^{\beta} + H_{21}^{\gamma})$. The last two spaces coincide because the lattice $\mathscr{L}(H_0)$ of all vector subspaces of H_0 is modular. In the same way, new spaces will appear at each successive generation. For instance, let $\{\alpha_n\}$, $\{\beta_n\}$ be two infinite sequences such that

$$0 < \cdots < \alpha_3 < \alpha_2 < \alpha_1 < \beta_1 < \beta_2 < \beta_3 < \cdots < 1,$$

and let $X^{(n)}$ be the following element of the lattice:

$$X^{(n)} = (H_{20}^{\alpha_1} + H_{10}^{\beta_1}) \cap (H_{20}^{\alpha_2} + H_{10}^{\beta_2}) \cap \cdots \cap (H_{20}^{\alpha_n} + H_{10}^{\beta_n}).$$

Then the sequence $X^{(1)}, X^{(2)}, X^{(3)}, \cdots$ is an infinite, strictly decreasing sequence of subspaces. As a consequence, we have the following result.

Proposition 4.3: Let H_I be the following discrete finite chain of Hilbert spaces:

$$H_2 \subset H_1 \subset H_0 \subset H_{\bar{1}} \subset H_{\bar{2}}.$$

Let A_j $(j = 1,2)$ be the interpolating operator between H_j and H_0. Let $H_{\tilde{I}}$ be the lattice generated from H_I by quadratic interpolation. Then:

(i) If for some γ, $0 < \gamma < 1$, $Q(A_2^{\gamma}) = H_1$, with equivalent norms, then $H_{\tilde{I}}$ is a continuous scale of Hilbert spaces.

(ii) If no such γ exists, $H_{\tilde{I}}$ is a continuous involutive lattice, but not a chain.

In both cases $H_{\tilde{I}}$ is a proper refinement of H_I.

Proof: For the part $H_2 \subset H_1 \subset H_0$, the alternative results from the discussion above. The same result holds for

$H_0 \subset H_{\bar{1}} \subset H_{\bar{2}}$ by the duality property of quadratic interpolation, Sec. II(iv), and the fact that, as in any PIP-space of type (H), $H_\alpha \cap H_\beta$ and $H_{\bar{\alpha}} + H_{\bar{\beta}}$ are dual of each other. The last statement is obvious.□

This construction shows clearly that quadratic forms are more natural than operators in a PIP-space context. For instance, one has:

$$H_{20}^{\alpha} \cap H_{10}^{\beta} \equiv Q(A_2^{\alpha}) \cap Q(A_1^{\beta}) = Q(A_2^{\alpha} \dot{+} A_1^{\beta}),$$

i.e., the form domain of the *form sum* of the two operators.

Remark: This is true only for the interpolating operators A_j, which are considered here as unbounded operators in H_0, and not as operators in the PIP-space sense. For these, the PIP-space sum $A + B$ may be a proper extension of the form sum $A \dot{+} B$; this will happen, for instance, if $A + B = c\mathbb{1}$, which is defined on the whole PIP-space.

We conclude this discussion with an example. Let $H_0 = L^2(\mathbb{R}, dx)$, $A_1 = 1 + x^2$, $A_2 = 2 + x^2 + p^2$. Then $B = 1 + (1 + x^2)^{-1}(1 + p^2)$, which is indeed self-adjoint in $H_1 = L^2(\mathbb{R}, (1 + x^2)\,dx)$. Thus the two chains generated by (B, A_1) and A_2, respectively, have no element in common, and so generate a genuine lattice, with typical elements of the form

$$H_{20}^{\alpha} \cap H_{10}^{\beta} = Q(A_2^{\alpha} \dot{+} A_1^{\beta}) = Q((2 + x^2 + p^2)^{\alpha} \dot{+} (1 + x^2)^{\beta}).$$

The same argument applies *a fortiori* to an infinite chain; so we may state in general:

Proposition 4.4: Let H_I be a discrete chain of Hilbert spaces $\{H_n, n \in I \subseteq \mathbb{Z}\}$. Then quadratic interpolation between H_j $(j > 0)$ and H_0 generates a PIP-space $H_{\tilde{I}}$ of type (H) with the following structure:

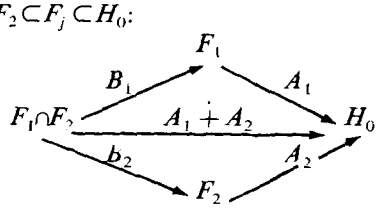
$$V^{\#} = \bigcap_{n \in I} H_n \subset \{H_\alpha\} \subset H_0 \subset \{H_{\bar{\alpha}}\} \subset V = \bigcup_{n \in I} H_n,$$

where both $\{H_\alpha\}$ and $\{H_{\bar{\alpha}}\}$ are infinite continuous lattices. These two lattices are chains iff, for every $j = 1,2,3,\cdots$, there exist γ_j $(0 < \gamma_j < 1)$ such that $Q(A_j^{\gamma_j}) = H_1$, with equivalent norms, where A_j is the interpolating operator between H_j and H_0 In any case, $H_{\tilde{I}}$ is a proper refinement of H_I.□

Combining this result with the construction of Sec. III we see that every chain of Hilbert spaces actually possesses arbitrarily many proper refinements by nested Hilbert spaces, which are in general no longer chains. One could go further by interpolating between arbitrary pairs (H_k, H_l), $k,l \in I$, but this will only complicate the situation further without changing the result.

V. REFINEMENT OF A LATTICE OF HILBERT SPACES

As a prelude to a general lattice, we consider first two noncomparable subspaces F_1, F_2 of H_0. We denote by A_j, B_j $(j = 1,2)$ the interpolating operators in the triplet $F_1 \cap F_2 \subset F_j \subset H_0$:



(5.1)

Then the analysis of Sec. IV gives immediately the following results:

(i) $B_1^{1/2*} A_1 B_1^{1/2} = B_2^{1/2*} A_2 B_2^{1/2} = A_1 \dot{+} A_2$. This operator is self-adjoint, with form domain $F_1 \cap F_2$, and defines an interpolating scale $\{Q((A_1 \dot{+} A_2)^{\alpha})\}$ between $F_1 \cap F_2$ and $H_0 (0 \leqslant \alpha \leqslant 1)$.

(ii) Let $\{Q(A_j^{\alpha})\}$ be the scale interpolating between F_j and H_0. Then these two scales generate again, by intersection and vector sum, a genuine lattice, and one has

$$Q(A_1^{\alpha}) \cap Q(A_2^{\beta}) = Q(A_1^{\alpha} \dot{+} A_2^{\beta}).$$

(iii) The spaces defined in (i) and (ii) are related by the following inclusion relations, which follow from the interpolation property[4]:

$$Q((A_1 \dot{+} A_2)^{\alpha}) \subseteq Q(A_1^{\alpha}) \cap Q(A_2^{\alpha}), \quad 0 \leqslant \alpha \leqslant 1. \tag{5.2}$$

Equality holds if A_1 and A_2 commute.

The classical example of this type of structure is given by $H_0 = L^2(\mathbb{R}^3, d\mathbf{x})$, $A_1 = (1 + |\mathbf{x}|^2)$, $A_2 = (1 + |\mathbf{p}|^2)$, which gives $B_1 = 1 + (1 + |\mathbf{x}|^2)^{-1}(1 + |\mathbf{p}|^2)$ and $B_2 = 1 + (1 + |\mathbf{p}|^2)^{-1}(1 + |\mathbf{x}|^2)$. Notice that $A_1^{1/2} A_2^{-1/2}$ and $A_2^{1/2} A_1^{-1/2}$ are both unbounded, so that, by Lemma 4.1, F_1 and F_2 are indeed noncomparable. This example, or variants thereof, constitutes an essential tool for scattering theory. It contains the Sobolev spaces (the scale built on A_2), the weighted spaces L^2_s (the scale built on A_1), and spaces of mixed type. We refer to Reed and Simon's treatise[16] for further details and specific applications, such as the derivation of trace ideal properties of operators of the form $f(\mathbf{x})g(\mathbf{p})$, or the proof of absence of singular continuous spectrum by the limiting absorption principle.

A situation reminiscent of the one just described may be found in a paper by Hunziker.[17] In the discussion above, the operator A_2 is essentially the nonrelativistic free Hamiltonian; for a general scattering system, Hunziker uses the lattice built instead on the powers of $\mathbf{x}$ and H, the full Hamiltonian. Since $\mathbf{x}$, A_2, H in general do not commute, this scattering theory indeed takes place in an infinite lattice as described in Sec. IV.

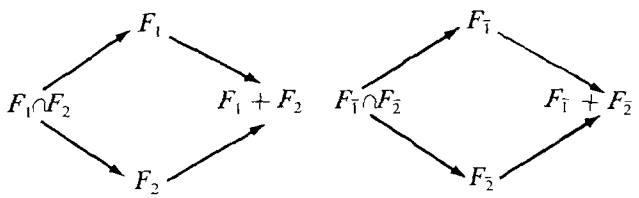
The standard examples of commuting operators are those that depend on different variables. Take again $H_0 = L^2(\mathbb{R}^3, d\mathbf{x})$ with $A_j = (1 + x_j^2)$, $j = 1,2,3$. Each of these generates a scale $Q(A_j^{\alpha})$ and those generate the lattice. Then Eq. (5.2) gives, for each $0 \leqslant \alpha \leqslant 1$,

$$Q((A_1 \dot{+} A_2 \dot{+} A_3)^{\alpha}) = \bigcap_{j=1}^{3} Q(A_j^{\alpha}),$$

i.e.,

$$L^2(\mathbb{R}^3, (1 + |\mathbf{x}|^2)^{\alpha}\,d\mathbf{x}) = \bigcap_{j=1}^{3} L^2(\mathbb{R}^3, (1 + x_j^2)^{\alpha}\,d\mathbf{x}).$$

From here on the generalization is obvious. Let H_I be a countable NHS, as in Sec. I. Then an arbitrary pair F_1, F_2 of assaying subspaces generates by quadratic interpolation a lattice as described above, and similarly for the duals $F_{\bar{1}}, F_{\bar{2}}$, according to the scheme



We may repeat the same operation for arbitrarily many pairs of assaying subspaces. Therefore we may conclude:

Theorem 5.1: Every countable nested Hilbert space admits infinitely many proper refinements of the same type.□

This result underlines once again the versatility of the PIP-space approach. The choice of a given structure is to a large extent a matter of taste and commodity, even for a given problem. For instance, the example discussed above leads to a whole class of NHS built around the Schwartz triplet $\mathscr{S} \subset L^2 \subset \mathscr{S}'$; among them one finds scales, discrete or continuous, such as $\{Q((1+p^2+x^2)^\alpha), \alpha \in \mathbb{Z} \text{ or } \mathbb{R}$, as well as various lattices built on powers of $(1+x^2)$ and of $(1+p^2)$. It should be noticed, however, that for a given Fréchet space $V^\# = \Phi$, as discussed in our previous work,[6] there does not always exist a *scale* interpolating between $V^\#$ and $V = \Phi^\times$. It may be necessary to use the powers of several self-adjoint operators for reproducing the topology of Φ. Lassner and Timmermann[18] have exhibited counterexamples. Thus the general case discussed in this section is necessary, even for countably Hilbert spaces.[6]

VI. CONCLUDING REMARKS

A. Enlargement of a PIP-space

The discussion above suggests several ways of enlarging a PIP-space V, that is, to build on a larger space $V_1 \supset V$ a PIP-space structure which is a refinement of the original one. For instance:

(i) Starting from a triplet $H_1 \subset H_0 \subset H_{\bar{1}}$ (or a finite scale), the self-adjoint operator A interpolated between H_1 and H_0 defines the full scale $\{H_n = Q(A^n), H_{\bar{n}} = (H_n)^\times, n = 1,2,3,\ldots\}$ thus getting $V^\# = \cap_{n\in\mathbb{N}} H_n \equiv Q^\infty(A)$, $V = \cup_{n\in\mathbb{N}} H_n$, with the natural PIP-space (NHS) structure.

(ii) The same construction can be performed in the case of two noncomparable subspaces F_1, F_2. The resulting NHS H_I is not a chain. Notice that $Q^\infty(A_1) \cap Q^\infty(A_2)$ is dense in H_0 since it contains $V^\# = \cap_{r\in I} H_r$.

(iii) There is no need to restrict oneself to powers of A, larger classes of functions can be used, such as exponentials $\exp(tA)$ or $\exp(tA^k)$, more generally powers $[f(A)]^t$, provided $f(A)$ is a self-adjoint operator larger than 1. An example is given by the Hilbert spaces of type S introduced by Grossmann,[19] which are generated by entire functions of (p^2+x^2). One could go further and study spaces generated by $*$-algebras of unbounded operators.[20]

B. Construction of nested Hilbert spaces

The lesson of this paper is that NHS's are intimately connected with positive self-adjoint operators and positive quadratic forms. This suggests another way of building an NHS. In any NHS H_I we consider an assaying subspace H_r, not comparable to H_0:

Let $A = 1 + R$ be the interpolating operator between $H_{r\wedge 0} = H_r \cap H_0$ and H_0:

$$\langle f,f\rangle_{r\wedge 0} = \langle f,f\rangle + \langle f,f\rangle_r = \langle f,(1+R)f\rangle. \quad (6.1)$$

By construction, R is a nonnegative self-adjoint operator, with form domain $Q(R) = H_r \cap H_0$. The space H_r is the completion of $H_r \cap H_0$ in the norm

$$\langle f,f\rangle_r = \langle f, Rf\rangle, \quad f \in H_r \cap H_0 \quad (6.2)$$

Thus $R^{1/2}$ is isometric from $H_r \cap H_0$ into H_0, so its closure $\overline{R^{1/2}}$ is unitary from H_r onto H_0. Since $\langle\cdot,\cdot\rangle$ is a norm, the operator R is invertible, i.e., 0 does not belong to its point spectrum. Indeed, the dual space $H_{\bar{r}} = (H_r)^\times$ corresponds precisely to the operator R^{-1}, and again $\overline{R^{-1/2}}$ is unitary from $H_{\bar{r}}$ onto H_0. Two cases may happen:

(i) R and R^{-1} are both unbounded (in H_0); then H_r and H_0, $H_{\bar{r}}$ and H_0 are mutually noncomparable. Take, for example, $H_0 = L^2(\mathbb{R},dx)$, $A = 1 + x^2$, then $R = x^2$, $H_r = L^2(\mathbb{R},x^2\,dx)$, $H_r \cap H_0 = L^2(\mathbb{R},(1+x^2)\,dx)$, $R^{-1} = x^{-2}$, $H_{\bar{r}} = L^2(\mathbb{R},x^{-2}\,dx)$.

(ii) One of them, say R^{-1}, is bounded. Then one has a triplet $H_r \subset H_0 \subset H_{\bar{r}}$, as discussed in Sec. III.

Of course, if both R and R^{-1} are bounded, the norms $\langle\cdot,\cdot\rangle_r$ and $\langle\cdot,\cdot\rangle$ are equivalent, i.e., $H_r = H_0 = H_{\bar{r}}$. We notice also that the norms $\langle\cdot,\cdot\rangle_r$ and $\langle\cdot,\cdot\rangle$ are automatically consistent on $H_r \cap H_0$. Finally, all that has been said can be repeated in terms of the closed, positive, nondegenerate quadratic form $r(f,g) \equiv \langle f, Rg\rangle$, with $Q(r) = Q(R)$.

These remarks suggest another method for constructing an NHS, starting from a given Hilbert space and a family of self-adjoint operators (or quadratic forms).[21] Let R_1, R_2 be two nonnegative invertible self-adjoint operators on H_0, with form domains $Q(R_1), Q(R_2)$. Let H_{R_j} $(j = 1,2)$ be the completion of $Q(R_j)$ with respect to the norm $\|f\|_j = \langle f, R_j f\rangle$. We say that R_1 and R_2 are *consistent* if:

(i) $Q(R_1) \cap Q(R_2)$ is dense in both H_{R_1} and H_{R_2},

(ii) the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ are consistent on $Q(R_1) \cap Q(R_2)$.

Let I be a family of nonnegative, invertible, self-adjoint operators on H_0. We say that I is *admissible* if one has:

(1) Any two operators of I are consistent;

(2) The inverse of any element of I belongs to I;

(3) The sum of any two elements of I belongs to I.

Given an admissible family I, denote by $\mathscr{S} = \{H_r\}_{R\in I}$ the corresponding family of Hilbert spaces. We may, of course, always assume that $H_0 \in \mathscr{S}$, that is $1 \in I$. Let $V = \Sigma_{R\in I} H_R$ be the algebraic inductive limit of $\mathscr{S}$.[7] It is clear that $\mathscr{S}$ is an involutive covering of V (see III), and this defines a linear compatibility on it, by the relation $f \# g \Leftrightarrow \exists R \in I$ such that $f \in H_R$, $g \in H_{\bar{R}} \equiv H_{R^{-1}}$, with the partial inner product

$$\langle f | g\rangle = \langle \overline{R^{1/2}} f | \overline{R^{-1/2}} g\rangle.$$

We therefore obtain a PIP-space of type (H) on V, and in fact an NHS. It follows that $V^\# = \cap_{R\in I} H_R$ is dense in every H_R and is a core for each $R \in I$. We thus state

Proposition 6.1: Any admissible family I of nonnegative,

invertible, self-adjoint operators on a Hilbert space H_0 defines an NHS on $V = \Sigma_{R\in I} H_R$, with central Hilbert space H_0.□

An entirely similar construction is obtained by starting from an admissible family of closed, nonnegative, nondegenerate, quadratic forms, the connection between the two approaches being, of course,

$$r(f,g) = \langle f, Rg\rangle, \quad f,g\in Q(r)\equiv Q(R).$$

As a standard example of this construction we may take the NHS of locally integrable functions described in III and IV.[5] Let $H_0 = L^2(\mathbb{R}^n, d\mathbf{x})$ and define R to be the operator of multiplication by the measurable, a.e. positive function r. This gives

$$H_r = \left\{ f:\mathbb{R}^n\to\mathbb{C},\ \text{measurable},\ \int |f|^2 r\, d\mathbf{x} < \infty \right\}.$$

If r and r^{-1} are both locally integrable, $Q(R^{\pm 1})$ contains the space L^{∞}_{comp} of essentially bounded functions of compact support, which is dense in H_r and $H_{r^{-1}}$. Thus if we take all such functions r, with $r^{\pm 1}\in L^1_{\text{loc}}$, we get an admissible family, which yields the usual PIP-space on $V = L^1_{\text{loc}}$. Similarly, if one takes instead functions r such that $r^{\pm 1}\in L^{\infty}_{\text{loc}}$, one gets a PIP-space on $V = L^2_{\text{loc}}$, with $V^{\#} = L^2_{\text{comp}}$.

C. Algebras of operators

We consider again a discrete *scale* of Hilbert spaces $\{H_n, n\in\mathbb{Z}\}$ or any symmetric portion H_I ($I\subseteq\mathbb{Z}$) of that scale, in particular, the pair $\{H_n, H_{\bar{n}}\}$. A useful object for the study of operators in H_I is the $*$-algebra $\mathscr{A}(H_I)$, introduced by Debacker-Mathot[22]; it consists of all the operators that map every H_n, $n\in I$, continuously into itself. Let $H_{\tilde{I}}$ be the continuous scale obtained from H_I by quadratic interpolation. Then the interpolation property [Sec. II(iii)] implies that every operator $T\in\mathscr{A}(H_I)$ maps every H_α, $\alpha\in\tilde{I}$, also continuously into itself. In other words, the algebras $\mathscr{A}(H_I)$ and $\mathscr{A}(H_{\tilde{I}})$ are isomorphic. In particular, they have the same idempotent elements, i.e., orthogonal projections: $\text{Proj}(H_I) \simeq \text{Proj}(H_{\tilde{I}})$. This in turn implies[23] that H_I and $H_{\tilde{I}}$ have the same orthocomplemented subspaces (this was known so far only for the finite dimensional subspaces[23]).

Remark: Let P be a projection in H_I. This means, in particular, that $PH_1\subset H_1$ or $A^{1/2}PA^{-1/2}$ is bounded in H_0. This is certainly true if P commutes with A (then P is absolute[23]), but that condition is not necessary. Take, for instance, a rank-1 projection on H_1, $P = |\psi\rangle\langle\psi|$ with $\psi\in H_1$, i.e., $\psi = A^{-1/2}\varphi$ for a unique $\varphi\in H_0$. Then $A^{1/2}PA^{-1/2} = |\varphi\rangle\langle\varphi|A^{-1}$, which is obviously bounded in H_0. On the other hand, P does not necessarily commute with A. Indeed, $AP = |A\psi\rangle\langle\psi|$, $PA = |\psi\rangle\langle A\psi|$. So if χ is such that $\langle\psi|\chi\rangle = 0$, but $\langle A\psi|\chi\rangle\neq 0$ (this is always possible unless ψ is an eigenvector of A, in which case $AP = PA$), then $AP = 0$, but $PA\neq 0$. In other words, H_I contains nondiagonal nonabsolute projections.

D. Connection with rosette spaces

There is an interesting connection between the spaces described in this paper and the *rosette spaces* introduced by Rzewuski.[24] Let $V^{\#}\subset H\subset V$ be a PIP-space with central Hilbert space, and G be a family (in particular, a group) of linear transformations on H which do not leave $V^{\#}$ invariant; for each $g\in G$, let $V_g^{\#}\equiv g(V^{\#})$ determine another PIP-space $\{V_g, \#_g, \langle\cdot,\cdot\rangle_g\}$ with $V_g^{\#}\subset H\subset V_g$ such that the self-compatible elements and their norm are independent of $g\in G$, i.e., the central Hilbert space H is the same for all $g\in G$. The collection of these spaces, for all $g\in G$, is called a rosette space. An interesting example is that of sequence spaces. Start with $\varphi\subset\ell^2\subset\omega$ corresponding to a given orthonormal basis $\{e_n\}$ in H (φ is the space of finite sequences, ω the space of arbitrary sequences). Let G be the group of all unitary transformations of H, with indexes the collection of all orthonormal bases in H. For any $g\in G$, $\varphi_g\subset\ell^2\subset\omega_g$ is a similar PIP-space, corresponding to the basis $\{ge_n\}$. It is interesting to notice that a very similar construction was used by Friedrichs and Shapiro[25] as a means of extending the Gaussian measures on H to a totally additive measure on a larger space $W\supset H$, called a *corona*. W is a vector space consisting of equivalence classes of elements of $\cup_i \omega_{g_i}$, for a countable family $\{g_i e_n\}$ of bases in H

Let us now come back to our standard example $H = L^2(\mathbb{R}, dx)$, $A_1 = 1 + x^2$, $A_2 = 1 + p^2$. We build the two corresponding scales $H_{I_j} = \{Q(A_j^n), n\in\mathbb{Z}\}$. Then we notice that Fourier transform $\mathscr{F}$ is unitary from $Q(A_1^n)$ onto $Q(A_2^n)$ for every $n\in\mathbb{Z}$. Hence the union $H_{I_1}\cup H_{I_2}$ is a rosette space with central space $H = L^2(\mathbb{R}, dx)$ and group $G = \{\mathbb{1}, \mathscr{F}\}$. We may conclude that our procedure of "filling in" the lattice generated by the two scales H_{I_j}, $j = 1,2$, could be applied to any rosette space, at least in the case of a countable G.

ACKNOWLEDGMENTS

W.K. acknowledges gratefully the hospitality of the Institut de Physique Théorique, Université Catholique de Louvain, where this work was begun, as well as financial support of the Administration des Relations Culturelles (Belgium). J.-P.A. thanks V. Enss for an interesting conversation. Both of us acknowledge gratefully the hospitality of the Centre de Physique Théorique, C.N.R.S., Luminy, Marseille, where this work was completed. We also wish to thank Alex Grossmann for interesting discussions and the communication of his unpublished results (see Sec. VI B).

[1] R. Palais, "Seminar on the Atiyah-Singer Index Theorem," *Annals of Mathematics Studies* (Princeton U. P., Princeton, N.J., 1965), Chaps. VII and XIV.

[2] S. G. Krein and Ju. I. Petunin, "Scales of Banach Spaces" (Russian), Usp. Mat. Nauk. **21**, 89–168 (1966) [Russian Math. Surveys **21**, 85–159 (1966)].

[3] Ju. M. Berezanskii, *Self-Adjoint Operators in Spaces of Functions of Infinitely Many Variables* (Russian) (Naukova Dumka, Kiev, 1978).

[4] J. L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications* (Dunod, Paris, 1968), Vol. 1.

[5] J.-P. Antoine and A. Grossmann, "Partial Inner Product Spaces. I. General Properties." J. Funct. Anal. **23**, 369–78 (1976); "II. Operators," *ibid.* **23**, 379–91 (1976); J.-P. Antoine, "III. Compatibility relations revisited," J. Math. Phys. **21**, 268–79 (1980); " IV. Topological considerations," *ibid.* **21**, 2067–79 (1980).

[6] J.-P. Antoine and W. Karwowski, " Countably Hilbert Spaces and Partial Inner Product Spaces, " Bull. Acad. Pol. Sci., Sér. Sci. Phys. **27**, 223–32 (1979).

[12]A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, " A class of explicitly soluble, local, many-center Hamiltonians for one-particle quantum mechanics in two and three dimensions. I, " J. Math. Phys. **21**, 2376–85 (1980); A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, " The one particle theory of periodic point interactions. Polymers, monomolecular layers and crystals, " Commun. Math. Phys. **77**, 87–110 (1980).
[13]J. Bergh and J. Löfström, *Interpolation Spaces : An Introduction* (Springer-Verlag, Berlin, 1976).
[14]J. P. Girardeau, " Sur l'interpolation entre un espace localement convexe et son dual, " Rev. Fac. Ciencias Univ. Lisboa, **11**, 165–86 (1964–65).
[15]S. Agmon and L. Hörmander, " Asymptotic Properties of Solutions of Differential Equations with Simple Characteristics, " J. Anal. Math. (Jerusalem) **30**, 1–38 (1976);J. M. Combes, "Recent Developments in Quantum Scattering Theory, " in *Mathematical Problems in Theoretical Physics*, Springer Lecture Notes in Physics Vol. 116, edited by K. Osterwalder (Springer-Verlag, Berlin, 1980); J. M. Combes, in *Rigorous Atomic and Molecular Physics* (Proceedings, Erice, 1980), edited by G. Velo and A. S. Wightman (D. Reidel, Dordrecht, 1981) to appear.
[16]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. III. Scattering Theory* (Academic, New York, 1979).
[17]W. Hunziker, " On the space–time behavior of Schrödinger wavefunctions, " J. Math. Phys. **7**, 300–4 (1966).
[18]G. Lassner and W. Timmermann, " Classification of Domains of Operator Algebras, " Rep. Math. Phys. **1**, 205–17 (1976).
[19]A. Grossmann, " Hilbert Spaces of Type S, " J. Math. Phys. **6**, 54–67 (1965).
[20]G. Lassner, Topological Algebras of Operators, Rep. Math. Phys. **3**, 279–93 (1978); M. Friedrich and G. Lassner, " Rigged Hilbert Spaces and Topologies on Operator Algebras, " *ibid.* (1981), to appear; J.-P. Antoine and F. Mathot, " Partial inner product spaces generated by algebras of unbounded operators, " preprint UCL-IPT-81-10 (Louvain-la-Neuve), to be published .
[21]A. Grossmann, unpublished manuscript and private communication.
[22]F. Debacker-Mathot, " Some Operator Algebras in Nested Hilbert Spaces, " Commun. Math. Phys. **42**, 183–93 (1975).
[23]J.-P. Antoine and A. Grossmann, " Orthocomplemented subspaces of nondegenerate partial inner product spaces, " J. Math. Phys. **19**, 329–35 (1978).
[24]J. Rzewuski, " On Hilbert Structure of Certain Linear Spaces, " Bull. Acad. Pol. Sci., Sér. Sci. Math. Astron. Phys. **20**, 85–94 (1972); " Rosette Spaces. I, " *ibid.* **22**, 617–24 (1974).
[25]K. O. Friedrichs and H. N. Shapiro, " Integration over Hilbert Space and Outer Extensions, " Proc. Nat. Acad. Sci. USA **43**, 336–38 (1957); *Integration of Functionals* (Courant Inst. Math. Sci., New York, 1967), Chap. XII.

Modified Hamiltonian systems and canonical transformations arising from the relationship between generalized Zakharov–Shabat and energy-dependent Schrödinger operators [a)]

L. Martínez Alonso and F. Guil Guerrero
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 18 February 1981; accepted for publication 29 May 1981)

A new family of completely integrable infinite-dimensional Hamiltonian systems is found. Two canonical maps are deduced which transform this family into the families of Hamiltonian systems associated with the generalized Zakharov–Shabat and energy-dependent Schrödinger operators. Through these maps several relevant Hamiltonian structures are connected. In particular, a simple explanation of the second Hamiltonian structure for AKNS equations is given.

PACS numbers: 02.30. + g

I. INTRODUCTION

Some recent results confirm the relevance of Hamiltonian formalism in the context of the inverse-scattering transform method. Gel'fand and Dikii[1] have found a procedure for calculating Hamiltonian systems with an infinite set of constants of motion which applies to a wide class of spectral problems. It has been proved by Adler[2] that the corresponding Hamiltonian structures can be described in terms of the Kirillov's symplectic form associated with Lie algebras. Moreover, in the case of scalar spectral problems, the second operator arising in the Lenard relations for Gel'fand–Dikii evolution equations is also a Hamiltonian operator.[3] On the other hand, the method used by Gel'fand and Dikii can also be applied to Schrödinger spectral problems with energy-dependent potentials leading us to new families of Hamiltonian systems having two Hamiltonian structures.[4] One of both Hamiltonian structures turns out to be also a Kirillov's symplectic form.[5] In this way, we are suddenly faced with a lot of important infinite-dimensional Hamiltonian structures whose analysis is undoubtedly fundamental in order to clarify our understanding about the integrability properties of nonlinear evolution equations. Especially interesting is the search of transformations relating different families of these Hamiltonian systems since it could help us to isolate a set of simple Hamiltonian structures from which all other can be obtained by means of canonical transformations. In fact, the second Hamiltonian structure for Gel'fand–Dikii equations, which nobody has been able to write down explicitly in the general case, is the image of a simple Hamiltonian structure under a generalization of the Miura map.[3]

This paper is concerned with the relationship between the families of Hamiltonian systems associated with the generalized Zakharov–Shabat spectral problem[6,7] (Z) and the Schrödinger spectral problem (S) with a potential which depends linearly on the spectral parameter.[4,8] An implicit transformation relating both families of evolution equations was derived by Jaulent and Miodek[9] by using the inverse-scattering transforms of (Z) and (S). Nevertheless, the algebraic aspects of such a transformation remained to be considered. We adopt here the Hamiltonian point of view to describe the algebraic content of the Jaulent–Miodek map in terms of canonical transformations. In this way, we find that the origin of the Jaulent–Miodek map is the existence of a new family of Hamiltonian systems which transforms into the families associated with (S) and (Z) under the action of two canonical transformations M and N, respectively. The main properties we deduce are:

(1) The new family of Hamiltonian systems contains as particular cases the modified equations for higher Korteweg–de Vries equations and the family of Burgers equations.

(2) The map M is a Miura transformation.

(3) The map N, when acting over Burgers equations, reduces to the Hopf–Cole transformation.

(4) The composition MN^{-1} is the Jaulent–Miodek transformation.

Incidentally, our analysis provides the following result: The second operator arising in the Lenard relation for the generalized Zakharov–Shabat spectal problem is a Hamiltonian operator. In fact, it is the image under N of the operator ∂_x acting on two-component functions. This implies that the evolution equations associated with (Z) are bi-Hamiltonian systems, and, as a consequence of it, we prove also that the Jaulent–Miodek map is a bicanonical transformation.

The paper is organized as follows. Sec. II deals with the description of the Hamiltonian systems associated with (S) and (Z). It contains also the derivation of the family of modified Hamiltonian systems and the definition of the map M. In Sec. III the spectral problems (S) and (Z) are related, the map N is defined, and their properties are analyzed. Finally, Sec. IV is devoted to describing the Jaulent–Miodek transformation in terms of M and N.

II. HAMILTONIAN STRUCTURES. MODIFIED EQUATIONS

A. The energy-dependent Schrödinger spectral problem

We shall consider here the following Schrödinger spec-

[a)]Partially supported by the Junta de Energía Nuclear, Madrid.

 0022-2488/81/112497-07$01.00

tral problem with an energy-dependent potential

$$(S):\quad (-\partial_{xx} + v_1(x) + kv_2(x) - k^2)f = 0. \tag{2.1}$$

It was found by Jaulent and Miodek[8] that an inverse-scattering transform for (S) exists which allows one to solve the family of nonlinear evolution equations

$$(\partial_t + \Omega(L^*)\partial_x)v = 0, \quad v \equiv \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \tag{2.2}$$

where $\Omega = \Omega(k)$ is a polynomial function and L^* is the operator

$$L^* = \begin{pmatrix} 0 & -\frac{1}{4}\partial_{xx} + v_1 + \frac{1}{2}v_{1x}I_+ \\ 1 & v_2 + \frac{1}{2}v_{2x}I_+ \end{pmatrix},$$
$$(I_+ g)(x) \equiv \int_{\infty}^{x} g(x')\,dx'. \tag{2.3}$$

These equations can be written as Hamiltonian systems[4,10] in a way we are going to describe in brief. Let $f_+(k,x)$ be the solution of (S) verifying

$$f_+(k,x) \underset{x \to +\infty}{\sim} e^{ikx}. \tag{2.4}$$

Consider the function $a = a(k)$ $(\mathrm{Im}k \geqslant 0)$ such that

$$f_+(k,x) \underset{x \to -\infty}{\sim} a(k)e^{ikx} + b(k)e^{-ikx}, \quad k \in \mathbb{R}. \tag{2.5}$$

The logarithm of $a(k)$ admits an asymptotic expansion

$$\ln a(k) = \frac{i}{2}\int_{-\infty}^{\infty} v_2(x)\,dx + \frac{i}{2k}\sum_{n=0}^{\infty}\frac{1}{k^n}H_n[v], \tag{2.6}$$

such that the coefficients $H_n[v]$ are functionals depending on v which satisfy the Lenard relation

$$J_1 \cdot \nabla H_{n+1} = J_2 \cdot \nabla H_n, \quad \nabla H_n \equiv \begin{pmatrix} \delta H_n/\delta v_1 \\ \delta H_n/\delta v_2 \end{pmatrix}, \tag{2.7}$$

where

$$J_1 \equiv \begin{pmatrix} -l(v_2) & \partial_x \\ \partial_x & 0 \end{pmatrix}, \quad J_2 \equiv \begin{pmatrix} -\frac{1}{4}\partial_{xxx} + l(v_1) & 0 \\ 0 & \partial_x \end{pmatrix},$$
$$l(v_i) \equiv v_i\partial_x + \tfrac{1}{2}v_{ix}. \tag{2.8}$$

The operators J_1 and J_2 are Hamiltonian operators. This means that two Lie algebra structures can be defined on the set of functionals depending on v given by (see Appendix A)

$$\{F,G\}_i[v] \equiv \langle \nabla F(v), J_i \cdot \nabla G(v) \rangle, \quad i = 1,2, \tag{2.9}$$

where $\langle\ ,\ \rangle$ denotes the bilinear form

$$\langle w,w' \rangle \equiv \int_{-\infty}^{\infty} dx\,(w_1(x)w_1'(x) + w_2(x)w_2'(x)). \tag{2.10}$$

The operations (2.9) are called the Poisson brackets associated with J_1 and J_2. As a consequence of (2.7), the family of functionals $\{H_n : n \geqslant 0\}$ turns out to be involutive with respect to both Poisson bracket operations.

The Hamiltonian formulation of Eqs. (2.2) is as follows.[10] The evolution equation

$$(\partial_t + \Omega(L^*)\partial_x)v = 0, \quad \Omega(k) = -\frac{1}{2}\sum_{n=0}^{N} c_n k^n \tag{2.11}$$

can be written as

$$\partial_t v = J_2 \cdot \nabla H(v), \quad H = \sum_{n=0}^{N} c_n H_n, \tag{2.12}$$

or equivalently in the form

$$\partial_t v = J_1 \cdot \nabla H'(v), \quad H' = \sum_{n=0}^{N} c_n H_{n+1}. \tag{2.13}$$

In our subsequent analysis we will use the Hamiltonian formulation (2.12) due to the transformation properties of J_2 under the Miura map.

B. Modified equations

Let us consider the transformation

$$v = M(u), \quad \begin{cases} v_1 = \frac{1}{2}u_{1x} + \frac{1}{4}u_1^2, \\ v_2 = u_2. \end{cases} \tag{2.14}$$

It is basically a Miura map from u_1 to v_1. Its Gateaux derivative M' is

$$M' = \begin{pmatrix} \frac{1}{2}(\partial_x + u_1) & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.15}$$

One finds easily that the Hamiltonian operator

$$D = \begin{pmatrix} \partial_x & 0 \\ 0 & \partial_x \end{pmatrix} \tag{2.16}$$

verifies

$$M'D\tilde{M}' = J_2, \tag{2.17}$$

where $\tilde{M}'$ denotes the transposed operator of M'. This implies (see Appendix A) that M is a canonical map relating Hamiltonian systems with respect to D and J_2:

$$\partial_t u = D \cdot \nabla(H \circ M)(u) \xrightarrow{M} \partial_t v = J_2 \cdot \nabla H(v). \tag{2.18}$$

In this way, from Eqs. (2.12) and (2.18) we obtain a new family of completely integrable Hamiltonian systems

$$\partial_t u = D \cdot \nabla(H \circ M)(u), \quad H = \sum_{n=0}^{N} c_n H_n. \tag{2.19}$$

They will be refered to as the "modified equations" associated with (S). The first few of them, corresponding to $H = 2H_0$, $4H_1$, $8H_2$, are

$$\partial_t u = u_x, \tag{2.20a}$$
$$\partial_t u_1 = -u_{2xx} + (u_1u_2)_x, \tag{2.20b}$$
$$\partial_t u_2 = u_{1xx} + \tfrac{1}{2}(u_1^2 + 3u_2^2)_x,$$
$$\partial_t u_1 = -u_{1xxx} - \tfrac{3}{2}(u_2^2)_{xx} + \tfrac{1}{2}(u_1^3)_x + \tfrac{3}{2}(u_1u_2^2)_x, \tag{2.21}$$
$$\partial_t u_2 = -u_{2xxx} + \tfrac{3}{2}(u_1^2u_2)_x + \tfrac{5}{2}(u_2^3)_x + 3(u_{1x}u_2)_x.$$

The functionals $\{H_n \circ M : n \geqslant 0\}$ are in involution with respect to the Poisson bracket induced by D, and hence they are constants of motion for each of the modified equations. Note also that, under the constraint $u_2 = 0$, Eq. (2.21) reduces to the modified KdV equation and therefore its image under M reduces to the KdV equation for v_1. Furthermore, one can prove that the modified equations (2.19) are compatible with the constraint $u_2 = 0$ when H is a linear combination of the functionals H_{2n} $(n \geqslant 0)$ only. In this case, their corresponding reductions over $u_2 = 0$ transform under M into the higher KdV flows for v_1. It justifies the term "modified equations" we are using here.

C. The generalized Zakharov–Shabat spectral problem

Since the work of Ablowitz, Kaup, Newell, and Segur,[6] we know that the inverse-scattering transform associated with the generalized Zakharov–Shabat spectral problem

$$(\mathrm{Z}):\quad \left(\partial_x - \begin{pmatrix} 0 & q_2(x) \\ q_1(x) & 0 \end{pmatrix} + ik\sigma_3\right)\varphi = 0, \qquad \varphi \equiv \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \tag{2.22}$$

provides an integration method for the family of nonlinear evolution equations

$$(\sigma_3\partial_t + 2\Lambda(\hat{L}^*))q = 0, \quad q \equiv \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \tag{2.23}$$

where $\Lambda = \Lambda(k)$ is a polynomial function and $\hat{L}^*$ is the operator defined by

$$\hat{L}^* \equiv i\begin{pmatrix} -\frac{1}{2}\partial_x + q_1 \cdot I_+ \cdot q_2 & -q_1 \cdot I_+ \cdot q_1 \\ q_2 \cdot I_+ \cdot q_2 & \frac{1}{2}\partial_x - q_2 \cdot I_+ \cdot q_1 \end{pmatrix}. \tag{2.24}$$

These equations will be refered to as the AKNS equations. They can also be expressed in Hamiltonian form.[7] The process which leads towards such Hamiltonian formulation is very similar to that followed in part A of this section for the Jaulent–Miodek equations. Its starting point is the solution $\varphi_+(k,x)$ of (Z) satisfying

$$\varphi_+(k,x) \underset{x \to +\infty}{\sim} \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{ikx}. \tag{2.25}$$

Then, one takes the function $\hat{a} = \hat{a}(k)$ $(\mathrm{Im}k \geqslant 0)$ such that

$$\varphi_+(k,x) \underset{x \to -\infty}{\sim} \hat{a}(k)\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{ikx} + \hat{b}(k)\begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-ikx}, \quad k \in \mathbb{R}. \tag{2.26}$$

The function $\ln\hat{a}(k)$ has an asymptotic expansion

$$\ln\hat{a}(k) = \frac{1}{2k}\sum_{n=0}^{\infty}\frac{1}{k^n}\hat{H}_n[q], \tag{2.27}$$

where the coefficients $\hat{H}_n[q]$ are functionals depending on q. In this case one finds[7] the Lenard relation

$$\hat{J}_1 \cdot \nabla\hat{H}_{n+1} = \hat{J}_2 \cdot \nabla\hat{H}_n, \quad \nabla\hat{H}_n \equiv \begin{pmatrix} \delta\hat{H}_n/\delta q_1 \\ \delta\hat{H}_n/\delta q_2 \end{pmatrix}, \tag{2.28}$$

$\hat{J}_1$ and $\hat{J}_2$ being the operators

$$\hat{J}_1 = i\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \hat{J}_2 = \begin{pmatrix} q_1 \cdot I_+ \cdot q_1 & \frac{1}{2}\partial_x - q_1 \cdot I_+ \cdot q_2 \\ \frac{1}{2}\partial_x - q_2 \cdot I_+ \cdot q_1 & q_2 \cdot I_+ \cdot q_2 \end{pmatrix}. \tag{2.29}$$

While $\hat{J}_1$ is clearly a Hamiltonian operator, it is not evident that the same property is shared by the operator $\hat{J}_2$. Nevertheless, the Hamiltonian character of $\hat{J}_2$ will be proved in Sec. III.

One may verify[7] that the evolution equation

$$(\sigma_3\partial_t + 2\Lambda(\hat{L}^*))q = 0, \quad \Lambda(k) = \tfrac{1}{2}\sum_{n=0}^{N} c_n k^{n+1} \tag{2.30}$$

can be put in the form

$$\partial_t q = \hat{J}_2 \cdot \nabla\hat{H}(q), \quad \hat{H} = \sum_{n=0}^{N} c_n \hat{H}_n \tag{2.31}$$

as well as

$$\partial_t q = \hat{J}_1 \cdot \nabla\hat{H}'(q), \quad \hat{H}' = \sum_{n=0}^{N} c_n \hat{H}_{n+1}. \tag{2.32}$$

This completes the information about Hamiltonian formalism aspects necessary to undertake the main task of this paper, which is to provide a link between the family of modified equations (2.19) and the AKNS equations.

III. A MAP TRANSFORMING MODIFIED EQUATIONS INTO AKNS EQUATIONS

A. Going to (Z) coming from (S)

The Miura map is strongly related to factorizing the Schrödinger operator.[3,11] Indeed, if the modified variables (u_1,u_2) are used, then (S) becomes

$$[(\partial_x + \tfrac{1}{2}u_1)(\partial_x - \tfrac{1}{2}u_1) + k(k - u_2)]f = 0. \tag{3.1}$$

This expression suggests us to write (S) as a first-order spectral problem

$$(\mathrm{M}):\quad \begin{cases} (\partial_x - \frac{1}{2}u_1)\theta_1 + ik\theta_2 = 0 \\ (\partial_x + \frac{1}{2}u_1)\theta_2 + i(k - u_2)\theta_1 = 0 \end{cases} \qquad (\theta_1 \equiv f). \tag{3.2}$$

We name it the "modified spectral problem" (M). Incidentally, it is interesting to note that, unlike the case (S), the potentials in (M) are free of k-dependent factors. The system (M) may be rewritten more concisely as

$$(\partial_x + U(k))\theta = 0, \quad U(k) \equiv \begin{pmatrix} -u_1/2 & ik \\ i(k - u_2) & u_1/2 \end{pmatrix}, \quad \theta \equiv \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}. \tag{3.3}$$

Similarly, we can put (Z) as

$$(\partial_x + Q(k))\varphi = 0, \quad Q(k) \equiv \begin{pmatrix} ik & -q_2 \\ -q_1 & -ik \end{pmatrix}, \quad \varphi \equiv \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}. \tag{3.4}$$

Now we look for an invertible matrix R relating (M) and (Z). That is, given a solution φ of (3.4), we demand that

$$\theta = R\varphi \tag{3.5}$$

be a solution of (3.3). One deduces easily what is required for it: The matrix R must verify the condition

$$R_x + UR - RQ = 0. \tag{3.6}$$

If we restrict ourselves to k-independent solutions, R, then (3.6) is equivalent to

$$\frac{\partial U}{\partial k}R - R\frac{\partial Q}{\partial k} = 0, \tag{3.7a}$$

$$R_x + U_0R - RQ_0 = 0, \tag{3.7b}$$

U_0 and Q_0 being the values of U and Q for $k = 0$, respectively. The solution of (3.7a) is

$$R = \begin{pmatrix} r & s \\ r & -s \end{pmatrix}. \tag{3.8}$$

On the other hand, Eq. (3.7b) reads

$$Q_0 = R^{-1}U_0R + R^{-1}R_x. \tag{3.9}$$

By using (3.8) and since the diagonal elements of Q_0 vanish, (3.9) implies

$$r_x = \tfrac{1}{2}iu_2r, \quad s_x = -\tfrac{1}{2}iu_2s. \tag{3.10}$$

Hence, we can take r and s as

$$r = \exp(\tfrac{1}{2} iI_+u_2), \quad s = \exp(-\tfrac{1}{2} iI_+u_2), \tag{3.11}$$

and we find at once that (3.9) is satisfied if and only if q and u are related in the following form

$$q = N(u), \quad \begin{cases} q_1 = \tfrac{1}{2}(u_1 - iu_2)\exp(iI_+u_2), \\ q_2 = \tfrac{1}{2}(u_1 + iu_2)\exp(-iI_+u_2). \end{cases} \tag{3.12}$$

Then we conclude that if $q = N(u)$, the spectral problems (M) and (Z) are linked by means of the matrix R determined in (3.8) and (3.11). This implies the following important consequence: Under the hypothesis $v = M(u)$ and $q = N(u)$, every solution $\varphi = \varphi(k,x)$ of (Z) determines a solution $f = f(k,x)$ of (S) given by

$$f = (R\varphi)_1 = \exp(\tfrac{1}{2} iI_+u_2)\varphi_1 + \exp(-\tfrac{1}{2} iI_+u_2)\varphi_2. \tag{3.13}$$

B. The Hamiltonian character of $\hat{J}_2$

We are now going to prove that the operator $\hat{J}_2$ appearing in the Lenard relation (2.28) is the image under N of the Hamiltonian operator D defined in (2.16). That is,

$$N' D \tilde{N}' = \hat{J}_2. \tag{3.14}$$

It implies (see Appendix A) that $\hat{J}_2$ is also a Hamiltonian operator.

The Gateaux derivative of N is easily calculated from (3.12), and it takes the form

$$N' = \begin{pmatrix} \bar{u}/2 & i(-\bar{u}/2 + q_1 \cdot I_+) \\ 1/2\bar{u} & i(1/2\bar{u} - q_2 \cdot I_+) \end{pmatrix}, \quad \bar{u} \equiv \exp(iI_+u_2), \tag{3.15}$$

Thus, we have

$$N'D\tilde{N}' = \begin{pmatrix} (\bar{u}/2)\partial_x & i((-\bar{u}/2)\partial_x + q_1) \\ (1/2\bar{u})\partial_x & i((1/2\bar{u})\partial_x - q_2) \end{pmatrix} \times \begin{pmatrix} \bar{u}/2 & 1/2\bar{u} \\ -i(\bar{u}/2 + I_+q_1) & i(1/2\bar{u} + I_+q_2) \end{pmatrix}. \tag{3.16}$$

By direct calculation, we obtain at once

$$(N'D\tilde{N}')_{11} = q_1 \cdot I_+ \cdot q_1, \quad (N'D\tilde{N}')_{22} = q_2 \cdot I_+ \cdot q_2, \tag{3.17}$$

and from (3.16) and (3.12) we get

$$\begin{aligned} (N'D\tilde{N}')_{12} &= -(N'D\tilde{N}')_{12}^{\sim} \\ &= \tfrac{1}{2}(\bar{u}\cdot\partial_x \cdot 1/\bar{u} - q_1/\bar{u} - q_2\bar{u}) - q_1 \cdot I_+ \cdot q_2 \\ &= \tfrac{1}{2}\partial_x - q_1 \cdot I_+ \cdot q_2. \end{aligned} \tag{3.18}$$

By comparing (3.17) and (3.18) with the explicit form of $\hat{J}_2$ given in (2.29), we conclude that (3.14) is satisfied.

According to this result, the transformation $q = N(u)$ defines a canonical map between the sets of Hamiltonian systems with respect to D and $\hat{J}_2$

$$\partial_t u = D \cdot \nabla(\hat{H} \circ N)(u) \xrightarrow{N} \partial_t q = \hat{J}_2 \cdot \nabla \hat{H}(q). \tag{3.19}$$

This raises an important question: Which of the Hamiltonian systems with respect to D transform under N into the AKNS equations? The answer is that they are indeed the modified equations (2.19). Let us prove it.

C. From modified equations to AKNS equations

Consider a modified equation

$$\partial_t u = D \cdot \nabla(H \circ M)(u), \quad H = \sum_{n=0}^{N} c_n H_n. \tag{3.20}$$

The functionals $H_n = H_n[v]$ are the coefficients of the asymptotic expansion (2.6) of $\ln[a(k)]$ and $v = M(u)$ is the Miura map defined in (2.14). We want to prove that the image of (3.20) under the map $q = N(u)$ defined in (3.12) is precisely the AKNS equation

$$\partial_t q = \hat{J}_2 \cdot \nabla \hat{H}(q), \quad \hat{H} = -i \sum_{n=0}^{N} c_n \hat{H}_n, \tag{3.21}$$

where the functionals $\hat{H}_n = \hat{H}_n[q]$ are the coefficients of the asymptotic expansion (2.27) of $\ln[\hat{a}(k)]$. In view of (3.14) all we have to prove is

$$iH_n \circ M = \hat{H}_n \circ N, \quad \forall\ n \geqslant 0. \tag{3.22}$$

To do it, let us return to the results of part A of this section. By assuming $v = M(u)$ and $q = N(u)$ we have that the solutions f and φ of (S) and (Z), respectively, are related in the form given in (3.13). Let us take $\varphi = \varphi_+(k,x)$; then by (2.25), (3.13), and the definition of I_+ [see (2.3)], we get

$$f(k,x) \underset{x \to +\infty}{\sim} e^{ikx}. \tag{3.23}$$

That is to say, $f = f_+(k,x)$. Consider now the limit of (3.13) when $x \to -\infty$. By using (2.26) it follows that

$$\begin{aligned} f_+(k,x) \underset{x \to -\infty}{\sim} &\exp\left(\frac{i}{2}\int_{-\infty}^{\infty} v_2\,dx\right)\hat{a}(k)e^{ikx} \\ &+ \exp\left(-\frac{i}{2}\int_{-\infty}^{\infty} v_2\,dx\right)\hat{b}(k)e^{-ikx}, \end{aligned} \tag{3.24}$$

which by comparison with (2.5) gives

$$\ln a(k) = \frac{i}{2}\int_{-\infty}^{\infty} v_2(x)\,dx + \ln\hat{a}(k). \tag{3.25}$$

If we substitute the asymptotic expansions (2.6) and (2.27) into (3.25), we deduce that $i \cdot H_n[v] = \hat{H}_n[q]$ for all $n \geqslant 0$ provided that $v = M(u)$ and $q = N(u)$. This proves (3.22) and therefore modified equations transform under N into AKNS equations.

D. A link between N and Hopf–Cole transformation

One of the elementary properties of AKNS equations is their compatibility with the constraint $q_1 = 0$. In fact, it is easily deduced from (2.23) and (2.24) that the generic equation (2.23) reduces when $q_1 = 0$ to the linear equation

$$\partial_t q_2 = 2\Lambda\,[(i/2)\partial_x]q_2. \tag{3.26}$$

But because of the definition of N we have that $q_1 = 0$ if and only if $u_1 = iu_2$. Hence, two consequences are immediately derived: modified equations preserve the constraint $u_1 = iu_2$ and their corresponding reductions transform under N into linear evolution equations. These facts remind us of the properties of the Hopf–Cole transformation. Indeed, by making $u_1 = iu_2$ the map $q = N(u)$ becomes

$$q_2 = -\partial_x \exp(-iI_+u_2), \tag{3.27}$$

which is nothing but the composition of ∂_x with the Hopf–

Cole transformation. From (3.26) and (3.27) it is easy to find the form of the modified equations when $u_1 = i u_2$. It turns out that (2.19) reduces to

$$\partial_t u_2 = \sum_{n=0}^{N} \frac{1}{2}\left(\frac{1}{2i}\right)^n c_n \partial_x (\partial_x + i u_2)^n u_2. \tag{3.28}$$

These equations are, as expected, Burgers equations. In this way, the map N is, when restricted to the manifold $u_1 = i u_2$, essentially the celebrated Hopf–Cole transformation.

It is also interesting to look at the constraint $u_1 = i u_2$ from the point of view of Jaulent–Miodek equations (2.11). By virtue of (2.14) it takes the form

$$v_1 = \tfrac{1}{2} i v_{2x} - \tfrac{1}{4} v_2^2, \tag{3.29}$$

and the resulting reduced equations coincide with (3.28) by replacing u_2 with v_2. Thus, we have that the Jaulent–Miodek equations reduce to Burgers equations provided v_1 and v_2 are related by (3.29).

IV. FROM AKNS EQUATIONS TO JAULENT–MIODEK EQUATIONS

In the previous sections two maps have been defined,

$$u \begin{array}{l} \nearrow v \\ \searrow q \end{array} \begin{array}{l} v = M(u), \\ q = N(u), \end{array} \tag{4.1}$$

which enable us to establish the following correspondences between evolution equations:

$$\partial_t u = D\cdot\nabla(H\circ M)(u) \begin{array}{l} \overset{M}{\nearrow} (\partial_t + \Omega(L^*)\partial_x)v = 0, \\ \underset{N}{\searrow} (\sigma_3\partial_t + 2\Lambda(\hat{L}^*))q = 0, \end{array} \tag{4.2}$$

where

$$H = \sum_{n=0}^{N} c_n H_n,$$

$$\Omega(k) = -\tfrac{1}{2}\sum_{n=0}^{N} c_n k_n, \quad \Lambda(k) = i k \Omega(k). \tag{4.3}$$

A question remains to be investigated: Is it possible to relate v with q by eliminating u into (4.1)? We are going to see here that it is, in fact, possible. Nevertheless, the correspondence between v and q turns out to be more complicated than M and N and it requires a careful analysis. The point is that until now, because of the simple explicit forms of M and N, we were allowed to study M and N without mentioning their domains of definition. Implicitly, we are assuming that u, v and q are rapidly decaying smooth functions, and clearly this assumption is consistent with the actions of M and N. The situation is different when we try to relate v and q since it compels us to consider the inverse transformations M^{-1} and N^{-1}.

Denote by U, V, and Q the phase spaces of u, v, and q, respectively. They will be assumed to be spaces of smooth two-component functions which vanish rapidly as $|x|\to\infty$ and such that M and N define maps from U to V and Q, respectively.[12] Let us now analyze whether NM^{-1} and MN^{-1} are or are not admissible maps. In what concerns to NM^{-1}, it turns out that its domain of definition is a quite restricted subset of V. The reason is the following:[13] Given $v\in V$, if we look for a two-component function u verifying $M(u) = v$, then we must take $u_2 = v_2$ and u_1 has to be a solution of the Riccati equation

$$\tfrac{1}{2} u_{1x} + \tfrac{1}{4} u_1^2 = v_1. \tag{4.4}$$

Since v_1 vanishes at infinity then $u_{1x} \sim u_1^2/2$ as $|x|\to\infty$. Hence the asymptotic behavior of a solution of (4.1) will be of the form

$$u_1(x) \underset{x\to\pm\infty}{\sim} \frac{2a_\pm}{a_\pm x + 1}. \tag{4.5}$$

Only when $a_+ = a_- = 0$, the function u_1 vanishes rapidly at infinity. We can impose either $a_+ = 0$ of $a_- = 0$ as an asymptotic boundary condition in order to fix univocally the solution u_1. But, except for very special v_1, there will not be any solution u_1 verifying both conditions simultaneously. This means that almost every $v\in V$ does not belong to the domain of M^{-1}, and consequently NM^{-1} cannot be considered as an admissible map. Consider now $q\in Q$, by virtue of (3.12) a function u satisfies $N(u) = q$ if and only if

$$u_1 = q_1/\bar{u} + q_2\bar{u}, \tag{4.6a}$$

$$u_2 = i(q_1/\bar{u} - q_2\bar{u}), \tag{4.6b}$$

where

$$\bar{u} \equiv \exp(iI_+ u_2). \tag{4.7}$$

From (4.6b) and (4.7) we find at once that $\bar{u}$ satisfies the Riccati equation

$$\bar{u}_x - q_2\bar{u}^2 + q_1 = 0. \tag{4.8}$$

Since q vanishes asymptotically, then $\bar{u}_x \sim 0$ as $|x|\to\infty$, which implies $\bar{u}(x)\sim a_\pm$ as $x\to\pm\infty$. By (4.7) we must impose $\bar{u}(x)\sim 1$ as $x\to+\infty$, and then it fixes univocally $\bar{u}$. Once having obtained $\bar{u}$, it is clear that the function u calculated by means of (4.6) is a rapidly decaying function. Nevertheless, the smooth character of u requires the absence of zeros and poles for $\bar{u}$, and it therefore sets a restriction over the domain of N^{-1}. But it is not so bad as the restriction we have found for M^{-1}. To see this, let us consider the following $k = 0$ solution of (Z):

$$\left.\begin{array}{l} \varphi_{1x} = q_2\varphi_2 \\ \varphi_{2x} = q_1\varphi_1 \end{array}\right\}, \quad \begin{pmatrix}\varphi_1(x)\\ \varphi_2(x)\end{pmatrix}_{x\to+\infty} \sim \begin{pmatrix}1\\ -1\end{pmatrix}. \tag{4.9}$$

One proves easily that $\bar{u} \equiv -\varphi_2/\varphi_1$ is our solution of (4.8) satisfying $\bar{u}\sim 1$ as $x\to+\infty$. Thus, in order that $\bar{u}$ have neither zeros or poles, the functions φ_1 and φ_2 must be zero-free. This condition may be conveniently studied by considering the Green's transform equation[14] of (4.9) which is given by

$$\varphi_1(x)^*\varphi_2(x) = -1 + \int_{\infty}^{x} (q_1|\varphi_1|^2 + q_2^*|\varphi_2|^2)\, dx'. \tag{4.10}$$

From this equation we deduce at once that φ_1 and φ_2 are zero-free if either of the following conditions if verified:

(i) $\mathrm{Re}\, q_1 \leqslant 0$, $\mathrm{Re}\, q_2 \leqslant 0$

or

(ii) $\mathrm{Im}\, q_1 > 0$ $(\leqslant 0)$, $\mathrm{Im}\, q_2 \leqslant 0$ (> 0).

These inequality restrictions determine subsets of Q having the same "dimensionality" as Q. In this way, we may conclude that MN^{-1} determines a transformation $q\to v$ defined on an appreciable domain of rapidly decaying smooth functions.

The mapping MN^{-1} transforming evolution equations associated with (Z) into those associated with (S) coincides with the one found by Jaulent and Miodek[9] by using a different approach based on the inverse-scattering transforms for (S) and (Z). These authors observed that the set of scattering data necessary in order to construct v and q by the inverse-scattering transform for (S) and (Z), respectively, are formally identical (see Appendix B):

$$s(v) = \{R^{\pm}(k), k_n^{\pm}, c_n^{\pm}\}, \quad k\in\mathbb{R}, \quad n = 1,...,N^{\pm},$$
$$\hat{s}(q) = \{\hat{R}^{\pm}(k), \hat{k}_n^{\pm}, \hat{c}_n^{\pm}\}, \quad k\in\mathbb{R}, \quad n = 1,...,\hat{N}^{\pm}. \tag{4.11}$$

On the other hand, written in terms of scattering data, Jaulent–Miodek equations and AKNS equations take the forms

$$\partial_t \ln R^{\pm} = \mp 2ik\Omega(\pm k), \quad \partial_t k_n^{\pm} = 0,$$
$$\partial_t \ln c_n^{\pm} = \mp 2ik_n^{\pm}\Omega(\pm k_n^{\pm}) \tag{4.12}$$

and

$$\partial_t \ln \hat{R}^{\pm} = \mp 2\Lambda(\pm k), \quad \partial_t \hat{k}_n^{\pm} = 0,$$
$$\partial_t \ln \hat{c}_n^{\pm} = \mp 2\Lambda(\pm \hat{k}_n^{\pm}), \tag{4.13}$$

respectively. Clearly, the identity map $s = \hat{s}$ transforms (4.13) into (4.12) provided $\Lambda(k) = ik\Omega(k)$.

The main result of Jaulent and Miodek can be described as follows (see Appendix B): Given $u\in U$, then $v = M(u)$ and $q = N(u)$ satisfy $s(v) = \hat{s}(q)$. In virtue of this fact, if we denote by S the common space of scattering data for (S) and (Z), then the following diagram is commutative:

$$\begin{array}{ccccc} & & V & & \\ & \overset{M}{\nearrow} & & \overset{s}{\searrow} & \\ U & & & & S \\ & \underset{N}{\searrow} & Q & \underset{\hat{s}}{\nearrow} & \end{array} \tag{4.14}$$

Therefore, we have that $MN^{-1} = s^{-1}\hat{s}$ over some suitable domain of Q.

We note that according with the results of Sec. III the Hamiltonian operator $\hat{J}_2$ transforms into J_2 under MN^{-1}. In a previous work[15] we proved that the image of $\hat{J}_1$ under the Jaulent–Miodek map is J_1. It shows that the Jaulent–Miodek map is canonical with respect to both pairs of Hamiltonian structures $(\hat{J}_1, J_1)$ and $(\hat{J}_2, J_2)$ defined on (V,Q).

APPENDIX A

Let V be a functional space whose elements are N-component functions $v(x) = (v_i(x))$ $(i = 1,...,N)$ depending smoothly on a real variable x and which vanish rapidly at infinity. On V we can define the bilinear form

$$\langle v,w\rangle = \int_{-\infty}^{\infty} \sum_{i=1}^{N} v_i(x)w_i(x)\,dx. \tag{A1}$$

Given a functional $F = F[v]$, we denote by $\nabla F(v)$ its gradient, that is, the N-component function (or distribution) verifying

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} F[v+\epsilon w] = \langle \nabla F(v), w\rangle. \tag{A2}$$

By a Hamiltonian operator on V we shall mean a linear operator J depending on v such that

$$\{F,G\}[v] = \langle \nabla F(v), J\cdot\nabla G(v)\rangle \tag{A3}$$

defines a Lie algebra structure on the set of functionals. The operation $\{\ ,\ \}$ is called the Poisson bracket associated with J. In this way, Hamiltonian systems on V are identified with those evolution equations which can be put in the form

$$\partial_t v = J\cdot\nabla H(v). \tag{A4}$$

Suppose we have a transformation T: $V\to\bar{V}$ from V into a similar functional space $\bar{V}$ with elements $\bar{v}(x) = (\bar{v}_i(x))$ $(i = 1,...,N)$. Let J and $\bar{J}$ be Hamiltonian operators on V and $\bar{V}$, respectively, then we say that T is a canonical transformation if given arbitrary functionals $\bar{F}$ and $\bar{G}$ on $\bar{V}$ the following condition is verified

$$\{\bar{F}\circ T,\bar{G}\circ T\} = \{\bar{F},\bar{G}\}^{-}\circ T, \tag{A5}$$

where $\{\ ,\ \}$ and $\{\ ,\ \}^{-}$ are the Poisson brackets operations of V and $\bar{V}$, respectively. By making use of the chain rule

$$\nabla(\bar{F}\circ T)(v) = \tilde{T}'\cdot\nabla\bar{F}(\bar{v}), \quad \bar{v} = T(v), \tag{A6}$$

where $\tilde{T}'$ denotes the transposed operator of the Gateaux derivative

$$T'\cdot w = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} T(v+\epsilon w), \tag{A7}$$

one finds easily that (A5) is equivalent to the condition

$$\bar{J} = T'\cdot J\cdot\tilde{T}'. \tag{A8}$$

It is important to observe that, given a Hamiltonian operator J on V, then the operator $\bar{J}$ defined on $\bar{V}$ by means of (A8) is a Hamiltonian operator on $\bar{V}$. Indeed, from (A8) it follows that the operation $\{\ ,\ \}^{-}$ associated with $\bar{J}$ satisfies (A5) and therefore

$$\{\{\bar{F},\bar{G}\}^{-},\bar{H}\}^{-}[\bar{v}] = (\{\{\bar{F},\bar{G}\}^{-},\bar{H}\}^{-}\circ T)(v)$$
$$= \{\{\bar{F},\bar{G}\}^{-}\circ T,\bar{H}\circ T\}[v] = \{\{\bar{F}\circ T,\bar{G}\circ T\},\bar{H}\circ T\}[v], \tag{A9}$$

which permits us to prove easily Jacobi identity for $\{\ ,\ \}^{-}$ from the Jacobi identity for $\{\ ,\ \}$. The rest of conditions necessary in order to conclude the Lie algebra character of $\{\ ,\ \}^{-}$ are trivial.

Given a canonical transformation T: $V\to\bar{V}$ and a Hamiltonian system on $\bar{V}$,

$$\partial_t\bar{v} = \bar{J}\cdot\nabla\bar{H}(\bar{v}), \tag{A10}$$

then from (A6), (A7), and (A8) it follows that

$$\partial_t\bar{v} - \bar{J}\cdot\nabla\bar{H}(\bar{v}) = T'(\partial_t v - J\cdot\nabla(\bar{H}\circ T)(v)). \tag{A11}$$

This means that the map $\bar{v} = T(v)$ establishes a correspondence between the solutions of the Hamiltonian systems

$$\partial_t v = J\cdot\nabla(\bar{H}\circ T)(v) \xrightarrow{T} \partial_t\bar{v} = \bar{J}\cdot\nabla\bar{H}(\bar{v}). \tag{A12}$$

APPENDIX B

The set of scattering data for the spectral problem (S) can be defined by considering the solutions (S) verifying

$$f_-(k,x) \underset{x\to-\infty}{\sim} e^{-ikx}, \quad \bar{f}_-(k,x) \underset{x\to-\infty}{\sim} e^{ikx}. \tag{B1}$$

Then we define the functions $R^{\pm}(k)$ and $T^{\pm}(k)$ through the asymptotic behavior

$$f_-(k,x) \underset{x\to+\infty}{\sim} \frac{1}{T^{+}(k)}(R^{+}(k)e^{ikx} + e^{-ikx}),$$
$$\bar{f}_-(k,x) \underset{x\to-\infty}{\sim} \frac{1}{T^{-}(-k)}(R^{-}(-k)e^{-ikx} + e^{ikx}). \tag{B2}$$

The relevant set of scattering data is provided by[8]

$s(v) = \{R^{\pm}(k), k_n^{\pm}, c_n^{\pm}\}$ $(k \in \mathbb{R},\ n = 1,..,N^{\pm})$, where $k_n^{\pm}$ are the zeros of $1/T^{\pm}(k)$ and $c_n^{\pm}$ are given by

$$c_n^{\pm} = i\left(\frac{R^{\pm}(k)}{T^{\pm}(k)}\left(\frac{\partial}{\partial k}\frac{1}{T^{\pm}(k)}\right)^{-1}\right)\Bigg|_{k=k_n^{\pm}}. \tag{B3}$$

Similarly, the scattering data for (Z) are defined from the solutions of (Z) verifying

$$\varphi_-(k,x) \underset{x\to-\infty}{\sim} \begin{pmatrix}1\\0\end{pmatrix} e^{-ikx}, \quad \bar{\varphi}_-(k,x) \underset{x\to-\infty}{\sim} \begin{pmatrix}0\\1\end{pmatrix} e^{ikx}. \tag{B4}$$

Then, the functions $\hat{R}^{\pm}(k)$ and $\hat{T}^{\pm}(k)$ are defined by means of the asymptotic form of these solutions at $x = +\infty$,

$$\varphi_-(k,x) \underset{x\to+\infty}{\sim} \frac{1}{\hat{T}^{+}(k)}\left(\hat{R}^{+}(k)\begin{pmatrix}0\\1\end{pmatrix} e^{ikx} + \begin{pmatrix}1\\0\end{pmatrix} e^{-ikx}\right), \tag{B5a}$$

$$\bar{\varphi}_-(k,x) \underset{x\to-\infty}{\sim} \frac{1}{\hat{T}^{-}(-k)}\left(\hat{R}^{-}(-k)\begin{pmatrix}1\\0\end{pmatrix} e^{-ikx} + \begin{pmatrix}0\\1\end{pmatrix} e^{ikx}\right). \tag{B5b}$$

Here the relevant set of scattering data is $\hat{s}(q) = \{\hat{R}^{\pm}(k), \hat{k}_n^{\pm}, \hat{c}_n^{\pm}\}$ $(k \in \mathbb{R},\ n = 1,...,\hat{N}^{\pm})$, where $\hat{k}_n^{\pm}$ and $\hat{c}_n^{\pm}$ obtain in the same way as their analogs for (S).

We are now going to prove the following statement mentioned in Sec. IV:

$$v = M(u), \quad q = N(u) \Rightarrow s(v) = \hat{s}(q). \tag{B6}$$

By virtue of (3.13) and (B4) we have

$$f(k,x) \equiv (R\varphi_-(k,x))_1 \underset{x\to-\infty}{\sim} \exp\left(-\frac{i}{2}\int_{-\infty}^{\infty} u_2\, dx'\right) e^{-ikx}. \tag{B7}$$

Hence

$$f(k,x) = \exp\left(-\frac{i}{2}\int_{-\infty}^{\infty} u_2\, dx'\right) f_-(k,x). \tag{B8}$$

If we consider now the limit of (3.13) when $x \to +\infty$ and we use (B2), (B5), and (B8), it is straightforward to see that

$$T^{+}(k) = \exp\left(-\frac{i}{2}\int_{-\infty}^{\infty} u_2\, dx'\right)\hat{T}^{+}(k),$$
$$R^{+}(k) = \hat{R}^{+}(k). \tag{B9}$$

It implies at once $(k_n^{\pm}, c_n^{\pm}) = (\hat{k}_n^{\pm}, \hat{c}_n^{\pm})$. The rest of the proof of (B6) follows by repeating the same procedure with $\bar{\varphi}_-$ instead of φ_-.

[1] I. M. Gel'fand and L. A. Dikii, Usp. Mat. Nauk. **30**, 67 (1975); Funk. Anal. Priloz. **10**, 18 (1976); Funk. Anal. Priloz. **10** (4), 13 (1976); Funk. Anal. Priloz. **11** (1), 11 (1977); Funk. Anal. Priloz. **12** (2), 8 (1978).
[2] M. Adler, Invent. Math. **50**, 219 (1979).
[3] B. A. Kupershmidt and G. Wilson, "Modifying Lax equations and the second Hamiltonian structure," Invent. Math. **62**, 403 (1981).
[4] L. Martínez Alonso, Lett. Math. Phys. **4**, 215 (1980); J. Math. Phys. **21**, 2342 (1980).
[5] L. Martínez Alonso, "A new infinite-dimensional Kirillov's structure related with nonlinear evolution equations," Lett. Math. Phys. **5**, 99 (1981).
[6] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 294 (1974).
[7] H. Flashka and A. C. Newell, "Integrable Systems of Nonlinear Evolution Equations," in *Lecture Notes in Physics*, Vol. 38, edited by J. Moser (Springer-Verlag, Berlin, 1975).
[8] M. Jaulent and I. Miodek, Lett. Math. Phys. **1**, 243 (1976).
[9] M. Jaulent and I. Miodek, Lett. Nuovo Cimento **20**, 655 (1977).
[10] F. Guil Guerrero and L. Martínez Alonso, Lett. Nuovo Cimento **27**, 28 (1980).
[11] A. P. Fordy and J. Gibbons, J. Math. Phys. **21**, 2508 (1980).
[12] For instance, U, V, and Q may be identified with the space of two-component smooth functions with compact support.
[13] M. J. Ablowitz and M. Kruskal, J. Math. Phys. **20**, 999 (1979).
[14] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[15] F. Guil Guerrero and L. Martínez Alonso, Lett. Nuovo Cimento **27**, 85 (1980).

Classification scheme for two-dimensional Ermakov-type systems and generalizations

W. Sarlet
Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 271-S9, B-9000 Gent, Belgium
J. R. Ray
Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631

(Received 21 May 1981; accepted for publication 20 July 1981)

We describe a classification scheme for certain pairs of coupled, nonlinear second-order ordinary differential equations, having a first integral which can be obtained by an Ermakov-type elimination procedure. The classification produces a hierarchy of systems having different levels of generality. Different cases within the same level are distinguished by the integrability of a certain one-form, which is related to a nonlinear superposition law between the two equations. The study of these one-forms and the classification are made systematic by using the general reduction theorem of Pfaffian forms. The paper contains illustrations of the way previous examples fit into the classification scheme, and a discussion of its generalization to higher dimensions.

PACS numbers: 03.20. + i,02.30.Hg

1. INTRODUCTION

Over the past five years, there has been a rapidly growing flow of new contributions to the problem of constructing first integrals for explicitly time-dependent, nonlinear, dynamical systems. The motivation for these studies goes back to Lewis' derivation of an invariant for the time-dependent harmonic oscillator,[1] and its successful use in solving exactly the quantum problem for this system.[2] Attempts to generalize Lewis' results within the same methodology (an application of Kruskal's perturbation method[3] in closed form), have so far failed to produce significantly broader examples.[4] Since then, a whole array of new methods have been used to rederive Lewis' result in a simpler way, or to construct different kinds of generalizations.

For a brief review of these different methods, we mention first the use of canonical transformations. In these approaches, some supplementary restriction has to be imposed from the outset. So far the following types of restrictions have been considered: the linearity of the transformation[5-7]; its time-independence[8]; or its dependence on time through some auxiliary function $\rho(t)$ and its derivative.[9] A second category of methods are the group-theoretical ones, which in this context mainly consist of Noether's theorem,[10-13] and the Lie method of extended groups.[14,15] In practically all cases so far discussed in the literature, the resulting invariants are quadratic in the velocities (or momenta). If a particular structure for the invariant can be surmised *a priori*, this structure can be used to search for the invariant by more *ad hoc* procedures.[16-18] In most of the above cited work, the starting point is a single second-order differential equation, and the search for a first integral introduces an auxiliary variable which itself must solve another second-order differential equation. Ray and Reid, in a series of papers,[11,19,20] have further broadened the picture by treating these equations as special cases of the situation where two fully coupled equations have a first integral of the same type. Their analysis traces back to an old paper by Ermakov.[21] After Lutzky[22] discussed the crucial role played by a Wronskian variable in some of these Ermakov-type systems, Sarlet showed[23] how in more general terms, the integrability of a differential one-form is an important feature of the theory, and may contain the clue for a better understanding of the cohesion between all examples, scattered over various papers, and derived by various methods. It is the purpose of the present paper to explore this idea in greater depth. We will arrive at a classification scheme for all Ermakov-type systems, in terms of a number of distinct levels of generality. The term classification is justified by the fact that we appeal to a well-known theorem about the reduction of Pfaffian forms, by which we can limit ourselves to a complete discussion of certain standard forms. The last level of generality will contain the previous ones. This will, however, not imply that these previous ones should not be mentioned separately, because a generalization sometimes only works at the expense of eliminating interesting properties. The main property we have in mind here is the existence of what has been called a "nonlinear superposition law" between the pair of equations (see, e.g., Ref. 19). This law is interesting, if only because it provides the link with those cases in which one of the equations plays a privileged role, and the other one is considered as auxiliary. The classification we make in this paper is restricted to the case of two dimensions (two dependent variables), plus possibly and extra dimension for the independent time-variable. We plan analogous studies for systems with several dependent variables in subsequent papers and comment briefly on such systems in the conclusions. In the next section, we first discuss an example, which will enable us to show the appearance of an integrable one-form, and the way we can transform the problem to a "standard representation."

2. PRELIMINARY EXAMPLE

Let us start with an example from previous contributions, and show how it can be cast into an appropriate form for our present investigations. Consider the system of equa-

 0022-2488/81/112504-08$01.00

tions (k is an arbitary constant),

$$\ddot{q} - \tfrac{1}{2}kq^{-1}\dot{q}^2 + (2-k)\omega^2(t)q = f(q,\rho,t), \tag{1a}$$

$$\ddot{\rho} + \tfrac{1}{2}(k-2)^2\omega^2(t)\rho = g(q,\rho,t), \tag{1b}$$

which would contain examples discussed by Sarlet,[8] if f and g were uncoupled (partially). Multiplying (1a) by $\frac{1}{2}(k-2)\rho q^{-k/2}$, (1b) by $q^{1-k/2}$, and adding the results, the left-hand side can be written in the form dw/dt, with

$$w = \tfrac{1}{2}(k-2)\rho q^{-k/2}\dot{q} + q^{1-k/2}\dot{\rho}. \tag{2}$$

So if f and g were zero, w would be a first integral of the system (1). Assuming that we want a first integral for the "perturbed case," which is quadratic in the velocities, it is therefore natural to ask under what conditions for f and g (1) will have a first integral of the form

$$I = \tfrac{1}{2}w^2 + F(q,\rho,t). \tag{3}$$

Expressing $\dot{I} = 0$ yields

$$w[\tfrac{1}{2}(k-2)\rho q^{-k/2}f + q^{1-k/2}g] + \frac{\partial F}{\partial q}\dot{q} + \frac{\partial F}{\partial \rho}\dot{\rho} + \frac{\partial F}{\partial t} = 0, \tag{4}$$

which can only be identically satisfied if $\partial F/\partial t = 0$, while $\partial F/\partial q$ and $\partial F/\partial \rho$ are proportional to the $\dot{q}$ and $\dot{\rho}$ part of the first term. The latter relations then imply

$$q^{1-k/2}\frac{\partial F}{\partial q} - \frac{1}{2}(k-2)\rho q^{-k/2}\frac{\partial F}{\partial \rho} = 0;$$

hence,

$$F = \bar{F}(u) \quad \text{with } u = \rho^{-1}q^{1-k/2}. \tag{5}$$

Substituting the structure (5) of F into condition (4) finally produces the following condition on the right-hand sides of (1),

$$\tfrac{1}{2}(k-2)\rho q^{-k/2}f + q^{1-k/2}g = \rho^{-2}\bar{F}'(u). \tag{6}$$

As an example we could take

$$f = bq^{2k-3}, \quad g = a\rho^{-3},$$

in which case (1) now exactly reduces to an example derived in Ref. 8 by a canonical transformation method. $\bar{F}(u)$ then has the form

$$\bar{F}(u) = -\tfrac{1}{4}(k-2)bu^{-2} + \tfrac{1}{2}au^2.$$

The manipulations which enabled us to pass from system (1) to the identification of the quantity w in (2) may seem rather artificial at first sight, but our later analysis eventually will produce precise rules for detecting the appropriate multipliers, of course, only for those systems that possess invariants of the type being discussed. At the present stage, we want to emphasize the fact that the required form (5) for F should not be a surprise, because there is a well-defined relationship between the quantities w and u. Indeed we have

$$w = -\rho^2\frac{du}{dt}; \tag{7}$$

in other words, the one-form in q and ρ, which can be associated to w, becomes exact ($= -du$) upon multiplication by ρ^{-2}. It is this feature, together with the form of the invariant I, which lies at the heart of what was called a "nonlinear superposition principle" in previous papers. To be precise, we can introduce a new, independent time-variable by

$$\tau = \int^t \rho^{-2}(t')\,dt', \tag{8}$$

by which the invariant I can be written in the form,

$$I = \tfrac{1}{2}\left(\frac{du}{d\tau}\right)^2 + \bar{F}(u). \tag{9}$$

Hence, a simple quadrature yields

$$\tau + C = \int du/[2(I - \bar{F}(u))]^{1/2},$$

which after inverting and taking account of the definition of τ and u finally gives

$$q^{1-k/2} = \rho U\left(I, \int \rho^{-2}dt + C\right) \tag{10}$$

for some function U. Since the right-hand side of (10) depends on two arbitrary constants, knowing a particular solution for $\rho(t)$ will produce the general solution for $q(t)$, which is what is meant by the terminology "nonlinear superposition" in this context. Examples are known where this procedure can be carried through to obtain the solution.

Now that we have seen how an integrable one-form emerges in our considerations, there remains the problem of the reduction to some kind of standard form. To that end, consider the coordinate transformation

$$x = \rho, \quad y = q^{1-k/2} \tag{11}$$

by which we obtain the following transformation of the one-form in question,

$$\tfrac{1}{2}(k-2)\rho q^{-k/2}dq + q^{1-k/2}d\rho = x\,dy - y\,dx. \tag{12}$$

The differential equations (1) accordingly transform to

$$\ddot{x} + \tfrac{1}{2}(k-2)^2\omega^2(t)x = f_1(x,y,t),$$
$$\ddot{y} + \tfrac{1}{2}(k-2)^2\omega^2(t)y = f_2(x,y,t), \tag{13}$$

in which we recognize for appropriate choices of f_1 and f_2 (and hence of f and g) the usual form of an Ermakov-type system (see, e.g., Ref. 17). It is this possibility which we will exploit for limiting most of our discussions to standard forms only. All permissible systems then belong to one of our categories up to an arbitrary transformation of the independent and dependent variables. Since we are primarily interested in the structure of permissible systems, we disregard problems of isolated singularities in the transformations involved. The fundamental reason why we can feel confident about looking at a number of standard forms only, is the fact that we can rely on the general theorem concerning the reduction of Pfaffian forms, which will even gain more importance, when we pass to higher dimensions. For our present needs, we will at most be dealing with two spatial dimensions and time. So let us recall the theorem for three dimensions.

Theorem (Reduction of Pfaffian forms): Consider a one-form

$$\alpha = a_1\,dx_1 + a_2\,dx_2 + a_3\,dx_3,$$

where the a_i are smooth functions of (x_1,x_2,x_3) in the neigh-

borhood of some point P. Let A denote the 4×3 matrix

$$A=\begin{pmatrix} a_1 & a_2 & a_3 \\ \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{pmatrix},$$

$$\text{with } \omega_{ij}=\frac{\partial a_i}{\partial x_j}-\frac{\partial a_j}{\partial x_i},$$

and let r denote the rank of A in the neighborhood of P. Then, there exists a new coordinate system (u_1,u_2,u_3) such that near P,

(i) $\alpha = u_1 du_2 + du_3$, if $r=3$,
(ii) $\alpha = u_1 du_2$, if $r=2$,
(iii) $\alpha = du_1$, if $r=1$.

Before embarking on our main section, it remains to explain what the relationship is between the statements in this reduction theorem, and the form $x\,dy - y\,dx$ which we termed the standard form in the above example. In that example, we are dealing with a two-dimensional one-form, for which the rank of the corresponding matrix A is 2. Hence, the reduction brings us to the form $u_1\,du_2$, which is in fact exactly what we did in (7). Of course, the one-form $xdy - ydx$ can equally be reduced to the same $u_1\,du_2$, so that an appropriate composition of coordinate transformations allows the transition expressed in (12). Now remember that the subject under investigation here is the study of certain types of (coupled) nonlinear second-order differential equations, and it turns out that $xdy - ydx$ in fact corresponds to the type of systems [like (13)], which were historically first considered as being of Ermakov-type. So it is mainly for historical reasons that we pick out $xdy - ydx$ as standard form in two dimensions. The fact that this form can be reduced to $u_1\,du_2 = x^2 d(y/x)$ is the key to the property of nonlinear superposition, as explained before. We will always try to choose the standard forms in a way such that they exhibit explicitly the same dimension (i.e., the same number of differentials) as the original forms, while the ultimate reduction provided by the theorem above will determine whether or not we have nonlinear superposition.

3. CLASSIFICATION OF DIFFERENT LEVELS OF GENERALITY

In the first two subsections, we will focus on classes of problems to which most of the examples, treated in the previously cited literature, belong. This essentially amounts to two assumptions: (i) First we have in mind systems like (1), where the right-hand sides (forces) do not depend on the velocities, so that all terms containing derivatives eventually determine w. (ii) Secondly, we are thinking of invariants which are quadratic in the velocities, or more precisely, which are of the "energy-type" $I=\frac{1}{2}w^2+F$.

We will later on consider dropping these assumptions.

A. Systems reducible to the standard form by time-independent coordinate transformations

In this category, we don't have to bother about time as adding a supplementary dimension to the problem. By a time-independent coordinate transformation we of course mean $x'=x'(x,y)$, $y'=y'(x,y)$, $(t'=t)$. The standard one-form, assuming $r=2$, is $xdy-ydx$ and we have

$$x^{-2}(xdy-ydx)=du, \quad u=y/x. \tag{14}$$

The invariants under consideration therefore have the structure

$$I=\tfrac{1}{2}(x\dot{y}-y\dot{x})^2+F(y/x), \tag{15}$$

and all systems of equations having I as a first integral are obtained by expressing $\dot{I}=0$. Denoting such systems by

$$\begin{aligned} \ddot{x}&=f_1(x,y,t), \\ \ddot{y}&=f_2(x,y,t), \end{aligned} \tag{16}$$

we find the condition

$$xf_2-yf_1+x^{-2}F'(y/x)=0. \tag{17}$$

This covers the results discussed by Lutzky.[22] It is of some interest to repeat[23] the corresponding formulae for all other systems which belong to this same class, namely, those systems which are obtained by an arbitrary time-independent coordinate transformation from the system just discussed. We then have a general one-form, for which an appropriate multiplier $f(x,y)$ exists, such that $f(adx+bdy)=du$, for some function $u(x,y)$.

The invariant is of the form,

$$I=\tfrac{1}{2}(a\dot{x}+b\dot{y})^2+F(u). \tag{18}$$

In agreement with assumption (i), the permissible differential equations in x and y must be such that

$$\begin{aligned} &a\ddot{x}+b\ddot{y}+a_x\dot{x}^2+(a_y+b_x)\dot{x}\dot{y}+b_y\dot{y}^2 \\ &\quad =aF_1+bF_2 \end{aligned} \tag{19}$$

(subscripts denote partial derivatives here), where F_1 and F_2 are functions of x,y,t, satisfying

$$aF_1+bF_2+fF'(u)=0 \tag{20}$$

which is the analog of Eq. (17).

Some remarks are in order now. First of all, it should be stressed that the above category already contains an infinite number of systems (even disregarding the possibility for transformations), because (17) or (19) and (20) only impose a condition on the combination of the two equations, and not on each equation separately. Of course, a more or less natural splitting of Eq. (19) (which explains the particular form we wrote down in its right-hand side) is given by

$$\frac{d}{dt}(a\dot{x})=aF_1, \quad \frac{d}{dt}(b\dot{y})=bF_2,$$

with F_1 and F_2 satisfying (20). Secondly, for the sake of completeness, we should say something about the case $r=1$ in this context. This case means that the one-form in question is already exact as given. One could also choose some standard representation for this case (like $xdx+ydy$), but this is not really relevant. Instead, it suffices to say that the case $r=1$ can be obtained from the situation (19), (20), for $f=1$, which in discussing nonlinear superposition simply means that the time variable τ then coincides with t.

B. The essentially time-dependent case

When we look, e.g., at (19) and (20), it is clear that we also should deal with the case that a and b depend explicitly on time, so that a possible coordinate transformation to standard form will also be time-dependent. For an appropriate generalization of the previous considerations, we therefore have to discuss one-forms in three dimensions (two space and one time dimension). With regard to the reduction theorem of Sec. 2, we have to distinguish between three possibilities.

Let us start with the most interesting one, namely the case $r = 2$. Then, a coordinate transformation can again bring us to a one-form like $u_1\,du_2$, or equivalently to $x\,dy - y\,dx$. However, if we want to keep track of the fact that our coordinate transformations depend explicitly on time now, it is better to choose a standard representation which exhibits this time involvement. To that end, it suffices to let the transformation, which turns $u_1\,du_2$ into $x\,dy - y\,dx$, depend parametrically on t, and the simplest way to do so is by taking

$$u_1 = x^2, \quad u_2 = y/x + \alpha(t). \tag{21}$$

Hence, the most appropriate choice for a standard one-form here seems to be

$$x\,dy - y\,dx + x^2\dot{\alpha}\,dt. \tag{22}$$

Indeed, every one-form like

$$a(x',y',t)\,dx' + b(x',y',t)\,dy' + e(x',y',t)\,dt, \tag{23}$$

for which a multiplier exists, yielding an exact form ($r = 2$), can be reduced to (22) by a time-dependent coordinate transformation. Moreover, (22) allows us to recover the results of the time-independent case by putting simply $\alpha = 0$.

With (22) as a standard form, and in the spirit of assumptions (i) and (ii), we now consider invariants of the form

$$I = \tfrac{1}{2}(x\dot{y} - y\dot{x} + x^2\dot{\alpha})^2 + F(x,y,t).$$

But looking at the result (15), and keeping in mind the minor modification expressed in (21), it is clear that we will end up with

$$I = \tfrac{1}{2}(x\dot{y} - y\dot{x} + x^2\dot{\alpha})^2 + F(y/x + \alpha(t)), \tag{24}$$

where the x and y equations must be such that

$$x\ddot{y} - y\ddot{x} + 2x\dot{x}\dot{\alpha} + x^2\ddot{\alpha} = xf_2(x,y,t) - yf_1(x,y,t), \tag{25}$$

with f_1 and f_2 satisfying

$$xf_2 - yf_1 + x^{-2}F'(y/x + \alpha) = 0. \tag{26}$$

As before, we list also here the formulae which must be taken into account, in case one deals with a general representation(23), omitting primes on the variables. Let $f(x,y,t)$ be the multiplier for which

$$f(a\,dx + b\,dy + e\,dt) = du \tag{27}$$

for some $u(x,y,t)$. Then,

$$I = \tfrac{1}{2}(a\dot{x} + b\dot{y} + e)^2 + F(u), \tag{28}$$

with

$$\begin{aligned} &a\ddot{x} + b\ddot{y} + a_x\dot{x}^2 + (a_y + b_x)\dot{x}\dot{y} + b_y\dot{y}^2 \\ &\quad + (a_t + e_x)\dot{x} + (b_t + e_y)\dot{y} + e_t \\ &= aF_1 + bF_2, \end{aligned} \tag{29}$$

and

$$aF_1 + bF_2 + fF'(u) = 0. \tag{30}$$

We leave it to the reader to verify that we still have, in this generalization, the kind of nonlinear superposition discussed before.

Concerning the situations where $r = 1$, we again limit ourselves to the remark that they are incorporated in the results (28), (29), and (30) for $f = 1$. So, there remains one new aspect to discuss here, namely the case $r = 3$. As in the $r = 1$ case, there is no real advantage in introducing a standard form, since the rather restrictive structure of a quadratic in w invariant can here easily be derived in all generality. Indeed, assuming the invariant has the form

$$I = \tfrac{1}{2}(a\dot{x} + b\dot{y} + e)^2 + F(x,y,t),$$

where the differential equations would again be such that (29) holds true, the requirement $\dot{I} = 0$ produces the relation.

$$(a\dot{x} + b\dot{y} + e)(aF_1 + bF_2) + \dot{F} = 0, \tag{31}$$

or

$$adx + bdy + edt = u_2 du_1,$$

with

$$u_1 = F \text{ and } u_2 = -(aF_1 + bF_2)^{-1}.$$

But since we insist that the rank of the related matrix A be 3 here, a relation like (31) is not permitted, hence we are led to the conclusion that F must be zero. So for $r = 3$, our invariants can only be of the form

$$I = \tfrac{1}{2}(a\dot{x} + b\dot{y} + e)^2. \tag{32}$$

Clearly, there does not exist any kind of nonlinear superposition in this case and we conclude that systems with $r = 3$ are not Ermakovian.

C. Nonquadratic in w invariants

For the last case discussed above, any function of w is of course an invariant. We can equally drop the requirement that I be of the quadratic type $\frac{1}{2}w^2 + F$ in the other cases. This will in general entail additional derivative terms in what we called the functions f_i or F_i before ($i = 1,2$). In this subsection, we shall consider invariants which give rise to nonlinear superposition. That means we are looking for first integrals of the form

$$I = F(u,w), \tag{33}$$

u and w still being related by the formula

$$\dot{u} = fw. \tag{34}$$

Let us take the standard representation in the time-dependent case ($r = 2$), for which

$$w = x\dot{y} - y\dot{x} + x^2\dot{\alpha}, \quad u = y/x + \alpha(t). \tag{35}$$

As in (25), we again put

$$\dot{w} = xf_2 - yf_1, \tag{36}$$

but f_1 and f_2 can depend on $\dot{x}$ and $\dot{y}$ now. Requiring $\dot{I} = 0$ then yields

$$xf_2 - yf_1 = -fw\frac{\partial F}{\partial u}\cdot\left(\frac{\partial F}{\partial w}\right)^{-1}. \tag{37}$$

All examples with "velocity-dependent potentials," discussed by Ray and Reid[20] belong to this class of problems (see later). It is a straightforward matter to write down the similar extension of the general representation (28), (29), and (30).

Remark: One may wonder how reasonable the assumption at the beginning of this section was, that our quadratic invariants be of the form $\frac{1}{2}w^2 + F$, and for instance not contain a term linear in w. The relevance of this assumption (ii) lies in its connection with assumption (i). Indeed, suppose we want f_1 and f_2 in (36) to depend on coordinates only. Then obviously, the right-hand side of (37) must be independent of w, i.e., a function $g(u)$ must exist, such that

$$wF_u - g(u)F_w = 0.$$

This condition precisely requires that F be an arbitrary function of the quantity

$$\tfrac{1}{2}w^2 + \int g(u)\,du,$$

which shows that quadratic invariants are indeed of the "energy-type" previously considered.

D. Further generalization

As a final stage in the classification of more general cases, let us now disregard the possibility of nonlinear superposition. We should, of course, keep some link with the previous cases, and we therefore consider invariants of the form (see also Ref. 24)

$$I = G(w,x,y,t), \tag{38}$$

where w is related to some one-form in (x, y, t) space,

$$w = a(x,y,t)\dot{x} + b(x,y,t)\dot{y} + e(x,y,t). \tag{39}$$

Since, from the point of view of nonlinear superposition, this generalization is actually going too far (except for certain rather artificial looking circumstances), it is irrelevant to make a distinction between existence or nonexistence of an integrating factor for w here. For the same reason then, it is not very meaningful to reduce the underlying one-form to some standard form here. So, if G is going to be an invariant for some second-order system in x and y, and if for comparison with previous situations we again introduce functions $F_i(x,y,\dot{x},\dot{y},t)$, $(i = 1,2)$, determined by the differential equations in such a way that

$$\dot{w} = aF_1 + bF_2, \tag{40}$$

we must have

$$aF_1 + bF_2 = -(G_x\dot{x} + G_y\dot{y} + G_t)(G_w)^{-1}. \tag{41}$$

As in all previous cases, the only thing that really matters is a condition on a certain *combination* of the two differential equations. Hence, an infinite number of examples could be constructed, for instance by choosing F_1 completely arbitrary, and then defining F_2, by (41) for some function G. Yet, the type of combined conditions like (29) and (30), or (40) and (41) can really be used constructively in practical applications, especially when the structure of one of the equations is given, and the problem then consists in selecting an appropriate auxiliary equation such that the combined system gives rise to an invariant. The next two sections illustrate this. In those examples, where we distinguish between a primary variable (y), and an auxiliary one (x), we will go back to the notations of Sec. 2 and put $y \equiv q$, $x \equiv \rho$.

4. EXAMPLES

4.1. Let us first go back to the example treated in Sec. 2, and show how it actually can be recognized as belonging to the class of systems discussed in Sec. 3 A. The essential point is that we have a system of differential equations of the form

$$\begin{aligned} &\ddot{q} - \tfrac{1}{2}kq^{-1}\dot{q}^2 = F_1, \\ &\ddot{\rho} = F_2, \end{aligned} \tag{42}$$

where F_1 and F_2 are free of derivative terms. Looking at (19), we can see that (42) is indeed a candidate for fitting into the class 3 A because of the $\dot{q}^2$ term and the fact that no linear terms in $\dot{q},\dot{\rho}$ appear [if there were such terms, we would have to look at the class 3.2, Eq. (29)]. We identify y with q and x with ρ. Then (42) can be combined to a form like (19) if and only if

$$a_x = 0, \tag{43a}$$

$$a_y + b_x = 0, \tag{43b}$$

$$b^{-1}b_y = -\tfrac{1}{2}ky^{-1}. \tag{43c}$$

Eq. (43c) yields $b = \phi(x)y^{-k/2}$ for some function ϕ. Next, (43b) has the solution $a = -\phi'(x)(1 - k/2)^{-1}y^{1-k/2}$. Finally, Eq. (43a) requires ϕ' to be constant, and if we choose $\phi' = k/2 - 1$, we have precisely obtained, in a constructive way, the multipliers which were mentioned in Sec. 2. The functions a and b having been determined, condition (20) of course coincides with the requirement (6). Thus (42) furnishes an example of an $r = 2$ system without explicit time dependence in w.

4.2. Consider the Hamiltonian for a particle moving in a time-dependent potential

$$H = \tfrac{1}{2}p^2 + V(q,t). \tag{44}$$

The equation of motion is

$$\ddot{q} = -\frac{\partial V}{\partial q}(q,t), \tag{45}$$

and assume we want to find invariants, quadratic in $\dot{q}$, which may depend on some auxiliary function $\rho(t)$. Since no first-order derivatives occur in (45), and we do not mind having ρ dependence in possible auxiliary equations, we can look at the standard representation (25), (26) of Sec. 3 B (with $y \equiv q$, $x \equiv \rho$). Let us allow complete freedom for the auxiliary equation, i.e., we set

$$\ddot{\rho} = P(t,\rho,\dot{\rho}), \tag{46}$$

with P arbitrary. Then, from (25) and (26), we see that permissible q equations fitting into that particular class, must be of the form

$$\ddot{q} = \rho^{-1}Pq - 2\dot{\alpha}\dot{\rho} - \ddot{\alpha}\rho^2 - \rho^{-3}F'(\rho^{-1}q + \alpha(t)), \tag{47}$$

while a quadratic invariant then follows from (24),

$$I = \tfrac{1}{2}(\rho\dot{q} - \dot{\rho}q + \dot{\alpha}\rho^2)^2 + F(\rho^{-1}q + \alpha). \tag{48}$$

Eqs. (47) and (48) exactly coincide with the results recently derived by Lewis and Leach,[18] concerning the form of all

permissible potentials for which an equation like (45) has a quadratic in $\dot{q}$ invariant. We therefore conclude that all these cases with a quadratic invariant belong to the class of problems discussed in Sec. 3 B, for the case $r = 2$; that is, they are examples of $r = 2$ systems with explicit time dependence in w.

Remark: The auxiliary equation (46) is arbitrary, as long as we merely look at the structure of permissible q equations. Once we are given a particular q equation of that structure, and construct the invariant (48), the auxiliary equation will be completely determined by the given (time-dependent) coefficient of q in (47).

4.3. In their recent paper on Ermakov systems with "velocity-dependent potentials,"[20] Ray and Reid concluded with the following type of coupled equations [their Eqs. (3.1) and (3.4) with a slight change in notations],

$$\ddot{y} + \omega^2 y + \frac{\dot{w}}{x}\frac{\partial^2 F}{\partial w^2} = \frac{1}{x^3}\frac{\partial F}{\partial u} - \frac{w}{x^3}\frac{\partial^2 F}{\partial u\,\partial w}, \tag{49a}$$

$$\ddot{x} + \omega^2 x - \frac{\dot{w}}{y}\frac{\partial^2 G}{\partial w^2} = \frac{1}{y^3}\frac{\partial G}{\partial \bar{u}} - \frac{w}{y^3}\frac{\partial^2 G}{\partial \bar{u}\,\partial w}, \tag{49b}$$

where $w = x\dot{y} - y\dot{x}$, $u = y/x$, $\bar{u} = u^{-1}$, $F = F(u,w)$, $G = G(\bar{u},w)$.

Equation (49) is not a very transparent way of writing differential equations, since the second-order derivatives are not isolated in the first terms (additional ones occur in $\dot{w}$). Rather than computing here the explicit form of these equations, we can just look at their combination, which is the only thing that matters for our discussions. Computing $x\cdot$ (49a) $- y\cdot$ (49b) and putting $\bar{G}(u,w) = G(\bar{u},w)$ we get

$$(x\ddot{y} - y\ddot{x})\left(1 + \frac{\partial^2 F}{\partial w^2} + \frac{\partial^2 \bar{G}}{\partial w^2}\right) = \frac{1}{x^2}\left[\frac{\partial F}{\partial u} + \frac{\partial \bar{G}}{\partial u} - w\left(\frac{\partial^2 F}{\partial u\,\partial w} + \frac{\partial^2 \bar{G}}{\partial u\,\partial w}\right)\right]. \tag{50}$$

This can be written in the form

$$x\ddot{y} - y\ddot{x} = -\frac{1}{x^2}\, w\,\frac{\partial I}{\partial u}\left(\frac{\partial I}{\partial w}\right)^{-1}, \tag{51}$$

with

$$I = \frac{1}{2}w^2 - (F + \bar{G}) + w\frac{\partial}{\partial w}(F + \bar{G}). \tag{52}$$

Comparing (51) with (37), we see that these examples belong to the category of problems discussed in Sec. 3 C. The expression (52) for I of course coincides with the result of Ray and Reid.[20]

5. APPLICATION

For an illustration of the constructive use of the formulae we obtained in Sec. 3 D, we go back to particle motions of the type (45), and try to find possibly nonquadratic in w invariants. For the sake of brevity, we do not attack this problem in all its generality. Since no $\dot{y}$ terms occur in the equation of motion (45) ($y \equiv q$), it seems natural not to start with a general w as in (39), but rather to take one of the standard forms. Moreover, we limit ourselves here to invariants (38), which do not have an extra time-dependence. So we look for invariants of the form

$$I = G(w,x,y), \quad w = x\dot{y} - y\dot{x}. \tag{53}$$

The equations for x and y then have to satisfy the condition

$$x\ddot{y} - y\ddot{x} = -(G_x\dot{x} + G_y\dot{y})(G_w)^{-1}.$$

Choosing x to satisfy some arbitrary auxiliary equation $\ddot{x} = f_1(x,\dot{x},t)$, the y equation acquires the structure

$$\ddot{y} = x^{-1}f_1 y - x^{-1}(G_x\dot{x} + G_y\dot{y})(G_w)^{-1}. \tag{54}$$

Eq. (54) is compatible with the form (45) if and only if the last term is independent of $\dot{y}$. This produces the following requirement,

$$\mathcal{F}(w,\dot{x},\dot{y},x,y) = \dot{x}A + \dot{y}B + C = 0, \tag{55}$$

with

$$A = \frac{\partial}{\partial w}\left(\frac{G_x}{G_w}\right), \quad B = \frac{\partial}{\partial w}\left(\frac{G_y}{G_w}\right), \quad C = \frac{1}{x}\frac{G_y}{G_w}.$$

Computing the derivatives of (55) with respect to $\dot{x}$ and $\dot{y}$ we get

$$\left.\begin{aligned} A - y\mathcal{F}_w &= 0 \\ B + x\mathcal{F}_w &= 0 \end{aligned}\right\} \Rightarrow xA + yB = 0.$$

Hence, there must exist a function $f(w,x,y)$ such that

$$\frac{\partial}{\partial w}\left(\frac{G_x}{G_w}\right) = yf, \tag{56a}$$

$$\frac{\partial}{\partial w}\left(\frac{G_y}{G_w}\right) = -xf, \tag{56b}$$

and condition (55) then finally reads

$$(1/x)G_y/G_w = wf. \tag{56c}$$

Conditions (56) are necessary and sufficient for (54) to be compatible with Eq. (45) for some appropriate potential V. Setting

$$F(w,x,y) = \int f\,dw, \tag{57}$$

(56a) and (56b) integrate to

$$G_x = (g_1(x,y) + yF)G_w, \tag{58}$$

$$G_y = (g_2(x,y) - xF)G_w, \tag{59}$$

where g_1 and g_2 are arbitrary functions of x and y. Matching (59) with (56c) yields

$$F_w + (1/w)F = g_2/xw,$$

which provides us with the following structure of F:

$$F = g_2/x + h(x,y)/w, \quad h \text{ arbitrary}.$$

Substituting this F into (59), we obtain

$$G_y + (xh/w)G_w = 0. \tag{60}$$

Case I: $h = 0$. Then $G = G(x,w)$ and (58) becomes

$$G_x - (g_1 + (y/x)g_2)G_w = 0. \tag{61}$$

Since $G_w \neq 0$, this implies that $g_1 + (y/x)g_2$ must be a function of x and w, and therefore can only depend on x. If we set

$$g_1 + (y/x)g_2 = \phi(x), \quad \Phi = \int \phi\,dx,$$

the general solution of (61) is given by

$$G = \bar{G}(w + \Phi(x)), \tag{62}$$

while the structure of allowable y equations follows from (54) as

$$\ddot{y} = x^{-1} f_1 y - x^{-1}\phi(x). \tag{63}$$

We see that the y equation is linear, and that all invariants provided by (62) are just trivially functions of the linear in w first integral $w + \Phi(x)$.

Case II: $h \neq 0$. Setting $h = g_y$, the general solution of (60) now is

$$G = \bar{G}(x,v), \quad v = \tfrac{1}{2}w^2 - xg(x,y). \tag{64}$$

Equation (58) then becomes

$$\bar{G}_x - \bar{G}_v[g + xg_x + yg_y + (g_1 + (y/x)g_2)w] = 0. \tag{65}$$

The terms between square brackets must be expressible as a function of x and v, which for $g_y \neq 0$ necessarily implies $g_1 + (y/x)g_2 = 0$, and $g + xg_x + yg_y = \phi(x)$. This last equation has as general solution

$$g = x^{-1}[\psi(y/x) + \Phi(x)], \tag{66}$$

with $\Phi' = \phi$ and ψ arbitrary.
Equation (65) finally becomes

$$\bar{G}_x - \phi(x)\bar{G}_v = 0,$$

which implies $\bar{G} = \bar{\bar{G}}(v + \Phi)$, or taking account of (64) and (66),

$$\bar{G} = \bar{\bar{G}}(\tfrac{1}{2}w^2 - \psi(y/x)). \tag{67}$$

We see that also in this case, the nonquadratic in w invariants we find are trivial ones, i.e., are functions of a quadratic first integral. This seems to confirm similar negative conclusions following from other methods.[25]

6. SOME HIGHER-DIMENSIONAL EXAMPLES AND CONCLUDING REMARKS

For the case of two spatial dimensions we have introduced a classification scheme for Ermakov-type systems based on the reduction theorem for the one-form $a\,dx + b\,dy + e\,dt$. The case $r = 2$ corresponds to Ermakov-type systems and we have defined the classification scheme by relating to the original Ermakov structure as much as possible. Within the $r = 2$ case there are two important and *useful* standard forms; one being independent of t, the other one having explicit t dependence. Associated with each of them is a certain class of invariants. Of course, we obtain interesting generalizations of the standard Ermakov systems by transforming them to a more general structure via coordinate transformations.

After introducing the classification scheme we have discussed several examples and applications to illustrate their utility. Next we present a brief discussion of the extension to multiple dimensions.

Consider a one-form in $2n$ dimensions, $\Sigma_{k=1}^{2n} a_k(z)\,dz_k$, and suppose that as before this one-form somehow plays a distinctive role in the construction of a first integral for an associated system of second-order ordinary differential equations in z. Again, we could discuss the case of existence of an integrating factor, i.e., the case where the rank of the associated matrix A (see Sec. 2) would be two. It is clear, however, that for an originally $2n$-dimensional situation, the case $r = 2$ cannot be the most important one. In other words, pursuing some nonlinear superposition cannot be the primary objective here. Instead, there is a much richer potentiality for having quadratic invariants, in which so to speak the known two-dimensional standard forms are used as building blocks. Assume for instance that $r = 2n$, then the given one-form can be reduced to something like

$$\sum_{i=1}^{n} v_i\,du_i,$$

or equivalently,

$$\sum_{i=1}^{n} (x_i\,dy_i - y_i\,dx_i).$$

We thus can introduce the quantities

$$w_i = x_i\dot{y}_i - y_i\dot{x}_i, \quad u_i = y_i/x_i, \tag{68}$$

and it will be natural to study systems having quadratic invariants of the form

$$I = \tfrac{1}{2}\sum_{i,j=1}^{n} w_i\alpha_{ij}w_j + F(u_k). \tag{69}$$

There is even a possibility for matrices of quadratic invariants such as

$$I_{ij} = \tfrac{1}{2}w_iw_j + F_{ij}(u_i, u_j). \tag{70}$$

Let us show that the invariants for higher-dimensional systems, mentioned by Ray and Reid,[26] indeed belong to these classes of problems (see also Lutzky[22] for similar remarks). The coupled system

$$\begin{aligned} \ddot{q}_1 + \omega^2(t)q_1 &= q_1G(q_1q_2/\rho^2)/\rho^4, \\ \ddot{q}_2 + \omega^2(t)q_2 &= q_2G(q_1q_2/\rho^2)/\rho^4, \end{aligned} \tag{71}$$

has the invariant

$$I = \tfrac{1}{2}[(\rho\dot{q}_1 - \dot{\rho}q_1)(\rho\dot{q}_2 - \dot{\rho}q_2) + q_1q_2/\rho^2 + \psi(q_1q_2/\rho^2)], \tag{72}$$

with $\psi(\lambda) = -2\int^{\lambda} G(\lambda)\,d\lambda$, and ρ satisfying the equation

$$\ddot{\rho} + \omega^2(t)\rho = \rho^{-3}. \tag{73}$$

Clearly, this example is covered by (69), where $y_1 \equiv q_1$, $y_2 \equiv q_2$, $x_1 \equiv x_2 \equiv \rho$,

$$\alpha = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \text{ and } F(u_k) = \tfrac{1}{2}u_1u_2 + \tfrac{1}{2}\psi(u_1u_2).$$

Similarly, taking in (69) the unit matrix for α, and

$$F(u_k) = \tfrac{1}{2}\sum_{i=1}^{n} u_i^2 + \tfrac{1}{2}\phi\left(\sum_{i=1}^{n} u_i^2\right),$$

we get the invariant[26] for the system

$$\ddot{q}_i + \omega^2(t)q_i = q_iF(r^2/\rho^2)/\rho^4,$$

with $r^2 = \Sigma_{i=1}^{n} q_i^2$, ρ still satisfying (73), and $\phi(\lambda) = -\int F(\lambda)\,d\lambda$. Finally, let us recall that the n-dimensional time-dependent oscillator

$$\ddot{q}_i + \omega^2(t)q_i = 0$$

has a matrix of invariants[5] of type (70), namely,

$$I_{ij} = \tfrac{1}{2}(w_iw_j + u_iu_j),$$

where we again have identified y_i and q_i, and all x_i with ρ, satisfying (73).

In conclusion, if we consider a generalization of the two-dimensional classification scheme to more space dimensions, it seems plausible that the reduction theorem of one-forms will lead us to adding and coupling a number of lower-dimensional standard forms. In the sketch above, the rank was taken to be even ($2n$). If we want to catalog systematically all cases, however, including those of odd rank, the two-dimensional possibilities will not provide us with enough "building blocks," since we will need to use the first odd-dimensional case, namely, the case of three space dimensions. This will be the subject of a forthcoming paper.

ACKNOWLEDGMENTS

One of us (W.S.) would like to thank Peter Leach and Ralph Lewis for stimulating discussions on related topics, which have inspired some parts of this paper. He also is grateful to R. Mertens for his continuous support. J. R. R. thanks Morton Lutzky for an important correspondence and Jim Reid for interesting discussions.

[1]H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[2]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[3]M. D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[4]W. Sarlet, Ann. Phys. **92**, 232 (1975); **92**, 248 (1975).
[5]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).
[6]P. G. L. Leach, J. Math. Phys. **18**, 1608 (1977).
[7]M. Crespo da Silva, Int. J. Non-Linear Mech. **9**, 241 (1974).
[8]W. Sarlet, J. Phys. A **11**, 843 (1978).
[9]H. R. Lewis and P. G. L. Leach, Preprint LA-UR-81-606, Los Alamos National Laboratory (1981).
[10]J. R. Ray, J. Phys. A **13**, 1969 (1980).
[11]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979); Nuovo Cimento A **59**, 134 (1980).
[12]G. H. Katzin and J. Levine, J. Math. Phys. **18**, 1267 (1977).
[13]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[14]P. G. L. Leach, J. Math. Phys. **22**, 465 (1981).
[15]G. E. Prince and C. J. Eliezer, J. Phys. A **13**, 815 (1980).
[16]W. Sarlet and L. Y. Bahar, Int. J. Non-Linear Mech. **15**, 133 (1980).
[17]J. R. Ray and J. L. Reid, Lett. Math. Phys. **4**, 235 (1980).
[18]H. R. Lewis and P. G. L. Leach, preprint (1981).
[19]J. L. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980).
[20]J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[21]V. P. Ermakov, Univ. Izv. Kiew 20 (1880), No. 9, 1.
[22]M. Lutzky, Phys. Lett. A **78**, 301 (1980).
[23]W. Sarlet, Phys. Lett. A **82**, 161 (1981).
[24]J. R. Ray, J. L. Reid, and M. Lutzky, Phys. Lett. A **84**, 42 (1981).
[25]H. R. Lewis (private communication).
[26]J. R. Ray and J. L. Reid, Phys. Lett. A **74** 23 (1979).

Equivalent potentials in classical mechanics

Peter Collas
Department of Physics and Astronomy, California State University—Northridge, Northridge, California 91330

(Received 17 March 1981; accepted for publication 29 May 1981)

We show, within the framework of classical conservative system mechanics, that there exist local diffeomorphisms which connect Hamiltonians of different potentials. Thus, solutions for a given potential may be transformed into solutions for a different potential. Most of these results can be extended to quantum mechanics.

PACS numbers: 03.20. + i, 03.65.Sq, 95.10.Eg

I. INTRODUCTION

In 1941 Schrödinger[1] transformed the radial part of the Schrödinger equation for s identical one-dimensional oscillators into the radial equation for the Kepler problem. Subsequently, this equivalence was rediscovered by Dulock and McIntosh[2] and Rockmore.[3] It is, of course, well known that the harmonic oscillator and the Kepler problem are the only central fields in which all bound orbits are closed and that, furthermore, these orbits are ellipses.[4] In this light the above equivalence is not too surprising. It is much more interesting, however, to realize that this type of equivalence is not restricted to these two potentials.

We examined the situation within the framework of classical mechanics (conservative systems) and found, in the one-dimensional case, two types of transformations, while, in the two-dimensional case, we found only one type. These transformations are, in general, local diffeomorphisms generated by the potential function itself. They transform a given Hamiltonian into a new Hamiltonian with a new potential function, and consequently they can be used to transform solutions of one problem into solutions of another problem (e. g., orbits of one potential into orbits of another). Unfortunately, one cannot transform a given potential into any other; in fact there are severe constraints restricting the equivalences.

Apart from the rather obvious applications in classical and quantum mechanics our results may be of interest to mathematicians in the area of "differentiable dynamics"[5] in connection with the so-called "equivalence problem." Loosely speaking, the problem deals with effective ways of analyzing dynamical systems so as to decide whether two such systems are in a certain sense equivalent.[6] Specifically, one is concerned with one-to-one mappings of a phase flow into another. These mappings may be *linear automorphisms*, *diffeomorphisms*, and *homeomorphisms*, connecting systems (phase flows) which are respectively *linearly equivalent*, *differentiably equivalent*, and *topologically equivalent*.[5,7] The mappings considered here are a very special kind of diffeomorphism, say g, such that if $H = K + V$, then $g^*K = K$, and $g^*\omega = \omega$, where H is the Hamiltonian, K the kinetic and V the potential energy terms, while ω is a certain differential 1-form. Leaving the kinetic energy term(s) form-invariant makes the physical identification of the new (transformed) system simple. The invariance of ω is necessary in order for solutions to transform correctly. Because our systems are much more than simply differentiably equivalent, we would like to propose the term *differentiably *-equivalent* for such systems. In Sec. II we deal with the one-dimensional case and work out briefly three examples (one involving the WKB approximation). In Sec. III we deal with the two-dimensional (planar) case.

II. ONE-DIMENSIONAL CASE

We adopt the following terminology and notation[4,5]: Configuration space is a smooth n-dimensional manifold M (in this section $n = 1$). Phase space is the cotangent bundle T^*M. T^*M has a natural symplectic structure, viz., the 2-form $d\omega^1 = \Sigma_i dy_i \wedge dx_i$, where the y_i are the canonical momenta and the x_i the canonical coordinates. Together, the $2n$ numbers y_i, x_i form a collection of local coordinates for points in T^*M. The Hamiltonian H is a function on T^*M. Since for all our cases H is equal to the total energy E, we use the letter H for both throughout.

Consider the Hamiltonian

$$H = y^2 + \alpha u(x), \tag{1}$$

where we let the mass $m = \frac{1}{2}$; α is the coupling constant, and $u(x)$ the potential. We assume, for simplicity, that $u(x)$ is a nonnegative function of x; the theory can be modified to take into account changes of sign of $u(x)$.

We rewrite (1) as

$$-\alpha = \frac{y^2}{u(x)} - \frac{H}{u(x)}.$$

If we now define the transformation g by

$$\bar{y} = \pm y/[u(x)]^{1/2}, \tag{2}$$

$$\bar{H} = -\alpha, \tag{3}$$

$$\bar{x} = \bar{x}(x), \tag{4}$$

we obtain a new Hamiltonian $\bar{H}$ given by

$$\bar{H} = \bar{y}^2 + \bar{\alpha}\bar{u}(\bar{x}), \tag{5}$$

where

$$\bar{\alpha}\bar{u}(\bar{x}) = -H/u(x). \tag{6}$$

We can complete the definition of g in two different ways, which we shall refer to as *type A* and *type B* transformations.

Type A: We require that the form

$$\omega_A \equiv y\,dx = \bar{y}\,d\bar{x}. \tag{7}$$

 0022-2488/81/112512-06$01.00

Form ω_A is called Poincaré's relative integral invariant. Equations (2) and (7) imply

$$d\bar{x} = \pm [u(x)]^{1/2}\, dx, \tag{8}$$

or

$$\bar{x} = \pm \int [u(x)]^{1/2}\, dx. \tag{9}$$

As a consequence of (7) we have for the abbreviated action

$$S = \int y\, dx = \int \bar{y}\, d\bar{x} = \bar{S}, \tag{10}$$

where

$$\frac{\partial S}{\partial H} = t - t_0.$$

Also if we introduce the action variable I, we have

$$I = (1/2\pi)\oint y\, dx = (1/2\pi)\oint \bar{y}\, d\bar{x} = \bar{I}. \tag{11}$$

Equation (11) is particularly useful for the calculation of eigenvalues of a quantum mechanical system in the WKB approximation.

It follows from Eq. (7) that

$$d\omega_A = dy \wedge dx = d\bar{y} \wedge d\bar{x}, \tag{12}$$

and so type A transformations are local symplectic diffeomorphisms (canonical transformations).

Type B: In this case we require that the form

$$\omega_B \equiv \frac{dx}{y} = \frac{d\bar{x}}{\bar{y}}. \tag{13}$$

Form $\omega_B = 2dt$. Equations (2) and (13) imply

$$d\bar{x} = \pm \frac{dx}{[u(x)]^{1/2}}, \tag{14}$$

or

$$\bar{x} = \pm \int \frac{dx}{[u(x)]^{1/2}}. \tag{15}$$

As a consequence of (13) we have

$$\int \frac{dx}{y} = \int \frac{d\bar{x}}{\bar{y}}, \tag{16}$$

and hence $x(t) \leftrightarrow \bar{x}(\bar{t})$. It also follows from Eq. (13) that

$$d\omega_B = -(1/y^2) dy \wedge dx = -(1/\bar{y}^2) d\bar{y} \wedge d\bar{x}. \tag{17}$$

Clearly, type B transformations are local diffeomorphisms but not canonical transformations.

We remark that had we allowed $\bar{x} = \bar{x}(y,x)$ instead of (4), requirements (7) and (13) would have forced us back to (4). Thus it is clear that we always map phase portraits which are symmetric about the x axis into phase portraits which are symmetric about the $\bar{x}$ axis. The complexity of the mapping depends on the behavior of the potential $u(x)$. If the potential generating the transformation is a simple power, $u(x) = x^a$, then the new potential

$$\bar{u}(\bar{x}) \propto \bar{x}^{-2a/(a+2)}, \quad a \neq -2, \quad \text{for type A,} \tag{18}$$

while

$$\bar{u}(\bar{x}) \propto \bar{x}^{-2a/(2-a)}, \quad a \neq 2, \quad \text{for type B.} \tag{19}$$

We wish to emphasize, however, that the potentials need not be restricted to single terms or algebraic functions. In fact let us consider a generalization of Hamiltonian (1)

$$H = y^2 + \sum_i \alpha_i u_i(x), \quad \alpha_i \neq \alpha_j \text{ for } i \neq j. \tag{20}$$

Proceeding as before, we have for either type of transformation

$$-\alpha_j = \frac{y^2}{u_j(x)} - \frac{H}{u_j(x)} + \sum_{i \neq j} \alpha_i \frac{u_i(x)}{u_j(x)},$$

and hence

$$\bar{H} = \bar{y}^2 + \bar{\alpha}_j \bar{u}_j(\bar{x}) + \sum_{i \neq j} \alpha_i \frac{u_i[x(\bar{x})]}{u_j[x(\bar{x})]}. \tag{21}$$

Any term in the potential energy of Hamiltonian (20) may be used to generate two transformations (type A or B); but now this process may be continued using a different potential term from the resulting $\bar{H}$ to generate another transformation (of the same type), and so on. However it is easy to show, for the case where the u_i are powers, that these sequences of transformations ultimately lead us back to the original H.

As an example, consider the case where the potential energy in (20) is $v = \alpha_1 u_1(x) + \alpha_2 u_2(x) = \alpha_1 x^a + \alpha_2 x^b$, $\alpha_1 \neq \alpha_2$, $a,b \neq -2$. Using a shorthand notation for v and the resulting potential energies, along with arrows to indicate the terms generating the transformations, we have, for a type A sequence,

$$\begin{array}{l} v{:}x^a, x^b \\ \quad \updownarrow \\ \bar{v}{:}\bar{x}^{-2a/(a+2)},\ \bar{x}^{2(b-a)/(a+2)} \\ \leftrightarrow \qquad\qquad \updownarrow \\ \bar{\bar{v}}\bar{\bar{x}}^{-2b/(b+2)},\ \bar{\bar{x}}^{-2(b-a)/(b+2)} \\ \quad \updownarrow \\ v{:}x^b, x^a. \end{array}$$

It is obvious from the sequence above that $*$-equivalence relations are actually equivalence relations, i. e., that

$$v \sim v, \quad v \sim \bar{v} \Rightarrow \bar{v} \sim v, \quad v \sim \bar{v}, \ \bar{v} \sim \bar{\bar{v}} \Rightarrow v \sim \bar{\bar{v}}.$$

We give now three examples in order to illustrate these ideas.

Example 1: This is the "one-dimensional Kepler problem"[8]

$$H = y^2 + \alpha/|x|. \tag{22}$$

Under a type A transformation we have from (2) and (9)

$$\bar{y}^2 = |x| y^2, \quad |x| = \bar{x}^2/4, \tag{23}$$

where we set the integration constant to zero. Equations (3), (5), and (6) give us

$$\bar{H} = \bar{y}^2 + \bar{\alpha}\bar{x}^2, \tag{24}$$

$$\bar{H} = -\alpha, \quad \bar{\alpha} = -H/4. \tag{25}$$

Evaluating $\bar{S}$ for $\bar{H} > 0$, $\bar{\alpha} > 0$, we obtain

$$\bar{S} = \int_0^{\bar{x}} \bar{y}\, d\bar{x}$$
$$= \frac{(\bar{\alpha})^{1/2}}{2}\left\{\bar{x}\left(\frac{\bar{H}}{\bar{\alpha}} - \bar{x}^2\right)^{1/2} + \frac{\bar{H}}{\bar{\alpha}} \sin^{-1}\left[\bar{x}\left(\frac{\bar{\alpha}}{\bar{H}}\right)^{1/2}\right]\right\}, \tag{26}$$

and from (11) we have the action variable

$$\bar{I} = \bar{H}/2(\bar{\alpha})^{1/2}. \tag{27}$$

We can transform now $\bar{S} \to S$ and $\bar{I} \to I$, for the case $H < 0$, $\alpha < 0$, by direct substitution of (23) and (25) into (26) and (27). We obtain

$$S = (|\alpha| x - |H|x^2)^{1/2} + \frac{|\alpha|}{|H|^{1/2}} \sin^{-1}\left[\left(\frac{x|H|}{|\alpha|}\right)^{1/2}\right] \tag{28}$$

and

$$I = |\alpha|/|H|^{1/2}. \tag{29}$$

Equations (28) and (29) are what one would obtain by performing the integrations

$$S = \int_0^x y\, dx \quad \text{and} \quad I = (1/2\pi)\oint y\, dx.$$

The equations of motion may be obtained from (26) and (28) using

$$\frac{\partial \bar{S}}{\partial \bar{H}} = \bar{t} - \bar{t}_0, \quad \frac{\partial S}{\partial H} = t - t_0, \tag{30}$$

respectively.

Under a type B transformation we have from (2) and (15)

$$\bar{y}^2 = |x| y^2, \quad |x| = \tfrac{3}{2}\bar{x}^{2/3}, \tag{31}$$

and so this time Hamiltonian of (22) is $*$-equivalent to

$$\bar{H} = \bar{y}^2 + \bar{\alpha}\bar{x}^{2/3}, \tag{32}$$

where

$$\bar{H} = -\alpha, \quad \bar{\alpha} = -\tfrac{3}{2}H. \tag{33}$$

Using (28) and (30), we can obtain $t = t(x)$ for $u(x) = 1/|x|$, $H < 0$, $\alpha < 0$; then by substitution of (31) and (33) we can obtain $t = t(\bar{x})$ for $\bar{u}(\bar{x}) = \bar{x}^{2/3}$, and $\bar{H} > 0$, $\bar{\alpha} > 0$. Because in this example we cannot solve explicitly for $x = x(t)$, we shall not carry these calculations further (see, however, Example 2).

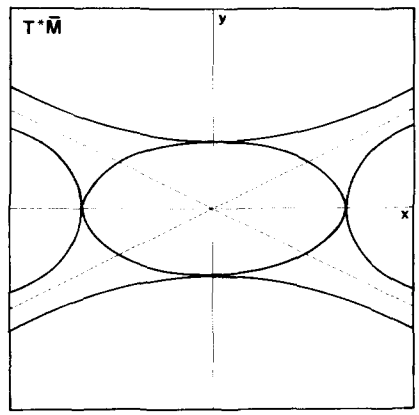


FIG. 2. Phase space of Hamiltonian (24), i. e., image of T^*M (Fig. 1) under the mapping (23).

Before leaving this example, it is interesting to see how the various families of phase curves transform under type A and B mappings. In Fig. 1 we show the phase space T^*M of (22). We have included phase curves for the cases (a) $H > 0$, $\alpha > 0$, (b) $H > 0$, $\alpha < 0$, and (c) $H < 0$, $\alpha < 0$. In Fig. 2 we show how T^*M maps under the type A transformation (23), while in Fig. 3 we show how T^*M maps under the type B transformation (31).

Example 2: Consider the Hamiltonian

$$H = y^2 + \alpha/x^2. \tag{34}$$

We define a local type A diffeomorphism as follows:

$$\bar{y}^2 = x^2 y^2, \tag{35}$$

$$\bar{x} = \int_\epsilon^x \frac{dx}{x}, \quad x \geqslant \epsilon, \ \bar{x} \geqslant 0, \tag{36}$$

$$\bar{x} = -\int_x^{-\epsilon} \frac{dx}{|x|}, \quad x \leqslant -\epsilon, \ \bar{x} \leqslant 0, \tag{37}$$

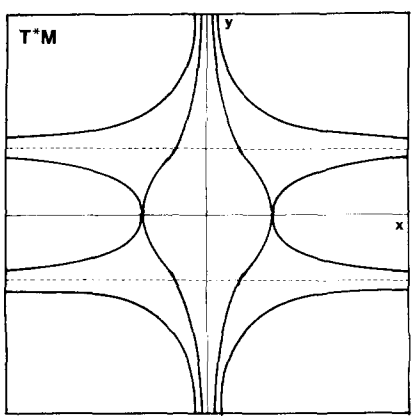


FIG. 1. Phase space of Hamiltonian (22). Shown are representative curves from the three families (a) $H > 0$, $\alpha > 0$, (b) $H > 0$, $\alpha < 0$, and (c) $H < 0$, $\alpha < 0$.

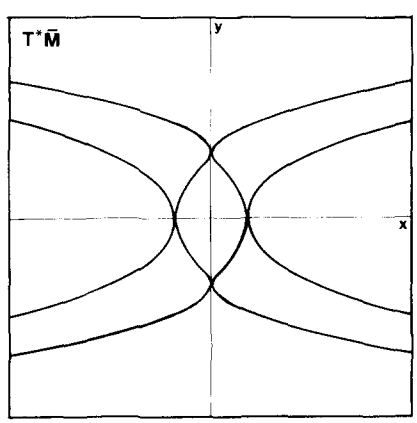


FIG. 3. Phase space of Hamiltonian (32), i. e., image of T^*M (Fig. 1) under the mapping (31).

where, in both (36) and (37), $\epsilon > 0$. Equations (36) and (37) give us

$$|\bar{x}| = \log(|x|/\epsilon), \quad |x| \geqslant \epsilon, \tag{38}$$

or, equivalently,

$$|x| = \epsilon e^{|\bar{x}|}. \tag{39}$$

Therefore the Hamiltonian of (34) is $*$-equivalent to

$$\bar{H} = \bar{y}^2 + \bar{\alpha} e^{2|\bar{x}|}, \tag{40}$$

where

$$\bar{H} = -\alpha, \quad \bar{\alpha} = -\epsilon^2 H. \tag{41}$$

For the case $\bar{H} < 0$, $\bar{\alpha} < 0$ we have, using (40),

$$\begin{aligned} \bar{S} = \int_0^{\bar{x}} \bar{y}\, d\bar{x} &= (|\bar{\alpha}| e^{2\bar{x}} - |\bar{H}|)^{1/2} - (|\bar{\alpha}| - |\bar{H}|)^{1/2} \\ &\quad - |\bar{H}|^{1/2} \tan^{-1}\left(\frac{|\bar{\alpha}|}{|\bar{H}|} e^{2\bar{x}} - 1 \right)^{1/2} \\ &\quad + |\bar{H}|^{1/2} \tan^{-1}\left(\frac{|\bar{\alpha}|}{|\bar{H}|} - 1 \right)^{1/2}. \end{aligned} \tag{42}$$

Now let $\epsilon = (\alpha/H)^{1/2}$; then using (39) and (41) to transform (42) back to the unbarred variables, we obtain

$$S = (Hx^2 - \alpha)^{1/2} - \alpha^{1/2} \tan^{-1}(Hx^2/\alpha - 1)^{1/2}. \tag{43}$$

Equation (43) is, of course, what we would have obtained by carrying out the integration

$$S = \int_{(\alpha/H)^{1/2}}^{x} y\, dx,$$

using Hamiltonian (34) with $H > 0$, $\alpha > 0$.

For later use we shall obtain the equation of motion from (43). We find

$$x = (\alpha/H + 4Ht^2)^{1/2}. \tag{44}$$

Let us now perform the type B diffeomorphism generated by the potential $u = 1/x^2$ of Eq. (34). With appropriate choice of signs and constants we have

$$\bar{y}^2 = x^2 y^2, \quad |\bar{x}| = x^2/2. \tag{45}$$

Under this transformation the Hamiltonian of (34) is $*$-equivalent to

$$\bar{H} = \bar{y}^2 + \bar{\alpha}|\bar{x}|, \tag{46}$$

where

$$\bar{H} = -\alpha, \quad \bar{\alpha} = -2H. \tag{47}$$

We choose $\bar{H} < 0$, $\bar{\alpha} < 0$ and obtain the equation of motion of (46) by direct integration of

$$2\bar{t} = \int \frac{d\bar{x}}{\bar{y}};$$

we find,

$$\bar{x} = |\bar{H}|/|\bar{\alpha}| + |\bar{\alpha}|\bar{t}^2, \tag{48}$$

for $\bar{x} > 0$. Using (45), and (47) to transform (48) back to the unbarred variables, we obtain (with $\bar{t} = t$) the equation of motion (44).

Example 3: As a final example we shall deduce the energy eigenvalues of the Hamiltonian[9,10]

$$H = y^2 + \alpha/\cosh^2 x, \tag{49}$$

from those of its $*$-equivalent Hamiltonian

$$\bar{H} = \bar{y}^2 + \bar{\alpha}\tan^2\bar{x}, \quad -\pi/2 < \bar{x} < \pi/2, \tag{50}$$

using the WKB (quasiclassical) approximation of the Schrödinger equation.

For a type A diffeomorphism we have

$$\bar{y}^2 = y^2 \cosh^2 x,$$

$$\bar{x} = \pm \int \frac{dx}{\cosh x} = \pm \tan^{-1}(\sinh x), \tag{51}$$

or, equivalently,

$$\sinh x = \pm \tan\bar{x}, \quad -\pi/2 < \bar{x} < \pi/2.$$

Thus we write

$$\cosh^2 x = 1 + \tan^2\bar{x}. \tag{52}$$

We note that (51) and (52) map the entire phase portrait of (49) onto the strip $-\pi/2 < \bar{x} < \pi/2$ of the $(\bar{x},\bar{y})$ plane.

The identifications

$$\bar{H} = H - \alpha, \quad \bar{\alpha} = -H, \tag{53}$$

along with (51) and (52) transform (49) into (50). Now we evaluate

$$\bar{I} = \frac{1}{2\pi} \oint \bar{y}\, d\bar{x} = \frac{2}{\pi} \int_0^{\tan^{-1}(\bar{H}/\bar{\alpha})^{1/2}} (\bar{H} - \bar{\alpha}\tan^2\bar{x})^{1/2}\, d\bar{x}, \qquad \bar{H} > 0, \ \bar{\alpha} > 0,$$

and obtain

$$\bar{I} = (\bar{H} + \bar{\alpha})^{1/2} \pm (\bar{\alpha})^{1/2}. \tag{54}$$

The $(\pm)$ ambiguity arises because we had to continue around a logarithmic branch point in performing the integration. At any rate, using the quantization rule $\bar{I} = \hbar(n + \frac{1}{2})$ and solving for $\bar{H}$ in (54), we obtain the energy eigenvalues in the WKB approximation for (50),

$$\bar{E}_n = \bar{H} = \hbar^2(n + \tfrac{1}{2})^2 \pm 2(\bar{\alpha})^{1/2}\hbar(n + \tfrac{1}{2}). \tag{55}$$

At this point the sign ambiguity is resolved; we have to choose the $(+)$ sign since otherwise large $\bar{\alpha}$ would give negative $\bar{E}_n$!

Going back to (54) and substituting (53), we immediately obtain

$$I = |\alpha|^{1/2} \pm |H|^{1/2}, \qquad H < 0, \ \alpha < 0, \ |\alpha| > |H|$$

which gives us

$$|H| = (I - |\alpha|^{1/2})^2.$$

Finally the quantization rule $I = \hbar(n + \frac{1}{2})$ gives us

$$E_n = -H = -\hbar^2\left[|\alpha|^{1/2}/\hbar - (n + \tfrac{1}{2}) \right]^2, \tag{56}$$

which is the result obtained in Ref. 10 for this problem.

III. TWO-DIMENSIONAL CASE

This is, of course, the central force problem. Because of angular momentum conservation, particle trajectories are planar and the configuration space M two-dimensional. We consider the Hamiltonian

$$H = y_1^2 + \frac{y_2^2}{x_1^2} + \alpha u(x_1), \tag{57}$$

which takes a more familiar appearance with the identifications $y_1 = p_r$, $y_2 = p_\phi$ (the angular momentum), $x_1 = r$, $x_2 = \phi$, ($m = \frac{1}{2}$, and $u(x_1) \geqslant 0$ again for simplicity).

We rewrite (57) as

$$-\alpha = \frac{y_1^2}{u(x_1)} + \frac{y_2^2}{x_1^2 u(x_1)} - \frac{H}{u(x_1)}.$$

We define our diffeomorphism by

$$\bar{y}_1 = \pm y_1/[u(x_1)]^{1/2}, \quad \bar{y}_2/\bar{x}_1 = \pm y_2/x_1[u(x_1)]^{1/2}, \tag{58}$$

$$\bar{H} = -\alpha,$$

$$\bar{\alpha}\bar{u}(\bar{x}_1) = -\frac{H}{u(x_1)};$$

furthermore, we require that the form [cf. Eq. (7)],

$$\omega = y_1\,dx_1 + y_2\,dx_2 = \bar{y}_1\,d\bar{x}_1 + \bar{y}_2\,d\bar{x}_2. \tag{59}$$

Equations (58) and (59) imply

$$d\bar{x}_1 = \pm\,[u(x_1)]^{1/2}\,dx_1, \tag{60}$$

and

$$\frac{y_2}{\bar{y}_2} = \frac{d\bar{x}_2}{dx_2} = \pm\,\frac{x_1[u(x_1)]^{1/2}}{\bar{x}_1} = \pm\beta, \tag{61}$$

where $\beta \neq 0$, a real constant.

This is a type A transformation. Type B is not possible because the conserved angular momentum term spoils things.

In order to check the self-consistency of Eqs. (60) and (61), we eliminate $[u(x_1)]^{1/2}$ between them, obtaining the differential equation

$$\frac{d\bar{x}_1}{dx_1} = \pm\,\frac{\bar{x}_1}{x_1}\beta.$$

The solution of this equation is $\bar{x}_1 = \text{const}\, x_1^{\pm\beta}$. Substituting this result back in (60), we find that

$$u(x_1) \propto x_1^{\pm 2\beta - 2},$$

therefore, things are consistent *only for power potentials.*

We shall write the resulting $*$-equivalent Hamiltonian as

$$\bar{H} = \bar{y}_1^2 + \frac{\bar{y}_2^2}{\bar{x}_1^2} + \bar{\alpha}\bar{u}(\bar{x}_1), \tag{62}$$

where, if $u(x_1) = x_1^a$, then

$$\bar{u}(\bar{x}_1) \propto \bar{x}_1^{-2a/(a+2)}, \quad a \neq -2. \tag{63}$$

We note that the potential $u(x_1) = x_1^{-2}$ cannot generate a transformation, since the diffeomorphism of Example 2 is inconsistent now.

The generalization corresponding to Eqs. (20) and (21), with the subsequent comments, holds for the present case also.

It follows from (59) that

$$d\omega = dy_1 \wedge dx_1 + dy_2 \wedge dx_2 = d\bar{y}_1 \wedge d\bar{x}_1 + d\bar{y}_2 \wedge d\bar{x}_2,$$

so again our mapping is a local canonical transformation.

It is easy to show that the orbits $x_1 = x_1(x_2)$ of system (57), do transform into the orbits $\bar{x}_1 = \bar{x}_1(\bar{x}_2)$ of system (62). From

$$y_1 = \frac{1}{2}\frac{dx_1}{dt} \quad \text{and} \quad y_2 = \frac{1}{2}x_1^2\frac{dx_2}{dt},$$

we obtain

$$y_1 = \frac{y_2}{x_1^2}\frac{dx_1}{dx_2}. \tag{64}$$

Substituting (64) into (57), we obtain a differential equation for the orbit, involving the potential $u(x_1)$,

$$H = \frac{y_2^2}{x_1^2}\left[\frac{1}{x_1^2}\left(\frac{dx_1}{dx_2}\right)^2 + 1\right] + \alpha u(x_1). \tag{65}$$

Now

$$x_1 = x_1(\bar{x}_1) = x_1[\bar{x}_1(\bar{x}_2)] = x_1[\bar{x}_1\{\bar{x}_2(x_2)\}];$$

therefore,

$$\frac{dx_1}{dx_2} = \frac{dx_1}{d\bar{x}_1}\frac{d\bar{x}_1}{d\bar{x}_2}\frac{d\bar{x}_2}{dx_2},$$

and, using (60) and (61), we have

$$\frac{dx_1}{dx_2} = \frac{x_1}{\bar{x}_1}\frac{d\bar{x}_1}{d\bar{x}_2}. \tag{66}$$

Substituting (66) in (65) and using (58), we obtain the differential equation for the orbit of the $*$-equivalent system, viz.,

$$\bar{H} = \frac{\bar{y}_2^2}{\bar{x}_1^2}\left[\frac{1}{\bar{x}_1^2}\left(\frac{d\bar{x}_1}{d\bar{x}_2}\right)^2 + 1\right] + \bar{\alpha}\bar{u}(\bar{x}_1).$$

Thus, for example, the Kepler conics

$$x_1 = \frac{-2y_2^2/\alpha}{1 - (1 + 4y_2^2 H/\alpha^2)^{1/2}\cos x_2},$$

under the transformation

$$x_1 = \bar{x}_1^2/4, \quad x_2 = 2\bar{x}_2, \quad y_2 = \bar{y}_2/2,$$

$$\alpha = -\bar{H}, \quad H = -4\bar{\alpha},$$

become the harmonic oscillator orbits[11]

$$\bar{x}_1^2 = \frac{2\bar{y}_2^2/\bar{H}}{1 - (1 - 4\bar{\alpha}\bar{y}_2^2/\bar{H}^2)^{1/2}\cos 2\bar{x}_2}.$$

These orbits are also conics whose center is at the origin ($\bar{x}_1 = 0$).[12]

These ideas can easily be extended into quantum mechanics. However, only the type A diffeomorphisms can be done on the Schrödinger equation, since coordinate transformations automatically induce canonical transformations in that framework.

ACKNOWLEDGMENTS

The author wishes to thank Professor John K. Lawrence, Professor Robert T. Park, and Professor Ryoichi Seki, for helpful and stimulating discussions.

[1] E. Schrödinger, Proc. Roy. Irish Acad. **46**, 183 (1941).
[2] V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39 (1966).
[3] R. Rockmore, Amer. J. Phys. **43**, 29 (1975).
[4] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978), see p. 38.

[5]S. Smale, *The Mathematics of Time* (Springer, New York, 1980).
[6]L. Markus, Trans. Amer. Math. Soc. **76**, 127 (1954); *Lectures in Differentiable Dynamics* (Amer. Math. Soc. Providence, R. I., 1971), Vol. 3.
[7]V. I. Arnold, *Ordinary Differential Equations* (MIT Press, Cambridge, Mass., 1973), see p. 141.
[8]For a very interesting paper dealing with the n-dimensional Kepler problem, see J. Moser, Comm. Pure Appl. Math. **23**, 609 (1970). Among other things he discusses the equivalence of the Kepler problem with the geodesic flow on the unit sphere.
[9]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1977), 3rd ed.
[10]I. I. Gol'dman, V. D. Krivchenkov, V. I. Kogan, and V. M. Galitskii, *Problems in Quantum Mechanics* (Academic, New York, 1960).
[11]See, for example, E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., New York, 1965), 4th ed., p. 82.
[12]It is amusing to recall at this point the intriguing fact that the potential outside a uniform spherically symmetric mass distribution is the Kepler, while inside we have the harmonic oscillator.

Bäcklund transformations connecting different isospectral deformation equations

D. V. Chudnovsky
CEN-Saclay, 91190 Gif-sur-Yvette, France[a]

G. V. Chudnovsky
CNRS, Université de Paris Sud, Orsay, France[a]

(Received 31 December 1980; accepted for publication 15 May 1981)

We study singular spectral problems on Riemann surfaces and isospectral deformation equations connected with them. We show that for a given class of reductions all spectral problems on a given Riemann surface give rise to the same class of isospectral deformation equations. These isospectral deformation equations can be solved using the Riemann boundary value problem. All possess local conservation laws and Bäcklund transformations.

PACS numbers: 03.40.Kf

INTRODUCTION

In this article we are trying to put together two different approaches to the so-called two-dimensional isospectral deformation equations.[1–6] These are equations that are connected with a spectral, possibly singular, problem on a Riemann surface Γ for a linear differential operator with matrix function coefficients. This linear spectral problem can be written in its general form as

$$\frac{d}{dx}\Phi(x,\lambda) = U(x,\lambda)\Phi(x,\lambda), \tag{0.1}$$

where $U(x,\lambda)$ is $n\times n$ matrix whose coefficients are rational functions of λ with singularities at fixed points $\lambda_1^0,\ldots,\lambda_m^0$ of fixed multiplicities $r_1,\ldots,r_m$. Sometimes the spectral problem (0.1) is considered in the so-called reduced form,[5] when one imposes algebraic restrictions on (0.1). They may be of two types: (a) conditions are imposed on $U(x,\lambda)$ for it to belong to a certain Lie algebra B of $n\times n$ matrices and (b) symmetry properties of eigenfunctions $\Phi(x,\lambda)$ of the form $\Phi(\lambda)\to G.\Phi(\tau(\lambda))$ for nonsingular G connected with the transformation $\tau(\lambda)$ of Γ, are generated, say, by a conformal transformation of $\mathbb{C}$.

In both reduced and nonreduced cases the spectral problem (0.1) gives rise to a large class of two-dimensional isospectral deformation equations according to the following two existing procedures. One of them generates a class of commuting Hamiltonians using the monodromy or scattering matrix of (0.1). The most efficient approach in this direction is connected with the factorized S-matrix method[7,8] and Baxter technique[9] and takes as a generating function for the isospectral deformation equations associated with (0.1) the quantities

$$\{\mathrm{tr}\Phi(x,x_0,\lambda)\} \tag{0.2}$$

for a fundamental solution $\Phi(x,x_0,\lambda)$ of (0.1) with norming condition $\Phi(x_0,x_0,\lambda) = \mathrm{I}$. Along these lines one expands the quantities like $(\partial/\partial\lambda)\log \mathrm{tr}\Phi(x,x_0,\lambda)$ at some point $\lambda = \lambda_0$ (usually, $\lambda_0 = 0$ or ∞) and gets an infinite family of Hamiltonians generating two dimensional completely integrable systems.[10] In the usual case of the regular spectral problem (0.1), when $U(x,\lambda)$ has a simple pole on ∞, $U(x,\lambda) = \lambda\mathcal{A} + U$, this infinite family of two-dimensional systems possesses a so called "Lax representation".[3,6] In other words, there are two linear spectral problems, the consistency condition of which defines a two-dimensional nonlinear equation. This is the second method for introduction of two-dimensional isospectral deformation equations, extensively discussed in.[2,3,5,6,11] In its most general setting this approach to two-dimensional equations means that one adds a new variable t and considers simultaneously two linear spectral problems for $\Phi(\lambda)$ on Γ: one, the main, is (0.1) and the second, called auxiliary, is of the form

$$\frac{d}{dt}\Phi(x,t,\lambda) = V(x,t,\lambda)\Phi(x,t,\lambda) \tag{0.3}$$

for $V(x,t,\lambda)$ being a rational function of λ or Γ with poles at $\lambda_1^{0'},\ldots,\lambda_{m'}^{0'}$ of multiplicities $r_1',\ldots,r_{m'}'$. The potential $U(\lambda)$ is now a function of x and t and the consistency condition of (0.1) and (0.3) takes the form of a two-dimensional system of equations:

$$\frac{d}{dt}U(x,t,\lambda) - \frac{d}{dx}V(x,t,\lambda) + [U(x,t,\lambda),V(x,t,\lambda)] = 0. \tag{0.4}$$

The equation (0.4) should be understood as a system of λ-independent equations on residues of $U(x,t,\lambda)$ at $\lambda = \lambda_i^0$ and on residues of $V(x,t,\lambda)$ at $\lambda = \lambda_i^{0'}$. Naturally, systems (0.4) are characterized by (i) positions and multiplicities of singularities of both $V(\lambda)$ and $U(\lambda)$; (ii) their reductions, i.e., the algebraic restrictions on $U(\lambda)$, $V(\lambda)$, and symmetries of $\Phi(x,t,\lambda)$.

These two approaches, one based solely on the spectral problem (0.1) and another one requiring two spectral problems (0.1) and (0.3) are indeed connected. Moreover, what we are trying to show, is the equivalence of both of them. Namely, there is no single spectral problem (0.1), and all of the equations (0.4) are, in fact, commuting Hamiltonians (within the same class of reductions and the same definition of Poisson brackets, naturally). We claim that for a fixed Riemann surface Γ and for a fixed class of reductions, one can consider all two-dimensional isospectral deformation equations ei-

[a]Present address: Department of Mathematics, Columbia University, New York, N.Y. 10027.

 0022-2488/81/112518-05$01.00

ther arising from (0.2), or written in the form (0.4) as commuting dynamical flows. This point of view has been in a different form emphasized in our earlier paper.[12]

I. COMMUTATIVITY OF DIFFERENT SPECTRAL PROBLEMS

The fact that different spectral problems (0.1) give rise to the same class of two-dimensional isospectral deformation equations can be illustrated on several known examples. For this we need some convention on the notations. From our point of view the variable x in (0.1) or the variable t in (0.3) is attached to the singularity data

$$\begin{pmatrix} \lambda_1^0,\ldots,\lambda_m^0 \\ r_1,\ldots,r_m \end{pmatrix}.$$

If $m = 1$ this variable is denoted at $t_{a,r}$ corresponding to a pole $\lambda = a$ of multiplicity r. The classical inverse scattering problem (0.1) takes the form

$$\frac{\partial}{\partial t_{\infty,1}} \Phi_\lambda = (\lambda \mathscr{A} + U)\Phi_\lambda. \tag{1.1}$$

The function Φ_λ has only one essential singularity on $\mathbb{PC}^1$ at $\lambda = \infty$, where the leading irregular part of Φ_λ is $\exp(\lambda \mathscr{A} t_{\infty,1})$. Moreover, following[13] one considers instead of (1.1) a more general family of commuting spectral problems, where the variable $t_{\infty,1}$ depends on a diagonal matrix $\mathscr{A}: t_{\infty,1} = t_{\infty,\mathscr{A},1}$:

$$\frac{\partial}{\partial t_{\infty,\mathscr{A},1}} \Phi_\lambda = (\lambda \mathscr{A} + U)\Phi_\lambda. \tag{1.2}$$

For example, one can consider two-dimensional equations of the form (0.4) that are consistency conditions of two linear problems (1.2) with different $\mathscr{A}$ and $\mathscr{A}'$. Another spectral problem that can be considered as commuting with (1.1) or (1.2) arises when one allows an irregular singularity of higher rank at $\lambda = \infty$ of Φ_λ:

$$\Phi_\lambda \sim \exp(\lambda \mathscr{A} t_{\infty,\mathscr{A},1} + \lambda^2 \mathscr{B} t_{\infty,\mathscr{B},2} + \cdots + \lambda^k C t_{\infty,C,k})[\mathbb{I} + O(\lambda^{-1})]$$

at $\lambda = \infty$. One gets Φ_λ in (1.1), (1.2) from this general Φ_λ, if one fixes $t_{\infty,\mathscr{B},i}$ for $i \geqslant 2$. The linear problem corresponding to this general structure of Φ_λ at $\lambda = \infty$ is of the form

$$\frac{\partial}{\partial t_{\infty,C,k}} \Phi_\lambda = (\lambda^k C + \sum_{j=0}^{k-1} U_{\infty,j} \lambda^j)\Phi_\lambda. \tag{1.3}$$

For example, one can write consistency conditions for linear problems (1.2) and (1.3) as two-dimensional linear equations on coefficients of U and $U_{\infty,j}: j = 0,\ldots,k-1$. These two-dimensional equations are naturally associated with the expansion of the monodromy matrix (0.2) of the linear problem (1.2) at $\lambda = \infty$. The consistency conditions for (1.2) and (1.3) generate the famous two-dimensional equations solvable via the inverse scattering technique (IST).[2,6] The most known example is the case of $n = 2$ and $\mathscr{A} = \sigma_3$. Then from (1.2) and (1.3) one gets the nonlinear Schrödinger equation, modified KdV equation, and their higher analogs. However, since all spectral problems (1.2), (1.3) are consistent, we can compose two of the linear problems (1.3) with different C's and k's. For example, for $k \geqslant 2$, C may be taken as $\mathbb{I}$. In this case we obtain the so called Gel'fand–Dikij equations[3] having Lax pair representation and canonical Hamiltonian structure. These equations, however, should be considered as commuting with ordinary isospectral deformation equations arising from (1.2). Indeed, since we don't have a distinguished variable t or x and all our variables commute,[14] all two-dimensional nonlinear equations arising as consistency conditions of (1.1), (1.2), and (1.3) are commuting. From this point of view a single linear problem (1.1) itself generates all the higher Gel'fand–Dikij equations in the sense that these equations arise as conservation laws for Hamiltonians generated by an expansion of the scattering matrix (1.1).

Example 1.4: This example deals with the higher KdV equations from.[1-3] The nth KdV equation can be written in a canonical way as,[3]

$$u_{t_n} = \frac{\partial}{\partial x} \frac{\delta}{\delta u} R_n[u], \tag{1.5}$$

so $t_0 = x$ and $R_n[u]$ is a polynomial in $u, u', u'',\ldots,u^{(2n-1)}$ and in (1.5) one puts $u^{(m)} = \partial^m/\partial x^m u$. However, from our point of view, when all variables are equal, $u^{(m)}$ is not necessarily $\partial^m/\partial x^m u$. In other words we can introduce the KdV equation of the type (n_1,n_2) in the following way. If

$$\frac{\partial}{\partial x} \frac{\delta}{\delta u} R_n[u] = R'_n(u,u',\ldots,u^{(2n-1)})$$

and

$$\frac{\partial^k}{\partial x^k} R'_n = R'_{n,k}(u,u',\ldots,u^{(2n-k-1)}),$$

then (n_1,n_2)th KdV in (y,t) variables can be written as $u_t = R'_{n_1}(u,u',\ldots,u^{(2n_1-1)})$, $u_y = R'_{n_2}(u,\ldots,u^{(2n_2-1)})$, $(u')_t = R_{n_1,1}(u,u',\cdots)$, $(u')_y = R_{n_2,1}(u,u',\cdots),\cdots$.

This is the system of $2n' - 1 = 2\min\{n_1,n_2\} - 1$ equations on $u,u',\ldots,u^{(2n'-1)}$ variables.

So far we considered only analytic properties of Φ_λ at $\lambda = \infty$. However one can consider Φ_λ as multivalued function on the whole Γ with the structure of regular and irregular singularities governed by monodromy matrices exponential matrices and Stokes multipliers.[12,15] For example, one can consider an expansion of the scattering matrix for (1.1) at $\lambda = 0$ or at any other point $\lambda = a \neq \infty$. This way we obtain a new spectral problem

$$\frac{d}{dt_0} \Phi_\lambda = \frac{U_{0,1}}{\lambda} \Phi_\lambda \tag{1.6}$$

or

$$\frac{d}{dt_a} \Phi_\lambda = \frac{U_{a,1}}{\lambda - a} \Phi_\lambda. \tag{1.7}$$

Equation (1.7) corresponds to the case, when Φ_λ has an essential singularity at $\lambda = a$. If one considers the consistency conditions (0.4) for two linear problems (1.1) and (1.5) one gets an interesting class of two-dimensional isospectral deformation equations, generalizing sine-Gordon equation.

Example 1.8: The two-dimensional Toda lattice (which includes sine-Gordon equation) arises as the consistency condition of Eq. (1.5) and (1.1), if one imposes certain reduc-

tions on Φ_λ. These reductions are the following:

$$\frac{\partial}{\partial x}\Phi_\lambda = \begin{pmatrix} u_1 & \lambda 0 & \cdots & 0 \\ 0 & u_2\lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda & 00 & \cdots & u_n \end{pmatrix}\Phi_\lambda, \tag{1.9}$$

$$\frac{\partial}{\partial t}\Phi_\lambda = \frac{1}{\lambda}\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ V_1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ V_n & 0 & 0 & V_{n-1} & 0 \end{pmatrix}\Phi_\lambda.$$

The equation of consistency (1.9) is equivalent to the two-dimensional Toda lattice.

Along the same lines one can consider consistency conditions of the form (0.4) for (1.1) and (1.7). The corresponding equation has the form

$$\frac{\partial}{\partial t_a} U_{\infty,1} + [\mathscr{A}, U_a] = 0,$$
$$-\frac{\partial}{\partial t_{\infty,\mathscr{A}}} U_{a,1} + [a\mathscr{A} + U_{\infty,1}, U_{a,1}] = 0, \tag{1.10}$$

where $\mathscr{A}$ is t_a — independent.

The system (1.10) contains, in particular, different versions of the two-dimensional Toda lattice. Naturally, instead of combining (1.7) with (1.1) one can combine two of the equations (1.7) with different a's:

$$\frac{d}{dt_{a_1}}\Phi_\lambda = \frac{U_{a_1}}{\lambda - a_1}\Phi_\lambda, \quad \frac{d}{dt_{a_2}}\Phi_\lambda = \frac{U_{a_2}}{\lambda - a_2}\Phi_\lambda \tag{1.11}$$

for $a_1, a_2 \in \mathbb{C}$, $a_1 \neq a_2$. The eigenfunctions Φ_λ satisfying (1.10) have in general two irregular singularities of the rank one at $\lambda = a_1, \lambda = a_2$.

The consistency condition (0.4) for (1.11) has the form[5-16] of principal chiral field equations

$$\frac{d}{dt_{a_2}} U_{a_1} = \frac{1}{a_1 - a_2}[U_{a_1}, U_{a_2}], \tag{1.12}$$
$$\frac{d}{dt_{a_1}} U_{a_2} = \frac{1}{a_1 - a_2}[U_{a_1}, U_{a_2}].$$

In particular, different reductions of (1.12) contain different principal chiral field models and different nonlinear σ-models for different matrix Lie algebras. In order to write them in a more familiar form,[4,16] usually the following representation for the potentials U_{a_i} is used

$$U_{a_i} = \frac{\partial}{\partial t_{a_i}} g \cdot g^{-1}. \tag{1.13}$$

From our point of view, when variables corresponding to different singularity data are treated in the same way, the system of nonlinear equations (1.10) is considered as the dynamical flow commuting with (1.12). Namely, (1.10) defines a conservation law for (1.12), if $t_a = t_{a_1}$, the Hamiltonian depends on derivatives with respect to t_a and time variables are $t_{\infty,\mathscr{A},1}$ and t_{a_2}.

The list of examples can be continued, e.g., our discussion shows how σ-model or principal chiral field equations (1.12) are connected through the linear problem (1.1) with equations of sine–Gordon type and, simultaneously, with higher-order KdV-type equations. The reason for this connection, as we see it, is the observation that two-dimensional isospectral deformation equations arising from a given linear spectral problem [say (1.7)] are connected with another linear spectral problem [say (1.11)] and, as the result, can be derived from entirely different spectral problems as, e.g., nonsingular spectral problem (1.1).

II. BÄCKLUND TRANSFORMATIONS VIA GAUGE TRANSFORMATIONS

The examples examined in Sec. 1 show that relations between different two-dimensional isospectral deformation equations are given by relations between different linear problems, when there exist common eigenfunctions Φ_λ of different spectral problems with different structure of singularities. To certain extent it is wrong to say that linear spectral problem (0.1) determines the structure of singularities of $\Phi(\lambda)$. Namely, eigenfunctions $\Phi(\lambda)$ of (0.1) should be considered up to a generalized gauge transformations $\Phi(\lambda) \to G(\lambda)\cdot\Phi(\lambda)\cdot C(\lambda)$. This gauge transformation may change the structure of singularities of $\Phi(\lambda)$ and the structure of potential $U(\lambda)$ in (0.1). For example, in the case of linear problem (1.1), (1.2), or (1.3) one has essential singularities of Φ_λ at $\lambda = \infty$ of different ranks: for (1.1) and (1.2) of rank 1; for (1.3) of rank k. Nevertheless Φ_λ in (1.2) and Φ_λ in (1.3) are essentially the same, since in order to get from canonical Φ_λ in (1.2):

$$\Phi_\lambda = D(\lambda)\exp\{\lambda, \mathscr{A} t_{\infty,\mathscr{A},1}\}, \tag{2.1}$$

[where $D(\lambda)$ is analytic at $\lambda = \infty$] the canonical Φ'_λ for (1.3), one should multiply Φ_λ in (2.1) by $t_{\infty,\mathscr{A},1}$-independent quantity $\exp\{\lambda^k C t_{\infty,C,k} + \ldots\}$ and to take into account $t_{\infty,C,k}$ dependence of $D(\lambda)$. In other words, in order to pass from one spectral problem (0.1) to another, one performs a gauge transformation of eigenfunctions. This way we are able to establish relations between different isospectral deformation equations and to show that, basically, within a given class of reductions all two-dimensional isospectral deformation equations arise from a single spectral problem. These gauge transformations of solutions of (0.1) can be at the same time used for construction of Bäcklund transformations for isospectral deformation equations. For this we apply those gauge transformations of solutions of (0.1) that do not change the structure of singularities of $U(x,\lambda)$. We can call this gauge transformation self-Bäcklund transformation. In particular, these self-Bäcklund transformations generate local conservation laws for isospectral deformation equations. If one considers a system with reduction, then there again are self-Bäcklund transformations that are compatible with reduction. They furnish Bäcklund transformations for reduced isospectral deformation equations.

We are going to describe the effect of gauge transformation on eigenfunctions. The gauge transformations we consider are usually λ-dependent. By a gauge transformation we understand a transformation of eigenfunctions $\Phi(\lambda)$ of (0.1) of the form

$$\Phi(\lambda) \to \Phi_1(\lambda) = G(\lambda)\Phi(\lambda)C(\lambda), \tag{2.2}$$

where $G(\lambda)$ is x dependent nonsingular $n \times n$ matrix and

$C(\lambda)$ is x-independent $n\times n$ matrix.

Multiplication by $G(\lambda)$ does change the form of (0.1) and new eigenfunction $\Phi_1(\lambda)$ satisfies the following differential equation:

$$\frac{d\Phi_1(\lambda)}{dx} = U_1(x,\lambda)\Phi_1(\lambda), \tag{2.3}$$

where

$$U_1(x,\lambda) = G(\lambda)U(x,\lambda)G(\lambda)^{-1} + G_x(\lambda)G(\lambda)^{-1}. \tag{2.4}$$

For λ independent transformation $G(\lambda)$ (2.2) is an ordinary gauge transformation.[4,16] Naturally, for λ-dependent $G(\lambda)$ the transformation (2.2) changes the singularities of (0.1). The most important class of λ-dependent gauge transformation (2.2) is the class of transformations that single out irregular singularities of $\Phi(\lambda)$. For this we use a very simple boundary value problem in the following fashion.

Lemma 2.5: Let $\Phi(\lambda)$ be a matrix valued function being an analytic function of λ on a Riemann surface Γ everywhere but at points $\lambda = \lambda_1^0,...,\lambda_m^0$, where $\Phi(\lambda)$ have essential or apparent singularities. Then for any λ_i^0 there exists a nonsingular matrix $G_i(\lambda)$ regular in the neighborhood of $\lambda = \lambda_i^0$ such that the function

$$\Phi_i(\lambda) = G_i(\lambda)\Phi(\lambda) \tag{2.6}$$

is analytic everywhere on Γ but at $\lambda = \lambda_i^0$.

Proof: We consider the Riemann boundary value problem on Γ taking for a closed curve Δ a small circle around λ_i^0. Then we formulate the Riemann boundary value problem in the following form:

$$\psi_+(\lambda) = \psi_-(\lambda)\Phi(\lambda) \quad \text{on } \lambda\in\Delta, \tag{2.7}$$

where $\psi_+(\lambda)$ is regular outside of Δ and $\psi_-(\lambda)$ is regular inside Δ. If problem (2.7) is a regular Riemann boundary value problem without zeroes,[17] then the solution $\psi_-(\lambda) = G_i(\lambda)$ of (2.7) is unique up to a norming

$$\Phi_1(\lambda_0) = \mathbb{I} \tag{2.8}$$

for an arbitrary λ_0 different from $\{\lambda_i^0\}$.

Remark 2.9: One can consider the Riemann boundary value problem (2.7) with zeroes.[17] This corresponds to the case when only essential singularities of $\Phi(\lambda)$ different from λ_i^0 are cancelled, while poles are preserved. In physical language this corresponds to solutions with a soliton part, since poles of $\Phi(\lambda)$ that provide no contribution to the poles of $U(\lambda)$ are usually associated with the soliton part.

Lemma 2.5 shows that, starting from a singular spectral problem with several singularities, one can always reduce it to a spectral problem with eigenfunction having only one singularity. Namely, let us assume that $\Phi(\lambda)$ satisfies equation (0.1), where $\Phi(\lambda)$ has singularities at $\lambda = \lambda_i^0: i = 1,...,m$, that are the singularities of $U(\lambda)$. We perform the gauge transformation (2.6) satisfying the conditions of Lemma 2.5. Then the eigenfunctions $\Phi_i(\lambda)$ satisfy a linear differential equation

$$\frac{d\Phi_i(\lambda)}{dx} = U_i(x,\lambda)\Phi_i(\lambda) \tag{2.10}$$

with $U_i(x,\lambda) = G_i(\lambda)U(x,\lambda)G_i(\lambda)^{-1} + G_{ix}(\lambda)G_i(\lambda)^{-1}$.

Since $\Phi_i(\lambda)$ is regular everywhere but at $\lambda = \lambda_i^0$ the function $U_i(x,\lambda)$ is regular everywhere but at $\lambda = \lambda_i^0$. Moreover, since $G_i(\lambda)$ is analytic (i.e., invertible) at $\lambda = \lambda_i^0$, the structure and multiplicity of singularity of $U_i(x,\lambda)$ at $\lambda = \lambda_i^0$ is the same as that of $U(x,\lambda)$ at $\lambda = \lambda_i^0$.

Consequently, using Lemma 2.5, one reduces an arbitrary spectral problem (0.1) with m singularities on a Riemann surface Γ to m separate spectral problems on Γ, each having only one singularity. Relations between these spectral problems are given by gauge transformations determined by the Riemann boundary value problem.

Moreover, we can proceed further and obtain gauge transformations that relate spectral problems (2.10) having one fixed singularity but of different ranks. For this one writes $U_i(\lambda)$ in the form

$$U_i(\lambda) = \sum_{j=0}^{r_i} U_{i,j}(\lambda - \lambda_i^0)^{-j}, \tag{2.11}$$

for $\lambda\to\lambda_i^0$. First one performs a gauge transformation that reduces U_{i,r_i} to its normal form. This way the leading part of irregular singularity is obtained. Let C_i be a normal form of U_{i,r_i}, which we assume to be a diagonal matrix with distinct entries (this is the so called nonreduced case). Let $U_{i,r_i} = M_i C_i M_i^{-1}$. We apply to $\Phi_i(\lambda)$ a gauge transformation

$$\Phi_i(\lambda)\to\Phi_i'(\lambda) = M_i\Phi_i(\lambda),$$

which is a λ-independent gauge transformation. In this case there is a canonical eigenfunction $\Phi_i(\lambda)$ whose form as $\lambda\to\lambda_i^0$ is the following:

$$\Phi_i'(\lambda) = D_i(\lambda)\exp\left[\frac{C_i x}{(\lambda-\lambda_i^0)^{r_i}}\right], \tag{2.12}$$

where $D_i(\lambda)$ has essential singularity at $\lambda = \lambda_i^0$ of rank less than r_i. Naturally, we change a variable x corresponding to C_i by a system of variables $x_{C_i,l}$ arising from $C_i = \operatorname{diag}(C_{i,l}: l = 1,...,n)$. The function $D_i(\lambda)$, which is a gauge transformation of $\Phi_i'(\lambda)$ [and so of $\Phi_i(\lambda)$], satisfies now Eq. (2.10) with $U_i'(\lambda)$ instead of $U_i(\lambda)$ and having in its expansion (2.11) at $\lambda = \lambda_i^0$ a pole of order at most $r_i - 1$ now. This way we reduce spectral problem (2.10) to a problem with a simple pole only. In the process of this reduction one sees that the role of external variable x can be played by an arbitrary system of spectral data built from: (a) positions of points $\lambda = \lambda_i^0$ on Γ; (b) diagonal matrices C_i characterizing normal forms of exponential matrices of regular singularities at $\lambda = \lambda_i^0$. These data are usually referred to in isomonodromy deformation theory[13,15,18,19] as varying monodromy data of a Riemann module.

III. "DRESSING OF THE VACUUM"

Let us show now how one can combine eigenfunctions corresponding to different varying monodromy data into a single eigenfunction. We explained above how a spectral problem with several singularities can be decoupled into several singular spectral problems each having only one singularity. Here we perform the opposite operation and show how a single gauge transformation can be simultaneously applied to several spectral problems.

Lemma 3.1: Let $\Phi_i(\lambda)$ be solutions of m spectral problems

$$\frac{d}{dt_i}\Phi_i(\lambda) = U_i(\lambda)\Phi_i(\lambda), \tag{3.2}$$

where $U_i(\lambda)$ have singularity on Γ only at $\lambda = \lambda_i^0$ and t_i correspond to λ_i^0, $i = 1,...,m$. Let $\lambda_1^0,...,\lambda_m^0$ be all distinct on Γ. Then there is an eigenfunction $\Phi(\lambda)$ regular everywhere but at $\lambda = \lambda_i^0$ and having in the neighborhood of $\lambda = \lambda_i^0$ the form

$$\Phi(\lambda) = \overline{G}_i(\lambda)\Phi_i(\lambda), \tag{3.3}$$

where $\overline{G}_i(\lambda)$ is analytic at $\lambda = \lambda_i^0$, $i = 1,...,m$. In other words $\Phi(\lambda)$ is a simultaneous gauge transformation of all $\Phi_i(\lambda)$. The function $\Phi(\lambda)$ satisfies linear differential equations

$$\frac{d\Phi(\lambda)}{dt_i} = \overline{U}_i(\lambda)\Phi(\lambda), \quad i = 1,...,m. \tag{3.4}$$

Proof: One considers (3.2) as a Riemann boundary value problem on Γ for a system of m closed curves Δ_i being small circles around distinct points λ_i^0, $i = 1,...,m$. Each of the equations (3.3) is considered on Δ_i only, where $\Phi(\lambda)$ is taken to be a regular function outside of all Δ_i, while each $\overline{G}_j(\lambda)$ is analytic inside Δ_j.

Corollary 3.5: Let us assume in the conditions of Lemma 3.1 that all functions $U_i(\lambda)$ are arbitrary t_i dependent functions, that are t_j independent for all $j \neq i$. Then the function $\Phi(\lambda)$ satisfies Eq. (3.3) with $\overline{U}_i(\lambda)$ having a singularity only at $\lambda = \lambda_i^0$ and of the same structure and multiplicity as $V_i(\lambda)$.

In particular, from Corollary 3.4 it follows that simple solutions of arbitrary isospectral deformation Eq. (0.4) can be constructed starting with arbitrary t independent $U(x,\lambda)$ and arbitrary x independent $V(t,\lambda)$. Similar construction was proposed in Refs. 5 and 16. It was called the "dressing of the vacuum." We see, however, that not only vacuum can be "dressed" but the Riemann boundary value problem allows us to write down a general solution for an arbitrary isospectral deformation equation that depends on a proper number of initial data. This work was supported in part by the Office of Naval Research.

[1] N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15,** 240 (1965).
[2] A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proc. IEEE **61,** 1443 (1973).
[3] I. M. Gel'fand, and L. A. Dikij, Russ. Math. Survey **20** (5), 77 (1975); Funct. Anal. Appl. **10** (1), 18 (1976).
[4] D. V. Chudnovsky and G. V. Chudnovsky, Z. Phys. D **5,** 55 (1980).
[5] V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **13** (3), 13 (1979).
[6] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53,** 249 (1974).
[7] L. D. Fadeev, L. Takhtadjan, and E. Sklanin, Theor. Math. Phys. **40** (3), (1979); E. K. Sklyanin, LOMI Preprint E-3-1979, 1979.
[8] D. V. Chudnovsky and G. V. Chudnovsky, Lett. Math. Phys. **4,** 485 (1980).
[9] R. Baxter, Ann. Phys. **76,** 1, 25, 48 (1973).
[10] One defines the Hamiltonian structure and Poisson brackets for these systems in such a way that the quantities (0.2) all commute. There are infinitely many Poisson brackets defined by (0.1). One of them, so called local commutation relation,[7] is defined using semiclassical S-matrix.[8] For the discussion of nonlocal commutation relation relations see Ref. 3. This method is connected with coadjoint orbits for symplectic structure.
[11] D. V. Chudnovsky, Phys. Lett. A **74,** 185 (1979).
[12] J. A. Lappo-Danilevsky, *Mémoires sur la théorie des systèmes des équations différentielles linéaires* (Chelsea, New York, 1950).
[13] D. V. Chudnovsky, in *Bifurcation Phenomena in Mathematical Physics and Related Topics: Proceedings*, edited by Claude Bardos and Daniel Bessis (Kluwer Boston, Hingham, Mass., 1980), pp. 387–447; Les Houches Lectures, August, 1979, *Lecture Notes in Physics* (Springer, New York, 1980),Vol. 126, pp. 352–416.
[14] The commutativity of all linear problems of the type (1.1), (1.2), (1.3) etc. follows from the existence of common solution Φ_λ, which is itself determined by its irregular part $\exp(\lambda\mathcal{A}t_{\infty,\mathcal{A},1} + \cdots + \lambda^k Ct_{\infty,C,k})$ and Stokes multipliers at $\lambda = \infty$.
[15] D. V. Chudnovsky and G. V. Chudnovsky, Phys. Lett. A **80,** 1 (1980).
[16] V. E. Zakharov and A. V. Mikhailov, Zh. Tech. Phys. **74,** 1953 (1978).
[17] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Groningen, The Netherlands, 1953).
[18] D. V. Chudnovsky and G. V. Chudnovsky, J. Math. Pure Appl. (to appear).
[19] M. Sato, T. Miwa, M. Jimbo, Publ. Res. Inst. Sci. **14,** 223 (1978).

Inverse problem for the reduced wave equation with fixed incident wave. II

V. H. Weston
Purdue University, West Lafayette, Indiana 47907

(Received 15 July 1980; accepted for publication 28 August 1980)

The reduced wave equation $\Delta u + k^2 n^2(x)u = 0$ for x in R^3 is considered where $[n^2 - 1]$ is a real function of compact support given by the domain D. Sources outside D produce an incident field u^i (satisfying the equation with $n \equiv 1$ for all x). This generates a scattered field $u^s(x,n)$ satisfying the reduced wave equation and the radiation condition. The scattered field is measured at a set of points $\{x_j\}_{j=1}^N$ outside D. If n_* is a known quantity for which $u^s(x,n_*)$ can be computed, the inverse problem considered here, consists of solving the nonlinear system of functional equations $u^s(x_j,n) - u^s(x_j,n_*) = b_j, j = 1,2,...,N$ for an $n(x)$ such that $\int_D (n^2 - n_*^2)^2 W\,dx$ is minimized. With restrictions on the data set $\{b_j\}_{j=1}^N$, a unique solution can be generated in a certain ball by iterating a system of nonlinear integral equations. It is shown that $n(x)$, with less stringent conditions than that of continuity, may be treated.

PACS numbers: 03.40.Kf

1. INTRODUCTION

The inverse problem for the reduced wave equation

$$\Delta u + k^2 n^2(x)u = 0, \quad x \in R^3 \tag{1}$$

associated with time dependence $\exp(-i\omega t)$ was examined[1] for the case where the measurements of the scattered field u^s (produced by an incident field u^i generated by a fixed source) were obtained at a finite number of points $\{x_j\}_{j=1}^N$. The index of refraction $n(x)$ was taken to be a real continuous function equal to unity outside some bounded region D_0 contained in a compact domain D (representing the scattering object). With $n_*(x)$ representing in *a priori* estimate of $n(x)$ and with

$$v(x) = n^2(x) - 1 \tag{2}$$

the set of scattered field measurements at the points $\{x_j\}_{j=1}^N$ outside D was represented by the system of N nonlinear equations with unknown $v(x)$,

$$u(x_j) - u_*(x_j) = b_j, \quad j = 1,2,...,N. \tag{3}$$

The complex numbers b_j correspond to the difference of the measured values of the total field $u(x_j) = u(x_j,v)$ [associated with the unknown value of $n(x)$ or $v(x)$] and the calculated values of the total field $u_*(x_j) = u(x_j,v_*)$ [associated with the index of refraction $n_*(x)$].

Here the results of the previous work,[1] henceforth known as Part I will be extended as follows.

We shall seek the solution $n(x)$ of Eqs. (1) and (3) that is closest to the *a priori* estimated value $n_*(x)$ and that lies in a certain domain of function space (to be specified later). Explicitly we shall impose the condition

$$\text{Min} \int_D (n^2 - n_*^2)^2 W\,dx, \tag{4}$$

where W is some continuous positive weight function bounded from zero. A very simple approximate version of this was employed in Part I (with W taken to be $|u_*|^4$). The condition will give us the desired uniqueness result.

We will also enlarge the class of functions to which $n(x)$ belongs by removing the continuity restriction given by assumption (i) of Part I and replace it by

(i') *$n(x)$ is real, bounded and is continuous everywhere except for a finite number of piecewise continuous surfaces across which $n(x)$ is discontinuous.*

With this extension, condition (ii) of Part I may be simplified by taking $D_0 = D$. [In Part I, D_0 was taken to be inside D so that a factor $\eta(x)$ could be added to expressions (24) and (25) of Part I so as to yield a solution $v(x)$ which vanishes on the boundary of D and thus is continuous across the boundary.]

It should be noted that with the introduction of the discontinuities on $n(x)$ the solution of Eq. (1) will still be required to be C^1 function everywhere except possibly at the source points of the incident field.

The *a priori* value $n_*(x)$ will now satisfy condition (i'). The additional assumption (iii) that was imposed on it in Part I will be modified. In place of it we take

(iii) *n_* is such that the Green's function $G(x,y,v_*)$ exists and the total field u_* is bounded from zero for x in $\bar{D}$.*

The nonsingular portion $\tilde{G}$ arising from the decomposition of the Green's function

$$G(x,y,v_*) = G_0(x,y) + \tilde{G}(x,y,v_*),$$

where

$$G_0(x,y) = e^{ik|x-y|}/4\pi|x-y|,$$

still satisfies the following integral equation given in Part I:

$$\tilde{G}(x,y,v_*) = k^2 \int_D G_0(x,y')G(y',y,v_*)v_*(y')\,dy'.$$

From this one can deduce that $G(x,y,v_*)$ still has the property that the integral operator with kernal $G(x,y,v_*)$, maps squared integrable functions over $\bar{D}$ into continuous functions over $\bar{D}$. This fact will be used later in the proofs.

2. REDUCTION OF THE INVERSE PROBLEM

The procedure for reducing the inverse problem to a more tractable mathematical from is initially similar to that given in Part I. These initial steps are briefly described as follows.

With the application of the Green's function the set of nonlinear functional equations involving the data has the

 0022-2488/81/112523-07$01.00

explicit form

$$k^2\int_D G(x_j,y,v_*)[v(y)-v_*(y)]u(y,v)\,dy = b_j, \tag{5}$$

with $j = 1,2,...,N$, and where the total field $u(x,v)$ satisfies the integral equation

$$u(x,v) = u(x,v_*) + k^2\int_D G(x,y,v_*) \times [v(y)-v_*(y)]u(y,v)\,dy. \tag{6}$$

With the introduction of the real functions $\varphi(x)$ and $\psi(x)$ through the relation

$$[v(x)-v_*(x)]u(x,v) = [\varphi(x)+i\psi(x)]/\bar{u}_*(x), \tag{7}$$

the complex nonlinear functional equations (5) reduce to the following system of $2N$ real linear function equations in φ and ψ

$$\int_D H_i(y)\varphi(y)\,dy = \sum_{j=1}^{2N} e_{ij}\int_D H_j(y)\psi(y)\,dy + B_i, \tag{8}$$

with $i = 1,2,...,2N$. Here B_i correspond to the real and imaginary parts of the data b_i by the relation

$$b_j = B_j + iB_{j+N}, \quad j = 1,2,...,N, \tag{9}$$

and H_j corresponds in a similar manner to the real and imaginary parts of h_j, where

$$h_j(y) = k^2 G(x_j,y,v_*)/\bar{u}(y,v_*), \tag{10}$$

with x_j corresponding to the points where the measurements were made. The numbers e_{ij} are related to the Kronecker delta

$$e_{ij} = \begin{cases} \delta_{(i+N)j}, & i = 1,2,...,N, \\ -\delta_{i(j+N)}, & i = N+1,...,2N. \end{cases} \tag{11}$$

By noting that $n(x)$ and hence $v(x)$ is a real quantity, integral equation (6) is decomposed to yield two real expressions. One expression relates φ and ψ through a quadratic integral equation which is given by

$$\psi = S(\psi,\varphi),$$
$$S(\psi,\varphi) = \psi(L_I\psi - L_R\varphi) + \varphi(L_R\psi + L_I\varphi), \tag{12}$$

where L_R and L_I are the real and imaginary parts of the integral operator with kernel

$$k^2 G(x,y,v_*)[u_*(x)\bar{u}_*(y)]^{-1}, \quad x,y\in D.$$

The second expression

$$v(x) - v_*(x) = \varphi(x)/|u_x|^2[1 + L_R\varphi - L_I\psi], \tag{13}$$

relates $v(x)$ to φ and ψ and allows one to recoup the value of $n(x)$ from knowledge of φ and ψ.

The inverse problem thus is reduced to solving Eqs. (8) and (12) for φ and ψ subject to the constraint that the integral $\int_D (v-v_*)^2 W\,dx$ be minimized, and then recovering $n(x)$ from Eq. (13).

It should be noted that in Part I an approximate version of condition (4) was used. There W was taken to be $|u_*|^4$ and the zeroth approximation (from a perturbation standpoint of view) was used for $v - v_*$ namely $v - v_* \sim \varphi/|u_*|^2$, hence the condition that was employed was that $\int_D \varphi^2\,dx$ be minimized.

The first step in the solution of system (8), (12), and (4), is to solve Eq. (8) for φ in terms of ψ. The procedure that is used here is similar to that of Part I, but because of the employment of a general weight function W some of the resulting expressions will be slightly different from that given in Part I.

Assumption (iv) of Part I, namely: $\{H_j\}_{j=1}^{2N}$ *from a linearly independent set*, will be imposed. Hence define the matrix $H = \{H_{ij}\}$ whose coefficients are given by

$$H_{ij} = \int_D H_i(y)H_j(y)W_1^{-1}(y)\,dy, \tag{14}$$

where

$$W_1 = W/|u_*|^4, \tag{15}$$

and let $\tilde{H} = \{\tilde{H}_{ij}\}$ be the inverse matrix

$$\tilde{H}H = H\tilde{H} = I. \tag{16}$$

Then set

$$\xi_i(x) = \sum_{j=1}^{2N} \tilde{H}_{ij}H_j(x)/W_1(x), \tag{17}$$

for $i = 1,2,...,2N$.

The solution of system (8) is then given by

$$\varphi(x) = \Phi(x) + \mathbb{K}\psi + \varphi^{\perp}(x), \tag{18}$$

where

$$\Phi(x) = \sum_{j=1}^{2N} B_j\xi_j(x), \tag{19}$$

$$\mathbb{K}\psi = \sum_{j,k=1}^{2N} \xi_j(x)\int_D e_{jk}H_k(y)\psi(y)\,dy, \tag{20}$$

and $\varphi^{\perp}$ is an unknown function which is perpendicular to the space spanned by $\{H_j\}_{j=1}^{2N}$ i.e., $\int_D \varphi^{\perp}H_j\,dx = 0$.

It should be noted that ξ_j, Φ, and $\mathbb{K}$ differ from the corresponding terms ξ_j, Φ_M, and $\mathbb{K}_M$ of Part I only in a factor W_1.

The advantage of the new form for these expressions is that we have the results

$$\int_D \Phi\varphi^{\perp}W_1\,dy = 0,$$
$$\int_D (\mathbb{K}\psi)\varphi^{\perp}W_1\,dy = 0, \tag{21}$$

which will simplify some of the later analysis. From Eqs. (14) and (17) it can be seen that

$$\int_D \xi_i(x)H_j(x)\,dx = \delta_{ij}.$$

From this it can be shown that the second iterate $\mathbb{K}^2$ has the explicit form

$$-\mathbb{K}^2\psi = \mathbb{P}\psi = \sum_{j=1}^{2N} \xi_j(x)\int_D H_j(y)\psi(y)\,dy,$$

were $\mathbb{P}$ is the projection operator on the subspace spanned by $\{\xi_j\}$. It has the property that

$$\|\mathbb{P}\|_w = \sup\left[\int_D (\mathbb{P}u)^2 W_1\,dx \Big/ \int_D u^2 W_1\,dx\right]^{1/2} = 1,$$

where the norm is that associated with the operator mapping

the space of squared integral functions with measure $W_1\,dx$ over $\bar{D}$ into itself. When $W_1 = 1$, the norm is just the usual squared integral norm $\|\cdot\|_2$.

It is thus seen that with $W_1 = 1$, $\|\mathbb{K}^2\|_2 = 1$. However $\|\mathbb{K}\|_2$ is not necessarily unity as was erroneously stated in the lemmas in Part I. The correct statement is that the spectral radius $\|\mathbb{K}\|_{sp} = 1$. The estimates given the Theorem 1 and 2 of Part I are valid when $\|\mathbb{K}\|_2 \leqslant 1$. For the case $\|\mathbb{K}\|_2 > 1$, the estimates have to be modified. However it should be pointed out that since the estimates on conditions for convergence are quite crude, the important fact is the demonstration of unique solution of the systems, and that nonlinear correction terms can be obtained to the linearized inverse problem.

With expression (18) inserted for φ into Eq. (12) it is seen that the inverse problem now reduces to finding ψ and $\varphi^{\perp}$ satisfying the nonlinear integral equation (12) such that the integral given by expression (4) is minimized.

More precisely, the resulting problem can be stated as follows,

Problem P: Find the real valued functions ψ and $\varphi^{\perp}$ that lie in the ball $\|\psi\|_2 + \|\varphi^{\perp}\|_2 < r$ (where the number r is to be determined so as to yield a unique solution) such that $\varphi^{\perp}$ is perpendicular to the space spanned by $\{H_j\}_{j=1}^{2N}$, and which satisfies the equation

$$\psi = S(\psi, \Phi + \mathbb{K}\psi + \varphi^{\perp}) \tag{22}$$

and minimizes the integral F, where

$$F = \int_D (v - v_*)^2 W\,dx = \int_D \left(\frac{\varphi}{1+p}\right)^2 W_1\,dx, \tag{23}$$

where W_1 is given by Eq. (15), and $\varphi = \Phi + \mathbb{K}\psi + \varphi^{\perp}$, and

$$p(x) = L_R\varphi - L_I\psi. \tag{24}$$

Once $\varphi^{\perp}$ and ψ are known, $v(x)$ is determined from Eqs. (13) and (18).

The formal solution of this problem will be obtained in the next section. A rigorous treatment including the determination of the value of r will be given in a later section for a simpler version of the problem (where the integral F is simplified by replacing the factor $1 + p$ by 1).

3. SOLUTION OF THE INVERSE PROBLEM *P*

The method of solving P is outlined as follows. It is seen that the solution of Eq. (22) yields ψ as a function of $\varphi^{\perp}$ and with the substitution of this solution in expression (23), the resulting form for F becomes a function of $\varphi^{\perp}$ only. Thus as a necessary condition for a minimum, stationary points or solutions of F (those points for which the first variation δF of F with respect to $\delta\varphi^{\perp}$ is zero) are sought. This in turn will yield an additional nonlinear equation involving ψ and $\varphi^{\perp}$ and combined with Eq. (22) will yield a coupled system of nonlinear integral equations. The solution of these coupled equations by the method of successive approximation is investigated. As a check to see if the stationary solution so obtained yields a minimum for F, the sign of the second variation of F with respect to $\delta\varphi^{\perp}$ is examined at the stationary point.

It may be deduced from Eq. (22) that as a function of $\varphi^{\perp}$ the variation of ψ with respect to $\varphi^{\perp}$ is given by

$$\delta\psi = \psi'\delta\varphi^{\perp} = (I - S_{\psi})^{-1}S_{\varphi^{\perp}}\delta\varphi^{\perp}, \tag{25}$$

where S_{ψ} and $S_{\varphi^{\perp}}$ are the Fréchet derivatives of $S(\psi, \Phi + \mathbb{K}\psi + \varphi^{\perp})$ with respect to ψ and $\varphi^{\perp}$, respectively (treated as independent variables). These linear operators have the explicit form

$$S_{\varphi^{\perp}}h = h(L_R\psi + L_I\varphi) + (\varphi L_I - \psi L_R)h, \tag{26}$$

$$S_{\psi}h = h(L_I\psi - L_R\varphi) + \psi(L_I - L_R\mathbb{K})h + (L_R\psi + L_I\varphi)\mathbb{K}h + \varphi(L_R + L_I\mathbb{K})h, \tag{27}$$

where φ is given by Eq. (18).

It then follows from Eq. (13) and Eq. (18) that as a function of $\varphi^{\perp}$ the variation of v with respect to $\varphi^{\perp}$ is given by $\delta v = v'\delta\varphi^{\perp}$, where the linear operator v' is given by

$$v' = \frac{1}{(1+p)|u_*|^2}\left(I + \mathbb{K}\psi' - \frac{\varphi}{1+p}\mathbb{A}\right), \tag{28}$$

with

$$\mathbb{A} = L_R(I + \mathbb{K}\psi') - L_I\psi'.$$

Finally, the first variation of $F(\varphi^{\perp})$ with respect to $\varphi^{\perp}$ is given by

$$\delta F = 2\int_D W(v - v_*)v'\delta\varphi^{\perp}\,dx$$

$$= 2\int_D \delta\varphi^{\perp}v^{*'}[W(v - v_*)]\,dx, \tag{29}$$

where the superscript star denotes the adjoint operator in the real Hilbert space with inner product $(u,v) = \int_D u(x)v(x)\,dx$.

Since we consider only variations $\delta\varphi^{\perp}$ which are perpendicular to the subspace spanned by $\{H_j\}_{j=1}^{2N}$, expression (29) can be written in the form

$$\delta F = 2\int_D \delta\varphi^{\perp}[\varphi^{\perp} - \tilde{R}(\psi, \varphi^{\perp})]W_1\,dx, \tag{30}$$

with

$$W_1\tilde{R}(\psi,\varphi^{\perp}) = \varphi W_1[1 - (1+p)^{-2}] - \psi^{*'}\mathbb{K}^*[\varphi W_1(1+p)^{-2}] + \mathbb{A}^*[\varphi^2 W_1(1+p)^{-3}]. \tag{31}$$

The stationary solutions of F thus are given by

$$\varphi^{\perp} - \tilde{R} = \sum_{j=1}^{2N} c_j\xi_j(x).$$

The constants c_j are determined by using the condition $\int_D \varphi^{\perp}H_i\,dx = 0$, $i = 1,2,\ldots,2N$. It follows that with $\mathbb{P}$ being the projection operator introduced earlier,

$$\mathbb{P}u = \sum_{j=1}^{2N}\xi_j(x)\int_D H_j(y)u(y)\,dy, \tag{32}$$

the stationary solutions of F are given by

$$\varphi^{\perp} = (I - \mathbb{P})\tilde{R}(\psi,\varphi^{\perp}), \tag{33}$$

and the inverse problem P has been reduced to finding the solutions of the coupled equations (22) and (33). However since such solutions yield only stationary solutions of F, we need to impose an extra condition to select the stationary solution which yields a minimum of F. This is obtained by requiring the second variation of F with respect to $\delta\varphi^{\perp}$, namely $\delta^2 F$, to be positive at the stationary point. From Eq. (30) the second variation of F is given by

$$\delta^2 F = 2\int_D \delta\varphi^{\perp}[\delta\varphi^{\perp} - (\tilde{R}_{\varphi^{\perp}}\delta\varphi^{\perp} + \tilde{R}_{\psi}\psi'\delta\varphi^{\perp})]W_1\,dx,$$

where $\tilde{R}_{\varphi^{\perp}}$ and $\tilde{R}_{\psi}$ are the Fréchet derivatives of $\tilde{R}$ with respect to $\varphi^{\perp}$ and ψ respectively (treated as independent variables). It follows that a sufficient condition for $\delta^2 F > 0$ is

$$\|\tilde{R}_{\varphi^{\perp}}\|_w + \|\tilde{R}_{\psi}\psi'\|_w < 1, \tag{34}$$

where the norms denote the norms of the operators mapping the space of square integral function, with measure $W_1\,dx$ over $\bar{D}$ into itself.

In order to examine the solution of the nonlinear system (22) and (33) consider the special case where the data $\{b_i\}$ (the difference between the measured calculated values) vanishes identically, giving $B_i = 0$ for $i = 1,2,\ldots,N$. From Eq. (19) this yields $\Phi(x) \equiv 0$. It can be shown that if $\varphi^{\perp}$ and ψ are set zero, then the right-hand sides of Eqs. (22) and (33) vanish. Thus a solution of the system is given by $\varphi^{\perp} \equiv \psi \equiv 0$. From Eq. (13) it follows that $v - v_* = 0$, hence the solution is given by $n(x) = n_*(x)$.

Of the many possible solutions of the nonlinear system we will be interested here only in the solution $v(x)$ close to $v_*(x)$ such that the number $\|\psi\|_2 + \|\varphi^{\perp}\|_2$ (or other suitable norms) is small. Hence we will be particularly interested in the solution of the nonlinear system (22) and (33) given by the method of successive approximations

$$\psi_{n+1} = S(\psi_n, \Phi + \mathbb{K}\psi_n + \varphi_n^{\perp}),$$
$$\varphi_{n+1}^{\perp} = (I - \mathbb{P})\tilde{R}(\psi_n, \varphi_n^{\perp})$$

starting with the initial approximation

$$\psi_0 = \varphi_0^{\perp} = 0.$$

However on physical grounds one would expect that in order for the method of successive approximations to converge the measured data corresponding to v must be close to the calculated data corresponding to v_*, i.e., the set $\{b_i\}$ is small in some sense. It was shown in Part I that the method of successive approximations applied to the single equation $\psi = S(\psi, \Phi + \mathbb{K}\psi)$ with $\varphi^{\perp} = 0$ converges if

$$(\|\Phi\|_2)^2 = \sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j \leqslant \rho^2,$$

where the number ρ depends upon the norms of the operators $\mathbb{K}$, L_R, and L_I. Hence for the case considered here we would expect a similar result, that $\|\Phi\|_2$ be sufficiently small. The rigorous verification of the method of succesive approximations for Eq. (22) and (33) will not be pursued here. However, in a later section a rigorous verification of the method of successive approximation applied to Eq. (22) and a simpler form of Eq. (33) is carried out with an estimate for the maximum value of $\|\Phi\|_2$ being given. In the remaining work we will concentrate on the case $W_1 = 1$.

4. FIRST-ORDER CORRECTION TO THE LINEARIZED SOLUTION OF THE INVERSE PROBLEM (for $W_1 = 1$)

Because of the complicated nature of expression (31) the analytical version of any iterative scheme employed for solving system (22) and (33) is in general not too useful. But under suitable approximations such schemes can lead to useful results. In this connection it is useful to look at the method of successive approximations of the systems starting from $\psi = \varphi^{\perp} = 0$ for the case where $W = |u_*|^4$ or $W_1 = 1$. In Part I it was pointed out that this could be interpreted as seeking the nonlinear correction to the linearized approximation of the inverse problem, which is given by $\psi = \varphi^{\perp} = 0$, $\varphi = \Phi$. The linearized approximation to v is $v - v_* \sim \Phi/|u_*|^2$.

The first iterate in the successive approximation scheme has the form (for $W_1 = 1$),

$$\psi_1 = \Phi L_I \Phi, \tag{35}$$

$$\varphi_1^{\perp} = (I - \mathbb{P})\tilde{R}(0,0). \tag{36}$$

In the expression $\varphi_1^{\perp}$ we will keep only quadratic and lower-order terms in Φ. From (26) and (18) it is seen that with $\varphi^{\perp} = \psi = 0$,

$$S^*_{\varphi^{\perp}} h = h(L_I \Phi) + L_I^*(\Phi h),$$

hence the operator $S^*_{\varphi^{\perp}}$ with $\varphi^{\perp} = \psi = 0$ is the order of Φ. From Eq. (27) it may be deduced that a similar result holds for S^*_{ψ}. Hence we can then make the approximation $(I - S^*_{\psi})^{-1} \sim I$ and keeping only the leading terms in Φ, we have as the following approximation to $\tilde{R}(0,0)$:

$$\tilde{R}(0,0) = -(L_I\Phi)(\mathbb{K}^*\Phi) - L_I^*(\Phi\mathbb{K}^*\Phi) + L_R^*(\Phi)^2 + 2\Phi L_R \Phi. \tag{37}$$

With $\varphi_1 = \Phi + \mathbb{K}\psi_1 + \varphi_1^{\perp}$ the first iterate of v is given by

$$v_1 - v_* \sim \varphi_1(1 + L_R\varphi_1 - L_I\psi_1)^{-1}|u_*|^{-2}.$$

Keeping only quadratic and lower-order terms in Φ this reduces to

$$v_1 - v_* \sim [\Phi + \mathbb{K}\psi_1 - \Phi L_R \Phi + (I - \mathbb{P})\tilde{R}(0,0)]|u_*|^{-2}. \tag{38}$$

Here ψ_1 and $\tilde{R}(0,0)$ are given by Eqs. (35) and (37) and Φ and $\mathbb{K}$ by Eqs. (19) and (20) with the auxilliary function $\xi_i(x)$ given by Eq. (17) with $W_1 \equiv 1$.

5. SIMPLE VERSION OF INVERSE PROBLEM

An alternative for choosing a unique solution of Eq. (22) is to select the solution which minimizes the integral

$$F_1 = \int_D (\Phi + \mathbb{K}\psi + \varphi^{\perp})^2 W_1\,dx, \tag{39}$$

instead of the integral F given by Eq. (23). The difference between F_1 and F is in the factor $(1 + p)$ in the integral of F. In F_1 this has been replaced by unity. For solutions $(\varphi^{\perp}, \psi)$ confined to a region with small norm (i.e., $\|\varphi^{\perp}\|_2 + \|\psi\|_2$ small) and with the norm $\|\Phi\|_2$ of the data term itself small, the maximum value of $|p(x)|$ in $\bar{D}$ will be much less than unity. Hence under these conditions, the integrands of the integrals F and F_1 will not be significantly different. However there is a great simplification in the use of F_1. Expression (39) is quadratic in the variables $\varphi^{\perp}$ and ψ. This makes it easier to use numerical or analytical methods in the minimization problem.

Thus we will consider in detail this simpler version of problem P for the case where the weight factor $W = |u_*|^4$ or $W_1 \equiv 1$, namely, to find the small norm solution for $\varphi^{\perp}$ and ψ of Eq. (22) which minimizes

$$F_1 = \int_D [(\Phi + \mathbb{K}\psi)^2 + (\varphi^{\perp})^2]\, dx. \tag{40}$$

Note expressions (21) are used to obtain Eq. (40) from Eq. (39).

As before it can be shown that the stationary solutions of expression (40) are given by the equation

$$\varphi^{\perp} = (I - \mathbb{P})R(\psi, \varphi^{\perp}), \tag{41}$$

where

$$R(\psi, \varphi^{\perp}) = -\psi'^{*} K^{*}(\Phi + K\psi), \tag{42a}$$

with

$$\psi'^{*} = S^{*}_{\varphi^{\perp}}(I - S^{*}_{\psi})^{-1}, \tag{42b}$$

and $\mathbb{P}$ is the projection operator given by Eq. (32).

A sufficient condition for stationary solution to give a minimum of expression (40) is that

$$\|R_{\varphi}\|_2 + \|R_{\psi}\|_2 \|\psi'\|_2 < 1, \tag{43}$$

where the linear operator ψ' is given by Eq. (25).

The simplified version of problem P is reduced to finding the small norm solutions of Eq. (22) and (41), subject to inequality (43). To rigorously investigate the solution of equations define the following vector functions

$$\chi = \begin{bmatrix} \psi \\ \varphi^{\perp} \end{bmatrix}, \tag{44}$$

$$T(\chi) = \begin{bmatrix} S(\psi, \Phi + \mathbb{K}\psi + \varphi^{\perp}) \\ (I - \mathbb{P})R(\psi, \varphi^{\perp}) \end{bmatrix}. \tag{45}$$

Then the system of Eqs. (22) and (41) can be rewritten in the concise form,

$$\chi = T(\chi). \tag{46}$$

For rigorous investigation of the solution of Eq. (46) we need to define an appropriate norm for χ, hence set

$$\|\chi\| = \|\psi\|_2 + \|\varphi^{\perp}\|_2. \tag{47}$$

with $\|L_I\|$ and $\|L_R\|$ representing the norms of the operators L_I and L_R mapping $\mathscr{L}_2(\bar{D})$ into $C(\bar{D})$ (the same notation as in Part I), define

$$\tilde{k} = \mathrm{Max}(1, \|\mathbb{K}\|_2), \tag{48}$$

$$l = 4\tilde{k}(\|L_I\| + \|L_R\|), \tag{49}$$

$$\alpha = \tfrac{1}{2} l \|\Phi\|_2. \tag{50}$$

We shall examine the solution of Eq. (46) in the ball

$$\|\chi\| < r/\tilde{k}, \quad lr = 0.2. \tag{51}$$

With much of the details left to the Appendix the following result will be shown.

Theorem: If $lr = 0.2$ and

$$\tilde{k}l\|\Phi\|_2 = \tilde{k}l\left(\sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j\right)^{1/2} \leqslant 0.24, \tag{52}$$

the Eq. (46) has a solution in the domain $\|\chi\| < r/\tilde{k}$ which is unique for $\|\chi\| \leqslant 0.95r/\tilde{k}^2$ and satisfies inequality (43). The sequence $\{\chi_n\}$ generated by the method of successive approximations

$$\chi_0 = 0, \quad \chi_{n+1} = T(\chi_n)$$

converges to a unique solution in $\|\chi\| \leqslant 0.95r/\tilde{k}^2$ and yields a minimum of the integral F_1 given by Eq. (40).

Proof: It is shown in Appendix A that if $\chi_0 = 0$ and $\|\chi - \chi_0\| = \tau$, and $\tilde{k}\tau = t$ then

$$\|\chi_0 - T(\chi_0)\| \leqslant C,$$

$$\|T'(\chi)\| \leqslant \tilde{k}At + \tilde{k}B, \quad \text{for } 0 \leqslant t \leqslant r,$$

where $T'(\chi)$ is the Frećhet derivative of $T(\chi)$ and

$$2A = l[1 + (3+\alpha)(1+\alpha)(1-\alpha)^{-1}(0.8-\alpha)^{-1}],$$

$$B = \alpha(3+\alpha)(1-\alpha)^{-1},$$

$$lC = \alpha^2(3-\alpha)(1-\alpha)^{-1}.$$

It follows from Vainberg[2] that in the domain $\|\chi\| \leqslant r/\tilde{k}$, the equation $\chi = T(\chi)$ is majorized by the equation

$$\tfrac{1}{2}A\tilde{k}^2\tau^2 + B\tilde{k}\tau + C = \tau,$$

which reduces to

$$\tfrac{1}{2}At^2 + (B - 1/\tilde{k})t + C = 0, \quad 0 \leqslant t < r. \tag{53}$$

If inequality (52) is satisfied then $\alpha \leqslant 0.12/\tilde{k}$. Since $B(\alpha)$ is an increasing function of α and since $\tilde{k} \geqslant 1$, it follows

$$B(\alpha) \leqslant B(0.12/\tilde{k}) \leqslant (1/\tilde{k})B(0.12)$$

giving

$$(1/\tilde{k} - B) \geqslant (1/\tilde{k})(0.5745) > 0.$$

In addition it is seen that

$$2A(\alpha)C(\alpha) \leqslant (1/\tilde{k}^2)2A(0.12)C(0.12) \leqslant (0.322)/\tilde{k}^2,$$

hence

$$2AC < (1/\tilde{k} - B)^2.$$

Thus both roots t_1^* and t_2^* of Eq. (53) are real and positive. As a consequence, the solution to Eq. (46) exists in the domain $\|\chi\| \leqslant r/\tilde{k}$, and is unique for $\|\chi\| \leqslant t_2^*/\tilde{k}$ (the largest root), and the method of successive approximations starting from $\chi_0 = 0$ converges to a unique solution in the ball $\|\chi\| \leqslant t_2^*/\tilde{k}$. For the range of values of $\|\Phi\|_2$ given by inequality (50), it can be shown that $\tilde{k}lt_2^* > 0.19$. Thus we have a unique solution for $\|\chi\| \leqslant 0.95r/\tilde{k}^2$.

From Eqs. (A3), (A6), and (A7) of the Appendix it follows that for $0 \leqslant lt \leqslant lr = 0.2$, and $0 \leqslant \alpha \leqslant 0.12/\tilde{k}$ that

$$\|R_{\varphi^{\perp}}\|_2 + \|R_{\psi}\|_2 \|\psi'\|_2 \leqslant 0.261,$$

hence inequality (43) is satisfied. Thus the stationary solution given by the solution of Eq. (46) yield a minimum for expression (40).

From Eq. (24) it can be shown for $lr = 0.2$ and $\alpha \leqslant 0.12/\tilde{k}$ that

$$|p| \leqslant \|L_R\|(\|\Phi\|_2 + \|\varphi^{\perp}\|_2) + \tfrac{1}{4}l\|\psi\|_2 \leqslant (\tfrac{1}{2}\alpha + \tfrac{1}{4}lr) \leqslant 0.11.$$

Thus for values of r and $\|\Phi\|_2$ as stated in the Theorem, unity is a fair approximation to $(1 + p)$.

Note, it should be emphasized that the values on the size of the domains (as indicated by the number r) and the size of the data (as indicated by the number $\|\Phi\|_2$) that are given in the Theorem for the results to hold, are crude estimates. A much sharper and precise analysis would yield much larger estimates for these numbers.

Finally the first iterate of v corresponding to the meth-

od of successive approximations applied to Eq. (46) is given by

$$v - v_* \sim (\Phi - \Phi L_R \Phi + \mathbb{K}(\Phi L_I \Phi) + \varphi_1^{\perp})/|u_*|^2,$$

with

$$\varphi_1^{\perp} = -(1 - \mathbb{P})[L_I^*(\Phi \mathbb{K}^* \Phi) + (L_I \Phi)(\mathbb{K}^* \Phi)]$$

and terms of higher order than the second in $\|\Phi\|_2$ are neglected.

COMMENTS

For the inverse problem associated with a finite set of measurements of the scattered field generated by a single incident wave at a single frequency, three strategies have been given so as to determine a unique solution $n(x)$ which is close to an *a priori* assigned value $n^*(x)$. The simplest procedure is given in Part I for it leads to the solution of a single quadratic integral equation. Here the most meaningful approach from the physical point of view given by employing the condition $\int_D (n^2 - n_*^2)^2 W\, dx$ is presented. This however leads to a much more difficult system of nonlinear equations to solve. Apart from the first iterate obtained by the method of successive approximations, practical solutions may not be too easily obtained. A simpler version of this is presented, namely, by the condition $\mathrm{Min} F_1$, where the integral F_1 is given by Eq. (39). This still leads to a system of nonlinear equations, but these are simpler. However, since F_1 is a quadratic form in the unknown variables, other techniques (numerical or otherwise) may be employed.

For the strategy given by the condition ($\varphi^{\perp} = 0$) in Part I, and the strategy given here by condition $\mathrm{Min} F_1$, crude estimates are given that insure a unique solution. The representative conditions, Eq. (50) of Part I and Eq. (52) of this paper, place a restraint on the size of the data.

Again we want to emphasize that condition (50) of Part I is valid only for $\|\mathbb{K}\|_2 \leqslant 1$. For the case $\|\mathbb{K}\|_2 \geqslant 1$, ρ on the right-hand side of Eq. (50) is replaced by $1/\rho = 2\tilde{k}\,[1 + (2\tilde{k})^{1/2}](\|L_I\| + \|L_R\|)$.

There still remains a considerable amount to be investigated from the theoretical standpoint (as indicated in Part I) as well as the practical implementation of the present analysis. Of particular interest would be the extension of these results to the case where only the amplitude of the scattered field is measured.

APPENDIX: ESTIMATE FOR SUCCESSIVE APPROXIMATIONS

Here we want to get estimates for the following quantities: $\|\chi_0 - T(\chi_0)\|$, where $\chi_0 = 0$ and $\|T'(\chi)\|$, for $\|\chi\| \leqslant r/\tilde{k}$.

From Part I, it was pointed out that as an operator mapping $\mathscr{L}_2(\bar{D})$ into the space $C(\bar{D})$, $\|L_R \psi\|_c \leqslant \|L_R\|\,\|\psi\|_2$, with a similar result for L_I. With

$$\|\chi\| = \|\psi\|_2 + \|\varphi^{\perp}\|_2 = \tau, \quad \tilde{k}\tau = t,$$

it follows from Eqs. (18), (26), and (27) that

$$\|\varphi\|_2 \leqslant \|\Phi\|_2 + t,$$

$$\tilde{k}\,\|S_{\varphi_1}\|_2 \leqslant \tfrac{1}{2} l t + \alpha, \tag{A1}$$

$$\|S_{\psi}\|_2 \leqslant l t + \alpha, \tag{A2}$$

where l is given by Eq. (49) and α by Eq. (50). Since

$$\|(I - S_{\psi})^{-1}\|_2 \leqslant (1 - \|S_{\psi}\|_2)^{-1}$$

it follows from Eq. (25) that as an operator mapping $\mathscr{L}_2(\bar{D})$ into itself, ψ' has norm

$$\tilde{k}\,\|\psi'\|_2 \leqslant (\tfrac{1}{2} l t + \alpha)/(1 - l t - \alpha). \tag{A3}$$

Since $\|\psi^*\|_2 = \|\psi'\|_2$, it follows from Eqs. (A3) and (42) that for $\varphi^{\perp} = \psi = 0$, and hence $t = 0$,

$$l\,\|R(0,0)\|_2 \leqslant 2\alpha^2 (1 - \alpha)^{-1},$$

and from Eq. (12)

$$l\,\|S(0,\Phi)\| \leqslant l\,\|L_I\|(\|\Phi\|_2)^2 \leqslant \alpha^2.$$

Thus with $\chi_0 = 0$, we have

$$\begin{aligned}\|\chi_0 - T(\chi_0)\| &= \|S(0,\Phi)\|_2 + \|(I - \mathbb{P})R(0,0)\|_2 \\ &\leqslant \alpha^2 (3 - \alpha)/l\,(1 - \alpha),\end{aligned} \tag{A4}$$

where we have made use of the fact that $\mathbb{P}$ is a projection operator, hence $\|I - \mathbb{P}\|_2 \leqslant 1$.

To get the appropriate expression for $T'(x)$ we will need the Fréchet derivatives of the linear operators S_{ψ}^* and $S_{\varphi_1}^*$ with respect to $\varphi^{\perp}$ and ψ treated as independent variables. Since

$$S_{\varphi_1}^* h = h(L_R \psi + L_I \varphi) + L_I^*(\varphi h) - L_R^*(\psi h), \tag{A5}$$

the bilinear operators $S_{\varphi_1 \psi}^*$ and $S_{\varphi_1 \varphi_1}^*$ mapping $\mathscr{L}_2(\bar{D}) \otimes \mathscr{L}_2(\bar{D})$ are given by

$$S_{\varphi_1 \varphi_1}^*[h,f] = h L_I f + L_I^*(hf),$$

$$\begin{aligned}S_{\varphi_1 \psi}^*[h,f] = &-L_R^*(hf) + h(L_R + L_I \mathbb{K}) f \\ &+ L_I^*(h \mathbb{K} f).\end{aligned}$$

And since it can be shown that

$$\|L_R^* h f\|_2 = \|h L_R f\|_2 \leqslant \|L_R\|\,\|f\|_2 \|h\|_2,$$

it follows

$$\|S_{\varphi_1 \varphi_1}^*\|_2 \leqslant 2\|L_I\| \leqslant l/2\tilde{k},$$

$$\|S_{\varphi_1 \psi}^*\|_2 \leqslant l/2.$$

Similar expressions can be derived for the bilinear operators $S_{\psi \varphi_1}^*$ and $S_{\psi \psi}^*$ and it can be deduced that

$$\|S_{\psi \varphi_1}^*\|_2 \leqslant l/2, \quad \|S_{\psi \psi}^*\|_2 \leqslant l\tilde{k}.$$

We are now in a position to get an estimate for $\|T'(\chi)\|$. Since

$$\begin{aligned}\|T'(\chi)\delta\chi\| = &\|S_{\psi}\delta\psi + S_{\varphi_1}\delta\varphi_1\|_2 \\ &+ \|(I - \mathbb{P})(R_{\psi}\delta\psi + R_{\varphi_1}\delta\varphi^{\perp})\|_2,\end{aligned}$$

where $\|\delta\chi\| = \|\delta\psi\|_2 + \|\delta\varphi^{\perp}\|_2$, it follows that

$$\|T'(\chi)\| \leqslant \mathrm{Max}[\|S_{\psi}\|_2 + \|R_{\psi}\|_2, \|S_{\varphi_1}\|_2 + \|R_{\varphi_1}\|_2].$$

From Eqs. (42a) and (42b) it is seen that the operators R_{ψ} and R_{φ_1} are given by

$$-R_{\psi} h = S_{\varphi_1 \psi}^*[\tilde{g},h] + \psi^* \mathbb{K}^* \mathbb{K} h + \psi^* S_{\psi \psi}^*[g,h],$$

$$-R_{\varphi_1} h = S_{\varphi_1 \varphi_1}^*[\tilde{g},h] + \psi^* S_{\psi \varphi_1}^*[g,h],$$

where

$$\mathbf{K}^*(\Phi + \mathbf{K}\psi) = g,$$

$$(I - S_{\psi}^*)^{-1} g = \tilde{g},$$

and thus

$$\|R_{\psi}\|_2 \leqslant \tilde{k}(lt + 2\alpha)(1 + \tfrac{1}{2}lt + \alpha)/(1 - lt - \alpha), \tag{A6}$$

$$\|R_{\varphi^1}\|_2 \leqslant (1/\tilde{k})\|R_{\psi}\|_2. \tag{A7}$$

Since $\|S_{\varphi^1}\|_2 \leqslant \|S_{\psi}\|_2$, it is seen that

$$\begin{aligned}\|T'(\chi)\| &\leqslant \|S_{\psi}\|_2 + \|R_{\psi}\|_2 \\ &\leqslant \tilde{k}(lt + 2\alpha)(1 + \tfrac{1}{2}lt + \alpha)/(1 - lt - \alpha) \\ &\quad + \tilde{k}lt + \alpha\tilde{k}.\end{aligned} \tag{A8}$$

To simplify analysis we want to replace the above estimate by an estimate whch is linear in t holding for $0 \leqslant t \leqslant r$. Such an upper bound estimate is obtained by taking a straight line approximation to the right-hand side of Eq. (A8) through the points with abcissa $t = 0$ and $t = r$, giving

$$\|T'(\chi)\| \leqslant \tilde{k}At + \tilde{k}B, \tag{A9}$$

$$2A = l[1 + (3 + \alpha)(1 + \alpha)(1 - \alpha)^{-1}(1 - lr - \alpha)^{-1}],$$

$$B = \alpha(3 + \alpha)(1 - \alpha)^{-1}.$$

[1] V. H. Weston, "Inverse Problem for the Reduced Wave Equation with Fixed Incident Wave," J. Math. Phys. **21**, 758–64 (1980).
[2] M. V. Vainberg, *Variational Methods for the Study of Non-Linear Operators* (Holden–Day, San Francisco, CA, 1964).

A note on helicity [a)]

Iwo Bialynicki-Birula, [b)] E. T. Newman, J. Porter, and J. Winicour
University of Pittsburgh, Pittsburgh, Pennsylvania 15260

B. Lukacs, Z. Perjes, and A. Sebestyen
Central Research Institute of Physics, Budapest, Hungary

(Received 23 March 1981; accepted for publication 19 June 1981)

We give a formal definition of the helicity operator for integral spin fields, which does not involve their momentum-space decomposition. We base our discussion upon a representation of the Pauli–Lubanski operator in terms of the action on tensor fields by the Killing vectors associated with the generators of the Poincaré group. This leads to an identification of the helicity operator with the duality operator defined by the space–time alternating tensor. Helicity eigenstates then correspond to self-dual or anti-self-dual fields, in agreement with usage implicit in the literature. In addition, we discuss the relationship between helicity eigenstates, which are intrinsically nonclassical, and states of right or left circular polarization in classical electrodynamics

PACS numbers: 03.50.De, 02.20. + b

I. INTRODUCTION

Helicity, which is defined as the projection of the total angular momentum on the direction of motion of the particle, is usually discussed in the context of particle theory and almost always with the use of momentum eigenstates.[1–4] In this note we shall extend the notion of helicity to classical fields, such as the Maxwell field or the linearized gravitational field, and shall furthermore show that the helicity operator can be represented as a single algebraic (local) operator on these fields, which is completely independent on the decomposition into momentum states. In the case of classical non-abelian gauge theory or general relativity, where the generators of the Poincaré group seem to be poorly defined, one can still define the helicity operator[5] as a natural generalization from the Maxwell case.

In Sec. II of this note we discuss, from the point of view of Lie derivatives, the action of the Pauli-Lubanski[3,6] vector on Maxwell and linearized gravitational fields and we discover that the helicity operator is essentially the (Hodge) duality operator. In Sec. III we discuss the relationship between helicity states and polarization states.

II. THE PAULI–LUBANSKI VECTOR, HELICITY AND DUALITY

We begin with listing the ten (Hermitian) generators of the Poincaré group P_a and $M_{ab} = -M_{ba}$ and their commutation relations. These generators can be represented as ten Killing vector fields on Minkowski space–time:

$$P_{\underline{a}} = ip^{a}_{\underline{a}} \frac{\partial}{\partial x^{a}}, \tag{2.1}$$

$$M_{\underline{ab}} = im^{ab}_{\underline{ab}} \eta_{ac} x^{c} \frac{\partial}{\partial x^{b}},$$

with

$$p^{a}_{\underline{a}} = \delta^{a}_{\underline{a}}, \tag{2.2}$$

$$m^{ab}_{\underline{ab}} = \delta^{a}_{\underline{a}}\delta^{b}_{\underline{b}} - \delta^{b}_{\underline{a}}\delta^{a}_{\underline{b}},$$

and $\eta_{ab} = \text{diag}\,(1, -1, -1, -1) = \eta_{ab}$. Note that underlined letters label vectors while the nonunderlined letters label components.

The commutators are defined as the Lie brackets between the Killing vectors

$$\begin{aligned} &(1/i)[P_{\underline{a}}, P_{\underline{b}}] = 0, \\ &(1/i)[M_{\underline{ab}}, P_{\underline{c}}] = P_{\underline{a}}\eta_{\underline{bc}} - P_{\underline{b}}\eta_{\underline{ac}}, \\ &(1/i)[M_{\underline{ab}}, M_{\underline{cd}}] = M_{\underline{ad}}\eta_{\underline{bc}} - M_{\underline{bd}}\eta_{\underline{ac}} + M_{\underline{bc}}\eta_{\underline{ad}} - M_{\underline{ac}}\eta_{\underline{bd}}. \end{aligned} \tag{2.3}$$

The action of a Poincaré generator on a tensor field is defined as the Lie derivative of the tensor field by the associated Killing vector. As an example, the Lie derivative of the Maxwell field by ξ^{a} is

$$\mathcal{L}_{\xi} F_{ab} \equiv \xi^{c}\nabla_{c}F_{ab} + F_{ac}\nabla_{b}\xi^{c} + F_{cb}\nabla_{a}\xi^{c}. \tag{2.4}$$

For classical zero real-mass particles the Pauli–Lubanski vector, defined as

$$S^{a} = \tfrac{1}{2}\epsilon^{abcd}P_{b}M_{cd}, \tag{2.5}$$

where ϵ^{abcd} is the alternating tensor ($\epsilon^{0123} = -1$) and P_b and M_{cd} are, respectively, the particle four-momentum and angular momentum, can be shown[6] to be proportional to the particle momentum P^{a}. The particle helicity S is defined as the proportionality coefficient:

$$S^{a} = SP^{a}. \tag{2.6}$$

With the motivation to find the analog of S for fields, we define the Pauli–Lubanski operator for tensor fields:

$$S^{a} = \tfrac{1}{2}\epsilon^{\underline{abcd}}P_{\underline{b}}M_{\underline{cd}} \tag{2.7}$$

$$= \tfrac{1}{2}\epsilon^{\underline{abcd}}\mathcal{L}_{P_{\underline{b}}}\mathcal{L}_{M_{\underline{cd}}}. \tag{2.8}$$

Note that $S^{\underline{a}}$ is independent of the order in which the two Lie derivatives appear, as well as the choice of origin for $M_{\underline{cd}}$.

We shall now apply this general definition to the Maxwell field. We have

[a)] This work has been supported by a grant from the National Science Foundation.

[b)] On leave from the Institute of Theoretical Physics, Warsaw, Poland.

 0022-2488/81/112530-03$01.00

$$\mathcal{L}_{P_{\underline{a}}} F_{ab} = i\nabla_{\underline{a}} F_{ab}, \tag{2.9}$$

$$\mathcal{L}_{M_{\underline{cd}}} \mathcal{L}_{P_{\underline{a}}} F_{ab} = -m^{\underline{cd}}_{\underline{cd}} \eta_{cf} x^f \nabla_d \nabla_{\underline{a}} F_{ab} + m^{\underline{cd}}_{\underline{cd}} \eta_{db} \nabla_{\underline{a}} F_{ac} + m^{\underline{cd}}_{\underline{cd}} \eta_{da} \nabla_{\underline{a}} F_{cb}. \tag{2.10}$$

It follows from the definition of $m^{ab}_{\underline{ab}}$ that the first term in (2.10) is a sum of two terms which are symmetric either in $\underline{bc}$ or $\underline{bd}$ and thus vanishes when contracted with ϵ^{abcd}. Hence,

$$S^{\underline{a}} F_{ab} = \tfrac{1}{2}\epsilon^{\underline{abcd}} m^{\underline{cd}}_{\underline{cd}} (\eta_{db} \nabla_{\underline{a}} F_{ac} + \eta_{da} \nabla_{\underline{a}} F_{cb}) = \epsilon^{\underline{abcd}} (\eta_{\underline{d}b} \nabla_{\underline{a}} F_{a\underline{c}} + \eta_{\underline{d}a} \nabla_{\underline{a}} F_{cb}). \tag{2.11}$$

Using $F^{*ab} = \tfrac{1}{2}\epsilon^{\underline{abcd}} F_{cd}$, $j_b = \nabla_a F^a{}_b$ and $j^*_b = \nabla_a F^{*a}{}_b$, we obtain

$$S_{\underline{a}} F_{ab} = \eta_{\underline{a}b} j^*_a - \eta_{\underline{a}a} j^*_b - \epsilon_{\underline{a}abc} j^c + \nabla_{\underline{a}} F^*_{ab}. \tag{2.12}$$

If the free Maxwell equations are satisfied, i.e., $j_a = 0 = j^*_a$, then

$$S_{\underline{a}} F_{ab} = -iP_{\underline{a}} F^*_{ab} = -iDP_{\underline{a}} F_{ab}, \tag{2.13}$$

where D is the duality operator

$$(DF)_{ab} \equiv \tfrac{1}{2}\epsilon_{abcd} F^{cd}. \tag{2.14}$$

Comparing (2.14) with (2.6) we see that we can take the helicity operator $\hat{S}$ as

$$\hat{S} = -iD. \tag{2.15}$$

If we had done this calculation with the Weyl tensor satisfying the linearized Einstein equations, the result would have been

$$S_{\underline{a}} C_{abcd} = -2iDP_{\underline{a}} C_{abcd} \tag{2.16}$$

with $DC_{abcd} \equiv \tfrac{1}{2}\epsilon_{abef} C^{ef}_{cd}$.
In general, one would have

$$\hat{S} = -|s| iD,$$

with $|s|$ an integer, the spin of the field.

Note that in order to obtain the operator equation

$$S_{\underline{a}} = \hat{S} P_{\underline{a}},$$

we needed to impose the field equations on the vector space of spin $|s|$ fields. However, the helicity operator $\hat{S}$ is a purely algebraic operation and, therefore, we can define it on the Weyl tensor for the full Einstein theory or for nonabelian gauge fields.

III. MEANING OF HELICITY EINGENSTATES

Returning to Maxwell fields we now examine helicity eigenstates. They are respectively the self-dual and anti-self-dual fields:

$$F^*_{+ab} = iF_{+ab}, \qquad F^*_{-ab} = -iF_{-ab}. \tag{3.1}$$

Unfortunately, aside from the zero field, they are never real and hence do not correspond to a classical Maxwell field. Nevertheless a classical "meaning" can still be assigned to (3.1). If we are given a real Maxwell field, it can be uniquely decomposed into the self- and anti-self-dual parts:

$$F_{ab} = F_{+ab} + F_{-ab}, \qquad F_{+ab} = \tfrac{1}{2}(F_{ab} - iF^*_{ab}), \quad F_{-ab} = \tfrac{1}{2}(F_{ab} + iF^*_{ab}). \tag{3.2}$$

Each part, one being the complex conjugate of the other, carries the full information about the original field. If however we ask for the positive frequency parts of the self- and anti-self-dual fields, they are independent and carry the full information of, respectively, the left and right circularly polarized parts of the original field.

To see this note that dualing and extracting the positive frequency parts are commuting linear operations and we can hence look at a single frequency plane wave. For a plane elliptically polarized wave traveling along the x axis we have[7,8]

$$\mathbf{E} = \mathrm{Re}\, \mathbf{b} e^{i\theta}, \qquad \mathbf{B} = \mathbf{i} \times \mathbf{E}, \tag{3.3}$$

with $\mathbf{i}, \mathbf{j}, \mathbf{k}$ unit vectors in the x, y, z directions and

$$\mathbf{b} = b_1 \mathbf{j} + ib_2 \mathbf{k}, \quad \theta = \omega(x - t).$$

This can be decomposed into right and left circularly polarized parts by

$$\mathbf{E} = \mathrm{Re}\{[b_{\mathrm{L}}(\mathbf{j} + i\mathbf{k}) + b_{\mathrm{R}}(\mathbf{j} - i\mathbf{k})] e^{i\theta}\}$$

with $b_1 = b_{\mathrm{L}} + b_{\mathrm{R}}$, $b_2 = b_{\mathrm{L}} - b_{\mathrm{R}}$. It is now straightforward to show that

$$\mathbf{E} - i\mathbf{B} = b_{\mathrm{L}}[\mathbf{j} - i\mathbf{k}] e^{-i\theta} + b_{\mathrm{R}}[\mathbf{j} - i\mathbf{k}] e^{i\theta}, \qquad \mathbf{E} + i\mathbf{B} = b_{\mathrm{L}}[\mathbf{j} + i\mathbf{k}] e^{i\theta} + b_{\mathrm{R}}[\mathbf{j} + i\mathbf{k}] e^{-i\theta}. \tag{3.4}$$

We thus have that the positive frequency part of $E - iB$ and $E + iB$ are

$$(\mathbf{E} - i\mathbf{B})^{(+)} = b_{\mathrm{R}}[\mathbf{j} - i\mathbf{k}] e^{i\theta}, \qquad (\mathbf{E} + i\mathbf{B})^{(+)} = b_{\mathrm{L}}[\mathbf{j} + i\mathbf{k}] e^{i\theta}, \tag{3.5}$$

representing the left and right circularly polarized fields. Since $\mathbf{E} + i\mathbf{B}$ and $\mathbf{E} - i\mathbf{B}$ are equivalent,[9] respectively, to the self- and anti-self-dual parts of F_{ab} we have proved our contention.

In addition to this classical "meaning", the positive-frequency self- and anti-self-dual fields are, from a quantum mechanical viewpoint, the one-particle wavefunctions for positive and negative helicity eigenstates. In contrast to the real-valued classical fields, these wavefunctions are necessarily complex, with $F_{\pm ab}$ having the form (3.2), where F_{ab} is real. This complex nature of $F_{\pm ab}$ combines with the axial-tensor property of ϵ_{abcd} to yield the well-known mapping of positive helicity wavefunctions into negative ones by a parity transformation. For instance,

$$PF_{+ab} = \tfrac{1}{2}[PF_{ab} + i(PF_{ab})^*],$$

where PF_{ab} is the image of F_{ab} under a parity transformation. Thus PF_{+ab} has negative helicity, in accordance with (3.2).

IV. DISCUSSION

In this note we have formalized the generally accepted (but no place to our knowledge explicitly stated) idea that helicity states of the Maxwell field are eigenstates of the duality operator. In the process of doing this we saw that the helicity operator could easily be extended to nonlinear gauge fields and general relativity.[11] A problem arises here in the interpretation of the classical helicity eingenstates—they

cannot be combined with their complex conjugates [as in Eq. (3.5)] to produce real solutions of the field equations. Only in the asymptotic domain (where the fields become weak) can this be done. It seems reasonable to try to interpret[9,10] these solutions as the one-particle quantum mechanical wavefunctions or, asymptotically, as classical circular-polarization states.

We conclude with a brief summary of the conventions used here. They have been adopted to coincide with those used by many workers in general relativity and in particular the Penrose group at Oxford University.

(1) The Lorentz metric is diag$(+\ -\ -\ -)$.

(2) The alternating tensor ϵ_{abcd} has

$$\epsilon_{0123} = 1,\ \epsilon^{0123} = -1.$$

(3) The duality operation is

$$F^{*}_{ab} = \tfrac{1}{2}\epsilon_{abcd}F^{cd}$$

and the decomposition of F into respectively self- and anti-self-dual parts is

$$F_{+\,ab} = \tfrac{1}{2}(F_{ab} - iF^{*}_{ab}),$$

$$F_{-\,ab} = \tfrac{1}{2}(F_{ab} + iF^{*}_{ab}),$$

from which

$$F^{*}_{+\,ab} = iF_{+\,ab}\ ,$$

$$F^{*}_{-\,ab} = -iF_{-\,ab}\ .$$

$\mathbf{E} + i\mathbf{B}$ is equivalent to the self-dual part and $\mathbf{E} - i\mathbf{B}$ to the anti-self-dual part.

(4) For spin $|s|$ fields the eigenvalues of the helicity operator $\hat{S} = -iD$ $|s|$ are $|s|$ for self-dual fields and $-|s|$ for the anti-self-dual fields.

(5) The electric vector of a right circularly polarized wave approaching an observer will be seen to rotate clockwise.

(6) Note that the traditional[8] definition of right (left) circular polarization coincides with negative (positive) helicity.

(7) Anti-self-dual fields correspond to unprimed spinor fields and self-dual fields correspond to primed spinor fields. Furthermore, one frequently encounters the terminology of self-dual fields being called left-flat and anti-self-dual fields being called right-flat. Note also that "right-handed" is left-circularly polarized, etc.

[1]William W. Frazer, *Elementary Particles* (Prentice-Hall, Englewood Cliffs, N.J., 1966).

[2]Stephen Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966).

[3]Paul Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969).

[4]V. Berestetskii, E. Lifshitz, and L. Pitaevskii, *Relativistic Quantum Theory* (Addison-Wesley, Reading, Mass., 1971).

[5]For another approach to this question see D. J. Duff and C. J. Isham, Nuc. Phys. B **162**, 271 (1980).

[6]R. Penrose and M. A. H. MacCallum, Phys. Lett. C **6**, 242 (1973). It should be noted that their η_{abcd} is $-\epsilon_{abcd}$.

[7]L. Landau and E. Lifshitz, *Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1951).

[8]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[9]If $F^{ab} \Leftrightarrow F^{oi} = E_i$, $F^{12} = B_3$, $F^{13} = -B_2$, $F^{23} = B_1$, then $F_{oi} = -E_i$ and $F^{*}_{oi} = \tfrac{1}{2}\epsilon_{oijk}F^{jk} = B_i$ and hence $F_{oi} - iF^{*}_{oi} = -(E_i + iB_i)$, etc.

[10]R. Penrose, Gen. Relativ. Gravit. **7**, 31 (1976).

[11]M. Ko, M. Ludvigsen, E. T. Newman, and K. P. Tod, "The Theory of H-Space," Phys. Rep. **71** No. 2 (May 1981).

Bound states of a classical charged nonlinear Dirac field in a Coulomb potential

Antonio F. Rañada and Juan M. Uson[a)]
Departamento de Física Teórica and GIFT, Universidad Complutense, Madrid-3, Spain

(Received 28 July 1980; accepted for publication 31 October 1980)

The electromagnetic self-interaction of a bound classical nonlinear Dirac field in a Coulomb potential is studied. It is shown that by a finite renormalization procedure the radiative corrections can be eliminated at order α^2, after which the energy levels coincide with those of the hydrogen atom.

PACS numbers: 03.50.Kk, 03.65.Ge

I. INTRODUCTION

The study of the nonlinear wave equations has been during the seventies a center of growing attention in theoretical physics. It seems that its interest will be even greater during during the eighties. One of the main reasons is the surprising behavior of the soliton and its applicability to many branches of physics. The extraordinary and unsuspected richness of the world of nonlinear partial differential equations advises physicists to reconsider many problems, focusing on aspects more directly related to the structure of the differential equations which are involved.[1] As a consequence the classical field theory appears today as a much more interesting problem than what it was thought to be ten years ago. More phenomena are being discovered in classical physics while the classical study of the wave equations is considered a useful tool in the understanding of quantum physics. In fact, had more effort been dedicated to the classical fields in the fifties, the soliton would had been discovered some years before.

One of the problems in which the solitons may be useful is study of the elementary particles. Unfortunately, its definition in more than one space dimension is not even clear.[2] But, in spite of that, several models of elementary particles have been proposed in which they appear represented as solitary waves of scalar or spinor fields. The first approach of this type is due to Rosen[3] and deals with the interaction of scalar and electromagnetic fields. Several authors considered classical Dirac Fields, self-coupled [4,5] or in interaction with its own electromagnetic field.[6] In 1970 Soler[7] considered a model of elementary fermion by means of a nonlinear Dirac field. Later on this model was developed to represent the nucleons[8–13] even allowing for the inclusion of the corresponding mesonic cloud. The main characteristics of the nucleons as the charge, magnetic moment, spin, energy, mean square charge radii, etc. appear fairly well represented.

These results seem to suggest that the use of nonlinear wave equations may be useful in the study of the elementary particles with structure. Although the present knowledge of the behavior of these particles cannot be compared to that of the electrons, they are supposed to be quantum objects and to obey some kind of quantum theory. It seems therefore that there must be some relation between nonlinear classical waves and quantum waves.[15,16] This relation is difficult to study in the case of the nucleons and other extended particles because the quantum laws that they follow are not completely known. The electron, on the other hand, offers a good opportunity to study the nonlinear terms in the partial differential equations in a system whose quantum properties are completely known. The purpose of this paper is therefore to ameliorate the knowledge of the nonlinear partial differential equations which appear in quantum physics. Although a nonlinear term of the form $(\bar{\psi}\psi)^2$ is considered at first, it is later neglected and we concentrate in the same equation which appears in the usual quantum theory but without a second quantization. In order to avoid any misunderstanding it must be stressed that we are not proposing a new theory of the hydrogen atom or a nonlinear quantum mechanics but only to improve the knowledge of the relation between a classical nonlinear theory and a quantum one.

The first question which naturally arises is the following. Is it possible to build a classical nonlinear theory which admits quantum mechanics as a certain kind of limit? Or, stated otherwise, can a classical nonlinear theory of the scalar or spinor fields be compatible with quantum mechanics? As this one is an essentially linear theory it is necessary that in some situations, for instance in atoms and molecules, the nonlinearities could be eliminated either because they are very small or by a renormalization procedure. This is the case in an interesting nonrelativistic proposal of Bialinicky-Birula and Mycielski.[14] It was found[15–17] that this is also the case in the relativistic hydrogen atom, when the radiative corrections are not considered. But the case in which the electron field interacts with its own electromagnetic field is more complicated. It is the object of this work.

II. THE CLASSICAL MODEL OF THE HYDROGEN ATOM

Let us consider the Lagrangian density

$$L = L_D + L_{NL} + L_{EM} + L_I, \tag{1}$$

where

$$\begin{aligned} L_D &= \tfrac{1}{2}\{\bar{\psi}\gamma^\mu\partial_\mu\psi - (\partial_\mu\bar{\psi})\gamma^\mu\psi\} - m\bar{\psi}\psi, \\ L_{NL} &= \lambda(\bar{\psi}\psi)^2, \\ L_{EM} &= -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \\ L_I &= -e\bar{\psi}\gamma^\mu\psi A_\mu - e\bar{\psi}\gamma^\mu\psi A_\mu^{(\mathrm{ext})}. \end{aligned} \tag{2}$$

[a)]Present address: Physics Department, Joseph Henry Laboratories, Princeton University, P. O. Box 708, Princeton, N. J. 08540.

 0022-2488/81/112533-06$01.00

It corresponds to a nonlinear Dirac field in minimal interaction with an external Coulomb potential $A_0^{(\mathrm{ext})}$ and with its own electromagnetic field A_μ. Although we could consider a more general form of L_{NL}, this is not necessary when the field is very extended with respect to its Compton wavelength,[15] as it will be in our case.

We will study in this work the solutions of the field equations corresponding to (1). To reduce the complexity of the problem we can neglect, as a first approximation, the electromagnetic field produced by the Dirac spinor. This was done in Refs. 15 and 17 and it was found that for each state of the hydrogen atom there is a family of solutions which depends continuously on the frequency ω. These families bifurcate from the zero solutions at the linear values of ω. In spite of that, a physical value of the frequency can be obtained by using the conservation of the norm. A classical model of the hydrogen atom can thus be constructed which, for a large interval of values of λ ($|\lambda|m^2 < 10^{-4}$) gives the same results as the usual one and in which there appears to exist a close relation and analogy between quantization and linearization.

Up to this level the theory could be considered as the classical counterpart of the standard relativistic quantum mechanics of the hydrogen atom. There is, however, a second level at which the radiative corrections must be considered. The standard quantum treatment gives infinite corrections which, fortunately, can be eliminated by the renormalization techniques,[18] after which an excellent agreement with the experimental results is obtained. In this paper we will undertake the study of this second level from the classical point of view. The corrections turn out to be finite. But its calculation as explicit analytic functions of α seems impossible. For this reason we have used numerical methods. Because the procedure is very long and time consuming we have only considered the α^2 corrections which are produced by electrostatic effects, the leading magnetic ones being of α^4 order.

III. THE RADIATIVE CORRECTIONS: SOLUTION OF THE FIELD EQUATIONS

As we will only consider electrostatic effects we can write the electromagnetic field of ψ as

$$A_\mu = (A_{e_0}, \vec{0}),$$

the field equations being

$$i\gamma^\mu \partial_\mu \psi - m\psi - e(A_{e_0} + A_0^{(\mathrm{ext})})\gamma^0\psi + 2\lambda(\bar{\psi}\psi)\psi = 0, \quad (3a)$$

$$\Box A_{e_0} - e\psi^+\psi = 0. \quad (3b)$$

It is easy to show that if $j = 1/2$ the solutions can be factorized in spherical coordinates as[20]

$$\psi = e^{-i\Omega m t} m^{3/2} \begin{pmatrix} G(\rho)\,\mathscr{Y}^{j}_{\frac{1}{2}l} \\ iF(\rho)\mathscr{Y}^{j}_{\frac{1}{2}l'} \end{pmatrix}, \quad (4a)$$

$$A_{e_0} = \frac{m}{e} V_e(\rho), \quad \rho = mr, \quad (4b)$$

where G, F, and V_e are dimensionless radial functions. If $j \neq 1/2$, the solution cannot be factorized and we will not consider this case. The field equations take the form:

$$F' + [(1 - \kappa/\rho)]F - (1 - \Omega + V_e + V_p)G + (\lambda m^2/2\pi)(G^2 - F^2)G = 0, \quad (5a)$$

$$G' + [(1 + \kappa/\rho)]G - (1 + \Omega - V_e - V_p)F + (\lambda m^2/2\pi)(G^2 - F^2)F = 0, \quad (5b)$$

$$V_e'' + (2/\rho)V_e' + \alpha(F^2 + G^2) = 0, \quad (5c)$$

where prime means $d/d\rho$, $V_p = -\alpha/\rho$, and κ is the eigenvalue of $-\beta(\Sigma \mathbf{L} + 1)$.[19]

The energy of the system can be calculated from the symmetric energy-momentum tensor

$$T^{\alpha\beta} = \tfrac{1}{4}i[\bar{\psi}\gamma^\alpha\partial^\beta\psi + \bar{\psi}\gamma^\beta\partial^\alpha\psi - (\partial^\alpha\bar{\psi})\gamma^\beta\psi - (\partial^\beta\bar{\psi})\gamma^\alpha\psi] + g^{\alpha\beta}\lambda(\bar{\psi}\psi)^2 + F^{\lambda\alpha}F^{\beta}{}_{\lambda} + \tfrac{1}{4}g^{\alpha\beta}F^{\mu\nu}F_{\mu\nu} - \tfrac{1}{2}e(\bar{\psi}\gamma^\alpha\psi A^\beta + \bar{\psi}\gamma^\beta\psi A^\alpha), \quad (6)$$

from which follows the expression for the energy,

$$E = m\left(\Omega\int_0^\infty (F^2 + G^2)\rho^2\,d\rho + \frac{\lambda m^2}{4\pi}\int_0^\infty (F^2 - G^2)^2\rho^2\,d\rho + \frac{1}{2\alpha}\int_0^\infty (V_e')^2\rho^2\,d\rho - \int_0^\infty V_e(F^2 + G^2)\rho^2\,d\rho\right), \quad (7)$$

while the norm and charge are

$$N = \int_{\mathbf{R}^3} \psi^+\psi\,d^3\mathbf{r} = \int_0^\infty (F^2 + G^2)\rho^2\,d\rho, \quad (8)$$

$$Q = eN. \quad (9)$$

The angular momentum $\mathbf{J}$ can be calculated by using the tensor

$$J^{\alpha\beta\gamma} = x^\alpha T^{\beta\gamma} - x^\beta T^{\alpha\gamma}, \quad (10)$$

from which it follows that

$$\mathbf{J} = (0,0,j^3 N). \quad (11)$$

The magnetic moment is

$$\mathbf{M} = \tfrac{1}{2}\int \mathbf{r}\times e\bar{\psi}\boldsymbol{\gamma}\psi\,d^3\mathbf{r} = (0,0,\mathscr{M}), \quad (12)$$

which can be written as

$$\mathscr{M} = -(2e/3m)(-1)^l \int_0^\infty FG\rho^3\,d\rho. \quad (13)$$

The finite energy solutions of (5) were obtained by using the following procedure.

If a solution is to be regular at the origin, V_e' and F in the case of S waves and V_e' and G in the case of P waves must vanish at $\rho = 0$. Let us take an S wave. We first give values to λm^2, Ω, and $V_e(0)$ and obtain a family of solutions which depend on $G(0)$ and which in general diverge when $\rho \to \infty$. However, it turns out that there is a special value G_0 of $G(0)$ which separates different types of divergence ($G \to +\infty$ from $G \to -\infty$ and so on). When $G(0) = G_0$ the behavior of the solution when $\rho \to \infty$ is

$$G,\ F \sim e^{-a\rho}/\rho,$$

$$V_e \sim V_e(\infty) = \text{const}, \quad (14)$$

where a is a certain constant. Once this solution is obtained, we make the subtraction

$$V_e \to V_e^T = V_e - V_e(\infty),$$
$$\Omega \to \Omega^T = \Omega - V_e(\infty) = \Omega_l + \Delta\Omega, \tag{15}$$

where Ω_l is the linear frequency corresponding to the same number of nodes. The functions G, F, V_e^T are the finite energy solutions, corresponding to the frequency Ω^T. This is a consequence of the gauge invariance of the theory which implies that Ω and V_e appear in (5) only through their difference. From now on we will only refer to the transformed quantities V_e^T, Ω^T but we will omit the superindex T for simplicity.

The numerical procedure must be considered very carefully because the interval of integration must be very long as the field is very extended in relation to its Compton wavelength. Moreover, the energy must be known with great accuracy. To reduce the singularity at $\rho = 0$ the change of independent variable $\rho^* = \ln\rho$ was made. Most of the calculations were made with a Runge–Kutta sixth order method but the results were checked with a Runge–Kutta fourth order and the Hamming predictor-corrector methods and by using different stepsizes. As a test, the following integral relations were verified.

First of all we have

$$\frac{d}{d\sigma} L\{\psi(\sigma r, t), A_{e_0}(\sigma r)\}_{\sigma=1} = 0, \tag{16}$$

which is a generalization of the Rosen[20] pseudovirial theorem due to Vázquez.[21] In our case it implies

$$\int_0^\infty (\Omega(F^2+G^2) - (G^2-F^2))\rho^2\,d\rho - \int_0^\infty (G^2+F^2)(V_e+V_p)\rho^2\,d\rho + \frac{1}{2\alpha}\int_0^\infty V_e'^2\rho^2\,d\rho - \frac{\lambda m^2}{4\pi}\int_0^\infty (G^2+F^2)^2\rho^2\,d\rho = 0. \tag{17}$$

From (5c) it can be easily shown that

$$V_e(0) = \alpha\int_0^\infty (F^2+G^2)\rho\,d\rho. \tag{18}$$

Multiplying (5a) by F, (5b) by G, and adding them leads to

$$\int_0^\infty FG\rho^3\,d\rho = \tfrac{1}{2}(1-\kappa)\int_0^\infty F^2\rho^2\,d\rho + \tfrac{1}{2}(1+\kappa)\int_0^\infty G^2\rho^2\,d\rho - \tfrac{3}{4}\int_0^\infty (F^2+G^2)\rho^2\,d\rho + (\lambda m^2/2\pi)\int_0^\infty (G^2-F^2)FG\rho^3\,d\rho. \tag{19}$$

Finally by integration of the equation $\mathrm{div}\mathbf{E} = e\psi^+\psi$ it can be shown that

$$\int_0^\infty V_e(F^2+G^2)\rho^2\,d\rho = (1/\alpha)\int_0^\infty V_e'^2\rho^2\,d\rho. \tag{20}$$

The integral conditions (17), (18), (19), and (20) were checked in all the calculations and they were found to be verified with six to eigth significant figures. This is an strong indication that the accuracy obtained is enough to support our conclusions.

Let us consider first the case $\lambda m^2 = 0$. It is possible to define the principal quantum number n by the number of nodes as usual. The results are typical of a bifurcation problem.[22,17] The solutions bifurcate from the zero solution $\psi = 0$ at $\Omega = \Omega_l$, the linear eigenvalues. We have thus a family of solutions which depend continuously on Ω, for each value of n and such that their norm N goes to zero when $\Omega \to \Omega_l$. Curiously enough we find linear results in this very nonlinear problem. We have found that the quantities $\Delta\Omega/N$, $\Delta E/N^2$ and $\mathcal{M}/N$ are linear functions of the norm N in all the waves that we have considered, $1S$,$2S$,$3S$,$2P$,$3P$. Let us take the ground state $1S$. We have

$$\Delta\Omega/N = a + bN, \tag{21}$$

with

$$a = 0.625\,635\,\alpha^2,$$
$$b = -0.172\,513\,\alpha^2. \tag{22}$$

Formula (21) interpolates eight solutions corresponding to different values of N and its correlation coefficient is $r = 0.9999$, which suggests that we have indeed a straight line.

The physical solution corresponds to unit norm and has a frequency

$$\Omega(N=1) = \Omega_{1S} + \Delta\Omega(N=1), \tag{23}$$

with $\Delta\Omega(N=1) = 0.453121\alpha^2$. This shift in the frequency is a consequence of the self-screening of the electron field. As $|\Omega| < 1$, in order to have a finite energy solution, it turns out that there is a maximum value of N which corresponds to $\Omega = 1$. This value is $N = 1.189$. In other words the self-repulsion of the Dirac field makes it impossible to bound more than 1.189 times the electron charge. This does not contradict the possiblity of an ion H^-, because in this case the two electrons are in different states. It should be noted that similar bounds can also be found in quantum mechanics.[23]

We have found for the energy

$$\Delta E/mN^2 = c + dN, \tag{24}$$

where

$$\Delta E = E - N m \Omega_{1S},$$
$$c = 0.313\,886\,\alpha^2,$$
$$d = -0.058\,0012\,\alpha^2. \tag{25}$$

This approximation has a correlation coefficient $r = 0.998$. The solution with $N = 1$ has an energy

$$E = m(1 - 0.243\,894\,\alpha^2), \tag{26}$$

which is greater than the experimental value, the difference being about $\alpha^2/4$. It is due to the Coulomb self-repulsion of ψ.

We have also calculated the magnetic field. From (20) and (14) we have

$$\frac{\mathcal{M}}{N} = -\left(\frac{g}{2} - \frac{4}{3}\frac{\int_0^\infty F^2\rho^2\,d\rho}{N} + \frac{4}{3}\frac{\lambda m^2}{2\pi}\frac{\int_0^\infty (G^2 - F^2)^2\,FG\rho^3\,d\rho}{N}\right)\mu_B, \tag{27}$$

where g is the Landé factor and μ_B is the Bohr magneton. As F and G are small and large respectively as in the usual linear theory the ratio $\mathcal{M}/N$ is correct at order zero in α. The magnetic moment depends on the norm N as

$$\mathcal{M}/N = e + fN, \tag{28}$$

where

$$e = -0.999\,982\,307\,\mu_B,$$
$$f = -0.939\,5 \times 10^{-5}\,\mu_B, \tag{29}$$

the correlation coefficient being $r = 0.99999$ and the value for $N = 1$,

$$\mathcal{M} = -(1 - 0.155\,827\,\alpha^2)\mu_B. \tag{30}$$

The mean square charge radius is also an interesting quantity. Near the bifurcation value it takes the Dirac linear value

$$\langle r^2\rangle^{1/2} = (\sqrt{3})/\alpha m = 237.35/m. \tag{31}$$

When the norm increases the growing Coulomb repulsion tends to make bigger this radius. For $N = 1$ it has the value $\langle r^2\rangle^{1/2} = 419/m$. All these results are represented in the Fig. 1.

In the waves $2S,3S,2P_{1/2},3P_{1/2}$ we found completely analogous results, with similar values of the correlation coefficients of the straight lines always very close to one. It seems therefore that the pattern is general.

The next thing to do is to study the effect of the value of λ, that is of the self-coupling of ψ. It is clear that if the fields are not submitted to an exterior field this effect is very important. However, in our case it turns out to be negligible because it is dominated by the self-screening. In fact if $|\lambda m^2| < 10^5$ the value of $\Delta\Omega/N$ keeps five significant figures unchanged. The smallness of this effect is due to the great extension of the Dirac field.

IV. THE RADIATIVE CORRECTIONS: CLASSICAL RENORMALIZATION

As we have seen the Coulomb self-repulsion of the electron field gives rise to an extra energy which in the ground state has a value close to $m\alpha^2/4$. In the other waves it is close to $m\alpha^2/4n^2$. This seems to imply that the predicted energies

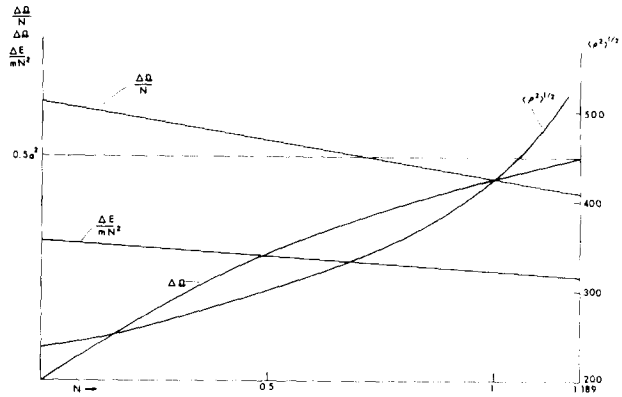


FIG. 1. $\Delta\Omega/N$, $\Delta\Omega$, $\Delta E/mN^2$ and $\langle\rho^2\rangle^{1/2}$ vs the norm N in the case of the wave $1S_{1/2}$.

are different from the experimental results, making meaningless this classical approach.

The same effect however gives an infinite energy when calculated by means of quantum field theory. Only after the theory is renormalized the correct result is obtained. This suggests that our extra energy could be eliminated by means of a classical renormalization scheme.

Let us take for the coupling constant e the value

$$e = Xe_0, \tag{32}$$

where e_0 is the experimental value and X an undetermined constant. We will see that it is possible to find a value of X such that the charge and energy have the right values, at least at order α^2.

As the charge Q has the value $Q = eN$ the condition $Q = eN$ the condition $Q = e_0$ implies $NX = 1$. In the case $X = 1$ it follows from (25) that the energy depends on the norm as

$$E = mN(1 - \alpha^2/2n^2 + cN + dN^2), \tag{33}$$

but, what happens if $X \neq 1$? In order to answer this question we will use the Ritz–Galerkin method[24] and expand the solutions in the basis of the usual Dirac eigenfunctions of the linear problem, with coupling constant Xe_0, which we call $\{\psi_k(X,\mathbf{r})\}$. Being only interested in α^2 corrections we can write

$$\psi_k(X,\mathbf{r}) = X^{3/2}\,\psi_k(1,X\mathbf{r}), \tag{34}$$

where $\psi_k(1,\mathbf{r})$ is the k th eigenfunction of the electron in the hydrogen atom in the Dirac theory. In fact (34) is only correct for the large components but it gives an approximation good enough because we will only consider α^2 corrections.

We thus express our solution as

$$\psi_X = e^{-i\omega_X t}\sum_k a_{kX}\,\psi_k(X,\mathbf{r}), \tag{35}$$

and after substitution in the field equations and projection on $\psi_i(X,\mathbf{r})$, we find

$$(\omega_X - \omega_{iX})a_{iX} = \sum_k \int \psi_i^+(X,\mathbf{r})Xe_0A_{e0}(X,\mathbf{r})a_{kX}\psi_k(X,\mathbf{r})\,d^3\mathbf{r}, \tag{36}$$

where $\omega_{iX} = [1 - X^2\alpha^2/(2n_i^2)]m$, n_i being the principal quantum number of the ith state. We have

$$A_{e_0} = \frac{Xe_0}{4\pi} \int \frac{\Sigma_{l,n} a^*_{lX} a_{nX} \psi_l^+(X,\mathbf{r}') \psi_n(X,\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}, \quad (37)$$

which can be written as

$$A_{e_0}(X,\mathbf{r}) = \frac{X^2 e_0}{4\pi} \int \frac{\Sigma_{l,n} a^*_{lX} a_{nX} \psi_1^+(1,X\mathbf{r}') \psi_n(1,X\mathbf{r}')}{|X\mathbf{r}' - X\mathbf{r}|} d^3(X\mathbf{r}')$$
$$= X^2 F_0(a^*_{lX},\ a_{nX},\ X\mathbf{r}), \quad (38)$$

where F_0 is a homogeneous quadratic function of a_{kX}. We now define $A_{ik}(a^*_{lX},\ a_{nX})$ as

$$A_{ik}(a^*_{lX},\ a_{nX}) = e_0 \int \psi_i^+(1,X\mathbf{r})\, F_0(a^*,a,X\mathbf{r}) \psi_k(1,X\mathbf{r})\, d^3(X\mathbf{r}), \quad (39)$$

after which we can write (36) as

$$(\omega_i - \omega_{iX}) A_{iX} = X^3 \sum_k A_{ik}(a^*_{lX},\ a_{nX}) a_{kX}. \quad (40)$$

If we write $\omega_X = (1 - \nu\alpha^2 X^2)m$ and $a_{kX} = b_k/\sqrt{X}$, we have

$$(1/2n_i^2 - \nu)\alpha^2 b_i = \sum_k A_{ik}(b^*, b) b_k, \quad (41)$$

which does not depend on X. In other words the dependence of the solution on X can be written explicitly as

$$\psi_X = e^{-i\Omega m t} \sum (a_k/\sqrt{X}) \psi_k(X,\mathbf{r}),$$
$$\Omega = 1 - X^2\alpha^2\nu. \quad (42)$$

This allows the extrapolation of all our results from $X = 1$ to any other value. In particular, it can be shown that the function $E(N,X)$ is

$$E = mN\,[1 - (\alpha^2/2n^2)X^2 + cNX^3 + dN^2X^4], \quad (43)$$

which if $X = 1$ reduces to (33). In the case of a neutral atom $NX = 1$, from which

$$E = mN(1 - X^2 C_{nl}\,\alpha^2). \quad (44)$$

If we take $m = Xm_0$, m_0 being the experimental value of the electron mass, we have

$$E = m_0(1 - X^2 C_{nl}\,\alpha^2). \quad (45)$$

The numbers C_{nl} were calculated numerically in five waves. The results were

$$C_{10} = 0.243\ 894, \quad C_{21} = 0.053\ 25,$$
$$C_{20} = 0.060\ 545, \quad C_{31} = 0.025\ 32.$$
$$C_{30} = 0.026\ 698, \quad (46)$$

If $X^2 = 2.0453$ the energy difference between the $2S$ and $1S$ states takes the experimental value. The energy differences $3S - 1S$ and $3S - 2S$ differ by $-4.0998\,\alpha^4$ from the experimental value. In other words the three S waves have the correct energies at α^2 order. The results concerning the p waves are worse. The difference between the renormalized and the experimental energies are $280\alpha^4$ $(2P - 1S)$ and $49\alpha^4$ $(3P - 1S)$. However, in these cases the computational problem seems to be more difficult. In the case of the P waves the distinction between the two modes of divergence of the solution, which is necessary to find the finite energy solutions, is more difficult than in S waves. For instance, if $l = 0$ it is enough to integrate from $\rho = 0$ to $\rho = 4{,}500$ ($r \simeq 20$ Å). For some unknown reason this is not enough if $l = 1$. In this case we have gone until $\rho \simeq 9000$ ($r \simeq 40$ Å). As one must take a steplength $h = 10^{-3}$ or even smaller a considerable amount of computer time is necessary. We have been able to obtain 13 figures of $G(0)$ in the S waves but only 8 of $F(0)$ in the P waves. This suggests that the worse behavior of the P waves has a computational origin and that the renormalization scheme may be correct in all the states. The renormalization also affects the value of $\langle r^2\rangle^{1/2}$ which was too high. The renormalized value is $204.86/m_0$ very close to the standard value $237.35/m_0$.

In this method not only the energy differences but also the energies have right values at order α^2. It is perhaps worth while to mention that there is a much simpler procedure which gives good differences but bad energies. It consists in using the experimental value of the coupling constant, $e = e_0$, but a renormalized value for the mass, $m = X^2 m_0$.

To go beyond α^2 order turns out to be very difficult. First of all the magnetic effects must be included and this implies an increase in the number of radial differential equations. As the accuracy has to be much better, an improved numerical scheme must be used as well as more computer time. Moreover as the solution is no longer separable in spherical coordinates, not even if $j = 1/2$, a multipole expansion must be used. It seems very difficult to obtain any result without the use of analytical techniques.

V. SUMMARY

We have studied the effect of the interaction of the Dirac spinor with its own electromagnetic field in the partial differential equation which appears in the hydrogen atom. The self-screening gives raise to an extra energy which can be eliminated by means of a classical renormalization procedure.

As the treatment is classical, this suggest that the classical study of the partial differential equations which appear in quantum mechanics may be a useful tool to understand the relation between classical and quantum physics.

ACKNOWLEDGMENT

We are very indebted to Dr. J. Campos, Dr. L. García, and Dr. L. Vázquez for discussions.

[1]See for instance, *Nonlinear Problems in Theoretical Physics*, edited by A. F. Rañada, Lecture Notes in Physics **98** (Springer, Berlin, 1979).
[2]V. G. Makhankov, G. Kummer, and A. B. Shvachka, Phys. Scr. **20**, 454.
[3]N. Rosen, Phys. Rev. **55**, 94 (1939).
[4]R. Finkelstein, R. Lelevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[5]R. Finkelstein, C. Fronsdal, and P. Kaus, Phys. Rev. **103**, 1571 (1956).
[6]M. Wakano, Prog. Theor. Phys. **35**, 1117 (1966).
[7]M. Soler, Phys. Rev. D **1**, 2766 (1970).
[8]M. Soler, Phys. Rev. D **8**, 3424 (1973).

[9]A. F. Rañada, M. R. Rañada, M. Soler, and L. Vazquez, Phys. Rev. D **10**, 517 (1974).
[10]A. F. Rañada and L. Vazquez, Prog. Theor. Phys. **56**, 311 (1976).
[11]L. Garcia and A. F. Rañada. Prog. Theor. Phys. (Kyoto) **64**, 671 (1980).
[12]A. F. Rañada, J. Phys. A: Gen. Phys. **11**, 341 (1978).
[13]A. F. Rañada, to be published in the proceedings of the Istanbul–Trabzon Colloquia on Mathematical Physics (1979).
[14]I. Bialinicky–Birula and J. Mycielski, Ann. Phys. (N.Y.) **100**, 62 (1976).
[15]A. F. Rañada, Int. J. Theor. Phys. **16**, 795 (1977).
[16]A. F. Rañada and L. Vazquez, Phys. Rev. D **19**, 493 (1979).
[17]A. F. Rañada and J. Uson, J. Math. Phys. **21**, 1205 (1980).
[18]A. I. Akhiezer and V. B. Beretetskii, *Quantum Electrodynamics* (Wiley, New York, 1965).
[19]J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading Mass., 1976).
[20]G. Rosen, J. Math. Phys. **7**, 2066 (1966).
[21]L. Vazquez, J. Math. Phys. **18**, 1343 (1977).
[22]D. Sattinger, *Topics in Stability and Bifurcation Theory*, Lecture Notes in Mathematics **309** (Springer, Berlin, 1973).
[23]A. Galindo and R. F. Alvarez–Estrada, Nuovo Cimento B **44**, 47 (1978).
[24]F. W. Byron and R. W. Fuller, *Mathematics of Classical and Quantum Physics* (Addison-Wesley, Reading, Mass., 1969).

On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. I

Asao Arai
Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan

(Received 4 February 1981; accepted for publication 18 May 1981)

As a first step towards giving a rigorous mathematical interpretation to the Lamb shift, a system of a harmonic oscillator coupled to a quantized, massless, scalar field is studied rigorously with special attention to the spectral property of the total Hamiltonian. It is proved that the point spectra of the harmonic oscillator disappear completely into the continuous spectrum due to the interaction with the quantized field.

PACS numbers: 03.65.Db, 03.65 − w, 03.70. + k

I. INTRODUCTION

The theory of the Lamb shift has been one of the brilliant achievements in quantum electrodynamics and has given a support for its validity. However, it has so far been formulated only in terms of the formal perturbation theory as a shift of point spectra and the rigorous mathematical foundation has not been given. Rigorous mathematical analysis is indeed necessary, because, in view of the finite lifetime of the excited states of the atom due to the spontaneous emission of light, one cannot expect that the Lamb shift can be described simply in terms of the shift of the point spectra. However, we do not as yet have a rigorous mathematical framework for quantum electrodynamics.

In this paper we consider the problem rigorously by using a simplified, exactly soluble model, first considered by Schwabl and Thirring,[1] which describes a system of a one-dimensional, nonrelativistic harmonic oscillator interacting with a quantized, massless, neutral scalar field in three space dimensions, where the harmonic oscillator may be compared to an atom and the quantized scalar field to a radiation field. All the point spectra of the unperturbed Hamiltonian are embedded in its continuous spectrum. Therefore, from the perturbation theoretical point of view, the analysis of the spectrum of the total Hamiltonian gives a problem of perturbation of point spectra embedded in a continuous spectrum, which is hard to analyze in general. We shall show that, under the perturbation, all the point spectra but the lowest one of the unperturbed Hamiltonian disappear into the continuous spectrum. Namely, for the total Hamiltonian, we prove the absence of point spectra for excited states and the existence and the uniqueness of ground state. We also prove that the continuous spectrum of the total Hamiltonian is $[E_0, \infty)$, where E_0 is the ground state energy.

The disappearance of the point spectra of the unperturbed Hamiltonian is natural in view of the spontaneous emission of the boson and gives a support for the expectation mentioned above in regard to the Lamb shift. Then we have a problem of how we should give a rigorous mathematical interpretation to the Lamb shift that the formal perturbation calculation gives. We conjecture, in the analogy to the other examples of perturbation of point spectra embedded in a continuous spectrum, such as the Auger effect of the helium atom (see Ref. 2, Sec. XII. 6), that the problem, together with that of light emission, should be formulated in terms of "resonances." The problem of resonances in our model will be discussed in a subsequent paper.

The outline of the paper is as follows. In Sec. II we define the model. In order to make the total Hamiltonian well defined, we introduce an ultraviolet cutoff in the interaction. In Sec. III we prove the self-adjointness of the total Hamiltonian. In Sec. IV we construct the exact solution of the Heisenberg equations for the dynamical variables, where some technical lemmas are needed. In Sec. V we find the explicit form of the asymptotic annihilation and creation operators by making use of the exact solution obtained in Sec. IV. In Sec. VI we analyze the spectrum of the total Hamiltonian, establishing the absence of point spectra for excited states and the existence and the uniqueness of ground state, where the explicit form of the asymptotic annihilation and creation operators is essentially used. In the last section we briefly remark on the other properties of the model: the scattering theory and the point limit of the interaction (the removal of the ultraviolet cutoff).

II. DEFINITION OF THE MODEL

The Hilbert space $\mathscr{H}$ for the model is defined as the tensor product of $L^2(\mathbf{R})$, the space of wavefunctions for the harmonic oscillator, and $\mathscr{F}$, the boson Fock space over $L^2(\mathbf{R}^3)$:

$$\mathscr{H} = L^2(\mathscr{R}) \otimes \mathscr{F}. \tag{2.1}$$

The Hamiltonian H_{0h} for the harmonic oscillator is given by

$$H_{0h} = \tfrac{1}{2}(p^2 + q^2), \quad p = -i\frac{d}{dq}, \tag{2.2}$$

where we take the mass and the spring constant of the harmonic oscillator both to be equal to 1. H_{0h} is self-adjoint with the domain $D(H_{0h}) = D(p^2) \cap D(q^2)$.

Let $a(f)$ and $a^*(f)$, $f \in L^2(\mathbf{R}^3)$, be the boson annihilation and creation operators, respectively, which are densely defined on $\mathscr{F}_F$, the set of finite particle vectors in $\mathscr{F}$, and leave it invariant, satisfying

$$(a(f)\Psi, \Phi) = (\Psi, a^*(\bar{f})\Phi), \tag{2.3}$$

$$[a(f), a^*(g)]\Psi = (\bar{f}, g)_2 \Psi, \tag{2.4}$$

$$[a(f), a(g)]\Psi = [a^*(f), a^*(g)]\Psi = 0, \tag{2.5}$$

 0022-2488/81/112539-10$01.00

for all Ψ,Φ in $\mathscr{F}_F$ and all f,g in $L^2(\mathbf{R}^3)$. The time zero field is given by

$$\phi_0(f) = (1/\sqrt{2})\{a^*(\hat{f}/\sqrt{\omega}) + a(\tilde{\hat{f}}/\sqrt{\omega})\}, \quad \hat{f}/\sqrt{\omega}\in L^2(\mathbf{R}^3), \tag{2.6}$$

where $\hat{f}$ denotes the Fourier transform of f and $\tilde{g}$ is defined by

$$\tilde{g}(\mathbf{k}) = g(-\mathbf{k}), \tag{2.7}$$

and

$$\omega(\mathbf{k}) = |\mathbf{k}|. \tag{2.8}$$

We denote the free Hamiltonian for massless bosons by H_{0b}, which is a nonnegative self-adjoint operator in $\mathscr{F}$ and is symbolically expressed as

$$H_{0b} = \int d^3\mathbf{k}|\mathbf{k}|a^*(\mathbf{k})a(\mathbf{k}). \tag{2.9}$$

All the operators in $L^2(\mathbf{R})$ or in $\mathscr{F}$ have natural extensions to $\mathscr{H}$ {e.g., p [resp. $a(f)$] is extended as $p\otimes I$ [resp. $I\otimes a(f)$]}. They will be denoted by the same notation.

The interaction Hamiltonian H_I of the harmonic oscillator with the quantized scalar field is defined by

$$H_I = \lambda q\phi_0(\rho), \tag{2.10}$$

where $\lambda\in\mathbf{R}\setminus\{0\}$ denotes the coupling constant and ρ a function of which Fourier transform is an ultraviolet cutoff factor. Throughout the paper we assume that $\hat{\rho}$ is a rotation invariant function satisfying

$$\hat{\rho}>0,\ \hat{\rho}\in C^2(\mathbf{R}^3),\ \hat{\rho}/\sqrt{\omega}\in L^2(\mathbf{R}^3). \tag{2.11}$$

The total Hamiltonian H for the model is given formally by

$$H = H_0 + H_I + R, \tag{2.12}$$

where

$$H_0 = H_{0h} + H_{0b}, \tag{2.13}$$

$$R = \tfrac{1}{2}\delta K\, q^2,\ \delta K = \lambda^2||\hat{\rho}/\omega||_2^2. \tag{2.14}$$

The operator H_0, which is nonnegative and self-adjoint, is the unperturbed Hamiltonian of the system, and the operator R corresponds to the spring constant renormalization of the harmonic oscillator.

Remark: All the eigenvalues of H_0 are embedded in the continuous spectrum. Namely, we have

$$\sigma(H_0) = \sigma_c(H_0) = [\tfrac{1}{2},\infty),$$
$$\sigma_p(H_0) = \{n+\tfrac{1}{2}\}_{n=0}^{\infty}, \tag{2.15}$$

where $\sigma(H_0)$ [resp. $\sigma_c(H_0)$, $\sigma_p(H_0)$] denotes the spectrum [resp. the continuous spectrum, the point spectrum] of H_0. Therefore, the analysis of the spectrum of H gives a problem of perturbation of point spectra embedded in a continuous spectrum.

III. SELF-ADJOINTNESS OF THE TOTAL HAMILTONIAN

In this section we prove the self-adjointness of the total Hamiltonian H given by (2.12).

By using the well-known estimates

$$||a(f)\Psi|| \leq ||f/\sqrt{\omega}||_2\,||H_{0b}^{1/2}\Psi||,$$
$$||a^*(f)\Psi|| \leq ||f/\sqrt{\omega}||_2\,||H_{0b}^{1/2}\Psi|| + ||f||_2||\Psi||, \tag{3.1}$$

we can show that

$$||H_I\Psi|| \leq |\lambda|\{(\sqrt{2}||\hat{\rho}/\omega||_2 + \epsilon||\hat{\rho}/\sqrt{\omega}||_2)||\tilde{H}_0\Psi|| + (1/2\epsilon)||\hat{\rho}/\sqrt{\omega}||_2||\Psi||\}, \tag{3.2}$$

$$||R\Psi|| \leq \delta K||\tilde{H}_0\Psi||, \tag{3.3}$$

for all Ψ in $D(\tilde{H}_0)$, where

$$\tilde{H}_0 = H_{0b} + \tfrac{1}{2}q^2, \tag{3.4}$$

and $\epsilon>0$ is arbitrary. Therefore, in particular, it follows that H_I is a well-defined symmetric operator on $D(\tilde{H}_0)$.

Lemma 3.1: Let

$$\tilde{H} = \tilde{H}_0 + H_I + R. \tag{3.5}$$

Then, $\tilde{H}$ is self-adjoint with $D(\tilde{H}) = D(\tilde{H}_0)$ and

$$\tilde{H}\geq\tfrac{1}{2}q^2. \tag{3.6}$$

The proof of this lemma is quite similar to that of Lemma 3.3 in Ref. 3 and is omitted.

In terms of $\tilde{H}_0$ and $\tilde{H}$ we can write

$$H_0 = \tilde{H}_0 + \tfrac{1}{2}p^2, \tag{3.7}$$

$$H = \tilde{H} + \tfrac{1}{2}p^2. \tag{3.8}$$

Lemma 3.2: H is a closed symmetric operator on $D(H_0)$.

Proof: It is clear that H is symmetric on $D(H_0)$. Since $\tilde{H}$ is nonnegative by (3.6), we have

$$\mathrm{Re}(\tilde{H}\Psi, p^2\Psi)\geq\mathrm{Re}([p,\tilde{H}]\Psi, p\Psi) \tag{3.9}$$

for all Ψ in $D(H_0^2)$. Furthermore, by using commutation relations, we have

$$[p,\tilde{H}]\Psi = -i\{(1+\delta K)q + \lambda\phi_0(p)\}\Psi,$$

which, together with (3.9), yields

$$\mathrm{Re}(\tilde{H}\Psi, p^2\Psi)\geq -\tfrac{1}{2}(1+\delta K)||\Psi||^2.$$

It is easy to see that H_I and p^2 are H_0-bounded [cf. (3.2) and (3.7)]. Therefore, the above inequality extends to all Ψ in $D(H_0)$. Thus, we obtain

$$||\tilde{H}||^2 + ||\tfrac{1}{2}p^2\Psi||^2 \leq ||H\Psi||^2 + \tfrac{1}{2}(1+\delta K)||\Psi||^2 \tag{3.10}$$

for all Ψ in $D(H_0)$. Since $\tilde{H}$ is a closed operator on $D(\tilde{H}_0)$ by Lemma 3.1, (3.10) shows that H is a closed operator on $D(\tilde{H}_0)\cap D(p^2) = D(H_0)$.

Theorem 3.1: H is self-adjoint with $D(H) = D(H_0)$ and

$$H\geq\tfrac{1}{2}. \tag{3.11}$$

Proof: Let $L = H_0 + 1$. Since H_I and R are $\tilde{H}_0$-bounded by (3.2) and (3.3) and hence are H_0-bounded, we have

$$||H\Psi|| \leq c||L\Psi||, \qquad \Psi\in D(H_0),$$

for some constant $c>0$. Suppose for the moment that $(\omega)^{1/2}\hat{\rho}\in L^2(\mathbf{R}^3)$. Then we have

$$[H,L]\Psi = \{(\lambda/\sqrt{2})q[a(\omega^{1/2}\hat{\rho}) - a^*(\omega^{1/2}\hat{\rho})] + i\lambda p\phi_0(\rho) + (i/2)\delta K(pq+qp)\}\Psi$$

for all Ψ in $D(H_0^2)$. Therefore, we get

$$|(H\Phi,L\Psi) - (L\Phi,H\Psi)| \leq d||L^{1/2}\Phi||\,||L^{1/2}\Psi||, \qquad \Phi\in D(H_0),\ \Psi\in D(H_0^2),$$

for some constant $d>0$. This inequality extends to all Ψ in $D(H_0)$. Thus, it follows from Nelson's commutator theorem

(see, e.g., Ref. 4, Sec.X.5) that H is essentially self-adjoint on $D(H_0)$, which, together with Lemma 3.2, implies that H is self-adjoint with $D(H) = D(H_0)$. Furthermore, by a limiting argument using strong resolvent convergence, we can extend the result to all $\hat{\rho}$ satisfying $\hat{\rho}/\sqrt{\omega} \in L^2(\mathbf{R}^3)$. Inequality (3.11) follows from (3.6) and the fact that $H_{0h} \geqslant \frac{1}{2}$.

Remarks: (1) As is seen from the proof of Theorem 3.1, the continuity and positivity of $\hat{\rho}$ are not needed for the proof of the self-adjointness of H. In fact, we can prove the theorem for every real-valued, rotation invariant function $\hat{\rho}$ satisfying $\hat{\rho}/\sqrt{\omega}, \hat{\rho}/\omega \in L^2(\mathbf{R}^3)$. The additional assumptions for $\hat{\rho}$ will be needed in the construction of the exact solution of the Heisenberg equations for the dynamical variables (see Sec.IV).

(2) By Lemma 3.2 and the closed graph theorem, there exists a constant $c > 0$ such that

$$||H_0\Psi|| \leqslant c(||H\Psi|| + ||\Psi||)$$

for all Ψ in $D(H_0)$ (see Ref.5, p.79). Since H_0 and H are self-adjoint, it follows (see, e.g., Ref.4,Sec.X.2) that H_0 is form-bounded with respect to H, i.e.,

$$||H_0^{1/2}\Psi|| \leqslant d\,||(H+1)^{1/2}\Psi||, \qquad \Psi \in D(H^{1/2}), \tag{3.12}$$

for some constant $d > 0$. On the other hand, it is easy to see that H is form-bounded with respect to H_0. Therefore, we have $D(H^{1/2}) = D(H_0^{1/2})$.

IV. CONSTRUCTION OF EXACT SOLUTION OF THE HEISENBERG EQUATIONS

In this section we shall give the explicit form of the dynamical variables defined by

$$\phi(f,t) = e^{itH}\phi_0(f)e^{-itH}, \qquad \hat{f}/\sqrt{\omega}, \hat{f}/\omega \in L^2(\mathbf{R}^3), \tag{4.1}$$

$$q(t) = e^{itH}qe^{-itH}, \tag{4.2}$$

which are well-defined on $D(H^{1/2})$ [see Remark (2) in Sec. III]. Formally the Heisenberg equations for $\phi(f,t)$ and $q(t)$ give the following equations of motion:

$$\left(\frac{d^2}{dt^2} + 1\right) q(t) = -\lambda \int d^3\mathbf{x}\, \rho(\mathbf{x})\phi(\mathbf{x},t) - \delta K\, q(t),$$

$$\left(\frac{\partial^2}{\partial t^2} - \Delta\right)\phi(\mathbf{x},t) = -\lambda\, \rho(\mathbf{x})\, q(t), \tag{4.3}$$

where $\phi(\mathbf{x},t)$ is the symbolic notation given by

$$\phi(f,t) = \int d^3\mathbf{x}\, \phi(\mathbf{x},t) f(\mathbf{x}). \tag{4.4}$$

Equations (4.3) can be solved exactly by using the theory of Fourier transform or Laplace transform. However, we do not write down the formal exact solution here (cf. Ref.1). In order to construct the exact solution in a rigorous manner, we need some technical lemmas.

A. Preliminaries

Lemma 4.1: Let

$$(G_\epsilon f)(\mathbf{k}) = \int d^3\mathbf{k}' \frac{f(\mathbf{k}')}{(|\mathbf{k}||\mathbf{k}'|)^{1/2}(\mathbf{k}^2 - \mathbf{k}'^2 + i\epsilon)}, \qquad \epsilon \in \mathbf{R}\setminus\{0\}. \tag{4.5}$$

Then, G_ϵ is a bounded operator on $L^2(\mathbf{R}^3)$ such that

$$||G_\epsilon|| \leqslant c \tag{4.6}$$

for some constant $c > 0$ independent of ϵ.

Before starting the proof, we define the function Ff on $\mathbf{R}$ by

$$(Ff)(s) = |s|^{1/4}[f](|s|^{1/2})\chi_{[0,\infty)}(s), \tag{4.7}$$

for each f in $L^2(R^3)$, where the function $[f]$ on $[0,\infty)$ is defined by

$$[f](k) = \int_{S^2} d\Omega\, f(k\Omega) \tag{4.8}$$

and $\chi_{[0,\infty)}(s)$ is the characteristic function for the half-line $[0,\infty)$. It is easy to see that F is a bounded operator from $L^2(\mathbf{R}^3)$ to $L^2(\mathbf{R})$ with

$$||Ff||_{L^2(\mathbf{R})} \leqslant (8\pi)^{1/2}||f||_{L^2(\mathbf{R}^3)}. \tag{4.9}$$

Let

$$(A_\epsilon^{(1)} f)(s) = (Q_\epsilon * Ff)(s),$$

$$(A_\epsilon^{(2)} f)(s) = (P_\epsilon * Ff)(s), \qquad s \in R, f \in L^2(\mathbf{R}^3), \tag{4.10}$$

where

$$Q_\epsilon(s) = \frac{s}{\pi(s^2+\epsilon^2)}, \qquad P_\epsilon(s) = \frac{\epsilon}{\pi(s^2+\epsilon^2)}. \tag{4.11}$$

We have from Young's inequality

$$||P_\epsilon * g||_{L^2(\mathbf{R})} \leqslant ||g||_{L^2(\mathbf{R})}, \qquad g \in L^2(\mathbf{R}),$$

and by the theory of Hilbert transform,[6]

$$||Q_\epsilon * g||_{L^2(\mathbf{R})} \leqslant ||g||_{L^2(\mathbf{R})}, \qquad g \in L^2(\mathbf{R}).$$

Therefore, by using (4.9), we get

$$||A_\epsilon^{(j)} f||_{L^2(\mathbf{R})} \leqslant \sqrt{8\pi}\,||f||_{L^2(\mathbf{R}^3)}, \qquad j = 1,2, \tag{4.12}$$

for all f in $L^2(\mathbf{R}^3)$.

Proof of Lemma 4.1: We can write

$$G_\epsilon = G_\epsilon^{(1)} - i\, G_\epsilon^{(2)}. \tag{4.13}$$

where

$$(G_\epsilon^{(j)} f)(\mathbf{k}) = (\pi/2)|\mathbf{k}|^{-1/2}(A_\epsilon^{(j)} f)(|\mathbf{k}|^2), \qquad j = 1,2. \tag{4.14}$$

By using (4.12), we get

$$||G_\epsilon^{(j)} f||_{L^2(\mathbf{R}^3)} = (\pi^3/2)^{1/2}||A_\epsilon^{(j)} f||_{L^2(\mathbf{R}_+)} \leqslant 2\pi^2||f||_{L^2(\mathbf{R}^3)}, \qquad j = 1,2,$$

for all f in $L^2(\mathbf{R}^3)$. Thus the lemma follows.

Let

$$(A_{(1)} f)(s) = \frac{1}{\pi} P \int ds' \frac{(Ff)(s')}{s - s'},$$

$$(A_{(2)} f)(s) = (Ff)(s), \qquad f \in L^2(\mathbf{R}^3). \tag{4.15}$$

Then, it follows from (4.10) and the theory of Hilbert transform[6] that

$$\operatorname*{s-lim}_{\epsilon \to +0} A_\epsilon^{(j)} f = A^{(j)} f, \qquad j = 1,2 \tag{4.16}$$

in $L^2(\mathbf{R})$, which, together with (4.12), implies that the $A_{(j)}, j = 1,2$, are bounded operators from $L^2(\mathbf{R}^3)$ to $L^2(\mathbf{R})$.

Define

$$Gf = G^{(1)}f - iG^{(2)}f, \qquad f \in L^2(\mathbf{R}^3), \tag{4.17}$$

where

$$(G^{(j)}f)(\mathbf{k}) = (\pi/2)|\mathbf{k}|^{-1/2}(A_{(j)}f)(|\mathbf{k}|^2). \tag{4.18}$$

Lemma 4.2: G is a bounded operator on $L^2(\mathbf{R}^3)$ and

$$\text{s-lim}_{\epsilon \to +0} G_\epsilon = G. \tag{4.19}$$

Furthermore, G is skew-symmetric.

Proof: (4.19) follows from the identity

$$||G_\epsilon^{(j)}f - G^{(j)}f||_{L^2(\mathbf{R}^3)} = (\pi^3/2)^{1/2}||A_\epsilon^{(j)}f - A^{(j)}f||_{L^2(\mathbf{R}_+)}$$

and (4.16). (4.19) and (4.6) imply that G is a bounded operator on $L^2(\mathbf{R}^3)$. The skew-symmetry of G follows from that of G_ϵ and (4.19).

For each $\alpha \in R$ we define a Hibert space $M_\alpha(\mathbf{R}^3)$ by

$$M_\alpha(\mathbf{R}^3) = \{f \mid ||f||_\alpha \equiv ||\omega^\alpha f||_{L^2(\mathbf{R}^3)} < \infty\}. \tag{4.20}$$

Henceforth we denote the L^2-norm by $||\ \ ||_0$.

Lemma 4.3: G is a bounded operator on $M_{-1/2}(\mathbf{R}^3)$.

Proof: We have from the definition of $G^{(j)}$

$$||G^{(j)}f||^2_{-1/2} = \pi^3 \int_0^\infty dk\, |A^{(j)}f(k^2)|^2,$$

$$j = 1,2, f \in M_{-1/2}(\mathbf{R}^3).$$

By direct computations we have

$$(A^{(1)}f)(k^2) = B^{(1)}(k) + B^{(2)}(k), \qquad k > 0,$$

where

$$B^{(1)}(s) = \frac{1}{\pi} P \int_0^\infty dt \frac{(Ff)(t^2)}{(s-t)},$$

$$B^{(2)}(s) = -\frac{1}{\pi} P \int_{-\infty}^0 dt \frac{(Ff)(t^2)}{s-t}, \qquad s \in \mathbf{R}.$$

Therefore, we get

$$||G^{(1)}f||^2_{-1/2} \leqslant 2\pi^3 \sum_{j=1}^2 ||B^{(j)}||^2_{L^2(\mathbf{R})}.$$

By the isometry of Hilbert transform we have

$$\sum_{j=1}^2 ||B^{(j)}||^2_{L^2(\mathbf{R})} = \int_{-\infty}^\infty ds\, |(Ff)(s^2)|^2$$

$$= 2\int_0^\infty ds\, s|[f](s)|^2$$

$$\leqslant 8\pi ||f||^2_{-1/2}.$$

Thus, we get

$$||G^{(1)}f||_{-1/2} \leqslant 4\pi^2 ||f||_{-1/2}.$$

On the other hand, it is easy to see that $G^{(2)}$ is a bounded operator on $M_{-1/2}(\mathbf{R}^3)$. Thus the lemma follows.

We define

$$D(z) = -z + 1 + \delta K + \lambda^2 \int d^3\mathbf{k} \frac{\hat\rho(\mathbf{k})^2}{z - \mathbf{k}^2}, \tag{4.21}$$

which is analytic on $\mathbb{C}\setminus[0,\infty)$.

Lemma 4.4: For each $s \in [0,\infty)$

$$D_\pm(s) \equiv \lim_{\epsilon \to +0} D(s \pm i\epsilon) \tag{4.22}$$

exist and are continuous. Furthermore, we have

$$\inf_{s \in [0,\infty)} |D_\pm(s)| > 0. \tag{4.23}$$

Proof: Since $\hat\rho$ is rotation invariant, we write $\hat\rho(\mathbf{k}) = \hat\rho(|\mathbf{k}|)$. It follows from the assumptions for $\hat\rho$ that $s^{1/2}\hat\rho(\sqrt{s})^2$ is Lipshitz continuous on $[0,\infty)$ with index 1/2. Therefore,

$$\lim_{\epsilon \to +0} \int d^3\mathbf{k} \frac{\hat\rho(\mathbf{k})^2}{s \pm i\epsilon - \mathbf{k}^2} = 2\pi \lim_{\epsilon \to +0} \int_0^\infty dt \frac{t^{1/2}\hat\rho(\sqrt{t})^2}{s \pm i\epsilon - t}$$

exist for each $s \in [0,\infty)$ and are continuous,[6] so that the first half of the lemma follows. (4.23) can be proved easily by using the assumption $\hat\rho > 0$ and the continuity of $D_\pm(s)$.

Lemma 4.5: There exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$|D(z) + z| < c_1, \qquad |D(z)| > c_2 \tag{4.24}$$

for all $z \in \mathbb{C}\setminus[0,\infty)$.

The proof is easy and is omitted.

Lemma 4.6: Let

$$Q(\mathbf{k}) = -\lambda \frac{\hat\rho(\mathbf{k})}{D_+(\mathbf{k}^2)}. \tag{4.25}$$

Then,

$$||Q||_0 = 1. \tag{4.26}$$

Proof: We have

$$||Q||_0 = 2\pi\lambda^2 \int_0^\infty ds \frac{s^{1/2}\hat\rho(\sqrt{s})^2}{|D_+(s)|^2}.$$

On the other hand, we have

$$D_-(s) - D_+(s) = i4\pi^2\lambda^2 s^{1/2}\hat\rho(\sqrt{s})^2. \tag{4.27}$$

Since D_+ is the complex conjugate of D_-, we get

$$||Q||_0^2 = \frac{1}{2\pi i}\int_0^\infty ds \left\{\frac{1}{D_+(s)} - \frac{1}{D_-(s)}\right\}.$$

The right-hand side can be evaluated by using a contour integral for $1/D(z)$ in the cut plane $\mathbb{C}\setminus[0,\infty)$ and applying the Cauchy integral theorem, where Lemma 4.5 is used, and we get (4.26).

Lemma 4.7: Let T be an operator given by

$$Tf = f + \lambda\omega^{1/2}QG\omega^{1/2}\hat\rho f. \tag{4.28}$$

Then, T is a bounded operator on $L^2(\mathbf{R}^3)$.

Proof: The lemma follows from the boundedness of G and the fact that $\omega^{1/2}Q$ and $\omega^{1/2}\hat\rho$ are in $L^\infty(\mathbf{R}^3)$.

Remark: The operator T is symbolically given by a distribution kernel:

$$(Tf)(\mathbf{k}) = \int d^3\mathbf{k}'\, T(\mathbf{k},\mathbf{k}')f(\mathbf{k}'), \tag{4.29}$$

where

$$T(\mathbf{k},\mathbf{k}') = \delta^3(\mathbf{k} - \mathbf{k}') - \frac{\lambda^2\hat\rho(\mathbf{k}')\hat\rho(\mathbf{k}')}{(\mathbf{k}^2 - \mathbf{k}'^2 + i0)D_+(\mathbf{k}^2)}. \tag{4.30}$$

Since G is skew-symmetric (Lemma 4.2), the adjoint operator T^* of T in $L^2(\mathbf{R}^3)$ is given by

$$T^*f = f - \lambda\omega^{1/2}\hat\rho G\omega^{1/2}\bar{Q}f, \qquad f \in L^2(\mathbf{R}^3). \tag{4.31}$$

Lemma 4.8: T and T^* are bounded operators on $M_\alpha(\mathbf{R}^3)$ for $\alpha = \pm\frac{1}{2}, -1$.

Proof: The lemma follows from Lemma 4.2, Lemma 4.3, and the fact that Q, ωQ, $\hat{\rho}$, and $\omega\hat{\rho}$ are in $L^\infty(\mathbf{R}^3)$.

Lemma 4.9: The operator T as a bounded operator on $L^2(\mathbf{R}^3)$ has the following properties:

$$(1)\ T^*T = I, \tag{4.32}$$

$$(2)\ TT^* + (Q,\cdot)_0 Q = I, \tag{4.33}$$

$$(3)\ T^*Q = 0, \tag{4.34}$$

$$(4)\ \overline{T}f = \gamma Tf + (1-\gamma)(f - [f]/4\pi), \quad f\in L^2(\mathbf{R}^3), \tag{4.35}$$

where the operator $\overline{T}$ is defined by

$$\overline{T}f = \overline{Tf} \tag{4.36}$$

and

$$\gamma(\mathbf{k}) = D_+(\mathbf{k}^2)/D_-(\mathbf{k}^2). \tag{4.37}$$

$$(5)\ T\overline{T}^* f + (\overline{Q},f)_0 Q = f - (1-\overline{\gamma})[f]/4\pi, \quad f\in L^2(\mathbf{R}^3). \tag{4.38}$$

(6) If h is a rotation invariant function on $\mathbf{R}^3$, then we have

$$T^*hT = \overline{T}^*h\overline{T} \tag{4.39}$$

as an operator identity and

$$T^*hQ = \overline{T}^*h\overline{Q} \quad \text{a.e.} \tag{4.40}$$

for h with $hQ\in L^2(\mathbf{R}^3)$.

Lemma 4.9 is proved in the Appendix.

Remark: By Lemma 4.8, (4.32) and (4.33) can be extended as operator identities on $M_\alpha(\mathbf{R}^3)$ for $\alpha = \pm\frac{1}{2}, -1$.

B. Exact solution

Let

$$\hat{\phi}_0(f) = (1/\sqrt{2})\{a^*(\tilde{f}/\sqrt{\omega}) + a(f/\sqrt{\omega})\},\ f\in M_{-1/2}(\mathbf{R}^3), \tag{4.41}$$

$$\hat{\pi}_0(f) = (i/\sqrt{2})\{a^*(\omega^{1/2}\tilde{f}) - a(\omega^{1/2}f)\},\ f\in M_{1/2}(\mathbf{R}^3), \tag{4.42}$$

and define

$$b(f) = (1/\sqrt{2})\{\hat{\phi}_0(T^*\omega^{1/2}f) + i\hat{\pi}_0(T^*f/\sqrt{\omega}) + (\omega^{1/2}Q,f)_0 q + i(Q/\sqrt{\omega},f)_0 p\}, \tag{4.43}$$

$$b^*(f) = (1/\sqrt{2})\{\hat{\phi}_0(\overline{T}^*\omega^{1/2}\tilde{f}) - i\hat{\pi}_0(\overline{T}^*\tilde{f}/\sqrt{\omega}) + (\omega^{1/2}\overline{Q},f)_0 q - i(\overline{Q}/\sqrt{\omega},f)_0 p\}\ f\in L^2(\mathbf{R}^3). \tag{4.44}$$

Let D be the linear subspace in $L^2(\mathbf{R})$ spanned by all the Hermite functions:

$$D = \left\{\sum_{j=0}^{n}\alpha_j h_j \,\middle|\, \alpha_j\in\mathbb{C},\ n\in\mathbb{N}, h_j\text{: the jth Hermite function}\right\} \tag{4.45}$$

and put

$$D_F = D\otimes\mathcal{F}_F. \tag{4.46}$$

Then, $b(f)$ and $b^*(f)$ are well-defined on D_F and leave it invariant, satisfying

$$(b(f)\Psi,\Phi) = (\Psi,b^*(\overline{f})\Phi), \tag{4.47}$$

$$[b(f),b^*(g)]\Psi = (\overline{f},g)_0\Psi, \tag{4.48}$$

$$[b(f),b(g)]\Psi = [b^*(f),b^*(g)]\Psi = 0, \tag{4.49}$$

for all Ψ, Φ in D_F and all f,g in $M_0(\mathbf{R}^3)$. (4.48) and (4.49) can be proved by using Lemma 4.9.

Lemma 4.10: Let $b^\#(f)$ be either $b(f)$ or $b^*(f)$. Then,

$$||b^\#(f)\Psi|| \leqslant c(||f||_{-1/2} + ||f||_0)||(H+1)^{1/2}\Psi|| \tag{4.50}$$

for all Ψ in $D(H^{1/2})$, all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$ and some constant $c>0$.

Proof: By using (3.1), Lemma 4.7, and Lemma 4.8, we can show that

$$||b^\#(f)\Psi|| \leqslant c(||f||_{-1/2} + ||f||_0)||(H_0+1)^{1/2}\Psi||, \quad \Psi\in D(H_0^{1/2}),$$

which, combined with (3.12), yields (4.50).

Lemma 4.11: Let f be in $M_{-1/2}(\mathbf{R}^3)\cap M_1(\mathbf{R}^3)$. Then, $b^\#(f)$ maps $D(H^{3/2})$ into $D(H)$ and we have

$$[H,b^\#(f)]\Psi = \pm b^\#(\omega f)\Psi, \quad \Psi\in D(H^{3/2}), \tag{4.51}$$

where $+$ (resp. $-$) goes with $b^*(\cdot)$ (resp. $b(\cdot)$].

Proof: We first prove (4.51) for Ψ in $D_F\cap D(H^{3/2})$ by using commutation relations for $H_{0b}, a^\#(\cdot), p$, and q, and then use a limiting argument to extend the result to all Ψ in $D(H^{3/2})$.

Now, we state the main result in this section:

Theorem 4.1: The dynamical variables $\phi(f,t)$ and $q(t)$ defined by (4.1) and (4.2) have the following explicit form:

$$\phi(f,t) = \frac{1}{\sqrt{2}}\left\{b^*\left(\frac{e^{i\omega t}}{\sqrt{\omega}}\overline{T}\hat{f}\right) + b\left(\frac{e^{-i\omega t}}{\sqrt{\omega}}\tilde{Tf}\right)\right\}, \quad f\in M_{-1}(\mathbf{R}^3)\cap M_{-1/2}(\mathbf{R}^3) \tag{4.52}$$

and

$$q(t) = \frac{1}{\sqrt{2}}\left\{b^*\left(\frac{e^{i\omega t}}{\sqrt{\omega}}\overline{Q}\right) + b\left(\frac{e^{-i\omega t}}{\sqrt{\omega}}Q\right)\right\}. \tag{4.53}$$

Furthermore,

(1) For each $t\in\mathbf{R}$ and all Ψ,Φ in D_F, $(\Psi,\phi(f,t)\Phi)$ is a tempered distribution as a functional of f in $\mathcal{S}(\mathbf{R}^3)$. If f is real, then $\phi(f,t)$ is essentially self-adjoint on D_F.

(2) $q(t)$ is essentially self-adjoint on D_F.

Proof: Let $\Psi\in D(H^{3/2})$ and $\hat{f}\in M_{-1}(\mathbf{R}^3)\cap M_{-1/2}(\mathbf{R}^3)\cap M_{1/2}(\mathbf{R}^3)$. We denote the right-hand side of (4.52) by $\psi(f,t)$. It is easy to see that $\psi(f,t)\Psi$ is strongly differentiable with respect to t. By using Lemma 4.11, we get

$$\frac{d}{dt}\psi(f,t)\Psi = i[H,\psi(f,t)]\Psi,$$

i.e., $\psi(f,t)$ satifies the Heisenberg equation. Furthermore, we have from Lemma 4.9 as applied to (4.43) and (4.44)

$$\psi(f,0) = \phi_0(f).$$

Thus, by the uniqueness of solution to the Heisenberg equation, we get

$$\psi(f,t)\Psi = \phi(f,t)\Psi. \tag{4.54}$$

By a limiting argument employing Lemma 4.10, we can extend (4.54) to all Ψ in $D(H^{1/2})$ and to all $\hat{f}$ in $M_{-1}(\mathbf{R}^3)\cap M_{-1/2}(\mathbf{R}^3)$, so that (4.52) holds. The proof for $q(t)$ is quite similar. By using standard estimates for $a^\#(\cdot)$ and the boson number operator (see, e.g., Ref.4, Sec.X.7), we can show that

$$||b^\#(f)\Psi|| \leqslant c_\Psi||f||_0 \tag{4.55}$$

for all Ψ in D_F and all f in $M_0(\mathbf{R}^3)$, where $c_\Psi>0$ is a constant depending on Ψ. It follows from (4.52), (4.55) and Lemma 4.8 that $(\Psi,\phi(f,t)\Phi)$ with $\Psi,\Phi\in D_F$ is a tempered distribution as a functional of f in $\mathcal{S}(\mathbf{R}^3)$. It is easy to see that if f is real, then

$\phi(f,t)$ is symmetric on D_F [cf. (4.47)]. Furthermore, by using the estimates (4.55) repeatedly (cf. Ref.4, Sec.X. 6–7), we can show that D_F is a set of analytic vectors for $\phi(f,t)$. Thus, it follows from Nelson's analytic vector theorem (see, e.g., Ref. 4, Sec. X. 6) that if f is real, then $\phi(f,t)$ is essentially self-adjoint on D_F. The proof for essential self-adjointnss of $q(t)$ is quite similar.

Finally we give the expressions of $a^{\#}(\cdot)$, p, and q in terms of $b^{\#}(\cdot)$ which are obtained by using Lemma 4.9 and Theorem 4.1. Namely, we have

$$a(f)=\frac{1}{2}\left\{b^*\left(\frac{1}{\sqrt{\omega}}\,\overline{T}\omega^{1/2}\tilde{f}-\omega^{1/2}\overline{T}\frac{\tilde{f}}{\sqrt{\omega}}\right)+b\left(\frac{1}{\sqrt{\omega}}\,T\omega^{1/2}f+\omega^{1/2}T\frac{f}{\sqrt{\omega}}\right)\right\}, \tag{4.56}$$

$$a^*(f)=\frac{1}{2}\left\{b^*\left(\frac{1}{\sqrt{\omega}}\overline{T}\omega^{1/2}f+\omega^{1/2}\overline{T}\frac{f}{\sqrt{\omega}}\right)+b\left(\frac{1}{\sqrt{\omega}}T\omega^{1/2}\tilde{f}-\omega^{1/2}T\frac{\tilde{f}}{\sqrt{\omega}}\right)\right\}, \tag{4.57}$$

$$q=\frac{1}{\sqrt{2}}\left\{b^*\left(\frac{\overline{Q}}{\sqrt{\omega}}\right)+b\left(\frac{Q}{\sqrt{\omega}}\right)\right\}, \tag{4.58}$$

$$p=\frac{i}{\sqrt{2}}\left\{b^*(\omega^{1/2}\overline{Q})-b(\omega^{1/2}Q)\right\}. \tag{4.59}$$

V. ASYMPTOTIC OPERATORS

In this section we consider the asymptotic limits, as $t\to\pm\infty$, of the operators $a_t^{\#}(f)$ given by

$$a_t^{\#}(f)=e^{itH}e^{-itH_0}a^{\#}(f)e^{itH_0}e^{-itH}, \tag{5.1}$$

for $t\in\mathbf{R}$ and $f\in M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$. Since $D(H^{1/2})=D(H_0^{1/2})$ [see Remark (2) in Sec. III], $a_t^{\#}(f)$ is well defined on $D(H^{1/2})$.

Theorem 5.1: Let Ψ be in $D(H^{1/2})$. Then, the strong limits

$$\operatorname*{s-lim}_{t\to\pm\infty}a_t^{\#}(f)\Psi\equiv a_{\text{out}}^{\#}(f)\Psi \tag{5.2}$$

exist and are given explicitly by

$$a_{\text{in}}^{\#}(f)=b^{\#}(f), \tag{5.3}$$

$$a_{\text{out}}(f)=b(f-i\lambda\pi\omega\hat{\rho}Q[f]), \tag{5.4}$$

$$a_{\text{out}}^*(f)=b^*(f+i\lambda\pi\omega\hat{\rho}\overline{Q}[f]). \tag{5.5}$$

Proof: The existence of the strong limits of $a_t^{\#}(f)$ as $t\to\pm\infty$ can be proved in the same way as in Ref.7. We prove (5.3), (5.4), and (5.5). Let $f\in\mathscr{S}(\mathbf{R}^3)$ and $\Psi\in D(H^{1/2})$. By differentiating first and then integrating the right-hand side of (5.1) with respect to t, we get

$$a_t(f)\Psi=a(f)\Psi-i\lambda\int_0^t ds\left(\frac{\hat{\rho}e^{-i\omega s}}{(2\omega)^{1/2}},f\right)_0 q(s)\Psi, \tag{5.6}$$

where the integral is the strong integral. Put

$$\alpha_s=\frac{i\lambda}{2}\left(\frac{\hat{\rho}e^{-i\omega s}}{\sqrt{\omega}},f\right)_0\frac{Q}{\sqrt{\omega}}e^{-i\omega s},$$

$$\beta_s=\frac{i\lambda}{2}\left(\frac{\hat{\rho}e^{-i\omega s}}{\sqrt{\omega}},f\right)_0\frac{\overline{Q}}{\sqrt{\omega}}e^{i\omega s}.$$

Since α_s and β_s are strongly continuous on s in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$, we get from (4.53) and (5.6)

$$a_t(f)\Psi=a(f)\Psi-b(u_t)\Psi-b^*(v_t)\Psi,$$

where

$$u_t=\int_0^t ds\,\alpha_s,\qquad v_t=\int_0^t ds\,\beta_s.$$

It is easy to see that the strong limits

$$\operatorname*{s-lim}_{t\to\pm\infty}u_t\equiv u_\pm,\qquad \operatorname*{s-lim}_{t\to\pm\infty}v_t\equiv v_\pm$$

exist in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$. Thus, we get

$$a_{\substack{\text{out}\\\text{in}}}(f)\Psi=a(f)\Psi-b(u_\pm)\Psi-b^*(v_\pm)\Psi.$$

By direct computations we can show that

$$u_-=\frac{1}{2}\left(\frac{1}{\omega^{1/2}}\,T\omega^{1/2}f+\omega^{1/2}T\frac{f}{\sqrt{\omega}}\right)-f,$$

$$u_+=u_-+i\lambda\pi\omega\hat{\rho}Q[f],$$

$$v_\pm=\frac{1}{2}\left(\frac{1}{\omega^{1/2}}\,\overline{T}\omega^{1/2}\tilde{f}-\omega^{1/2}\,\overline{T}\frac{\tilde{f}}{\sqrt{\omega}}\right).$$

Thus, by using (4.56), we get

$$a_{\text{out}}(f)\Psi=b(f-i\lambda\pi\omega\hat{\rho}Q[f])\Psi,$$

$$a_{\text{in}}(f)\Psi=b(f)\Psi.$$

By a limiting argument, the result can be extended to all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$. The expressions for $a_{\text{in}}^*(f)$ and $a_{\text{out}}^*(f)$ are obtained from the fact that

$$(a_{\substack{\text{out}\\\text{in}}}^*(f)\Psi,\Phi)=(\Psi,a_{\substack{\text{out}\\\text{in}}}(\bar{f})\Phi)$$

for all Ψ,Φ in $D(H^{1/2})$.

VI. SPECTRUM OF THE TOTAL HAMILTONIAN

A. Absence of point spectra for excited states

The proof of the absence of point spectra for excited states will be based upon the following lemma:

Lemma 6.1: If Ψ is an eigenvector of H, then $b(f)\Psi=0$ for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$.

Proof: If Ψ is an eigenvector of H, then we have $a_{\text{in}}(f)\Psi=0$ for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$ (see Ref.7), and the lemma follows from (5.3).

Lemma 6.2 (absence of point spectra for excited states): Let

$$E_0=\inf\sigma(H). \tag{6.1}$$

Then, we have either $\sigma_p(H)=\varnothing$ or $\sigma_p(H)=\{E_0\}$.

Proof: Suppose that there exists a nonzero vector Ω such that $H\Omega=\alpha\Omega$ with some $\alpha\geq E_0$. We need only to show that $\alpha=E_0$. Let $\mathscr{H}'$ be the linear subspace in $\mathscr{H}$ spanned by

$$\{b^*(f_1)\cdots b^*(f_n)\Omega\mid f_j\in M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3),\ n\in\mathbb{N}\}.$$

By Lemma 6.1 and the commutation relation (4.48), $b^{\#}(f)$ leaves $\mathscr{H}'$ invariant. Therefore, it follows from (4.56)–(4.59) that $a^{\#}(f)$, p, and q also leave $\mathscr{H}'$ invariant. On the other hand, it is well known that the operator algebra $\{a^*(f),a(f)\mid f\in M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)\}$ (resp. $\{p,q\}$) is irreducible in $\mathscr{F}$ (resp. in $L^2(\mathbf{R})$) (see, e.g., Ref.8). Therefore, the

operator algebra $\{a^*(f),a(f),p,q \mid f\in M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)\}$ is irreducible in $\mathscr{H}$. Thus, we get $\overline{\mathscr{H}'}=\mathscr{H}$. For $f_j\in[\cap_{\alpha>0}M_\alpha(\mathbf{R}^3)]\cap M_{-1/2}(\mathbf{R}^3)$, we define

$$\Psi(f_1,\ldots,f_n)=b^*(f_1)\ldots b^*(f_n)\Omega.$$

By using Lemma 4.11, we can show that $\Psi(f_1,\ldots,f_n)$ is in $D(H)$ for all $n\in\mathbf{N}$. We compute

$$(\Psi(f_1,\ldots,f_n),H\Psi(f_1,\ldots,f_n))$$
$$=\sum_{j=1}^{n}(b^*(f_1)\ldots b^*(f_n)\Omega,\, b^*(f_1)\cdots b^*(\omega f_j)\ldots b^*(f_n)\Omega)$$
$$+\alpha\|\Psi(f_1,\ldots,f_n)\|^2.$$

It is not difficult to see that the first term in the right-hand side is nonnegative. Therefore, we get

$$(\Psi(f_1,\ldots,f_n),H\Psi(f_1,\ldots,f_n))\geqslant\alpha\|\Psi(f_1,\ldots,f_n)\|^2.$$

Since the linear subspace spanned by $\{\Psi(f_1,\ldots,f_n)\mid n\in\mathbf{N}\}$ is also dense by $\overline{\mathscr{H}'}=\mathscr{H}$ and is invariant under the action of the unitary group $\{e^{itH}\mid t\in\mathbf{R}\}$ (cf. Lemma 4.11), it is a core for H. Furthermore, we can show that

$$(\Psi(g_1,\ldots,g_n),H\Psi(f_1,\ldots,f_m))=0,$$
$$(\Psi(g_1,\ldots,g_n),\Psi(f_1,\ldots,f_m))=0$$

for $m\neq n$. Thus, we conclude that

$$(\Psi,H\Psi)\geqslant\alpha\|\Psi\|^2$$

for all Ψ in $D(H)$, which means that $E_0\geqslant\alpha$. Therefore, we get $\alpha=E_0$.

B. Existence and uniqueness of ground state

Lemma 6.3 (uniqueness of ground state): The ground state, if it exists, is unique up to scalar multiples.

Proof: Suppose that there exist two linear independent nonzero vectors Ω_i, $i=1,2$, such that $H\Omega_i=E_0\Omega_i$. Without loss of generality we assume that $(\Omega_1,\Omega_2)=0$. Let $\mathscr{H}_i$, $i=1,2$, be the linear subspace spanned by

$$\{b^*(f_1)\cdots b^*(f_n)\Omega \mid f_j\in M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3),\ n\in\mathbf{N}\}.$$

Then, by the proof of Lemma 6.2 we have $\overline{\mathscr{H}_i}=\mathscr{H}$, $i=1,2$. However, by Lemma 6.1 we have $\{\Omega_1\}\perp\mathscr{H}_2$, which is a contradiction.

Lemma 6.4: If $b(f)\Psi=0$ $(\Psi\neq0)$ for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$, then $H\Psi=E_0\Psi$ and hence Ψ is the ground state.

Proof: If the assumption of the lemma is satisfied, then we have

$$[b(\omega f)+Hb(f)]\Psi=0$$

for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)\cap M_1(\mathbf{R}^3)$. Therefore, it follows that $\Psi\in D(H^{3/2})$ and $b(f)H\Psi=0$ (cf. Lemma 4.11). On the other hand, we can show in the same way as in the proof of Lemma 6.2 that the vector Φ satisfying $b(f)\Phi=0$ for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$ is unique up to scalar multiples. Therefore, $H\Psi=\alpha\Psi$ with some $\alpha\geqslant E_0$. By Lemma 6.3, α must be equal to E_0.

Since we have Lemma 6.4, we prove the existence of the ground state by constructing a vector Ω such that $b(f)\Omega=0$ for all f in $M_{-1/2}(\mathbf{R}^3)\cap M_0(\mathbf{R}^3)$.

Let

$$W_+f=\frac{1}{2}\left\{\frac{1}{\sqrt{\omega}}T^*\omega^{1/2}+\omega^{1/2}T^*\frac{1}{\sqrt{\omega}}\right\}f, \tag{6.2}$$

$$W_-f=\frac{1}{2}\left\{\frac{1}{\sqrt{\omega}}T^*\omega^{1/2}-\omega^{1/2}T^*\frac{1}{\sqrt{\omega}}\right\}\tilde{f}, \tag{6.3}$$

$$A=\frac{1}{\sqrt{2}}(q+ip), \tag{6.4}$$

$$Q_\pm=\frac{1}{2}\left(\omega^{1/2}\pm\frac{1}{\sqrt{\omega}}\right)Q. \tag{6.5}$$

Since T^* is a bounded operator on $M_\alpha(\mathbf{R}^3)$ for $\alpha=\pm\frac{1}{2},0,-1$ (see Lemma 4.8), $W_\pm$ are bounded operators on $M_\alpha(\mathbf{R}^3)$ for $\alpha=-\frac{1}{2},0$. In terms of $W_\pm$, A, and A^* we can write [see (4.43), (4.44), (4.56) and (4.57)]

$$b(f)=a^*(W_-f)+(W_+f)+(Q_-,f)_0A^*+(Q_+,f)_0A, \tag{6.6}$$

$$b^*(f)=a^*(\bar{W}_+f)+a(\bar{W}_-f)+(\bar{Q}_+,f)_0A^*+(\bar{Q}_-,f)_0A, \tag{6.7}$$

$$a(f)=-b^*(\bar{W}_-^*f)+b(W_+^*f), \tag{6.8}$$

$$a^*(f)=b^*(\bar{W}_+^*f)-b(W_-^*f). \tag{6.9}$$

The commutation relations for $b^\#$, $a^\#$, and $A^\#$ give the following relations for $W_\pm$ and $Q_\pm$:

$$W_+^*W_+-W_-^*W_-+P_+-P_-=I, \tag{6.10}$$

$$\bar{W}_+^*W_--\bar{W}_-^*W_++P_{+-}-P_{-+}=0, \tag{6.11}$$

$$W_+W_+^*-\bar{W}_-\bar{W}_-^*=I, \tag{6.12}$$

$$W_-W_+^*-\bar{W}_+\bar{W}_-^*=0, \tag{6.13}$$

where

$$P_\pm f=(Q_\pm,f)_0Q_\pm, \tag{6.14}$$

$$P_{+-}f=(Q_-,f)_0\bar{Q}_+,\qquad P_{-+}f=(Q_+,f)_0\bar{Q}_-. \tag{6.15}$$

In order to construct the ground state, we must know some properties of the operators $W_\pm$.

Lemma 6.5: W_- is a Hilbert-Schmidt operator on $M_0(\mathbf{R}^3)$.

Proof: It follows from the definition of T that

$$(W_-f)(\mathbf{k})=\int d^3\mathbf{k}'\ W_-(\mathbf{k},\mathbf{k}')f(\mathbf{k}'),$$

where

$$W_-(\mathbf{k},\mathbf{k}')=\frac{\lambda\hat{\rho}(\mathbf{k})\,\overline{Q(\mathbf{k}')}}{2(|\mathbf{k}||\mathbf{k}'|)^{1/2}(|\mathbf{k}|+|\mathbf{k}'|)}.$$

It is easy to see that $W_-(\cdot,\cdot)\in L^2(\mathbf{R}^6)$. Therefore, the lemma follows.

Lemma 6.6: $\dim\operatorname{Ker}W_+=1$.

Proof: Let

$$Lf=\frac{1}{\sqrt{\omega}}T\omega T^*\frac{f}{\sqrt{\omega}}.$$

Then, it follows from Lemma 4.8 that L is a bounded, nonnegative, self-adjoint operator on $M_0(\mathbf{R}^3)$. Put

$$g=(1+L)^{-1}\frac{Q}{\sqrt{\omega}}.$$

By using the identity

$$W_+f=(1/2\sqrt{\omega})\,T^*\omega^{1/2}(1+L)f,$$

we have

$$W_+ g = (1/2\sqrt{\omega})\, T^* Q = 0,$$

so that g is in $\mathrm{Ker} W_+$, i.e., $\mathrm{Ker} W_+ \neq \emptyset$. Let $f \in \mathrm{Ker} W_+$. Then, by using (4.33), we get

$$(1 + L) f = (\omega^{1/2} Q, f)_0 (Q/\sqrt{\omega}), \tag{6.16}$$

i.e.,

$$f = (\omega^{1/2} Q, f)_0\, g,$$

which implies that $\dim \mathrm{Ker} W_+ = 1$.

Lemma 6.7: If $f \in \mathrm{Ker} W_+$, then (i) $W_- f \neq 0$ and (ii) $(Q_+, f)_0 \neq 0$ $(f \neq 0)$.

Proof: Let $f \in \mathrm{Ker} W_+$ $(f \neq 0)$ and suppose that $W_- f = 0$. Then, it follows from (6.10) and (6.11) that

$$(P_+ - P_-) f = f, \quad (P_{+-} - P_{-+}) f = 0,$$

which imply that $f = 0$. This is a contradiction. Thus, $W_- f \neq 0$. By (6.16) and the positivity of L, we have

$$(\omega^{1/2} Q, f)_0 (f, Q/\sqrt{\omega})_0 \geq 1.$$

Therefore, we get

$$(Q_+, f)_0 (f, Q/\sqrt{\omega})_0 = \tfrac{1}{2}\{(\omega^{1/2} Q, f)_0 (f, Q/\sqrt{\omega})_0 + |(f, Q/\sqrt{\omega})_0|^2\} > 1,$$

which implies that $(Q_+, f)_0 \neq 0$.

Henceforth we fix a vector f_+ in $\mathrm{Ker} W_+$.

We define an operator Y on $M_0(\mathbf{R}^3)$ by

$$Yf = W_- f - \frac{(Q_+, f)_0}{(Q_+, f_+)_0} W_- f_+, \qquad f \in M_0(\mathbf{R}^3), \tag{6.17}$$

which is a Hilbert-Schmidt operator by Lemma 6.5.

It follows from (6.12) that $(W_+ W_+^*)^{-1}$ exists and therefore that $W_+ \restriction (\mathrm{Ker} W_+)^\perp$ is a one-one map from $(\mathrm{Ker} W_+)^\perp$ onto $M_0(\mathbf{R}^3)$. Thus,

$$Z \equiv (W_+ \restriction (\mathrm{Ker} W_+)^\perp)^{-1} \tag{6.18}$$

is a bounded operator on $M_0(\mathbf{R}^3)$ with $\mathrm{Ran} Z = (\mathrm{Ker} W_+)^\perp$. Put

$$C = YZ. \tag{6.19}$$

Then, C is a Hilbert-Schmidt operator on $M_0(\mathbf{R}^3)$.

Lemma 6.8: *Let $C(\mathbf{k},\mathbf{k}')$ denote the Hilbert-Schmidt* kernel of the operator C. Then,

$$C(\mathbf{k},\mathbf{k}') = C(\mathbf{k}',\mathbf{k}). \tag{6.20}$$

Proof: We need only to show that $C^* = \bar{C}$, i.e.,

$$Z^* Y^* = \bar{Y}\, \bar{Z}. \tag{6.21}$$

By using (6.11), we can show that

$$W_+^*\, \bar{Y} = Y^* \bar{W}_+,$$

which, combined with $W_+ Z = 1$, yields (6.21).

We now construct the ground state. Let

$$\Omega_0 = h_0 \otimes \Omega_F, \tag{6.22}$$

where h_0 is the zeroth Hermite function and Ω_F is the Fock vacuum in $\mathcal{F}$:

$$h_0(q) = \pi^{-1/4} e^{-q^2/2}, \tag{6.23}$$

$$\Omega_F = \{1, 0, 0, \ldots\}. \tag{6.24}$$

Then, it follows that

$$A\Omega_0 = 0, \qquad a(f)\Omega_0 = 0 \tag{6.25}$$

for all f in $M_0(\mathbf{R}^3)$. Put

$$u = Q_- - \frac{(f_+, Q_-)_0}{(f_+, Q_+)_0} Q_+ \tag{6.26}$$

and define

$$V = -\tfrac{1}{2}\int d^3\mathbf{k} d^3\mathbf{k}' C(\mathbf{k},\mathbf{k}') a^*(\mathbf{k}) a^*(\mathbf{k}') - a^*(\overline{Z^* u}) A^* - \frac{1}{2}\frac{(Q_-, f_+)_0}{(Q_+, f_+)_0} A^{*2}. \tag{6.27}$$

The operator V is well defined on D_F and leaves it invariant. Furthermore, we can show that D_F is a set of analytic vectors for V. Thus we can define

$$\Omega = c \sum_{n=0}^{\infty} \frac{V^n \Omega_0}{n!}, \tag{6.28}$$

where $c > 0$ is the normalization constant. We want to prove that

$$b(f)\Omega = 0 \tag{6.29}$$

for all f in $M_0(\mathbf{R}^3)$. Then, it follows from Lemma 6.4 that Ω is the ground state.

Equation (6.29) is equivalent to the following equations:

$$[a^*(W_- f) + a(W_+ f) + (Q_-, f)_0 A^* + (Q_+, f)_0 A]\Omega = 0 \quad \text{for } f \in (\mathrm{Ker} W_+)^\perp, \tag{6.30}$$

$$[a^*(W_- f_+) + (Q_-, f_+)_0 A^* + (Q_+, f_+)_0 A]\Omega = 0. \tag{6.31}$$

It is easy to see that

$$[V, a(W_+ f)]\Psi = [a^*(Yf) + (u, f)_0 A^*]\Psi$$

for all f in $(\mathrm{Ker} W_+)^\perp$ and all Ψ in D_F. Therefore, we have

$$a(W_+ f) V^n \Omega_0 = -n\{a^*(Yf) + (u, f)_0 A^*\} V^{n-1} \Omega_0$$

for all $n \geq 1$. Thus, we get

$$[a^*(Yf) + a(W_+ f) + (u, f)_0 A^*]\Omega = 0 \tag{6.32}$$

for all f in $(\mathrm{Ker} W_+)^\perp$. Similarly, we can prove

$$[a^*((Q_+, f_+)_0\, \overline{Z^* u}) + (Q_-, f_+)_0 A^* + (Q_+ f_+)_0 A]\Omega =$$

By using (6.11) and $W_+ Z = 1$, we can show that

$$(Q_+, f_+)_0\, \overline{Z^* u} = W_- f_+. \tag{6.33}$$

Therefore, we get (6.31). (6.31) and (6.32) imply (6.30), so that (6.29) holds. Thus, we have proved the following:

Lemma 6.9 (existence of ground state): The ground state exists and is given by (6.28) up to scalar multiples.

C. Spectrum of the total Hamiltonian

Theorem 6.1: We have

$$\sigma(H) = \sigma_c(H) = [E_0, \infty), \qquad \sigma_p(H) = \{E_0\}.$$

The eigenvalue E_0 is simple.

Proof: From Lemma 6.2 and Lemma 6.9 we get $\sigma_p(H) = \{E_0\}$, where E_0 is simple. Since $E_0 = \inf \sigma(H)$ is an eigenvalue of H and the boson is massless, it follows that $\sigma_c(H) = [E_0, \infty)$ (see Ref.3,Proposition 4.1).

Remarks: (1) By using (6.28) and (6.33), we can evaluate the ground state energy:

$$E_0 = \frac{(\Omega_0, H\Omega)}{(\Omega_0,\Omega)}$$
$$= \tfrac{1}{2} + \tfrac{1}{2}\delta K(Q/\sqrt{\omega}, g_+)_0 - \tfrac{1}{2}\lambda(\hat{\rho}/\sqrt{\omega}, W_- g_+)_0,$$

where

$$g_+ = f_+/(Q_+, f_+)_0.$$

(2) The theorem shows that, under the perturbation, all the point spectra but the lowest one of the unperturbed Hamiltonian H_0 completely disappear into the continuous spectrum. However, we can prove that if we introduce an infrared cutoff in the interaction, then the total Hamiltonian has a countable set of eigenvalues embedded in the continuous spectrum. Thus apparently it is essential for the disappearance of the point spectra of the unperturbed Hamiltonian that the boson is massless.

VII. OTHER PROPERTIES OF THE MODEL

We briefly remark on the other properties of the model.

A. Scattering theory

Since we have established the existence of the ground state (Lemma 6.9) and the asymptotic field (Theorem 5.1), we can construct the S matrix in the usual way and show that it is nontrivial and that the n-boson S-matrix element is written as a combinatorial sum of products of the one-boson S-matrix element $S(\mathbf{k}',\mathbf{k})$ given by

$$S(\mathbf{k}',\mathbf{k}) = \delta^3(\mathbf{k}' - \mathbf{k}) + i\pi\delta(|\mathbf{k}'| - |\mathbf{k}|)\frac{\lambda^2\hat{\rho}(\mathbf{k})^2}{|\mathbf{k}|D_+(\mathbf{k}^2)}.$$

B. Point limit

We can consider the point limit $\rho(\mathbf{x}) \to \delta^3(\mathbf{x})$ of the interaction, which corresponds to removing the ultraviolet cutoff in momentum space, in terms of the Wightman functions $\{W_\rho^{(m,n)}\}$ given by

$$W_\rho^{(m,n)}(f_1,t_1;\ldots;f_m,t_m;t_{m+1},\ldots,t_{m+n})$$
$$= (\Omega, \phi(f_1,t_1)\cdots\phi(f_m,t_m)\,q(t_{m+1})\cdots q(t_{m+n})\Omega)$$

with $f_i \in \mathcal{S}(\mathbf{R}^3)$, $t_j \in \mathbf{R}$, $i = 1,\ldots,m$, $j = 1,\ldots,m+n \geqslant 0$. By using (4.48),(4.52), and (4.53), we can explicitly evaluate $W_\rho^{(m,n)}$. Let

$$\rho_\kappa(\mathbf{x}) = \frac{\kappa^2}{4\pi}\frac{e^{-\kappa|\mathbf{x}|}}{|\mathbf{x}|}, \qquad \kappa > 0,$$

which tends to $\delta^3(\mathbf{x})$ in the distribution sense as $\kappa \to \infty$. Then, we can prove that

$$\lim_{\kappa\to\infty} W_{\rho_\kappa}^{(m,n)} \equiv W^{(m,n)}$$

exists and that the limit theory defined in terms of the Wightman functions $\{W^{(m,n)}\}$ is identified with a definite theory, which is easily constructed. The details are omitted.

ACKNOWLEDGMENT

The author would like to thank Professor H. Ezawa of Gakushuin University for the suggestion of the problem and helpful discussions.

APPENDIX: PROOF OF LEMMA 4.9

Throughout the proof we set

$$\eta = \omega^{1/2}\hat{\rho}, \qquad \Lambda = \omega^{1/2}Q,$$

and denote the inner product of $L^2(\mathbf{R}^3)$ by $(\ ,\ )$.

(1) From the definition of T, Lemma 4.1 and Lemma 4.2, we have

$$(Tf,Tg) = (f,g) + \lambda\{(f,\Lambda G\eta g) + (\Lambda G\eta f, g)\} + \lim_{\epsilon\to+0} I_\epsilon^{(1)} \tag{A1}$$

for all f,g in $L^2(\mathbf{R}^3)$, where

$$I_\epsilon^{(1)} = \lambda^2(\Lambda G_\epsilon \eta f, \Lambda G_\epsilon \eta g).$$

Putting

$$K_\epsilon^{(1)}(\mathbf{k}',\mathbf{k}'') = \int d^3\mathbf{k}\,\frac{|Q(\mathbf{k})|^2}{(\mathbf{k}^2 - \mathbf{k}'^2 - i\epsilon)(\mathbf{k}^2 - \mathbf{k}''^2 + i\epsilon)},$$

we can write

$$I_\epsilon^{(1)} = \lambda^2\int d^3\mathbf{k}'\,d^3\mathbf{k}''\,\hat{\rho}(\mathbf{k}')\hat{\rho}(\mathbf{k}'')\,\overline{f(\mathbf{k}')}\,g(\mathbf{k}'')K_\epsilon^{(1)}(\mathbf{k}',\mathbf{k}''),$$

By using (4.27), we have

$$K_\epsilon^{(1)}(\mathbf{k}',\mathbf{k}'') = \frac{1}{2\pi i}\int_0^\infty ds \frac{1}{(s - \mathbf{k}'^2 - i\epsilon)(s - \mathbf{k}''^2 + i\epsilon)} \times\left\{\frac{1}{D_+(s)} - \frac{1}{D_-(s)}\right\}.$$

The right-hand side can be evaluated by using a contour integral for $1/[(z - \mathbf{k}'^2 - i\epsilon)(z - \mathbf{k}''^2 + i\epsilon)D(z)]$ in the cut plane $\mathbb{C}\setminus[0,\infty)$ and we get

$$K_\epsilon^{(1)}(\mathbf{k}',\mathbf{k}'') = \frac{1}{(\mathbf{k}'^2 - \mathbf{k}''^2 + 2i\epsilon)D(\mathbf{k}'^2 + i\epsilon)} + \frac{1}{(\mathbf{k}''^2 - \mathbf{k}'^2 - 2i\epsilon)D(\mathbf{k}''^2 - i\epsilon)}.$$

Therefore, we obtain

$$I_\epsilon^{(1)} = \lambda\left\{\left(\frac{\lambda\eta}{D_\epsilon}f, G_{2\epsilon}\eta g\right) + \left(G_{2\epsilon}\eta f, \frac{\lambda\eta}{D_\epsilon}g\right)\right\},$$

where

$$D_\epsilon(s) = D(s - i\epsilon).$$

Since the operator norm of G_ϵ is bounded uniformly with respect to ϵ and $G_\epsilon \xrightarrow{s} G$, $\lambda\eta f/D_\epsilon \to -\bar{\Lambda}f$ in $L^2(R^3)$ as $\epsilon \to +0$, we get

$$\lim_{\epsilon\to+0} I_\epsilon^{(1)} = -\lambda\{(f,\Lambda G\eta g) + (\Lambda G\eta f, g)\},$$

which, together with (A 1), yields

$$(Tf,Tg) = (f,g),$$

implying $T^*T = I$.

(2) By Lemma 4.1, Lemma 4.2, and (4.31), we have

$$(T^*f,T^*g) = (f,g) - \lambda\{(\eta G\bar{\Lambda}f, g) + (f,\eta G\bar{\Lambda}g)\} + \lim_{\epsilon\to+0} I_\epsilon^{(2)} \tag{A2}$$

for all f,g in $L^2(\mathbf{R}^3)$, where

$$I_\epsilon^{(2)} = \lambda^2(\eta G_\epsilon\bar{\Lambda}f, G_\epsilon\bar{\Lambda}g).$$

Putting

$$K_\epsilon^{(2)}(\mathbf{k}',\mathbf{k}'') = \int d^3\mathbf{k} \frac{\lambda^2 \hat{\rho}(\mathbf{k})^2}{(\mathbf{k}^2 - \mathbf{k}'^2 - i\epsilon)(\mathbf{k}^2 - \mathbf{k}''^2 + i\epsilon)},$$

we can write

$$I_\epsilon^{(2)} = \int d^3\mathbf{k}' d^3\mathbf{k}'' \, Q(\mathbf{k}') \, \overline{Q(\mathbf{k}'')} \, \overline{f(\mathbf{k}')} \, g(\mathbf{k}'') K_\epsilon^{(2)}(\mathbf{k}',\mathbf{k}'').$$

By direct algebraic computations, we have

$$K_\epsilon^{(2)}(\mathbf{k}',\mathbf{k}'') = \frac{D(\mathbf{k}''^2 - i\epsilon) - D(\mathbf{k}'^2 + i\epsilon)}{\mathbf{k}'^2 - \mathbf{k}''^2 + 2i\epsilon} - 1.$$

Therefore, we get

$$I_\epsilon^{(2)} = \lambda \{(\eta G_{2\epsilon} \overline{\Lambda} f, F_\epsilon g\} + (F_\epsilon f, \eta G_{2\epsilon} \overline{\Lambda} g) - (f,Q)(Q,g),$$

where

$$F_\epsilon(\mathbf{k}) = D(\mathbf{k}^2 - i\epsilon)/D_-(\mathbf{k}^2).$$

Thus, we obtain

$$\lim_{\epsilon \to +0} I_\epsilon^{(2)} = \lambda \{(\eta G \overline{\Lambda} f, g) + (f, \eta G \overline{\Lambda} g)\} - (f,Q)(Q,g),$$

which, together with (A2), yields

$$(T^* f, T^* g) - (f,(Q,g)Q) = (f,g).$$

Thus, (4.33) follows.

(3) We can write

$$(Tf,Q) = (f,q) + \lim_{\epsilon \to +0} I_\epsilon^{(3)}, \tag{A3}$$

for all f in $L^2(\mathbf{R}^3)$, where

$$I_\epsilon^{(3)} = \lambda \int d^3\mathbf{k} \, \hat{\rho}(\mathbf{k}) \overline{f(\mathbf{k})} K_\epsilon^{(3)}(\mathbf{k}),$$

$$K_\epsilon^{(3)}(\mathbf{k}) = \int d^3\mathbf{k}' \frac{\lambda^2 \hat{\rho}(\mathbf{k}')^2}{|D_+(\mathbf{k}'^2)|(\mathbf{k}'^2 - \mathbf{k}^2 - i\epsilon)}.$$

By using (4.27), we have

$$K_\epsilon^{(3)}(\mathbf{k}) = \frac{1}{2\pi i} \int_0^\infty ds \frac{1}{(s - \mathbf{k}^2 - i\epsilon)} \left\{ \frac{1}{D_+(s)} - \frac{1}{D_-(s)} \right\}$$
$$= \frac{1}{D(\mathbf{k}^2 + i\epsilon)},$$

where the last equality follows from an evaluation using a contour integral. Therefore, we get

$$\lim_{\epsilon \to 0} I_\epsilon^{(3)} = -(f,Q),$$

which, together with (A3), implies $(Tf,Q) = 0$. Thus, we obtain $T^*Q = 0$.

(4) This is proved by direct algebraic computations.

(5) This follows from direct computations using (4.33) and (4.35).

(6) This follows from the property (4.35).

[1] F.Schwabl and W.Thirring, Ergeb. Exacten Naturwiss. **36**, 219 (1964).
[2] M.Reed and B.Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol.IV.
[3] A.Arai, "Self-adjointness and spectrum of Hamiltonians in nonrelativistic quantum electrodynamics," J. Math. Phys. **22**, 534 (1981).
[4] M.Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol.II.
[5] K.Yosida, *Functional Analysis* (Springer-Verlag, Berlin, 1965).
[6] E.C.Titchmarsh, *Introduction to the Theory of Fourier Intergrals* (Oxford U. P., Amen House, London, 1937).
[7] Y.Kato and N.Mugibayashi, Prog.Theor.Phys. **30**, 103-133 (1963).
[8] F.A.Berezin, *The Method of Second Quantization* (Academic, New York, 1966).

On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. II

Asao Arai
Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan

(Received 23 April 1981; accepted for publication 12 June 1981)

In terms of the resonance pole of the S matrix, which is also a complex pole associated with the 2-point function for the harmonic oscillator, a rigorous mathematical interpretation is given to the formal perturbation calculation for the "level shifts" (the Lamb shifts in QED) of the harmonic oscillator atom and for the "decay probabilities" of the excited states due to the spontaneous emission of bosons.

PACS numbers: 03.65.Db,03.70. + k

I. INTRODUCTION

This paper is a continuation of the author's previous paper,[1] which is henceforth referred as [I]; we shall use the same notations as in [I].

In [I], we studied a quantum system of a harmonic oscillator coupled to a quantized, massless, scalar field whose total Hamiltonian is formally given by

$$H = \frac{1}{2}\left(-\frac{d^2}{dq^2} + q^2\right) + \int d^3\mathbf{k}\,|\mathbf{k}|a^*(\mathbf{k})a(\mathbf{k}) + \lambda q \int d^3\mathbf{k}\,\frac{\hat{\rho}(\mathbf{k})}{(2|\mathbf{k}|)^{1/2}}\{a^*(\mathbf{k}) + a(\mathbf{k})\} + \tfrac{1}{2}\delta K q^2.$$

We proved

$$\sigma(H) = [E_0, \infty), \quad \sigma_p(H) = \{E_0\},$$

which shows that, under the perturbation, all point spectra but the lowest one of the unperturbed Hamiltonian completely disappear into the continuous spectrum. Then, a problem arises: What are the "level shifts" (the Lamb shifts in QED) of the harmonic oscillator atom that the formal perturbation calculation gives? Or, in other words, what does the formal perturbation calculation approximate? It is the purpose of the present paper to give a solution to the problem together with that of the spontaneous emission of bosons.

In Sec. II, we consider analytic properties of the n-point functions for the harmonic oscillator. After showing that the study can be reduced to that of 2-point function, we shall prove (i) that the complex Fourier transform with respect to positive time of the 2-point function, which is originally defined as an analytic function in the upper half-plane, has a meromorphic continuation into the second Riemann sheet and (ii) that the continuation has a unique pole for sufficiently small values of the coupling constant $|\lambda|$, and further (iii) that the real and the imaginary part of the pole give a rigorous mathematical interpretation to the "level shifts" and the "decay probabilities" from the formal perturbation calculations, at least up to the second order in the coupling constant; the decay is of course due to the spontaneous emission of bosons. We also give the time decay law for the 2-point function, which is not exactly exponential for large times. In Sec. III, we consider the analyticity of the S matrix with respect to the energy variables, showing, after a reduction to the analyticity of the one-boson S matrix, (i) that the one-boson S-matrix element is meromorphically continued into the second Riemann sheet and (ii) that the continuation has a unique pole for sufficiently small values of the coupling constant; the pole is located at exactly the same position as that found in Sec. II. We obtain an asymptotic form of the one-boson S-matrix element as the coupling constant tends to zero, which shows that the pole is the resonance pole in boson scattering.

II. ANALYTIC PROPERTIES OF THE N-POINT FUNCTIONS FOR THE HARMONIC OSCILLATOR

The n-point function for the harmonic oscillator is defined by

$$W_n(t_1,\ldots,t_n) = (\Omega, q(t_1)\cdots q(t_n)\Omega), \tag{2.1}$$

where Ω is the ground state of the total Hamiltonian constructed explicitly in [I]. By using the explicit form of $q(t)$ (see [I], Theorem 4.1), we can prove

$$W_{2n+1}(t_1,\ldots,t_{2n+1}) = 0,$$

$$W_{2n}(t_1,\ldots,t_{2n}) = \sum_{\text{comb}} W_2(t_{i_1},t_{j_1})\cdots W_2(t_{i_n},t_{j_n}) \tag{2.2}$$

for all $n \geqslant 0$, where Σ_{comb} indicates the sum over all $(2n)!/2^n n!$ ways of writing $1,\ldots,2n$ as n distinct unordered pairs $(i_1,j_1),\ldots,(i_n,j_n)$. Thus, the problem of the analytic properties of $W_{2n}(t_1,\ldots,t_{2n})$ is reduced to that of the 2-point function $W_2(t_1,t_2)$, which is explicitly given by

$$W_2(t_1,t_2) = \frac{1}{2}\int d^3\mathbf{k}\,\frac{|Q(\mathbf{k})|^2}{|\mathbf{k}|}e^{-i|\mathbf{k}|(t_1-t_2)} \equiv W_2(t_1-t_2). \tag{2.3}$$

Throughout the paper, we shall assume, in addition to the assumptions in [I], that $\hat{\rho}(k)$ has an analytic continuation $\hat{\rho}(z)$ onto the set $\{z \in \mathbb{C} | \mathrm{Re}(z) > 0, \mathrm{Im}(z) < 0\}$ and

$$\hat{\rho}(z) = O(1/|z|^{1+\epsilon})(|z| \to \infty) \quad \text{for some } \epsilon > 0. \tag{2.4}$$

A. Complex pole associated with the 2-point function

We shall first consider the analytic properties of

$$\hat{W}_2(z) = \int_0^\infty dt\, W_2(t)e^{itz}, \tag{2.5}$$

 0022-2488/81/112549-04$01.00

which is originally defined and is analytic in the upper half-plane.

Let $\Psi_n^{(0)}$ be the eigenvector of the unperturbed Hamiltonian H_0 with the eigenvalue

$$E_n^{(0)} = n + \tfrac{1}{2}, \quad n = 0,1,\ldots . \tag{2.6}$$

and $\{P^{(0)}(E)\}$ be the spectral family associated with H_0. For each Borel set B in R^1, we define the projection operator

$$P_n^{(0)}(B) = P^{(0)}(B \setminus \{E_n^{(0)}\}). \tag{2.7}$$

Let

$$A_n(z) = -(\Psi_n^{(0)}, H_I(H_0 - z)^{-1} P_n^{(0)}(R^1) H_I \Psi_n^{(0)}), \tag{2.8}$$

which is analytic in the cut plane $\mathbb{C}\setminus[\frac{1}{2},\infty)$. Then, we have

Lemma 2.1: The function $A_n(z)$ has an analytic continuation from the upper half-plane into a neighborhood of $E_n^{(0)}$ in the complex plane. In particular,

$$A_n \equiv \lim_{z \to E_n^{(0)}} A_n(z), \quad \mathrm{Im}(z) > 0 \tag{2.9}$$

exists for each $n \geqslant 0$ and is explicitly given by

$$A_n = \frac{\lambda^2}{2}\left\{ nP\int d^3\mathbf{k}\,\frac{\hat\rho(\mathbf{k})^2}{1-\mathbf{k}^2} - \frac{1}{2}\int d^3\mathbf{k}\,\frac{\hat\rho(\mathbf{k})^2}{|\mathbf{k}|(1+|\mathbf{k}|)}\right\} - in\lambda^2\pi^2\hat\rho(1)^2. \tag{2.10}$$

Proof: Since $P_n^{(0)}(R^1) = I - P^{(0)}(E_n^{(0)})$ and $P^{(0)}(E_n^{(0)}) H_I \Psi_n^{(0)} = 0$, we have

$$A_n(z) = -(\Psi_n^{(0)}, H_I(H_0 - z)^{-1} H_I \Psi_n^{(0)}). \tag{2.8'}$$

By direct computations, we have

$$(\Psi_n^{(0)}, H_I e^{itH_0} H_I \Psi_n^{(0)}) = \tfrac{1}{4}\lambda^2\{(n+1)e^{i(E_n^{(0)}+1)t} + ne^{it(E_n^{(0)}-1)t}\}(\hat\rho/\omega^{1/2}, (\hat\rho/\omega^{1/2})e^{i\omega t})_{L^2(R^3)},$$

so that

$$A_n(z) = -\tfrac{1}{4}\lambda^2\{(n+1)J^{(+)}(z) + nJ^{(-)}(z)\}, \tag{2.11}$$

where

$$J^{(\pm)}(z) = 4\pi\int_0^\infty dk\,\frac{k\hat\rho(k)^2}{k + E_n^{(0)} \pm 1 - z}. \tag{2.12}$$

From assumption (2.4), we can easily see that $J^{(-)}(z)$ has an analytic continuation from the upper half-plane into a neighborhood of $E_n^{(0)}$ in the complex plane, while $J^{(+)}(z)$ is analytic in the cut plane $\mathbb{C}\setminus[1 + E_n^{(0)}, \infty)$. Thus, the first half of the lemma follows. The limit (2.10) follows from direct computations using (2.11) and (2.12).

Let

$$E_n(\lambda) = E_n^{(0)} + \mathrm{Re}(A_n) + (\Psi_n^{(0)}, \tfrac{1}{2}\delta K q^2 \Psi_n^{(0)}), \quad n = 0,1,\ldots, \tag{2.13}$$

which are the formal perturbation expansions (see, e.g., Ref. 2, Chap. II)—up to the second order in the coupling constant λ—for the "perturbed levels" of the harmonic oscillator under the perturbation $H_I + \frac{1}{2}\delta K q^2$. The following theorem, which is one of our main results, gives a rigorous mathematical interpretation to the formal perturbation expansions.

Theorem 1: Let

$$\Pi_\pm = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0, \quad \mathrm{Im}(z) \gtrless 0\}. \tag{2.14}$$

The function $\hat W_2(z)$ originally defined as an analytic function in the upper half-plane has a meromorphic continuation from Π_+ onto Π_- across the positive real axis. Furthermore, for each ϵ, $0 < \epsilon < 1$, there exists a constant $r(\epsilon) > 0$ such that if $|\lambda| < r(\epsilon)$, then the meromorphic continuation has a unique pole $\zeta(\lambda)$ in the region

$$\Pi_\epsilon = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > \epsilon, \mathrm{Im}(z) < 0\},$$

which is simple, analytic in λ, and satisfies $\zeta(\lambda) \to 1$ as $\lambda \to 0$. Let

$$\zeta(\lambda) = 1 + a_1\lambda + a_2\lambda^2 + \cdots, \quad |\lambda| < r(\epsilon).$$

Then:

(1) $a_1 = 0, \mathrm{Im}(a_2) < 0$;

(2) (Fermi golden rule) the function

$$f_n(E) = (\Psi_n^{(0)}, H_I P_n^{(0)}(E_n^{(0)} - \delta, E) H_I \Psi_n^{(0)})$$

with $0 < \delta < 1$ is C^∞ near $E = E_n^{(0)}$ and

$$-n\lambda^2\,\mathrm{Im}(a_2) = \pi\left.\frac{df_n}{dE}\right|_{E = E_n^{(0)}}; \tag{2.15}$$

(3) $1 + \lambda^2\,\mathrm{Re}(a_2) = E_n(\lambda) - E_{n-1}(\lambda), \quad n = 1,2,\cdots$. (2.16)

Remarks: (1) This theorem shows that the formal perturbation expansions for the "level shifts" of the harmonic oscillator and for the "decay probabilities" of the excited states are characterized by the complex pole associated with the 2-point function for the perturbed harmonic oscillator, at least up to the second order in the coupling constant.

(2) As is easily seen, we have

$$\hat W_2(z) = -i(q\Omega, (\hat H - z)^{-1} q\Omega), \tag{2.17}$$

where

$$\hat H = H - E_0. \tag{2.18}$$

Therefore, the analytic properties of $\hat W_2(z)$ are translated to that of the resolvent function $(q\Omega, (\hat H - z)^{-1} q\Omega)$.

(3) For the mathematical formulation of the Fermi golden rule, see Ref. 3.

To prove Theorem 1, we need

Lemma 2.2: The function $D(z^2)$ (see [I]) originally defined as an analytic function in the cut plane $\mathbb{C}\setminus R^1$ has an analytic continuation from Π_+ onto Π_- across the positive real axis. Furthermore, for each ϵ, $0 < \epsilon < 1$, there exists a constant $r(\epsilon) > 0$ such that if $|\lambda| < r(\epsilon)$, then the analytic continuation has a unique zero $\eta(\lambda)$ in the region Π_ϵ, which is simple, analytic in λ, and satisfies $\eta(\lambda) \to 1$ as $\lambda \to 0$. Let

$$\eta(\lambda) = 1 + b_1\lambda + b_2\lambda^2 + \cdots, \quad |\lambda| < r(\epsilon).$$

Then,

$$b_1 = 0, \tag{2.19}$$

$$b_2 = \tfrac{1}{2}\left\{\delta K\lambda^{-2} + P\int d^3\mathbf{k}\,\hat\rho(\mathbf{k})^2/(1-\mathbf{k}^2)\right\} - i\pi^2\hat\rho(1)^2. \tag{2.20}$$

Proof: It is easy to see that the function

$$D_{II}(z) = D(z^2) - i4\pi^2\lambda^2 z\hat\rho(z)^2, \quad z \in \Pi_-, \tag{2.21}$$

defines an analytic continuation of $D(z^2)$ from Π_+ onto Π_-

across the positive real axis. Let

$$F(z,\lambda)=\begin{cases}D(z^2), & z\in\Pi_+,\\ D_+(z^2), & z\in(0,\infty),\\ D_{II}(z), & z\in\Pi_-.\end{cases}$$

Then, $F(z,\lambda)$ is analytic as a function of two complex variables in the region $\{\mathrm{Re}(z)>0\}\times\mathbb{C}$ and $F(z,0)$ has a simple zero at $z=1$. Therefore, by the Weierstrass preparation theorem (see, e.g., Ref. 3, p. 188), there exists a neighborhood N of (1,0) and a function $G(z,\lambda)$ which is analytic and nonvanishing in N such that in N, $F(z,\lambda)$ can be written in the form

$$F(z,\lambda)=(z-\eta(\lambda))G(z,\lambda),$$

where $\eta(\lambda)$ is analytic in a neighborhood U of $\lambda=0$. Since $D(z^2)$ with nonzero real λ has no zero in $\overline{\Pi}_+$ (see Lemma 4.4 in [I]), we get

$$D_{II}(\eta(\lambda))=0 \tag{2.22}$$

for nonzero real $\lambda\in U$.

We next prove the uniqueness of $\eta(\lambda)$ satisfying (2.22) in the region Π_ϵ. We can write

$$F(z,\lambda)=-z^2+1+\lambda^2(B(z)+\lambda^{-2}\delta K),$$

where $B(z)$ is analytic in $\{\mathrm{Re}(z)>0\}$ and satisfies $|B(z)|\to 0$ as $|z|\to\infty$. Therefore, for each ϵ, $0<\epsilon<1$, we have

$$\sup_{z\in\Pi_\epsilon}|F(z,\lambda)+z^2-1|\leqslant\lambda^2 c_\epsilon,$$

from which it follows easily that $z(\lambda)$ satisfying $F(z(\lambda),\lambda)=0$ must be equal to $\eta(\lambda)$ for $|\lambda|<r(\epsilon)$, where $r(\epsilon)>0$ is a constant depending on ϵ. Thus, the uniqueness follows. Equations (2.19) and (2.20) can be easily proved by using (2.22).

Proof of Theorem 1: From (2.3), (2.5), and the identity

$$D_-(k^2)-D_+(k^2)=i4\pi^2\lambda^2 k\hat{\rho}(k)^2,$$

we have

$$\hat{W}_2(z)=\frac{1}{2\pi}\int_0^\infty dk\,\frac{1}{k-z}\left[\frac{1}{D_-(k^2)}-\frac{1}{D_+(k^2)}\right],\qquad \mathrm{Im}(z)>0,$$

with which we continue $\hat{W}_2(z)$ analytically over the cut plane $\mathbb{C}\setminus[0,\infty)$. It is easy to see that the function

$$\hat{W}_2(z)_{II}=\hat{W}_2(z)+i\{1/D(z^2)-1/D_{II}(z)\},\quad z\in\Pi_-,$$

defines a meromorphic continuation of $\hat{W}_2(z)$ from Π_+ onto Π_- across the positive real axis. The singularities of $\hat{W}_2(z)_{II}$ come only from that of $1/D_{II}(z)$, so that the first half of the theorem follows from Lemma 2.2, and in fact $\zeta(\lambda)=\eta(\lambda)$. By direct computations, we have

$$(\Psi_n^{(0)},\tfrac{1}{2}\delta K q^2\Psi_n^{(0)})=\tfrac{1}{2}\delta K E_n^{(0)},$$

$$\mathrm{Re}(A_n-A_{n-1})=\tfrac{1}{2}\lambda^2 P\int d^3\mathbf{k}\,\frac{\hat{\rho}(\mathbf{k})^2}{1-\mathbf{k}^2},$$

from basic commutation relations and (2.10), respectively. Thus, comparing (2.13) and (2.20), we get (2.16). Finally, we prove (2). The idea for the proof is similar to that in Ref. 4. By Stone's formula as applied to (2.8′) and Lemma 2.1, $f_n(E)$ is written as an integral of the imaginary part of an analytic function in a neighborhood of $E_n^{(0)}$, so that $f_n(E)$ is C^∞ near $E=E_n^{(0)}$ and we get

$$\pi\frac{df_n}{dE}\bigg|_{E=E_n^{(0)}}=-\mathrm{Im}(A_n),$$

which proves (2.15) by (2.10) and (2.20).

B. Time decay property of the 2-point function

Theorem 2: For sufficiently small $|\lambda|$, the 2-point function for the harmonic oscillator is expressed in the form

$$W_2(t)=-[D_{II}'(\zeta(\lambda))]^{-1}e^{-i\zeta(\lambda)t}+r(t),\quad t>0, \tag{2.23}$$

where

$$r(t)=(1/t)r_1(t)+r_2(t),$$

$$\lim_{t\to\infty}|r_1(t)|=c_1(\lambda)\neq 0,\quad |r_2(t)|\leqslant c_2(\lambda), \tag{2.24}$$

with some constants $c_j(\lambda)$ depending on λ.

Proof: From (2.3), we have

$$W_2(t)=\frac{1}{2\pi i}\sum_{j=1}^{3}I_j(t),$$

where

$$I_1(t)=\int_0^\epsilon dk\,\frac{e^{-ikt}}{D_+(k^2)},\quad I_2(t)=\int_\epsilon^\infty dk\,\frac{e^{-ikt}}{D_+(k^2)},$$

$$I_3(t)=-\int_0^\infty dk\,\frac{e^{-ikt}}{D_-(k^2)},\quad 0<\epsilon<1.$$

Since $1/D_+(k^2)$ is uniformly bounded, we have

$$|I_1(t)|\leqslant(\mathrm{const})\epsilon. \tag{2.25}$$

Noting that $D_{II}(z)$ is an analytic continuation of $D_+(k^2)$ onto Π_-, we can evaluate the integral $I_2(t)$ by using the contour integral of $1/D_{II}(z)$ along the curve $\{\epsilon-ik\,|0\leqslant k\leqslant L\}\cup\{\epsilon+Le^{i\theta}|-\frac{1}{2}\pi\leqslant\theta\leqslant 0\}\cup\{k\,|\epsilon\leqslant k\leqslant L\}(L\to\infty)$. Using Lemma 2.2 and $\zeta(\lambda)=\eta(\lambda)$, we get
we get

$$I_2(t)=-2\pi i\{D_{II}'(\zeta(\lambda))\}^{-1}e^{-i\zeta(\lambda)t}-i\int_0^\infty dk\,\frac{e^{-i\epsilon t-kt}}{D_{II}(\epsilon-ik)} \tag{2.26}$$

for $|\lambda|<r(\epsilon)$, where we take ϵ such that $D_{II}(\epsilon-ik)$ has no zero. By changing the integral variable, we have

$$\int_0^\infty dk\,\frac{e^{-i\epsilon t-kt}}{D_{II}(\epsilon-ik)}=\frac{1}{t}e^{-i\epsilon t}\int_0^\infty dk\,\frac{e^{-k}}{D_{II}(\epsilon-ik/t)}. \tag{2.27}$$

The integral $\overline{I_3(t)}$ can be evaluated by using the contour integral of $1/D(z^2)$ along the curve $\{k\,|0\leqslant k\leqslant L\}\cup\{Le^{i\theta}|0\leqslant\theta\leqslant\frac{1}{2}\pi\}\cup\{ik\,|0\leqslant k\leqslant L\}\,(L\to\infty)$, and we get

$$\overline{I_3(t)}=i\frac{1}{t}\int_0^\infty dk\,\frac{e^{-k}}{D(-k^2/t^2)}, \tag{2.28}$$

where we have used the fact that $D(z^2)$ has no zero in $\overline{\Pi}_+$. Equations (2.25)–(2.28) imply the theorem.

Remark: Since $W_2(t)=(q\Omega,e^{-it\hat{H}}q\Omega)$, the theorem gives the time decay law for the state $q\Omega$, which is not exactly exponential for large times. The time decay laws for other states, e.g., $q^n\Omega$, can be obtained through the formula (2.2) or in a similar way.

III. RESONANCE POLE OF THE S MATRIX

We shall give another characterization for the formal perturbation expansions by considering the analytic properties of the S matrix.

It can be proved that the n-boson S-matrix element $S_\lambda^{(n)}(\mathbf{k}_1',...,\mathbf{k}_n',\mathbf{k}_1,...,\mathbf{k}_n)$ in our model is given by

$$S_\lambda^{(n)}(\mathbf{k}_1',...,\mathbf{k}_n',\mathbf{k}_1,...,\mathbf{k}_n) = \sum_\pi S_\lambda^{(1)}(\mathbf{k}_1',\mathbf{k}_{\pi(1)})\cdots S_\lambda^{(1)}(\mathbf{k}_n',\mathbf{k}_{\pi(n)}), \tag{3.1}$$

where $\mathbf{k}_j'$(resp.$\mathbf{k}_j$), $j = 1,...,n$, denote the momenta of outgoing (resp. incoming) bosons and the sum is taken over all permutations of n suffices of $\mathbf{k}$'s. Therefore, the problem of the analyticity of $S_\lambda^{(n)}$ is reduced to that of $S_\lambda^{(1)}$.

It is easy to see that

$$S_\lambda^{(1)}(\mathbf{k}',\mathbf{k}) = \delta^3(\mathbf{k}'-\mathbf{k}) - 2\pi i\delta(|\mathbf{k}'|-|\mathbf{k}|)\mathscr{T}_\lambda(|\mathbf{k}|) \tag{3.2}$$

where $\mathscr{T}_\lambda(|\mathbf{k}|)$ is the transition matrix element for one boson scattering:

$$\mathscr{T}_\lambda(k) = -\lambda^2\hat{\rho}(k)^2/2kD_+(k^2), \quad k>0. \tag{3.3}$$

Theorem 3: (1) The function $\mathscr{T}_\lambda(k)$ originally defined in $(0,\infty)$ has a meromorphic continuation onto Π_-, which, for $|\lambda|<r(\epsilon)$, has the unique pole $\zeta(\lambda)$ in the region Π_ϵ.

(2) Let $\zeta(\lambda) = E(\lambda) - i\Gamma(\lambda)$, so that $\Gamma(\lambda)>0$, and let $\{I_\lambda\,|\,\lambda>0\}$ be an arbitrary family of open intervals containing $E(\lambda)$ and such that the Lebesgue measure of I_λ is $O(\Gamma(\lambda)^{1/2})$ as $\lambda\to+0$. Then, for $k\in I_\lambda$,

$$|\mathscr{T}_\lambda(k)|^2 = \frac{\lambda^2\hat{\rho}(k)^2}{8\pi^2k^3}\left\{\frac{\gamma(\lambda)\Gamma(\lambda)}{[((k-E(\lambda))^2+\Gamma(\lambda)^2]}\times(1+O(1))+O(1)\right\} \tag{3.4}$$

as $\lambda\to+0$, where

$$\gamma(\lambda) = -\mathrm{Re}[D_{\mathrm{II}}'(\zeta(\lambda))]^{-1}.$$

Remark: The theorem shows that $\zeta(\lambda)$ is the resonance pole of the S matrix.

Proof: (1) By the proof of Lemma 2.2, the function

$$\mathscr{T}_\lambda(z)_{\mathrm{II}} = -\lambda^2\hat{\rho}(z)^2/2zD_{\mathrm{II}}(z)$$

defines a meromorphic continuation of $\mathscr{T}_\lambda(k)$ onto Π_- and hence the assertion follows.

(2) The method of the proof is quite similar to that of Ref. 5, Theorem 4. With the notations in the proof of Lemma 2.2, we can write

$$[F(z,\lambda)]^{-1} = (z-\eta(\lambda))^{-1}\alpha(\lambda) + O(1)$$

as $\lambda\to0$, uniformly on N, where $\alpha(\lambda)$ is the residue of $1/D_{\mathrm{II}}(z)$ at $z=\eta(\lambda)$:

$$\alpha(\lambda) = [D_{\mathrm{II}}'(\eta(\lambda))]^{-1}.$$

Let

$$\alpha(\lambda) = -\gamma(\lambda) + i\beta(\lambda).$$

Then, for $k\in I_\lambda$, we have, using $\eta(\lambda)=\zeta(\lambda)$,

$$\frac{2\pi^2\lambda^2k\hat{\rho}(k)^2}{|D_+(k^2)|^2} = \mathrm{Im}[F(z,\lambda)]^{-1} = \frac{\gamma(\lambda)\Gamma(\lambda)}{(k-E(\lambda))^2+\Gamma(\lambda)^2} + \frac{(k-E(\lambda))\beta(\lambda)}{(k-E(\lambda))^2+\Gamma(\lambda)^2} + O(1) \tag{3.5}$$

as $\lambda\to0$. By considering the cases $k = E(\lambda)\pm\Gamma(\lambda)^{1/2}$, we can prove that $\beta(\lambda) = O(\Gamma(\lambda)^{1/2})$ as $\lambda\to0$, so that the second term in the right-hand side of (3.5) is $O(1)$ uniformly on $k\in I_\lambda$ as $\lambda\to+0$. Thus, (3.4) follows.

ACKNOWLEDGMENT

The author would like to thank Professor H. Ezawa of Gakushuin University for helpful suggestions.

[1] A. Arai, "On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. I," J. Math. Phys. **22**, 2539 (1981).

[2] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer, Berlin, 1976).

[3] S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton U.P., Princeton, N. J., 1948).

[4] B. Simon, Ann. Math. **97**, 247–274 (1973); M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV (Academic, New York, 1978), Sec. XII.6.

[5] J. S. Howland, Pacific J. Math. **29**, 565–582 (1962).

Generalized monotone convergence and Radon–Nikodym theorems

S. Gudder and J. Zerbe
Mathematics & Computer Science Department, University of Denver, Denver, Colorado 80208

A measure and integration theory is presented in the quantum logic framework. A generalization of the monotone convergence theorem is proved. Counterexamples are used to show that the dominated convergence theorem, Fatou's lemma, Egoroff's theorem, and the additivity of the integral do not hold in this framework. Finally, a generalization of the Radon–Nikodym theorem is proved.

PACS numbers: 03.65.Bz, 02.50. + s

I. INTRODUCTION

In this paper we consider a generalized measure and integration theory. We use the so called "quantum logic" framework which generalizes the traditional Hilbert space approach to quantum mechanics.

There are other reasons, besides applications to quantum mechanics, for developing a generalized measure and integration theory. As pointed out in Ref. 1, if there is an "elementary length" or an accuracy limitation for a measurement, then a generalized measure space (Ω, Σ, μ) is necessary for a proper description. In such a space, Σ is only closed under countable disjoint unions instead of the arbitrary countable unions for a σ algebra of a measure space. There are also possible applications of such a theory to computer science where roundoff error is important and to pattern recognition where coarse measurements are necessary.

Even in classical probability theory, a generalization may be useful.[2] Suppose a statistician is interested in two events A and B, and collects data to determine their probabilities $P(A)$, $P(B)$. The σ algebra construct then demands that an assignment of the probability be made to $A\cap B$. However, in general $P(A\cap B)$ is not uniquely determined by $P(A)$ and $P(B)$, and our statistician may have no interest in the event $A\cap B$ anyway. Also, there may not exist an extension of P from the events of interest to their intersections which satisfies the usual properties of a probability. Such an extension is not necessary in the generalized theory.

In Sec. II we give the notation and basic definitions that are needed in the sequel. We also present three important examples to which the theory is applicable. In Sec. III we prove a generalization of the monotone convergence theorem. Section IV uses counterexamples to show that generalizations of the dominated convergence theorem, Fatou's lemma, and Egoroff's theorem do not hold. Moreover, we show that the generalized integral need not be additive. In Sec. V, we prove a generalization of the Radon–Nikodym theorem.

II. DEFINITIONS AND NOTATION

Let $(P, \leqslant)$ be a partially ordered set (poset) with least and greatest elements 0 and 1, respectively. Denote the greatest lower and least upper bound of $a, b\in P$ (if they exist) by $a\wedge b$ and $a\vee b$, respectively. A map $':P\to P$ is called an *orthocomplementation* if

(1) $a'' = a$ for all $a\in P$;

(2) $a\leqslant b$ implies $b'\leqslant a'$;

(3) $a\vee a' = 1$ for all $a\in P$.

If $(P, \leqslant)$ possesses an orthocomplementation, we say that P is *orthocomplemented*. We say that $a, b\in P$ are *orthogonal* and write $a\perp b$ if $a\leqslant b'$. An orthocomplemented poset P is *σ orthocomplete* if for any sequence of mutually orthogonal elements $a_i\in P$, we have $\vee a_i\in P$. An orthocomplemented poset P is *orthomodular* if $a\leqslant b$ implies that $b = a\vee(b\wedge a')$. We call a σ orthocomplete orthomodular poset a *logic*.[3–5] In the sequel, $P = (P, \leqslant, ')$ will always denote a logic.

A *measure* on P is a map μ from P to the nonnegative extended real numbers $[0, \infty]$ such that $\mu(0) = 0$ and $\mu(\vee a_i) = \Sigma\mu(a_i)$ for any sequence of mutually orthogonal elements $a_i\in P$. We call the pair (P, μ) a *measure logic*. We say that μ is *finite* if $\mu(1) < \infty$.

An *observable* is a homomorphism $x: B(\mathbf{R})\to P$ where $B(\mathbf{R})$ is the Borel σ algebra on $\mathbf{R}$. That is,

(1) $x(\mathbf{R}) = 1$;

(2) $x(E)\perp x(F)$ if $E\cap F = \phi$;

(3) $x(\cup E_i) = \vee x(E_i)$ if $E_i\cap E_j = \phi$, $i\neq j = 1, 2, \cdots$.

An observable preserves all the Boolean algebra operations on $B(\mathbf{R})$ and the range of an observable is a Boolean σ algebra. If x is an observable and μ is a measure on P, define the measure μ_x on $B(\mathbf{R})$ by $\mu_x(E) = \mu[x(E)]$, $E\in B(\mathbf{R})$. If $\int_{\mathbf{R}}\lambda\mu_x(d\lambda)$ exists (i.e., either $\int_{[0,\infty]}\lambda\mu_x(d\lambda) < \infty$ or $\int_{(-\infty,0]}\lambda\mu_x(d\lambda) > -\infty$), we call this the *integral* of x and denote it by $\mu(x)$ or $\int x\, d\mu$.

We say that $a, b\in P$ are *compatible* and write $a\leftrightarrow b$ if there exist mutually orthogonal elements $a_1, b_1, c\in P$ such that $a = a_1\vee c$, $b = b_1\vee c$. Two observables x, y on P are compatible (written $x\leftrightarrow y$) if $x(E)\leftrightarrow y(F)$ for all $E, F\in B(\mathbf{R})$.

If x is an observable and $f:\mathbf{R}\to\mathbf{R}$ a real Borel function, we define the observable $f(x)$ by $f(x)(E) = x[f^{-1}(E)]$ for all $E\in B(\mathbf{R})$. It is easy to check that $u(x)\leftrightarrow v(x)$ for any real Borel functions u, v and that $\mu[f(x)] = \int f(\lambda)\mu_x(d\lambda)$. If $K\in\mathbf{R}$, then K can be thought of as an observable. In fact, we define K to be the unique observable satisfying $K(\{K\}) = 1$. Clearly $K = u(x)$, where $u(\lambda)\equiv K$, for any observable x. If u, v are real Borel functions and $g:\mathbf{R}^2\to\mathbf{R}$ is a Borel function, we define

$$g[u(x), v(x)](E) = x\{\lambda\in\mathbf{R}: (u(\lambda), v(\lambda))\in g^{-1}(E)\}.$$

If follows that

$$\mu[g(u(x), v(x))] = \int g(u(\lambda), v(\lambda))\mu_x(d\lambda).$$

 0022-2488/81/112553-09$01.00

In particular,

$$\mu[u(x)+v(x)] = \int [u(\lambda)+v(\lambda)]\mu_x(d\lambda)$$
$$= \int u(\lambda)\mu_x(d\lambda) + \int v(\lambda)\mu_x(d\lambda)$$
$$= \mu[u(x)] + \mu[v(x)].$$

In the sequel we shall need the following two theorems due to V. Varadarajan.

Theorem 2.1: (Refs. 3 and 5) Let (P,μ) be a measure logic. (a) If $a_1 \leqslant a_2 \leqslant \cdots \in P$, then $\wedge a_i$ exists, $\vee a_i = a_1 \vee (a_2 \wedge a_1') \vee (a_3 \wedge a_2') \vee \cdots$, and $\mu(\vee a_i) = \lim \mu(a_i)$. (b) If $a \leftrightarrow a_i$, $\vee a_i$, and $\vee (a_i \wedge a)$ exist, then $a \leftrightarrow \vee a_i$ and $a \wedge (\vee a_i) = \vee (a_i \wedge a)$.

Theorem 2.2: If $x \leftrightarrow y$, then there exists an observable z and two real Borel functions u, v such that $x = u(z)$ and $y = v(z)$.

There are three important examples of logics.

Example 1: Let (Ω, Σ) be a measurable space. Denoting set-theoretic complementation by ^{c}, it is easy to check that $(\Sigma, \subseteq, ^c)$ is a logic. In this case, any pair $A, B \in \Sigma$ is compatible. Also, the notion of measure reduces to the usual concept so a measure space (Ω, Σ, μ) is a measure logic. If $f:\Omega \to \mathbf{R}$ is measurable, then $f^{-1}: B(\mathbf{R}) \to \Sigma$ is an observable. Conversely, any observable on Σ has the form $x(E) = f^{-1}(E)$, $E \in B(\mathbf{R})$, where $f:\Omega \to \mathbf{R}$ is measurable.[5] If x is an observable, and μ a measure on Σ, then the integral reduces to the usual definition

$$\mu(x) = \int f\, d\mu, \text{ where } f^{-1}(E) = x(E),\ E \in B(\mathbf{R}).$$

Example 2: Let C be a collection of subsets of a set Ω satisfying

(1) $\phi \in C$;
(2) $A \in C$ implies $A^c \in C$;
(3) if $A_i \in C$, $A_i \cap A_j = \phi$, $i \neq j = 1,2,\cdots$, then $\cup A_i \in C$.

We call C a *σ class* in Ω and we call (Ω, C) a *generalized measurable space*.[1,3,6] Again, $(C, \subseteq, ^c)$ is a logic.[1,3,6] We say that a function $f:\Omega \to \mathbf{R}$ is *measurable* if $f^{-1}(E) \in C$ for every $E \in B(\mathbf{R})$. As before, there is a one-to-one correspondence between measurable functions and observables on C. Let μ be a measure on C, $f:\Omega \to \mathbf{R}$ a measurable function, and $\Sigma(f) \subseteq C$, the σ algebra generated by f. Then $(\Omega, \Sigma(f), \mu)$ is a measure space, so $\int f\, d\mu$ may be defined in the usual way on this space. If x is the corresponding observable, we have $\mu(x) = \int f\, d\mu$. In the sequel we shall frequently not distinguish between a measurable function f and its corresponding observable f^{-1}. For example, when we write $f \leftrightarrow g$, we mean $f^{-1} \leftrightarrow g^{-1}$. In the present example, $A, B \in C$ are compatible if and only if $A \cap B \in C$ (Ref. 1, 3, 6)

Example 3: Let H be a Hilbert space, let $P(H)$ be the set of closed subspaces of H, and let $M^\perp$ denote the orthogonal complement of $M \in P(H)$. Then $(P(H), \subseteq, ^\perp)$ is a logic which we call a *Hilbertian logic*. It follows from Gleason's theorem[5,7] that any finite measure μ on $P(H)$ has the form $\mu(M) = \mathrm{tr}(WP_M)$, where P_M is the orthogonal projection onto M and W is a unique positive trace class operator. By the spectral theorem,[8] observables correspond to self-adjoint operators on H. If μ is a finite measure with corresponding trace class operator W and x is an observable with corresponding self-adjoint operator T, then $\mu(x) = \mathrm{tr}(WT)$. Two closed subspaces are compatible if and only if their corresponding projections commute and two observables are compatible if and only if their corresponding self-adjoint operators commute.[3,5]

It can be shown that Examples 2 and 3 are generalizations of Example 1 but that neither is a generalization of the other.[1,3]

In this work, most of our examples and counterexamples will be from generalized measure spaces. Some examples and counterexamples from Hilbertian logics may be found in Ref. 9.

III. MONOTONE CONVERGENCE THEOREM

If f, g are real-valued functions on a set Ω, then it is straightforward to show that $f(\omega) \leqslant g(\omega)$ for all $\omega \in \Omega$ if and only if $f^{-1}(\alpha, \infty) \subseteq g^{-1}(\alpha, \infty)$ for all $\alpha \in \mathbf{R}$. Motivated by this, if x, y are observables on P, we write $x \leqslant y$ whenever $x(\alpha, \infty) \leqslant y(\alpha, \infty)$ for all $\alpha \in \mathbf{R}$. It is not hard to show that $\leqslant$ is a partial order relation. In general, $x \leqslant y$ need not imply that $x \leftrightarrow y$. For example, let (Ω, C) be the generalized measurable space with $\Omega = \{1, 2, 3, 4\}$ and C the σ class of subsets of Ω with an even number of elements. Define $f(1) = f(4) = 0$, $f(2) = f(3) = 1$, and $g(1) = g(2) = 1$, $g(3) = g(4) = 2$. Then $f \leqslant g$ but $f \not\leftrightarrow g$.

If $x_1 \leqslant x_2 \leqslant \cdots$ are observables, then $\vee x_i(\alpha, \infty)$ exists for every $\alpha \in \mathbf{R}$. If there exists an observable x such that $x(\alpha, \infty) = \vee x_i(\alpha, \infty)$ for every $\alpha \in \mathbf{R}$, we write $x = \lim x_i$. For measurable functions on a generalized measurable space, this definition reduces to the usual pointwise limit when the limit exists. For an example in which the limit does not exist, let f_i be the measurable function on $\mathbf{R}$ defined by $f_i(\lambda) = i$ for all $\lambda \in \mathbf{R}$, $i = 1,2,\cdots$. Then $\vee f_i^{-1}(\alpha, \infty) = \mathbf{R}$, but there is no observable x such that $x(\alpha, \infty) = \mathbf{R}$ for all $\alpha \in \mathbf{R}$, so $\lim f_i^{-1}$ does not exist.

Theorem 3.1. If $x_1 \leqslant x_2 \cdots \leqslant y$ are observables on P, then there exists a unique observable x such that $x(\alpha, \infty) = \vee x_n(\alpha, \infty)$ for all $\alpha \in \mathbf{R}$.

Proof: The latter part of this proof is patterned after (Ref. 5; Theorem 1.4, p. 15). Let $a_\alpha = \vee x_n(\alpha, \infty)$, $\alpha \in \mathbf{R}$. If $\alpha \leqslant \beta$, then $x_n(\alpha, \infty) \geqslant x_n(\beta, \infty)$ so that

$$a_\alpha = \vee x_n(\alpha, \infty) \geqslant \vee x_n(\beta, \infty) = a_\beta. \tag{1}$$

Denoting $a_{-\infty} = 1$ and $a_\infty = 0$, (1) still holds for $\alpha, \beta = \pm\infty$. Let B be the set of all finite orthogonal suprema of elements of the form $a_\alpha \wedge a_\beta'$, $-\infty \leqslant \alpha < \beta \leqslant \infty$. It is straightforward to show that B is a Boolean subalgebra of P. By Zorn's lemma, there is a maximal Boolean subalgebra B_0 of P containing B. Let a_i be an orthogonal sequence in B_0 and let $a = \vee a_i$. If $b \in B_0$, then $a_i \leftrightarrow b$, $i = 1,2,\cdots$, and by Theorem 2.1, $b \leftrightarrow a$. Let B_1 be the Boolean subalgebra $\{0, 1, a, a'\}$. Since $B_0 \leftrightarrow B_1$, there exists a Boolean subalgebra B_2 such that $B, B_1 \subseteq B_2$ (Refs. 3 and 5). Since B_0 is maximal, $B_0 = B_2$ and hence $a \in B_0$. Therefore, B_0 is a Boolean sub σ algebra containing B.

By Loomis's theorem[5,10] there exists a measurable space (Ω, Σ) and a surjective σ homomorphism $h:\Sigma \to B_0$. Let $r_1, r_2, \cdots$ be an enumeration of the rationals in $\mathbf{R}$ and let D_i

$= (r_i, \infty)$, $a_i = a_{r_i}$. As in (1), if $r_i \leqslant r_j$, then $a_i \geqslant a_j$. We shall now construct sets $A_1, A_2, \cdots \in \Sigma$ such that (a) $h(A_i) = a_i$, $i = 1,2,\cdots$; (b) $A_i \supseteq A_j$ whenever $r_i \leqslant r_j$. Let A_1 be any set in Σ such that $h(A_1) = a_1$. Suppose, $A_1, A_2, \cdots A_n \in \Sigma$ have been constructed satisfying (a) and (b). We now construct A_{n+1} as follows. Let $\{i_1, i_2, \cdots i_n\}$ be the permutation of $\{1, 2, \cdots, n\}$ such that $r_{i_1} < r_{i_2} < \cdots < r_{i_n}$. Then there exists a unique k such that $r_{i_k} < r_{n+1} < r_{i_{k+1}}$ (we define $r_{i_0} = -\infty$ and $r_{i_{n+1}} = +\infty$, $A_{i_0} = \Omega$, $A_{i_{n+1}} = \phi$). Now $a_{i_k} \geqslant a_{n+1} \geqslant a_{i_{k+1}}$. Since h is surjective there is an $S \in \Sigma$ such that $h(S) = a_{n+1}$. Let $A_{n+1} = (S \cap A_{i_k}) \cup A_{i_{k+1}}$ Then $A_{i_k} \supseteq A_{n+1} \supseteq A_{i_{k+1}}$ and

$$\begin{aligned} h(A_{n+1}) &= [h(S) \wedge h(A_{i_k})] \vee h(A_{i_{k+1}}) \\ &= (a_{n+1} \wedge a_{i_k}) \vee a_{i_{k+1}} = a_{n+1}. \end{aligned}$$

Then $A_1, A_2, \cdots A_{n+1} \in \Sigma$ satisfies (a) and (b). The construction now follows by induction. As

$$h(\cap A_j) = \wedge a_j \leqslant \wedge y(D_j) = y(\cap D_j) = y(\phi) = 0,$$

we may, by replacing A_k by $A_k - \cap A_j$ if necessary, assume that $\cap A_j = \phi$. Moreover,

$$\begin{aligned} h(\cup A_j) &= \vee a_j = \vee_j \vee_n x_n(D_j) \\ &= \vee_n x_n(\vee_j D_j) \\ &= \vee x_n(\mathbf{R}) = 1. \end{aligned}$$

Hence, if $N = \Omega - \cup A_j$, then $h(N) = 0$. We now define a function $f: \Omega \to (\mathbf{R})$ as follows

$$f(\omega) = \begin{cases} 0 & \text{if} \quad \omega \in N \\ \sup\{r_j : \omega \in A_j\} & \text{if} \quad \omega \in \cup A_j. \end{cases}$$

Then f is finite everywhere. Also, for all k

$$f^{-1}(D_k) \cap N^c = \cup\{A_j : r_j > r_k\},$$

so f is Σ measurable. Further,

$$\begin{aligned} h[f^{-1}(D_k)] &= h[\cup\{A_j : r_j > r_k\}] \\ &= \vee\{a_j : r_j > r_k\} = \vee_j \vee_n \{x_n(D_j) : r_j > r_k\} \\ &= \vee_n x_n(D_k) = a_k. \end{aligned}$$

Now define the observable $x: B(\mathbf{R}) \to P$ by $x(E) = h[f^{-1}(E)]$, $E \in B(\mathbf{R})$. If $\alpha \in R$, approximating α by rationals from above and using a similar limiting procedure gives $x(\alpha, \infty) = a_\alpha$. For uniqueness, suppose z is an observable satisfying $z(\alpha, \infty) = x(\alpha, \infty)$ for every $\alpha \in (\mathbf{R})$. Since $\{(\alpha, \infty) : \alpha \in (\mathbf{R})\}$ generates $B(\mathbf{R})$ it follows that $z = x$. □

Corollary 3.2: Let $f_1 \leqslant f_2 \leqslant \cdots$ be measurable functions on a generalized measurable space (Ω, Σ), and let $f(\omega) = \lim f_i(\omega)$ for all $\omega \in \Omega$. If $f(\omega) < \infty$ for all $\omega \in \Omega$, then f is measurable.

In Sec. IV, we shall give an example in which the pointwise limit of a sequence of measurable functions is not measurable.

A logic P satisfies *condition C* if for any three mutually compatible elements $a, b, c \in P$ we have $a \leftrightarrow b \vee c$. We now give an example of a generalized measurable space that does not satisfy condition C (Ref. 11). Let $\Omega = \{1, 2, \cdots, 8\}$ and let C be the σ class of subsets of Ω with an even number of elements. Let $A = \{1, 2, 3, 4\}$, $B = \{1, 2, 5, 6\}$, and $D = \{1, 3, 6, 8\}$. Then A, B, and D are mutually compatible but $(A \cup B) \cap D = \{1, 3, 6\} \notin C$. Thus, $A \vee B = A \cup B \leftrightarrow D$. There are, however, many logics that do satisfy condition C. For example, Hilbertian logics satisfy condition C. In fact, it follows from theorem 2.1 that if P is a lattice, then P satisfies condition C.

Theorem 3.3: Suppose that P satisfies condition C, that $x_1 \leqslant x_2 \leqslant \cdots$, and $x = \lim x_n$. If $y \leftrightarrow x_n$, $n = 1, 2, \cdots$, then $y \leftrightarrow x$.

Proof: Since $y \leftrightarrow x_n$, we have $y(E) \leftrightarrow x_n(\alpha, \infty)$ for all $E \in B(\mathbf{R})$, $\alpha \in (\mathbf{R})$, $n = 1, 2, \cdots$. Now $\vee x_n(\alpha, \infty) = x(\alpha, \infty)$ and $\vee[x_n(\alpha, \infty) \wedge y(E)]$ exists for all $\alpha \in (\mathbf{R})$. Applying theorem 2.1, we have $y(E) \leftrightarrow x(\alpha, \infty)$ for every $\alpha \in (\mathbf{R})$. Hence, $A = \{y(E), x(\alpha, \infty) : \alpha \in (\mathbf{R})\}$ is a set of mutually compatible elements. It follows[3,5] that there exists a Boolean sub σ algebra B of P with $A \subseteq B$. Then B must contain the range of x since the range is generated by $\{x(\alpha, \infty) : \alpha \in (\mathbf{R})\}$. Hence, $y(E) \leftrightarrow x(F)$ for all $E, F \in B(\mathbf{R})$ and $x \leftrightarrow y$. □

As we have seen, a generalized measurable space need not satisfy condition C. However, theorem 3.3 still holds.

Theorem 3.4: Let $f_1 \leqslant f_2 \leqslant \cdots \leqslant f$ be measurable functions on a generalized measurable space (Ω, C) and suppose $\lim f_n = f$. If g is measurable and $g \leftrightarrow f_n$ for all n, then $g \leftrightarrow f$.

Proof: The proof is similar to that of Theorem 3.3. Instead of using condition C we use (Ref. 1; Theorem 2.3, p. 127). □

In Sec. IV, we give an example which shows that the monotonicity in Theorem 3.4 is necessary.

Let (P, μ) be a measure logic and let x be an observable on P. Define the nonincreasing function $F(x, \cdot)$ by $F(x, \alpha) = \mu[x(\alpha, \infty)]$, $\alpha \in \mathbf{R}$. For two observables x and y we define $x \leqslant y[\mu]$ if $F(x, \alpha) \leqslant F(y, \alpha)$ for all $\alpha \in \mathbf{R}$. This relation is reflexive and transitive, but need not be anti-symmetric. Notice that $x \leqslant y$ implies that $x \leqslant y[\mu]$. However, the converse need not hold as the following example shows.

Let (Ω, C) be the generalized measurable space with $\Omega = \{1, 2, 3, 4\}$ and C the σ class of subsets of Ω with an even number of elements. Define the measure μ by $\mu\{1, 4\} = 1/3$, $\mu\{2, 3\} = 2/3$ and $\mu(A) = 1/2$, where A is any other doubleton set. Define the measurable functions f and g by $f(1) = f(2) = 1, f(3) = f(4) = 0, g(1) = g(4) = 0$, $g(2) = g(3) = 1$. Then $f \leqslant g[\mu]$ but $f \not\leqslant g$.

Lemma 3.5: If $x \leqslant y[\mu]$ and $\mu(x)$, $\mu(y)$ exist, then $\mu(x) \leqslant \mu(y)$.

Proof: First assume that $0 \leqslant x \leqslant y[\mu]$. Then $\mu(x) = \int_{[0, \infty)} \lambda \mu_x(d\lambda)$ and $\mu(y) = \int_{[0, \infty)} \lambda \mu_x(d\lambda)$. Choose an increasing sequence of nonnegative simple functions on $\mathbf{R}$ of the form $\phi_n = \Sigma_i c_i^{(n)} \chi_{(a_i, \infty)}$ such that $\phi_n(\lambda) \uparrow \lambda$, $\lambda \geqslant 0$. Then by the classical monotone convergence theorem we have

$$\begin{aligned} \mu(x) &= \int_{[0, \infty)} \lambda \mu_x(d\lambda) = \lim_{n \to \infty} \int_{[0, \infty)} \phi_n(\lambda) \mu_x(d\lambda) \\ &= \lim_{n \to \infty} \sum_i c_i^{(n)} \mu_x(a_i^{(n)}, \infty) \\ &\leqslant \lim_{n \to \infty} \sum_i c_i^{(n)} \mu_y(a_i^{(n)}, \infty) = \int_{[0, \infty)} \mu_y(d\lambda) = \mu(y). \end{aligned}$$

For general $x \leqslant y[\mu]$, let $f^+(\lambda) = \lambda$ for $\lambda \geqslant 0$, $f^+(\lambda) = 0$ for $\lambda < 0$ and $f^-(\lambda) = -\lambda$ for all $\lambda \leqslant 0$, $f^-(\lambda) = 0$ for $\lambda > 0$. Then $x = f^+(x) - f^-(x)$, $y = f^+(y) - f^-(y)$, $f^+(x) \leqslant f^+(y)[\mu]$, and $f^-(y) \leqslant f^-(x)[\mu]$. By the above we have

$$\mu(x) = \mu[f^+(x)] - \mu[f^-(x)] \leqslant \mu[f^+(y)] - \mu[f^-(y)] = \mu(y). \square$$

By a slight modification of the above proof we have:

Corollary 3.6: If $0\leqslant x\leqslant y[\mu]$, then

$$\int_{[M,\infty)}\lambda\mu_x(d\lambda)\leqslant\int_{[M,\infty)}\lambda\mu_y(d\lambda)$$

for all $M\in\mathbf{R}$.

We write $\lim x_i = x[\mu]$ if $\lim\mu[x_i(\alpha,\infty)] = \mu[x(\alpha,\infty)]$ for all $\alpha\in\mathbf{R}$. Notice that if $x_1\leqslant x_2\leqslant\cdots$ and $\lim x_i = x$, then $\lim x_i = x[\mu]$.

The *center* of P is the set $Z(P) = \{a: a\leftrightarrow b \text{ for all } b\in P\}$. A measure μ is *σ-finite* if there exists an orthogonal sequence $a_i\in Z(P)$ such that $\vee a_i = 1$ and $\mu(a_i) < \infty$, $i = 1, 2, \cdots$.

Even for the classical case, the next theorem is a generalization of the monotone convergence theorem.

Theorem 3.7: Let (P,μ) be a measure logic with μ σ-finite. If $y\leqslant x_1\leqslant x_2\leqslant\cdots[\mu]$, where $\mu(y) > -\infty$ and $\lim x_i = x[\mu]$, then $\lim\mu(x_i) = \mu(x)$.

Proof: *Step* 1. Assume μ is finite. Since

$\lim F(x_n,\alpha) = \lim\mu[x_n(\alpha,\infty)] = \mu[x(\alpha,\infty)] = F(x,\alpha)$ for all $\alpha\in\mathbf{R}$, the measures μ_{x_i} converge weakly to the measure μ_x.[12]

Step 2. Assume μ is finite and $x_1, x_2, \cdots$ are uniformly bounded. That is, there exist $M, K\in\mathbf{R}$ such that $M\leqslant x_i\leqslant K$, $i = 1, 2, \cdots$. Then $x_i[M,K]^c = 0$ so the measures μ_{x_i}, μ_x are concentrated on $[M,K]$, $i = 1, 2, \cdots$. Let $f:\mathbf{R}\to\mathbf{R}$ be the function

$$f(\lambda) = \begin{cases}\lambda & \text{if } M\leqslant\lambda\leqslant K\\ M & \text{if } \lambda < M\\ K & \text{if } \lambda > K\end{cases}.$$

Since μ_{x_i} converges weakly to μ_x and f is bounded and continuous, we have[9]

$$\lim\mu(x_i) = \lim\int\lambda\mu_{x_i}(d\lambda) = \lim\int f(\lambda)\mu_{x_i}(d\lambda)$$
$$= \int f(\lambda)\mu_x(d\lambda) = \int\lambda\mu_x(d\lambda) = \mu(x).$$

Step 3. Assume μ is finite and $x_1, x_2, \cdots$ are uniformly bounded below. Then there exists an $M\in\mathbf{R}$ such that $x_i(-\infty, M) = 0$, $i = 1, 2, \cdots$. For $K > M$, let $f_K(\lambda) = \lambda$ if $\lambda\leqslant K$ and $f_K(\lambda) = K$ if $\lambda > K$. Then $f_K^{-1}(\alpha,\infty) = \phi$ if $\alpha > K$ and $f_K^{-1}(\alpha,\infty) = (\alpha,\infty)$ if $\alpha\leqslant K$. It is clear that $f_K(x_1)\leqslant f_K(x_2)\leqslant\cdots[\mu]$, $M\leqslant f_K(x_i)\leqslant K$ and $\lim f_K(x_i) = f_K(x)[\mu]$. By Step 2, $\lim\mu[f_K(x_i)] = \mu[f_K(x)]$ By the classical monotone convergence theorem

$$\lim_{K\to\infty}\mu[f_K(x)] = \lim_{K\to\infty}\int f_K(\lambda)\mu_x(d\lambda)$$
$$= \int\lambda\mu_x(d\lambda) = \mu(x).$$

If $\mu(x) = +\infty$, then

$$\lim\mu((x_i)\geqslant\lim\mu[f_K(x_i)] = +\infty.$$

If $\mu(x) < \infty$, then given an $\epsilon > 0$ there exists a $K\in\mathbf{R}$ such that $0\leqslant\mu(x) - \mu[f_K(x)] < \epsilon/2$. Also, there exists an integer N such that $i\geqslant N$ implies $0\leqslant\mu[f_K(x)] - \mu[f_K(x_i)]\leqslant\epsilon/2$. Hence, $i\geqslant N$ implies

$$0\leqslant\mu(x) - \mu[f_K(x_i)] = \{\mu(x) - \mu[f_K(x)]\}$$
$$+ \{\mu[f_K(x)] - \mu[f_K(x_i)]\} < \epsilon.$$

Since $\mu[f_K(x_i)]\leqslant\mu(x_i)\leqslant\mu(x)$ we have $\lim\mu(x_i) = \mu(x)$.

Step 4. Assume μ is finite and $y\leqslant x_i\leqslant 0[\mu]$. Since $\mu(y) > -\infty$ we have $\int_{(-\infty,M]}\lambda\mu_y(d\lambda)\to 0$ as $M\to-\infty$. Choose M sufficiently small so that $|\int_{(-\infty,M]}\lambda\mu_y(d\lambda)| < \epsilon$. Since $y\leqslant x_i\leqslant x\leqslant 0[\mu]$ it follows from taking the negative of Corollary 3.6 that

$$-\epsilon < \int_{(-\infty,M]}\lambda\mu_y(d\lambda)$$
$$\leqslant\int_{(-\infty,M]}\lambda\mu_{x_i}(d\lambda)$$
$$\leqslant\int_{(-\infty,M]}\lambda\mu_x(d\lambda)\leqslant 0,$$

which implies that for all i

$$\left|\int_{(-\infty,M]}\lambda\mu_{x_i}(d\lambda) - \int_{(-\infty,M]}\lambda\mu_x(d\lambda)\right| < \epsilon. \tag{2}$$

Define $g_M(\lambda) = \lambda$ if $\lambda\geqslant M$ and $g_M(\lambda) = M$ if $\lambda < M$. Then $M\leqslant g_M(x_1)\leqslant g_M(x_2)\leqslant\cdots[\mu]$ and $\lim g_M(x_i) = g_M(x)[\mu]$. By Step 3,

$$\lim_{i\to\infty}\mu[g_m(x_i)] = \mu[g_M(x)]. \tag{3}$$

Moreover,

$$\mu[g_M(x)] = \int g_M(\lambda)\mu_x(d\lambda) = \int_{(-\infty,M]} g_M(\lambda)\mu_x(d\lambda) + \int_{(M,\infty)} g_M(\lambda)\mu_x(d\lambda) = M\mu_x(-\infty,M] + \int_{(M,\infty)}\lambda\mu_x(d\lambda).$$

Similarly,

$$\mu[g_M(x_i)] = M\mu_{x_i}(-\infty,M] + \int_{(M,\infty)}\lambda\mu_{x_i}(d\lambda).$$

Since $\lim x_i = x[\mu]$ we have

$$\mu_{x_i}(-\infty,M]\downarrow\mu_x(-\infty,M]. \tag{4}$$

Now

$$|\mu(x_i) - \mu(x)| = \left|\int_{(-\infty,M]}\mu_{x_i}(d\lambda) + \int_{(M,\infty)}\lambda\mu_{x_i}(d\lambda) - \int_{(-\infty,M]}\lambda\mu_x(d\lambda) - \int_{(M,\infty)}\lambda\mu_x(d\lambda)\right|$$
$$= \left|\int_{(-\infty,M]}\lambda\mu_{x_i}(d\lambda) + \mu[g_M(x_i)] - M\mu_{x_i}(-\infty,M] - \int_{(-\infty,M]}\lambda\mu_x(d\lambda) - \mu[g_M(x)] + M\mu_x(-\infty,M]\right|$$
$$\leqslant\left|\int_{(-\infty,M]}\lambda\mu_{x_i}(d\lambda) - \int_{(-\infty,M]}\lambda\mu_x(d\lambda) + |\mu[g_M(x_i)] - \mu[g_M(x)]|\right| + |M|\,|\mu_{x_i}(-\infty,M] - \mu_x(-\infty,M]|.$$

Using (2), (3), and (4) we have $\lim |\mu(x_i) - \mu(x)| < \epsilon$, and since $\epsilon > 0$ was arbitrary, $\lim \mu(x_i) = \mu(x)$.

Step 5. Assume that μ is finite. The observables $f^+(x_i)$ and $f^+(x)$ are bounded below by 0 and $\lim f^+(x_i) = f^+(x)[\mu]$. Hence, by Step 3, $\lim \mu[f^+(x_i)] = \mu[f^+(x)]$. Also, $-\infty < \mu(y)$ implies that $-\infty < \mu[-f^-(y)]$. Then $-f^-(y) \leqslant -f^-(x_i) \leqslant 0[\mu]$ and $\lim[-f^-(x_i)] = -f^-(x)[\mu]$ together with Step 4 give $\lim \mu[-f^-(x_i)] = \mu[-f^-(x)]$. Therefore,

$$\begin{aligned}\lim \mu(x_i) &= \lim \mu[f^+(x_i)] - \lim \mu[f^-(x_i)] \\ &= \mu[f^+(x)] - \mu[f^-(x)] = \mu(x).\end{aligned}$$

Step 6. Now assume that μ is σ finite. Then there exists an orthogonal sequence $a_i \in Z(P)$ such that $\vee a_i = 1$ and $\mu(a_i) < \infty$, $i = 1, 2, \cdots$. Define $\mu_i(a) = \mu(a \wedge a_i)$ for all $a \in P$. Then μ_i is a finite measure and

$$\mu(a) = \mu[\vee(a \wedge a_i)] = \sum \mu(a \wedge a_i) = \sum \mu_i(a)$$

for all $a \in P$. Let z be an observable for which $\mu(z)$ exists. Define the Borel measures ν_n by $\nu_n(E) = \sum_{i=1}^{n} \mu_{iz}(E)$, $E \in B(\mathbf{R})$. Then ν_n is an increasing sequence of Borel measures and

$$\begin{aligned}\lim_{n\to\infty} \nu_n(E) &= \sum_{i=1}^{\infty} \mu_{iz}[(E)] = \sum_{i=1}^{\infty} \mu_i[z(E)] \\ &= \mu[z(E)] = \mu_z(E)\end{aligned}$$

for every $E \in B(\mathbf{R})$. It follows from Ref. 8; Theorem IV. 9.5 that

$$\begin{aligned}\sum_{i=1}^{\infty} \mu_i(z) &= \lim_{n\to\infty} \sum_{i=1}^{n} \int \lambda \mu_{iz}(d\lambda) = \lim_{n\to\infty} \int \lambda \nu_n(d\lambda) \\ &= \int \lambda \mu_z(d\lambda) = \mu(z).\end{aligned}$$

Now by Step 5, $\lim_{i\to\infty} \mu_j(x_i) = \mu_j(x)$ for every j. Since $\mu_j(y) \leqslant \mu_j(x_1) \leqslant \mu_j(x_2) \leqslant \cdots$, and $\sum_{j=1}^{\infty} \mu_j(y) > -\infty$, by the classical monotone convergence theorem we have

$$\begin{aligned}\lim_{i\to\infty} \mu(x_i) &= \lim_{i\to\infty} \sum_{j=1}^{\infty} \mu_j(x_i) = \sum_{j=1}^{\infty} \lim_{i\to\infty} \mu_j(x_i) \\ &= \sum_{j=1}^{\infty} \mu_j(x) = \mu(x). \qquad \square\end{aligned}$$

In the classical theory one can go immediately from Step 3 to Step 5 as follows. Assume $y \leqslant x_1 \leqslant x_2 \leqslant \cdots [\mu]$, $\mu(y) > -\infty$, and $\lim x_i = x[\mu]$. If $\mu(y) = +\infty$, the result is immediate. Now assume that $\mu(y) < \infty$ and subtract y to get $0 \leqslant x_1 - y \leqslant x_2 - y \leqslant \cdots [\mu]$, and $\lim(x_i - y) = x - y[\mu]$. Applying Step 3 we obtain $\lim \mu(x_i - y) = \mu(x - y)$. By the additivity of the integral $\lim \mu(x_i) - \mu(y) = \mu(x) - \mu(y)$ and Step 5 is accomplished. However, we cannot use this simple method in our proof for two reasons. First, $x_i - y$ does not make sense in general. In the case of a generalized measure space, if x_i and y are measurable functions and $x_i - y$ happens to be measurable, then $x_i - y$ does have a meaning. But now (this is the second reason), the integral need not be additive. In Sec. IV we shall give a counterexample to show this. We also shall give counterexamples to show that generalizations of the dominated convergence theorem, Fatou's lemma, and Egoroff's theorem do not hold.

IV. COUNTEREXAMPLES

In Corollary 3.2 we showed that the limit of an increasing sequence of measurable functions in a generalized measurable space is measurable. In our first counterexample we show that this is not true for an arbitrary pointwise limit.

Example 4.1: Let $\Omega = [-1, 1]$. We define a sequence of subsets B_n, $n = 0, 1, 2, \ldots$, inductively.

Let $c_j(0) = 1/j, j = 1, 2, \ldots$ and $B_0 = (c_1(0), c_2(0), \ldots)$.

$$\text{Let } c_j(n) = \begin{cases} \dfrac{c_j(n-1) + 1/(j+1)}{2}, & j = 1, 2, \ldots, n \\ c_j(n-1), & j = n+1, \quad n+2, \ldots \end{cases}$$

and $B_n = (c_1(n), c_2(n), \ldots)$. For $n \neq m$, the sequences B_n and B_m agree eventually, but differ on some initial finite set of points. Let $C_n = [-1, 0] \cup B_n$, and $A_n = C_n^c$, $n = 0, 1, 2, \ldots$ By construction $A_n \cap C_m \neq \phi$ for $n \neq m$, and it follows that $C = \{\Omega, \phi, C_n, A_n, n = 0, 1, \ldots\}$ is a σ class.

Define the functions

$\phi_n = \chi_{A_n} = 1$ on A_n

0 on C_n, $n = 0, 1, 2, \ldots$.

We will show that $\phi_n \to \phi = \chi_{(0,1]}$.

(1) Let $\omega \in [-1, 0]$. Then $\omega \in C_n$ for every n, and $\phi_n(\omega) = 0$ for every n. Hence $\phi_n(\omega) \to 0 = \phi(\omega)$.

(2) Let $\omega \in (0, 1]$. Then $\omega \in (1/(k+1), 1/k]$ for some k. Since $c_k(n) \downarrow n\ 1/(k+1)$, $c_k(n) < \omega$ eventually. Further, $\omega \leqslant 1/k < c_{k-1}(n)$ for every n. Therefore, $c_k(n) < \omega < c_{k-1}(n)$ for n sufficiently large. Thus $\omega \in A$ eventually, and $\phi_n(\omega) = 1$ eventually. Hence $\lim_{n\to\infty} \phi_n(\omega) = 1 = \phi(\omega)$. $\phi = \chi_{(0,1]}$ is not measurable since $(0, 1] \notin C$. $\square$

Theorem 3.4 showed that if g is compatible with an increasing sequence of measurable functions, then g is compatible with the limit. The next counterexample shows that the monotonicity of the sequence cannot be removed.

Example 4.2: Let $\Omega = [-1, 2]$. Let B_n and A_n be defined as in Example 4.1. B_n, A_n are subsets of $(0, 1]$. Let B_n', A_n' be the analogous subsets of $(1, 2]$, i.e., $B_n' = B_n + 1$, $A_n' = A_n + 1$. Further let $I_1 = [-1, -1/2)$, $I_2 = [-1/2, 0]$, and $U = [-1, 0]$. Let $C(E)$ be the σ class generated by the following collection of subsets of Ω: $E = \{U, A_n, A_m', I_2 \cup B_n, I_1 \cup B_m', B_n \cup B_m', \text{ for every } n, m\}$. We will show that C, the collection of finite disjoint unions of sets of E (including the empty set), is a σ class, and hence $C(E) = C$. Except for $B_n \cup B_m'$ the Lebesgue measure of each set of C is greater than $1/2$ and $B_n \cup B_m'$ is not disjoint from any other $B_i \cup B_j'$. Therefore no countably infinite disjoint union of sets from E is possible. Also $\Omega = U \cup A_n \cup A_n' \cup (B_n \cup B_n) \in C$. Therefore it suffices to show C is closed under complementation, i.e., the complement of every finite disjoint union of sets from E is a finite disjoint union of sets from E. Table I verifies this by listing all the possibilities.

Then let $\phi_n = \chi_U \cup (B_n \cup B_n')$, $h = \chi_{[-1/2, 1]}$, $([-1/2, 1] = (I_2 \cup B_n) \cup A_n)$, and $\phi = \chi_U$. By an analogous argument to that used in Example 4.1, $\phi_n \to \chi_U = \phi$. Also $\phi_n \not\longrightarrow h$ for every n since $[U \cup (B_n \cup B_n')] \cap [-1/2, 1] = I_2 \cup B_n \in C$. But $\phi \not\longrightarrow h$, since $U \cap [-1/2, 1] = I_2 \notin C$. $\square$

The next counterexample shows that generalizations of

TABLE I. Each possible finite disjoint union is listed alongside its complement. We remind the reader that $A_n(A'_n)$ and $B_n(B'_n)$ are not disjoint from $B_m(B'_m)$ $m \neq n$.

U	$A_n \cup A'_m \cup (B_n \cup B'_m)$
$U \cup A_n$	$A'_m \cup (B_n \cup B_m)$
$U \cup A_n \cup A'_m$	$B_n \cup B'_m$
$U \cup A_n \cup A'_m \cup (B_n \cup B'_m) =$ $A_n \cup A'_m \cup (I_2 \cup B_n) \cup (I_1 \cup B'_m)$	ϕ
$U \cup A'_m \cup (B_n \cup B'_m)$ $= A'_m \cup (I_2 \cup B_n \cup (I_1 \cup B'_m)$	A_n
$U \cup (B_n \cup B'_m) = (I_2 \cup B_n) \cup (I_1 \cup B'_m)$	$A_n \cup A'_m$
$A_n \cup A'_m \cup (I_2 \cup B_n)$	$(I_1 \cup B'_m)$
$A_n \cup A'_m \cup (I_1 \cup B'_m)$	$(I_2 \cup B_n)$
$A_n \cup (I_2 \cup B_n) \cup (I_1 \cup B'_m) =$ $U \cup A_n \cup (B_n \cup B'_m)$	A'_m
$A_n \cup (I_1 \cup B'_m)$	$A'_m \cup (I_2 \cup B_n)$
$A_n \cup (B_n \cup B'_m)$	$U \cup (A'_m)$
$A'_m \cup (I_1 \cup B'_m)$	$A_n \cup (I_2 \cup B_n)$

the dominated convergence theorem and Fatou's lemma fail.

Example 4.3: Let A_n, B_n be as in Example 4.1. Recall that $A_n \cup B_n = (0, 1]$. Define the σ algebras $\Sigma_n = (\phi, A_n, B_n, \Omega)$. It is easy to see that $C = \cup_{n=1}^{\infty} \Sigma_n$ is a σ class. The assignment $\mu(A_n) = \mu(B_n) = 1/2$ gives a measure on C, since the only requirement on the measure is $\mu(A_n) + \mu(B_n) = \mu(\Omega)$ for every n. Let $f_n = \chi_{A_n}$. Since $\omega \in A_n$ eventually, $f \equiv 1$ is the pointwise limit of f_n. Then $\int f_n \, d\mu = \mu(A_n) = 1/2$, $\int f \, d\mu = \mu(\Omega) = 1$. Now $|f_n|$ is dominated by the integrable function 1, but $\int f_n \, d\mu \neq \int f \, d\mu$. Hence, the dominated convergence theorem fails. Also, f_n is dominated below by the integrable function 0, but

$$\liminf \int f_n \, d\mu = 1/2 \not> \int f \, d\mu = 1.$$

Hence, Fatou's lemma fails. □

Our next counterexample gives two measurable functions f, g on a generalized measure space (Ω, C, μ) for which $f + g$ is measurable and yet $\int (f + g) \, d\mu \neq \int f \, d\mu + \int g \, d\mu$.

Example 4.4: Let $\mathbf{I}$ denote the integers and let $\Omega = \{(i, j), i, j \in I\}$. Let $R_m = \{(i,j): i = m\}$, $C_n = \{(i,j): j = n\}$, and $D_t = \{(i,j): i + j = t\}$. Let Σ_R, Σ_C, Σ_D be the σ-algebras generated by $\{R_m : m \in \mathbf{I}\}$, $\{C_n : n \in \mathbf{I}\}$, $\{D_t : t \in \mathbf{I}\}$, respectively. Since $R_m \cap C_n \neq \phi$, $R_m \cap D_t \neq \phi$, $C_n \cap D_t \neq \phi$ for all m, m, t it follows that $C = \Sigma_R \cup \Sigma_C \cup \Sigma_D$ is a σ class. Define the measure μ on Σ by $\mu(R_0) = \mu(C_0) = \mu(D_1) = 1$, $\mu(R_m) = \mu(C_n) = \mu(D_t) = 0$ for $m, n \neq 0, t \neq 1$. Define the measurable functions f, g by $f(i, j) = i$, $g(i, j) = j$. Then $(f + g)(i, j) = i + j$ is measurable. Now $\int f \, d\mu = 0 = \int g \, d\mu$ but $\int (f + g) \, d\mu = 1$. □

Let (Ω, C, μ) be a generalized measure space. A sequence of measurable functions f_n converges to a measurable function f *almost everywhere* if $A \subseteq \{\omega : \lim f_n(\omega) \neq f(\omega)\}$ and $A \in C$ implies that $\mu(A) = 0$. We say that f_n converges to f *almost uniformly* if for every $\epsilon > 0$ there exists an $A_\epsilon \in C$ such that $\mu(A_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on A_ϵ^c. It is straightforward to show that convergence almost uniformly implies convergence almost everywhere. The next counterexample shows that the converse (Egoroff's Theorem) need not hold on a finite generalized measure space.

Example 4.5: Let $\Omega = (0, 1]$ and let A_n, B_n be as in Example 4.1. Then $C = \{\phi, \Omega, A_n, B_n, n = 1, 2, \ldots\}$ is a σ class. If we define $\mu(A_n) = \mu(\phi) = 0$, $\mu(B_n) = \mu(\Omega) = 1$, $n = 1, 2, \ldots$, then μ is a measure and (Ω, C, μ) is a generalized measure space. Let $f_n = \chi_{B_n}$. Then $f_n(\omega) \to 0$ for every $\omega \in \Omega$. For $0 < \epsilon < 1$, $\{\omega : f(\omega) > \epsilon\} = \{\omega : f(\omega) = 1\} = B_n$, which implies that $\mu\{\omega : f(\omega) > \epsilon\} = 1$. Hence, $f_n \not\to 0$ almost uniformly.

V. A RADON–NIKODYM THEOREM

An element $a \in p$ is *central* if $a \in Z(P)$. A logic P is *irreducible* if $Z(P) = \{0, 1\}$. In a Hilbertian logic $P(H)$ the only central elements are $\{0\}$ and H so $P(H)$ is irreducible. There are many examples of sublogics of $P(H)$ which are reducible; for example, the projections in a von Neumann algebra which is not a factor. In a generalized measurable space (Ω, C) every singleton set in C is central.

Lemma 5.1: $Z(P)$ is a Boolean sub σ algebra of P.

Proof: If $a \in Z(P)$, then $a' \in Z(P)$ and $0, 1 \in Z(P)$. We now show that $Z(P)$ is a lattice. Suppose $a, b \in Z(P)$ and $c \in P$. Since $b \leftrightarrow c$, $b \wedge c$ exists, and since $a \leftrightarrow b \wedge c$, $a \wedge b \wedge c$ exists. Similarly, $a \wedge c \wedge b'$ and $b \wedge c \wedge a'$ exist. Now

$$a \wedge b \wedge c \leqslant b \leqslant a' \vee c' \vee b = (a \wedge c \wedge b')'.$$

In this way one shows that $a \wedge b \wedge c, a \wedge c \wedge b'$, and $b \wedge c \wedge a'$ are mutually orthogonal. Let

$$d = (a \wedge b \wedge c) \vee (a \wedge c \wedge b') \vee (b \wedge c \wedge a').$$

Clearly, $d \leqslant a \vee b, d \leqslant c$. Hence, $a \vee b = d \vee [(a \vee b) \wedge d']$ and $c = d \vee (c \wedge d')$. Now $d \perp (a \vee b) \wedge d'$ and $d \perp c \wedge d'$. We now show that $c \wedge d' \perp (a \vee b) \wedge d'$. If $e \in P$, it follows from Theorem 2.1 that

$$\begin{aligned}(e \vee b') \wedge (e \vee b) &= [(e \vee b') \wedge e] \vee [(e \vee b') \wedge b] \\ &= e \vee [(e \wedge b) \vee (b' \wedge b)] \\ &= e \vee (e \wedge b) = e.\end{aligned}$$

Hence,

$$(a' \vee c' \vee b') \wedge (a' \vee c' \vee b) = a' \vee c'$$

and

$$(b' \vee c' \vee a') \wedge (b' \vee c' \vee a) = b' \vee c'.$$

Therefore, applying Theorem 2.1 again, gives

$$\begin{aligned}c \wedge d' &= c \wedge (a' \vee b' \vee c') \wedge (a' \vee c' \vee b) \wedge (b' \vee c' \vee a) \\ &= c \wedge (a' \vee c') \wedge (b' \vee c') = (c \wedge a') \wedge (c \wedge b') \\ &= c \wedge a' \wedge b' \leqslant (a' \wedge b') \vee d = [(a \vee b) \wedge d']'.\end{aligned}$$

It follows that $c \leftrightarrow a \vee b$ and $a \vee b \in Z(P)$. In a similar way $c \leftrightarrow a' \vee b'$ and hence, $c \leftrightarrow (a' \vee b')' = a \wedge b$. Thus, $Z(P)$ is a lattice. Applying Theorem 2.1 shows that $Z(P)$ is distributive and σ orthocomplete. We conclude that $Z(P)$ is a Boolean sub σ-algebra of P. □

A *signed measure* ν on P is a map from P to $\mathbf{R} \cup \{-\infty, +\infty\}$ which does not assume both the values $\pm\infty$, and which satisfies $\nu(0) = 0$, $\nu(\bigvee a_i) = \Sigma \nu(a_i)$, where a_i is any orthogonal sequence. An element $a \in p$ is *positive* (*negative*) with respect to a signed measure ν if $b \leqslant a$ implies that $\nu(b) \geqslant 0 (\leqslant 0)$. A signed measure ν has a *central Hahn decomposition* if there exists an $a \in Z(P)$ such that a is positive and a' is negative.

As an example, let $\Omega = \{1, 2, 3, 4\}$ and

$$C = \{\phi, \{1, 2\}, \{3, 4\}, \{1, 3\}, \{2, 4\}, \Omega\}.$$

Then (Ω, C) is an irreducible generalized measurable space. Let $\nu\{1,2\} = \nu\{1,3\} = 1$, $\nu\{2,4\} = \nu\{3,4\} = -1$. Then C contains both positive and negative sets but a central Hahn decomposition for ν is impossible.

An observable x is *central* if the elements of the range of x are central. If x is a central observable, $a \in P$, and μ is a measure on P, then $E \to \mu[x(E) \wedge a]$, $E \in B(\mathbf{R})$, is a Borel measure on $\mathbf{R}$. We use the notation $\int_a x d\mu \equiv \int \lambda \mu[x(d\lambda) \wedge a]$.

Lemma 5.2: Let μ be a measure on P and let x be a central observable such that $\mu(x)$ exists. Then $\nu(a) = \int_a x d\mu$ is a signed measure on P.

Proof: Clearly, $\nu(0) = 0$, and since $\mu(x)$ exists, ν cannot have both $\pm\infty$ as values. Let $a_i \in P$, $i = 1, 2, \ldots$, be an orthogonal sequence. Now there exist mutually disjoint sets $B_i \in B(\mathbf{R})$ and an observable z such that $z(B_i) = a_i$, $i = 1, 2, \ldots$, (Refs. 3 and 5) and since x is central, $x \leftrightarrow z$. Applying Theorem 2.2, there exist Borel functions u, v and an observable y such that $x = u(y)$, $z = v(y)$. Let $A_i = v^{-1}(B_i)$, $i = 1, 2, \ldots$. Then the A_i's are mutually disjoint and

$$y(A_i) = y[v^{-1}(B_i)] = v(y)(B_i) = z(B_i) = a_i.$$

Hence,

$$\begin{aligned}\int_{a_i} x d\mu &= \int \lambda \mu[x(d\lambda) \wedge a_i] = \int \lambda \mu[u(y)(d\lambda) \wedge a_i] \\ &= \int u(\lambda) \mu[y(d\lambda) \wedge a_i] \\ &= \int u(\lambda) \mu[y(d\lambda) \wedge y(A_i)] \\ &= \int u(\lambda) \mu[y(d\lambda \cap A_i)] = \int_{A_i} u(\lambda) \mu[y(d\lambda)] \\ &= \int_{u^{-1}(A_i)} \lambda \mu_x(d\lambda).\end{aligned}$$

In a similar way

$$\int_{\vee a_i} x d\mu = \int_{u^{-1}(\cup A_i)} \lambda \mu_x(d\lambda)$$

Hence,

$$\begin{aligned}\nu(\vee a_i) &= \int_{\cup u^{-1}(A_i)} \lambda \mu_x(d\lambda) \\ &= \sum \int_{u^{-1}(A_i)} \lambda \mu_x(d\lambda) \\ &= \sum \nu(a_i). \end{aligned}$$ □

Corollary 5.3: Let μ be a measure on P, $a \in P$, and x a central observable such that $\mu(x)$ exists. Then there exists an observable y whose range contains a and a Borel function u such that $x = u(y)$ and

$$\int_a x \, d\mu = \int_A u(\lambda) \mu_y(d\lambda) = \int_{u^{-1}(A)} \lambda \mu_x(d\lambda)$$

for all $A \in B(\mathbf{R})$ with $y(A) = a$.

The main concern of this section is to characterize measures of the from $a \to \int_a x \, d\mu$ where x is a central observable. Let ν and μ be two measures on P. If there exists a central observable x such that $\nu(a) = \int_a x d\mu$ for all $a \in P$, then we say that x is a *central Radon–Nikodym* derivative for ν with respect to μ. We say that ν is *absolutely continuous* relative to μ ($\nu \ll \mu$) if $\mu(a) = 0$ implies that $\nu(a) = 0$. We use the notation $a \Delta b = (a \wedge b') \vee (a' \wedge b)$ if the right-hand side exists for a, $b \in P$. The next lemma shows that a central Radon–Nikodym derivative is unique in a certain sense.

Lemma 5.4: Let x and y be central observables such that $\int_a x d\mu = \int_a y d\mu$ for all $a \in P$. Then $x = y$ almost everywhere $[\mu]$; that is, $\mu[x(E) \Delta y(E)] = 0$ for all $E \in B(\mathbf{R})$.

Proof: Suppose x is a central observable and $\int_a x d\mu = 0$ for all $a \in P$. Then

$$\begin{aligned}0 &= \int_{x(0,\infty)} x d\mu = \int \lambda \mu[x(d\lambda) \cap x(0,\infty)] \\ &= \int \lambda \mu[x(d\lambda \cap (0,\infty))] = \int_{(0,\infty)} \lambda \mu[x(d\lambda)] \\ &= \int_{(0,\infty)} \lambda \mu_x(d\lambda).\end{aligned}$$

Hence, $\mu_x(0,\infty) = 0$ and similarly $\mu_x(-\infty, 0) = 0$. Therefore, $\mu_x(\{0\}^c) = 0$. Now let x and y satisfy the hypothesis of the theorem. By theorem 2.2 there exists a central observable z and Borel functions u, v such that $x = u(z)$, $y = v(z)$. Then

$$\begin{aligned}\int_a (u - v)(z) d\mu &= \int [u(\lambda) - v(\lambda)] \mu[z(d\lambda) \wedge a] \\ &= \int u(\lambda) \mu[z(d\lambda) \wedge a] - \int v(\lambda) \mu[z(d\lambda) \wedge a] \\ &= \int \lambda u[x(d\lambda) \wedge a] - \int \lambda \mu[y(d\lambda) \wedge a] \\ &= \int_a x d\mu - \int_a y d\mu = 0.\end{aligned}$$

By the above, $\mu[(u - v)(z)\{0\}^c] = 0$.

Therefore,

$$\begin{aligned}\mu[x(E) \wedge y(E)'] &= \mu[x(E) \wedge y(E^c)] \\ &= \mu[u(z)(E) \wedge v(z)(E^c)] \\ &= \mu[z(u^{-1}(E)) \wedge z(v^{-1}(E^c))] \\ &= \mu[z(u^{-1}(E)) \wedge v^{-1}(E^c)] \\ &\leqslant \mu[z((u - v)(\{0\}^c))] = 0.\end{aligned}$$

Hence, $\mu[x(E) \wedge y(E)'] = 0$ and similarly $\mu[x(E)' \cap y(E)] = 0$. □

Until now we have only considered observables based on $B(\mathbf{R})$. For generality in the next theorem we shall allow observables based on $B(\mathbf{R}_+)$, $\mathbf{R}_+ \equiv \mathbf{R} \cup \{\infty\}$.

Theorem 5.5: Let μ be a finite measure on P and let ν be a measure on P. A central Radon–Nikodym derivative for ν with respect to μ exists if and only if $\mu \ll \mu$ and the signed measures $\nu - \alpha\mu$ admit a central Hahn decomposition for every rational number α.

Proof: Assume that a central Radon–Nikodym derivative x for ν with respect to μ exists. Hence, $\nu(a) = \int_a x d\mu$. That $\nu \ll \mu$ is clear. For $\alpha \in \mathbf{R}$, let $a_\alpha = x(\alpha, \infty)$. Since x is central, $a_\alpha \in Z(P)$. Let $a \leqslant a_\alpha$. Then by Corollary 5.3 there exists an observable y whose range includes a, and a Borel function u such that $x = u(y)$ and

$$\nu(a) = \int_a x d\mu = \int_A u(\lambda) \mu_y(d\lambda)$$

for every $A\in B(\mathbf{R})$ with $y(A)=a$. Now suppose that $y(B)=a$, and let $A=B\cap u^{-1}(\alpha,\infty)$. Then

$$y(A)=y(B)\wedge u(y)(\alpha,\infty)=a\wedge a_\alpha=a.$$

Since $A\subseteq u^{-1}(\alpha,\infty)$ we have

$$(\nu-\alpha\mu)(a)=\nu(a)-\alpha\mu(a)=\int_A u(\lambda)\mu_y(d\lambda)-\alpha\int_A\mu_y(d\lambda)$$
$$=\int_A[u(\lambda)-\alpha]\mu_y(d\lambda)\geqslant 0.$$

Hence, a_α is positive with respect to $\nu-\alpha\mu$. A similar argument shows that a'_α is negative with respect to $\nu-\alpha\mu$.

Now assume that $\nu\ll\mu$ and the signed measures $\nu-\alpha\mu$ admit a central Hahn decomposition a_α for every $\alpha\in Q$, where Q denotes the rationals. By Lemma 5.1, $Z(P)$ is a Boolean sub σ algebra of P. Hence, by Loomis' theorem[10,14] there is a measurable space (Ω,Σ) and a surjective σ-homomorphism $h:\Sigma\to Z(P)$. Since $a_0=1$, letting $A_0=\Omega$, we have $h(A_0)=1$. For $0\neq\alpha\in Q$, let $A_\alpha,B_\alpha\in\Sigma$ satisfy $h(A_\alpha)=a_\alpha$, $h(B_\alpha)=a'_\alpha$. Define the measures $\bar\nu,\bar\mu$ on Σ by $\bar\nu(A)=\nu[h(A)],\bar\mu(A)=\mu[h(A)]$ for all $A\in\Sigma$. Since $\bar\nu\ll\bar\mu$, by the classical Radon–Nikodym theorem, there exists a nonnegative Σ-measurable function $f:\Omega\to\mathbf{R}$ such that $\bar\nu(A)=\int_A f d\bar\mu$ for all $A\in\Sigma$. Let $A\subseteq A_\alpha$. Since $h(A)\leqslant h(A_\alpha)=a_\alpha$ we have

$$\int_A(f-\alpha)d\bar\mu=\int_A f d\bar\nu-\alpha\int_A d\bar\mu=\bar\nu(A)-\alpha\bar\mu(A)$$
$$=\nu[h(A)]-\alpha\mu[h(A)]\geqslant 0.$$

Hence, $f\geqslant\alpha$ a.e.$[\bar\mu]$ on A_α. In a similar way, $f\leqslant\alpha$ a.e.$[\bar\mu]$ on B_α. By adjusting f to be zero on a set of $\bar\mu$ measure zero we may assume that $f\geqslant\alpha$ on A_α and $f<\alpha$ on B_α. Then $A_\alpha=B^c_\alpha$ and $h(A_\alpha)=a_\alpha$, $h(B_\alpha)=a'_\alpha$ (except for an element of μ measure zero). Now define the central observable $x:B(\mathbf{R})\to Z(P)$ by $x(E)=h[f^{-1}(E)]$, $E\in B(\mathbf{R})$.

Notice that $\alpha<\beta$ implies that $B_\alpha\subseteq B_\beta$. Let N be a fixed integer and let $E_k=B_{k+1/N}\cap B^c_{k/N}$, $k=0,1,2,\ldots,E_\infty=\bigcap_{k=0}^{\infty}A_{k/N}$. Then $\Omega=E_\infty\cup(\bigcup_{k=0}^{\infty}E_k)$ is a disjoint union. Defining $b_k=a'_{k+1/N}\wedge a_{k/N}$, $b_\infty=\bigwedge_{k=0}^{\infty}a_{k/N}$ it follows that there exists an element $b_{-\infty}\in Z(P)$ with $\mu(b_{-\infty})=0$ such that $\{b_{-\infty},b_\infty,b_0,b_1,\ldots\}$ are mutually orthogonal and $1=b_{-\infty}\vee b_\infty\vee\bigvee_{k=0}^{\infty}b_k$. Let $a\in P$, and let $l_k=a\wedge b_k$, $k=-\infty,\infty,0,1,\ldots$. Note that $l_k\leqslant a'_{k+1/N}$ implies that $(\nu-(k+1)/N\mu)(l_k)\leqslant 0$, $k=0,1,2,\ldots$. Hence,

$$\nu(l_k)\leqslant[(k+1)/N]\mu(l_k).\tag{5}$$

Also, $l_k\leqslant a_{k/N}$ implies that $(\nu-(k/N)\mu)(l_k)\geqslant 0$, $k=0,1,2,\ldots$. Hence,

$$\nu(l_k)\geqslant(k/N)\mu(l_k).\tag{6}$$

Since $a=l_{-\infty}\vee l_\infty\vee\bigvee_{k=0}^{\infty}l_k$ and $\nu\ll\mu$ we have

$$\mu(a)=\nu(l_\infty)+\sum_{k=0}^{\infty}\nu(l_k).\tag{7}$$

Now $x[\alpha,\infty]=h[f^{-1}[\alpha,\infty]]=h(A_\alpha)=a_\alpha$ a.e.$[\mu]$. Hence,

$$\mu\left[x\left(\frac{k}{N},\frac{k+1}{N}\right)\wedge l_k\right]$$
$$=\mu\left[x\left(-\infty,\frac{k+1}{N}\right)\wedge x\left[\frac{k}{N},\infty\right]\wedge l_k\right]$$
$$=\mu(a_{(k+1)/N}\wedge a_{k/N}\wedge l_k)=\mu(l_k)=\mu[x(\mathbf{R}_+)\wedge l_k].$$

Hence, the Borel measure $E\to\mu[x(E)\wedge l_k]$, $E\in B(\mathbf{R}_+)$ is concentrated on $[k/N,(k+1/N)$. Therefore,

$$(k/N)\mu(l_k)\leqslant\int\lambda\mu[x(d\lambda)\wedge l_k]\leqslant[(k+1)/N]\mu(l_k).\tag{8}$$

From (5) we have that $\nu(l_k)-(1/N)u(l_k)\leqslant(k/N)\mu(l_k)$. This together with (8) gives

$$\nu(l_k)-(1/N)\mu(l_k)\leqslant\int\lambda\mu[x(d\lambda)\wedge l_k]=\int_{l_k}x\,d\mu.\tag{9}$$

Similarly, from (6), $[(k+1)/N]\mu(l_k)\leqslant\nu(l_k)+(1/N)\mu(l_k)$. This together with (8) and (9) gives

$$\nu(l_k)-(1/N)\mu(l_k)\leqslant\int_{l_k}x\,d\mu\leqslant\nu(l_k)+(1/N)\mu(l_k).\tag{10}$$

If $\mu(l_\infty)=0$, then $\nu(l_\infty)=0$ so $\int_{(l_\infty)}x d\mu=0=\nu(l_\infty)$. Now suppose that $\mu(l_\infty)>0$. Since $l_\infty=a\wedge b_\infty\leqslant\wedge_{k=0}^{\infty}a_{k/N}$ we have $\nu(l_\infty)\geqslant(k/N)\mu(l_\infty)$, $k=0,1,\ldots$. Hence, $\nu(l_\infty)=\infty$. Now

$$x\{\infty\}=x(\cap[k/N,\infty])=\wedge x[k/N,\infty]=\wedge a_{k/N}\geqslant l_\infty$$

which gives

$$\mu[x\{\infty\}\wedge l_\infty]=\mu(l_\infty)=\mu[x(\mathbf{R}_+)\wedge l_\infty].$$

Thus, the Borel measure $\mu[x(\cdot)\wedge l_\infty]$ is concentrated at ∞. Hence,

$$\int_{l_\infty}x\,d\mu=\int\lambda\mu[x(d\lambda)\wedge l_\infty]=\int_{|\infty|}\lambda\mu[x(d\lambda)]\wedge l_\infty$$
$$=\infty\mu(l_\infty)=\nu(l_\infty).$$

Thus, in either case

$$\int_{l_\infty}x\,d\mu=\nu(l_\infty).\tag{11}$$

Summing (10) over k and using (7), (11), and Lemma 5.2 gives

$$\nu(a)-(1/N)\mu(a)\leqslant\int_a x\,d\mu\leqslant\nu(a)+(1/N)\mu(a).$$

Since μ is finite and N arbitrary, $\nu(a)=\int_a x\,d\mu$. □

Theorem 5.5 can be easily generalized to the case in which μ is σ finite and ν is a signed measure. Also Q can be replaced by any countable dense subset of R.

For an example in which the central Radon–Nikodym derivative does not exist, let (Ω,C) be as in the example preceding Lemma 5.2. Define μ to be 1 on the doubleton sets of C and $\nu\{1,2\}=\nu\{1,3\}=1$, $\nu\{3,4\}=\nu\{2,4\}=2$. Clearly $\nu\ll\mu$. Since C is irreducible, the only central observables (functions) are constants. Hence, the central Radon–Nikodym derivative of ν with respect to μ cannot exist because ν is not a constant multiple of μ.

As another example, let (Ω,C,μ) be a generalized measure space, where C contains a singleton set $\{\omega\}$ and $\mu\{\omega\}=\delta>0$. Let ν be the point measure concentrated at $\{\omega\}$. Then $\nu\ll\mu$ and the function $f=\delta^{-1}\chi_{\{\omega\}}$ is central. Further, $\nu(E)=\int_E f d\mu=1$ if $\omega\in E$, 0 if $\omega\notin E$. Hence, f^{-1} is the

central Radon–Nikodym derivative of ν with respect to μ. The central Hahn decompostion of $\nu - \alpha\mu$ is $a_\alpha = \{\omega\}$ for $\alpha \leqslant \delta^{-1}$ and $a_\alpha = \phi$ for $\alpha > \delta^{-1}$.

We say that P has the *Radon–Nikodym property* if $\nu \ll \mu$ (μ σ finite) implies that the central Randon–Nikodym derivative of ν with respect to μ exists. We have seen an example of a logic which does not have the Radon–Nikodym property. Another example is a Hilbertian logic $P(H)$, dim $H \geqslant 2$.

Lemma 5.6: If P is a Boolean σ- algebra, then P has the Radon–Nikodym property.

Proof: Suppose $\nu \ll \mu$ on P. By Loomis' theorem[10,14] there exists a measurable space (Ω, Σ) and a σ homomorphism h from Σ onto P. Define the measures $\bar{\nu}(A) = \nu[h(A)]$, $\bar{\mu}(A) = \mu[h(A)]$, $A \in \Sigma$. Then $\bar{\nu} \ll \bar{\mu}$ and by the Radon–Nikodym theorem, there exists a measurable function f such that $\bar{\nu}(A) = \int_A f d\bar{\mu}$ for all $A \in \Sigma$. Let x be the observable $x(E) = h[f^{-1}(E)], E \in B(\mathbf{R})$. Then for any $a \in P$, there is an $A \in \Sigma$ with $h(A) = a$, and

$$\begin{aligned}\int_a x d\mu &= \int \lambda\mu[x(d\lambda) \wedge a] = \int \lambda\mu[h(f^{-1}(d\lambda) \wedge h(A)] \\ &= \int \lambda\mu[h(f^{-1}(d\lambda) \cap A)] = \int \lambda\bar{\mu}[f^{-1}(d\lambda) \cap A] \\ &= \int_A \lambda\bar{\mu}[f^{-1}(d\lambda)] = \int_A f(\lambda)\bar{\mu}(d\lambda) = \bar{\nu}(A) \\ &= \nu[h(A)] = \nu(a). \qquad \square\end{aligned}$$

There are non-Boolean logics which have the Radon–Nikodym property. Indeed, there are non-Boolean logics which have no nontrivial measures.[13] Such logics would have the Radon–Nikodym property vacuously. Less trivial examples would be a non-Boolean logic P which has essentially only one nontrivial measure μ.[14] That is, every measure is a constant multiple of μ. This shows that the Radon–Nikodym property on P does not imply that P is Boolean. However, we conjecture that if C is a σ class of subsets of Ω, then if C has the Radon–Nikodym property, C must be Boolean. We have been able to prove this conjecture only in the case in which $Z(C)$ is generated by a countable partition. This result will follow from a series of lemmas.

Lemma 5.7: Let (Ω, C) be a generalized measurable space with C irreducible. If C has the Radon–Nikodym property, then $C = Z(C)$.

Proof: Assume $C \neq Z(C) = \{\phi, \Omega\}$. Then there exist A, $B \in C$ with $A \not\leftrightarrow B$. It follows that $A \cap B$ and $A' \cap B$ are two disjoint, nonempty sets, and we can choose two distinct points ω_1 and ω_2, one from each set, respectively. Let ν_i denote the point measure at ω_i, $i = 1, 2$. Letting $\mu = \nu_1 + \nu_2$ we see that $\nu_1 \ll \mu$. If a central Radon–Nikodym derivative f exists it must be a constant. But

$$\nu_1(A)/\mu(A) = 1 \neq 1/2 = \nu_1(B)/\mu(B)$$

so ν_1 is not a constant multiple of μ. Hence, f does not exist and $C = Z(C)$. $\square$

For $A \in C$, we use the notation

$$C(A) = \{B \in C : B \subseteq A\}.$$

It is straightforward to show that $C(A)$ is a σ-class A.

Lemma 5.8: If $A \in Z(C)$ then $Z[C(A)] = Z(C) \cap A$. The central functions of $C(A)$ are exactly the central functions of C restricted to A.

Proof: Suppose that $D \in Z[C(A)]$. Let $E \in C$. Since $A \in Z(C)$, $E \cap A \in C$ and hence, $E \cap A \in C(A)$. Since $D \in Z[C(A)]$, $D \cap (E \cap A) \in C(A) \subseteq C$. But $D \cap (E \cap A) = D \cap E$ so $D \cap E \in C$. Hence, $D \in Z(C) \cap A$, and $Z[C(A)] \subseteq Z(C) \cap A$. Now suppose that $D \in Z(C) \cap A$. Since $Z(C)$ is a Boolean algebra and $A \in Z(C)$, we have $Z(C) \cap A \subseteq Z(C)$. Hence, $D \in Z(C)$. If $E \in C(A)$, then $D \cap E \in C(A)$. Therefore, $D \in Z[C(A)]$ and $Z(C) \cap A \subseteq Z[C(A)]$. The second statement of the lemma follows directly from the first. $\square$

Lemma 5.9: If $A \in Z(C)$ and the Radon–Nikodym property holds on (Ω, C), it also holds on $(A, C(A))$.

Proof: Let μ_A, ν_A be two measures on $C(A)$ such that $\nu_A \ll \nu_A$. Define the measures, ν, μ on C by $\nu(D) = \nu_A(D \cap A)$, $\mu(D) = \mu_A(D \cap A)$ for all $D \in C$. It is clear that $\nu \ll \mu$. By the Radon–Nikodym property for (Ω, C) there exists a central function f such that $\nu(D) = \int_D f d\mu$ for all $D \in C$. By Lemma 5.8, $f|A$ is central in $C(A)$. Hence, for $B \in C(A)$, $B \subseteq A$ and $\nu_A(B) = \nu(B) = \int_B f|A \, d\mu$. $\square$

Theorem 5.10: If $Z(C)$ is generated by a countable partition of Ω ($\Omega = \cup A_i$), and C has the Radon–Nikodym property, then $C = Z(C)$.

Proof: Applying Lemma 5.9, $(A_i, C(A_i))$ has the Radon–Nikodym property. By Lemma 5.8, $Z[C(A_i)] = Z(C) \cap A_i = \{\phi, A_i\}$. By Lemma 5.7, $C(A_i) = \{\phi, A_i\}$. Since the A_i's are central every $B \in C$ has the form $B = \cup(B \cap A_i)$, where $B \cap A_i \in C(A_i)$. Hence, $B \cap A_i = \phi$ or A_i, and it follows that every $B \in C$ is a disjoint union of the A_i sets. Hence, $C = Z(C)$. $\square$

[1] S. Gudder, Found. Phys. **3**, 399, (1973).
[2] T. Fine, *Theories of Probability: An Examination of Foundation* (Academic, New York, 1973).
[3] S. Gudder, *Stochastic Methods in Quantum Mechanics* (North-Holland, New York, 1979).
[4] G. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[5] V. Varadarajan, *Geometry of Quantum Theory*, Vol, I (Van Nostrand, Princeton, N. J., 1968).
[6] S. Gudder, Proc. Am. Math. Soc. **21**, 296 (1969).
[7] A. Gleason, J. Rat. Mech. Anal. **6**, 885 (1965).
[8] N. Dunford and J. Schwartz, *Linear Operators*, Parts I and II (Wiley, New York, 1958, 1963).
[9] H. Mullikin and S. Gudder, J. Math. Phys. **14**, 234 (1973).
[10] L. Loomis, Bull. A. Math. Soc. **53**, 757 (1947).
[11] A. Ramsey, J. Math. Mech. **15**, 227 (1966).
[12] M. Loeve, *Probability Theory* (Van Nostrand, Princeton, N. J., 1963).
[13] R. Greechie, J. Comb. Theory **10**, 119 (1971).
[14] R. Greechie, Synthese **29**, 113 (1974).

A model of a quantum mechanical treatment of measurement with a physical interpretation

Franklin E. Schroeck, Jr.
Department of Mathematics, University of Denver, Denver, Colorado 80208
Department of Mathematics, Florida Atlantic University, Boca Raton, Florida 33432

(Received 15 May 1980; accepted for publication 1 May 1981)

Projections onto minimum uncertainty states for Weyl systems are interpreted as being associated with elementary particles. This identification yields physically motivated definitions of measurements, instruments, and observables, which, in turn, lead to a quantum mechanical treatment of instruments, a divergence from the von Neumann collapse scheme of measurement, a framework for simultaneous measurement of position and momentum, and a derivation of action-at-a-distance—all consequences of the uncertainty relations.

PACS numbers: 03.65.Bz, 12.20.Hn

I. INTRODUCTION

In what follows, we approach measurement theory from the point of view that instruments are composed of fundamental particles. A particle, in turn, is associated with an irreducible representation of some appropriate group of transformations. We shall deal here exclusively with irreducible representations of a Weyl system. The corresponding particles will be shown to be indexed by points in fuzzy phase space, the fuzziness being a natural consequence of the noncommuting property of position and momentum (i.e., of the uncertainty principle). (A similar fuzzy phase space treatment of the Galilei group and the Poincaré group has been developed.[1]) This fuzziness is passed on to the instruments and to all physically meaningful functions of position and momentum in an inherent fashion. The precise form of this fuzziness is explicitly computed in all that follows, providing a precise description of uncertainty! Some physical consequences of this formulation are (i) a framework for the simultaneous measurement of position and momentum,[2] (ii) a quantum mechanical treatment of instruments, (iii) an explicit computation based on uncertainty showing conflict with the von Neumann collapse scheme of measurement, (iv) the result that action-at-a-distance is an inherent consequence of the uncertainty principle. Some mathematical results include an example of a joint spectral family for noncommuting operators and an example of a probability measure on a noncommuting algebra satisfying all marginality conditions.

In Sec. II we introduce the basic operators projecting out irreducible representations of a Weyl system, which we parametrize with the coordinates of fuzzy phase space. In Sec. III we discuss the corresponding instruments and conditioning of states. In Sec. IV the uncertainty principle is reinvestigated. Section V deals with the physically meaningful observable functions of position and momentum, which correspond with the instruments in the canonical fashion. Here we also display the joint spectral family for position and momentum. In Sec. VI we discuss sequential conditioning and discuss the conflict with the von Neumann collapse scheme. In contrast, in Sec. VII, we discuss simultaneous conditioning from both of the standpoints of coherent and incoherent addition of states, and then compare with the results of Sec. VI for an infinite alternating sequence of conditionings. The classical limit, $\hbar \to 0$, of our results is studied in Sec. VIII. In Sec. IX free dynamics as well as dynamics in the presence of N harmonic oscillator potentials is computed. Finally in Sec. X comparisons with the C^*-algebraic approach to physics and comparisons with quantum logic are discussed.

II. PROJECTIONS ON IRREDUCIBLE REPRESENTATIONS OF THE WEYL ALGEBRA

Let $\mathscr{H}$ be a Hilbert space, and let $\{W(x,y) | x, y \in \mathbb{R}\} \equiv \mathscr{W}$ be a set of unitary operators on $\mathscr{H}$ satisfying the Weyl relations

$$W(x,y)W(a,b) = \exp[-i\hbar(ay - xb)/2] \times W(x+a, y+b). \qquad (*)$$

If $W(x,y)$ is strongly continuous from $\mathbb{R}^2$ into the set of linear operators on $\mathscr{H}$, then, by Stone's theorem, there exist self-adjoint operators P,Q such that $[P,Q] = -i\hbar\mathbb{1}$ and $W(x,y) = \exp(ixP + iyQ)$; i.e., P and Q satisfy the canonical commutation relations for position and momentum.

Since $\mathscr{W}$ is isomorphic with $\mathbb{R}^2$, we may think of complex-valued functions g over $\mathscr{W}$ as functions over $\mathbb{R}^2$ and may therefore consider integrals over $\mathscr{W}$ in the form

$$T_g \equiv \int_{\mathbb{R}^2} g(x,y) W(x,y)\, dx dy, \qquad (1)$$

the integral being meaningful in the strong operator topology for any Lebesgue measurable function g.

If T_g not only exists, but in fact is self-adjoint, then $g(x,y) = \overline{g(-x,-y)}$, the bar denoting complex conjugation. If $g \in \mathscr{L}_1(\mathbb{R}^2)$, then T_g exists as a uniform integral and $\|T_g\| \leq \|g\|_1$. In this case the necessary condition for self-adjointness is also sufficient. T_g is positive iff for every $f \in \mathscr{L}^2(\mathbb{R})$,

$$0 \leq \int dz\, dz'\, dy\, \overline{f(z)} g((z'-z)/\hbar, y) \times \exp[i(z'+z)y/2] f(z').$$

If one prefers to work with position and momentum in n dimensions, then the x, y in $W(x,y)$ should be elements of $\mathbb{R}^n$, P and Q should be n component operators, and most of the

 0022-2488/81/112562-11$01.00

products above should be dot products. This just leads to more cumbersome notation and may be introduced in any of the following sections in an obvious manner. For simplicity of notation, we shall continue the discussion in the one-dimensional language, a decision which the reader will soon applaud.

If we let $a,b,c\in\mathbb{R}$, $c>0$, and take for $g(x,y)$ the particular function

$$g(x,y;a,b,c) = (\hbar/2\pi)\exp\left[-\tfrac{1}{4}\hbar c(x^2+y^2/c^2) - ibx - iay/c\right], \quad (2)$$

then $\|g\|_1 = 2$, and T_g exists as a uniform integral, is positive and, furthermore, satisfies $T_g = T_g{}^2$ (a projection) (hence $\|T_g\| = 1$); also

$$[P - icQ + (ia - b)\mathbf{1}]T_g = 0. \quad (3)$$

T_g may therefore be interpreted as a projection onto the subspace spanned by a coherent state in the case of an irreducible representation. Now, for any positive trace class operator ρ (not necessarily of trace 1), if we define
$\langle P\rangle = \mathrm{Tr}(PT_g\,\rho T_g)$, $\langle Q\rangle = \mathrm{Tr}(QT_g\,\rho T_g)$,
$\langle \mathbf{1}\rangle = \mathrm{Tr}(T_g\,\rho T_g) = \mathrm{Tr}(\rho T_g)$, then we have, from (3),

$$\langle P\rangle - ic\langle Q\rangle + (ia - b)\langle \mathbf{1}\rangle = 0. \quad (3')$$

Since $P,Q,\mathbf{1}$ are self-adjoint, $\langle Q\rangle,\langle P\rangle,\langle \mathbf{1}\rangle$ are real, so $b = \langle P\rangle/\langle \mathbf{1}\rangle$ and $a = c\langle Q\rangle/\langle \mathbf{1}\rangle$, assuming $\langle \mathbf{1}\rangle \neq 0$. If $\langle \mathbf{1}\rangle = 0$, then $T_g\,\rho T_g = 0$, and so $\langle P\rangle = \langle Q\rangle = 0$ and (3') is trivial. To realize $T_g\,\rho T_g = 0$, suppose that the wavefunction ψ is orthogonal to the ground state ϕ_0 of a harmonic oscillator potential relative to the canonical pair $P + b\mathbf{1}$, $Q - a\mathbf{1}$. Then ψ is a linear combination of wavefunctions of excited states, i.e., of functions of the form $[(P + b\mathbf{1}) + i(Q - a\mathbf{1})]^n\phi_0, n\geqslant 1$. Thus we may write $\psi = [(P + b\mathbf{1}) + i(Q - a\mathbf{1})]\psi'$ for some (unnormalized) wavefunction ψ'. Now from (3) for $c = 1$, taking adjoints, we have

$$T_g[(P + b\mathbf{1}) + i(Q - a\mathbf{1})] = 0.$$

Hence $T_g\psi = T_g[P + b\mathbf{1} + i(Q - a\mathbf{1})]\psi' = 0$; so T_g is a projection onto the ground state of the harmonic oscillator. Later we shall see that excited states of the harmonic oscillator, and in fact arbitrary states, may be expanded as linear combinations of ground states of harmonic oscillator systems parametrized by a,b above. The condition $c = 1$ may be relaxed to $c > 0$ by considering the canonical pair P', Q':

$$P' = c^{-1/2}[P + b\mathbf{1}], \quad Q' = c^{1/2}Q - c^{-1/2}a\mathbf{1}. \quad (3'')$$

For the remainder of this paper we shall lose no generality in assuming $\langle \mathbf{1}\rangle \neq 0$ wherever it appears. Now, to make $T_g\,\rho T_g$ a state, we need to normalize it; that is, $[\mathrm{Tr}(T_g\,\rho T_g)]^{-1}T_g\,\rho T_g = \langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g$ is a state. In this state, the expected value $\bar{P}$ of P is $\bar{P} = \mathrm{Tr}(\langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g P) = \langle P\rangle/\langle \mathbf{1}\rangle = b$ and the expected of Q is $\bar{Q} = \mathrm{Tr}(\langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g Q) = \langle Q\rangle/\langle \mathbf{1}\rangle = a/c$.

Premultiplying (3) by Q, postmultiplying by $\rho\,T_g$, taking the trace, and solving for c yields

$$c = \hbar/2\mathrm{Var}Q,$$

where $\mathrm{Var}Q$ is the expected value of $(Q - \bar{Q}\mathbf{1})^2$ in the state $\langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g$. In this way g is parametrized by $a = c\bar{Q}$, $b = \bar{P}$, and $c = \hbar(2\mathrm{Var}Q)^{-1}$. Furthermore, having chosen and fixed these parameters, then the expectations of Q, P, and $(Q - \bar{Q}\mathbf{1})^{-2}$ in the state $\langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g$ are independent of ρ, so that the map

$$\rho \to \langle \mathbf{1}\rangle^{-1}T_g\,\rho T_g$$

projects out the part of ρ having $\bar{Q}$, $\bar{P}$ as the respective averages of $\bar{Q}$, $\bar{P}$ with fixed dispersion of Q.

If we postmultiply (3) by P and take traces, we recover a second expression for c, namely

$$c = [\hbar \pm (\hbar^2 - 4\mathrm{Var}P\,\mathrm{Var}Q)^{1/2}]/2\mathrm{Var}Q,$$

which agrees with the previous result iff

$$\mathrm{Var}P\,\mathrm{Var}Q = \hbar^2/4$$

that is, the state $T_g\,\rho T_g$ satisfies the minimum uncertainty relation. We shall henceforth refer to T_g as a projection onto a minimum uncertainty state, and to g as a minimum uncertainty function. One may reverse this logic: it is proven in Ref. 3 that any projection T onto a minimum uncertainty state satisfies (3), from which one may show that T is of the form T_g as in (1) with g as given in (2).

Subsequently, we shall treat c (i.e., $\mathrm{Var}Q$ and $\mathrm{Var}P$) as a nonvarying parameter, and shall abbreviate

$$T_{g(\cdot,\cdot;a,b,c)} \quad \text{by} \quad T_{\bar{Q},\bar{P}}.$$

From (*) one may directly calculate the following:

$$T_{\bar{Q},\bar{P}}T_{\bar{Q}',\bar{P}'} = \exp[-i(\bar{Q}'\bar{P} - \bar{Q}\bar{P}')/\hbar] \times \exp\{-(c/4\hbar)[(\bar{Q} - \bar{Q}')^2 + (\bar{P} - \bar{P}')^2/c^2]\} \times W(\hbar^{-1}(\bar{Q}' - \bar{Q}), \hbar^{-1}(\bar{P} - \bar{P}'))T_{\bar{Q}',\bar{P}'}; \quad (4)$$

$$T_{\bar{Q},\bar{P}}W(a,b)T_{\bar{Q},\bar{P}} = \exp[i(a\bar{P} + b\bar{Q}) - \hbar c4^{-1}(a^2 + b^2/c^2)]\,T_{\bar{Q},\bar{P}}; \quad (5)$$

$$T_{\bar{Q},\bar{P}}T_{\bar{Q}',\bar{P}'}T_{\bar{Q},\bar{P}} = \exp\{-(c/2\hbar)[(\bar{Q} - \bar{Q}')^2 + (\bar{P} - \bar{P}')^2/c^2]\}T_{\bar{Q},\bar{P}}; \quad (6)$$

$$\int T_{\bar{Q},\bar{P}}\,d\bar{Q}\,d\bar{P} = 2\pi\hbar\mathbf{1}. \quad (7)$$

We now proceed in the spirit of Ref. 4 to show the sense in which $T_{\bar{Q},\bar{P}}$ projects out irreducible representations of the Weyl algebra. Let $\mathscr{H}_{\bar{Q},\bar{P}} \equiv T_{\bar{Q},\bar{P}}\mathscr{H}$. Let $\{e_\alpha\}$ be an orthonormal basis for $\mathscr{H}_{\bar{Q},\bar{P}}$, and let $\mathscr{H}_\alpha$ be the closed linear subspace spanned by $\{W(a,b)e_\alpha \mid a,b\in\mathbb{R}\}$. Then with the change (3'') we may use Ref. 4 to obtain $\mathscr{H} = \oplus_\alpha \mathscr{H}_\alpha$. Since no nontrivial subspace of $\mathscr{H}_\alpha$ is $\mathscr{W}$-invariant, each $\mathscr{H}_\alpha$ hosts an irreducible representation of $\mathscr{W}$. If we let E_α denote the projection onto $\mathscr{H}_\alpha$, then the irreducible representation is just $\{W(a,b)E_\alpha\}$. In Ref. 4 it is shown that any two representations of $\mathscr{W}$ in an $\mathscr{H}_\alpha$ are unitarily equivalent, giving the celebrated von Neumann theorem on representations of a Weyl system. As in particle physics, we will identify a vector in $\mathscr{H}_\alpha$ (any fixed α) as an elementary Weyl particle. Then, in view of (4), $W(\hbar^{-1}(\bar{Q}' - \bar{Q}), \hbar^{-1}(\bar{P} - \bar{P}'))$ converts a particle at fuzzy point $(\bar{Q}',\bar{P}')$ to one at point $(\bar{Q},\bar{P})$. Operating in $\mathscr{H} = \oplus_\alpha \mathscr{H}_\alpha$, $T_{\bar{Q}',\bar{P}}$ projects out a mixture of elementary Weyl particles at fuzzy point $(\bar{Q}',\bar{P}')$ and $W(\hbar^{-1}(\bar{Q}' - \bar{Q}), \hbar^{-1}(\bar{P} - \bar{P}'))$ moves the mixture to fuzzy point $(\bar{Q},\bar{P})$. Henceforth, we will omit the terms "mixture of ...(s)" as if we were working in an irreducible representation space for $\mathscr{W}$.

Another consequence of (4) is that we may now quantize phase space: Suppose that the $(\bar{Q},\bar{P})$ values are restricted to a lattice, $\mathscr{L}$. Then the subalgebra $\mathscr{W}_{\mathscr{L}} = \{W(\hbar^{-1}(\bar{Q},\bar{P}))|(\bar{Q},\bar{P})\in\mathscr{L}\}$ will shift elementary Weyl particles from one lattice point to another.

We conclude this section with two remarks concerning the physical interpretation of this formalism. We caution the reader that we have singled out the position and momentum as the observables of prime importance. The "particles" we so obtain are not endowed with any other features. If instead we were to focus on the electromagnetic properties, we would obtain analogs of the minimum uncertainty states discussed by Marburger and Power.[5] If we insist on considering only momentum and position and (semiclassically) exclude regions of available phase space by means of barriers, etc. (thereby changing the spectra of P,Q), we again have to derive our new set of minimum uncertainty states. In this sense, we are considering here the "free" minimum uncertainty states. For the method for treating the general case, see Ref. 3.

The question arises as to the treatment of (linear combinations of) excited states of the harmonic oscillator. From the discussion following (3'), they are not considered as "elementary particles" here. In view of (7), the wavefunction for such a state (or any state, for that matter) may be written as an integral over fuzzy phase space of "elementary particles" centered at other points in fuzzy phase space. Thus excited states, temperature states, etc. are easy to handle in this formalism, but are not endowed with the label "elementary."

III. CONDITIONING OF STATES; INSTRUMENTS, EXPECTATIONS, AND MEASUREMENTS

Let B be a bounded operator on $\mathscr{H}$, E a Borel subset of $\mathbb{R}^2$ and define

$$\mathscr{E}(E,B)\equiv\int_E T_{\bar{Q},\bar{P}}BT_{\bar{Q},\bar{P}}\,d\bar{Q}\,d\bar{P}. \tag{8}$$

It is shown in Ref. 3 that $\mathscr{E}$ then satisfies the conditions
(i) $B\geqslant 0\Rightarrow\mathscr{E}(E,B)\geqslant 0$ for all E,
(ii) $B\to\mathscr{E}(E,B)$ is linear for all E,
(iii) $E\to\mathscr{E}(E,B)$ is σ-additive for all B,
(iv) $\mathscr{E}$ is normal,
(v) $\mathscr{E}(\mathbb{R}^2,1)=2\pi\hbar\mathbb{1}$. (9)

Thus $\mathscr{E}$ is an "instrument" and an "expectation" in the notation of Refs. 6 and 7 and a "measurement" in the notation of Ref. 8. We obtain similar results if μ is any measure on $\mathbb{R}^2$ and we define

$$\mathscr{E}(\mu,B)+\int T_{\bar{Q},\bar{P}}BT_{\bar{Q},\bar{P}}\,d\mu(\bar{Q},\bar{P}). \tag{10}$$

Notice also that if ρ is any density operator and B is bounded,

$$\mathrm{Tr}[\mathscr{E}(\mu,\rho)B]=\mathrm{Tr}[\rho\mathscr{E}(\mu,B)]; \tag{11}$$

so that we need not distinguish between "instruments" and "expectations."

In view of the interpretation of $T_{\bar{Q},\bar{P}}$ as a projection onto the state of an elementary particle with coordinates $(\bar{Q},\bar{P})$ in fuzzy phase space, $\mathrm{Tr}(T_{\bar{Q},\bar{P}}\rho T_{\bar{Q},\bar{P}})$ is the probability that ρ is in fuzzy state $(\bar{Q},\bar{P})$ and $\mathrm{Tr}\mathscr{E}(\mu,\rho)$ the probability that ρ consists of elementary particles distributed according to the classical distribution μ on fuzzy phase space; $\mathscr{E}(\mu,\rho)$ represents the conditioning of ρ onto the (unnormalized) state consisting of a classical distribution μ of elementary particles.

To say that ρ is a known state is equivalent to saying that $\mathrm{Tr}[\rho W(a,b)]$ is known for all $(a,b)\in\mathbb{R}^2$. Thus, one may compute, using (11) and (5),

$$\begin{aligned}
&\mathrm{Tr}[\mathscr{E}(\mu,\rho)W(a,b)]\\
&=\mathrm{Tr}\Big[\rho\int T_{\bar{Q},\bar{P}}W(a,b)T_{\bar{Q},\bar{P}}\,d\mu(\bar{Q},\bar{P})\Big]\\
&=\int\exp[i(a\bar{P}+b\bar{Q})-\hbar c4^{-1}(a^2+b^2/c^2)]\\
&\quad\times\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}})\,d\mu(\bar{Q},\bar{P})\\
&=\int\exp[i(a\bar{P}+b\bar{Q})-\hbar c4^{-1}(a^2+b^2/c^2)]g(x,y;c\bar{Q},\bar{P},c)\\
&\quad\times Tr[\rho W(x,y)]\,dxdyd\mu(\bar{Q},\bar{P})\\
&=\int\exp[i(a-x)\bar{P}+i(b-y)\bar{Q}]\,d\mu(\bar{Q},\bar{P})\,g(x,y;0,0,c)\\
&\quad\times\mathrm{Tr}[\rho W(x,y)]\,dxdy\exp[-\hbar c4^{-1}(a^2+b^2/c^2)].
\end{aligned}$$

In particular, if the measure is linear in $\bar{Q}$, the Fourier transformation of μ yields a delta function in $b-y$, and similarly for $\bar{P}$. Consequently, one of the x and y integrals becomes trivial and a simpler Gaussian integral remains. The conditioning operators for Weyl systems found in Refs. 7, 8 are of this special category.

A state ρ is called a "coherent state" if there is $(\bar{Q},\bar{P})$ such that $T_{\bar{Q},\bar{P}}\rho T_{\bar{Q},\bar{P}}=\rho$. In this case, $\rho=\mathscr{E}(\mu,\rho)$, where μ is a probability measure concentrated on $(\bar{Q},\bar{P})$.

IV. THE UNCERTAINTY RELATIONS ON CONDITIONED STATES

For μ a probability measure, we have

$$\begin{aligned}
\mathrm{Tr}[\mathscr{E}(\mu,\rho)P]&=\mathrm{Tr}\Big[\mathscr{E}(\mu,\rho)\Big(-i\frac{\partial}{\partial x}W(x,y)\Big)\Big]\Big|_{x=y=0}\\
&=-i\frac{\partial}{\partial x}\mathrm{Tr}[\mathscr{E}(\mu,\rho)W(x,y)]|_{x=y=0}\\
&=-i\frac{\partial}{\partial x}\mathrm{Tr}[\rho\mathscr{E}(\mu,W(x,y))]|_{x=y=0}\\
&=-i\frac{\partial}{\partial x}\int d\mu(\bar{Q},\bar{P})\exp[i(x\bar{P}+y\bar{Q})]\\
&\quad\times\exp[-\hbar c4^{-1}(x^2+y^2/c^2)]\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}})|_{x=y=0};
\end{aligned}$$

so

$$\mathrm{Tr}[\mathscr{E}(\mu,\rho)P]=\int d\mu(\bar{Q},\bar{P})\bar{P}\,\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}}) \tag{12}$$

and

$$\begin{aligned}
\mathrm{Tr}[\mathscr{E}(\mu,\rho)P^2]&=-\frac{\partial^2}{\partial x^2}\mathrm{Tr}[\rho\mathscr{E}(\mu,W(x,y))]|_{x=y=0}\\
&=\int d\mu(\bar{Q},\bar{P})\,(\bar{P}^2+\hbar c/2)\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}}).
\end{aligned} \tag{13}$$

Normalizing the state $\mathscr{E}(\mu,\rho)$ by division by

$$\mathrm{Tr}[\mathscr{E}(\mu,\rho)\mathbb{1}] = \int d\mu(\bar{Q},\bar{P})\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}}) \tag{14}$$

and using (12) and (13) allows one to calculate the expected value and variance of P (and similarly for Q) in any conditioned state. In particular, if μ is a point measure concentrated on $(\bar{Q}_0,\bar{P}_0)$, then the expected value of P is

$$\bar{P}_0\mathrm{Tr}(\rho T_{\bar{Q}_0,\bar{P}_0}) \div \mathrm{Tr}(\rho T_{\bar{Q}_0,\bar{P}_0}) = \bar{P}_0.$$

the variance of P is $\hbar c/2$, the expected value of Q is $\bar{Q}_0$, and the variance of Q is $\hbar/(2c)$, in agreement with the previous discussion of c in Sec. II and reconfirming the fact that $T_{\bar{Q}_0,\bar{P}_0}\rho T_{\bar{Q}_0,\bar{P}_0}$ is an (unnormalized) state of minimum uncertainty. In other terminology, an elementary Weyl particle is a state of minimum uncertainty. In the language of measurement theory, even the measurement of an elementary Weyl particle at a precise point in (fuzzy) phase space statistically obeys the uncertainty principle.

We now show that if μ is concentrated at more than a single point, the uncertainty principle is satisfied with a strict inequality. For this we shall generalize a result of Ref. 9 from the case where μ has only jump discontinuities.

Let $z = (\bar{Q},\bar{P}), \rho = \int \rho(z)\, d\mu(z), \rho(z) = T_z\, \rho_0 T_z, X$ a self-adjoint linear operator on $\mathscr{H}$, $\{E_t\}$ its spectral family,

$$m_x^z = \left[\int t\, d(\mathrm{Tr}\,\rho(z)E_t)\right]/\mathrm{Tr}\,\rho(z),$$

$$\mathrm{Var}X = \int (t - m_x)^2 d(\mathrm{Tr}\,\rho E_t)/\mathrm{Tr}\,\rho,$$

$$\mathrm{Var}_x^z = \int (t - m_x^z)^2 d(\mathrm{Tr}\,\rho(z)E_t)/\mathrm{Tr}\,\rho(z).$$

Then $m_x = \int d\mu(z) m_x^z \mathrm{Tr}\,\rho(z)/\mathrm{Tr}\,\rho$, and letting $d\mu'(z) = d\mu(z)\,\mathrm{Tr}\,\rho(z)/\mathrm{Tr}\,\rho$,

$$\begin{aligned}
\mathrm{Var}X &= \int t^2 d(\mathrm{Tr}(\rho E_t)/\mathrm{Tr}\,\rho) - {m_x}^2 \\
&= \int d\mu(z)\, t^2 d(\mathrm{Tr}[\rho(z)E_t]/\mathrm{Tr}\,\rho) \\
&\quad - \int d\mu'(z)\, d\mu'(z') m_x^z m_x^{z'} \\
&= \int d\mu(z)\, (t^2 - m_x^{z^2}) d(\mathrm{Tr}[\rho(z)E_t]/\mathrm{Tr}\,\rho) \\
&\quad + \tfrac{1}{2}\int d\mu'(z)\, d\mu'(z')\, (m_x^z - m_x^{z'})^2 \\
&= \int d\mu(z)\, \mathrm{Tr}\,\rho(z)\mathrm{Var}_x^z/\mathrm{Tr}\,\rho \\
&\quad + \tfrac{1}{2}\int d\mu'(z)\, d\mu'(z')\, (m_x^z - m_x^{z'})^2.
\end{aligned} \tag{14'}$$

Thus if μ is concentrated at more than a single point, $\mathrm{Var}X > \int d\mu(z)\,\mathrm{Tr}\,\rho(z)\mathrm{Var}_x^z/\mathrm{Tr}\,\rho = \int d\mu'(z)\,\mathrm{Var}_x^z$.

In particular,

$$\begin{aligned}
\mathrm{Var}P\,\mathrm{Var}Q &> \int d\mu'(z)\mathrm{Var}_P^z d\mu'(z')\mathrm{Var}_Q^{z'} \\
&\geqslant \left|\int d\mu'(z)[\mathrm{Var}_P^z\ \mathrm{Var}_Q^z]^{1/2}\right|^2 \\
&= \left|\int d\mu'(z)\hbar/2\right|^2 \\
&= (\hbar/2)^2.
\end{aligned}$$

V. THE OBSERVABLES CORRESPONDING TO THE INSTRUMENTS; A JOINT SPECTRAL FAMILY FOR P AND Q

The observable $A(\mu)$ corresponding to the instrument $\mathscr{E}(\mu,\cdot)$ is defined in Ref. 6 by

$$\mathrm{Tr}[\mathscr{E}(\mu,\rho)\mathbb{1}] = \mathrm{Tr}[\rho A(\mu)].$$

Hence we have

$$A(\mu) = \mathscr{E}(\mu,\mathbb{1}) = \int d\mu(\bar{Q},\bar{P}) T_{\bar{Q},\bar{P}}. \tag{15}$$

$A(\mu)$ may therefore be interpreted as a classical distribution of free elementary Weyl particles in fuzzy phase space with (probability) measure μ. These are the physically meaningful observables in our simple model.

Now $\mathrm{Tr}[\mathscr{E}(\mu,\rho)A(\sigma)] \div \mathrm{Tr}[\mathscr{E}(\mu,\rho)]$ represents the probability of finding elementary particles distributed according to probability measure σ in a state ρ which has been conditioned with an instrument composed of elementary particles distributed according to probability measure μ. We analyze this quantity:

$$\begin{aligned}
&\mathrm{Tr}[\mathscr{E}(\mu,\rho)A(\sigma)] \\
&\quad = \mathrm{Tr}[\rho\mathscr{E}(\mu,A(\sigma))] \\
&\quad = \int d\mu(Q,P) d\sigma(\bar{Q}',\bar{P}')\,\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}} T_{\bar{Q}',\bar{P}'} T_{\bar{Q},\bar{P}}) \\
&\quad = \int d\mu(\bar{Q},\bar{P}) d\sigma(\bar{Q}',\bar{P}') \\
&\quad\quad \times \exp\{-(c/2\hbar)[(\bar{Q}-\bar{Q}')^2 + (\bar{P}-\bar{P}')^2/c^2]\}\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}}).
\end{aligned}$$

Thus, even if μ and σ have disjoint supports in fuzzy phase space, $\mathrm{Tr}[\mathscr{E}(\mu,\rho)A(\sigma)]$ is nonvanishing, although the expression decreases exponentially in the square of the Euclidean distance between the supports.

Introducing the function E by

$$E(x,y) = \exp[-c/2\hbar(x^2 + y^2/c^2)], \tag{16}$$

we then have

$$\mathrm{Tr}[\mathscr{E}(\mu,\rho)A(\sigma)] = \mathrm{Tr}[\rho A(\mu\cdot\sigma * E)],$$

where $*$ denotes convolution. More generally we have

$$\mathscr{E}(\mu,A(\sigma)) = A(\mu\cdot\sigma * E). \tag{17}$$

The restrictions of "the physically meaningful observables" to operators of the form $A(\mu)$ is in fact not much of a restriction. We shall show that $\mathscr{A} \equiv \{A(\mu)\}$ includes most functions of P and Q and in fact provides what we shall label "the joint spectral family for P and Q."

In Ref. 3 it is shown that

(i) $\int T_{\bar{Q},\bar{P}}\, d\bar{Q}$ depends only on P,
(ii) $\int T_{\bar{Q},\bar{P}}\, d\bar{P}$ depends only on Q,
(iii) $\int T_{\bar{Q},\bar{P}}\, d\bar{Q}d\bar{P} = 2\pi\hbar\mathbb{1}$. (7)

Integrating (3) yields, in fact,

(iv) $\int \bar{P} T_{\bar{Q},\bar{P}}\, d\bar{Q}d\bar{P} = 2\pi\hbar P$, (18)

(v) $\int \bar{Q} T_{\bar{Q},\bar{P}}\, d\bar{Q}d\bar{P} = 2\pi\hbar Q$. (19)

More generally, if we take, for Δ a Borel set in $\mathbb{R}$, $d\mu_{\mathbb{R}\times\Delta}(\bar{Q},\bar{P}) = \chi_{\mathbb{R}\times\Delta}(\bar{Q},\bar{P})d\bar{Q}d\bar{P}$, $\chi_{\mathbb{R}\times\Delta}$ the characteristic function of $\mathbb{R}\times\Delta$, we obtain, by direct computation,

$$A(\mu_{\mathbb{R}\times\Delta}) = 2(\pi\hbar/c)^{1/2}\int_{\Delta} d\bar{P} \times \int_{\mathbb{R}} dE_t^P \exp[-(\hbar c)^{-1}(t+\bar{P})^2] \tag{20}$$

and, similarly,

$$A(\mu_{\Delta\times\mathbb{R}}) = 2(\pi\hbar c)^{1/2}\int_{\Delta} d\bar{Q}\int_{\mathbb{R}} dE_x^Q \exp[-c\hbar^{-1}(\bar{Q}+x)^2], \tag{21}$$

where $\{E_t^P\}$, $\{E_x^Q\}$ are the spectral families for P, Q, respectively. Since $\int z^n \exp[-\gamma(z+w)^2]\,dz$ is a polynomial in w of order n, from measures

$$d\mu(\bar{Q},\bar{P}) = \bar{Q}^n d\bar{Q}\,d\bar{P} \quad \text{or} \quad \bar{P}^n\,d\bar{Q}d\bar{P}$$

all polynomials in Q or P may be constructed from the $A(\mu)$, and more generally, all measurable functions. The particular expressions (19) and (20) are identical with the corresponding expressions found in Ref. 8 so that our $\{A(\mu)\}$ generalizes the results in Refs. 8 and 15.

In view of (7), (18), (19), (20), (21), and the above discussion, we may consider the set $\{T_{\bar{Q},\bar{P}}\}$ as the joint spectral family for P and Q; however, we recognize from (4) that $(\bar{Q},\bar{P})\neq(\bar{Q}',\bar{P}')$ does not imply $T_{\bar{Q},\bar{P}}T_{\bar{Q}',\bar{P}'} = 0$.

VI. SEQUENTIAL CONDITIONING; THE COLLAPSE SCHEME

Let ρ be a state and let μ be a positive measure such that $\mathrm{Tr}[\mathscr{E}(\mu,\rho)] < \infty$. Then $\{\mathrm{Tr}[\mathscr{E}(\mu,\rho)]\}^{-1}\mathscr{E}(\mu,\rho)$ defines a new state. For σ another measure such that $\mathscr{E}(\sigma,\mathscr{E}(\mu,\rho))$ is trace-class, we then have from (4)

$$\begin{aligned}
&\mathscr{E}(\sigma,\mathscr{E}(\mu,\rho))\\
&\quad = \int d\sigma(\bar{Q}',\bar{P}')d\mu(\bar{Q},\bar{P})\,T_{\bar{Q}',\bar{P}'}T_{\bar{Q},\bar{P}}\rho T_{\bar{Q},\bar{P}}T_{\bar{Q}',\bar{P}'}\\
&\quad = \int d\sigma(\bar{Q}',\bar{P}')d\mu(\bar{Q},\bar{P})\\
&\qquad \times\exp\{-(c/2\hbar)[(\bar{Q}-\bar{Q}')^2+(\bar{P}-\bar{P}')^2/c^2]\}\\
&\qquad \times W(\hbar^{-1}(\bar{Q}-\bar{Q}'),\hbar^{-1}(\bar{P}'-\bar{P}))T_{\bar{Q},\bar{P}}\rho T_{\bar{Q},\bar{P}}\\
&\qquad \times W(h^{-1}(\bar{Q}-\bar{Q}'),h^{-1}(\bar{P}'-\bar{P}))^{\dagger}.
\end{aligned}$$

By using (16) and (17), the corresponding normalized state is therefore

$$\begin{aligned}
&\{\mathrm{Tr}[\rho A(\mu\cdot\sigma * E)]\}^{-1}\\
&\times\int d\sigma(\bar{Q}'\bar{P}')d\mu(\bar{Q},\bar{P})\,E(\bar{Q}-\bar{Q}',\bar{P}-\bar{P}')\\
&\qquad \times W(\hbar^{-1}(\bar{Q}-\bar{Q}'),\hbar^{-1}(\bar{P}'-\bar{P}))T_{\bar{Q},\bar{P}}\rho T_{\bar{Q},\bar{P}}\\
&\qquad \times W(\hbar^{-1}(\bar{Q}-\bar{Q}'),\hbar^{-1}(\bar{P}-\bar{P}'))^{-1}.
\end{aligned} \tag{22}$$

In the case $\mu=\sigma$ this does not reduce to the normalized $\mathscr{E}(\mu,\rho)$ unless μ is concentrated at a single point. That may be interpreted physically by saying that the only measurement that gives the same result "immediately after" the first attempt at the same measurement is the physically unattainable measurement of locating precisely a single elementary Weyl particle, the optimally precise experiment. For any realizeable experiment this repeatability axiom fails; i.e., in this sense, the von Neumann collapse scheme of measurement is violated. For further discussion of this point, see Refs. 6 and 8. The philosophical difficulty of the noncontinuous nature of the change in the state due to measurement and whether it occurs at the "moment the result of the measurement enters the consciousness of the observer" is partially rectified by the present measurement scheme. If the measurement is introduced when the state is centered "far" from the apparatus, the exponential tails of the interaction discussed in Sec. V will produce a continuous change as the state approaches the apparatus with time. Even the assembly of the apparatus, if done in a continuous manner, will yield a continuous change in state. If, however, we simply "turn on" the measurement by an instantaneous application of $\mathscr{E}(\mu,\cdot)$, the philosophical difficulties will remain. We shall soon see that the limit of a sequence of such sudden measurements also leads to mathematical difficulties, in particular a change of representation space.

It is suggested in Ref. 10 that if one works with the von Neumann collapse scheme, so that measurements correspond to projectors, and if R_1, R_2 are two such projectors, then $R_1\wedge R_2 \equiv \lim_{n\to\infty}(R_1R_2)^n$ represents the joint measurement (simultaneous measurement). $R_1\wedge R_2$, where it exists, then has the properties (i) $R_1(R_1\wedge R_2) = R_2(R_1\wedge R_2) = R_1\wedge R_2$, (ii) $R_1\wedge R_1 = R_1$, and (iii) $R_1\wedge R_2 = R_2\wedge R_1$. The first result implies (i') $(R_1\wedge R_2)(R_1\wedge R_2) = R_1\wedge R_2$. We, by analogy, consider the infinite sequence $\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\sigma,\cdot)$ of alternating applications of $\mathscr{E}(\mu,\cdot)$ and $\mathscr{E}(\sigma,\cdot)$ to some state ρ. (i') is then the repeatability of measurements which we expect to fail unless $\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\sigma,\cdot)$ represents an optimally precise measurement, or if $\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\sigma,\cdot)$ cannot be written in the form $\mathscr{E}(\omega,\cdot)$, where ω is some measure determined by μ, σ. In view of (22), even the analogy of (i) seems hopeless. (ii) is integrally tied to the collapse scheme. In terms of a realistic interpretation of simultaneous measurement, only the commutativity principle (iii) $[\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\sigma,\cdot) = \mathscr{E}(\sigma,\cdot)\wedge(\mathscr{E}(\mu,\cdot)]$ seems justified. Since properties (i), (i'), and (ii) are the ones in doubt, we will simplify the computation to the consideration of only $\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\mu,\cdot)$. We shall see that, in at least one case, the result, applied to state ρ, is independent of ρ and cannot be written in the form $\mathscr{E}(\omega,\rho)$. This indicates that, viewed as a Markov process, there is a single ergodic class. The general computation of $\mathscr{E}(\mu,\cdot)\wedge\mathscr{E}(\sigma,\cdot)$ remains unsolved.

Define $[\mathscr{E}(\mu,\cdot)]^n$ by

$$\begin{aligned}
&[\mathscr{E}(\mu,\cdot)]^1 = \mathscr{E}(\mu,\cdot) \quad \text{and}\\
&[\mathscr{E}(\mu,\cdot)]^n(B) = \mathscr{E}(\mu,[\mathscr{E}(\mu,\cdot)]^{n-1}(B)), \quad n>1,\\
&B \text{ bounded.}
\end{aligned} \tag{23}$$

We must compute $[\mathscr{E}(\mu,\cdot)]^n[W(a,b)]$, and since

$$\begin{aligned}
&[\mathscr{E}(\mu,\cdot)]^n[W(a,b)]\\
&\quad = \int d\mu(\bar{Q},\bar{P})\exp[i(a\bar{P}+b\bar{Q}) - \hbar c 4^{-1}(a^2+b^2/c^2)]\\
&\qquad \times[\mathscr{E}(\mu,\cdot)]^{n-1}(T_{\bar{Q},\bar{P}}),
\end{aligned}$$

it suffices to compute $[\mathscr{E}(\mu,\cdot)]^{n-1}(T_{\bar{Q},\bar{P}})$. If we change variables of integration, $x = \bar{Q}c^{1/2}(2\hbar)^{-1/2}$, $y = \bar{P}(2\hbar c)^{-1/2}$, and set $T_{\bar{Q},\bar{P}} = S_{x,y}$, $\mu(\bar{Q},\bar{P}) = \mu'(x,y)$, we then obtain

$$\mathscr{E}(\mu,T_{Q,P}) = \int d\mu'(x',y')\exp[-(x-x')^2-(y-y')^2] \times S_{x',y'}, \tag{24}$$

and more generally

$$[\mathscr{E}(\mu,\cdot)]^n(T_{\bar{Q},\bar{P}}) = \int d\mu'(x_n,y_n)\,\phi_n(x,y;x_n,y_n)S_{x_n,y_n}, \tag{25}$$

where

$$\phi_n(x,y;x_n,y_n) = \int\prod_{i=1}^{n-1} d\mu'(x_i,y_i)\prod_{j=1}^{n}\exp[-(x_{j-1}-x_j)^2-(y_{j-1}-y_j)^2] \tag{26}$$

$$= \int d\mu'(x_{n-1},y_{n-1})\phi_{n-1}(x,y;x_{n-1},y_{n-1}) \times\exp[-(x_n-x_{n-1})^2-(y_n-y_{n-1})^2]. \tag{26'}$$

As an example, if we set $d\mu'(x,y) = \exp[-\beta(x^2+y^2)]\,dxdy, \beta>0$, then we obtain by induction

$$\phi_n(x,y;x_n\,y_n) = a_n \times\exp[-b_n(x^2+y^2)+c_n(xx_n+y\,y_n)-d_n(x_n^2+y_n^2)]$$

where

$$a_n = a_{n-1}\pi/(d_{n-1}+\beta+1),$$
$$b_n = b_{n-1}-(c_{n-1})^2/4(d_{n-1}+\beta+1),$$
$$c_n = c_{n-1}/(d_{n-1}+\beta+1),$$
$$d_n = 1+(d_{n-1}+\beta+1)^{-1}.$$

As $n\to\infty$, $a_n\to 0$, and $c_n\to 0$, b_n converges, and $1\leqslant d_n\leqslant 1+(1+\beta)^{-1}$. From this we obtain

$$\begin{aligned}\mathrm{Tr}\{[\mathscr{E}(\mu,\cdot)]^{n+1}(\rho)W(a,b)\}/\mathrm{Tr}\{[\mathscr{E}(\mu,\cdot)]^{n+1}(\rho)\}&\\ = \exp\{-\hbar c4^{-1}[1+2c(\beta+b_n)^{-1}](a^2+b^2/c^2)\}&\\ \times\int dx_n dy_n\exp\{-[\beta+d_n-c_n^2 4^{-1}(\beta+b_n)^{-1}]&\\ \times(x_n^2+y_n^2)\}&\\ \times\exp\{-i[c_n(\hbar c/2)^{1/2}/(\beta+b_n)](ay_n+bx_n/c)\}&\\ \times\mathrm{Tr}(\rho S_{x_n,y_n})&\\ \times\left(\int dxdy\exp\{-[\beta+d_n-c_n^2 4^{-1}(\beta+b_n)^{-1}]\right.&\\ \left.\times(x^2+y^2)\}\mathrm{Tr}(\rho S_{x,y})\right)^{-1}.&\end{aligned}$$

Since $\{d_n\}$ is a bounded sequence, it has a converging subsequence; along any such subsequence the above has the limit

$$\exp[-\hbar c4^{-1}(1+2c\beta+b_\infty)^{-1}(a^2+b^2/c^2)],$$

where $b_\infty = \lim_{n\to\infty} b_n$. Since these subsequential limits all coincide, the limit exists. It is manifestly independent of ρ.

Now suppose there were a measure ω depending only on μ such that

$$\mathrm{Tr}[\mathscr{E}(\omega,\rho)W(a,b)] = \exp\{-\hbar c4^{-1}[1+2c(\beta+b_\infty)^{-1}](a^2+b^2/c^2)\}. \tag{27}$$

Taking the Fourier transformation of the right-hand side then gives

$$\begin{aligned}\int d\omega(\bar{Q},\bar{P})\,\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}})\exp[i(a\bar{P}+b\bar{Q})]&\\ = \int d\bar{Q}d\bar{P}\frac{\beta+b_\infty}{2\pi\hbar c}\exp\left[\frac{(\beta+b_\infty)}{2\hbar}(\bar{Q}^2+\bar{P}^2/c^2)\right]&\\ \times\exp[i(a\bar{P}+b\bar{Q})].&\end{aligned}$$

Since this is to hold for all $(a,b)\in\mathbb{R}^2$, we may identify $d\omega(\bar{Q},\bar{P})\,\mathrm{Tr}(\rho T_{\bar{Q},\bar{P}})$

$$= \frac{\beta+b_\infty}{2\pi\hbar c}\exp\left[\frac{(\beta+b_\infty)}{2\hbar}((\bar{Q}^2+\bar{P}^2/c^2)\right]d\bar{Q}d\bar{P}$$

so that there is no measure ω which is independent of ρ and which satisfies (27) for all ρ.

This result, of course, depends on the choice $d\mu'(x,y) = \exp[-\beta(x^2+y^2)]\,dxdy$. If instead we take a polynomial times a Gaussian and use the generating function for Hermite polynomials on the exponential terms in (26), similar results for the class of measures absolutely continuous with respect to Lebesgue measure could be obtained.

VII. SIMULTANEOUS CONDITIONING

In combining two pure quantum mechanical states, one distinguishes between coherent superpositions (addition of amplitudes) and incoherent superpositions (addition of probability densities). For two mixtures with density matrices ρ_1 and ρ_2 we could form the convex combination, which in the case where ρ_1 and ρ_2 are pure states would correspond to an incoherent superposition. If we take $\rho_i = \mathscr{E}(\mu_i,\rho)$, then the convex combination is $\lambda\rho_1+(1-\lambda)\rho_2 = \mathscr{E}(\lambda\mu_1+(1-\lambda)\mu_2,\rho)$ so that we may define "the incoherent simultaneous composition of instruments," $\mathscr{E}(\mu_1,\cdot)$ and $\mathscr{E}(\mu_2,\cdot)$, by $\mathscr{E}(\mu_1+\mu_2,\cdot)$ which corresponds to choosing $\lambda=\frac{1}{2}$ and ignoring the fact that the ρ_i need not be normalized in general. If ρ_1 and ρ_2 were normalized, the factor $\lambda=\frac{1}{2}$ would be recovered upon normalizing $\mathscr{E}(\mu_1+\mu_2,\rho)$. In other notation, $\mathscr{E}(\mu_1+\mu_2,\rho)$ represents "either $\mathscr{E}(\mu_1,\rho)$ or $\mathscr{E}(\mu_2,\rho)$."

We now search for a composition of instruments which describes conditioning with "both $\mathscr{E}(\mu_1,\cdot)$ and $\mathscr{E}(\mu_2,\cdot)$" and in some sense is a "coherent composition of the instruments." A further consideration in terms of elementary Weyl particles is useful in determining axioms for a coherent composition of instruments. Recall that $\mathrm{Tr}[\mathscr{E}(\mu,\rho)] = \mathrm{Tr}[\rho A(\mu)]$ gives the probability of observing Weyl particles in probability distribution μ on state ρ. Thus $\mathrm{Tr}[\frac{1}{2}\mathscr{E}(\mu_1+\mu_2,\rho)]$ gives the probability of measuring particles distributed in either probability distribution μ_1 or μ_2 (incoherent superposition). By comparison, if ρ describes a single particle, μ_1 a detector in region Δ_1 of phase space and μ_2 a detector in region Δ_2 of phase space, then we search for an instrument which describes the simultaneous measurement of the particle in both regions Δ_1 and Δ_2 (coherent superposition). Because of the inherent fuzziness of the elementary Weyl particles, the probability of such simultaneous detection should (a) be nonzero even if μ_1 and μ_2 are carried on disjoint sets Δ_1 and Δ_2, (b) tend to zero as the distance between Δ_1 and Δ_2 tends to infinity, and (c) be inde-

pendent of the order in which μ_1 and μ_2 are listed. If such a method of composition of instruments could be found, then it perhaps should augment the concept of "addition of amplitudes." We also remark here that in (a) and (b) above we are only concerned with the detection of some influence of the particle. If "it" were absorbed simultaneously at both detectors, then we might have to interpret "it" as "they" in order to avoid contradicting some conservation laws. However, contiguous cloud chambers might detect the "same" particle coincidentally, for example.

Let $\mathscr{E}(\mu_1 \circ \mu_2, \cdot)$ represent the simultaneous coherent superposition (composition) of instruments $\mathscr{E}(\mu_1, \cdot)$ and $\mathscr{E}(\mu_2, \cdot)$. Then we require of $\circ$ that

(a) $\mu_1 \circ \mu_2 = \mu_2 \circ \mu_1$; (28)

(b) $\mu_1 \circ \mu_2 \neq 0$ (even if support $\mu_1 \cap$ support $\mu_2 = \phi$); (29)

(c) If $\{\mu_t\}$ is a family of probability measures such that the distance of the support μ_t from the origin goes to infinity as $t \to \infty$, then $\mu_t \circ \mu_0 \to 0$ as $t \to \infty$. (30)

If μ_1, μ_2 are absolutely continuous with respect to Lebesque measure with Radon–Nikodym derivatives (densities) μ_1', μ_2', then for $\mu_1 \circ \mu_2$ one could consider the measure with density $\mu_1' \mu_2'$, or $\min(\mu_1', \mu_2')$, or $\max(\mu_1', \mu_2')$, or the convolution of μ_1', and μ_2'. All satisfy (28); $\max(\mu_1', \mu_2')$ satisfies (29), and $\mu_1' \mu_2'$ and $\min(\mu_1', \mu_2')$ satisfy (30). Thus, none will satisfy all the requirements of $\circ$. We present a less obvious candidate and show that it is "natural" in some sense. So define

$$d(\mu_1 \circ \mu_2)(\overline{Q}, \overline{P}) = d\overline{Q} d\overline{P} \int dx dy\, \mu_1'(\overline{Q} + x, \overline{P} + y) \times \exp[-(2c/\hbar)(x^2 + y^2/c^2)] \mu_2'(\overline{Q} - x, \overline{P} - y). \quad (31)$$

Then (31) satisfies (28)–(30).

Recalling the notation (1), suppose T_f and T_h are positive operators. We search for a commutative bilinear operation $\bullet$ such that $T_f \bullet T_h = T_{f \otimes h}$ is positive, where $\otimes$ is another commutative bilinear operation which is distinct from the noncandidates for $\circ$ above. From the linearity of T_f in f and the bilinearity of $\bullet$ we expect

$$T_f \bullet T_h = \int dx dx' f(x, y) h(x', y') K(x, x'; y, y') W(x, y) W(x', y'),$$

where K is some complex-valued kernel. For $\bullet$ to preserve self-adjointness, K must satisfy

$$K(x, x'; y, y') = K(-x, -x', -y, -y') \times \exp\{i\hbar(x'y - x y')\}.$$

Thus for

$$L(x, x' y, y') = K(x, x', y, y') \exp[-(i\hbar/2)(x'y - x y')],$$
$$L(x, x, t, y') = L(-x, -x'; -y, -y')$$

and

$$T_f \bullet T_h = \int dx d y dx' dy' f(x, y) h(x', y') L(x, x', y, y') \times W(x + x', y + y').$$

In other words, preservation of self-adjointness requires the dropping of the phase in (*); we are dealing with ray representations of $\mathscr{W}$.

Commutativity of $\bullet$ then requires $L(x, x', y, y') = L(x', x, y', y)$. The choice $L(x, x', y, y') = l(x + x', y + y')$ yields

$$T_f \bullet T_h = \int dx d y dx' d y' f(x, y) h(x', y') l(x + x', y + y') = T_{f \otimes h},$$

where

$$(f \otimes h)(s, t) = l(s, t) f * h(s, t), \quad * \text{ the usual convolution.}$$

Preservation of positivity is more complicated; however, we shall be able to preserve positivity of the physically meaningful positive operators. To start with, let f and h be minimum uncertainty functions (2) at fuzzy phase space points $(\overline{Q}, \overline{P})$ and $(\overline{Q}', \overline{P}')$, respectively. Then

$$(f * h)(u, v) = (\hbar/2\pi) \exp\{-(\hbar c/8)[u^2 + v^2/c^2]\} \times \exp\{-i[\tfrac{1}{2}(\overline{P} + \overline{P}')u + \tfrac{1}{2}(\overline{Q} + \overline{Q}')v]\} \times \exp\{-(c/2\hbar)[(Q - Q')^2 + (P - P')^2/c^2]\}.$$

If we choose $l(u, v) = \exp\{-(\hbar c/4)[\sigma u^2 + \gamma v^2/c^2]\}$, where $(\sigma + \tfrac{1}{2})(\gamma + \tfrac{1}{2}) = 1$, we then obtain, using (16),

$$(f \otimes h)(s, t) = g(s, t, (\sigma + \tfrac{1}{2})c[(\overline{Q} + \overline{Q}')/2], \times [(\overline{P} + \overline{P}')/2], (\sigma + \tfrac{1}{2})c) \times E(\overline{Q} - \overline{Q}', \overline{P} - \overline{P}')$$

and $T_{\overline{Q}, \overline{P}} \bullet T_{\overline{Q}', \overline{P}'} = E(\overline{Q} - \overline{Q}', \overline{P} - \overline{P}') T^{\sigma}_{(\overline{Q} + \overline{Q}')/2, (\overline{P} + \overline{P}')/2}$, where the superscript σ denotes that instead of c we use $(\sigma + \tfrac{1}{2})c$ in the minimum uncertainty function. Positivity is clearly preserved. Since we are taking c as a fixed, fundamental constant inherent to an elementary Weyl particle, we choose $\sigma = \tfrac{1}{2}$ in $l(u, v)$ to obtain

$$T_{\overline{Q}, \overline{P}} \bullet T_{\overline{Q}', \overline{P}'} \equiv E(\overline{Q} - \overline{Q}', \overline{P} - \overline{P}') T_{(\overline{Q} + \overline{Q}')/2, (\overline{P} + \overline{P}')/2}. \quad (32)$$

Then

$$A(\mu_1) \bullet A(\mu_2) \equiv A(\mu_1 \circ \mu_2) = \int d\mu_1(\overline{Q}, \overline{P}) d\mu_2(\overline{Q}', \overline{P}') E(\overline{Q} - \overline{Q}', \overline{P} - \overline{P}') \times T_{(\overline{Q} + \overline{Q}')/2, (\overline{P} + \overline{P}')/2} \quad (33)$$

$$= \int d\mu_1(x + s, y + t) d\mu_2(x - s, y - t) E(2s, 2t) T_{x, y}, \quad (33')$$

of which (31) is a special case.

Since $\mathrm{Tr}[\mathscr{E}(\mu_1, \cdot) \circ \mathscr{E}(\mu_2, \cdot)](\rho) = \mathrm{Tr}[\rho A(\mu_1 \circ \mu_2)]$ for all ρ, we then have

$$\mathscr{E}(\mu_1, \cdot) \circ \mathscr{E}(\mu_2, \cdot) \equiv \mathscr{E}(\mu_1 \circ \mu_2, \cdot), \quad (34)$$

where

$$d(\mu_1 \circ \mu_2)(x, y) = \int_{s, t} d\mu_1(x + s, y + t) d\mu_2(x - s, y - t) \times E(2s, 2t) \quad (35)$$

as in the above sense.

There is, unfortunately, a second candidate for $\circ$, call it

$\circ'$, given by

$$[\mathscr{E}(\mu_1,\cdot)\circ'\mathscr{E}(\mu_2,\cdot)](\rho)$$

$$\equiv\int d\mu_1(\bar{Q},\bar{P})d\mu_2(\bar{Q}',\bar{P}')\,T_{\bar{Q}',\bar{P}'}\bullet T_{\bar{Q},\bar{P}}\,\rho T_{\bar{Q},\bar{P}}\bullet T_{\bar{Q}',\bar{P}}$$

$$=\int d\mu_1(\bar{Q},\bar{P})d\mu_2(\bar{Q}',\bar{P}')\,E^2(\bar{Q}-\bar{Q}',\bar{P}-\bar{P}')$$

$$\times T_{(\bar{Q}+\bar{Q}')/2,(\bar{P}+\bar{P}')/2}\,\rho T_{(\bar{Q}+\bar{Q}')/2,(\bar{P}+\bar{P}')/2} \tag{36}$$

$$=\int d\mu_1(x+s,y+t)d\mu_2(x-s,t-y)\,E^2(2s,2t)T_{x,y}\,\rho T_{x,y}. \tag{36'}$$

(33') and (35) are similarly altered by using E^2 instead of E, to give definition to $A(\mu_1)\bullet' A(\mu_2)=A(\mu_1\circ'\mu_2)$.

Both $\circ$ and $\circ'$ satisfy (28)–(30). Another reasonable property is also satisfied: If $\mu_1=\mu_2$ is carried at a single point in phase space, then $\mathscr{E}(\mu_1,\cdot)\circ\mathscr{E}(\mu_1,\cdot)=\mathscr{E}(\mu_1,\cdot)$ and similarly for $\circ'$, so that there is one case in which one expects no interference: when there is only one particle under consideration. A single particle should not interfere with itself.

In Sec. VI we computed one case of $[\mathscr{E}(\mu,\cdot)]^n(\rho)$ as $n\to\infty$ and found that it could not be written as $\mathscr{E}(\omega,\rho)$ for some measure ω independent of ρ. Thus

$$\lim_{n\to\infty}\frac{[\mathscr{E}(\mu,\cdot)]^n(\rho)}{\mathrm{Tr}[\mathscr{E}(\mu,\cdot)]^n(\rho)}\neq\mathscr{E}(\mu\circ\mu,\rho),$$

and the limit is distinctly different from the $\circ$ composition of instruments.

As an example, take μ_1 of the form

$$d\mu_1(\bar{Q},\bar{P})=d\nu_1(\bar{Q})d\bar{P}=\nu_1'(\bar{Q})\,d\bar{Q}d\bar{P}$$

and μ_2 of the form

$$d\mu_2(\bar{Q},\bar{P})=d\bar{Q}\,d\nu_2(\bar{P})=\nu_2'(\bar{P})\,d\bar{Q}d\bar{P}.$$

Then (36) [and similarly (34)] becomes

$$[\mathscr{E}(\mu_1,\cdot)\circ'\mathscr{E}(\mu_2,\cdot)](\rho)$$

$$=\int dx\,dy\,ds\,dt\,\nu_1'(x+s/2)\nu_2'(y-t/2)$$

$$\times\exp[-(c/\hbar)(s^2+t^2/c^2)]T_{x,y}\,\rho T_{x,y}.$$

In the extreme case where $d\nu_1(Q)$ is a δ function at Q_0 and $d\nu_2(P)$ at P_0, then we get

$$[\mathscr{E}(\mu_1,\cdot)\circ'\mathscr{E}(\mu_2,\cdot)](\rho)=\int dx\,dy\exp[-4c(x^2+y^2/c^2)/\hbar]$$

$$\times T_{Q_0+x/2,P_0+y/2}\,\rho T_{Q_0+x/2,P_0+y/2}.$$

We may then use (12 and (13) to compute the residual variance in Q and P. More importantly, we see than an optimally precise measurement of Q and one of P, if simultaneously composed in the $\circ'$ (or $\circ$) manner, yield an instrument (which is not optimal). A corresponding analysis of (22) yields a similar result for an optimally precise measurement of Q followed by one of P, and conversely.

We have introduced $\circ$ so as to augment the concept of "addition of amplitudes" as a form of coherent interference. $\circ$ will not replace the "addition of amplitude" concept, however, as it does not lead to an adequate description of interference effects such as appear in the double slit experiment, as a lengthy computation, which we do not reproduce here, will verify.

VIII. THE CLASSICAL LIMIT

In the limit $\hbar\to 0$, the Weyl algebra becomes commutative, so that measurement theory should then resemble classical measurement theory. By (16), $\lim_{\hbar\to 0^+}E(x,y)=1$. By (2), $\lim_{\hbar\to 0^+}g(x,y;c\bar{Q},\bar{P},C)=\exp(-i\bar{P}x-i\bar{Q}y)$ (ignoring normalization). Then, by (1), T_g becomes

$$\sim\int dx\,dy\exp(-i\bar{P}x-i\bar{Q}y)\exp(ixP)\exp(iyQ)\,\partial$$

$$=\int dx\,dy\exp[-ix(\bar{P}-\lambda)]\exp[-iy(\bar{Q}-\mu)]$$

$$\times dE^P_\lambda dE^Q_\mu$$

$$\sim E^P_{\bar{P}}E^Q_{\bar{Q}},$$

these spectral families now commuting. $\mathscr{E}(\mu,\cdot)$ thus becomes a classical conditioning and obeys von Neumann's collapse scheme. Also, in the classical limit,

$d\mu_1\circ\mu_2(x,y)$ and $d\mu_1\circ'\mu_2(x,y)$ are both given by

$$\int_{s,t}d\mu_1(x+s,y+t)d\mu_2(x-s,y-t),$$

which is a symmetrized convolution. Finally, Eq. (22) for sequential measurements reduces to a result compatible with the repeatability axiom for "measurement immediately after."

IX. DYNAMICS

If we work in the Heisenberg representation, we wish to be able to compute

$$\alpha_t[A(\mu)]\equiv\exp(iHt/\hbar)A(\mu)\exp(-iHt/\hbar), \tag{37}$$

where H is the Hamiltonian for the system. If we work instead in the Schrödinger representation, we would want

$$\beta_t(\rho)\equiv\exp(-iHt/\hbar)\,\rho\,\exp(iHt/\hbar) \tag{38}$$

for any state ρ. Since the two representations are related in the canonical fashion

$$\mathrm{Tr}[\beta_t(\rho)A(\mu)]=\mathrm{Tr}[\rho\alpha_t(A(\mu))], \tag{39}$$

it suffices to compute only (37). Now $\alpha_t(A(\mu))$ will be some function of the Weyl operators of the form $A(\mu_t)$ in view of the joint spectral family nature of the $A(\mu)$, as detailed in Sec. V. In general the t dependence of μ_t will be complicated, but we shall shortly present two simple examples in detail. μ_t is always linear in μ since

$$\alpha_t(A(\mu))=\int d\mu(\bar{Q},\bar{P})\,\alpha_t(T_{\bar{Q},\bar{P}}). \tag{40}$$

This may also be written in the form

$$\alpha_t(A(\mu))$$

$$=\int d\mu(\bar{Q},\bar{P})\,dx\,dy\,g(x,y;c\bar{Q},\bar{P},c)\alpha_t(W_{x,y}) \tag{40'}$$

$$=\int dx\,dy\,\hat{\mu}(y,x)g(x,y;0,0,c)\,\alpha_t(W(x,y)) \tag{40''}$$

where $\hat{\mu}(y,x)\equiv\int d\mu(\bar{Q},\bar{P})\exp(-i\bar{P}x-i\bar{Q}y)$. (41)

Then (39) becomes

$$\mathrm{Tr}[\rho\alpha_t(A(\mu))] = \int dx\,dy\,\hat{\mu}(y,x)\,g(x,y;0,0,c) \times \mathrm{Tr}[\rho\alpha_t(W(x,y))], \tag{42}$$

the reversal of the roles of y,x in μ occurring since the Fourier transform of the momentum variable is the position variable, and conversely. Again we note that $\alpha_t(W(x,y))$ is some function of Q and P so that it may also be written in the form

$$\begin{aligned}\alpha_t(W(x,y)) &= \int da_t(\bar{Q}',\bar{P}';x,y)\,T_{\bar{Q}',\bar{P}'} \\ &= \int da_t(\bar{Q}',\bar{P}';x,y)\,dx'\,dy'\,g(x',y';c\bar{Q}',\bar{P}',c)\,W(x',y') \\ &= \int dx'\,dy'\,da_t(\bar{Q}',\bar{P}';x,y) \times \exp\{-i\bar{P}'x' - i\bar{Q}'y'\}g(x',y';0,0,c)\,W(x',y') \\ &= \int dx'\,dy'\hat{a}_t(y',x',x,y)\,g(x',y';0,0,c)\,W(x',y') \end{aligned} \tag{43}$$

for some x,y,t-dependent measure $a_t(\cdot,\cdot;x,y)$. If we define

$$g^{\#}(y,x) = g(x,y;0,0,c), \tag{44}$$

we have

$$\begin{aligned}&\mathrm{Tr}[\rho\alpha_t(A(\mu))] \\ &= \int dx\,dy\,dx'\,dy'\,(\hat{\mu}g^{\#})(y,x)\hat{a}_t(y',x';x,y) \times g^{\#}(y',x')\mathrm{Tr}[\rho W(x',y')],\end{aligned} \tag{45}$$

which may be viewed as a classical correlation of the transformed, regularized densities $\hat{\mu}\,g^{\#}$ and $\hat{\rho}\,g^{\#}$, where

$$\hat{\rho}(-y',-x') = \mathrm{Tr}[\rho W(x',y')], \tag{46}$$

all dynamics being contained in the term $\hat{a}_t(y',x',;x,y)$. For $t=0$ we obtain

$$\mathrm{Tr}[\rho A(\mu)] = \int dx\,dy\,(\hat{\mu}g^{\#})(y,x)\,\hat{\rho}(-y,-x). \tag{47}$$

As examples, take the Hamiltonian to describe a system in the presence of N harmonic oscillators: $H = (2m)^{-1}P^2 + \Sigma_{i=1}^{N} c_i(Q - a_i\mathbb{1})^2$; i.e.,

$$H = (2m)^{-1}P^2 + \delta Q^2 + \beta Q + \gamma\mathbb{1} \tag{48}$$

for m,δ,β,γ real scalars. The case $\delta=\beta=\gamma=0$ is the free dynamics case. In the Appendix, by applying a slight extension of a result in Ref. 11, we calculate

$$\begin{aligned}&\alpha_t(W(x,y)) \\ &\quad = \exp(i(\beta/2\delta)\{[c(t)-1]\,y - (2\delta m)^{1/2}s(t)x\}) \\ &\quad\times W(s(t)(2\delta m)^{-1/2}y + c(t)x,\, c(t)y - s(t)(2\delta m)^{1/2}x),\end{aligned} \tag{49}$$

where

$$c(t) = \cos[(2\delta m^{-1})^{1/2}t], \tag{50}$$

$$s(t) = \sin[2\delta m^{-1})^{1/2}t], \tag{51}$$

and

$$\alpha_t^{\mathrm{free}}(W(x,y)) = W(x + tym^{-1}, y), \tag{52}$$

a simple displacement in position. Consequently,

$$\begin{aligned}&\mathrm{Tr}[\rho\alpha_t(A(\mu))] = \int dx\,dy\,(\hat{\mu}g^{\#})(y,x) \\ &\times\exp(i\{[c(t)-1]y - (2\delta m)^{1/2}s(t)x\}\,\beta/2\delta) \\ &\times\hat{\rho}(-c(t)y + s(t)(2\delta m)^{1/2}x, -s(t)(2\delta m)^{-1/2}y - c(t)x),\end{aligned} \tag{53}$$

and in the free case

$$\begin{aligned}&\mathrm{Tr}[\rho\alpha_t^{\mathrm{free}}(A(\mu))] \\ &\quad = \int dx\,dy\,(\hat{\mu}g^{\#})(y,x)\hat{\rho}(-y,-x-tym^{-1}).\end{aligned} \tag{54}$$

If $\rho = \mathscr{E}(\sigma,\rho')$, then

$$\hat{\rho}(-y,-x) = \mathrm{Tr}[\rho W(x,y)] = \frac{2\pi}{\hbar}[\sigma(\rho')]\,\hat{}\,g^{\#}\}(-y,-x), \tag{55}$$

where $[d\sigma(\rho')](\bar{Q},\bar{P}) = d\sigma(\bar{Q},\bar{P})\mathrm{Tr}(\rho' T_{\bar{Q},\bar{P}})$. Inserting (55) into (53) and (54) gives a more symmetric appearing result.

For general translations in the position and momentum domains, as opposed to the time domain, see Ref. 3, p. 324, in which it is shown that the $\{A(\mu)\}$ under the action of $\exp(-iaP)\exp(-iQb)$ form a generalized system of imprimitivity in the sense of Ref. 8.

X. IMPLICATIONS

In Ref. 12 in order to view the set of physical observables of a system as a Jordan *-algebra, and later as a C^*-algebra, it is necessary to have a (large) number of dispersion-free states for each observable (so that powers of the observable are defined). As we have seen, in this model the physically relevant observables in a Weyl system are the $A(\mu)$ and the experimentally preparable states are the $\mathscr{E}(\sigma,\rho)$. If we choose μ as in (18) or (19), we have $A(\mu) = P$ or Q, which in Sec. IV were shown to have no dispersion-free states of the form $\mathscr{E}(\sigma,\rho)$. If we take μ to be concentrated at a single point, then $A(\mu)$ is dispersion-free on the optimal experiments $\mathscr{E}(\mu,\rho)$. If we could generate all the $A(\mu)$ from those with μ a point mass by taking (limits of) linear combinations, then this would be sufficient. However, in any of the usual topologies (uniform, $\mathscr{L}^P$, etc.) one does not so generate measures absolutely continuous with respect to Lebesegue measure on any interval. So we search for a larger class of observables $A(\mu)$ with dispersion-free states. From (14'), $A(\mu)$ will have zero variance on $\mathscr{E}(\sigma,\rho)$ only if σ is a point mass. Thus it suffices to consider the variance of $A(\mu)$ in state $\rho = T_{\bar{Q}'',\bar{P}''}\,\rho\,T_{\bar{Q}'',\bar{P}''}$. We compute

$$\begin{aligned}&\mathrm{Var}_{\rho}(A(\mu)) \\ &\quad = \int d\mu(\bar{Q},\bar{P})d\mu(\bar{Q}',\bar{P}')\exp\{-c(2\hbar)^{-1} \\ &\qquad\times[(\bar{Q}-\bar{Q}'')^2 + (\bar{Q}'-\bar{Q}'')^2]\} \\ &\qquad\times\exp\{-(2\hbar c)^{-1}[(\bar{P}-\bar{P}'')^2 + (\bar{P}'-\bar{P}'')^2]\} \\ &\quad\times(\exp\{-(c/4\hbar)[(\bar{Q}-\bar{Q}')^2 - (\bar{Q}'-\bar{Q}'')^2 - (\bar{Q}-\bar{Q}'')^2] \\ &\qquad - (4\hbar c)^{-1}[(\bar{P}-\bar{P}')^2 - (\bar{P}'-\bar{P}'')^2 - (\bar{P}-\bar{P}'')^2]\} \\ &\qquad\times\cos\{(2\hbar)^{-1}[\bar{Q}''(\bar{P}'-\bar{P}) + \bar{P}''(\bar{Q}-\bar{Q}') \\ &\qquad + \bar{P}\bar{Q}' - \bar{P}'\bar{Q}]\} - 1).\end{aligned} \tag{56}$$

We conjecture that this is zero iff μ is concentrated on

$(\bar{Q}'', \bar{P}'')$, in which case one of the basic axioms of the algebraic approach to physics is not realized in this model.

Turning to the quantum logic approach, we recall the following: The elements of a quantum logic are the so-called "yes–no" or "elementary" physical propositions and form an orthocomplemented lattice. In a Hilbert space representation, the elementary propositions become spectral projectors for self-adjoint operators. By comparison, in our setting the observables are the self-adjoint $A(\mu)$; however, it is not the spectral projectors of the $A(\mu)$ which are of physical interest—instead it is the decomposition of the $A(\mu)$ into the $T_{\bar{Q},\bar{P}}$. If we are to make a lattice of the $\{A(\mu)\}$, we must identify the $\wedge$ lattice operation (both/and) as well as the $\vee$ lattice operation (either/or/both) and find the corresponding partial order on $\{A(\mu)\}$.

In quantum logic $a \leqslant b$, which is equivalent to $a^c \geqslant b^c$, may be interpreted "whenever proposition b is not observed (has probability zero), then a is not observed (has probability zero)." Hence we, by analogy, take $A(\mu) \leqslant A(\nu)$ to mean μ is absolutely continuous with respect to ν:

$$A(\mu) \leqslant A(\nu) \leftrightarrow \mu \ll \nu. \tag{57}$$

This suggests that all of the putative lattice structure would then be expressible solely on the set of measures. This is further supported by the correspondence (15) between the observables $A(\mu)$ and the conditionings $\mathscr{E}(\mu,\cdot)$, and the discussion in Sec. VI and VII. In fact, we have seen that there are three relevant binary operations on the instruments: sequential composition, simultaneous incoherent composition, and simultaneous coherent composition. The first has no present analog in quantum logic. The other two lead to the following natural definitions:

$$\mathscr{E}(\mu,\cdot) \vee \mathscr{E}(\nu,\cdot) = \mathscr{E}(\mu+\nu,\cdot), \tag{58}$$

or

$$A(\mu) \vee A(\nu) = A(\mu+\nu); \tag{58'}$$

and

$$\mathscr{E}(\mu,\cdot) \wedge \mathscr{E}(\nu,\cdot) = \mathscr{E}(\mu \circ \nu,\cdot), \tag{59}$$

or

$$A(\mu) \wedge A(\nu) = A(\mu \circ \nu) \tag{59'}$$

with $\circ$ given either by (35) or by (35) with E^2 replacing E. However, although $\circ$ is (i) commutative, it is (ii) nonassociative, (iii) nonidempotent ($\mu \circ \mu \neq \mu$), and (iv) $\circ$ has no identity. Consequently, (v) no complementation is defined either. Hence this attempt to mimic quantum logic does not yield a lattice at all. And, in retrospect, since no piece of experimental apparatus has sharp boundaries due to the uncertainty principle and other phenomena, not to mention what is being measured, the actual existence of the "yes–no" propositions seems dubious; and even if they existed in some sense, they wouldn't have complements because of the nonsharp boundaries. From considerations of complementarity and compatibility, the existence of a greatest lower bound in quantum logic has also been called into question.[13] For a discussion of a restricted form of fuzzy logic, see Ref. 14, although it does not quite apply: Our fuzzy sets are minimum uncertainty functions so the support of each g in (2) is all of $\mathbb{R}$, and the complement in Ref. 14 degenerates.

Irrespective of any interpretation of the $\circ$ operation or of the existence of the greatest lower bound, in quantum logic one assumes that for every physical yes/no proposition there is a physical state in which the probability of obtaining a "yes" for the yes/no proposition is one. Such a state is necessarily dispersion-free for that proposition; yet we have seen that dispersion-free physical states for the general $A(\mu)$ may not exist.

Concerning causality, locality, and determinism, we recall that in Sec. V we showed that $\mathrm{Tr}[\mathscr{E}(\mu,\rho)A(\sigma)]$ is nonvanishing even if μ and σ have disjoint supports. Thus one consequence of working in fuzzy phase space is that we obtain this form of action-at-a-distance (nonlocality). Because the conditionings and observables are described in fuzzy phase space, we have a nondeterministic theory, although the dynamics of Sec. IX does give a definite result for the (nonlocal) interaction and the time evolution of expected values of observables: the theory is *both* nondeterministic and nonrandom. Hence, if one identifies the concepts "nonrandom theory" and "causal theory," then one must conclude that causality and determinism are not identical. This suggests the need for a definition of causality independent of the concept of determinism.

ACKNOWLEDGMENTS

The author is indebted to Professor J.-P. Marchand and Professor S. P. Gudder for innumerable useful discussions of this work. He would also like to thank Mr. S. Blumenthal for prodding him into a consideration of some of the implications. Furthermore, several comments from the referee have lead to pleasant improvements.

APPENDIX

Since α_t is an automorphism, we have

$$\alpha_t(W(x,y)) = \alpha_t(\exp(ixP + iyQ)) = \exp[ix\alpha_t(P) + iy\alpha_t(Q)].$$

Thus we need only compute $\alpha_t(P)$ and $\alpha_t(Q)$. From Lemma 4 of Ref. 11, if the multiple commutators $(\mathrm{ad}H)^K(P)$, $(\mathrm{ad}H)^K(Q)$ vanish for sufficiently large K, then

$$\alpha_t(P) = \sum_{x=0}^{\infty} (it/\hbar)^K (\mathrm{ad}H)^K(P)/K!.$$

This termination condition is satisfied in the case of free dynamics and yields $\alpha_t(P) = P$, $\alpha_t(Q) = Q + tP/m$. For the general case, we work initially on a common dense domain of analytic vectors for P and Q and eventually extend $\alpha_t(W(x,y))$ to a unitary operator on all of $\mathscr{H}$. Now, we prove by induction, from $[P,Q] = -i\hbar\mathbb{1}$ that

$$(\mathrm{ad}H)^{2n-1}(Q) = (-i\hbar/m)(2\delta\hbar^2/m)^{n-1}P, \quad n = 1,2,\cdots, \tag{A1}$$

$$(\mathrm{ad}H)^{2n}(Q) = (2\delta\hbar^2/m)^n[Q + (\beta/2\delta)\mathbb{1}], \quad n = 1,2,\cdots, \tag{A2}$$

$$[H,P] = i2\delta\hbar Q + i\hbar\beta\mathbb{1}, \tag{A3}$$

$$(\mathrm{ad}H)^n(P) = (2i\delta\hbar)(\mathrm{ad}H)^{n-1}(Q), \quad n = 2,3,\cdots. \tag{A4}$$

From (59)–(61) one obtains

$$\alpha_t(Q) = (\beta/2\delta)\{\cos[(2\delta/m)^{1/2}t\,] - 1\}\mathbb{1} + \cos[(2\delta/m)^{1/2}t\,]Q + (2\delta m)^{-1/2}\sin[(2\delta m)^{1/2}t\,]P, \tag{A5}$$

and from (61) and (62)

$$\alpha_t(P) = \cos[(2\delta/m)^{1/2}t\,]P - (2\delta m)^{1/2}\sin[(2\delta/m)^{1/2}t\,] \times [Q + (\beta/2\delta)\mathbb{1}]. \tag{A6}$$

Checking the limit as $\beta, \delta \to 0$, one reobtains the free dynamics.

If we now let

$$c(t) = \cos[(2\delta/m)^{1/2}t], \tag{A7}$$

$$s(t) = \sin[(2\delta/m)^{1/2}t], \tag{A8}$$

we then have

$$\alpha_t\{W(x,y)\} = \exp\{i(\beta/2\delta)\{[c(t)-1]y - (2\delta m)^{1/2}s(t)x\}\} \times W(s(t)\,y/(2\delta m)^{1/2} + c(t)x, c(t)\,y - s(t)\,(2\delta m)^{1/2}x), \tag{A9}$$

which has a unique unitary extension off the dense analytic domain.

In the limit $\beta \to 0$, $\delta \to 0$ we then obtain the corresponding free dynamics

$$\alpha_t^{\text{free}}(W(x,y)) = W(x + t\,y)/m, y). \tag{A10}$$

[1] E. Prugovecki, "Quantum Spacetime Operationally Based on Propagators for Extended Test Particles," Hadronic J. **4** (1981), to appear.

[2] For a discussion of an experimental realization of the simultaneous measurement of position and momentum, see G. F. Cerofolini, Nuovo Cimento **588**, 286–300 (1980), esp. pp. 295–6.

[3] F. E. Schroeck, Jr., "Measures with Mimimum Uncertainty," in *Mathematical Foundations of Quantum Theory*, edited by A. R. Marlow (Academic, New York, 1978), pp. 299–327.

[4] E. Prugovecki, *Quantum Mechanics in Hilbert Space* (Academic, New York, 1971), pp. 331–47, which includes many of the results of Sec. II for the case $\bar{Q} = \bar{P} = 0$. See also J. von Neumann, Math. Ann. **104**, 570–8 (1931).

[5] J. H. Marburger and E. W. Power, Found. Phys. **10**, 865–74 (1980).

[6] E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239–60 (1970).

[7] E. B. Davies, J. Funct. Anal. **6**, 318–46 (1970).

[8] S. T. Ali and G. G. Emch, J. Math. Phys. **15**, 176–182 (1974).

[9] T. Louton, Pac. J. Math. **59**, 147–59 (1975).

[10] J. M. Jauch and C. Piron, Helv. Phys. Acta **40** 559–70 (1967).

[11] F. E. Schroeck, Jr., J. Math. Phys. **16**, 729–39 (1975), Lemma 4.

[12] G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley-Interscience, New York, 1972), esp. p. 40, Axiom 4.

[13] Pekka J. Lahti, "Uncertainty and Complementarity in Axiomatic Quantum Mechanics," preprint, 1979.

[14] G. Gattanco, J. Math. Anal. Appl. **75**, 523–48 (1980).

[15] E. Prugovecki, J. Math. Phys. **17**, 517–523 (1976), esp. Eqs. (5.20) and (5.21).

An expansion for power-law wavefunctions

Bruce R. Johnson
Theoretical Chemistry Institute, University of Wisconsin–Madison, Madison, Wisconsin 53706

(Received 19 May 1981; accepted for publication 19 June 1981)

The regular and irregular solutions of the radial wave equation for r^α potentials are given in terms of simultaneous power series in r^2 and $r^{\alpha+2}$. These results are shown to generalize to a potential with an arbitrary number of power-law terms.

PACS numbers: 03.65.Ca

1. INTRODUCTION

It has been noted before[1–4] that the solutions to the radial wave equations for power-law potentials r^α and r^β, $\beta = -2\alpha/(\alpha+2)$, are related by a change of variable. This is quite interesting since knowledge of the exact functional form of these wavefunctions is fairly limited except for special values of the exponents. The possibility is raised that the case of general exponent α (> -2) may be more tractable than usually supposed.

Now, in principle, one first wants to have at hand the two standard solutions to the second-order differential equation, from which both scattering and bound state wavefunctions may be obtained by imposition of the proper boundary conditions. As an initial step toward this end, the present paper addresses the construction of these solutions in the small r limit, with primary focus on the regular one.

This seems a difficult task since the nonanalyticity of the potential r^α for general α immediately precludes an analytic power series. It will turn out, however, that it is possible to express the solutions as double power series, one in r^2 and one in $r^{\alpha+2}$, reducing the nonanalyticity to a relatively mild form. (These results have been anticipated in part by Müller-Kirsten, Hite, and Bose.[5])

The coefficients in this expansion obey a three-term recursion relation that can be solved rather explicitly. In the process, a natural generalization occurs of the connection with partitioning theory that Phares and co-workers[6] used to expand the wavefunctions for the linear potential. This connection also pervades the problem of a potential containing several distinct powers of r, which is discussed near the end. Such a potential merely requires one more simultaneous power series for each additional power of r, although the algebra grows more complicated.

The program is the following. In Sec. 2, the expansion of the regular solution for a single-term potential is set down, and the simple recursion relation [Eq. (2.10)] which the coefficients must obey is derived. This relation is solved iteratively in Sec. 3, where the structure of the coefficients is discussed. The distinction between rational and irrational exponents is pursued. In Sec. 4, some examples are given. Section 5 gives the expansion and resulting recursion relation for the irregular solution. The generalization to a potential with an arbitrary number of powers of r is indicated in Sec. 6, followed by a discussion in Sec. 7.

2. EXPANSION OF THE REGULAR SOLUTION AT THE ORIGIN

The radial Schrödinger equation for a particle moving in a central potential (in an arbitrary number of dimensions) may be cast into the form[4]

$$\left[\frac{d^2}{dr^2} - \frac{L^2-\frac{1}{4}}{r^2} - U(r) + \mathscr{E}\right]\chi(r) = 0. \tag{2.1}$$

Our intent is to determine the expansion about $r=0$ for the regular solution when the potential is a single power of r,

$$U(r) = ar^\alpha, \quad (\alpha > 0). \tag{2.2}$$

From earlier results, this range is sufficient to also obtain solutions for $-2 < \alpha < 0$.[1–4]

The function $\chi(r)$ behaves as $r^{L+1/2}$ near $r=0$,[7] so this is factored off immediately,

$$\chi(r) = r^{L+1/2}\Phi(r), \tag{2.3}$$

and Φ obeys

$$\left[\frac{d^2}{dr^2} + \frac{2L+1}{r}\frac{d}{dr} - ar^\alpha + \mathscr{E}\right]\Phi(r) = 0. \tag{2.4}$$

A power series expansion of Φ is clearly inappropriate in the normal sense since the various terms are of (in general) incommensurate powers of the variable. The appropriate form of the expansion is suggested by temporarily changing the scale of r, $r \to cr$, which takes Eq. (2.4) into

$$\left[\frac{d^2}{dr^2} + \frac{2L+1}{r}\frac{d}{dr} - c^{\alpha+2}ar^\alpha + c^2\mathscr{E}\right]\Phi(cr) = 0. \tag{2.5}$$

To explicitly account for these scaling properties, it is necessary to have both powers of r^2 and $r^{\alpha+2}$ in Φ. Thus, after defining for notational convenience

$$\lambda = (\alpha+2)/2, \tag{2.6}$$

Φ is taken to be of the form

$$\Phi(r) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} d_{ij} r^{2i+2j\lambda}. \tag{2.7}$$

At this point, λ is assumed to be irrational. This is unnecessary, as will be discussed in Sec. 3, but allows of the following simplification. If λ is irrational, Eq. (2.7) may be substituted into Eq. (2.4) and each power of $r^{2i+2j\lambda}$ is linearly independent (i.e., the coefficient of $r^{2i+2j\lambda}$ must vanish for each i and each j). The d_{ij} must then obey the three-term recursion relation

 0022-2488/81/112573-08$01.00

$$4(i+j\lambda)(L+i+j\lambda)d_{ij} = ad_{i,j-1} - \mathscr{E}d_{i-1,j}. \tag{2.8}$$

This is reduced further by eliminating a and $\mathscr{E}$:

$$d_{ij} = (-\mathscr{E}/4)^i(a/4)^j c_{ij}, \tag{2.9}$$

where the c_{ij} are related by

$$(i+j\lambda)(L+i+j\lambda)c_{ij} = c_{i-1,j} + c_{i,j-1}, \tag{2.10}$$

with the conventions adopted that $c_{00} = 1$ and $c_{ij} = 0$ for $i<0$ or $j<0$.

Equation (2.10) is a remarkably simple recursion relation, constituting a generalization of the two-term one satisfied by the coefficients of Bessel functions. In this connection, the function $\chi(r)$ may be regarded as a double-series Bessel function, as will be made a bit more precise in Sec. 3, where the explicit form of c_{ij} is obtained.

3. SOLUTION OF THE RECURSION EQUATION

A. λ irrational

The coefficients c_{ij} may be determined iteratively from Eq. (2.10). The first thing to notice is that a two-term recur- occurs if either i or j is set equal to zero:

$$j\lambda(L+j\lambda)c_{0,j} = c_{0,j-1}, \tag{3.1}$$

$$i(L+i)c_{i,0} = c_{i-1,0}. \tag{3.2}$$

Equations (3.1) and (3.2) are trivially iterated to give

$$c_{0,j} = [\lambda^{2j} j!(L/\lambda+1)_j]^{-1}, \tag{3.3}$$

$$c_{i,0} = [i!(L+1)_i]^{-1}. \tag{3.4}$$

These coefficients are the ones entering into the well-known solutions of Eq. (2.4) when $a=0$ or $\mathscr{E}=0$, respectively. These are the Bessel functions J_ν and I_ν[8]:

$$\Phi_{a=0}(r) \propto r^{-L} J_L(\pm \mathscr{E}^{1/2} r) \propto \frac{1}{\Gamma(L+1)} \sum_{i=0}^{\infty} c_{i0} \left(-\frac{\mathscr{E}}{4}\right)^i r^{2i}, \tag{3.5}$$

$$\Phi_{\mathscr{E}=0}(r) \propto r^{-L} I_{L/\lambda}(a^{1/2} r^\lambda/\lambda) \propto \frac{1}{\Gamma(L/\lambda+1)} \sum_{j=0}^{\infty} c_{0j} \left(\frac{a}{4}\right)^j r^{2j\lambda}. \tag{3.6}$$

The next sept in solving Eq. (2.10) is to iterate in either i or j. If we assume that $i>0$, then

$$\begin{aligned} c_{ij} &= [(i+j\lambda-1)(L+i+j\lambda)]^{-1}(c_{i,j-1} + c_{i-1,j}) \\ &= [(i+j\lambda)(L+i+j\lambda)]^{-1} c_{i,j-1} \\ &\quad + [(i+j\lambda-1)(i+j\lambda)(L+i+j\lambda-1)(L+i+j\lambda)]^{-1} \\ &\quad \times (c_{i-1,j-1} + c_{i-2,j}) \\ &\quad \cdots \\ &= \sum_{k=0}^{i} [(k+j\lambda)_{i+1-k}(L+k+j\lambda)_{i+1-k}]^{-1} c_{k,j-1}. \end{aligned} \tag{3.7}$$

This may be iterated in the k index, giving at the end a rather complicated form for c_{ij} in terms of j summations and the coefficients in Eq. (3.4),

$$\begin{aligned} c_{ij} &= \sum_{k_1=0}^{i} [(k_1+j\lambda)_{i+1-k_1}(L+k_1+j\lambda)_{i+1-k_1}]^{-1} c_{k_1,j-1} \\ &= \sum_{k_1=0}^{i} \sum_{k_2=0}^{k_1} [(k_1+j\lambda)_{i+1-k_1}(L+k_1+j\lambda)_{i+1-k_1} \\ &\quad \times (k_2+(j-1)\lambda)_{k_1+1-k_2}(L+k_2+(j-1)\lambda)_{k_1+1-k_2}]^{-1} \\ &\quad \times c_{k_2,j-2} \\ &= \cdots \\ &= \sum_{k_1=0}^{i} \sum_{k_2=0}^{k_1} \sum_{k_3=0}^{k_2} \cdots \sum_{k_j=0}^{k_{j-1}} [(k_1+j\lambda)_{i+1-k_1} \\ &\quad \times (L+k_1+j\lambda)_{i+1-k_1} (k_2+(j-1)\lambda)_{k_1+1-k_2} \\ &\quad \times (L+k_2+(j-1)\lambda)_{k_1+1-k_2} \\ &\quad \cdots \\ &\quad \times (k_j+\lambda)_{k_{j-1}+1-k_j}(L+k_j+\lambda)_{k_{j-1}+1-k_j}]^{-1} c_{kj0}. \end{aligned} \tag{3.8}$$

A similar form is obtained if i and j are iterated in the opposite order. Equation (3.8) is a little cumbersome and obscures some of the order in the coefficients. To condense the notation, we introduce the quantity

$$\begin{aligned} (\gamma|\mu_1^{n_1}\mu_2^{n_2}\cdots\mu_m^{n_m}) &= (\gamma+\mu_1)(\gamma+2\mu_1)\cdots(\gamma+n_1\mu_1) \\ &\quad \times (\gamma+n_1\mu_1+\mu_2)\cdots(\gamma+\sum_{l=1}^{m} n_l\mu_l), \end{aligned} \tag{3.9}$$

which is a generalization of the usual shifted factorial in which successive factors are incremented by different amounts (n_i is the number of factors increased by length μ_i). The case of only one μ reduces to

$$(\gamma|\mu^n) = \mu^n(\gamma/\mu+1)_n. \tag{3.10}$$

Thus, using Eq. (3.4) for c_{kj0}, Eq. (3.8) is

$$\begin{aligned} c_{ij} = \sum_{k_1=0}^{i} \sum_{k_2=0}^{k_1} \cdots \sum_{k_j=0}^{k_{j-1}} &[(0|1^{k_j}\lambda^{\,1}1^{k_{j-1}-k_j}\cdots\lambda^{\,1}1^{i-k_1}) \\ &\times (L\,|1^{k_j}\lambda^{\,1}1^{k_{j-1}-k_j}\cdots\lambda^{\,1}1^{i-k_1})]^{-1}. \end{aligned} \tag{3.11}$$

An alternative form, produced by the other order of iteration, is

$$\begin{aligned} c_{ij} = \sum_{l_1=0}^{j} \sum_{l_2=0}^{l_1} \cdots \sum_{l_i=0}^{l_{i-1}} &[(0|\lambda^{\,l_i}1^{\,1}\lambda^{\,l_{i-1}-l_i}\cdots 1^{\,1}\lambda^{\,j-l_1}) \\ &\times (L\,|\lambda^{\,l_i}1^{\,1}\lambda^{\,l_{i-1}-l_i}\cdots 1^{\,1}\lambda^{\,j-l_1})]^{-1}. \end{aligned} \tag{3.12}$$

The coefficients c_{ij} can be described fairly easily, despite the complicated appearance of Eqs. (3.11) and (3.12). Each c_{ij} is a sum over

$$\binom{i+j}{j}$$

terms, with each term represented by the inverse of a pair of generalized shifted factorials, defined in Eq. (3.9). Each member of this pair consists of $(i+j)$ total factors with different starting points [$\gamma=0$ and L in Eq. (3.9)] but identical geneaologies. The increments between succeeding factors consist of i of length unity and j of length λ, with the ordering of these increments determined by the geneaology of the term, which can be represented by the set of k's in the particular term in Eq. (3.11), or by the set of l's in Eq. (3.12).

It is clear from either of these equations that each of the

$$\binom{i+j}{j}$$

terms corresponds to one of the possible ways that an interval of total length $i+j\lambda$ can be partitioned into i intervals of length unity and j of length λ, considering intervals of these same length as indistinguishable. The order in which the interval is partitioned translates into the order in which increments are made in the succession of factors, and the multiple summations then produce each possible order once and only once. To illustrate, the first few c_{ij} for which $i\neq 0$ and $j\neq 0$ are, explicitly,

$$c_{1,1} = [(\lambda)(1+\lambda)(L+\lambda)(L+1+\lambda)]^{-1} + [(1)(1+\lambda)(L+1)(L+1+\lambda)]^{-1}, \tag{3.13}$$

$$\begin{aligned} c_{1,2} = {} & [(\lambda)(2\lambda)(1+2\lambda)(L+\lambda)(L+2\lambda)(L+1+2\lambda)]^{-1} \\ & + [(\lambda)(1+\lambda)(1+2\lambda)(L+\lambda)(L+1+\lambda)(L+1+2\lambda)]^{-1} \\ & + [(1)(1+\lambda)(1+2\lambda)(L+1)(L+1+\lambda)(L+1+2\lambda)]^{-1}, \end{aligned} \tag{3.14}$$

$$\begin{aligned} c_{2,1} = {} & [(\lambda)(1+\lambda)(2+\lambda)(L+\lambda)(L+1+\lambda)(L+2+\lambda)]^{-1} \\ & + [(1)(1+\lambda)(2+\lambda)(L+1)(L+1+\lambda)(L+2+\lambda)]^{-1} \\ & + [(1)(2)(2+\lambda)(L+1)(L+2)(L+2+\lambda)]^{-1}, \end{aligned} \tag{3.15}$$

$$\begin{aligned} c_{2,2} = {} & [(\lambda)(2\lambda)(1+2\lambda)(2+2\lambda)(L+\lambda)(L+2\lambda)(L+1+2\lambda)(L+2+2\lambda)]^{-1} \\ & + [(\lambda)(1+\lambda)(1+2\lambda)(2+2\lambda)(L+\lambda)(L+1+\lambda)(L+1+2\lambda)(L+2+2\lambda)]^{-1} \\ & + [(1)(1+\lambda)(1+2\lambda)(2+2\lambda)(L+1)(L+1+\lambda)(L+1+2\lambda)(L+2+2\lambda)]^{-1} \\ & + [(\lambda)(1+\lambda)(2+\lambda)(2+2\lambda)(L+\lambda)(L+1+\lambda)(L+2+\lambda)(L+2+2\lambda)]^{-1} \\ & + [(1)(1+\lambda)(2+\lambda)(2+2\lambda)(L+1)(L+1+\lambda)(L+2+\lambda)(L+2+2\lambda)]^{-1} \\ & + [(1)(2)(2+\lambda)(2+2\lambda)(L+1)(L+2)(L+2+\lambda)(L+2+2\lambda)]^{-1}. \end{aligned} \tag{3.16}$$

Although the general form of the coefficients c_{ij} was not clarified, Müller-Kirsten, Hite and Bose[5] have previously derived the leading terms in the double expansion of Eq. (2.7).

As an aside, it may be verified that c_{ij} obeys the symmetry relation

$$c_{ij}(\lambda,L) = \lambda^{-2i+2j} c_{ij}(1/\lambda, L/\lambda). \tag{3.17}$$

Reference 4 shows that $\chi(r)$ in Eq. (2.3) can be transformed by a change of variable into a radial solution for a different power-law potential ($\alpha \to -\alpha/\lambda$) with $L \to L/\lambda$, and Eq. (3.17) arises as a necessary consequence of this fact.

Since Eq. (2.7) is a double power series, discussions of absolute convergence are a little removed from the normal realm of single series. However, we can still find upper and lower bounds for each c_{ij} by noting that the smallest of the

$$\binom{i+j}{j}$$

terms in Eqs. (3.8) or (3.11) corresponds to $k_1 = k_2 = \cdots = k_j = 0$, while the largest has $k_1 = k_2 = \cdots = k_j = i$. Thus we obtain the crude, but rigorous, bounds

$$c_{ij}^{-} \leqslant c_{ij} \leqslant c_{ij}^{+}, \tag{3.18}$$

where

$$c_{ij}^{-} = \lambda^{-2j}\binom{i+j}{j}[j!(1+j\lambda)_i (L/\lambda+1)_j (L+1+j\lambda)_i], \tag{3.19}$$

$$c_{ij}^{+} = \lambda^{-2j}\binom{i+j}{j}[i!(i/\lambda+1)_j (L+1)_i ((L+i)/\lambda+1)_j]. \tag{3.20}$$

With Eq. (3.20), we find that $c_{i+1,j}^{+}/c_{ij}^{+} \to i^{-2}$ as $i \to \infty$ for fixed j, and $c_{i,j+1}^{+}/c_{ij}^{+} \to (\lambda j)^{-2}$ as $j \to \infty$ for fixed i, showing Bessel-function-like decrease of the envelope of upper bounds in each index.

B. λ rational

The case of rational λ is really just a slight modification of the irrational case. Although the latter allowed us to associate a unique power of r with each (i,j) pair [leading to Eqs. (2.8) or (2.10)], the resulting power series is still valid for λ rational. If we take $\lambda = q/p$, where $q > p$ and q and p are relatively prime integers, then

$$i + j\lambda = (i - kq) + (j + kq)q/p, \quad k = 0, \pm 1, \cdots \tag{3.21}$$

and there is a many-to-one correspondence between indices and powers of r.

To eliminate this nonuniqueness, i is still allowed to range from zero to infinity, while j is restricted to $j = 0,1,\ldots,p-1$. One may plug this as an ansatz into Eq. (2.7).

$$\Phi = \sum_{i=0}^{\infty} \sum_{j=0}^{p-1} d'_{ij} r^{2i+2j\lambda}, \tag{3.22}$$

and then solve the resulting (more complicated) recursion relations for the d'_{ij}. This is unnecessary, however, since we already have at hand the complete solution.

A review of the treatment of the irrational case shows that the rationality of λ was really not a consideration after getting to Eq. (2.8). The form of Φ in Eq. (2.7) with coefficients extended to general λ is perfectly valid. To make the connection with Eq. (3.22), we replace the double sum over $0 \leqslant i, j < \infty$ by a triple one where $0 \leqslant j \leqslant p-1$ and k accounts for equivalent powers of r according to Eq. (3.21),

$$\Phi = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} d_{ij} r^{2i+2j\lambda}$$
$$= \sum_{i=0}^{\infty}\sum_{j=0}^{p-1}\sum_{k=-\infty}^{\infty} d_{i-kq,\,j+kp}\, r^{2(i-kq)+2(j+kp)\lambda}$$
$$= \sum_{i=0}^{\infty}\sum_{j=0}^{p-1} r^{2i+2j\lambda} \sum_{k=-\infty}^{\infty} d_{i-kq,\,j+kp}. \tag{3.23}$$

Remembering the convention that $d_{mn} = 0$ for $m < 0$ or $n < 0$, comparison with Eq. (3.22) yields

$$d'_{ij} = \sum_{k=0}^{[i/q]} d_{i-kq,\,j+kp}, \tag{3.24}$$

where $[i/q]$ is the integral part of i/q.

It is readily seen from Eqs. (2.9) and (3.24) that d'_{ij} contains mixed powers of the parameters a and $\mathscr{E}$:

$$d'_{ij} = \sum_{k=0}^{[i/q]} \left(-\frac{\mathscr{E}}{4}\right)^{i-kq} \left(\frac{a}{4}\right)^{j+kp} c_{i-kq,\,j+kp}. \tag{3.25}$$

Although this holds for the case of general $\mathscr{E}$, an eigenvalue $\mathscr{E}_N$ of Eq. (2.1) scales as[4]

$$\mathscr{E}_n = a^{1/\lambda}\mathscr{F}_n, \tag{3.26}$$

where $\mathscr{F}_n$ is independent of a. Equation (3.25) reduces to

$$d'_{ij} = a^{i/\lambda + j} \sum_{k=0}^{[i/q]} \left(\tfrac{1}{4}\right)^{i-kq+j+kp} (-\mathscr{F}_n)^{i-kq} \times c_{i-kq,\,j+kp}, \tag{3.27}$$

and all of the a dependence is isolated as a multiplicative factor, a scale factor for r.

4. SPECIAL CASES

In this section, the preceding results are examined briefly in few a particular cases of the parameters.

A. $L = \frac{1}{2}$

When $L = \frac{1}{2}$, Eq. (2.1) reduces to an equation with no singular centrifugal term. This corresponds to S states in three dimensions, or else to a one-dimensional problem with a potential $|x|^{\alpha}$. In the latter case, the solutions we have been considering constitute only one-half of the normal solutions [after suitable matching of $\chi(x)$ and $\chi(-x)$ and their derivations at $x = 0$]; the solutions which do not vanish at the origin must be sought in the other (leading behavior $r^{-L+1/2}$) functions satisfying the second-order equation.

In the context of the coefficients c_{ij} given in Eq. (3.8), the case $L = \frac{1}{2}$ is interesting because each factor involving L can be combined with the similar factor not involving L by means of the Γ-function duplication formula.[8] Then

$$c_j(L=\tfrac{1}{2}) = \frac{2^{2i+2j}}{\Gamma(2i+2j\lambda+2)} \sum_{k_1=0}^{i} \cdots \sum_{k_j=0}^{k_{j-1}} \left\{ \frac{\Gamma(2k_j+2\lambda)}{\Gamma(2k_j+2)} \frac{\Gamma(2k_{j-1}+4\lambda)}{\Gamma(2k_{j-1}+2\lambda+2)} \cdots \frac{\Gamma(2k_1+2j\lambda)}{\Gamma(2k_1+2(j-1)\lambda+2)} \right\}. \tag{4.1}$$

B. $\alpha = 0$

The limiting case $\alpha = 0$ has as its solution Eq. (3.5) with $\mathscr{E}$ replaced by $\mathscr{E} - a$. From Eqs. (3.11) or (3.12), one sees that *each* of the terms in c_{ij} is identical:

$$c_{ij} = \sum_{k_1=0}^{i} \sum_{k_2=0}^{k_1} \cdots \sum_{k_j=0}^{k_{j-1}} [(i+j)!(L+1)_{i+j}]^{-1} = \binom{i+j}{j} \Big/ [(i+j)!(L+1)_{i+j}]. \tag{4.2}$$

Thus, from Eqs. (2.7) and (2.9),

$$\Phi(r) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \left(\frac{-\mathscr{E}}{4}\right)^{i} \left(\frac{a}{4}\right)^{j} \binom{i+j}{j} r^{2i+2j} \Big/ [(i+j)!(L+1)_{i+j}] = \sum_{j=0}^{\infty}\sum_{k=j}^{\infty} \left(\frac{-\mathscr{E}}{4}\right)^{k-j} \left(\frac{a}{4}\right)^{j} \binom{k}{j} r^{2k} \Big/ [k!(L+1)_k]$$
$$= \sum_{k=0}^{\infty} \left(\frac{a-\mathscr{E}}{4}\right)^{k} \frac{r^{2k}}{k!(L+1)_k}, \tag{4.3}$$

where we have used the binomial formula.

C. $\alpha = 2$

This is the exactly solvable radial harmonic oscillator problem. According to Sec. 3 B, we find for Φ only a single series,

$$\Phi(r) = \sum_{i=0}^{\infty} d'_i r^{2i}, \tag{4.4}$$

$$d'_i = \sum_{k=0}^{[i/2]} \left(\frac{-\mathscr{E}}{4}\right)^{i-2k} \left(\frac{a}{4}\right)^{k} c_{i-2k,k}, \tag{4.5}$$

where

$$c_{ij} = \sum_{k_1=0}^{i} \sum_{k_2=0}^{k_1} \cdots \sum_{k_j=0}^{k_{j-1}} [(0|1^{k_j}2^1 1^{k_{j-1}-k_j} \cdots 2^1 1^{i-k_1})(L\,|1^{k_j}2^1 1^{k_{j-1}-k_j} \cdots 2^1 1^{i-k_1})]^{-1}$$
$$= [(i+2j)!\Gamma(L+i+2j+1)]^{-1} \sum_{k_1=0}^{i} (k_1+2j-1)(L+k_1+2j-1)$$
$$\times \sum_{k_2=0}^{k_1} (k_2+2j-3)(L+k_2+2j-3) \cdots \sum_{k_j=0}^{k_{j-1}} (k_j+1)(L+k_j+1). \tag{4.6}$$

This is to be contrasted with the solution traditionally given in terms of Whittaker functions,[9] which leads to

$$d'_i = (-\tfrac{1}{2}a^{1/2})^i \sum_{k=0}^{i} \frac{(-\frac{1}{2})^{-k}}{k!(i-k)!}\left(\frac{L+1}{2} - \frac{\mathscr{E}}{4a^{1/2}}\right)_k \Big/ (L+1)_k. \tag{4.7}$$

Making the explicit connection between Eqs. (4.5) and appears to be difficult: this is a combinatorics problem that the author has not pursued for long. However, it may be faintly possible that this connection would suggest a generalization to other special values of α, as for instance integral values or even-integral values. The special feature of $\alpha = 2$ is of course the ability to factor out a single overall exponential dependence of ϕ, leading to a two-term recursion relation.

D. $\alpha = 1$

The linear rising potential $U(r) = ar$ has, for $L = \frac{1}{2}$, solutions which are just the familiar Airy functions. It is only recently that Phares and co-workers[6] have solved exactly the recursion relations for general L resulting from a single power series in r. There use was made of the *combinatorics function technique*, which allowed them to show explicitly a connection to partitioning an interval into lengths 2 and 3 ($\lambda = \frac{3}{2}$ for $\alpha = 1$; our partitions are correspondingly 1 and $\frac{3}{2}$). From Sec. 3 B, Φ is

$$\Phi = \sum_{i=0}^{\infty} \sum_{j=0}^{1} d'_{ij} r^{2i+3j}, \tag{4.8}$$

$$d'_{ij} = \sum_{k=0}^{[i/3]} d_{i-3k,j+2k} = \sum_{k=0}^{[i/3]} \left(\frac{-\mathscr{E}}{4}\right)^{i-3k} \left(\frac{a}{4}\right)^{j+2k} c_{i-3k,j+2k}, \tag{4.9}$$

$$c_{ij} = \sum_{k_1=0}^{i} \cdots \sum_{k_j=0}^{k_j} \left[(0|1^{k_j}(\tfrac{3}{2})^1 1^{k_{j-1}-k_j} \cdots (\tfrac{3}{2})^1 1^{i-k_1}) \times (L\,|1^{k_j}(\tfrac{3}{2})^1 1^{k_{j-1}-k_j} \cdots (\tfrac{3}{2})^1 1^{i-k_1})\right]^{-1}. \tag{4.10}$$

This may also be written in terms of ratios of Γ functions, whereby one makes the connection with, say, Eq. (16a) of the third of Refs. 6. This places the results for the linear potential into the broader context of general power-law potentials.

The partitioning approach was not used in the present derivation, but the two methods are quite equivalent. The real need for couching the formalism in terms of partitioning theory is likely to arise in considering potentials containing several powers of r (Sec. 6).

E. $\alpha = 4$

The quartic oscillator is hardly in need of preamble. In this case, as for all even-integral α, there is only one infinite series:

$$\Phi = \sum_{i=0}^{\infty} d'_{i0} r^{2i}, \tag{4.11}$$

$$d'_{i0} = \sum_{k=0}^{[i/3]} d_{i-3k,k} = \sum_{k=0}^{[i/3]} \left(\frac{-\mathscr{E}}{4}\right)^{i-3k} \left(\frac{a}{4}\right)^{k} c_{i-3k,k}, \tag{4.12}$$

$$c_{ij} = \sum_{k_1=0}^{i} \cdots \sum_{k_j=0}^{k_{j-1}} \left[(0|1^{k_j}3^1 1^{k_{j-1}-k_j} \cdots 3^1 1^{i-k}) \times (L\,|1^{k_j}3^1 1^{k_{j-1}-k_j} \cdots 3^1 1^{i-k})\right]^{-1}. \tag{4.13}$$

For the one-dimensional case, with $L = \frac{1}{2}$, we obtain the odd functions by replacing λ with 3 in Eq. (4.1),

$$c_{ij} = \frac{2^{2i+2j}}{\Gamma(2i+6j+2)} \times \sum_{k_1=0}^{i} \cdots \sum_{k_j=0}^{k_{j-1}} \left\{ \frac{\Gamma(2k_j+6)}{\Gamma(2k_j+2)} \frac{\Gamma(2k_{j-1}+12)}{\Gamma(2k_{j-1}+8)} \cdots \times \frac{\Gamma(2k_1+6j)}{\Gamma(2k_1+6(j-1)+2)} \right\}. \tag{4.14}$$

It does not seem at all unlikely that Eq. (4.13) [or (4.14), which is of even simpler form] could be reduced. It is similar in nature to Eq. (4.6), which gives c_{ij} for $\alpha = 2$. In fact, for all even-power potentials, Φ reduces to a single series with coefficients which are in principle calculable in the sense that the multiple sums in c_{ij} for fixed i and j could be done.[10] Failing analytical results (which should be possible to obtain for $\alpha = 2$), a numerological approach might prove fruitful.

5. THE IRREGULAR SOLUTION

For general values of L, a second linearly independent solution of Eq. (2.1) can be constructed which behaves as $r^{-L+1/2}$ near $r \cong 0$. This is designated $\psi(r)$ and is written as[11]

$$\psi(r) = \mathscr{K} r^{L+1/2} \Phi(r) \ln r + r^{-L+1/2} \Theta(r), \tag{5.1}$$

with $\mathscr{K}$ a constant and $\Theta(0) = 1$. A constant multiple of $r^{L+1/2}\Phi(r)$ may be added to the definition without changing the essentials of the treatment to follow. The function $\psi(r)$ with $L = \frac{1}{2}$ and $\mathscr{K} = 0$ is, as previously noted, important for one-dimensional problems in the guise of the "even" solutions, meaning here that they do not vanish at the origin.

Inserting Eq. (5.1) into Eq. (2.1), and using the fact that $\chi = r^{L+1/2}\Phi$ also satisfies the latter,

$$\left[\frac{d^2}{dr^2} + \frac{1-2L}{r}\frac{d}{dr} - ar^{2\lambda-2} + \mathscr{E}\right]\Theta + 2\mathscr{K}\sum_{i,j} d_{ij}(2i+2j\lambda+L) r^{2i+2j\lambda+2L-2} = 0. \tag{5.2}$$

Expanding Θ in a double power series ($f_{0,0} = 1$),

$$\Theta(r) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_{ij} r^{2i+2j\lambda}, \tag{5.3}$$

leads to

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r^{2i+2j\lambda-2}\{4(i+j\lambda)(i+j\lambda-L) f_{ij} - a f_{i,j-1} + \mathscr{E} f_{i-1,j}\} + 2\mathscr{K} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} r^{2i+2j\lambda+2L-2}(2i+2j\lambda+L) d_{ij} = 0. \tag{5.4}$$

Here λ will be assumed irrational as a convenience. If it is not, then modifications must be made as was done for the regular solution. We now focus upon L to see which powers of r are independent. If L is such that there exists no pair i' and j' for which $i' + j'\lambda = L$, then $\mathscr{K}$ must be set equal to zero in Eq. (5.4), giving

$$4(i+j\lambda)(i+j\lambda-L)f_{ij}=af_{i,j-1}-\mathscr{E}f_{i-1,j}. \tag{5.5}$$

This recursion relation has already been solved in Secs. 2 and 3 since it differs from Eq. (2.8) only by reversing the sign of L. Thus,

$$f_{ij}=d_{ij}(L\to -L)=\left(\frac{-\mathscr{E}}{4}\right)^i\left(\frac{a}{4}\right)^j c_{ij}(L\to -L). \tag{5.6}$$

In the event that there *does* exist a pair i' and j' for which $L=i'+j'\lambda$, then the indices are shifted in the last term of Eq. (5.4):

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}r^{2i+2j\lambda-2}\{4(i+j\lambda)(i+j\lambda-L)f_{ij}-af_{i,j-1}+\mathscr{E}f_{i-1,j}\}+2\mathscr{K}\sum_{i=i'}^{\infty}\sum_{j=j'}^{\infty}r^{2i+2j\lambda-2}(2i+2j\lambda-L)d_{i-i',j-j'}=0. \tag{5.7}$$

There results the recursion

$$4(i+j\lambda)(i+j\lambda-L)f_{ij}-af_{i,j-1}+\mathscr{E}f_{i-1,j}+2\mathscr{K}(2i+2j\lambda-L)d_{i-i',j-j'}=0. \tag{5.8}$$

The constants a and $\mathscr{E}$ are eliminated from Eq. (5.8) by using Eq. (2.9) and the definitions

$$f_{ij}=\left(\frac{-\mathscr{E}}{4}\right)^i\left(\frac{a}{4}\right)^j e_{ij}, \tag{5.9}$$

$$\mathscr{K}=\left(\frac{-\mathscr{E}}{4}\right)^{i'}\left(\frac{a}{4}\right)^{j'}\mathscr{J}. \tag{5.10}$$

Thus,

$$(i+j\lambda)(i+j\lambda-L)e_{ij}-e_{i,j-1}-e_{i-1,j}+2\mathscr{J}(2i+2j\lambda-L)c_{i-i',j-j'}=0. \tag{5.11}$$

Since c_{ij} and e_{ij} are zero for $i<0$ or $j<0$, Eq. (5.11) for $i<i'$ *or* $j<j'$ reduces to

$$(i+j\lambda)(i+j\lambda-L)e_{ij}=e_{i,j-1}+e_{i-1,j} \tag{5.12}$$

so that, once again, for these limits on i, and j,

$$e_{ij}=c_{ij}(L\to -L). \tag{5.13}$$

For $(i,j)=(i',j')$ in Eq. (5.11), using $c_{00}=1$, an equation is obtained wherein $e_{i'j'}$ has vanished:

$$2L\mathscr{J}=e_{i',j'-1}+e_{i'-1,j'}=c_{i',j'-1}(L\to -L)+c_{i'-1,j'}(L\to -L). \tag{5.14}$$

This yields an exact specification of $\mathscr{J}$ [or $\mathscr{K}$ through Eq. (5.10)] in terms of known quantities. In turn, $e_{i'j'}$ is indeterminate; this just corresponds to the previously noted fact that $\psi(r)$ may contain an arbitrary multiple of the regular solution $\chi(r)$. Once $e_{i'j'}$ has been fixed, Eq. (5.11) provides the means of obtaining explicitly all of the remaining e_{ij}, although we do not pursue this further here.

The analogy to ordinary Bessel functions is quite strong here, and of course $\psi(r)$ and $\chi(r)$ reduce to these in special cases. For instance, when we demanded that L and λ were such that the equation $L=i'+j'\lambda$ could be satisfied for *no* i' and j', then ψ was obtained from χ by just reversing the sign of L everywhere. If the scalar a is set equal to zero, we see from Eq. (3.5) that χ reduces to the standard Bessel function J_L, and the corresponding condition is that L is not an integer. The irregular solution *can* be taken as J_{-L} (which is our χ), but this is inconvenient because of the unique behavior when L assumes integral values. Then, due to the presence of the Γ function in Eq. (3.5), J_L and J_{-L} are no longer linearly independent. Instead, one may take as the second standard solution an L-dependent linear combination of J_L and J_{-L} which behaves continuously as L approaches integral values. The next step for the case where a is nonzero appears to be to find such a (second) standard solution. In particular, it is clear that a generalization of the factor $\Gamma(L+1)$ in Eq. (3.5) must be constructed to make $\psi(r)$ and $\chi(r)$ entire in L. Perhaps the necessary generalization of the Γ function to two dimensions may be constructed in terms of double integrals, but we have not yet been able to find such.

If $L=\frac{1}{2}$ and λ is such that the logarithmic terms do not appear, a formula for e_{ij} may be obtained that is similar to Eq. (4.1).

6. GENERALIZATION TO ARBITRARY NUMBER OF POWERS

One of the most fascinating aspects of the multiple series representation is the immediate generalization that occurs when the potential of Eq. (2.1) contains a sum of n power-law terms. Making the definitions $a_0=-\mathscr{E}$ and $\lambda_0=1$ for notational expendiency, the equation to be solved is

$$\left[\frac{d^2}{dr^2}-\frac{L^2-\frac{1}{4}}{r^2}-\sum_{i=0}^{n}a_i r^{2\lambda_i-2}\right]\chi(r)=0, \tag{6.1}$$

where $\lambda_i>0$ and $\lambda_i\neq\lambda_j$ for $i\neq j$. The solution is then of the form

$$\chi(r)=r^{L+1/2}\Phi(r). \tag{6.2}$$

$$\Phi(r)=\sum_{i=0}^{\infty}\sum_{i_1=0}^{\infty}\cdots\sum_{i_n=0}^{\infty}d(i_0,i_1,\ldots,i_n)r^{2(i_0\lambda_0+i_1\lambda_1+\cdots+i_n\lambda_n)}. \tag{6.3}$$

The scale parameters are factored from d by defining

$$d(i_0,i_1,\ldots,i_n)=\left(\frac{a_0}{4}\right)^{i_0}\left(\frac{a_1}{4}\right)^{i_1}\cdots\left(\frac{a_n}{4}\right)^{i_n}c(i_0,i_1,\ldots,i_n). \tag{6.4}$$

This leads, from Eq. (6.1), to the simple relation

$$\left(\sum_{j=0}^{n}i_j\lambda_j\right)\left(\sum_{j=0}^{n}i_j\lambda_j+L\right)c(i_0,i_1,\ldots,i_n)=c(i_0-1,i_1,\ldots,i_n)+c(i_0,i_1-1,\ldots,i_n)+\cdots+c(i_0,i_1,\ldots,i_n-1). \tag{6.5}$$

For $n=1$, Eq. (6.5) has already been solved. For higher n, it may be solved iteratively precisely as before. To see this, we set $i_2=i_3=\cdots i_n=0$. Since $c(i_0,i_1,\ldots,i_n)=0$ for any of the indices less than zero, Eq. (6.5) reduces to

$$\left(\sum_{j=0}^{1}i_j\lambda_j\right)\left(\sum_{j=0}^{1}i_j\lambda_j+L\right)c(i_0,i_1,0,\ldots,0)=c(i_0-1,i_1,0,\ldots,0)+c(i_0,i_1-1,0,\ldots0). \tag{6.6}$$

This is Eq. (2.10), and, consequently, $c(i_0,i_1,0,\ldots,0)$ is given by c_{i_0,i_1} in Eq. (3.11).

Now let i_2 also be nonzero. Then

$$\left(\sum_{j=0}^{2}i_j\lambda_j\right)\left(\sum_{j=0}^{2}i_j\lambda_j+L\right)c(i_0,i_1,i_2,0,\ldots,0)=c(i_0-1,i_1,i_2,0,\ldots,0)=c(i_0,i_1-1,i_2,0,\ldots0)+c(i_0,i_1,i_2-1,0,\ldots,0). \tag{6.7}$$

By setting $i_0 = i_1 = 0$, we have a two-term recursion which can be immediately solved to yield $c(0,0,i_2,0,...,0)$. By setting $i_0 = 1$ and $i_1 = 0$, we have a three-term recursion involving $c(1,0,i_2,0,...,0)$, $c(0,0,i_2,0,...,0)$, and $c(1,0,i_2 - 1,0,...,0)$, which allows us to find the form of $c(1,0,i_2,0,...,0)$. By bootstrapping in this fashion, all of the $c(i_0,i_1,i_2,0,...,0)$ can be obtained. Then one continues on in like manner to find all of the $c(i_0,i_1,...,i_n)$. As is evident, the final form will be too unwieldly to write to write down explicitly. The important point is that all of the coefficients can be determined exactly from Eq. (6.5).

Here is where the value of a partitioning-theoretic analogy becomes important in describing the structure of $c(i_0,i_1,...,i_n)$. It is not too difficult to see (or verify) that each coefficient will be a sum of many terms, and that each term will involve a product of two generalized shifted factorials as defined in Eq. (3.9); these will have factors incremented by $\lambda_0(=1), \lambda_1 ... \lambda_n$ in various orders, but the total number of increments by λ_0 will be i_0, that by λ_1 will be i_1, etc. For example, if $n = 2$, then $c(1,1,1)$ is

$$\begin{aligned} c(1,1,1) = {} & [(0|\lambda_0{}^1\lambda_1{}^1\lambda_2{}^1)(L\,|\lambda_0{}^1\lambda_1{}^1\lambda_2{}^1)]^{-1} \\ & + [(0|\lambda_0{}^1\lambda_2{}^1\lambda_1{}^1)(L\,|\lambda_0{}^1\lambda_2{}^1\lambda_1{}^1)]^{-1} \\ & + [(0|\lambda_1{}^1\lambda_0{}^1\lambda_2{}^1)(L\,|\lambda_1{}^1\lambda_0{}^1\lambda_2{}^1)]^{-1} \\ & + [(0|\lambda_1{}^1\lambda_2{}^1\lambda_0{}^1)(L\,|\lambda_1{}^1\lambda_2{}^1\lambda_0{}^1)]^{-1} \\ & + [(0|\lambda_2{}^1\lambda_0{}^1\lambda_1{}^1)(L\,|\lambda_2{}^1\lambda_0{}^1\lambda_1{}^1)]^{-1} \\ & + [(0|\lambda_2{}^1\lambda_1{}^1\lambda_0{}^1)(L\,|\lambda_2{}^1\lambda_1{}^1\lambda_0{}^1)]^{-1}. \end{aligned} \tag{6.8}$$

Each term corresponds to exactly one of the distinct orders of incrementing once each by λ_0, λ_1, and λ_2. In general, each term in $c(i_0,i_1,...,i_n)$ corresponds to one of the distinct ways in which i_0 increments of λ_0, i_1 of $\lambda_1 \cdots$ and i_n of λ_n may be ordered, and thus has a one-to-one correspondence with the partitions of an interval of total length $\Sigma_{j=0}^{n} i_j\lambda_j$ into i_0 intervals of length λ_0, etc.

Now the role of the ratios λ_j/λ_k has to be considered. With $n = 1$, $\lambda_1/\lambda_0 = \lambda_1$ was first considered to be irrational; then this restriction was relaxed, and when λ_1 was rational, the coefficient d'_{i_0,i_1} of each unique power of r was composed of a sum of d_{l_0,l_1}'s. This sum extended over all (l_0,l_1) for which

$$i_0 + i_1\lambda_1 = l_0 + l_1\lambda_1. \tag{6.9}$$

In interval parlance, d'_{i_0,i_1} had one term for each of the possible ordered partitionings of an interval of length $i_0 + i_1\lambda_1$ into subintervals of lengths unity and λ_1, with no other restrictions on the numbers of each type.

Again, the general case may be regarded similarly. The various powers of r that arise may be such that there are high degrees of degeneracy among the exponents $2(i_0\lambda_0 + ...i_n\lambda_n)$ in Eq. (6.3). The coefficient of each *unique* power of r then involves a sum over the coefficients $d(l_0,...,l_n)$ for all l's such that

$$\sum_{k=0}^{n} i_k\lambda_k = \sum_{k=0}^{n} l_k\lambda_k. \tag{6.10}$$

(Here we assume that the ranges of the i's, but not the l's, have been chosen to provide a unique indexing of powers of r, as in Sec. 3 B.) Thus, the coefficient of each unique power is a sum over many terms, each of which corresponds to one of the possible ordered partitionings of an interval of length $\Sigma_{k=0}^{n} i_k\lambda_k$ into subintervals of lengths $\lambda_0,\lambda_1,...,\lambda_n$, regardless of the relative numbers of each length.

At the very least, this clarifies the counting procedure associated with each coefficient. At best, consequences of the partitioning analogy might exist which could shed some more light on the structure of the standard solutions.

The existence of these solutions yields some insight into the form of perturbed wavefunctions as well. Taking $n = 2$, the eigenvalue problem is (suppressing the L dependence of $\mathscr{E}_m$ and χ_m)

$$\left[\frac{d^2}{dr^2} - \frac{L-\frac{1}{4}}{r} - a_1 r^{2\lambda_1 - 2} - a_2 r^{2\lambda_2 - 2} + \mathscr{E}_m(a_1,a_2)\right] \times \chi_m(r) = 0. \tag{6.11}$$

The complete solution is, with N_m a normalization constant,

$$\chi_m(r) = N_m r^{L+1/2} \sum_{i_0 i_1 i_2} c(i_0,i_1,i_2) \left(-\frac{\mathscr{E}_m}{4}\right)^{i_0} \left(\frac{a_1}{4}\right)^{i_1} \left(\frac{a_2}{4}\right)^{i_2} \times r^{2(i_0 + i_1\lambda_1 + i_2\lambda_2)}. \tag{6.12}$$

If the problem with $a_2 = 0$ is exactly solvable, the usual procedure is to expand the wavefunction and energy in powers of a_2 (or some multiple). We can see imediately that the primary functional distinction between, say, the zeroth- and first-order wavefunctions, will be that the former contains only terms with $i_2 = 0$, while the latter will have ones with both $i_2 = 0$ and $i_2 = 1$. We must, however, beg the questions here of whether or not the perturbation series is convergent and whether or not the total energies and normalization constants can be expanded analytically in some multiple of a_2.

7. DISCUSSION

It has been demonstrated that expansions around the origin can be constructed for formal solutions of the radial Schrödinger equation with power-law potentials. These multiple-series generalizations of the standard Bessel functions are, properly speaking, expansions in the *linear coefficients* of the differential equation. They are not generally analytic in the radial variable, but the nature of the nonanalyticity is very visible.

It is also quite clear that a great deal has been left unbroached. In particular, it remains to be seen whether or not asymptotic expansions may be explicitly developed and matched to the power series expansions.

There is the possibility that the series form may be represented more compactly. As is apparent from the discussion of the structure of the coefficients and the connection with partitioning an interval into subintervals of different lengths, this would require a generalization of the Γ function defined such that its inverse vanished only on points of a two-dimensional (or higher) lattice. If such is available, then c_{ij} [or $c(i_0,i_1,...,i_n)$] might be expressed as a Mellin–Barnes type integral, the kernel having one pole for each of the terms in c_{ij}. Each pole would correspond to one of the different routes from (0,0) to (i,j) along a lattice. Presumably, the simplicity of the recursion relation [Eq. (2.10)] implies some correspondingly simple functional dependence of each kernel upon i and j.[12]

It should be mentioned that solutions can also be writ-

ten in terms of a two-dimensional generalized Laplace transform.[13] The form of the expansion suggests that χ can be expressed as

$$\chi(r) = r^{L+1/2}\int_{c_s}\int_{c_t} ds dt\, v(s,t)e^{sr+tr^{\lambda}}, \tag{7.1}$$

where $v(s,t)$ is to be determined by inserting Eq. (7.1) into Eq. (2.1), converting powers of r into derivatives in s and t, then integrating by parts to determine the equation satisfied by v. The contours c_s and c_t in the complex s and t planes are chosen so that all semi-integrated terms vanish. In the present circumstances, we already know the expanded version of v. From Eq. (2.7) and the integral[14]

$$\zeta^{\mu} = \frac{\Gamma(\mu+1)}{2\pi i}\int_c z^{-\mu-1}e^{\zeta z}\,dz, \tag{7.2}$$

it follows that

$$v(s,t) = (2\pi i)^{-2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} d_{ij}\frac{(2i)!}{s^{2i+1}}\frac{(2j)!}{t^{2j+1}}. \tag{7.3}$$

This allows i and j to be associated with two different variables, with easy generalization to multiple potentials.

There are also a couple points of interest about quantization of the energy. The first is that the "collapsing" of the double series for rational λ is mirrored by a similar collapsing of the WKB equation for the energy eigenvalues. This was shown by Feldman, Fulton, and Devoto[15] for the lowest-order equation in $\hbar$, but the same is true if higher-order corrections are included.

The other point pertains to the fact that certain artificial multiple potentials admit of exact square-integrable eigensolutions. This was discussed for a generalized anharmonic potential $U(x) = \Sigma_{n=1}^{2N-1} b_n x^{2n}$ by Magyari.[16] The essential point is that, with certain restrictions on the b_n, $\chi(x)$ can be put into the form of a polynomial in x times an exponential $\exp[-\frac{1}{2}\Sigma_{n=1}^{N} a_n x^{2n}]$. This is not restricted to integral or even-integral polynomials, at least when the radial problem is concerned. To see this, we take a wavefunction of the form

$$\chi(r) = r^{L+1/2}f(r)e^{-S(r)}, \tag{7.4}$$

where $(\mu_i > 0,\ a_i > 0)$

$$S(r) = \sum_{i=1}^{n} a_i r^{\mu_i}. \tag{7.5}$$

The function f must then satisfy

$$f'' + \left[-2S' + \frac{2L+1}{r}\right]f' + \left[(S')^2 - S'' - \frac{2L+1}{r}S' - U(r) + \mathscr{E}\right]f = 0, \tag{7.6}$$

with $f(0) = 1$. $U(r)$ is then much restricted by the requirement that f be a finite sum of various powers of r. For instance, if μ_n is the largest exponent in Eq. (7.5) and $\mu_n > 2$, then it is seen that the potential must have a term $a_n{}^2\mu_n{}^2 r^{2\mu_n-2}$. One of the simplest examples is

$$\chi(r) = r^{L+1/2}\exp(-ar - br^{\mu}) \tag{7.7}$$

leading to

$$U(r) - \mathscr{E} = -a\frac{2L+1}{r} - b\mu(\mu+2L)r^{\mu-2} + 2abr^{\mu-1} + b^2\mu^2 r^{2\mu-2} - a^2. \tag{7.8}$$

This aspect is not considered further, but may be interesting because it gives specific cases for which the multiple expansions in Sec. 6 can in principle be rearranged into an exponentially decreasing factor times a finite sum of (inequivalent) powers of r.

ACKNOWLEDGMENT

This work was supported by National Science Foundation Grant CHE80-12399.

[1] A. Joseph, Int. J. Quant. Chem. **1**, 615 (1967).
[2] C. Quigg and J. L. Rosner, Phys. Rep. **56**, 167 (1979).
[3] J. P. Gazeau, Phys. Lett. A **75**, 159 (1980).
[4] B. R. Johnson, J. Math. Phys. **21**, 2640 (1980).
[5] H. J. W. Müller-Kirsten, G. E. Hite, and S. K. Bose, J. Math. Phys. **20**, 1878 (1979).
[6] A. F. Antippa and A. J. Phares, J. Math. Phys. **19**, 308 (1978); A. J. Phares, *ibid.* **19**, 2239 (1978); P. A. Maurone and A. J. Phares, *ibid.* **21**, 830 (1980).
[7] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1977), 3rd ed., Chap. V.
[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965).
[9] H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969).
[10] O. Brander, Phys. Lett. A **75**, 471 (1980), discusses a limit-case nature of the asymptotic behavior for $\alpha = 4$, which does not seem to be reflected here.
[11] F. Brauer and J. A. Noel, *Ordinary Differential Equations* (Benjamin, New York, 1967), pp. 151–156.
[12] A related idea occurs in the "generalized Bessel functions" in K. I. Gross and R. A. Kunze, J. Funct. Anal. **22**, 73 (1976); **25**, 1 (1977); W. J. Holman, III, J. Math. Phys. **21**, 1977 (1980).
[13] W. Bührung, J. Math. Phys. **18**, 1121 (1977).
[14] A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, Chap. 1.
[15] G. Feldman, T. Fulton, and A. Devoto, Nucl., Phys. B **154**, 441 (1979).
[16] E. Magyari, Phys. Lett. A **81**, 116 (1981).

Monodromy preserving deformation of ramified solutions to $(\Delta - m^2)u = 0$

M. Jimbo, M. Kashiwara, and T. Miwa
Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, Japan

(Received 24 February 1981; accepted for publication 5 June 1981)

Monodromy preserving deformation is discussed for solutions of $(\Delta - m^2)u = 0$, which are axially symmetric and with a given monodromy around a circle in $\mathbb{R}^3$. A holonomic system of linear differential equations is derived for appropriately normalized solutions. Their integral representations are also given, through which the deformation equations reduce to the fifth Painlevé equation. This problem can be related to the scattering of a spherical wave by a circular plate.

PACS numbers: 03.65.Ge, 02.30.Hq

1. INTRODUCTION

Recent progress in monodromy preserving deformations has found fruitful applications to various problems of mathematical physics.[1-4] As they suggest, further researches would bring a still enlarged list of problems solvable in terms of the Painlevé transcendents and their generalizations. This paper aims to present another such example related to the solutions of the Euclidean Klein–Gordon equation

$$(\Delta - m^2)u = 0. \tag{1.1}$$

Let C be a circle on the xy plane in $\mathbb{R}^3$, with its center at the origin. Suppose one wants to find multivalued solutions of (1.1) on $\mathbb{R}^3 - C$, which change their sign when analytically continued around C. Under appropriate growth conditions at C and at ∞, along with the symmetry requirement around the z axis

$$(x\partial_y - y\partial_x)u = 0, \tag{1.2}$$

there are only two such linearly independent solutions. If $m = 0$, these are simply

$$\left.\begin{matrix} u(x, y, z) \\ \tilde{u}(x, y, z) \end{matrix}\right\} = \frac{1}{[x^2 + y^2 + (z \mp ia)^2]^{1/2}}, \tag{1.3}$$

where $a > 0$ is the radius of C. If $m > 0$, the solutions are no longer elementary. The problem is to find a holonomic system of linear differential equations that characterizes the suitably normalized solutions, and to describe its deformation with respect to a.

This type of problem has been treated by Sato *et al.*[5] in a two-dimensional situation, where the circle C is replaced by a finite number of points on the plane. Originally the case of two points goes further back to the work of Myers[6] on the wave scattering in $\mathbb{R}^2$ by a finite strip. As early as in 1965 he showed that the effect of deforming the length of the strip is described by a Painlevé equation of the third kind. In a similar way the present problem in three dimensions also admits of a physical interpretation by putting $m = ik \in i\mathbb{R}$. Consider the scattering by a circular plate $P = \{x^2 + y^2 \leqslant a^2, z = 0\}$ of the spherical wave in $\mathbb{R}^3$, emitted from a point source (x^0, y^0, z^0). Let $u_D(x, y, z)$, $u_N(x, y, z)$ denote the unique solutions, satisfying the Dirichlet or the Neumann condition on P,

$$u_D|_P = 0, \quad \left.\frac{\partial u_N}{\partial z}\right|_P = 0,$$

respectively, and the Sommerfeld radiation condition

$$|u| = O\left(\frac{1}{r}\right), \quad \left|\frac{\partial u}{\partial r} - iku\right| = O\left(\frac{1}{r^2}\right), \quad r = (x^2 + y^2 + z^2)^{1/2} \to \infty.$$

If the source is on the z axis $x^0 = 0$, $y^0 = 0$, then $\partial u_D/\partial a$, $\partial u_N/\partial a$ provide solutions (without having the source) of (1.1) with the correct monodromy property around $C = \partial P$ and the axial symmetry (1.2).

In the problem of Myers, the deformation theory is reduced to that of an ordinary differential equation (whence appears the Painlevé equation). This is also the case in three dimensions. By introducing the integral transformation

$$u(x, y, z) = \int \frac{e^{-m[x^2 + y^2 + (z - t)^2]^{1/2}}}{[x^2 + y^2 + (z - t)^2]^{1/2}} v(t)\, dt, \tag{1.4}$$

the holonomic system for the normalized solutions is transformed into a 2×2 system of ordinary differential equations

$$\frac{d}{dt}\mathbf{v}(t) = \left(\frac{A_+}{t - ia} + \frac{A_-}{t + ia} + B\right)\mathbf{v}(t). \tag{1.5}$$

Accordingly the deformation equations reduce to a Painlevé equation of the fifth kind. In this sense both Myers' and the present problems are no more than deformation theories in one dimension. To find an "essentially higher dimensional" example is a problem of great interest yet to be solved.

The plan of this paper is as follows.

In Sec. 2 the problem is formulated in a slightly generalized way. Namely the circle in $\mathbb{R}^3$ and the monodromy -1 are replaced here by an $(n-2)$-dimensional sphere in $\mathbb{R}^n$ and by a phase factor $-e^{2\pi i l}$, respectively.

In Sec. 3 the holonomic system for the normalized solutions is derived, and its reduction to (1.5) is discussed. For $n = 3$ these results are shown to be also valid if one replaces (1.2) by

$$(x\partial_y - y\partial_x)u = i\nu u$$

with ν an integer. In Sec. 4 the monodromy group for (1.5) and the choice of the integration contour in (1.4) are determined in the three-dimensional case. Remarkably, when the monodromy of $u(x)$ is -1, the relevant Painlevé transcendent coincides with the one related to the random matrices.[2] It is worth mentioning that in this case the associated τ function decays exponentially like[7,8]

$$\tau(a)/\tau(0) = ce^{-S/2\pi}a^{-1/4}[1 + O(a^{-1})] \quad (a \to \infty),$$

 0022-2488/81/112581-07$01.00

where $c = 2^{1/12}e^{3\zeta'(-1)}$ and $S = \pi a^2$ is the area of the disk P. Finally in Sec. 5 the massless case is discussed briefly and explicit formulas for the normalized solutions are given.

2. STATEMENT OF THE PROBLEM

Let C be an $(n-2)$-dimensional sphere of radius a in the Euclidean space $\mathbb{R}^n$;

$$C:\ x_1^2 + \cdots + x_{n-1}^2 = a^2,\quad x_n = 0.$$

It is the intersection of two light cones $s(x) = 0$ and $\bar{s}(x) = 0$, where

$$s(x) = x_1^2 + \cdots + x_{n-1}^2 + (x_n - ia)^2,$$
$$\bar{s}(x) = x_1^2 + \cdots + x_{n-1}^2 + (x_n + ia)^2.$$

The aim of the present article is to study solutions of Euclidean Klein–Gordon equation

$$[\mathrm{W1}]\quad (\Delta - m^2)u(x) = 0\quad (\Delta = \partial_1^2 + \cdots + \partial_n^2,\ m \geqslant 0),$$

that have a certain monodromy property around C.

Let l be a real number such that $|l| < \frac{1}{2}$. We consider multivalued solutions to [W1] on $\mathbb{R}^n - C$ satisfying the following conditions.

[W2] In a neighborhood of C, $u(x)$ behaves like

$$u(x) = s(x)^{l-1/2}f(x) + \bar{s}(x)^{-l-1/2}g(x)$$

with real analytic functions $f(x)$ and $g(x)$.

[W3] As $|x| \to \infty$,

$$|u(x)| = \begin{cases} O(e^{-m|x|}) & (m > 0), \\ O(1/|x|^{n-2}) & (m = 0). \end{cases}$$

Let $\gamma u(x)$ denote the analytic continuation of $u(x)$ along a closed path shown in Fig. 1.

Then [W2] implies the monodromy property of $u(x)$:

$$\gamma u(x) = -e^{2\pi i l}u(x). \tag{2.1}$$

A solution $u(x)$ to [W1] satisfying [W2] and [W3] is uniquely determined if we specify $f(x)|_C$ and $g(x)|_C$. In the sequel we shall consider the simplest case, where $f(x)$ and $g(x)$ are constant on C. This amounts to imposing on $u(x)$ the axial symmetry

$$[\mathrm{W4}]\quad (x_i\partial_j - x_j\partial_i)u(x) = 0\quad (1 \leqslant i, j \leqslant n-1).$$

The space W of solutions to [W1] satisfying [W2]~[W4] is two-dimensional. In fact, one can show the existence of the following normalized basis $u_l(x)$ and $\bar{u}_{-l}(x)$:

$$u_l(x) = s(x)^{l-1/2}f_l(x) + \bar{s}(x)^{-l+1/2}\bar{g}_l(x)$$
$$\bar{u}_{-l}(x) = s(x)^{l+1/2}g_{-l}(x) + \bar{s}(x)^{-l-1/2}\bar{f}_{-l}(x)$$

with the normalization $f_l(x)|_C = 1$ and $\bar{f}_{-l}(x)|_C = 1$.

Our goal is to derive a holonomic system of linear differential equations for the above basis and to discuss its deformation with respect to the radius a.

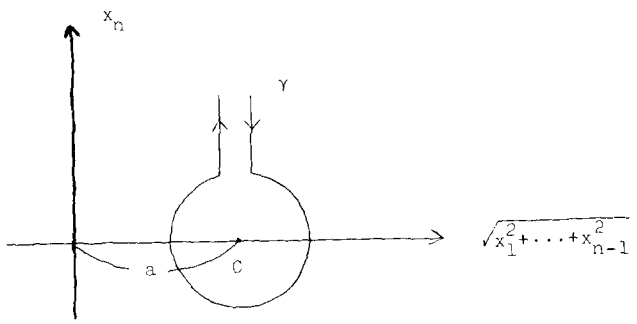


FIG. 1. The path γ in (2.1).

3. HOLONOMIC SYSTEM FOR THE NORMALIZED BASIS

Our strategy in deriving the holonomic system for $u_l(x)$ and $\bar{u}_{-l}(x)$ is as follows. Suppose we find a pth order differential operator $P(x, \partial)$ that has the following property:

[A] If $u(x)$ solves [W1] and [W4], then so does $P(x, \partial)u(x)$.

Then $P(x, \partial)u_l(x)$ and $P(x, \partial)\bar{u}_{-l}(x)$ are linear combinations of $\partial_n^j u_l(x)$ $(0 \leqslant j \leqslant p)$ and $\partial_n^k \bar{u}_{-l}(x)$ $(0 \leqslant k \leqslant p)$.

In fact, if we denote by $w(x)$ either $P(x, \partial)u_l(x)$ or $P(x, \partial)\bar{u}_{-l}(x)$, then $w(x)$ has the development

$$w(x) = s(x)^{l-1/2} \sum_{j \geqslant -p, k \geqslant 0} f_{jk}s(x)^j \bar{s}(x)^k + \bar{s}(x)^{-l-1/2} \sum_{j \geqslant 0, k \geqslant -p} g_{jk}s(x)^j \bar{s}(x)^k. \tag{3.1}$$

If we subtract a suitable linear combination of $\partial_n^j u_l(x)$ $(0 \leqslant j \leqslant p)$ and $\partial_n^k \bar{u}_{-l}(x)$ $(0 \leqslant k \leqslant p)$ from $u(x)$, the resulting solution $\tilde{w}(x)$ has an expression like (3.1) with $f_{-p0} = \cdots = f_{00} = 0$ and $g_{0-p} = \cdots = g_{00} = 0$. Since $\tilde{w}(x)$ solves [W1], we have $f_{jk} = 0$ ($-p \leqslant j \leqslant 0$, all k) and $g_{jk} = 0$ (all j, $-p \leqslant k \leqslant 0$) for $\tilde{w}(x)$. Along with [W3] this implies that $\tilde{w}(x)$ is identically zero.

One can verify that the operators L and M enjoy the property [A].

$$L = \vartheta\partial_n - m^2x_n.$$
$$M = \vartheta^2 + (n-2)\vartheta - m^2(x_1^2 + \cdots + x_n^2),$$

where $\vartheta = x_1\partial_1 + \cdots + x_n\partial_n$.

Applying the above argument with $P = L$, M and $\partial_a = \partial/\partial a$, we have

Theorem 3.1: The normalized basis $u_l(x)$ and $\bar{u}_{-l}(x)$ satisfies the following system of linear partial differential equations with unknown 2×2 matrix coefficients F, G, J, K, and E, which are independent of x.

$$L\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix} = \left[\begin{pmatrix} ia & \\ & -ia \end{pmatrix}\partial_n^2 + F\partial_n + G\right]\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix}, \tag{3.2}$$

$$M\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix} = (-a^2\partial_n^2 + J\partial_n + K)\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix}, \tag{3.3}$$

$$a\partial_a\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix} = \left[\begin{pmatrix} -ia & \\ & ia \end{pmatrix}\partial_n + E\right]\begin{pmatrix} u_l \\ \bar{u}_{-l} \end{pmatrix}. \tag{3.4}$$

These equations along with [W1] and [W4] constitute a holonomic system for the normalized basis. In order to compute the compatibility condition for this system we reduce it to a system of ordinary differential equations.

Let $K_\nu(z)$ be the modified Bessel function of the second kind. We set

$$K(x) = (x_1^2 + \cdots + x_n^2)^{-(n-2)/4} K_{(n-2)/2}(m(x_1^2 + \cdots + x_n^2)^{1/2}),$$

and consider an integral transformation

$$u(x) = \int K(x_1, \ldots, x_{n-1}, x_n - t)v(t)\,dt.$$

Since the kernel $K(x_1, \ldots, x_{n-1}, x_n - t)$ satisfies [W1] and [W4], an operator $P(x, \partial)$ of the property [A] is transformed into an ordinary differential operator $Q(t, \partial_t)$ acting on $v(t)$. In fact, we have the following table of correspondence:

$P(x, \partial)$	$Q(t, \partial_t)$
∂_n	∂_t
L	$t(\partial_t^2 - m^2) - (n-3)\partial_t$
M	$[t(\partial_t^2 - m^2) - (n-3)\partial_t]t$

Thus, by setting

$$\mathrm{u}(x) = \int K(x_1, \ldots, x_{n-1}, x_n - t)\mathrm{v}(t)dt,$$

$$\mathrm{u}(x) = \begin{pmatrix} u_l(x) \\ \bar{u}_{-l}(x) \end{pmatrix}, \quad \mathrm{v}(t) = \begin{pmatrix} v_l(t) \\ \bar{v}_{-l}(t) \end{pmatrix}, \tag{3.5}$$

we can transform (3.2)~(3.4) as follows.

$$\left[\begin{pmatrix} t - ia & \\ & t + ia \end{pmatrix}\partial_t^2 - (F + n - 3)\partial_t - G - m^2 t\right]\mathrm{v} = 0. \tag{3.6}$$

$$\{(t^2 + a^2)\partial_t^2 - [J + (n-5)t]\partial_t - K - m^2 t - (n-3)\}$$
$$\times \mathrm{v} = 0. \tag{3.7}$$

$$a\partial_a \mathrm{v} = \left[\begin{pmatrix} -ia & \\ & ia \end{pmatrix}\partial_t + E\right]\mathrm{v}. \tag{3.8}$$

Let us rewrite (3.6) and (3.7) in a simpler form. We set

$$u_l(x) = \sum_{j,k\geq 0} f_{l,jk} s(x)^{l-1/2+j}\bar{s}(x)^k + \sum_{j,k\geq 0} \bar{g}_{l,jk} s(x)^j \bar{s}(x)^{-l+1/2+k} \quad (f_{l,00} = 1),$$

$$\bar{u}_{-l}(x) = \sum_{j,k\geq 0} g_{-l,jk} s(x)^{l+1/2+j}\bar{s}(x)^k + \sum_{j,k\geq 0} \bar{f}_{-l,jk} s(x)^j \bar{s}(x)^{-l+1/2+k} \quad (\bar{f}_{-l,00} = 1). \tag{3.9}$$

Proposition 3.2: Suppose that [W1], [W4], (3.2), and (3.3) admit a solution of the form (3.9) such that

$$-l^2 + (n-3)^2/4 + 16a^4(l^2 - \tfrac{1}{4})g_{-l,00}\bar{g}_{l,00} \neq 0. \tag{3.10}$$

Then (3.6) and (3.7) are equivalent to the following:

$$\partial_t \mathrm{v} = \left(\frac{A_+}{t - ia} + \frac{A_-}{t + ia} + B\right)\mathrm{v}. \tag{3.11}$$

The 2×2 matrices $A_\pm$ and B are of the form

$$A_+ = \begin{pmatrix} l + (n-5)/2 & \xi \\ 0 & 0 \end{pmatrix}, \quad A_- = \begin{pmatrix} 0 & 0 \\ \eta & -l + (n-5)/2 \end{pmatrix},$$

$$B = \begin{pmatrix} \kappa & \lambda \\ \mu & -\kappa \end{pmatrix}, \tag{3.12}$$

with the algebraic constraints

$$B^2 = m^2, \quad [B, A_+ + A_-]_+ + (5-n)B = 0, \tag{3.13}$$

or equivalently

$$\kappa^2 + \lambda\mu = m^2, \quad 2l\kappa + \xi\mu + \eta\lambda = 0. \tag{3.14}$$

In terms of $\xi, \eta, \kappa, \lambda$, and μ, the coefficients F, G, J, K, and E are determined by the following formulas:

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} = \begin{pmatrix} l - \dfrac{n+1}{2} & \xi - 2ia\lambda \\ \eta + 2ia\mu & -l - \dfrac{n+1}{2} \end{pmatrix}, \tag{3.15}$$

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} -iam^2 & \\ & iam^2 \end{pmatrix} - (F+2)B, \tag{3.16}$$

$$J = \begin{pmatrix} 2ia(l-1) & \\ & 2ia(l+1) \end{pmatrix}, \tag{3.17}$$

$$K = \begin{pmatrix} 2iaG_{11} + F_{12}F_{21} - m^2a^2 + [l + (n-3)/2][l - (n-1)/2] & -F_{12} \\ -F_{21} & -2iaG_{22} + F_{12}F_{21} - m^2a^2 + [l - (n-3)/2][l - (n-1)/2] \end{pmatrix}, \tag{3.18}$$

$$E = \begin{pmatrix} l + (n-3)/2 & -F_{12} \\ -F_{21} & -l + (n-3)/2 \end{pmatrix}. \tag{3.19}$$

Proof: Substituting (3.9) into (3.2), (3.3), and (3.4) we obtain the identities (3.17), (3.19), and

$$F = \begin{pmatrix} l - (n+1)/2 & 4a^2(-l + \frac{1}{2})\bar{g}_{l,00} \\ 4a^2(l + \frac{1}{2})g_{-l,00} & -l - (n+1)/2 \end{pmatrix}.$$

Then

$$(3.7) - \begin{pmatrix} t + ia & \\ & t - ia \end{pmatrix} \times (3.6)$$

reads

$$(F+2)\begin{pmatrix} t - ia & \\ & t + ia \end{pmatrix}\partial_t \mathrm{v} + \left[\begin{pmatrix} t + ia & \\ & t - ia \end{pmatrix} \times (G + m^2 t) - K - m^2 t^2 - (n-3)\right]\mathrm{v} = 0.$$

The assumption (3.10) means that $F + 2$ is invertible. Hence we have a system of the form (3.11). Comparing (3.11) with (3.6) and (3.7), we obtain (3.12), (3.13), (3.15), (3.16), and (3.18).

Remark 1: The coefficients $f_{l,jk}$, $\bar{g}_{l,jk}$, $g_{-l,jk}$, and $\bar{f}_{-l,jk}$ in (3.7) and (3.8) are uniquely determined from [W1], [W4], and (3.2). Conversely E, F, G, J, and K are determined in terms of these coefficients.

Remark 2: The constraint (3.13) means that at $t=\infty$, we have formal solutions to (3.11) of the form

$$\mathbf{v}_{\pm}(t)=(\mathbf{v}_{0\pm}+\mathbf{v}_{1\pm}t^{-1}+\cdots)t^{(n-5)/2}e^{\pm mt}.$$

It is known[2,9] that the deformation equations for the system of the form (3.11) reduce to the fifth Painlevé equation. This fact is equally valid although the present normalization (3.12) differs from Refs. 2 and 9.

The compatibility condition between (3.4) and (3.11) gives us the following first order system of nonlinear differential equations.

$$\begin{aligned}
\frac{d\xi}{da}&=\left(2i\kappa+\frac{4l}{a}\right)\xi-2i\left(l+\frac{n-5}{2}\right)\lambda,\\
\frac{d\eta}{da}&=\left(2i\kappa-\frac{4l}{a}\right)\eta+2i\left(-l+\frac{n-5}{2}\right)\mu,\\
\frac{d\kappa}{da}&=2i\lambda\mu+\frac{-\xi\mu+\eta\lambda}{a},\\
\frac{d\lambda}{da}&=\frac{2l}{a}\lambda-2\kappa\left(i\lambda-\frac{\xi}{a}\right),\\
\frac{d\mu}{da}&=-\frac{2l}{a}\mu-2\kappa\left(i\mu+\frac{\eta}{a}\right).
\end{aligned}\tag{3.20}$$

Following Ref. 10, let us introduce the τ function by

$$\frac{d\log\tau(a)}{da}=\frac{\xi\eta}{a}+i(\xi\mu-\eta\lambda)+i(n-5)\kappa.$$

We set

$$\sigma(a)=a\frac{d\log\tau(a)}{da}.$$

The system (3.20) reduces to the following single equation for σ, which is equivalent to a Painlevé equation of the fifth kind.[2,9]

$$\begin{aligned}
\left(\frac{a}{m}\frac{d^2\sigma}{da^2}\right)^2=&\\
&-\left[4\left(\sigma-a\frac{d\sigma}{da}\right)-\frac{1}{2}\left(\frac{1}{m}\frac{d\sigma}{da}\right)^2+2l^2-\frac{(n-5)^2}{2}\right]^2\\
&+\frac{1}{4}\left[\left(\frac{1}{m}\frac{d\sigma}{da}\right)^2+(2l+n-5)^2\right]\\
&\times\left[\left(\frac{1}{m}\frac{d\sigma}{da}\right)^2+(2l-n+5)^2\right].
\end{aligned}$$

Note that the system (3.20) is invariant by the scale transformation $(\xi,\eta,\kappa,\lambda,\mu)\longmapsto(c\xi,c^{-1}\eta,\kappa,c\lambda,c^{-1}\mu)$. Scale invariant quantities such as κ, $\xi\eta$, $\xi\mu$, and $\mu\lambda$ can be recovered from σ by using (3.14) and the following identities.

$$\begin{aligned}
&\frac{d\sigma}{da}=i(\xi\mu-\eta\lambda)+i(n-5)\kappa,\quad \sigma-a\frac{d\sigma}{da}=\xi\eta,\\
&a\frac{d^2\sigma}{da^2}=-4i\kappa\xi\eta+2i\left(l-\frac{n-5}{2}\right)\xi\mu+2i\left(l+\frac{n-5}{2}\right)\eta\lambda.
\end{aligned}$$

Now assume that $n=3$ and consider the equation

$$(x_1\partial_2-x_2\partial_1)u(x)=i\nu u(x)\quad(\nu:\text{integer}),\tag{3.21}$$

instead of [W4]. The above arguments can be traced in this case, as we shall see below. Without loss of generality we can assume that $\nu>0$. Theorem 3.1 is equally valid for the normalized basis $u_{l\nu}(x)$ and $\bar{u}_{-l\nu}(x)$,

$$\begin{aligned}
u_{l\nu}(x)=&(x_1+ix_2)^{\nu}\Bigg(\sum_{j,k\geq 0}f_{l\nu,jk}s(x)^{l-1/2+j}\bar{s}(x)^k\\
&+\sum_{j,k\geq 0}\bar{g}_{l\nu,jk}s(x)^j\bar{s}(x)^{-l+1/2+k}\Bigg)\quad(f_{l\nu,00}=1),\\
\bar{u}_{-l\nu}(x)=&(x_1+ix_2)^{\nu}\Bigg(\sum_{j,k\geq 0}g_{-l\nu,jk}s(x)^{l+1/2+j}\bar{s}(x)^k\\
&+\sum_{j,k\geq 0}\bar{f}_{-l\nu,jk}s(x)^j\bar{s}(x)^{-l+1/2+k}\Bigg)\quad(\bar{f}_{-l\nu,00}=1),
\end{aligned}\tag{3.22}$$

with appropriate matrices F, G, J, K, and E. The kernel of the integral transformation (3.5) is modified to

$$\begin{aligned}
K(x)=&(x_1+ix_2)^{\nu}[(x_1^2+x_2^2+x_3^2)^{1/2})]^{-\nu-1/2}\\
&\times K_{\nu+1/2}(m(x_1^2+x_2^2+x_3^2)^{1/2}).
\end{aligned}$$

The correspondence of operators is as follows:

∂_n	∂_t
L	$t(\partial_t^2-m^2)-\nu\partial_t$
M	$t(\partial_t^2-m^2)t-2\nu t\partial_t-\nu(1-\nu)$.

The transformed equations read

$$\left[\begin{pmatrix}t-ia & \\ & t+ia\end{pmatrix}\partial_t^2-(F+\nu)\partial_t-G-m^2t\right]\mathbf{v}=0,\tag{3.23}$$

$$\begin{aligned}
&\{(t^2+a^2)\partial_t^2-[J-2(1-\nu)t]\partial_t\\
&\quad-K-m^2t^2-\nu(1-\nu)\}\mathbf{v}=0.
\end{aligned}\tag{3.24}$$

Proposition 3.3: Suppose that [W1], (3.21), (3.2), and (3.3) admit a solution of the form (3.22) such that

$$-l^2+\nu^2+16a^4(l^2-\tfrac{1}{4})g_{-l\nu,00}\bar{g}_{l\nu,00}\neq 0.$$

Then (3.23) and (3.24) are equivalent to the system of the form (3.11), where

$$A_+=\begin{pmatrix}l+\nu-1 & \xi\\ 0 & 0\end{pmatrix},\quad A_-=\begin{pmatrix}0 & 0\\ \eta & -l+\nu-1\end{pmatrix},$$

$$B=\begin{pmatrix}\kappa & \lambda\\ \mu & -\kappa\end{pmatrix},$$

with different ξ,η,κ,λ, and μ. The algebraic constraints now read

$$B^2=m^2,\quad [B,A_++A_-]_+-2(1-\nu)B=0.$$

The coefficients E, F, G, J, and K are given by

$$F=\begin{pmatrix}l-2 & \xi-2ia\lambda\\ \eta+2ia\mu & -l-2\end{pmatrix},$$

$$K=\begin{pmatrix}2iaG_{11}+F_{12}F_{21}-m^2a^2+(l+\nu-1)(l-\nu) & -F_{12}\\ -F_{21} & -2iaG_{22}+F_{12}F_{21}-m^2a^2+(l-\nu+1)(l+\nu)\end{pmatrix},$$

$$G = \begin{pmatrix} -iam^2 & \\ & iam^2 \end{pmatrix} - (F + 2 - \nu)B,$$

$$J = \begin{pmatrix} 2ia(l-1) & \\ & 2ia(l+1) \end{pmatrix},$$

$$E = \begin{pmatrix} l+\nu & -F_{12} \\ -F_{21} & -l+\nu \end{pmatrix}.$$

4. INTEGRAL REPRESENTATION OF THE NORMALIZED SOLUTIONS IN THREE DIMENSIONS

In order to give a precise meaning to the integral transformation (3.5), one must specify the integration contour and also the vector $\mathbf{v}(t)$. In the sequel we restrict our discussion to the case $n = 3$, $\nu = 0$.

As a function of t, the integrand of (3.5) contains 4 regular singularities $t = \pm ia$, $x_3 \pm i(x_1^2 + x_2^2)^{1/2}$ and an irregular singularity of rank one at $t = \infty$. The contour Γ and the vector $\mathbf{v}(t)$ should be so chosen as to satisfy the two requirements, namely, (i) the monodromy property (2.1) for $\mathbf{u}(x)$, and (ii) that $\mathbf{u}(x)$ should be regular on the x_3 axis $x_1 = x_2 = 0$. The latter is obviously fulfilled if we choose Γ as in Fig. 2. For this to be possible we require (3.11) to admit a solution $\mathbf{v}(t)$ which has singularities at $t = \pm ia$ but is singlevalued on Γ.

First consider the case $l \neq 0$. Let $V(t) = (\mathbf{v}_1(t), \mathbf{v}_2(t))$ be a matrix with the following monodromy data:

$$V(t) = \hat{V}_{\pm}(t) \cdot (t \mp ia)^{\mathrm{diag}(\pm l - 1, 0)} \quad \text{at} \quad t = \pm ia, \tag{4.1}$$

$$V(t)\begin{pmatrix} 1 & 1 \\ e^{2\pi i l} & 1 \end{pmatrix} = \hat{V}_{\infty}(t) \cdot t^{-1}\begin{pmatrix} e^{mt} & \\ & e^{-mt} \end{pmatrix} \tag{4.2}$$

in the full neighborhood of $t = \infty$. Here $\hat{V}_{\pm}(t)$, $\hat{V}_{\infty}(t)$ are invertible and holomorphic at $t = \pm ia$ or $t = \infty$, respectively. That is, the monodromy group for $V(t)$ is abelian, the Stokes multipliers are trivial, and the only nontrivial monodromy datum in the sense of Ref. 10 is the connection matrix $\begin{pmatrix} 1 & 1 \\ e^{2\pi i l} & 1 \end{pmatrix}$.

Let us verify that the choice $\mathbf{v}(t) = \mathbf{v}_1(t)$ leads to the correct monodromy property (2.1) for $\mathbf{u}(x)$. Regarding $\mathbf{u}(x)$ as a single-valued function on $\mathbb{R}^3 - \{x_1^2 + x_2^2 \leqslant a^2, x_3 = 0\}$, we consider its restrictions $\mathbf{u}^{(\pm)}(0, 0, x_3)$ on the x_3 axis for $x_3 \gtrless 0$. Explicitly they are given by

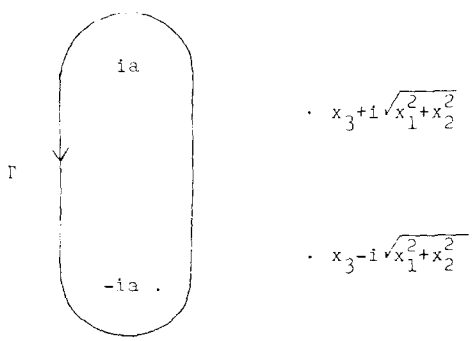
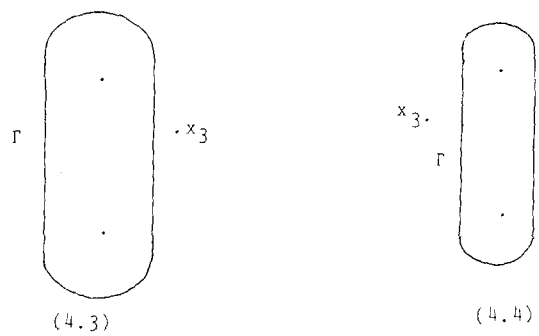


FIG. 2. The integration contour in (3.5).



(4.3) (4.4)

FIG. 3. The mutual positions of x_3 and Γ.

$$\mathbf{u}^{(+)}(0, 0, x_3) = \int_{\Gamma} \frac{e^{-m(x_3 - t)}}{x_3 - t} \mathbf{v}_1(t)\, dt$$

$$= \int_{\tilde{\Gamma}} \frac{e^{-m(x_3 - t)}}{x_3 - t} \mathbf{v}_1(t)\, dt - 2\pi i \mathbf{v}_1(x_3), \tag{4.3}$$

$$\mathbf{u}^{(-)}(0, 0, x_3) = \int_{\Gamma} \frac{e^{m(x_3 - t)}}{-(x_3 - t)} \mathbf{v}_1(t)\, dt$$

$$= -\int_{\tilde{\Gamma}} \frac{e^{m(x_3 - t)}}{x_3 - t} \mathbf{v}_1(t)\, dt + 2\pi i e^{2\pi i(l-1)} \mathbf{v}_1(x_3), \tag{4.4}$$

where $\tilde{\Gamma}$ is the contour obtained by letting x_3 inside Γ. The mutual positions of x_3 and Γ in (4.3), (4.4) are shown in Fig. 3. The monodromy property (2.1) implies $\mathbf{u}^{(+)}(0, 0, x_3) = -e^{-2\pi i l}\mathbf{u}^{(-)}(0, 0, x_3)$. The converse is also true, for one can show without difficulty

Lemma: If $u(x)$ satisfies [W1], [W4], and $u(0, 0, x_3) = 0$, then $u(x) \equiv 0$.

The equation $\mathbf{u}^{(+)}(0, 0, x_3) = -e^{-2\pi i l}\mathbf{u}^{(-)}(0, 0, x_3)$ reduces to the following (4.5).

Proposition 4.1:

$$(1 - e^{2\pi i l})\mathbf{v}_1(t) = \frac{1}{2\pi i} \int_{\Gamma} \frac{e^{m(t-t')} - e^{2\pi i l} e^{-m(t-t')}}{t - t'} \mathbf{v}_1(t')\, dt' \tag{4.5}$$

$$(1 - e^{2\pi i l})\mathbf{v}_2(t) = \frac{-1}{2\pi i} \int_{\Gamma} \frac{e^{m(t-t')} - e^{-m(t-t')}}{t - t'} \mathbf{v}_1(t')\, dt'. \tag{4.6}$$

Proof: Denote by $(1 - e^{2\pi i l})\mathbf{v}_i^*(t)$ $(i = 1, 2)$ the right-hand sides of (4.5)–(4.6), and set $V^*(t) = (\mathbf{v}_1^*(t), \mathbf{v}_2^*(t))$. We shall show $V^*(t) - V(t) = 0$. From the definition, $(1 - e^{2\pi i l})[\mathbf{v}_1^*(t) - \mathbf{v}_1(t)]$ is given by (4.5) with $\tilde{\Gamma}$ in place of Γ, so that $V^*(t) - V(t)$ is entire holomorphic. On the other hand, one has for $t \to \infty$

$$V^*(t)\begin{pmatrix} 1 & 1 \\ e^{2\pi i l} & 1 \end{pmatrix} = \frac{1}{2\pi i} \int_{\Gamma} dt' \frac{-1}{t' - t} \left(e^{-m(t'-t)} \mathbf{v}_1(t'), e^{m(t'-t)} \mathbf{v}_1(t') \right)$$

$$= O(1) \times t^{-1}\begin{pmatrix} e^{mt} & \\ & e^{-mt} \end{pmatrix}. \tag{4.7}$$

From (4.2) and (4.7) one sees that

$$[V^*(t)-V(t)]\begin{pmatrix}1 & 1\\ e^{2\pi i l} & 1\end{pmatrix}\begin{pmatrix}e^{-mt} & \\ & e^{mt}\end{pmatrix}$$

is an entire function which behaves like $O(t^{-1})$ as $t\to\infty$. This implies $V^*(t)-V(t)=0$ and the proof is over.

Making use of the freedom $v_i(t)\to Gv_i(t)$, one can normalize $v_1(t)$ at $t=\pm ia$,

$$\begin{aligned}v_1(t) &= \left[(2ia)^{-l}c_{3l}(a)\times\begin{pmatrix}1\\0\end{pmatrix}+O(|t-ia|)\right]\times(t-ia)^{l-1}\\ &= \left[(-2ia)^{-l}\bar{c}_{3,-l}(a)\times\begin{pmatrix}0\\1\end{pmatrix}+O(|t+ia|)\right]\\ &\quad\times(t+ia)^{-l-1},\end{aligned}$$

so that the normalization $f_l|_C=1, \bar{f}_{-l}|_C=1$ is achieved. The constants $c_{3l}(a), \bar{c}_{3,-l}(a)$ will be given in (5.3). From (4.5) and (4.6), $\mathfrak{w}^{(+)}(0,0,x_3)$ is equal to $2\pi i(v_1(x_3)+v_2(x_3))$. Hence we have also the inversion formula to (3.5)

$$\begin{aligned}v(t) &= \frac{-1}{1-e^{2\pi i l}}\int_\Gamma \frac{e^{-m(t'-t)}-e^{2\pi i l}e^{m(t'-t)}}{t'-t}\\ &\quad\times\mathfrak{w}^{(+)}(0,0,t')dt',\end{aligned}\tag{4.8}$$

where $\mathfrak{w}^{(+)}(0,0,t')$ signifies the branch for $t'>0$.

Next consider the case $l=0$. The relevant monodromy problem coincides with the one previously studied in connection with the impenetrable bose gas.[2] We set

$$\pi i v_1(t)=\begin{pmatrix}\dfrac{e^{m(t-ia)}}{2(t-ia)}\\ \dfrac{e^{m(t+ia)}}{2(t+ia)}\end{pmatrix}+\pi i\lambda\int_\Gamma dt_1\frac{e^{m(t-t_1)}}{2(t-t_1)}\frac{1}{2\pi i}\log\frac{t_1-ia}{t_1+ia}v_2(t_1),\tag{4.9}$$

$$\pi i v_2(t)=\begin{pmatrix}\dfrac{\sinh m(t-ia)}{t-ia}\\ \dfrac{\sinh m(t+ia)}{t+ia}\end{pmatrix}+\pi i\lambda\int_\Gamma dt_1\frac{\sinh m(t-t_1)}{t-t_1}\frac{1}{2\pi i}\log\frac{t_1-ia}{t_1+ia}v_2(t_1).\tag{4.10}$$

The latter is a Fredholm integral equation of the second kind, and can be solved by iteration at least if $a\ll 1$. {In the notation of Ref. 2, $v_1(t)=(m/\pi i)^l(R_l^+(-imt, ma; -i\lambda), R_l^+(-imt, -ma; -i\lambda))$, $v_2(t)=(m/\pi i)^l(R_l(-imt, ma; -i\lambda), R_l(-imt, ma; -i\lambda))$ with $I=[-ma, ma]$.} By a similar calculation as above, the condition $\mathfrak{w}^{(+)}(0,0,x_3)=-\mathfrak{w}^{(-)}(0,0,x_3)$ is reduced to

$$\begin{aligned}&\int_{\tilde\Gamma}\frac{e^{-m(t-t_1)}}{t-t_1}v_1(t_1)dt_1+2\pi i v_1(t)\\ &=\int_{\tilde\Gamma}\frac{e^{m(t-t_1)}}{t-t_1}v_1(t_1)dt_1\\ &\quad+2\pi i[v_1(t)+2\pi i\cdot(\lambda/2)v_2(t)].\end{aligned}\tag{4.11}$$

Here we have used

$$v_1(t)-\tilde{v}_1(t)=\frac{\lambda}{2}\log\frac{t-ia}{t+ia}v_2(t),\tag{4.12}$$

$\tilde{v}_1(t)$ being (4.9) with Γ replaced by $\tilde\Gamma$. Substituting (4.12) into (4.11) and using (4.9), (4.10) we see that (4.11) is further reduced to

$$(1+\pi i\lambda)v_2(t)=0.$$

Hence the parameter λ is determined to be $-1/\pi i$. This choice is exactly the same as the case of $E_\nu(t)$ in the theory of random matrices [see (7.117)[2]].

5. MASSLESS CASE

When $m=0$, one can work out all the procedures in Secs. 2 and 3 to find explicit representations of the solutions. Naturally the deformation theory is trivial in this case, for it reduces to the scaling of variables $u_l(ax;a)=a^{2l-1}u_l(x;1)$.

As the kernel of the integral transformation (3.5), we now choose

$$K(x)=\begin{cases}\log(x_1^2+x_2^2) & (n=2),\\ (x_1^2+\cdots+x_n^2)^{-(n-2)/2} & (n\geqslant 3).\end{cases}\tag{5.1}$$

Assume first that $n\neq 2$, and that $l\neq 0$ for odd $n\geqslant 5$. Under the same choice of the contour Γ as in Sec. 3, the transform $v(t)$ is explicitly given by

$$v(t)=\begin{pmatrix}c_{nl}(a)\cdot(t-ia)^{l+(n-5)/2}(t+ia)^{-l+(n-3)/2}\\ \bar{c}_{n,-l}(a)\cdot(t-ia)^{l+(n-3)/2}(t+ia)^{-l+(n-5)/2}\end{pmatrix}.\tag{5.2}$$

The normalization constants $c_{nl}(a), \bar{c}_{n,-l}(a)$ are so determined to satisfy $f_l|_C=1, \bar{f}_{-l}|_C=1$, as follows.

$$\begin{aligned}c_{nl}(a) &= -\frac{(2a)^{2l}}{2i^n\pi}\frac{\Gamma((n-2)/2)\Gamma(-l-(n-5)/2)}{\Gamma(-l+\frac{1}{2})},\\ \bar{c}_{n,-l}(a) &= -\frac{(2a)^{-2l}}{2(-i)^n\pi}\frac{\Gamma((n-2)/2)\Gamma(l-(n-5)/2)}{\Gamma(l+\frac{1}{2})}.\end{aligned}\tag{5.3}$$

In particular, for $n=3$ and $l=0$, a residue calculus will recover (1.3).

If $l=0$ and $n\geqslant 5$ is odd, (5.2), (5.3) are modified as

$$v(t)=\begin{pmatrix}c'_{n0}(a)\cdot(t-ia)^{(n-5)/2}(t+ia)^{(n-3)/2}\log\dfrac{t-ia}{t+ia}\\ \bar{c}'_{n0}(a)\cdot(t-ia)^{(n-3)/2}(t+ia)^{(n-5)/2}\log\dfrac{t+ia}{t-ia}\end{pmatrix},\tag{5.4}$$

where $c'_{n0}(a)=\lim_{l\to 0}lc_{nl}(a)$, $\bar{c}'_{n0}(a)=\lim_{l\to 0}(-l)\bar{c}_{n,-l}(a)$ are given by

$$c'_{n0}(a)=\frac{\Gamma((n-2)/2)}{2i\pi^{3/2}\Gamma((n-3)/2)}=-\bar{c}'_{n0}(a).\tag{5.5}$$

If $n=2$, the normalized solutions themselves can be written out explicitly.

$$\begin{aligned}u_l(x) &= (2a)^{2l-1}\\ &\quad\times\left[\left(\frac{x_1-a-ix_2}{x_1+a-ix_2}\right)^{l-1/2}+\left(\frac{x_1+a+ix_2}{x_1-a-ix_2}\right)^{l-1/2}\right].\\ \bar{u}_{-l}(x) &= (2a)^{-2l-1}\\ &\quad\times\left[\left(\frac{x_1-a-ix_2}{x_1+a-ix_2}\right)^{l+1/2}+\left(\frac{x_1+a+ix_2}{x_1-a+ix_2}\right)^{l+1/2}\right].\end{aligned}\tag{5.6}$$

Finally in the case $n=3$ with nonzero ν (which we take to be positive), the solutions are as follows. For $l\neq 0$

$$\mathbf{v}(t) = \begin{pmatrix} c^{r(\nu)}(a)\cdot(t-ia)^{l+\nu-1}(t+ia)^{-l+\nu} \\ \vec{c}^{r(\nu)}_{-l}(a)\cdot(t-ia)^{l+\nu}(t+ia)^{-l+\nu-1} \end{pmatrix}, \tag{5.7}$$

$$c^{r(\nu)}_{l}(a) = \frac{(2a)^{2l-\nu}\Gamma(\nu+\frac{1}{2})}{2i\Gamma(l+\nu)\Gamma(-l+\frac{1}{2})\sin\pi l},$$

$$\vec{c}^{r(\nu)}_{-l}(a) = \frac{(2a)^{-2l-\nu}\Gamma(\nu+\frac{1}{2})}{2i\Gamma(-l+\nu)\Gamma(l+\frac{1}{2})\sin\pi l},$$

and for $l=0$, $\nu\neq 0$

$$\mathbf{v}(t) = \frac{(2a)^{-\nu}\Gamma(\nu+\frac{1}{2})}{2\pi i\pi^{1/2}\Gamma(\nu)}\begin{pmatrix} (t-ia)^{\nu-1}(t+ia)^{\nu} \\ (t-ia)^{\nu}(t+ia)^{\nu-1} \end{pmatrix}\log\frac{t-ia}{t+ia}. \tag{5.8}$$

[1] M. Sato, T. Miwa, and M. Jimbo, Proc. Jpn. Acad. A **53**, 147, 153, 183 (1977); Publ. Res. Inst. Math. Sci. **15**, 871 (1979).
[2] M. Jimbo, T. Miwa, Y. Môri, and M. Sato, Physica D **1**, 80 (1980).
[3] M. Jimbo and T. Miwa, Proc. Jpn. Acad. A **56**, 405 (1980).
[4] K. Ueno, Proc. Jpn. Acad. A **56**, 103, 210 (1980).
[5] M. Sato, T. Miwa, and M. Jimbo, Publ. Res. Inst. Math. Sci. **15**, 577 (1979).
[6] J. M. Myers, J. Math. Phys. **6**, 1839 (1965).
[7] J. Des Cloiseaux and M. L. Mehta, J. Math. Phys. **14**, 1648 (1973).
[8] F. J. Dyson, Commun. Math. Phys, **47**, 171 (1976).
[9] M. Jimbo and T. Miwa, Physica D **2**, 407 (1981).
[10] M. Jimbo, T. Miwa, and K. Ueno, Physica D **2**, 306 (1981).

Stochastic quantization of the electromagnetic field

Mark Davidson
Department of Physics, San Jose State University, San Jose, California 95192

(Received 15 July 1980; accepted for publication 12 September 1980)

Stochastic field equations for electromagnetism are presented. The covariance in the ground state is calculated. The question of Lorentz covariance is examined. The connection with ordinary quantum field theory is made. The theory is compared with stochastic electrodynamics.

PACS numbers: 03.70. + k, 11.10.Np, 03.65.Bz, 12.20. − m

I. INTRODUCTION

Stochastic quantization provides a realistic model of quantum systems while reproducing the measurable assertions of the usual quantum theory. The germinal work was a paper by Fenyes,[1] but Nelson's elegant analysis[2,3] has been the primary inspiration in the field. Extensive contributions have also been made by de la Pena and Cetto.[4–6] It is now established that the procedure of stochastic quantization is not unique, and that the diffusion parameter can be chosen to any positive value.[7–9]

Guerra and his coworkers have analyzed the relativistic scalar field and the free electromagnetic field[10,11] using the quantization prescription of Nelson. Moore[12] has also considered the Maxwell field as part of a general program of stochastic field theory. In these studies, only the fixed value of the diffusion parameter of the original Fényes–Nelson theory was considered. In this paper, the more general theory is presented. As for the scalar field[13] the moments of the electromagnetic field in the ground state are equal to the Schwinger functions of the theory, but with the times scaled by a common factor which depends upon the diffusion parameter. Agreement is obtained with Refs. 11 and 12 when the diffusion parameter is restricted to Nelson's value, even though in the present work quantization is carried out in terms of vector potentials rather than directly in terms of electric and magnetic fields as in Ref. 11.

As is well known, it is impossible to quantize the electromagnetic field is such a way that the vector potentials are manifestly Lorentz covariant while at the same time the quantization procedure is strictly canonical. Since stochastic quantization, in its present form, must parallel a canonical quantization procedure, the noncovariant Coulomb or radiation gauge has been selected in the present work. Ultimately, Lorentz covariance of the theory is established by means of an analytic continuation in the diffusion parameter. In fact, it is found that the measurable predictions of the theory are the same as the usual quantum electromagnetism. This is in spite of the fact that the moments of the stochastic theory are generally different from those of the quantum theory and are not manifestly Lorentz covariant.

In the Coulomb gauge it turns out that the magnetic field is a well-defined stochastic process but that the electric field is not because it is expressed in terms of a time derivative of the vector potential. This difference between electric and magnetic fields is not fundamental because of the symmetry between these fields in electromagnetism. Other methods of stochastic quantization are illustrated, which directly exploit this symmetry, and which yield different conclusions.

Finally, some of the properties of stochastic quantization are compared with a different stochastic model for electromagnetism.

II. THE ELECTROMAGNETIC FIELD AS A STOCHASTIC PROCESS

In an appropriate system of units the Maxwell equations take the form

$$\nabla\cdot\mathbf{E}=\rho,\quad \nabla\cdot\mathbf{B}=0,$$
$$\nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t},\quad \nabla\times\mathbf{B}=\mathbf{J}+\frac{\partial\mathbf{E}}{\partial t}, \tag{1}$$

where c has been set to unity. In the Coulomb gauge one has

$$\nabla\cdot\mathbf{A}=0, \tag{2}$$

$$\mathbf{E}=-\nabla\phi-\frac{\partial\mathbf{A}}{\partial t},\quad \mathbf{B}=\nabla\times\mathbf{A}, \tag{3}$$

$$\Delta\phi=-\rho. \tag{4}$$

The Maxwell equations become

$$\left[\frac{\partial^2}{\partial t^2}-\Delta\right]\mathbf{A}=\mathbf{J}-\nabla\frac{\partial\phi}{\partial t}. \tag{5}$$

The energy density of the the field in these units is given by

$$\mathscr{H}=\tfrac{1}{2}[\mathbf{E}^2+\mathbf{B}^2]. \tag{6}$$

Quantization shall be carried out in a cubic volume of edge length L with periodic boundary conditions in L. Consistency requires that $\mathbf{J}$ and ρ have this same periodicity, but little generality is lost by this since ultimately the volume will tend to infinity. Because of the periodicity requirement, the following decompositions are allowed

$$\mathbf{A}=1/\left(\sqrt{2}L^{3/2}\right)\sum_{\lambda,\mathbf{k}}\epsilon(\lambda,\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}Q_{\lambda,\mathbf{k}}(t), \tag{7}$$

$$\phi=\sum_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\rho(\mathbf{k},t)/\mathbf{k}^2, \tag{8}$$

$$\rho=\sum_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\rho(\mathbf{k},t), \tag{9}$$

$$\mathbf{J}=\sum_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\mathbf{J}(\mathbf{k},t), \tag{10}$$

where the components of $\mathbf{k}$ are restricted to be integral multiples of $2\pi/L$ and

$$\epsilon(\lambda,-\mathbf{k})=\epsilon(\lambda,\mathbf{k}),\quad \epsilon\cdot\mathbf{k}=0,\quad \lambda=1,2, \tag{11}$$

 0022-2488/81/112588-06$01.00

$$Q_{\lambda,-\mathbf{k}} = Q^*_{\lambda,\mathbf{k}}, \quad \rho(-\mathbf{k},t) = \rho^*(\mathbf{k},t), \quad \mathbf{J}(-\mathbf{k},t) = \mathbf{J}^*(\mathbf{k},t), \tag{12}$$

$$\partial\rho/\partial t = -i\mathbf{k}\cdot\mathbf{J}. \tag{13}$$

It is customary to define the transverse current by

$$\mathbf{J}_T = \mathbf{J} - \nabla\partial\phi/\partial t \tag{14}$$

and it is easy to show that its divergence vanishes

$$\nabla\cdot\mathbf{J}_T = 0, \tag{15}$$

so that one may write

$$\mathbf{J}_T = (1/\sqrt{2}L^{3/2})\sum_{\lambda,\mathbf{k}} \epsilon(\lambda,\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} Z_{\lambda,\mathbf{k}}(t). \tag{16}$$

Since the scalar potential's time derivatives do not enter into Eqs. (3) and (4), it can be immediately solved for as in Eq. (8) and it need not be considered as a dynamic variable. The only canonical coordinates, then, are the real and imaginary parts of the Q's. Equation (5) leads to the following equations for the Q's

$$\left[\frac{\partial^2}{\partial t^2} + \mathbf{k}^2\right] Q_{\lambda,\mathbf{k}}(t) = Z_{\lambda,\mathbf{k}}(t). \tag{17}$$

A suitable Hamiltonian from which these equations may be derived is

$$H = \sum_{\lambda,\mathbf{k}} \{\tfrac{1}{4}(|\dot{Q}_{\lambda,\mathbf{k}}|^2 + \mathbf{k}^2|Q_{\lambda,k}|^2) - \tfrac{1}{2}\mathrm{Re}(Z^*_{\lambda,\mathbf{k}} Q_{\lambda,\mathbf{k}})\}, \tag{18}$$

which may also be written as

$$H = \int_\Omega \left[\tfrac{1}{2}\dot{A}^2 + \tfrac{1}{2}\sum_{i=1}^{3} (\nabla A_i)^2 - \mathbf{J}_T\cdot\mathbf{A}\right] d^3x, \tag{19}$$

where Ω denotes the cubic volume of edge L.

When there are no charges or currents, the Hamiltonian of Eq. (18) or (19) is equal to the total energy of the field obtained from the energy density in Eq. (6). Otherwise, it is not. The canonical coordinates in Eq. (18) are the real and imaginary parts of the Q's, and the canonical momenta are their time derivatives. Since the gauge constraint has been taken into account in the Fourier decomposition of $\mathbf{A}$, the Q's and their conjugate momenta may be varied freely.

Since the equations for the Q's are those for forced harmonic oscillators, in principle the quantization of the field can be achieved by solving a Schrödinger equation for uncoupled but forced harmonic oscillators. Thus, there is some wave function in Q space which satisfies

$$H\psi = i\hbar\frac{\partial\psi}{\partial t} \tag{20}$$

and where the Hamiltonian operator has the general form

$$H = \sum_i \left[-\frac{\hbar^2}{2}\frac{\partial^2}{\partial Q_i^2} - Z_i Q_i + \tfrac{1}{2}\mathbf{k}_i^2 Q_i^2\right], \tag{21}$$

where the Q_i's and Z_i's are real and represent the real and imaginary parts of the Q's and Z's in Eq. (18). The sum in Eq. (21) is only over independent Q's [in the sense of Eq. (12)] which explains the factor-of-two difference between Eqs. (18) and (21).

The basis of the stochastic formulation of quantum mechanics rests on the following theorem. Let

$$\psi = e^{R+iS} \text{ and define } \Delta_Q \equiv \sum_i \frac{\partial^2}{\partial Q_i^2}, \tag{22}$$

where R and S are real functions. Then,

Theorem: Let R and S be bounded, continuous, real functions of Q and t with first and second Q-derivatives and first time-derivative in a region U in (Q,t) space. Let $z \neq 0$ below. Then for (Q,t) in U, the following two equations are equivalent:

$$\left[-\frac{\hbar^2}{2}\Delta_Q + V(Q)\right] e^{R+iS} = i\hbar\frac{\partial}{\partial t} e^{R+iS}, \tag{23}$$

$$\left[-\frac{(z\hbar)^2}{2}\Delta_Q + \left(V(Q) + \frac{\hbar^2}{2}(z^2-1)\frac{\Delta_Q e^R}{e^R}\right)\right] e^{R+iS/z} = i(z\hbar)\frac{\partial}{\partial t} e^{R+iS/z}, \tag{24}$$

where V is a real function which can also depend explicitly on the time without affecting the result. The parameter z must be a constant, but it may be complex. The proof of the theorem is elementary and follows simply by expanding (24) into real and imaginary parts. The proof is essentially the same as that given in Ref. 8 for the three-dimensional case.

The theorem shows that the Schrödinger equation has an interesting nonlinear symmetry. If we scale $\hbar$ by the factor z and simultaneously add an extra (nonlinear) term to the potential and also divide S by z, then the equation remains invariant. If z is chosen to be imaginary, then Schrödinger's equation is transformed into the form of the heat equation with an extra term in the potential. This fact is essentially why a stochastic model of quantum mechanics is possible at all, for a diffusion equation with a suitable nonlinear term in the potential is the same as Schrödinger's equation. Actually, Nelson's dynamics[2,3] or the more general dynamics of Refs. 7–9 have the effect of adding an extra term to the potential in the diffusion equation of exactly the form needed in the theorem. It would be nice to have a deeper understanding of this theorem and the symmetry it reveals, especially since it plays such a crucial role in the stochastic formulation. Unfortunately, at the present state of knowledge, the theorem seems to reveal a nonapparent symmetry of Schrödinger's equation with no deeper mathematical meaning, except that it happens to allow a view of nature entirely different from the more tranditional interpretations of quantum mechanics.

The stochastic interpretation rests on the assertion that all of the experimentally verifiable predictions of quantum mechanics are contained in the Schrödinger equation together with the statistical interpretation of ψ and the phenomenon of reduction of the wave packet upon measurement. A justification for this claim shall not be attempted here, but it is a reasonable if not compelling premise.

The basic techniques for developing a stochastic model for a quantum system have been set forth by Nelson,[2,3] and have been applied to the free Maxwell field by Guerra and Loffredo.[11] Here, the techniques of the generalized Fényes–Nelson model will be followed as introduced in Refs. 7 and 8. Each of the canonical coordinates is described by a stochastic process and satisfies a stochastic differential equation of the form

$$dQ_i = b_i(Q,t)dt + dW_i. \tag{25}$$

The W's are Wiener processes which satisfy

$$E(dW_i dW_j) = 2\nu\delta_{ij}dt. \tag{26}$$

The solutions to Eq. (25) are called Ito processes or generlized Brownian motion. Background information may be found in Refs. 14–21.

If the b's satisfy a global Lipschitz condition (Ref. 1, p. 43) then a unique solution to (25) is ensured in the finite-dimensional case. The Q's are then found to be continuous Markov processes on configuration space whose sample trajectories are so jagged as to be nowhere differentiable almost surely. The restriction to finite-dimensionality is not of major concern since any field theory can be made finite-dimensional by imposing a momentum cutoff and working in a finite volume. The global Lipschitz condition, however, presents some more fundamental difficulties since it is not satisfied when the quantum wave function has zeros as it always does in excited stationary states. The results of Albeverio and Hoegh–Krohn for the stationary case show that the global Lipschitz condition can be relaxed considerably if existence is all that is required,[22] but a similar treatment of the nonstationary case has not yet been given. These complications can be avoided by working only with wave functions which do not have exact zeros.

Following Nelson,[2,3] forward and backward time derivatives may be defined

$$Df(Q,t) = \lim_{h\to 0,\ h} \frac{1}{h} E\left[f(Q(t+h),t+h) - f(Q(t),t) \mid Q(t) = Q \right], \tag{27}$$

$$D_* f(Q,t) = \lim_{h\to 0,\ h} \frac{1}{h} E\left[f(Q(t),t) - f(Q(t-h),t-h) \mid Q(t) = Q \right]. \tag{28}$$

Some important relations which may be derived are

$$DQ_i = b_i, \tag{29}$$

$$D_* Q_i = b_{i*}, \tag{30}$$

$$b_i - b_{i*} = 2\nu \frac{\partial}{2Q_i} \ln(\rho(Q,t)), \tag{31}$$

$$D = \frac{\partial}{\partial t} + \sum_i b_i \frac{\partial}{\partial Q_i} + \nu\Delta_Q, \tag{32}$$

$$D_* = \frac{\partial}{\partial t} + \sum_i b_{i*} \frac{\partial}{\partial Q_i} - \nu\Delta_Q. \tag{33}$$

In these formulae, ν is termed the diffusion parameter or the diffusion constant.

In order to complete the specification of a stochastic model, some dynamical assumption must be made. Nelson[2,3] showed that the assumption

$$\tfrac{1}{2}[DD_* + D_* D]Q_i = -\frac{\partial}{\partial Q_i} V \tag{34}$$

leads to a system of equations which can be transformed into Schrödinger's equation where

$$\psi = e^{R + iS}, \tag{35}$$

$$b_i = 2\nu \frac{\partial}{\partial Q_i}(R + S_N), \tag{36}$$

$$b_{i*} = 2\nu \frac{\partial}{\partial Q_i}(-R + S_N), \tag{37}$$

$$\nu = \hbar/2. \tag{38}$$

It is now established[7–9] that Nelson's dynamical assumption is not the only possibility. The following equation also leads to Schrödinger's equation

$$\tfrac{1}{2}[DD_* + D_* D]Q_i + \tfrac{1}{8}\beta(D - D_*)^2 Q_i = -\frac{\partial}{\partial Q_i} V, \tag{39}$$

where β is a constant and where

$$\psi = e^{R + iz S_N}, \tag{40}$$

$$z = 1/\sqrt{1 - \beta/2}, \tag{41}$$

$$\nu = \hbar/\left(2\sqrt{1 - \beta/2}\right), \tag{42}$$

and where b_i and b_{i*} are still given by Eqs. (36) and (37) but with the value of ν in Eq. (42). If ν in Eq. (42) is to be real, which is necessary for a physical interpretation, then β must be real and satisfy the condition

$$\beta < 2. \tag{43}$$

By choosing β appropriately, any value of the diffusion parameter can be used in a stochastic model of quantum mechanics. It is possible to choose ν to be arbitrarily small by letting β be large in magnitude but negative. In the limit of zero ν the model becomes deterministic and in fact becomes the familiar hidden variable model of Bohm,[23] a point first made by Shucker.[9]

With these tools, the stochastic quantization of the electromagnetic field is straightforward. It is convenient to define

$$\delta(x,t) = 1/\left(\sqrt{2}L^{3/2}\right) \sum_{\lambda,\mathbf{k}} \epsilon(\lambda,\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \times \left(\frac{\partial}{\partial \mathrm{Re} Q_{\lambda,\mathbf{k}}} + i \frac{\partial}{\partial \mathrm{Im} Q_{\lambda,\mathbf{k}}} \right). \tag{44}$$

The forward and backward time derivatives can be expressed in terms of δ as

$$D = \frac{\partial}{\partial t} - \int_\Omega d^3x\, (\mathbf{E}_+ + \nabla\phi)\cdot\delta + \nu \int_\Omega d^3x\, \delta^2, \tag{45}$$

$$D_* = \frac{\partial}{\partial t} - \int_\Omega d^3x\, (\mathbf{E}_- + \nabla\phi)\cdot\delta - \nu \int_\Omega d^3x\, \delta^2, \tag{46}$$

where

$$\mathbf{E}_+ = -D\mathbf{A} - \nabla\phi, \tag{47}$$

$$\mathbf{E}_- = -D_*\mathbf{A} - \nabla\phi. \tag{48}$$

The Maxwell equations become

$$\nabla\cdot\mathbf{E}_+ = \nabla\cdot\mathbf{E}_- = \rho, \tag{49}$$

$$\nabla\cdot\mathbf{B} = 0, \tag{50}$$

$$\nabla\times\mathbf{E}_+ = -D\mathbf{B}, \quad \nabla\times\mathbf{E}_- = -D_*\mathbf{B}, \tag{51}$$

$$\tfrac{1}{2}[D\mathbf{E}_- + D_*\mathbf{E}_+] + (\beta/8)(D - D_*)(\mathbf{E}_+ - \mathbf{E}_-) = \nabla\times\mathbf{B} - \mathbf{J}. \tag{52}$$

With the aid of the following stochastic field,

$$\mathbf{W}(x,t) = 1/(\sqrt{2}L^{3/2})\sum_{\lambda,\mathbf{k}} \epsilon(\lambda,\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} W_{\lambda,k}(t), \tag{53}$$

the stochastic differential equations can be written in the compact and suggestive form

$$d\mathbf{A} = -(\mathbf{E}_+ + \nabla\phi)dt + d\mathbf{W}, \tag{54}$$

$$d\mathbf{A} = -(\mathbf{E}_- + \nabla\phi)dt + d\mathbf{W}_*, \tag{55}$$

where Eq. (55) is the "backwards" version of the stochastic differential equations and $\mathbf{W}_*$ is another field of the form (53), but with different Wiener processes.

When the present theory is restricted to the special case $\nu = \hbar/2$, and when there are no charges or currents present, then Eqs. (49)–(52) become precisely those derived by Guerra and Loffredo.[11] Therefore, quantization in the Coulomb gauge is equivalent to quantization directly in terms of the fields as in Ref. 11. It is interesting to note, as is pointed out in Ref. 11, that the electric field does not exist as a well-defined stochastic process since the time derivatives of **A** do not exist. Despite this fact, the two functions E_+ and E_- may be thought of as a stochastic replacement for the electric field. The set of stochastic Maxwell equations are a richer theory than the classical theory. Besides being equivalent to the quantized field, they open the interesting possibilty of studying the theory of classical point particles from a new perspective.

The equations take a particularly simple form when there are no charges or currents present.

III. THE GROUND STATE OF THE FREE FIELD

For the free field, the salar potential vanishes in the Coulomb gauge. When the Schrödinger equation is solved in the ground state and the Markov process for the field examined, it is found that the vector potential is a Gaussian Markov field with the following covariance:

$$E(A^i(x)A^j(y)) = \hbar\int \frac{d^4k}{(2\pi)^4} \frac{(\delta_{ij} - k_i k_j/\mathbf{k}^2)}{k_0^2 + \mathbf{k}^2} \times e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) + (ik_0 2\nu/\hbar)|t_x - t_y|}, \tag{56}$$

where the infinite volume limit has been taken. The covariance for the magnetic field is

$$E(B^i(x)B^j(y)) = \hbar\int \frac{d^4k}{(2\pi)^4} \frac{(\delta_{ij}\mathbf{k}^2 - k_i k_j)}{k_0^2 + \mathbf{k}^2} \times e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) + (ik_0 2\nu/\hbar)|t_x - t_y|}. \tag{57}$$

As has been stated before, the electric field does not, strictly speaking, exist. However, because the following expressions are equal,

$$\frac{\partial}{\partial t_x}\frac{\partial}{\partial t_y} E(A^i(x)A^j(y)) = E(E^i_{\pm}(x)E^j_{\pm}(y)), \quad t_x \neq t_y, \tag{58}$$

it is natural to study the properties of the left-hand side of (58) and identify these with an electric field. One finds

$$\frac{\partial}{\partial t_x}\frac{\partial}{\partial t_y} E(A^i(x)A^j(y)) = 2\nu\delta(t_x - t_y)\delta^{ij}_{tr}(\mathbf{x}-\mathbf{y}) + (2\nu/\hbar)^2 E(B^i(x)B^j(y)), \tag{59}$$

where $\delta(t)$ is the Dirac delta function and where $\delta^{ij}_{tr}(x)$ is the transverse delta function defined by

$$\delta^{ij}_{tr}(x) = \int \frac{d^3k}{(2\pi)^3}(\delta_{ij} - k_i k_j/\mathbf{k}^2)e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{60}$$

Finally, terms with both electric and magnetic fields may be evaluated. The result is

$$-\frac{\partial}{\partial t_x} E(A^i(x)B^j(y)) = E(E^i_{\pm}(x)B^j(y)) = i\hbar(2\nu/\hbar)\epsilon_{ijl}(1 - 2\theta(t_y - t_x)) \times \int \frac{d^4k}{(2\pi)^4} k_l\, e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y}) + (ik_0 2\nu/\hbar)|t_x - t_y|}, \tag{61}$$

where ϵ in (61) is the totally antisymmetric tensor, and θ equals 1 when its argument is positive and zero otherwise. When both fields are at the same point in space, one obtains the results

$$E(B^i(t_x)B^j(t_y)) = \hbar(\delta_{ij}/\pi^2)(\hbar/2\nu)^4|t_x - t_y|^{-4}, \tag{62}$$

$$E(E^i_{\pm}(t_x)E^j_{\pm}(t_y)) = \hbar(\delta_{ij}/\pi^2)(\hbar/2\nu)^2|t_x - t_y|^{-4}, \tag{63}$$

$$E(E^i_{\pm}(t_x)\mathbf{B}^j_{\pm}(t_y)) = 0. \tag{64}$$

Higher-order correlations may be calculated from those above by means of the well-known combinatoric rule for Gaussian processes. For example,

$$E(A^{i1}(x_1)\times\cdots\times A^{i2N}(x_{2N})) = \sum_{\pi} E(A^{i1}(x_1)A^{i2}(x_2)) \times\cdots\times E(A^{i2N-1}(x_{2N-1})A^{i2N}(x_{2N})), \tag{65}$$

where the sum is over all distinct permutations and where all odd correlations vanish.

IV. THE CONNECTION WITH QUANTUM FIELD THEORY AND RELATIVISTIC COVARIANCE

The Green's functions of the usual quantum field theory are obtained by analytically continuing the diffusion parameter to imaginary values, as was done in Ref. 13. The physical interpretation of the stochastic theory is of course lost when the diffusion parameter becomes imaginary. However, since Schrödinger's equation is satisfied at every point along the analytic continuation, and since it is reasonable to assume that this is enough to ensure that all of the measurable predictions of quantum theory are reproduced, then it follows that the diffusion parameter cannot be deduced by any measurement except perhaps one which proved quantum mechanics to be wrong or incomplete. Therefore, the numerical and physical predictions of the theory (cross sections, line spectrum, etc.) must be independent of the diffusion parameter so that by continuing it to imaginary values,

these experimental predictions won't be affected. The analytically-continued expectations are denoted by a subscript. The following results are found when the usual quantization is carried out in the Coulomb gauge:

$$E_{\nu=+i\hbar/2}(A^i(x)A^j(y)) = \langle 0|TA^i(x)A^j(y)|0\rangle, \tag{66}$$

$$E_{\nu=-i\hbar/2}(A^i(x)A^j(y)) = \langle 0|T^*A^i(x)A^j(y)|0\rangle, \tag{67}$$

where $T(T^*)$ denotes time ordering (anti time ordering) with later times to the left (right). The two-point function for the quantum field is given by

$$\langle 0|A^i(x)A^j(y)|0\rangle = i\hbar\int \frac{d^4k}{(2\pi)^4}(\delta_{ij} - k_ik_j/\mathbf{k}^2) \times e^{-ik^\mu(x_\mu - y_\mu)}/[k^\mu k_\mu + i\epsilon],$$

$$t_x > t_y, \tag{68}$$

where the metric convention $g^{00} = +1$ is used and where ϵ is positive and is taken to zero after the k_0 integration is performed. The other time ordering is obtained by taking the complex conjugate of Eq. (68). Owing to the Gaussian combinatorics of both the quantum theory and the stochastic theory, we may write in general

$$E_{\nu=i\hbar/2}(A^i(x_1)\cdots A^{iN}(x_N)) = \langle 0|TA^i(x_1)\cdots A^{iN}(x_N)|0\rangle, \tag{69}$$

$$E_{\nu=-i\hbar/2}(A^i(x_1)\cdots A^{iN}(x_N)) = \langle 0|T^*A^i(x_1)\cdots A^{iN}(x_N)|0\rangle. \tag{70}$$

The conclusion to be drawn is that the stochastic theory is physically equivalent to the Lorentz covariant quantum field theory, even though the stochastic theory is not itself manifestly Lorentz covariant because the expectations in Eqs. (56)–(65) do not transform in a covariant way. It seems that one of the tenets which must be given up in exchange for a realistic picture of nature is manifest Lorentz covariance. One way to interpret this noncovariance is to imagine that the properties of the vacuum are not Lorentz invariant, in reality, but nevertheless the differences are so subtle between different Lorentz frames as to prevent a real observer from noticing any difference between the frames no matter how hard he looks. This question of covariance needs further clarification before most physicists will embrace a stochastic viewpoint. It seems as if two very basic principles are at odds: manifest and microscopic Lorentz covariance and the existence of an underlying reality to quantum phenomena.

As a final point, it should be recognized that the various expectations for a real diffusion parameter are just the Schwinger functions of the usual quantum field (time-ordered analytic continuations to imaginary time) with the times scaled by a factor of $2\nu/\hbar$.

V. ON THE NONUNIQUENESS OF THE STOCHASTIC QUANTIZATION PROCEDURE AS APPLIED TO ELECTROMAGNETISM

It was found in Sec. III that there was an essential difference between the electric and magnetic fields after the procedure of stochastic quantization was carried out. The magnetic field became a Markov field whereas the electric field became ill-defined. This was because the electric field was expressed as a time derivative of a vector potential plus a scalar potential term. The time derivative in the electric field was the source of the difference.

The apparent difference between the electric and magnetic fields is rather spurious for the following reason. Consider first the free Maxwell field. It is well known that Maxwell's equations possess a symmetry which mixes the electric and magnetic fields, and this symmetry is the primary justification for the possible existence of magnetic monopoles. The Maxwell equations are invariant under the change of field variables

$$\mathbf{E}' = \alpha\mathbf{E} + \beta\mathbf{B}, \quad \mathbf{B}' = \alpha\mathbf{B} - \beta\mathbf{E}, \tag{71}$$

$$\mathbf{E} = \alpha\mathbf{E}' - \beta\mathbf{B}', \quad \mathbf{B} = \alpha\mathbf{B}' + \beta\mathbf{E}', \tag{72}$$

$$\alpha^2 + \beta^2 = 1. \tag{73}$$

One could choose to work with the primed fields rather than the original ones and write

$$\mathbf{E}' = -\partial\mathbf{A}/\partial t, \quad \mathbf{B}' = \nabla\times\mathbf{A}, \quad \nabla\cdot\mathbf{A} = 0. \tag{74}$$

The result would be that now the field $\mathbf{B}'$ would be a Markov field, but the field $\mathbf{E}'$ would be ill-defined. Using Eq. (72) to solve for the actual electric and magnetic fields, it would turn out that these were mixtures of a well-defined Markov field and an ill-defined quantity. In particular, if one chooses

$$\alpha = 0, \quad \beta = 1, \tag{75}$$

then

$$\mathbf{B} = \mathbf{E}', \quad \mathbf{E} = -\mathbf{B}', \tag{76}$$

and one obtains the exactly opposite conclusion. The electric field is now well-defined, but the magnetic field is not.

In general, if Eq. (74) is used, then the covariances of the primed fields are given by Eqs. (56)–(64) and the covariances of the actual fields may be found by using these together with Eq. (72).

If there are charges present, the stochastic quantization prescription will still be nonunique because, due to the linearity of electromagnetism, the fields can always be written as a particular, nonrandom solution plus a free-field solution. The equations for the free-field part are invariant under Eqs. (71)–(73), and so again it is found that the difference between electric and magnetic fields is spurious.

VI. SOME COMMENTS ON STOCHASTIC ELECTRODYNAMICS

Despite the similar name, stochastic electrodynamics (SED) is a completely different theory from the one considered here. Good sources for SED are Boyer,[24] de la Pena and Cetto,[25] and Claverie.[26] In the SED world view, the vacuum contains a random radiation field with a spectral density equal to that of Planck radiation (including the zero point term) at zero temperature

$$\rho(w) = \hbar\omega^3/(2\pi^2). \tag{77}$$

The SED field can be thought of as a superposition of free-field plane waves whose phases are random. The stochastic

theory presented here is not a superposition of plane waves, but rather, the time dependence of each normal mode of the field is a complicated stochastic process. The covariances in SED are manifestly Lorentz covariant as was shown by Boyer.[24] Despite the differences between the two theories, there is one interesting similarity. If one evaluates the covariance is SED with the fields at the same space point, one finds

$$E_{\mathrm{SED}}(E^i(t_x)E^j(t_y)) = E_{\mathrm{SED}}(B^i(t_x)B^j(t_y)) = \frac{\hbar}{6\pi^2}\delta_{ij}\int_0^\infty dw\cos(w(t_x-t_y))w^3e^{-\epsilon w} = (\hbar/\pi^2)\delta_{ij}|t_x-t_y|^{-4} \tag{78}$$

$$E_{\mathrm{SED}}(E^i(t_x)B^j(t_y)) = 0, \tag{79}$$

where the limit $\epsilon\to 0$ has been taken in the last expression in (78). These formulas should be compared with similar ones for the present theory, Eqs. (62)–(64). When the Nelson condition is satified, so that $\nu = \hbar/2$, then the SED expectations agree with those obtained from stochastic quantization. This result must be qualified, however, because Eq. (63) does not include the singular term in Eq. (59). If this term is included in the present theory, one finds

$$\frac{\partial}{\partial t_x}\frac{\partial}{\partial t_y}E(A^i(t_x)A^j(t_y)) = 2\nu\delta(t_x-t_y)\delta^{ij}_{tr}(0) + (\hbar/\pi^2)(\hbar/2\nu)^2\delta_{ij}|t_x-t_y|^{-4}. \tag{80}$$

The transverse delta function in this equation could be made finite by putting a cutoff in k space in the theory. Comparing Eq. (80) with Eq. (78), one sees that even for Nelson's value of the diffusion coefficient, there is an extra term in (80). In the method of Secs. II and III, the expression (80) would represent the covariance of the electric field. However, using the arguments of Sec. V, the extra term could be shifted into the covariance for the magnetic field. Thus, when comparing SED with stochastic quantization, one can choose the covariance for either the electric fields or the magnetic fields to be the same, but not both.

VII. CONCLUSION

Stochastic quantization applied to the electromagnetic field yields a rich and interesting theory. It provides a realistic alternative to the usual quantization procedure, although it does not seem to offer any calculational advantages, except perhaps in numerical investigations or in stationary state problems.[27]

Working in the Coulomb gauge yields the same results as the method of Guerra and Loffredo[11] when the diffusion parameter has Nelson's value. The magnetic field is well defined as a stochastic process, but the electric field is not. This difference is not fundamental, however, because owing to the symmetry between electric and magnetic fields, it is possible to alter this conclusion by a different quantization method, and in particular one can quantize the field so that the electric field is well defined but the magnetic field isn't.

Although not manifestly Lorentz covariant, the stochastic field theory is experimentally equivalent to quantum field theory. The moments of the stochastic theory are equal to Schwinger functions, but with the times scaled by a common factor $2\nu/\hbar$.

When Nelson's value for the diffusion parameter is chosen, the covariance in the stochastic theory, with the fields at the same point, is the same as the covariance in stochastic electrodynamics except for an extra rather spurious term.

[1] I. Fényes, Z. Physik **132**, 81–106 (1952).
[2] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P., Princeton, N. J., 1967).
[3] E. Nelson, Phys. Rev. **150**, 1079 (1966).
[4] L. de la Pena-Auerbach and A. M. Cetto, Phys. Rev. D **3**, 795–800 (1971); Phys. Lett. A **29**, 562–63 (1969); Rev. Mex. Fis. **18**, 253–64 (1969).
[5] L. de la Pena-Auerbach, Rev. Mex. Fis. **19**, 133–45 (1970); Phys. Lett. A **31**, 403–4 (1970).
[6] T. A. Brody, A. M. Cetto, and L. de la Pena, Rev. Mex. Fis. **26**, 59–87 (1979).
[7] M. Davidson, Physica A **96**, 465–86 (1979).
[8] M. Davidson, Lett. Math. Phys. **3**, 271–7 (1979).
[9] D. Shucker, Lett. Math. Phys. **4**, 61–5 (1980).
[10] F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).
[11] F. Guerra and M. I. Loffredo, Lett. Nuovo Cimento **27**, 41–5 (1980).
[12] S. M. Moore, Found. Phys. **9**, 237–59 (1979).
[13] M. Davidson, "The Generalized Fényes–Nelson Model for Free Scalar Field Theory," to appear in Lett. Math. Phys.
[14] P. Levy, *Processus Stochastique et Mouvement Brownien* (Gauthier–Villars, Paris, 1965).
[15] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[16] K. Ito, *Lectures on Stochastic Processes* (Tata, Bombay, 1961).
[17] K. Ito and H. P. Mckean, Jr., *Diffusion Processes and Their Sample Paths* (Springer, Berlin, 1965).
[18] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1971).
[19] L. Breiman, *Probability* (Addison–Wesley, Reading, Mass., 1968).
[20] E. B. Dynkin, *Markov Processes* (Springer, Berlin, 1965).
[21] D. W. Stroock and S. S. Varadhan, *Multidimensional Diffusion Processes* (Springer, Berlin, 1979).
[22] S. Albeverio and R. Hoegh–Krohn, J. Math. Phys. **15**, 1745–7 (1974).
[23] D. Bohm, Phys. Rev. **85**, 180–93 (1952).
[24] T. H. Boyer, Phys. Rev. D **11**, 790–808, 809–30 (1975).
[25] L. de la Pena and A. M. Cetto, Found. Phys. **8**, 191–210 (1978); J. Math. Phys. **20**, 469–83 (1979).
[26] P. Claverie, "Stochastic Electrodynamics: Methods and Results," to appear in Proceedings, International Seminar on Mathematical Theory of Dynamical Systems and Microphysics, held at CISM (Centre International des Sciences Mecaniques) Udine, Sept. 3–4 (1979).
[27] B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).

Trajectories of the S-matrix poles in the complex wavenumber plane [a)]

P. Sanz, J. Sañudo, and J. Sesma
Departamento de Fisica Teorica, Facultad de Ciencias, Zaragoza, Spain

(Received 29 April 1981; accepted for publication 26 June 1981)

The singularities of the S matrix for a square potential, and their trajectories in the complex wavenumber plane as the angular momentum varies through continuous real values, are presented. Some enlightenment about the physical interpretation of Regge trajectories is obtained.

PACS numbers: 03.80. + r, 11.60. + c, 11.20. − e

I. INTRODUCTION

The study of singularities of the S matrix in the complex angular momentum plane (Regge poles) is established as a useful tool in the analysis of certain aspects of potential scattering. In what concerns nuclear physics, Regge poles can contribute to explaining many features of heavy-ion reaction.[1]

Instead of looking for poles of the S matrix for complex values of the angular momentum and their (Regge) trajectories as the (physical) energy varies, it is equally possible to analyze these singularities in the complex wavenumber plane and their trajectories as the angular momentum takes continuously varying real values (k trajectories). To our knowledge, this possibility, pointed out by Newton[2] many years ago, has not been sufficiently exploited, in spite of the fact that Regge trajectories and k trajectories are two different mathematical aspects of the same physical phenomena and would in principle be equivalent. Regge poles have deserved considerably more attention, probably due to their success in elementary particle physics and relevance in applying the Sommerfeld–Watson transform. However, it is our feeling that k trajectories can contribute to supplement and clarify some applications of Regge poles.

As a first step toward considering more realistic ion–ion optical potentials, we discuss in this paper the k trajectories for a real square (well or barrier) potential. For the lowest physical values of the angular momentum, the poles of the S matrix in the complex k plane were explicitly obtained by Nussenzveig[3] a long time ago. Here we discuss how the pole pattern found by Nussenzveig evolves when the angular momentum takes continuous real values.

In Sec. 2 we recall the S matrix for the square potential and its symmetry properties in the complex k plane. The trajectories of the k poles for continuously varying real angular momentum are presented in Sec. 3. Their large angular momentum behavior is considered in Sec. 4. Finally, some aspects of Regge and k trajectories are discussed in Sec. 5.

II. THE S MATRIX

We are considering a particle of mass m and energy E in a square potential of intensity V and range b. Let us denote by k and k' the wavenumbers outside and inside the potential;

$$k^2 = 2mE/\hbar^2, \quad k'^2 = 2m(E - V)/\hbar^2. \tag{2.1}$$

Using dimensionless wavenumbers

$$\alpha = kb, \quad \beta = k'b, \tag{2.2}$$

the l-wave part of the S matrix reads[3]

$$S_l(k) = -\frac{\alpha H_\lambda^{(2)\prime}(\alpha) - H_\lambda^{(2)}(\alpha)\beta J_\lambda'(\beta)/J_\lambda(\beta)}{\alpha H_\lambda^{(1)\prime}(\alpha) - H_\lambda^{(1)}(\alpha)\beta J_\lambda'(\beta)/J_\lambda(\beta)}, \tag{2.3}$$

where we have denoted, as usual,

$$\lambda = l + \tfrac{1}{2}. \tag{2.4}$$

Of course, H and J in Eq. (2.3) refer to the Hankel and Bessel functions[4] and the prime indicates derivative with respect to the argument. Obviously, the poles of $S_l(k)$ are given by the solutions of

$$\alpha H_\lambda^{(1)\prime}(\alpha)/H_\lambda^{(1)}(\alpha) - \beta J_\lambda'(\beta)/J_\lambda(\beta) = 0, \tag{2.5}$$

a relation which merely expresses the matching, at the edge of the potential, of the regular internal solution of the Schrödinger equation with the external solution properly behaved at infinity.

The complex values of λ satisfing Eq. (2.5) for physical (real or pure imaginary) α, i.e., the Regge poles, have been discussed by several authors.[5–7] Here we are interested in the solutions of Eq. (2.5) in the complex k plane for real λ.

Before entering in the detailed analysis of k trajectories, let us recall some general facts about the location, in the complex k plane, of singularities of the S matrix for a real potential. It is well known[2] that, for real positive λ, all poles in the upper half-plane must lie on the imaginary axis. In the lower half-plane, nonreal poles, besides the pure imaginary ones, can appear. Usually it is mentioned that these poles occur in pairs symmetric with respect to the imaginary axis. It should be emphasized, however, that this is so for physical (half-integer) values of λ, but not for arbitrary real λ. Actually, the property[8]

$$S^*(\lambda^*, -k^*) = S(\lambda, k), \tag{2.6}$$

which contains the symmetry of the poles with respect to the imaginary axis, is valid for physical λ, but ceases to be true as λ assumes unphysical values.

III. THE k TRAJECTORIES

Using numerical procedures we have looked for solutions of Eq. (2.5) for given real λ. The reduced logarithmic derivative of the Hankel function was calculated as a quotient of series. Double precision was used in the summation of these series. The reduced logarithmic derivative of the Bessel function was evaluated using a continued fraction ex-

[a)]Work supported by Instituto de Estudios Nucleares.

 0022-2488/81/112594-04$01.00

pansion. The potential was characterized by the dimensionless parameter

$$U = 2mVb^2/\hbar^2. \tag{3.1}$$

The results for a barrier and a well, both of parameter $|U| = 25$, are presented in Figs. 1 and 2.

Two infinite families of k poles are found, corresponding to the two classes of Regge poles for the square potential reported by Nussenzveig.[9] Class I poles are determined by the interior of the potential, whereas Class II poles correspond to surface waves.

The Class I k poles lie in the fourth quadrant near the positive real semiaxis. Obviously, their positions at physical values of λ correspond to resonances. The trajectories followed by these poles as λ increases become nearly coincident. As λ decreases, the k poles reach the origin and go up along the positive imaginary semiaxis. The values of λ at which a k pole enters the origin are given[5] by the solutions of

$$J_{\lambda-1}[(-U)^{1/2}] = 0 \quad \text{for } \lambda > 0, \tag{3.2a}$$

$$J_{\lambda+1}[(-U)^{1/2}] = 0 \quad \text{for } \lambda < 0. \tag{3.2b}$$

In the case of a barrier ($U > 0$) the k poles enter the origin at negative values of λ. For a sufficiently deep well a finite number of solutions, $\lambda_{0,n} > \frac{1}{2}$, of Eq. (3.2a) exist. The poles reaching the origin at these values of $\lambda = \lambda_{0,n}$ appear as bound states for smaller physical values of λ. The motion of the k poles in the vicinity of the origin can be deduced from the threshold behavior of the Regge poles and has been described by Newton.[2]

The Class II k poles are almost regularly distributed on the third and fourth quadrants. As λ increases they follow trajectories going from the third quadrant to the fourth one. These trajectories intersect the negative imaginary semiaxis at odd (even) values of the angular momentum l in the case of barrier (well). As λ decreases, the k poles approach the cut existing on the negative real semiaxis due to the multivaluedness of the reduced logarithmic derivative of the Hankel function for unphysical values of λ. If the intensity $|U|$ of the potential is sufficient, the Class II k poles cross the cut, appear in the second quadrant of a different Riemann sheet, and go up to infinity as λ tends to a physical value. A similar behavior is found for scattering by a hard sphere, the k poles in this case being given by the zeros[10] of the Hankel function $H_\lambda^{(1)}(kb)$, b denoting the radius of the sphere. As λ decreases further, the k poles at infinity make a discontinuous jump of $-\pi/2$ and go down, crossing the cut and coming back to the principal Riemann sheet, where they occupy, at physical values of λ, positions in the third quadrant symmetrical to those of the Class I k poles in the fourth quadrant. These parts of the k trajectories are not shown in Figs. 1 and 2 as they seem to be of less interest from the physical point of view.

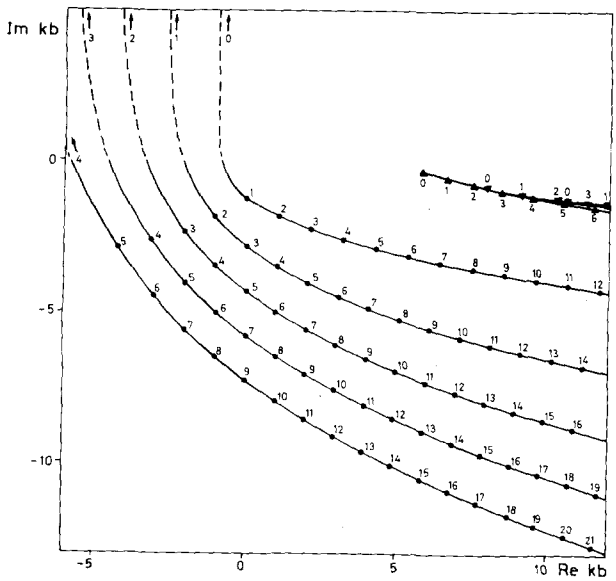


FIG. 1. Trajectories of the S matrix poles in the complex wavenumber plane for a square barrier of parameter $U = 25$. The positions of the poles corresponding to physical values of the angular momentum are denoted by small triangles, inverted triangles, and squares for the first three Class I trajectories and by dots for the first five Class II ones. The numbers along the trajectories correspond to values of the angular momentum.

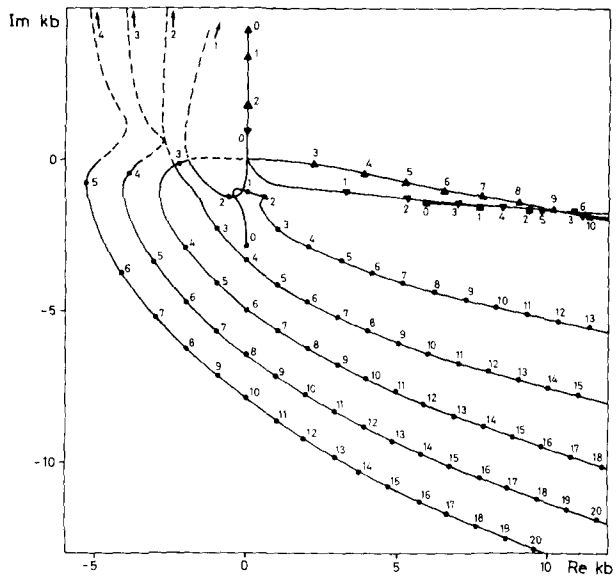


FIG. 2. Trajectories of the S matrix poles in the complex wavenumber plane for a square well of parameter $U = -25$. The notation is the same as in Fig. 1.

The values of λ at which the Class II k poles may cross the cut are such that the reduced logarithmic derivative of the Hankel function (of order λ and variable of argument $-\pi$) becomes real, in order to satisfy Eq. (2.5). This happens for $\lambda = n \pm \frac{1}{3}$, n integer.

It is interesting to observe how the symmetry of the positions of the poles with respect to the imaginary axis is lost for unphysical λ and recovered again when λ takes physical values.

IV. POLES AT LARGE ANGULAR MOMENTUM

Now we are going to consider the behavior of the k trajectories discussed in the preceding section as the angular momentum takes large positive values. A simple inspection of Figs. 1 and 2 suggests that all k poles go to infinity as λ increases without limit. So, we look for solutions of Eq. (2.5) corresponding to $\lambda \to \infty$, $|\alpha| \to \infty$. Moreover, Regge pole studies have revealed[9,11] the existence of two families of poles

of the S matrix located, respectively, near the zeros of $H_\lambda^{(1)}(\alpha)$ and $J_\lambda(\beta)$.

The first family of solutions is more easily obtained by writing the equations of the poles in the form

$$H_\lambda^{(1)\prime}(\alpha) + FH_\lambda^{(1)}(\alpha) = 0, \tag{4.1}$$

where

$$F = -(\beta/\alpha)J_\lambda'(\beta)/J_\lambda(\beta) \tag{4.2}$$

is obviously a function of λ and α. Equation (4.1) belongs to a type already discussed by other authors.[12,13] It can be solved by using the asymptotic expansions of the Hankel function and its derivative, to obtain

$$\alpha_n = \lambda - 2^{-1/3}\exp(-2\pi i/3)c_n\lambda^{1/3} + (\tfrac{3}{10})2^{-2/3}\exp(-4\pi i/3)c_n^2\lambda^{-1/3} + \cdots, \tag{4.3}$$

where c_n denotes the nth solution of the equation

$$\mathrm{Ai}'(c) + 2^{-1/3}\exp(\pi i/3)\lambda^{1/3}F\,\mathrm{Ai}(c) = 0. \tag{4.4}$$

Here Ai denotes the Airy function and the product $\lambda^{1/3}F$ remains nearly constant as λ tends to infinity. The other family of solutions arises analogously from the equation of the poles written in the form

$$J_\lambda'(\beta) + GJ_\lambda(\beta) = 0, \tag{4.5}$$

with

$$G = -(\alpha/\beta)H_\lambda^{(1)\prime}(\alpha)/H_\lambda^{(1)}(\alpha). \tag{4.6}$$

Following a procedure entirely similar to that used to solve Eq. (4.1), one can easily obtain

$$(\alpha_n^2 - U)^{1/2} = \beta_n = \lambda - 2^{-1/3}d_n\lambda^{1/3} + \tfrac{3}{10}2^{-2/3}d_n^2\lambda^{-1/3} + \cdots, \tag{4.7}$$

where d_n is the nth solution of the equation

$$\mathrm{Ai}'(d) + 2^{-1/3}\lambda^{1/3}G\,\mathrm{Ai}(d) = 0. \tag{4.8}$$

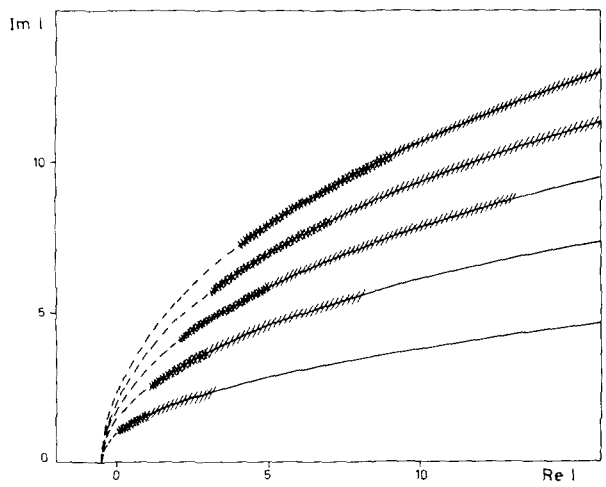


FIG. 3. Class II Regge trajectories for a square barrier of parameter $U = 25$. The unshadowed and simply shadowed sections of the trajectories correspond to k poles in the fourth quadrant located respectively above and below the bisector, the doubly shadowed ones to k poles in the third quadrant, and the dashed sections to k poles in a different Riemann sheet. Only the first five of the infinite set of Regge trajectories are shown.

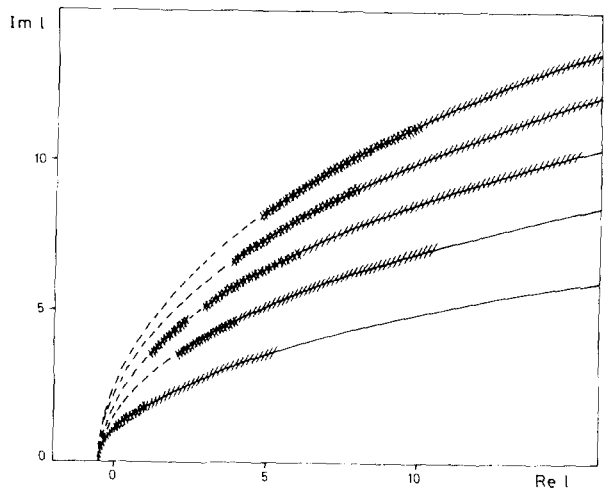


FIG. 4. Class II Regge trajectories for a square well of parameter $U = -25$. The notation is the same as in Fig. 3.

Equations (4.3) and (4.7) give the asymptotic positions of the Class II and Class I k poles, respectively.

V. DISCUSSION

We have found two families of k poles corresponding precisely to the two classes of Regge poles reported earlier[9] for a square potential. In fact, a one-to-one correspondence can be established between Regge trajectories and k trajectories.

Both Class I and Class II k trajectories are related to resonant states at high values of the angular momentum. For each angular momentum an infinity of Class I resonances exists. They are narrow and their energies lie above the discontinuity in the effective (well or barrier plus centrifugal) potential, i.e., they are produced inside the square potential. Class II resonant states are much broader and have energies below such discontinuity: they are formed at the potential surface and are due to the presence of the centrifugal barrier, just as in a similar way to what happens for scattering on a hard sphere.[14]

In what concerns bound states and narrow resonances, i.e., Class I poles, k trajectories contain the same kind of information as Regge trajectories do and they do not seem to add new significant knowledge. In the case of Class II poles, instead, k trajectories suggest some comments.

The physical meaning of complex k poles was thoroughly discussed by Beck and Nussenzveig.[14] It appears that a distinction must be made between poles located above and below the bisector of the fourth quadrant. Actually, only poles above the bisector can be interpreted as decaying states. In view of this, a given Class II k pole related to broad resonances at very large values of the angular momentum loses its meaning as its k trajectory crosses the bisector. As a matter of fact, the associated resonances have increasing relative width (quotient of the imaginary to the real parts of the energy) as the angular momentum decreases. The relative width becomes infinite as the k pole crosses the bisector.

In view of the one-to-one correspondence between Regge and k trajectories it is possible to distinguish, on a

Class II Regge trajectory, between points corresponding to k poles above and below the bisector, respectively. The correspondence procedure suggested by Newton[2] has allowed us to obtain the results shown in Figs. 3 and 4 for barrier and well potential, respectively. Regge poles corresponding to k poles above the bisector can be interpreted as surface creep waves,[1] whereas the meaning of Regge poles corresponding to k poles below the bisector is obscure.

ACKNOWLEDGMENT

The authors acknowledge the Secretaría de Estado de Universidades e Investigación for financial help during the completion of this work.

[1]K. W. McVoy, "Semiclassical methods and surface creep waves in heavy-ion reactions," Lectures presented at the Escuela Latino-Americana de Física, Universidad Nacional Autónoma de México, 1977.
[2]R. G. Newton, *The complex j-plane* (Benjamin, New York, 1964).
[3]H. M. Nussenzveig, Nucl. Phys. **11**, 499 (1959).
[4]F. W. J. Olver, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Dover, New York, 1965), p. 355.
[5]A. O. Barut and F. Calogero, Phys. Rev. **128**, 1383 (1962).
[6]C. J. Bollini and J. J. Giambiagi, Nuovo Cimento **26**, 619 (1962); *ibid.* **28**, 341 (1963).
[7]A. Z. Patashinskii, V. L. Pokrovskii, and I. M. Khalatnikov, J. Exptl. Theoret. Phys. **44**, 2062 (1963) [Sov. Phys. JETP **17**, 1387 (1963)].
[8]R. G. Newton, J. Math. Phys. **3**, 867 (1962).
[9]H. M. Nussenzveig, J. Math. Phys. **10**, 82 (1969).
[10]E. Jahnke, F. Emde, and F. Lösch, *Tables of higher functions* (McGraw-Hill, New York, 1960), p. 229.
[11]R. Anni and L. Taffara, Nuovo Cimento A **31**, 321 (1976).
[12]J. B. Keller, S. I. Rubinow, and M. Goldstein, J. Math. Phys. **4**, 829 (1963).
[13]J. A. Cochran, Numer. Math. **7**, 238 (1965).
[14]G. Beck and H. M. Nussenzveig, Nuovo Cimento **16**, 416 (1960).

Spacetime G structures and their prolongations[a)]

Robert Alan Coleman
Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, S4S OA2, Canada
Herbert Korte
Department of Philosophy, University of Regina, Regina, Saskatchewan, S4S OA2, Canada

(Received 30 December 1980; accepted for publication 11 May 1981)

An elementary procedure for obtaining the prolongations of arbitrary G structures is described and applied to the affine, projective, Lorentz, conformal, and Weyl structures. Also, the condition that a projective and conformal structure be compatible and thereby define a unique Weyl structure is discussed from the G-structure viewpoint. Throughout, an invariant notation for jets is employed.

PACS numbers: 04.20.Cv, 02.40. + m

1. INTRODUCTION

The geometric structures required for the formulation of the constructive axioms of general relativity theory proposed by Ehlers, *et al.*[1] are the affine, projective, Lorentz, conformal, and Weyl structures.[2,3] In this paper, these geometric[4] structures are all discussed in a uniform manner as group or G structures of first and higher order. This approach has a number of advantages. First, G structures are naturally described by gauge fields in a way that is described below in Sec. 2. Second, the theory of prolongations of G structures reveals how a G structure of a given order determines an infinite sequence of G structures of progressively higher order and leads to the notion of normal coordinates for each type of G structure. Also, the interrelationships of the geometric structures are more apparent because the groups which characterize the geometric structures play a more prominent role in the G-structure approach, a fact which also emphasizes the role of symmetry which is of central importance for spacetime theories.

The standard approach to the theory of G structures[5-7] and their prolongations evolved from the analysis of automorphism groups of G structures. Kobayashi[8] has given a quite readable account of this approach. A very elegant and quite thorough presentation of the standard theory of G structures has been given by P. Molino.[9]

The theory of G structures may also be subsumed under the broader theory of systems of partial differential equations and Lie pseudogroups developed in the work of Spencer,[10] Guillemin and Sternberg.[11] Important new contributions to this theory have been presented by J. F. Pommaret[12] in a recent book. Some of the G structures mentioned above are discussed therein from this more advanced and general but less intuitive perspective.[13]

The presentation given below focuses solely on the description of the above mentioned G structures and the process of prolongation of these G structures to higher order. Although the results obtained are the same as those obtained by the standard approach mentioned above, the conceptual basis for the prolongation procedure is quite different. Moreover, the prolongation procedure used below rests on *entirely elementary considerations* which have a *vividly intuitive interpretation*, and it is simple to apply. The procedure may be described using either frames or coframes, but it is more simply stated using coframes.

Briefly, if $h:U\to\mathbb{R}^n$ is a local diffeomorphism of an open neighborhood U of a manifold M into $\mathbb{R}^n$, then an n^k coframe at a point $p\in M$ is a k-jet $j^k_p[h-h(p)]$ for some h, where $h(p)$ is the constant function on U with image $h(p)$. Roughly, the chart (U,h) provides an image of the neighborhood U of M and the n^k coframe at p is the k th order Taylor polynomial approximation to $[h-h(p)]$ at p. It is clear that the Taylor polynomial approximation of order k at a point p determines at an infinitesimally near point q the Taylor polynomial approximation of order $k-1$ at q to first order in the coordinate difference $x^i(q)-x^i(p)$; that is, an n^k coframe at p determines an n^{k-1} coframe at q. Since a G structure of order $k-1$ specifies *at each* $p\in M$ an equivalence class of n^{k-1} coframes, it is clear that a G structure of order k must satisfy the constraint that it induce the same equivalence classes as the lower order G structure in order to be a prolongation. In fact, this simple constraint generates not only the prolonged G structure but also the prolonged gauge group which characterizes it. This prolongation procedure which rests on the notion of approximation to the manifold is discussed more fully in Sec. 2.

In Secs. 3–7, the affine, projective, Lorentz, conformal, and Weyl structures are treated. The sequence of presentation reflects the increase in algebraic complexity. The affine, projective, and conformal structures are prolonged to 3rd order, beyond which order no novel features appear for these G structures. For the Lorentz and Weyl structures, one may stop at 2nd order because these structures define a unique affine structure and further prolongations may be made by appeal to the results for the affine case.

In Sec. 8, the condition that a conformal and a projective structure be compatible and hence define a unique Weyl structure, a condition originally derived by Ehlers *et al.*[1] is discussed from the G structure viewpoint.

Although the prolongation procedure is applied to only five G structures in this paper, it may be applied with equal facility to any G structure such as symplectic, complex, Hermitian, and Kaehler structures. Also, one of the authors (Coleman) and R. Mann[14] have applied the procedure to subspace G structures of which the Galilean structure is a

a)Supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/81/112598-14$01.00

special case.

Indeed, the procedure is applicable in more general circumstances; for example, Rogers[15] has presented a view of supermanifolds in which k jets of mappings are well defined and the k jet of a composition is the composition of the k jets. Since the constructions needed rest on these properties, the theory of super coframe bundles, super G structures and their prolongations could be carried through in a manner directly analogous to that used for the theory of such structures on ordinary manifolds.

Throughout the paper, an invariant notation for jets is used which is explained in Appendix A. The germ of the idea behind the notation may be found in the work of Ehresmann.[16] The notation used for the jet bundles that are employed is described in Appendix B and the notation used for the various groups that appear is described in Appendix C.

2. G STRUCTURES AND PROLONGATIONS

Let $\mathscr{P}(M)$ be a principal fiber bundle with total space $P(M)$, base space M, projection $\pi_P: P(M)\to M$ and structure group G; that is, $\mathscr{P}(M)$ is the structure

$$\mathscr{P}(M) = \langle P(M), \pi_P, M, G\rangle. \tag{2.1}$$

Let H be a closed subgroup of G. An H structure on M is a reduction to the subgroup H of the structure group G of $\mathscr{P}(M)$. Such structures are in bijective correspondence with cross sections of the associated fiber bundle

$$\mathscr{P}(M)/H = \langle P(M)/H, \pi_{P/H}, M, G/H, \mathscr{P}(M)\rangle \tag{2.2}$$

of orbits under the action of H on elements of $P(M)$. In the original context, $\mathscr{P}(M)$ was the principal bundle of linear frames and the closed subgroup of the structure group $G^1_n(\equiv \mathrm{GL(n)})$ was denoted by G rather than H.

An H structure on M given by a cross section of the bundle (2.2) specifies an equivalence class of H related elements of $P(M_p)$ for every $p\in M$. Since equivalence classes are difficult to work with, the cross section $\sigma: M\to P(M)/H$ is represented by a family of local cross sections $\sigma_U: U\to P(U)$ such that

$$\forall\, p\in U,\quad \sigma_U(p)\in\sigma(p). \tag{2.3}$$

Moreover, if σ_{U_1} and σ_{U_2} are any two such local cross sections such that $U_1\cap U_2\neq\phi$, then there must exist a cross section $\rho_{12}: U_1\cap U_2\to(U_1\cap U_2)\times H$ such that

$$\forall\, p\in U_1\cap U_2,\quad \sigma_{U_2}(p) = \sigma_{U_1}(p)\circ\rho_{12}(p). \tag{2.4}$$

The transformation (2.4) is a local gauge transformation. Note that it is an active transformation and not a passive coordinate transformation. For a given computation, it is often convenient to work with a particular local cross section σ_U. Such a representative is usually selected by imposing a gauge-fixing condition which amounts to choosing a particular set of coordinates for the coset space G/H. Such conditions are coordinate dependent because the local trivializing maps depend on the chart (U,x). Suitable examples of such conditions will appear below in the analysis of the various G structures.

The relevant principal bundles for the analysis of spacetime G structures and their prolongations are the bundles of frames or coframes of order k. While the prolongation procedure used in this paper can be stated either in terms of frames or coframes, it is more natural to use coframes. The bundle of n^k coframes is denoted by

$$\mathscr{H}^k{}^*(M) = \langle H^k{}^*(M), \pi_{H^k{}^*}, M, G^k_n\rangle. \tag{2.5}$$

Let $p\in U\subset M$ and let $h: U\to\mathbb{R}^n$ be a local diffeomorphism, and let $h(p)$, the image of p in $\mathbb{R}^n$, also denote the constant map $h(p): U\to\mathbb{R}^n$ such that $\forall q\in U[h(p)(q) = h(p)]$. A pair (U,h), called a chart, provides an image of a local neighborhood U in the standard space $\mathbb{R}^n$. If only one such image were allowed, the local region would have all the rigidity of $\mathbb{R}^n$; however, for a manifold with only a differentiable structure, all such images are allowed provided that they are smoothly related and compatible. An element of $H^k{}^*(M_p)$, called an n^k coframe at p, is a k jet $j^k_p(h - h(p))$ which is the equivalence class of local diffeomorphisms which agree with $h - h(p)$ at p and have the same Taylor expansion of order k at p with respect to some arbitrarily chosen chart (U,x) for a neighborhood of p. Thus an n^k coframe is essentially the k th order polynomial approximation to a chart at a given point p.

A local n^k coframe field is given by a local cross section $h^k: U\to H^k{}^*(U)$ which is represented with respect to a given chart (U,x) by

$$h^k(p) = \partial^k_0 I_n(h^i_j d^k_p x^j + (1/2!)\, h^i_{j_1 j_2} d^k_p x^{j_1} d^k_p x^{j_2} + \cdots + (1/k!)\, h^i_{j_1 j_2\cdots j_k} d^k_p x^{j_1} d^k_p x^{j_2}\cdots d^k_p x^{j_k}), \tag{2.6}$$

where the coefficients $(h^i_j, h^i_{j_1 j_2},\ldots, h^i_{j_1 j_2\cdots j_k})$ are functions of p and are symmetric with respect to their lower indices.

A field of n^1 coframes is a rather restrictive geometric structure because at each point p only a single linear image of the neighborhood of p is regarded as giving an appropriate representation. More flexible geometric structures may be defined by specifying at each p a class of distinguished n^1 coframes rather than a unique one; for example, for any closed *Lie* subgroup SG^1_n of G^1_n, a cross section of the associated fiber bundle $SG^1_n\backslash\mathscr{H}^1{}^*(M)$ defines a field of equivalence classes of SG^1_n related n^1 coframes, an SG^1_n structure. If $SG^1_n = G^1_n$, then all linear images of the neighborhood of any point p are regarded as equivalent and the manifold M acquires none of the structural features of $\mathbb{R}^n$. At the other extreme, SG^1_n consists of the identity element alone and the manifold M acquires a great deal of the structure of $\mathbb{R}^n$. For other choices of SG^1_n, such as $O^1_{p,q}$ and $C^1_{p,q}$, the infinitesimal neighborhoods of points $p\in M$ acquire structural attributes from $\mathbb{R}^n$ that have the microsymmetry[17] of the corresponding group.

Structures of higher order are defined in a similar way. If SG^k_n is a closed Lie subgroup of G^k_n, then an SG^k_n structure on a manifold M is given by a cross section of the associated fiber bundle $SG^k_n\backslash\mathscr{H}^k{}^*(M)$. Such a cross section specifies at each $p\in M$ an equivalence class of SG^k_n related n^k coframes. As in the first-order case, structure is pulled back from $\mathbb{R}^n$ by virtue of the fact that not every k th order approximate image of the infinitesimal neighborhood of a given point p is considered faithful. However, for $k>1$, there is an additional self-consistency condition which the SG^k_n structure must satisfy, a condition which underlies the prolongation procedure used throughout the paper.

The self-consistency condition follows from the elementary observation that the k th order Taylor expansion of a function at some given point determines the $(k-1)$th order Taylor expansion of that function at an infinitesimally near point accurately to first order in the infinitesimal vector which describes the separation of the two points. Let $w^i = x^i(q) - x^i(p)$, where p and q are infinitesimally near points in M. Then

$$x^i - x^i(p) = x^i - x^i(q) + w^i. \tag{2.7}$$

An n^k coframe at p given by (2.6) determines an n^{k-1} coframe at q by means of the replacement

$$d_p^k x^i \to d_q^{k-1} x^i + w^i. \tag{2.8}$$

Note that the constant term is discarded because a coframe by definition has zero constant term. The result is

$$\begin{aligned}\partial_0^{k-1} I_n &((h^i_j + h^i_{j,\ell} w^\ell) d_q^{k-1} x^{j_1} \\ &+ (1/2!)(h^i_{j_1 j_2} + h^i_{j_1 j_2,\ell} w^\ell) d_q^{k-1} x^{j_1} d_q^{k-1} x^{j_2} + \cdots \\ &+ (1/(k-1)!)(h^i_{j_1 j_2 \cdots j_{k-1}} + h^i_{j_1 j_2 \cdots j_{k-1},\ell} w^\ell) \\ &\times d_q^{k-1} x^{j_1} d_q^{k-1} x^{j_2} \cdots d_q^{k-1} x^{j_{k-1}}.\end{aligned} \tag{2.9}$$

An n^k-coframe field determines an n^{k-1}-coframe field simply by the discarding of the k th order term. The n^k coframe at q determines an n^{k-1} coframe at q; namely,

$$\begin{aligned}\partial_0^{k-1} I_n &(h^i_j + h^i_{j,\ell} w^\ell d_q^{k-1} x^j \\ &+ (1/2!) h^i_{j_1 j_2} + h^i_{j_1 j_2,\ell} w^\ell) d_q^{k-1} x^{j_1} d_q^{k-1} x^{j_2} \\ &+ \cdots \\ &+ (1/(k-l)!)(h^i_{j_1 j_2 \cdots j_{k-1}} + h^i_{j_1 j_2 \cdots j_{k-1},\ell} w^\ell) \\ &\times d_q^{k-1} x^{j_1} d_q^{k-1} x^{j_2} \cdots d_q^{k-1} x^{j_{k-1}}),\end{aligned} \tag{2.10}$$

where the coefficients describing the n^{k-1} coframe at q have been reexpressed in terms of the field coefficients and their derivatives at p means of a 1st order Taylor expansion. If the n^k coframe field is coherent, then each of the n^k coframes should be the k th order Taylor approximation at the appropriate point in M of one and the same chart. Thus self-consistency requires that the n^{k-1} coframes (2.9) and (2.10) be identical. However, there are two additional complications to consider.

First, one is usually not concerned with a simple n^k-coframe field, but rather with an SG_n^k structure which specifies at each $p \in M$ not a unique n^k coframe but an equivalence class of SG_n^k related coframes. Thus the self-consistency condition is weakened to the requirement that the n^{k-1}-coframes (2.9) and (2.10) are the same up to an SG_n^{k-1} transformation where SG_n^{k-1} is the group obtained by discarding the k th order terms of the elements of SG_n^k. This complication may be efficiently handled by transforming (2.9) and (2.10) to standard representatives before making the identification.

Second, it is clear that the coefficients of w^ℓ in (2.9) are necessarily symmetric with respect to all of their lower indices since they arise as the coordinates of a k jet. However, the corresponding coefficients of w^ℓ in (2.10) need not be totally symmetric in their lower indices since they arise by differentiation of the coefficient functions describing a $(k-1)$-jet field. In general, (2.9) and (2.10) differ by a term of the form $B^i_{j_1 j_2 \cdots j_r \ell} w^\ell$ for each $r \in \{1,2,\ldots,k-1\}$, where $B^i_{j_1 j_2 \cdots j_r \ell}$ has the symmetry with respect to the lower indices obtained by first antisymmetrizing with respect to j_r and ℓ and then symmetrizing with respect to $j_1, j_2, \ldots j_r$.

The prolongation procedure may now be described as follows: assume that a self-consistent SG_n^{k-1} structure is given; for any given point $p \in M$, assume for an arbitrary n^{k-1}-coframe at p an extension to an n^k coframe; compute the n^{k-1}-coframes at q corresponding to (2.9) and (2.10) and reduce these to standard form using SG_n^{k-1} transformations; equate the results allowing for the terms of the form $B^i_{j_1 j_2 \cdots j_r \ell} w^\ell$. This procedure results in a system of linear equations for the k th order coefficient of the n^k coframe. These equations determine both the typical element of the prolonged group SG_n^k and the standard representative of the equivalence class of n^k coframes belonging to the SG_n^k structure at p with respect to the chosen coordinate system. It may happen that the system of equations is underdetermined in which case the group SG_n^k has new parameters in addition to those of SG_n^{k-1}. Such new parameters arise in the prolongation of the conformal or $C^1_{p,q}$ structure to a $C^2_{p,q}$ structure.

For every case, the process must begin at some lowest order. Not every case begins in the first order. The affine and projective structures begin at second order, and although the conformal and Weyl structures are indistinguishable in first order, they are distinguished at second order by the imposition of additional structure in the Weyl case. In all of the cases discussed below, no new group parameters arise beyond second order; however, in the case of subspace structures which will be discussed elsewhere,[14] new group parameters arise at each successive order.

We conclude this section with a formal characterization of the notion of prolongation. For any bundle $\mathscr{E}(M)$, denote by $J^r\mathscr{E}(M)$ the bundle of r jets of local cross sections of $\mathscr{E}(M)$. An SG_n^k structure is determined by a cross section $\sigma^k: M \to SG_n^k \backslash H^{k*}(M)$, which in turn may be described by a family of local cross sections $h^k: U \to H^{k*}(U)$ which satisfy $h^k(p) \in \sigma^k(p)$ for $p \in U$. Such a local cross section is given explicitly by (2.6). The natural projection maps $G_n^k \to G_n^{k-1}$ and $H^{k*}(M) \to H^{k-1*}(M)$ defined by the discarding of the k th order term induce projection maps $SG_n^k \to SG_n^{k-1}$ and $SG_n^k \backslash H^{k*}(M) \to SG_n^{k-1} \backslash H^{k-1*}(M)$. Thus the compositions of the maps σ^k and h^k with the appropriate projection maps yield maps $\sigma^{k-1}: M \to SG_n^{k-1} \backslash H^{k-1*}(M)$ and $h^{k-1}: U \to H^{k-1*}(U)$. In this way, every SG_n^k structure uniquely determines an SG_n^{k-1} structure by truncation.

In an infinitesimal neighborhood of $p \in U \subset M$, h^{k-1} is determined to first order by $j_p^1 h^{k-1} \in J^1 H^{k-1*}(M_p)$. Explicitly,

$$\begin{aligned}j_p^1 h_{k-1} = \partial_0^{k-1} I_n &((h^i_j + h^i_{j,\ell} d_p^1 x^\ell) j_p^1 d^{k-1} x^j + \cdots \\ &+ \frac{1}{(k-1)!} (h^i_{j_1 \cdots j_{k-1}} + h^i_{j_1 \cdots j_{k-1},\ell} d_p^1 x^\ell) \\ &\times j_p^1 d^{k-1} x^{j_1} \cdots j_p^1 d^{k-1} x^{j_{k-1}}).\end{aligned} \tag{2.11}$$

The expression (2.10) is obtained by composing the 1-jet aspect of (2.11) with the 1-jet $\partial_p^1 x(w^i d_0^1 I)$.

Define a mapping $e: H^{k*}(M) \to J^1 H^{k-1*}(M)$ by means of the substitution

$$d_p^k x \to j_p^1 d^{k-1} x + d_p^1 x \tag{2.12}$$

with the understanding that the constant term of the $(k-1)$ jet is discarded. One obtains

$$e[h^k(p)] = \partial_0^{k-1} I_n((h^i_j + h^i_{j'} d_p^1 x^{j'}) j_p^1 d^{k-1} x^j + \ldots + (1/(k-1)!)(h^i_{j_1 \ldots j_k} + h^i_{j_1 \ldots j_{k-1} j'} d_p^1 x^{j'}) j_p^1 d^{k-1} x^{j_1} \ldots j_p^1 d^{k-1} x^{j_k}). \tag{2.13}$$

The expression (2.9) is related to (2.13) in the same way that (2.10) is related to (2.11).

Both $j_p^1 h^{k-1}$ and $e(h^k(p))$ may be regarded as generalized elements of $H^{k-1*}(M)$ with 1-jet coefficients. The inverse $e(h^k(p))^{-1}$ is a generalized element of $H^{k-1}(M)$. The compostion $j_p^1 h^{k-1} \circ e(h^k(p))^{-1}$ [regarded as generalized $(k-1)$ jets] has the form

$$\partial_0^{k-1} I_n [(\delta^i_j + E^i_{j'} d_p^1 x^{j'}) d_0^{k-1} I_n^j + \ldots + \frac{1}{(k-1)!} E^i_{j_1 \ldots j_k j'} d_p^1 x^{j'} d_0^{k-1} I_n^{j_1} \ldots d_0^{k-1} I_n^{j_k}]. \tag{2.14}$$

This is an infinitesimal element of G_n^{k-1}, where the sense of infinitesimal is made precise by the use of the 1 jet $d_p^l x$.

The concept of a self-consistent SG_n^k structure may now be formulated as follows.

Definition: An SG_n^k structure is *self-consistent* iff $j_p^1 h^{k-1} \circ e(h^k(p))^{-1}$ is the composition of an infinitesimal element of SG_n^{k-1} and an infinitesimal element of G_n^{k-1} of the form (2.14) with E replaced by $\hat{B}$ such that the $\hat{B}^i_{j_1 \ldots j_r q} h^{-1q}_{j'}$ have the same type of symmetry with respect to the lower indices as $B^i_{j_1 \ldots j_r j'}$ discussed above.

Finally, the definition of "prolongation" may be stated as follows.

Definition: An SG_n^k structure is the *prolongation* of an SG_n^{k-1} structure iff the SG_n^{k-1} structure is determined by the SG_n^k structure by truncation and the SG_n^k structure is self-consistent.

Note that the self-consistency of the SG_n^k structure entails the self-consistency of the induced SG_n^{k-1} structure.

3. THE AFFINE STRUCTURE

It is customary to describe the affine structure of spacetime as a connection on the principal bundle of linear frames or n-frames $\mathscr{H}^1(M)$. In this section, however, the affine structure will be presented as a G structure defined by a reduction to the affine subgroup Γ_n^2 of the structure group G_n^2 of the principal bundle of n^2 coframes

$$\mathscr{H}^{2*}(M) = \langle H^{2*}(M), \pi_{H^{2*}}, M, G_n^2 \rangle. \tag{3.1}$$

This approach has the advantages that other related structures can be described in a similar way, that the structure is described explicitly by local Γ_n^2 gauge fields, and that the concept of approximating the manifold by second-order polynomials is emphasized. Moreover, the relationship of the affine structure to affine curves and parallel transport and to the higher-order prolongations of these structures is quite direct and intuitive from the viewpoint of higher-order approximation of the manifold.

The group Γ_n^2 is the subgroup of G_n^2 of elements of the form

$$\partial_0^2 I_n(a^i_j d_0^2 I^j_n). \tag{3.2}$$

The restriction of the left, free action of G_n^2 on $H^{2*}(M)$ to the subgroup Γ_n^2 may be used to construct the associated bundle of equivalence classes of Γ_n^2 related n^2 coframes

$$\Gamma_n^2 \backslash \mathscr{H}^{2*}(M) = \langle \Gamma_n^2 \backslash H^{2*}(M), \pi_{\Gamma_n^2 \backslash H^{2*}}, M, \Gamma_n^2 \backslash G_n^2, \mathscr{H}^{2*}(M) \rangle, \tag{3.3}$$

where the typical fiber $\Gamma_n^2 \backslash G_n^2$ is the set of left cosets. The affine structure Γ is a cross section $\Gamma: M \to \Gamma_n^2 \backslash H^{2*}(M)$. Since for each $p \in M$, $\Gamma(p)$ is an equivalence class of n^2 coframes, the affine structure may be represented locally by cross sections $h: U \to H^{2*}(U)$,

$$h(p) = \partial_0^2 I_n(h^i_j d_p^2 x^j + (1/2!) h^i_{jk} d_p^2 x^j d_p^2 x^k), \tag{3.4}$$

which are determined up to a Γ_n^2 left-acting gauge transformation

$$L: U \to U \times \Gamma_n^2,$$
$$L(p) = \partial_0^2 I_n(a^i_j d_0^2 I^j_n). \tag{3.5}$$

Thus under a gauge transformation,

$$\tilde{h}(p) = L(p) \circ h(p) \times \partial_0^2 I_n(a^i_{j'} h^{j'}_j d_p^2 x^j + \frac{1}{2!} a^i_{j'} h^{j'}_{jk} d_p^2 x^j d_p^2 x^k). \tag{3.6}$$

Whenever desired, the gauge may be chosen by the coordinate dependent gauge fixing condition

$$a^i_{j'} h^{j'}_j = \delta^i_j. \tag{3.7}$$

Then, the only nontrivial coefficient in (3.6) is

$$\Gamma^i_{jk} = h^{-1i}_r h^r_{jk}. \tag{3.8}$$

The bundle of second order curve elements is

$$\mathscr{L}_1^2(M) = \langle L_1^2(M), \pi_{L_1^2}, M, L_{1,n}^2, \mathscr{H}^2(M) \rangle. \tag{3.9}$$

If $\gamma: \mathbb{R} \to M$ and $\gamma(0) = p$, then an element $j_0^2 \gamma \in L_1^2(M_p)$ is given by

$$j_0^2 \gamma = \partial_p^2 x(\gamma^i_1 d_0^2 I + (1/2!) \gamma^i_2 (d_0^2 I)^2). \tag{3.10}$$

The n^2 coframe (3.4) may be used to give an image of this curve element in $\mathbb{R}^n$.

$$h(p) \circ j_0^2 \gamma = \partial_0^2 I_n [h^i_j \gamma^j_1 d_0^2 I + (1/2!)(h^i_r \gamma^r_2 + h^i_{rs} \gamma^r_1 \gamma^s_1)(d_0^2 I)^2]. \tag{3.11}$$

Thus the image line element is linear to 2nd order iff

$$\gamma^i_2 + \Gamma^i_{jk} \gamma^j_1 \gamma^k_1 = 0, \tag{3.12}$$

where (3.8) has been used.

The idea that n^k coframes are k th order Taylor approximations to local diffeomorphisms can be used to explicate the relationship of the affine structure to the affine connection and the prolongation of these structures to higher order. Let $f: U \to V \subset \mathbb{R}^n$ be a local diffeomorphism such that $f(p) = 0$. Let q be infinitesimally near p and set $\hat{f} = f - f(q)$. Then $j_q^{k-1} \hat{f}$ is determined by $j_p^k f$ to first order in the coordinate difference $w = x(q) - x(p)$, where (U, x) is a chart. Set $F = f \circ x^{-1}$ and $\hat{F} = \hat{f} \circ x^{-1}$. Then to third order

$$F^i(x) = F^i_j x^j_p + \frac{1}{2!} F^i_{jk} x^j_p x^k_p + \frac{1}{3!} F^i_{jk'} x^j_p x^k_p x^{k'}_p, \tag{3.13}$$

where $x_p^i = x^i - x^i(p)$. Since

$$x_p^i = x_q^i + w^i \quad \text{and} \quad f^i(q) = F_j^i w^j,$$

$$\begin{aligned}\hat{F}^i(x) &= F^i(x) - F_j^i w^j \\ &= F_j^i x_q^j + \frac{1}{2!} F_{jk}^i x_q^j x_q^k + \frac{1}{3!} F_{jk\ell}^i x_q^j x_q^k x_q^\ell + F_{js}^i x_q^j w^s \\ &\quad + \frac{1}{2!} F_{jks}^i x_q^j x_q^k w^s,\end{aligned} \tag{3.14}$$

where only terms up to first order in w^i have been kept. An affine structure defines an equivalence class of second-order coframes; namely,

$$\partial_0^2 I_n(a_j^i d_p^2 x^j + \frac{1}{2!} a_\ell^i \Gamma_{jk}^\ell d_p^2 x^j d_p^2 x^k) \tag{3.15}$$

for $(a_j^i) \in G_n^1$. Each such second-order coframe at p determines a first-order coframe at p

$$\partial_0^1 I_n(a_j^i d_p^1 x^j) \tag{3.16}$$

by projection and a first-order coframe at an infinitesimally near point q

$$\partial_0^1 I_n((a_j^i + a_\ell^i \Gamma_{js}^\ell w^s) d_q^1 x^j) \tag{3.17}$$

by the substitution (3.14). The *parallel transport* of coframes is defined by mapping (3.16) into (3.17).

Now consider the problem of prolonging the affine structure to the next order. An extension of an arbitrary n^2 coframe (3.15) to a n^3 coframe has the form

$$\partial_0^3 I_n\left[a_j^i d_p^3 x^j + \frac{1}{2!} a_s^i \Gamma_{jk}^s d_p^3 x^j d_p^3 x^k + \frac{1}{3!} h_{jk\ell}^i d_p^3 x^j d_p^3 x^k d_p^3 x^\ell\right], \tag{3.18}$$

which by the substitution (3.14) defines a n^2-coframe equivalence class at q near p which contains

$$\partial_0^2 I_n((a_j^i + a_s^i \Gamma_{jr}^s w^r) d_q^2 x^j + \frac{1}{2!}(a_s^i \Gamma_{jk}^s + h_{jks}^i w^s) d_q^2 x^j d_q^2 x^k). \tag{3.19}$$

However, the affine structure also defines another equivalence class of n^2 coframes at q, namely, the equivalence class containing

$$\partial_0^2 I_n(a_j^i d_q^2 x^j + \tfrac{1}{2} a_s^i (\Gamma_{jk}^s + \Gamma_{jk,r}^s w^r) d_q^2 x^j d_q^2 x^k). \tag{3.20}$$

It is natural to require that the two equivalence classes defined by (3.19) and (3.20) be the same and to impose this condition by reducing each of the n^2 coframes to standard form and equating the results; however, it is necessary to allow for a term of the form $B_{jkr}^i w^r$, where B_{jkr}^i has the symmetry in the lower indices obtained by first antisymmetrizing in the indices k and r and then symmetrizing in the indices j and k. Such a term arises because the coefficients describing the 1 jet of the local 2-jet field need not be symmetric in the lower indices. Similar terms arise in higher orders. One obtains

$$\begin{aligned}&\Gamma_{jk}^i + \Gamma_{jk,r}^i w^r + B_{jkr}^i w^r \\ &\quad = \Gamma_{jk}^i + a^{-1i}{}_s h_{jkr}^s w^r - \Gamma_{sr}^i \Gamma_{jk}^s w^r.\end{aligned} \tag{3.21}$$

Set $\Gamma_{jk\ell}^i = a^{-1i}{}_s h_{jk\ell}^s$. Then, the relation (3.21) gives, after appropriate symmetrization, the result

$$\begin{aligned}\Gamma_{jk\ell}^i &= \tfrac{1}{3}(\Gamma_{jk,\ell}^i + \Gamma_{k\ell,j}^i + \Gamma_{\ell j,k}^i \\ &\quad + \Gamma_{sj}^i \Gamma_{k\ell}^s + \Gamma_{sk}^i \Gamma_{\ell j}^s + \Gamma_{s\ell}^i \Gamma_{jk}^s),\end{aligned} \tag{3.22}$$

which upon substitution into (3.18) gives the affine frame of third order corresponding to the affine frame of second order (3.15). Clearly, this procedure could be iterated order-by-order to define a reduction of the structure group G_n^k of $\mathscr{H}^{k*}(M)$ to the affine subgroup Γ_n^k. Moreover, such a reduction defines a connection on the principal bundle $\mathscr{H}^{k-1*}(M)$; for example, suppose affine n^3 coframes have been defined by (3.18) with $h_{jk\ell}^i = a_s^i \Gamma_{jk\ell}^s$, where $\Gamma_{jk\ell}^i$ is given by (3.22). Then by projection one obtains the frames (3.15) and by substitution the frames (3.19). Taking the composition of (3.19) with the inverse of (3.15) [applied on the left of (3.19)], one obtains an invertible map from $H^{2*}(M_p)$ to $H^{2*}(M_q)$; namely,

$$\partial_p^2 x\left[(\delta_j^i + \Gamma_{jr}^i w^r) d_q^2 x^j + \frac{1}{2!}(\Gamma_{jkr}^i w^r - \Gamma_{sj}^i \Gamma_{kr}^s w^r - \Gamma_{sk}^i \Gamma_{jr}^s w^r) d_q^2 x^j d_q^2 x^k\right]. \tag{3.23}$$

To parallel transport the second-order curve element $j_0^2\gamma$ given by (3.10) from p to q, the inverse of (3.23) is applied on the left of $j_0^2\gamma$. Also, note that the image of the third order curve element

$$j_0^3\gamma = \partial_p^3 x\left[\gamma_1^i d_0^3 I + \frac{1}{2!}\gamma_2^i (d_0^3 I)^2 + \frac{1}{3!}\gamma_3^i (d_0^3 I)^3\right] \tag{3.24}$$

under an affine n^3 coframe is

$$\partial_0^3 I_n\left[a_s^i \gamma_1^s d_0^3 I + \frac{1}{2!}(a_s^i \gamma_2^s + a_s^i \Gamma_{jk}^s \gamma_1^j \gamma_1^k)(d_0^3 I)^2 + \frac{1}{3!}(a_s^i \gamma_3^s + 3 a_s^i \Gamma_{jk}^s \gamma_2^j \gamma_1^k + a_s^i \Gamma_{jk\ell}^s \gamma_1^j \gamma_1^k \gamma_1^\ell)(d_0^3 I)^3\right]. \tag{3.25}$$

This image curve is linear to third order provided that (3.12) holds and

$$\gamma_3^i + 3\Gamma_{jk}^i \gamma_2^j \gamma_1^k + \Gamma_{jk\ell}^i \gamma_1^j \gamma_1^k \gamma_1^\ell = 0 \tag{3.26}$$

holds. The result agrees with the usual result obtained by differentiating the equation for affine curves. In addition, the prolongation to order k gives the k jet of the coordinate transformation that relates the given coordinate system to the normal coordinate system for a given point p. In the normal coordinate system all of the functions $\Gamma_{jk}^i, \Gamma_{jk\ell}^i, \ldots$ vanish at p.

4. THE PROJECTIVE STRUCTURE

The projective structure of space-time is a second-order G structure defined by a reduction of the group G_n^2 of $\mathscr{H}^{2*}(M)$ to the subgroup P_n^2 consisting of elements of the form

$$\partial_0^2 I_n(a_j^i d_0^2 I^j + \frac{1}{2!}(a_j^i a_k + a_k^i a_j) d_0^2 I_n^j d_0^2 I_n^k). \tag{4.1}$$

As in the affine case, the restriction of the action of G_n^2 on $H^{2*}(M)$ to the subgroup P_n^2 may be used to construct the associated bundle of equivalence classes of P_n^2 related n^2 coframes

$$P_n^2 \backslash \mathscr{H}^{2*}(M)$$

$$= \langle P_n^2 \backslash H^{2*}(M), \pi_{P_n^2 \backslash H^{2*}}, M, P_n^2 \backslash G_n^2, \mathscr{H}^{2*}(M) \rangle. \tag{4.2}$$

A projective structure on space time is a cross section $\Pi: M \to P_n^2 \backslash H^{2*}(M)$. Such a cross section may be represented locally by a family of cross sections $h: U \to H^{2*}(U)$ [see (3.4)], which are determined up to a P_n^2 gauge transformation $L: U \to U \times P_n^2$, where $L(p)$ is given by (4.1) and where the a_j^i and a_k depend smoothly on $p \in U$. The cross sections h satisfy

$$\forall\, p \in U,\ h(p) \in \Pi(p) \tag{4.3}$$

and any two such are related by a gauge transformation of the form

$$\begin{aligned} \tilde{h}(p) &= L(p) \cdot h(p) \\ &= \partial_0^2 I_n \left[(a_r^i h_j^r d_p^2 x^j \right. \\ &\quad + \frac{1}{2!}(a_r^i h_{jk}^r + a_r^i h_j^r a_m h_k^m + a_r^i h_k^r a_m h_j^m) d_p^2 x^j \\ &\quad \left. \times d_p^2 x^k \right]. \end{aligned} \tag{4.4}$$

The gauge fixing condition (3.7) may also be used in this case. Then the coefficient for the second-degree term in (4.4) becomes

$$h^{-1}{}_r^i h_{jk}^r + \delta_j^i a_m h_k^m + \delta_k^i a_m h_j^m. \tag{4.5}$$

Since a_m is still undetermined, the additional gauge-fixing condition

$$h^{-1}{}_r^i h_{ik}^r + (n+1) a_m h_k^m = 0 \tag{4.6}$$

may be imposed. The coefficient of the second-degree term is then

$$\Pi_{jk}^i = h^{-1}{}_r^i h_{jk}^r - \frac{1}{n+1}(\delta_j^i h^{-1}{}_r^m h_{mk}^r + \delta_k^i h^{-1}{}_r^m h_{mj}^r). \tag{4.7}$$

The Π_{jk}^i which satisfy $\Pi_{ij}^i = 0$ are called the *projective coefficients* and relative to a given coordinate system uniquely determine and are uniquely determined by the projective structure. Clearly, the process of choosing a particular gauge is just the selection of standard representatives for the cosets $P_n^2 \backslash G_n^2$. For a n^2 coframe $h(p) \in \Pi(p)$,

$$h(p) = \partial_0^2 I_n (a_j^i d_p^2 x^j + \frac{1}{2!}(a_r^i \Pi_{jk}^r + a_j^i a_k + a_k^i a_j) d_p^2 x^j d_p^2 x^k), \tag{4.8}$$

where $a_j^i = h_j^i$ and $a_k = \dfrac{1}{n+1} h^{-1}{}_r^m h_{mk}^r$.

The n^2 coframe (4.8) may be composed with the second-order curve element (3.10) to give an image of the curve element in $\mathbb{R}^n$.

$$\begin{aligned} h(p) \circ j_0^2 \gamma = \partial_0^2 I_n &(a_j^i \gamma_1^j d_0^2 I \\ &+ \frac{1}{2!}(a_r^i \gamma_2^r + a_r^i \Pi_{jk}^r \gamma_1^j \gamma_1^k + 2 a_j^i \gamma_1^j a_k \gamma_1^k)(d_0^2 I)^2). \end{aligned} \tag{4.9}$$

The condition that this image curve element is linear to second order is just

$$\gamma_2^i + \Pi_{jk}^i \gamma_1^j \gamma_1^k + 2 a_k \gamma_1^k \gamma_1^i = 0. \tag{4.10}$$

Since a_k is arbitrary at every p, the factor $a_k \gamma_1^k$ may be chosen freely in a smooth way along any integral curve. It is readily shown that this freedom allows an arbitrary choice of parameter for the integral curves; consequently, (4.10) describes a path element rather than a curve element.

The projective structure determines only a family of connections on the manifold. By projection and substitution, the coframe (4.8) determines the two coframes

$$\partial_0^1 I_n (a_j^i d_p^1 x^j)$$

and

$$\partial_0^1 I_n ((a_j^i + a_r^i \Pi_{js}^r w^s + a_s^i w^s a_j + a_j^i a_s w^s) d_q^1 x^j). \tag{4.11}$$

The map from $H^{1*}(M_p)$ to $H^{1*}(M_q)$ is then

$$\partial_p^1 x((\delta_j^i + \Pi_{js}^i w^s + w^i a_j + \delta_j^i a_s w^s) d_q^1 x^j) \tag{4.12}$$

which depends on the parameters a_i. Consider a 1^1 frame $j_0^1 \gamma \in H_1^1(M_p)$,

$$j_0^1 \gamma = \partial_p^1 x(\gamma_1^i d_0^1 I). \tag{4.13}$$

Under parallel transport by (4.12) for any a_i, the 1^1 frame becomes

$$\partial_q^1 x((\gamma_1^i - \Pi_{jk}^i w^k \gamma_1^j - w^i a_j \gamma_1^j - a_j w^j \gamma_1^i) d_0^1 I). \tag{4.14}$$

For autoparallel transport, $w^i = \lambda \gamma_1^i$ and (4.14) becomes

$$\partial_q^1 x((\gamma_1^i - \lambda \Pi_{jk}^i \gamma_1^j \gamma_1^k - 2\lambda a_j \gamma_1^j \gamma_1^i) d_0^1 I). \tag{4.15}$$

It is clear that the autoparallel transport of projective 1^1 frames (directions) is uniquely defined by the projective structure because the term (4.15) which depends on a_i is proportional to γ_1^i and consequently may be removed by a parameter, G_1^1, transformation.

The above discussion may be readily extended to 2^1 frames and projective 2^1 frames (a set of all 2^1 frames for a given two-dimensional subspace of the tangent space). Let $j_0^1 f \in H_2^1(M_p)$, where $f: \mathbb{R}^2 \to M$ such that $f(\mathbf{0}) = p$. Then

$$j_0^1 f = \partial_p^1 x(f_\alpha^i d_0^1 I_2^\alpha). \tag{4.16}$$

A projective 2^1 frame is an equivalence class of 2^1-frames under the action of the group G_2^2 of parameter transformations. Now, choose any vector which belongs to the subspace of $L_1^1(M_p)$ corresponding to a given projective 2^1 frame, say $w^i = \lambda_1 f_1^i + \lambda_2 f_2^i$. Next, parallel transport (4.16) along w^i using (4.12). The result is

$$\begin{aligned} \partial_q^1 x [\{ &f_\alpha^i - \Pi_{jk}^i f_\alpha^i (\lambda_1 f_1^k + \lambda_2 f_2^k) \\ &- (\lambda_1 f_1^i + \lambda_2 f_2^i) a_j f_\alpha^j + a_j (\lambda_1 f_1^j + \lambda_2 f_2^j) f^i \} d_0^1 I_2^\alpha]. \end{aligned} \tag{4.17}$$

Again, the terms which depend on the a_i may be removed by a G_2^2 parameter transformation; consequently, a projective structure determines the parallel transport of projective 2^1 frames. Also, the autoparallel transport of the initial direction lies in the subspace determined by the new projective 2^1 frame; therefore, the procedure may be iterated to define a parallel field of tangential projective 2^1 frames along a geodesic. This procedure corresponds to the *strip-forming* construction discussed by Ehlers and Schild.[18] It would probably be useful to reformulate the rest of their results in the language of jets.

The problem of determining the prolongations of the projective structure can be solved by the same procedure used in the affine case; however, the algebra is slightly more complicated. Consider an extension of the projective n^2 co-

frame (4.8),

$$\partial_0^3 I_n\Big[a_j^i d_p^3 x^j + \frac{1}{2!}(a_s^i \Pi_{jk}^s + a_j^i a_k + a_k^i a_j) d_p^3 x^j d_p^3 x^k + \frac{1}{3!} a_s^i \Gamma_{jk\prime}^s d_p^3 x^j d_p^3 x^k d_p^3 x^\prime\Big]. \tag{4.18}$$

Such a frame defines an equivalence class of n^2 coframes at q near p containing

$$\partial_0^2 I_n\Big[(a_j^i + a_s^i \Pi_{jr}^s w^r + a_r^i a_j w^r + a_j^i a_r w^r) d_q^2 x^j + \frac{1}{2!}(a_s^i \Pi_{jk}^s + a_j^i a_k + a_k^i a_j + a_s^i \Gamma_{jk\prime}^s w^\prime) d_q^2 x^j d_q^2 x^k\Big]. \tag{4.19}$$

This n^2 coframe should define the same P_n^2 equivalence class at q as the n^2 coframe

$$\partial_0^2 I_n\Big[a_j^i d_q^2 x^j + \frac{1}{2!}(a_s^i \Pi_{jk}^s + a_s^i \Pi_{jk,r}^s w^r + a_j^i a_k + a_k^i a_j) d_q^2 x^j d_q^2 x^k\Big]. \tag{4.20}$$

The standard representative corresponding to the n^2 coframes given by (4.20) is

$$\partial_0^2 I_n (d_q^2 x^i + \frac{1}{2!}(\Pi_{jk}^i + \Pi_{jk,\prime}^i w^\prime) d_q^2 x^j d_q^2 x^k). \tag{4.21}$$

The reduction of (4.19) to standard form is somewhat more complicated. If the first-order coefficient is reduced to δ_j^i, the second-order coefficient becomes

$$\Pi_{jk}^i + \Gamma_{jk\prime}^i w^\prime - \Pi_{r\prime}^i \Pi_{jk}^r w^\prime - w^i a_r \Pi_{jk}^r - a_\prime w^\prime \Pi_{jk}^i - a_k \Pi_{j\prime}^i w^\prime - a_j \Pi_{k\prime}^i w^\prime - 2 w^i a_j a_k. \tag{4.22}$$

The next step is to remove the trace from the expression (4.22). The resulting expression is equated with the coefficient of the second-order part of (4.21) allowing for a term $B_{jk\prime}^i$ with the same symmetries as in the affine case and also traceless, $B_{rk\prime}^r = 0$. The compatibility condition so obtained is

$$\begin{aligned}\Gamma_{jk\prime}^i - \frac{1}{n+1}(\delta_j^i \Gamma_{rk\prime}^r + \delta_k^i \Gamma_{rj\prime}^r) &= B_{jk\prime}^i + \Pi_{jk,\prime}^i + \Pi_{r\prime}^i \Pi_{jk}^r \\ &\quad - \frac{1}{n+1}(\delta_j^i \Pi_{sk}^r \Pi_{r\prime}^s + \delta_k^i \Pi_{sj}^r \Pi_{r\prime}^s) \\ &\quad + \delta_\prime^i a_r \Pi_{jk}^r - \frac{2}{n+1}(\delta_j^i a_r \Pi_{k\prime}^r + \delta_k^i a_r \Pi_{j\prime}^r) \\ &\quad + a_j \Pi_{k\prime}^i + a_k \Pi_{\prime j}^i + a_\prime \Pi_{jk}^i \\ &\quad + 2\delta_\prime^i a_j a_k - \frac{2}{n+1}(\delta_j^i a_\prime a_k + \delta_k^i a_\prime a_j).\end{aligned} \tag{4.23}$$

Take the totally symmetric part of this equation. The term $B_{jk\prime}^i$ drops out and one obtains

$$\begin{aligned}\Gamma_{jk\prime}^i - \frac{2}{3(n+1)}(\delta_j^i \Gamma_{rk\prime}^r + \delta_k^i \Gamma_{r\prime j}^r + \delta_\prime^i \Gamma_{rjk}^r) &= \frac{1}{3}(\Pi_{jk,\prime}^i + \Pi_{k\prime,j}^i + \Pi_{\prime j,k}^i + \Pi_{rj}^i \Pi_{k\prime}^r \\ &\quad + \Pi_{rk}^i \Pi_{\prime j}^r + \Pi_{r\prime}^i \Pi_{jk}^r) \\ &\quad - \frac{2}{3(n+1)}(\delta_j^i \Pi_{sk}^r \Pi_{r\prime}^s + \delta_k^i \Pi_{s\prime}^r \Pi_{rj}^s + \delta_\prime^i \Pi_{sj}^r \Pi_{rk}^s) \\ &\quad + \frac{n-3}{2(n+1)}(\delta_j^i a_r \Pi_{k\prime}^r + \delta_k^i a_r \Pi_{\prime j}^r + \delta_\prime^i a_r \Pi_{jk}^r) \\ &\quad + a_j \Pi_{k\prime}^i + a_k \Pi_{\prime j}^i + a_\prime \Pi_{jk}^i \\ &\quad + \frac{2(n-1)}{3(n+1)}(\delta_j^i a_k a_\prime + \delta_k^i a_\prime a_j + \delta_\prime^i a_j a_k).\end{aligned} \tag{4.24}$$

The trace of this equation gives

$$\begin{aligned}\frac{1(n-1)}{3(n+1)} \Gamma_{rk\prime}^r &= \frac{1}{3} \Pi_{k\prime,r}^r - \frac{2}{3} \frac{1}{(n+1)} \Pi_{sk}^r \Pi_{r\prime}^s \\ &\quad + \frac{(n-1)(n+3)}{3(n+1)} a_r \Pi_{k\prime}^r \\ &\quad + \frac{2}{3} \frac{(n-1)(n+2)}{(n+1)} a_k a_\prime.\end{aligned} \tag{4.25}$$

Then (4.24) gives the final result

$$\begin{aligned}\Gamma_{jk\prime}^i &= \frac{1}{3}(\Pi_{jk,\prime}^i + \Pi_{k\prime,j}^i + \Pi_{\prime j,k}^i) \\ &\quad + \frac{2}{3(n-1)}(\delta_j^i \Pi_{k\prime,r}^r + \delta_k^i \Pi_{\prime j,r}^r + \delta_\prime^i \Pi_{jk,r}^r) \\ &\quad + \frac{1}{3}(\Pi_{rj}^i \Pi_{k\prime}^r + \Pi_{rk}^i \Pi_{\prime j}^r + \Pi_{r\prime}^i \Pi_{jk}^r) \\ &\quad - \frac{2}{3(n-1)}(\delta_j^i \Pi_{sk}^r \Pi_{r\prime}^s + \delta_k^i \Pi_{s\prime}^r \Pi_{rj}^s + \delta_\prime^i \Pi_{sj}^r \Pi_{rk}^s) \\ &\quad + a_j \Pi_{k\prime}^i + a_k \Pi_{\prime j}^i + a_\prime \Pi_{jk}^i \\ &\quad + \delta_j^i a_r \Pi_{k\prime}^r + \delta_k^i a_r \Pi_{\prime j}^r + \delta_\prime^i a_r \Pi_{jk}^r \\ &\quad + 2(\delta_j^i a_k a_\prime + \delta_k^i a_\prime a_j + \delta_\prime^i a_j a_k).\end{aligned} \tag{4.26}$$

Define $\Pi_{jk\prime}^i$ to be the expression given by (4.26) for $a_j = 0$. Then, the general n^3-coframe belonging to the P_n^3 structure may be expressed as the composition of

$$\partial_0^3 I_n\Big[a_j^i d_0^3 I_n^j + \frac{1}{2!}(a_j^i a_k + a_k^i a_j) d_0^3 I_n^j d_0^3 I_n^k + \frac{2}{3!}(a_j^i a_k a_\prime + a_k^i a_\prime a_j + a_\prime^i a_j a_k) d_0^3 I_n^j d_0^3 I_n^k d_0^3 I_n^\prime\Big] \tag{4.27}$$

and

$$\partial_0^3 I_n (d_p^3 x^i + \frac{1}{2!} \Pi_{jk}^i d_p^3 x^j d_p^3 x^k + \frac{1}{3!} \Pi_{jk\prime}^i d_p^3 x^j d_p^3 x^k d_p^3 x^\prime). \tag{4.28}$$

The 3 jet (4.27) is the typical element of the prolonged group P_n^3.

5. RIEMANN AND LORENTZ STRUCTURES

The local Lorentz gauge invariant vierbein formulation of Einstein's general relativity theory is well known. It was first developed by Weyl[19] in 1929. Much later in 1950, he discussed the coupling of the vierbein fields to the spin density of matter fields and showed that an equivalent metric theory could be formulated.[20] Thereafter, the theory was analyzed further by Utiyama,[21] Sciama[22] and Kibble.[23] The natural setting for this formulation of general relativity is the theory of an $O_{1,n-1}^1$ structure on spacetime. In the following presentation, the bundles and formulas needed will be given for the case of a Lorentz structure; however, the corresponding bundles and formulas for the case of a Riemann structure may be obtained simply by substituting the group O_n^1 for the group $O_{1,n-1}^1$. Both the first-order $O_{1,n-1}^1$ structure and its prolongation to the second-order $O_{1,n-1}^2$ structure are discussed.

The restriction of the action of G_n^1 on $H^{1*}(M)$ to the

subgroup $O^1_{1,n-1}$ may be used to construct the bundle of equivalence classes of $O^1_{1,n-1}$ related n^1 coframes

$$O^1_{1,n-1}\backslash\mathscr{H}^{1*}(M) = \langle O^1_{1,n-1}\backslash H^{1*}(M), \pi_{O^1_{1,n-1}\backslash H^{1*}}, M, O^1_{1,n-1}\backslash G_n, \mathscr{H}^{1*}(M)\rangle. \quad (5.1)$$

An $O^1_{1,n-1}$ structure on M is a (global) cross section of $O^1_{1,n-1}\backslash\mathscr{H}^{1*}(M)$. Such a cross section may be locally represented by a family of local cross sections of $\mathscr{H}^{1*}(M)$. Any two such local cross sections $h: U\to H^{1*}(U)$ and $\tilde{h}: U\to H^{1*}(U)$ are related by a local $O^1_{1,n-1}$ gauge transformation defined by a cross section $\Lambda: U\to U\times O^1_{1,n-1}$

$$h(p) = \partial^1_0 I_n[h^i_j(p) d^1_p x^j],$$
$$\tilde{h}(p) = \Lambda(p)\circ h(p) = \partial^1_0 I_n[\Lambda^i_r(p) h^r_j(p) d^1_p x^j], \quad (5.2)$$

where

$$\eta_{ij}\Lambda^i_{i'}(p)\Lambda^j_{m}(p) = \eta_{i'm}. \quad (5.3)$$

For both the Lorentz and Riemann cases, a suitable gauge-fixing condition is

$$h^i_j(p) = 0;\ i>j. \quad (5.4)$$

For such a choice of gauge, the matrix $[h^i_j(p)]$ is upper triangular.

The local cross sections h are usually called *n-bein* fields. The Lorentz metric tensor defined by the $O^1_{1,n-1}$ structure is locally given by

$$g = \eta_{ij} h^i_k h^j_{i'} dx^k \otimes dx^{i'}. \quad (5.5)$$

Since these local metric tensor fields are $O^1_{1,n-1}$ gauge independent, they can be patched together to define a global metric tensor field.

Now, consider the problem of prolonging an $O^1_{1,n-1}$ structure to an $O^2_{1,n-1}$ structure. At $p\in M$, a general element of the equivalent class of frames defined by the $O^1_{1,n-1}$ structure is

$$\partial^1_0 I_n(\Lambda^i_r h^r_j d^1_p x^j). \quad (5.6)$$

A general n^2 coframe having the same first-order part and belonging to some $O^2_{1,n-1}$ structure is

$$\partial^2_0 I_n(\Lambda^i_r h^r_j d^2_p x^j + \frac{1}{2!}\Lambda^i_r h^r_s \Gamma^s_{jk} d^2_p x^j d^2_p x^k). \quad (5.7)$$

Note that factoring out h^r_s in the second-order part is a mere convenience. If $q\in M$ is near $p\in M$ and $w^i = x^i(q) - x^i(p)$, then (5.7) defines at q an equivalence class of n^1 coframes containing the element

$$\partial^1_0 I_n((\Lambda^i_r h^r_j + \Lambda^i_r h^r_s \Gamma^s_{j'} w^{j'}) d^1_q x^j). \quad (5.8)$$

However, the $O^1_{1,n-1}$ structure or Lorentz structure already determines at q an equivalence class of n^1 coframes which contains the standard element [some gauge fixing condition such as (5.4) is assumed]

$$\partial^1_0 I_n((h^i_j + h^i_{j,i'} w^{i'}) d^1_q x^j). \quad (5.9)$$

The condition that the $O^2_{1,n-1}$ structure is compatible with the $O^1_{1,n-1}$ structure is that the two equivalence classes of $O^1_{1,n-1}$ related n^1 coframes determined by (5.8) and (5.9) are the same. In order to make the desired comparison, it is necessary to transform the n^1 coframe to standard form. Now, the matrix

$$h^i_j + h^i_r \Gamma^r_{j'} w^{i'} \quad (5.10)$$

differs only infinitesimally from standard form. Hence, for some infinitesimal Lorentz transformation, say

$$\Lambda^i_j = \delta^i_j + \lambda^i_{j'} w^{i'},$$
$$\eta_{ir}\lambda^r_{j'} = \lambda_{ij'} = -\lambda_{ji'} \quad (5.11)$$

the matrix

$$\Lambda^i_s(h^s_j + h^s_r \Gamma^r_{j'} w^{i'}) = h^i_j + h^i_r \Gamma^r_{j'} w^{i'} + h^r_j \lambda^i_{r'} w^{i'} \quad (5.12)$$

is in standard form. If the particular gauge-fixing condition (5.4) is used, then

$$h^i_r \Gamma^r_{j'} + h^r_j \lambda^i_{r'} = 0;\ i>j. \quad (5.13)$$

Since $h^r_j = 0$ unless $j\geq r$, Eqs. (5.13) contain $\lambda^i_{r'}$ for $i>r$ and may be solved successively for $\lambda^i_{j'}(i>j)$ in terms of h^i_j and $\Gamma^i_{j'}$. The values for $i<j$ follow from the skew symmetry of $\lambda_{ij'}$ in i and j. Although Eq. (5.13) are easy to solve, it is somewhat tedious to write the solution out explicitly. Fortunately, in this case, it is not necessary to do so; nevertheless, the expression (5.12) should be regarded as a function of h^i_j, $\Gamma^i_{j'}$, and $w^{i'}$. If this expression is equated with the matrix of the standard form (5.9), one obtains the condition

$$h^i_{j,i'} = h^i_r \Gamma^r_{j'} + h^r_j \lambda^i_{r'}. \quad (5.14)$$

Since $\Gamma^i_{j'} = \Gamma^i_{ji'}$, the skew part of (5.14) gives

$$h^i_{j,i'} - h^i_{i',j} = h^r_j \lambda^i_{r'} - h^r_{i'} \lambda^i_{rj}. \quad (5.15)$$

Define

$$\hat{\lambda}_{ijk} = \eta_{is}\lambda^s_{jr} h^{-1r}_k. \quad (5.16)$$

The (5.15) may be rewritten as

$$\eta_{ii'}(h^{i'}_{r,s} - h^{i'}_{s,r}) h^{-1r}_j h^{-1s}_k = \hat{\lambda}_{ijk} - \hat{\lambda}_{ikj}. \quad (5.17)$$

Cyclic permutation of the indices and use of the skew symmetry (5.11) yields the solution given by De Witt[24]; namely,

$$2\hat{\lambda}_{ijk} = \eta_{ii'}(h^{i'}_{r,s} - h^{i'}_{s,r}) h^{-1r}_j h^{-1s}_k + \eta_{ji'}(h^{i'}_{r,s} - h^{i'}_{s,r}) h^{-1r}_k h^{-1s}_i - \eta_{ki'}(h^{i'}_{r,s} - h^{i'}_{s,r}) h^{-1r}_i h^{-1s}_j. \quad (5.18)$$

This result may then be used in (5.14) to give the usual result for Γ^i_{jk} which may be written as

$$\Gamma^i_{jk} = h^{-1i}_r h^r_{j,k} - \eta^{rs} h^{-1i}_r \hat{\lambda}_{sab} h^a_j h^b_k. \quad (5.19)$$

Although the standard gauge defined by (5.4) was assumed in order to make the discussion more concrete, the result (5.18) is gauge-independent provided that the quantities $\lambda^i_{j'}$ transfrom in the manner appropriate to an O(1,3) gauge connection under a local gauge transformation.

It is useful to discuss an alternate approach to the Lorentz structure of spacetime. Consider the bundle of 1^2 cospeeds

$$\mathscr{L}^2_1{}^*(M) = \langle L^2_1{}^*(M), \pi_{L^2_1{}^*}, M, L^2_{n,1}, \mathscr{H}^{2*}(M)\rangle. \quad (5.20)$$

If $f: M\to\mathbb{R}$ such that f(p) = 0, then $j^2_p f\in L^2_1{}^*(M_p)$ and

$$j^2_p f = \partial^2_0 I(f_i d^2_p x^i + \frac{1}{2!} f_{ij} d^2_p x^i d^2_p x^j). \quad (5.21)$$

If only those functions f are considered which are singular but Morse with signature 2 − n at p, then a subbundle $\hat{\mathscr{L}}_1^2{}^*(M)$ of $\mathscr{L}_1^2{}^*(M)$ is obtained consisting of elements of the form

$$j_p^2 g = \partial_0^2 I\left(\frac{1}{2!} g_{ij} d_p^2 x^i d_p^2 x^j\right), \tag{5.22}$$

where the matrix (g_{ij}) assumes in some coordinate system the standard Minkowski form. Clearly, a Lorentz structure may also be defined on M by a global cross section of the $\hat{\mathscr{L}}^2{}^*(M)$ as is evident from a comparison of (5.5) and (5.22).

The principal bundle associated with $\hat{\mathscr{L}}^2{}^*(M)$ is $\mathscr{H}^1{}^*(M)$ rather than $\mathscr{H}^2{}^*(M)$; consequently, if (5.22) is parallel transported using (3.23), only the first-order part will have any effect. The element (5.22) is mapped into

$$\partial_0^2 I\left[\frac{1}{2!}(g_{ij} + g_{rj}\Gamma^r_{i\prime} w^\prime + g_{ir}\Gamma^r_{j\prime} w^\prime) d_q^2 x^i d_q^2 x^j\right]. \tag{5.23}$$

If the Lorentz structure defined by the cross section of $\hat{\mathscr{L}}_1^2{}^*(M)$ is horizontal under the connection induced by the second-order structure, the element (5.23) must equal the element

$$\partial_0^2 I\left[\frac{1}{2!}(g_{ij} + g_{ij,\prime} w^\prime) d_q^2 x^i d_q^2 x^j\right]. \tag{5.24}$$

Hence

$$g_{ij,k} = g_{ir}\Gamma^r_{jk} + g_{rj}\Gamma^r_{ik}. \tag{5.25}$$

This relation may also be derived using (5.5) and (5.14). The well-known solution is

$$\Gamma^i_{jk} = \tfrac{1}{2} g^{ir}(g_{rj,k} + g_{rk,j} - g_{jk,r}), \tag{5.26}$$

where g^{ij} is the inverse of g_{ij}.

The compatibility conditions (5.14) and (5.25) are not the most general possible. Just as in the affine and projective cases, allowance should be made for a possible term B^i_{jk} with $B^i_{jk} = -B^i_{kj}$. Thus (5.14) becomes

$$h^i_{j,\prime} + h^i_r B^r_{j\prime} = h^i_r \Gamma^r_{j\prime} + h^r_j \lambda^i_{r\prime}, \tag{5.27}$$

and (5.25) becomes

$$g_{ij,k} + g_{ir}B^r_{jk} + g_{rj}B^r_{ik} = g_{ir}\Gamma^r_{jk} + g_{rj}\Gamma^r_{ik}. \tag{5.28}$$

This equation leads to the formula given by Schrödinger.[25]

$$\begin{aligned}\Gamma^i_{jk} = {} & \tfrac{1}{2} g^{ir}(g_{rj,k} + g_{rk,j} - g_{jk,r}) \\ & + g_{js}g^{ir}B^s_{rk} + g_{ks}g^{ir}B^s_{rj}.\end{aligned} \tag{5.29}$$

Finally, since the affine connection induced by the $O^2_{1,n-1}$ structure is completely defined, further prolongations are determined by the formulas that apply in the affine case. Also under infinitesimal parallel transport, the connection induces infinitesimal $O^1_{1,n-1}$ transformations.

6. THE CONFORMAL STRUCTURE

The analysis of a first-order conformal or $C^1_{1,n-1}$ structure on M and of its prolongation to a second-order conformal or $C^2_{1,n-1}$ structure on M is similar to the corresponding analysis of the Lorentz structure. The further prolongation of the $C^2_{1,n-1}$ structure to a $C^3_{1,n-1}$ structure may be derived by an analysis entirely analogous to that given for the corresponding prolongation in the affine and projective cases; however, the algebra required is rather tedious although in principle straight forward. Thus, the explicit solution for the $C^3_{1,n-1}$ structure will not be given although the problem will be set up and a solution shown to exist.

A first-order conformal or $C^1_{1,n-1}$ structure on M is a reduction of the structure group G^1_n of the bundle of n^1 coframes $\mathscr{H}^1{}^*(M)$ to the subgroup $C^1_{1,n-1}$; that is, a $C^1_{1,n-1}$ structure is defined by a cross section of the bundle of equivalence classes of $C^1_{1,n-1}$ related n^1 coframes

$$\begin{aligned}C^1_{1,n-1}\backslash\mathscr{H}^1{}^*(M) = {} & \langle C^1_{1,n-1}\backslash H^1{}^*(M), \pi_{C^1_{1,n-1}\backslash H^1{}^*}, M, \\ & C^1_{1,n-1}\backslash G^1_n, \mathscr{H}^1{}^*(M)\rangle.\end{aligned} \tag{6.1}$$

Such a cross section is locally represented by a family of local cross sections of $\mathscr{H}^1{}^*(M)$. Any two local cross sections $h: U \to H^1{}^*(U)$ and $\tilde{h}: U \to H^1{}^*(U)$ are related by a local $C^1_{1,n-1}$ gauge transformation defined by a cross section $c: U \to U \times C^1_{1,n-1}$. Thus

$$\begin{aligned}h(p) &= \partial_0^1 I_n(h^i_j(p) d_p^1 x^j), \\ \tilde{h}(p) &= c(p) \circ h(p) = \partial_0^1 I_n(c^i_r(p) h^r_j(p) d_p^1 x^j),\end{aligned} \tag{6.2}$$

where

$$\begin{aligned}& c^i_j(p) = e^{\lambda(p)} \Lambda^i_j(p), \\ & \eta_{ij}\Lambda^i_k(p)\Lambda^j_\prime(p) = \eta_{k\prime}.\end{aligned} \tag{6.3}$$

As for the Riemann and Lorentz cases, many gauge-fixing conditions are possible. One possible choice is

$$\begin{aligned}& h^i_j = 0;\ i > j \\ & h^0_0 = 1.\end{aligned} \tag{6.4}$$

The metric tensor fields g and $\tilde{g}$ defined by h and $\tilde{h}$ by (5.5) are related by

$$\tilde{g}(p) = e^{2\lambda(p)} g(p). \tag{6.5}$$

To prolong the $C^1_{1,n-1}$ structure to a $C^2_{1,n-1}$ structure consider a general n^2 coframe belonging to a $C^1_{1,n-1}$ equivalence class of n^2 coframes and having a first-order part belonging to the $C^1_{1,n-1}$ structure; namely

$$\partial_0^2 I_n(c^i_r h^r_j d_p^2 x^j + \frac{1}{2!} c^i_r h^r_s \Gamma^s_{jk} d_p^2 x^j d_p^2 x^k). \tag{6.6}$$

This n^2-coframe at p defines a n^1 coframe at a neighboring point q; namely,

$$\partial_0^1 I_n((c^i_r h^r_j + c^i_r h^r_s \Gamma^s_{j\prime} w^\prime) d_q^1 x^j), \tag{6.7}$$

where $w^i = x^i(q) - x^i(p)$. However, the $C^1_{1,n-1}$ structure already determines at q an equivalence class of n^1 coframes at q determined by the standard n^1-coframe

$$\partial_0^1 I_n((h^i_j + h^i_{j,\prime} w^\prime) d_q^1 x^j). \tag{6.8}$$

For compatibility, the equivalence classes of n^1 coframes defined by (6.7) and (6.8) must be the same. To reduce the n^1 coframe to standard form, one must determine the infinitesimal conformal transformation

$$\begin{aligned}& c^i_j = \delta^i_j + \lambda_\prime w^\prime \delta^i_j + \lambda^i_{j\prime} w^\prime, \\ & \eta_{ir}\lambda^r_{j\prime} = \lambda_{ij\prime} = -\lambda_{ji\prime},\end{aligned} \tag{6.9}$$

such that the matrix

$$c^i_r(h^r_j + h^r_s \Gamma^s_{j\prime} w^\prime) = h^i_j + h^i_r \Gamma^r_{j\prime} w^\prime + h^r_j \lambda^i_{r\prime} w^\prime + \lambda_\prime w^\prime \delta^i_j \tag{6.10}$$

satisfies the conditions (6.4); that is,

$$\begin{aligned} &h^i_r \Gamma^r_{j\prime} + h^r_j \lambda^i_{r\prime} = 0;\ i > j \\ &h^0_r \Gamma^r_{0\prime} + \lambda_\prime = 0. \end{aligned} \tag{6.11}$$

Note that $h^r_0 \lambda^0_{r\prime} = 0$ since $h^r_0 = 0$ unless $r = 0$ and $\lambda_{ij\prime}$ is skew in i and j. The discussion given for the Lorentz case applies equally to Eqs. (6.11). The solutions for $\lambda^i_{j\prime}$ and $\lambda_\prime$ are substituted into (6.10) which is then in standard form. Equating the matrix of (6.8) with the matrix (6.10) gives the compatibility condition

$$h^i_{j,\prime} = h^i_r \Gamma^r_{j\prime} + h^r_j \lambda^i_{r\prime} + \lambda_\prime h^i_j. \tag{6.12}$$

As for the Lorentz case, the possibility of a torsion term is for simplicity not considered.

The algebraic difficulties are considerably reduced if the problem is reformulated in terms of metric tensor fields. For this purpose, it is necessary to introduce the bundles $G^2_1 \backslash \mathscr{L}^2_1{}^*(M)$ and $G^2_1 \backslash \hat{\mathscr{L}}^2_1{}^*(M)$ of equivalence classes of elements of $\mathscr{L}^2_1{}^*(M)$ and $\hat{\mathscr{L}}^2_1{}^*(M)$ under the action of the parameter group G^2_1. If $\mu: \mathbb{R} \to \mathbb{R}$ is a diffeomorphism with $\mu(0) = 0$, then $j^2_0 \mu \in G^2_1$ and

$$j^2_0 \mu = \partial^2_0 I (\mu_1 d^2_0 I + \frac{1}{2!} \mu_2 (d^2_0 I)^2), \tag{6.13}$$

where μ_1 and μ_2 are the first and second derivatives of μ at $0 \in \mathbb{R}$. Also, if $j^2_p f \in L^2_1{}^*(M_p)$ is given by (5.21) and $\tilde{f} = \mu \circ f$, the action of G^2_1 is given by

$$j^2_p \tilde{f} = j^2_0 \mu \circ j^2_p f. \tag{6.14}$$

In component form, this action is given by

$$\begin{aligned} &\tilde{f}_i = \mu_1 f_i, \\ &\tilde{f}_{ij} = \mu_1 f_{ij} + \mu_2 f_i f_j. \end{aligned} \tag{6.15}$$

For elements of the subbundle $\hat{\mathscr{L}}^2_1{}^*(M)$ given by (5.22), this action simplifies to

$$\tilde{g}_{ij} = \mu_1 g_{ij}. \tag{6.16}$$

A $C^1_{1,n-1}$ structure on M may also be defined by a (global) cross section of $G^2_1 \backslash \hat{\mathscr{L}}^2_1{}^*(M)$ which may be represented by local cross sections of $\mathscr{L}^2_1{}^*(M)$, any two of which are related by a gauge transformation (6.16). Numerous gauge-fixing conditions are possible. A simple choice [which is not consistent with (6.4)] is

$$g_{00} = 1. \tag{6.17}$$

Again, under the parallel transport induced by the prolonged structure, the standard representative (5.22) is transformed into (5.23). For some infinitesimal gauge transformation ($\mu_1 = 2\lambda_\prime w^\prime$) the element (5.23) is transformed into standard form

$$g_{ij} + g_{rj} \Gamma^r_{i\prime} w^\prime + g_{ir} \Gamma^r_{j\prime} w^\prime + 2 g_{ij} \lambda_\prime w^\prime. \tag{6.18}$$

If the gauge-fixing condition (6.17) is used, then

$$\lambda_\prime = -g_{0r} \Gamma^r_{0\prime}. \tag{6.19}$$

Define the conformal connection coefficients K^i_{jk} by

$$\Gamma^i_{jk} = K^i_{jk} + \frac{1}{n}(\delta^i_j \Gamma_k + \delta^i_k \Gamma_j - g_{jk} g^{ir} \Gamma_r), \tag{6.20}$$

$$K^i_{ik} = 0, \quad \Gamma_k = \Gamma^i_{ik}.$$

Substitute (6.20) into (6.18). The terms involving Γ_k cancel. Thus the matrix of the standard metric representative at q is

$$g_{ij} + g_{rj} K^r_{i\prime} w^\prime + g_{ir} K^r_{j\prime} w^\prime - 2 g_{ij} g_{0r} K^r_{0\prime} w^\prime, \tag{6.21}$$

which must equal the matrix of the standard element (5.24); consequently,

$$g_{ij,k} = g_{rj} K^r_{ik} + g_{ir} K^r_{jk} - 2 g_{ij} g_{0r} K^r_{0k}. \tag{6.22}$$

Then

$$g^{ij} g_{ij,k} = -2n g_{0r} K^r_{0k} \tag{6.23}$$

and

$$g_{ir} K^r_{jk} + g_{rj} K^r_{ik} = g_{ij,k} - \frac{1}{n} g_{ij} g^{rs} g_{rs,k}. \tag{6.24}$$

This equation may be solved to give

$$\begin{aligned} K^i_{jk} = {} & \tfrac{1}{2} g^{ir} (g_{rj,k} + g_{rk,j} - g_{jk,r}) \\ & - \frac{1}{2n} (\delta^i_j g^{rs} g_{rs,k} + \delta^i_k g^{rs} g_{rs,j} - g_{jk} g^{it} g^{rs} g_{rs,t}). \end{aligned} \tag{6.25}$$

Although a particular gauge was used for the computation, the result (6.25) is in fact gauge independent. Note that the trace Γ_k of Γ^i_{jk} is completely undetermined. If these results are now substituted into (6.6), it is readily shown that the typical n^2 coframe belonging to the prolonged structure may be expressed as the composition of

$$\partial^2_0 I_n \left[c^i_j d^2_0 I^j_n + \frac{1}{2!} (c^i_j c_k + c^i_k c_j - \eta_{jk} \eta^{ir} c_r) d^2_0 I^j_n d^2_0 I^k_n \right] \tag{6.26}$$

and

$$\partial^2_0 I_n (h^i_j d^2_p x^j + \frac{1}{2!} h^i_r K^r_{jk} d^2_p x^j d^2_p x^k), \tag{6.27}$$

where

$$\Gamma_k = c_r h^r_k. \tag{6.28}$$

The 2 jet (6.26) is a general element of the prolonged group $C^2_{1,n-1}$. The general n^2 coframe belonging to the $C^2_{1,n-1}$ structure is

$$\begin{aligned} \partial^2_0 I_n \Big[& a^i_j d^2_p x^j + \frac{1}{2!} a^i_r (K^r_{jk} + \delta^r_j a_k + \delta^r_k a_j \\ & - g_{jk} g^{rs} a_s) d^2_p x^j d^2_p x^k \Big], \end{aligned} \tag{6.29}$$

where

$$a^i_j = c^i_r h^r_j \qquad a_k = c_r h^r_k. \tag{6.30}$$

The prolongation of the $C^2_{1,n-1}$ structure to a $C^3_{1,n-1}$ structure follows the same pattern. Consider the most general n^3 coframe which projects to the n^2 coframe (6.29).

$$\begin{aligned} \partial^3_0 I_n \Big[& a^i_j d^3_p x^j + \frac{1}{2!} a^i_r (K^r_{jk} + \delta^r_j a_k \\ & + \delta^r_k a_j - g_{jk} g^{rs} a_s) d^3_p x^j d^3_p x^k \\ & + \frac{1}{3!} a^i_r \Gamma^r_{jk\prime} d^3_p x^j d^3_p x^k d^3_p x^\prime \Big]. \end{aligned} \tag{6.31}$$

The n^2-coframe at q determined by (6.31) by the substitution process is

$$\partial_0^2 I_n \left[a_r^i (\delta_j^r + K_{js}^r w^s + \delta_j^r a_s w^s + w^r a_j - g^{rt} a_t g_{js} w^s) d_q^2 x^j + \frac{1}{2!} a_r^i (K_{jk}^r + \delta_j^r a_k + \delta_k^r a_j - g_{jk} g^{rt} a_t + \Gamma_{jks}^r w^s) d_q^2 x^j d_q^2 x^k \right]. \tag{6.32}$$

The next task is to bring the first-order part of (6.32) to a form satisfying the standard gauge conditions (6.4). The problem is essentially the same as that discussed above following (6.9) in connection with the prolongation from a $C_{1,n-1}^1$ structure to a $C_{1,n-1}^2$ structure; that is, the form of the linear equations for the parameters of the infinitesimal $C_{1,n-1}^1$ transformation is the same. However, the number of terms involved is even greater than it was before. Moreover, it does not seem possible to artfully dodge the problem as before because it is necessary to know the additional terms that will appear in the second-order part of (6.32) when the first-order part is brought to standard form. Once the first-order part is in standard form, it is easy to bring the second-order part to standard form by applying (6.26) with $c_j^i = \delta_j^i$ and c_j chosen so that the coefficient of h_r^i in the second-order term will have zero trace.

The standard n^2 coframe at q obtained by the above procedure must agree with the appropriate n^2 coframe at q determined by the $C_{1,n-1}^2$ structure. The n^2 coframe corresponding to (6.27) evaluated at q is

$$\partial_0^2 I_n \left[(h_j^i + h_{j,s}^i w^s) d_q^2 x^j + \frac{1}{2!} h_r^i (K_{jk}^r + K_{jk,s}^r w^s + h^{-1}{}_a^r h_{b,s}^a K_{jk}^b w^s) d_q^2 x^j d_q^2 x^k \right]. \tag{6.33}$$

However, the coefficient of $h_r^i = h_r^i(p)$ in (6.33) does not have zero trace. This trace must be removed before the comparison with the standard form of (6.32) described above is made.

The first-order terms must necessarily agree because the K_{jk}^i were defined by that condition at the previous stage. The demand that the second-order terms agree up to a term of the form $h_r^i B_{jks}^r w^s$, where B_{jks}^i is traceless and has the same symmetries as in the affine and projective cases, gives an equation for the Γ_{jkl}^i which may be solved in a manner similar to that used in the projective case. The solution gives $\Gamma_{jk\ell}^i$ in terms of h_j^i, K_{jk}^i, and a_j. The general n^3 coframe belonging to the $C_{1,n-1}^3$ structure may be obtained by substituting this solution into (6.31) which may in turn be factored into a $C_{1,n-1}^3$ gauge group element and the standard n^3 coframe for the $C_{1,n-1}^3$ structure.

7. THE WEYL STRUCTURE

A Weyl structure on M is a reduction of the structure group G_n^2 of $\mathscr{H}^{2*}(M)$ to the Weyl subgroup $W_{1,n-1}^2$ (see Appendix C). Since $W_{1,n-1}^2$ is a Lie subgroup of $C_{1,n-1}^2$ the reduction may be accomplished by first defining a $C_{1,n-1}^1$ structure and prolonging this structure to a $C_{1,n-1}^2$ structure and then reducing this structure to a $W_{1,n-1}^2$ structure. The general element of an equivalence class belonging to a $C_{1,n-1}^2$ structure is given by (6.6) together with (6.20) and (6.25). The result is

$$\partial_0^2 I_n \left[c_r^i h_j^r d_p^2 x^j + \frac{1}{2!} c_r^i h_s^r (K_{jk}^s + \frac{1}{n} (\delta_j^s \Gamma_k + \delta_k^s \Gamma_j - g_{jk} g^{st} \Gamma_t) d_p^2 x^j d_p^2 x^k \right], \tag{7.1}$$

where $\Gamma_k = c_r h_k^r$. Clearly, the reduction to $W_{1,n-1}^2$ is determined locally by specifying the Γ_k (or c_k) as functions of $p \in U \subset M$. The connection determined by the Γ_{jk}^i given by (6.20) is then fully specified; consequently, the formulas for the higher-order prolongations of an affine structure may be used to determine the higher-order prolongations of a Weyl structure. The Weyl connection induces infinitesimal conformal transformations under infinitesimal parallel transport.

8. THE CONFORMAL–PROJECTIVE COMPATIBILITY CONDITION

In this section, the algebraic condition that expresses the compatibility between a conformal or $C_{1,n-1}^2$ structure and a projective or P_n^2 structure is derived using the analysis of Secs. 4 and 6. The result was originally derived by Ehlers, *et al.*[1], from constructive axioms for general relativity theory. The derivation given here is not a replacement for their argument, but is rather a supplement which clarifies the geometric interpretation of the condition as the necessary and sufficient condition that a $C_{1,n-1}^2$ structure and a P_n^2 structure are compatible and uniquely define a Weyl or $W_{1,n-1}^2$ structure.[26]

A general element belonging to an equivalence class of P_n^2 related n^2 coframes defined by a projective structure is given by (4.8)

$$\partial_0^2 I_n \left[a_j^i d_p^2 x^j + \frac{1}{2!} a_r^i (\Pi_{jk}^r + \delta_j^i a_k) + \delta_k^i a_j \right] d_p^2 x^j d_p^2 x^k. \tag{8.1}$$

A general element belonging to an equivalence class of $C_{1,n-1}^2$ related n^2 coframes is given by (7.1)

$$\partial_0^2 I_n \left[c_r^i h_j^r d_p^2 x^j + \frac{1}{2!} c_r^i h_s^r (K_{jk}^s + \frac{1}{n} (\delta_j^s \Gamma_k + \delta_k^s \Gamma_j - g_{jk} g^{sa} \Gamma_a)) d_p^2 x^j d_p^2 x^k \right]. \tag{8.2}$$

The two structures are compatible iff the equivalent classes defined by (8.1) and (8.2) have a nonempty intersection. Consideration of the first-order parts shows that a projective coframe is a conformal coframe only if

$$a_j^i = c_r^i h_j^r. \tag{8.3}$$

Comparison of the second-order parts then yields the relation

$$\Pi_{jk}^i + \delta_j^i a_k + \delta_k^i a_j = K_{jk}^i + \frac{1}{n} (\delta_j^i \Gamma_k + \delta_k^i \Gamma_j - g_{jk} g^{ir} \Gamma_r). \tag{8.4}$$

Define

$$\Delta_{jk}^i = \Pi_{jk}^i - K_{jk}^i. \tag{8.5}$$

Then take the trace of (8.4) to obtain

$$(n+1)a_k = \Gamma_k. \tag{8.6}$$

Then

$$\Delta^i_{jk} = \frac{1}{n(n+1)}(\delta^i_j\Gamma_k + \delta^i_k\Gamma_j) - \frac{1}{n}g_{jk}\,g^{ir}\Gamma_r \tag{8.7}$$

and

$$g^{rs}\Delta^i_{rs} = -\frac{(n-1)(n+2)}{n(n+1)}g^{ir}\Gamma_r \tag{8.8}$$

or

$$\Gamma_k = -\frac{n(n+1)}{(n-1)(n+2)}g_{ki}g^{rs}\Delta^i_{rs}. \tag{8.9}$$

Substitution of (8.9) into (8.7) yields a system of linear equations for Δ^i_{jk} which are satisfied iff the projective and conformal structures are compatible; moreover, if they are compatible, Γ_k given by (8.9) and hence Γ^i_{jk} are well defined and the conformal and projective structures define a unique Weyl structure.

APPENDIX A: OF GERMS AND JETS

Let M and N be C^∞ differentiable manifolds. Denote by $C^\infty(M,N)$ the set of C^∞ differentiable maps $f:M\to N$. Any two maps $f_1, f_2\in C^\infty(M,N)$ are said to be germ equivalent at $p\in M$ iff there exists an open neighborhood U of p on which f_1 and f_2 agree. The *germ* j_pf of a map $f\in C^\infty(M,N)$ is the equivalence class to which f belongs. The set of such equivalence classes will be denoted by $J(M_p,N)$. In certain applications, maps defined only on some open submanifold U of M are considered. The set of germs of such maps is then denoted by $J(U_p,N)$. The map $j_p:C^\infty(M,N)\to J(M_p,N)$ is the canonical projection.

Any two maps $f_1, f_2\in C^\infty(M,N)$ are said to be k-jet equivalent iff they agree *at* p

$$f_1(p) = f_2(p) = q, \tag{A1}$$

and for any charts $(U,x)_p$ and $(V,y)_q$ where $p\in U$ and $q\in V$, the maps $y\circ f_1\circ x^{-1}$ and $y\circ f_2\circ x^{-1}$ have the same partial derivatives of all types at $x(p)$ up to and including order k. That this equivalence relation is independent of the choice of coordinate charts follows by applying the chain rule to the relation

$$\bar{y}\circ f\circ\bar{x}^{-1} = (\bar{y}\circ y^{-1})\circ(y\circ f\circ x^{-1})\circ(x\circ\bar{x}^{-1}). \tag{A2}$$

The k jet $j^k_p f$ of a map $f\in C^\infty(M,N)$ is the equivalence class to which f belongs. The set of k jets with *source* $p\in M$ and *target* $q\in N$ is denoted by $J^k(M_p,N_q)$, and the corresponding sets with unrestricted source, target, or both are denoted by $J^k(M,N_q)$, $J^k(M_p,N)$, and $J^k(M,N)$, respectively. Again, for some applications it may be convenient to consider the set of maps $C^\infty(U,V)$ where U and V are open submanifolds of M and N. Then the corresponding sets of k jets are denoted by $J^k(U_p,V_q)$ and so forth. The map $j^k_p:C^\infty(M,N)\to J^k(M_p,N)$ is the canonical projection. Since $j_pf\supset j^k_pf$ one may also use j^k_p to denote the projection $j^k_p: J(M_p,N)\to J^k(M_p,N)$ and write

$$j^k_p(j_pf) = j^k_pf. \tag{A3}$$

If M, L and N are C^∞ manifolds, then the composition map $\circ:C^\infty(L,N)\times C^\infty(M,L)\to C^\infty(M,N)$ defined by $g\circ f(p) = g(f(p))$ induces a number of compositions between germs and jets; namely,

$$\begin{aligned} j_p(g\circ f) &= j_{f(p)}g\circ j_pf \\ j^k_p(g\circ f) &= j^k_{f(p)}g\circ j_pf \\ &= j_{f(p)}g\circ j^k_pf \\ &= j^k_{f(p)}g\circ j^k_pf. \end{aligned} \tag{A4}$$

An algebraic structure on the manifold N will in general induce an algebraic structure on $J(M_p,N)$ and $J^k(M_p,N)$; for example, these sets inherit algebra structures if N is $\mathbb{R}$. Define $\forall j^k_pf, j^k_pf_1, j^k_pf_2\in J^k(M_p,\mathbb{R})$ and $\forall\lambda_1,\lambda_2\in\mathbb{R}$

$$\begin{aligned} &\lambda_1j^k_pf_1 + \lambda_2j^k_pf_2 = j^k_p(\lambda_1f_1 + \lambda_2f_2), \\ &j^k_pf_1\,j^k_pf_2 = j^k_p(f_1f_2), \end{aligned} \tag{A5}$$

where the functions $\lambda_1f_1 + \lambda_2f_2$ and f_1f_2 are defined by pointwise scalar multiplication, addition, and multiplication. The definitions of the algebra operations for $J(M_p,\mathbb{R})$ are similar. Also, if N is a vector space, then so are $J(M_p,N)$ and $J^k(M_pN)$, and if N is finite dimensional, then so is $J^k(M_p,N)$.

Any element $j^k_pf\in J^k(M_p,N_q)$ may be invariantly represented with respect to any coordinate charts $(U,x)_p$ and $(V,y)_q$. Denote by $x(p)$ the constant map on U which maps every point of U into $x(p)$, and set $x_p = x - x(p)$. Similarly, define the map $y_q = y - y(q)$. Then

$$\begin{aligned} j^k_pf &= j^k_p(y_q^{-1}\circ y_q\circ f\circ x_p^{-1}\circ x_p) \\ &= j^k_0\,y_q^{-1}\circ j^k_p(F\circ x_p), \end{aligned} \tag{A6}$$

where $F = y_q\circ f\circ x_p^{-1}$. Note that $F(\mathbf{0}) = \mathbf{0}$. Since $F\circ x_p$ may be expanded in powers of the functions x^i_p on U and since j^k_p projects products of the functions x^i_p with more than k factors onto the zero equivalence class, the k jet (A6) may be expanded in the form

$$\begin{aligned} j^k_pf = j^k_0\,y_q^{-1}(&F_{i_1}j^k_px^{i_1} \\ &+ \cdots + \frac{1}{k!}F_{i_1\cdots i_k}j^k_px^{i_1}\cdots j^k_px^{i_k}), \end{aligned} \tag{A7}$$

where the coefficients $F_{i_1\cdots i_r}$ are the partial derivatives of the function F at $\mathbf{0}$. The transformation law for these coefficients may readily be determined with the aid of the formulas

$$\begin{aligned} &j^k_px_p = j^k_p(x_p\circ\bar{x}_p^{-1}\circ\bar{x}_p) = j^k_0X_p\circ j^k_p\bar{x}_p, \\ &j^k_0\,y_q^{-1} = j^k_0(\bar{y}_q^{-1}\circ\bar{y}_q\circ y_q^{-1}) = j^k_0\,\bar{y}_q^{-1}\circ j^k_0\,\bar{Y}_q, \end{aligned} \tag{A8}$$

where $X_p = x_p\cdot\bar{x}_p^{-1}$ and $\bar{Y}_q = \bar{y}_q\circ y_q^{-1}$.

Both simplicity and flexibility are greatly enhanced by the following notational convention. Define

$$\begin{aligned} &d^k_px = j^k_px_p, \\ &\partial^k_px = j^k_0x_p^{-1}. \end{aligned} \tag{A9}$$

Then

$$\begin{aligned} &d^k_px\circ\partial^k_px = j^k_0I_n, \\ &\partial^k_px\circ d^k_px = j^k_pI_M. \end{aligned} \tag{A10}$$

The expansion (A7) may be written

$$\begin{aligned} j^k_pf = \partial^k_py(&F_{i_1}d^k_px^{i_1} \\ &+ \cdots + \frac{1}{k!}f_{i_1\cdots i_k}d^k_px^{i_1}\cdots d^k_px^{i_k}). \end{aligned} \tag{A11}$$

For a change of variables (A8), the following formulas are useful.

$$d_p^k\bar{x}\circ\partial_p^k x = j_0^k\bar{X}_p,$$
$$d_p^k x\circ\partial_p^k\bar{x} = j_0^k X_p. \tag{A12}$$

For mappings from or into $\mathbb{R}^n$, the standard chart $(\mathbb{R}^n, I_n)$ is used. Thus for a curve $\gamma:\mathbb{R}\to M$ such that $\gamma(0)=p$,

$$j_0^k\gamma = \partial_p^k x(\gamma_1 d_0^k I + \frac{1}{2!}\gamma_2(d_0^k I)^2 + \ldots + \frac{1}{k!}\gamma_k(d_0^k I)^k), \tag{A13}$$

where

$$\gamma_\alpha^i = \frac{d^\alpha}{dt^\alpha}(x_p^i\circ\gamma)\Big|_{t=0} \qquad \alpha = 1,2,\ldots,k. \tag{A14}$$

Note that the argument of $\partial_p^k x$ in (A13) is an element of $\mathbb{R}^n$; so that each γ_α is an n-tuple. A similar remark applies to the argument of $\partial_p^k y$ in (A11).

Finally, note that $\partial_p^k x$ and $d_p^k x$ denote k jets at a given point, and that $\partial^k x$ and $d^k x$ denote k jet fields.

APPENDIX B: NOTATION FOR JET BUNDLES

It is useful to have special notations for particular sets of jets and for particular jet bundle structures. A principal bundle structure will be denoted by

$$\mathscr{P}(M) = \langle P(M), \pi_P, M, G\rangle, \tag{B1}$$

where $P(M)$ is the *total* space, π_P is the *projection* map, M is the *base* space, and G is the *structure* group of the bundle. A bundle structure associated with a given principal bundle structure will be denoted by

$$\mathscr{E}(M) = \langle E(M), \pi_E, M, F, \mathscr{P}(M)\rangle, \tag{B2}$$

where $E(M)$ is the total space and F is the *typical fiber*.

Set

$$L_{m,n}^k = J^k(\mathbb{R}_0^m,\mathbb{R}_0^n),$$
$$A_{m,n}^k = J^k(\mathbb{R}_0^m,\mathbb{R}^n) = \mathbb{R}^n\oplus L_{m,n}^k. \tag{B3}$$

Note that $A_{m,n}^k$ may be regarded as the nth power of the algebra $A_{m,1}^k$. Denote by $G_{m,n}^k$ the subset of $L_{m,n}^k$ consisting of jets of maximal rank. For the case $m=n$, set $G_n^k = G_{n,n}^k$. The elements of G_n^k are jets of local diffeomorphisms and form a Lie group under jet composition. The group G_n^1 is customarily denoted by $\mathrm{GL}(n)$.

Let M be an n-dimensional C^∞ manifold and let $p\in M$. Set

$$L_m^k(M_p) = J^k(\mathbb{R}_0^m,\mathrm{M}_p),$$
$$L_m^{k*}(M_p) = J^k(M_p,\mathbb{R}_0^m),$$
$$A_m^k(M_p) = J^k(M_p,\mathbb{R}^m). \tag{B4}$$

The submanifolds of $L_m^k(M_p)$, the m^k speeds, and $L_m^{k*}(M_p)$, the m^k cospeeds, which consist of jets of maximal rank are the m^k frames, $H_m^k(M_p)$, and the m^k coframes, $H_m^{k*}(M_p)$. Normally, $m\leqslant n$, and for $m=n$, one writes simply $H^k(M_p)$ and $H^{k*}(M_p)$ for the sets of n^k frames and n^k coframes, respectively.

The sets obtained by taking the union of the above sets for all $p\in M$ are denoted by $L_m^k(M)$, $A_m^k(M)$, $H_m^k(M)$, and so forth. For an open subset $U\subset M$, the sets obtained by taking the union only for $p\in U$, the *portions* over U, are denoted by $L_m^k(U)$, $A_m^k(U)$, $H_m^k(U)$, and so forth. For a given coordinate chart $(U,x)_p$ for M, the locally trivializing maps may be used to give the sets $L_m^k(M)$, $A_m^k(M)$, $H_m^k(M)$, and so forth their usual topological and differentiable structures; for example, for $L_1^k(M)$, the map $\psi_U:L_1^k(U)\to L_{1,n}^k$ is defined by

$$\psi_U(j_0^k\gamma) = d_p^k x\circ j_0^k\gamma, \tag{B5}$$

where $j_0^k\gamma$ is given by (A13) and

$$d_p^k x\circ j_0^k\gamma = \partial_0^k I_n(\gamma_1 d_0^k I + \ldots + \frac{1}{k!}\gamma_k(d_0^k I)^k). \tag{B6}$$

Then $\pi_{L_1^k}\times\psi_U:L_1^k(U)\to U\times L_{1,n}^k$ is bijective and may be used together with all other such maps corresponding to the charts of an altas for M to pull back the topological and differentiable structures of the spaces $U\times L_{1,n}^k$.

The principal bundle structure for n^k frames is denoted by

$$\mathscr{H}^k(M) = \langle H^k(M), \pi_{H^k}, M, G_n^k\rangle, \tag{B7}$$

with a similar notation for the principal bundle of n^k coframes $\mathscr{H}^{k*}(M)$. There are many associated bundle structures that could be written down; for example,

$$\mathscr{L}_m^k(M) = \langle L_m^k(M), \pi_{L_m^k}, M, L_{m,n}^k, \mathscr{H}^k(M)\rangle. \tag{B8}$$

There is a natural isomorphism between $\mathscr{L}_1^1(M)$ and the tangent bundle $\mathscr{T}(M)$ and between $\mathscr{H}^1(M)$ and the principal bundle of linear frames $\mathscr{L}(M)$.

APPENDIX C: NOTATION FOR GROUPS

Let $a:\mathbb{R}^n\to\mathbb{R}^n$ be a diffeomorphism such that $a(\mathbf{0})=\mathbf{0}$. The group G_n^k is the set of k jets $j_0^k a$ together with the group product defined by k-jet composition

$$j_0^k c = j_0^k a\circ j_0^k b. \tag{C1}$$

For the case $k=3$, an element of G_n^3 may be written

$$j_0^3 a = \partial_0^3 I_n(a_j^i d_0^3 I_n^j + \frac{1}{2!}a_{jk}^i d_0^3 I_n^j d_0^3 I_n^k + \frac{1}{3!}a_{jk\ell}^i d_0^3 I_n^j d_0^3 I_n^k d_0^3 I_n^\ell) \tag{C2}$$

and the product (C1) is given explicitly by the formulas

$$c_j^i = a_r^i b_j^r,$$
$$c_{jk}^i = a_{rs}^i b_j^r b_k^s + a_r^i b_{jk}^r,$$
$$c_{jk\ell}^i = a_{rst}^i b_j^r b_k^s b_\ell^t + a_{rs}^i(b_{jk}^r b_\ell^s + b_{k\ell}^r b_j^s + b_{\ell j}^r b_k^s) + a_r^i b_{jk\ell}^r. \tag{C3}$$

The subgroup of G_n^k which consists of elements of the form

$$\partial_0^k I_n(a_j^i d_0^k I_n^j) \tag{C4}$$

will be called the affine group (of order k) and will be denoted by Γ_n^k. Let η_{ij} be the flat metric tensor with signature $s=p-q$, where $n=p+q$. Then $O_{p,q}^k$ is the subgroup of Γ_n^k consisiting of those elements (C4) of Γ_n^k for which

$$\eta_{ij}a_k^i a_\ell^j = \eta_{k\ell}. \tag{C5}$$

The orthogonal subgroup $\mathscr{O}_n^k$ of Γ_n^k is the group $\mathscr{O}_{p,q}^k$ for $p=n$ and $q=0$. The Lorentz subgroup is $O_{1,n-1}^k$. The Weyl subgroup $W_{p,q}^k$ of Γ_n^k consists of those elements (C4) of Γ_n^k which have the form

$$a^i_j = e^b b^i_j, \tag{C6}$$

where $b \in \mathbb{R}$ and b^i_j satisfies (C5).

Let $(\xi^\alpha) \in \mathbb{R}^{n+1}$ and consider the invertible linear transformations

$$\hat{\xi}^\alpha = A^\alpha_\beta \xi^\beta. \tag{C7}$$

Set

$$x^i = \frac{\xi^i}{\xi^{n+1}}, \qquad i = 1,2,\ldots n. \tag{C8}$$

The transformations (C7) induce the projective transformations

$$x^i = \frac{a^i + a^i_j x^j}{1 - a_k x^k}, \tag{C9}$$

where

$$a^i = \frac{A^i_{n+1}}{A^{n+1}_{n+1}}, \quad a_i = -\frac{A^{n+1}_i}{A^{n+1}_{n+1}}, \quad a^i_j = \frac{A^i_j}{A^{n+1}_{n+1}}. \tag{C10}$$

The projective transformations which preserve the origin ($x^i = 0$) have the form

$$\tilde{x}^i = \frac{a^i_j x^j}{1 - a_k x^k}. \tag{C11}$$

The subgroup P^k_n of G^k_n consists of the group of k jets of transformations of the form (C11). In particular, an element of P^2_n has the form

$$\partial^2_0 I_n (a^i_j d^2_0 I^j_n + \frac{1}{2!}(a^i_j a_k + a^i_k a_j) d^2_0 I^j_n d^2_0 I^k_n). \tag{C12}$$

In a slightly more complicated but similar manner, the conformal group in $\mathbb{R}^n$ may be obtained by projecting invertible linear transformations from a higher dimensional space. This group consists of transformations of the form

$$\tilde{x}^i = \frac{a^i_j x^j - \frac{1}{2}\eta^{pq} a^i_p a_q \eta_{jk} x^j x^k}{1 - a_l x^l + \frac{1}{4}\eta^{rs} a_r a_s \eta_{lm} x^l x^m}, \tag{C13}$$

where the a^i_j are given by (C6). The subgroup $C^k_{p,q}$ of G^k_n consists of k jets of transformations of the form (C13). In particular, an element of $C^2_{p,q}$ has the form

$$\partial^2_0 I_n (a^i_j d^2_0 I^j_n + \frac{1}{2!}(a^i_j a_k + a^i_k a_j - \eta^{pq} a^i_p a_q \eta_{jk}) d^2_0 I^j_n d^2_0 I^k_n). \tag{C14}$$

The following isomorphisms obtain among the groups introduced above:

$$\begin{aligned} &G^1_n \equiv P^1_n \equiv \Gamma^1_n \cong \Gamma^k_n, \quad C^1_{p,q} \equiv W^1_{p,q} \cong W^k_{p,q}, \\ &O^1_{p,q} \cong O^k_{p,q}, \\ &C^2_{p,q} \cong C^k_{p,q}, \qquad k \geqslant 2 \\ &P^2_n \cong P^\kappa_n, \qquad k \geqslant 2. \end{aligned} \tag{C15}$$

In spite of these isomorphisms, it is useful to preserve the distinctions because they permit precise and ready reference to a G structure of a given order in the theory of prolongations of G structures.

[1] J. Ehlers, F. A. E. Pirani, and A. Schild, in *General Relativity, Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Oxford, New York, 1972), pp. 63–84.

[2] Applications of these structures to mathematical physics have been discussed by J. Ehlers, in *The Physicists Conception of Nature*, edited by J. Mehra (Reidel, Dordrecht-Holland, 1973), pp. 71–91; F. A. E. Pirani and A. Schild, in *Perspectives in Geometry and Relativity, Essays in honour of Václav Hlavatý*, edited by B. Hoffman (Indiana, Bloomington, 1966), pp. 291–309; J. Ehlers and A. Schild, Commun. Math. Phys. **32**, 119 (1973); J. Ehlers and E. Köhler, J. Math. Phys. **18**, 2014 (1977); R. A. Coleman and H. Korte, J. Math. Phys. **21**, 1340 (1980); H. P. Künzle, Ann. Inst. Henri Poincaré, **XVII**, 337 (1972).

[3] Mathematical analyses of these structures have been given by: S. Kobayashi, in *Proceedings of Symposia in Pure Mathematics, Vol.* III: *Differential Geometry* (Amer. Math. Soc., Providence, RI, 1961), pp. 186–193; S. Kobayashi and T. Nagano, J. Math. Mech. **13**, 215 (1964); N. Tanaka, Nagoya Math. J. **12**, 1 (1957); K. Ogiue, Kodai Math. Sem. Rep. **19**, 193 (1967); N. Tanaka, Trans. Am. Math. Soc. **92**, 168 (1959); K. Ogiue, Kodai Math. Sem. Rep. **19**, 488 (1967).

[4] For an elegant overview of the notions of "geometric object" and "geometric object fields," see A. Nijenhuis, in *Differential Geometry*, in honour of K. Yano (Kinokuniya, Tokyo, 1972), pp. 317–334.

[5] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N. J., 1964).

[6] V. Guillemin, Trans. Am. Math. Soc. **116**, 544 (1965).

[7] V. Guillemin and S. Sternberg, J. Differen. Geom. **1**, 127 (1967).

[8] S. Kobayashi, *Transformation Groups in Differential Geometry* (Springer, Berlin, 1972).

[9] P. Molino, Theorié des G Structures: Le Probléme d'Equivalence, *Lecture Notes in Mathematics* **588** (Springer, 1977).

[10] D. C. Spencer, Ann. Math. **76**, 306 (1962); **76**, 399 (1962); **81**, 389 (1965).

[11] V. Guillemin and S. Sternberg, Mem. Am. Math. Soc. **64**, 1 (1966).

[12] J. F. Pommaret, *Systems of Partial Differential Equations and Lie Pseudogroups* (Gordon and Breach, New York, 1968).

[13] See Ref. 12, pp. 130, 132, and 400.

[14] R. A. Coleman and R. Mann, "Spacetime Metric and Yang Mills Fields Unified in a Galilean Subspace Structure," University of Toronto, preprint, 1981.

[15] A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[16] C. Ehresmann, C. R. Acad. Sci. **233**, 598, 777, 1081 (1951); **234**, 1028, 1424 (1952).

[17] For a definition of this term, see R. A. Coleman and H. Korte, Ref. 2, p. 1345. See also Ref. 4, p. 318.

[18] See J. Ehlers and A. Schild, Ref. 2, p. 128.

[19] H. Weyl, Z. Phys. **56**, 330 (1929).

[20] H. Weyl, Phys. Rev. **77**, 699 (1950).

[21] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[22] D. W. Sciama, J. Math. Phys. **2**, 472 (1961).

[23] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).

[24] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965), p. 115.

[25] E. Schrödinger, *Space-Time Structure* (Cambridge, 1954), p. 66, the symmetric part of Eq. (9.8).

[26] A discussion of this relationship has also been given by J. D. Hennig, "G-Structures and Space-Time Geometry," I: *Geometric Objects of Higher Order*, Trieste preprint, 1978; II: *Conformal and Projective Structure Imply Weyl-Structure*, Köln preprint, 1978.

Dual-mass in general relativity

Sriram Ramaswamy
James Franck Institute, University of Chicago, Chicago, Illinois 60637

Amitabha Sen[a)]
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 17 March 1981; accepted for publication 5 June 1981)

A general notion of dual-mass, the gravitational analog of the magnetic monopole, is formulated in space-times that are asymptotically empty and flat at null infinity and in which the Bondi news vanishes. Dual-mass is specified by a real valued linear function on the asymptotic infinitesimal translation symmetries which, furthermore, depends on the asymptotic dual Weyl curvature tensor. It is shown that space-times with nonzero dual-mass are characterized by a null boundary (null infinity) having the structure of a principal S^1 fiber bundle over S^2 such that the dual-mass is proportional to the number of twists, n, in the bundle. Thus the topology of null infinity is that of a lens space $L(n,1)$. A consequence of the existence of dual-mass is that the space-time is acausal. The NUT space-time is shown to be an example exhibiting these features, with a null infinity having the three-sphere topology.

PACS numbers: 04.20.Cv, 02.40. + m

I. INTRODUCTION

General relativity shares with gauge theories some formal similarities which one might hope to exploit in quantum gravity. A feature of certain gauge theories is that they admit classical solutions which are interpreted as representing magnetic monopoles and whose existence has interesting implications at the quantum level.[1] One might ask if Einstein's equations similiarly admit solutions that can be said to possess "magnetic" or "dual" mass in analogy with magnetic monopoles.

Several authors[2] have suggested that the NUT space-time,[3] a stationary solution of the vacuum Einstein equations, may be regarded as a gravitational "dyon" on account of the fact that the NUT metric is characterized by two parameters: one which can be interpreted as the mass and another (called the NUT parameter) which displays properties reminiscent of the magnetic monopole moment. However, the precise sense in which the NUT parameter is to be regarded as dual mass has remained obscure.

We shall here introduce a general characterization of dual mass in general relativity in a framework that closely parallels the description of mass of isolated gravitating systems. Recall that isolated systems in general relativity are represented by space-times which possess an asymptotic region similar to that of Minkowski space-time at (suitably defined) large spacelike or null separations from the sources. In each of these asymptotic regimes (called spatial infinity and null infinity) physical attributes such as mass and angular momentum associated with the gravitational field can be recovered from its asymptotic properties.

An elegant and useful technique is available for the treatment of the asymptotic structure of space-times.[4] Consider the description of null infinity for example. Briefly, null infinity is represented by a three-dimensional null manifold, $\mathscr{I}$,[5] with a degenerate metric and having topology $S^2 \times R$ (with R being the null generators). Fields on space-time having the "correct" asymptotic properties are registered as tensor fields on $\mathscr{I}$. Further, there exists an asymptotic symmetry group, called the BMS[6] group, acting on $\mathscr{I}$. The asymptotic symmetries give rise to integrals similar to the conserved integrals arising from symmetries of flat space-time. For example, a real valued linear function on the asymptotic translations, expressed as an integral over a cross section of $\mathscr{I}$ and involving the asymptotic Weyl tensor, yields the (Bondi) mass evaluated at the instant of "time" represented by the cross section.

In this framework, an expression similar to that for the Bondi mass but involving the asymptotic *dual* Weyl tensor presents itself as a natural candidate for dual mass. This quantity, which we shall henceforth call dual-mass, differs from the Bondi mass in two important respects. First, the expression for dual-mass is well defined if and only if the Bondi news (which measures the gravitational radiation flux) vanishes. Second, for $\mathscr{I}$ with topology $S^2 \times R$, e.g., in asymptotically Minkowskian space-times, the dual-mass vanishes identically. This suggests that the usual definition of asymptotic flatness at null infinity (which requires $\mathscr{I}$ to have $S^2 \times R$ topology) must be suitably generalized to incorporate the notion of dual-mass.

We investigate the possibility of obtaining a nonzero value for the dual-mass, N_0, by considering alternative topologies for $\mathscr{I}$, requiring only that $\mathscr{I}$ be a line bundle over S^2. We find that this is possible if and only if $\mathscr{I}$ is an S^1 bundle over S^2. Furthermore, N_0 is proportional to the number of twists, n, in the bundle. The topology of $\mathscr{I}$ is that of a lens space $L(n,1)$ which is obtained by making certain identifications on the S^1 (Hopf) fibers on S^3.[7] Since the null generators of $\mathscr{I}$ in this case are closed, a simple consequence of the existence of dual-mass ($N_0 \neq 0$) is that the space-time (with zero news) is not stably causal.[8] Moreover, this acausality cannot be removed by going to a covering space since the covering of $L(n,1)$ is S^3.

Throughout this paper the expression for dual-mass has been restricted to the asymptotic *translations* and we shall

[a)]Fellow of Center for Theoretical Physics, University of Maryland.

 0022-2488/81/112612-08$01.00

often refer to it as *dual-Bondi-mass* to emphasize this fact. It will be useful to generalize the definition of dual-mass as a real valued linear function on the infinitesimal asymptotic symmetries known as supertranslations, which form an infinite dimensional abelian subalgebra of the BMS Lie algebra and which contains the translations as a four-dimensional ideal. One can trivially extend the expression for dual-Bondi-mass to the infinitesimal supertranslations to obtain such a generalization. Unfortunately, this extension seems to be unreasonable because it can give an infinite number of conserved quantities, the supertranslation-dual-masses, when the news vanishes and $\mathscr{I}$ has topology $S^2\times R$, i.e., even when dual-Bondi-mass vanishes. It turns out that a reasonable generalization is available which, furthermore, describes dual-mass even in the presence of Bondi news. But it leads to a somewhat different structure on $\mathscr{I}$ than the one considered here and will be discussed in a later paper.

While the existence of the quantity we call dual-mass is unquestionably an immanent feature of general relativity, one would like to have some evidence in favor of its interpretation as dual-mass and develop a physical intuition for it. There are essentially two pieces of evidence. First, when the Bondi news vanishes, the expressions for mass and dual-mass differ only in that one involves the asymptotic Weyl tensor and the other its dual. Thus the "duality" here is the usual (Hodge) $*$-duality as in electromagnetism. Second, as we shall show, the NUT space-time has an asymptotic regime which can be described by a $\mathscr{I}$ with S^3 topology[9] and the dual-mass turns out to be precisely the NUT parameter. Now the NUT parameter can be regarded as "dual" to the mass for the following reason, quite different from the one described above. Consider a stationary, vacuum solution of Einstein's equations such that the manifold Σ of orbits of the stationary Killing vector field is asymptotically flat in the sense of Hansen.[10] Then there exists two functions ϕ_M and ϕ_J on Σ, called mass and angular momentum potentials, respectively, whose multipole moments evaluated at "infinity" uniquely[11] characterize the local structure of the space-time. The NUT space-time (with Σ having topology $S^2\times R$) is an example[12] and the *monopole* moments corresponding to ϕ_M and ϕ_J are precisely the mass and the NUT parameter, respectively.[13] Thus, the dual-mass in this case is to be regarded as a monopole source of angular momentum and this is the only available clue to its physical interpretation. The duality transformation in this case is a rotation in the (ϕ_M,ϕ_J) plane and is induced by the Geroch transformation[10,14] which maps one stationary vacuum solution to another.

We begin in Sec. II with a brief review of the asymptotic techniques (at null infinity) mainly to fix our notation and to assemble the results we shall need. In Sec. III, dual-mass is introduced and its general characteristics are established. A summary of the essential ideas of this section in the Newman–Penrose formalism is also given. An example of these considerations, the NUT space-time, is presented in Sec. IV. We conclude in the final section with a brief discussion. Finally, an appendix summarizes a description of dual-mass for stationary, vacuum space-times.

II. PRELIMINARIES

We briefly outline here the technique for analyzing the asymptotic structure of space-times at null infinity[15] in the formalism developed by Geroch. For details of the results in this section see Ref. 4.

Definition 1: A space-time $(\tilde{M}, \tilde{g}_{ab})$[16] will be said to be asymptotically *flat* at null infinity if there exists a manifold with boundary $M = \tilde{M}\cup I$ consisting of $\tilde{M}$ with boundary I (the ideal points at "null infinity") together with a smooth Lorentz metric g_{ab} and a smooth scalar field Ω on M such that (i) on M, $g_{ab} = \Omega^2\tilde{g}_{ab}$; (ii) at I, $\Omega = 0$, $n_a := \nabla_a\Omega \neq 0, g^{ab}\nabla_a\nabla_b\Omega = 0$ [here ∇_a is the derivative operator on (M, g_{ab})]; (iii) There exists a neighborhood N of I in M such that in $N\cap M$, g_{ab} satisfies the vacuum Einstein equations; (iv) The vector field $n^a := g^{ab}\nabla_b\Omega$ is complete and the space B of orbits of n^a is diffeomorphic to S^2. The first two conditions are meant to capture the "correct" local behavior of the gravitational field as one recedes from the isolated source in null directions. The third condition ensures that the neighborhood of infinity is a vacuum solution while the fourth specifies the global structure of null infinity.

In the last condition we permit the possibility that the integral curves of the vector field n^a on I can be closed, thus allowing the topology of I to be different from $S^2\times R$. In this sense Definition 1 differs from the usual ones.[17] Thus, for example, I may have topology S^3 with the integral curves of n^a being the S^1 (Hopf) fibers of the three-sphere. Indeed, it is crucial that the restrictions on the topology of I be relaxed because, as we shall show, any reasonably defined expression for dual-mass vanishes identically if the topology is $S^2\times R$. The motivation for Definition 1 is then to suitably generalize the criteria of asymptotic flatness so as to admit a notion of dual-mass.

The fundamental field equations are the Einstein equations, which in the matter free region $N\cap M$ take the form

$$\Omega S_{ab} + 2\nabla_a n_b - \Omega^{-1}(n^c n_c)g_{ab} = 0, \tag{1}$$

and the Bianchi identity

$$\nabla_{[a}R_{bc]d}{}^{e} = 0, \tag{2}$$

where ∇_a is the derivative operator on (M, g_{ab}), $R_{abc}{}^{d}$ its Riemann tensor,[18] and $S_{ab} = R_{ab} - \frac{1}{6}Rg_{ab}$, where R_{ab} and R are, respectively, the Ricci tensor and the scalar curvature. One may use the decomposition

$$R_{abcd} = C_{abcd} + g_{a[c}S_{d]b} - g_{b[c}S_{d]a}, \tag{3}$$

where C_{cbcd} is the Weyl tensor, together with (1) and (2) to obtain some useful equations. One such equation is, in N,

$$\nabla_{[a}S_{b]c} + \Omega^{-1}C_{abcd}n^d = 0. \tag{4}$$

Fields at infinity are conveniently studied by introducing a three-manifold $\mathscr{I}$ and a diffeomorphism $\psi:\mathscr{I}\to M$ such that $\mathscr{I}$ is a diffeomorphic copy of I. There is a unique "pullback" operator ψ^* which maps covariant tensor fields $T_{a\ldots b}$ to a tensor of the same rank on $\mathscr{I}$, denoted by $\psi^*(T_{a\ldots b}) := \mathbf{T}_{a\ldots b}$ such that (i) for a scalar field $\phi:M\to R$, $\psi^*(\phi) = \phi\circ\psi$, (ii) pullback of the gradient of a scalar field is the gradient of its pullback, and (iii) ψ^* is distributive and commutative with respect to taking sums and outer products, respectively, of covariant tensor fields. The action of

ψ^* can be extended to contravariant tensor field $T^{a\ldots b}$ if and only if the contravariant indices are tangent to I. For example, since $n^a n_a = 0$ at I, n^a is tangent to I and so $\mathbf{n}^a = \psi^*(n^a)$ exists on $\mathscr{I}$.

The geometry inherited by $\mathscr{I}$ is described by the vector field $\mathbf{n}^a := \psi^*(n^a)$ and a (degenerate) metric $\mathbf{g}_{ab} := \psi^*(g_{ab})$ satisfying

$$\text{(i)}\mathscr{L}_{\mathbf{n}}\mathbf{g}_{ab} = 0, \quad \text{(ii)}\mathbf{g}_{ab}\mathbf{n}^b = 0 \tag{5}$$

and with signature $(0, +, +)$. Alternating tensors on $\mathscr{I}$, ϵ^{abc}, and ϵ_{abc}, are fixed (up to sign) by

$$\epsilon^{abc}\epsilon^{pqr}\mathbf{g}_{ap}\mathbf{g}_{bq} = 2\mathbf{n}^c\mathbf{n}^r,$$
$$\epsilon^{abc}\epsilon_{abc} = 6. \tag{6}$$

There are essentially four objects on $\mathscr{I}$ that capture the asymptotic properties of the gravitational field at null infinity. They are (i) a derivative operator D_a on $\mathscr{I}$ satisfying

$$D_a\mathbf{n}^b = 0, \quad D_a\mathbf{g}_{bc} = 0, \tag{7}$$

(ii) $\mathbf{S}_a{}^b := \psi^*(S_{ac}g^{bc})$, (iii) the asymptotic Weyl tensor $K^{ab} := \epsilon^{apq}\epsilon^{blm}\psi^*(\Omega^{-1}C_{pqlm})$, (iv) the asymptotic dual Weyl tensor ${}^*K^{ab} := \epsilon^{apq}\epsilon^{blm}\psi^*(\Omega^{-1}\frac{1}{2}\epsilon_{pq}{}^{cd}C_{cdlm})$. Their existence is ensured by Definition 1 and the field equations (1) and (2). In particular, the existence of (iii) and (iv) is guaranteed if the manifold of orbits, B, of the vector field $\mathbf{n}^a$ on $\mathscr{I}$ is *simply connected* which is the case here (B is a two-sphere). Note that $\mathscr{I}$ itself is *not* required to be simply connected (e.g., $\mathscr{I}$ can have topology $S^1\times S^2$). The properties of these fields that we shall use are listed below.

$$D_a[\mathbf{S}_b{}^a - \mathbf{S}\delta_b{}^a] = 0, \quad \mathbf{S} := \mathbf{S}_m{}^m, \tag{8a}$$
$$D_{[a}\mathbf{S}_{b]}{}^c = \tfrac{1}{4}\epsilon_{abm}{}^*K^{mc}, \tag{8b}$$
$$D_a{}^*K^{ab} = 0, \quad D_aK^{ab} = 0. \tag{8c}$$

K^{ab} and ${}^*K^{ab}$ are symmetric and trace-free and satisfy

$$\mathbf{g}_{am}K^{mb} = -\epsilon_{amp}\mathbf{n}^p{}^*K^{mb}, \quad \mathbf{g}_{am}{}^*K^{mb} = \epsilon_{amp}\mathbf{n}^pK^{mb}. \tag{8d}$$

The global structure of $\mathscr{I}$ is that of a fiber bundle over S^2 (the base space) with the integral curves of $\mathbf{n}^a$ as the fibers. Let $\pi:\mathscr{I}\to S^2$ denote the projection that maps each fiber of $\mathscr{I}$ onto a point of S^2. Then one can pullback tensor fields on S^2 via π^* defined in the same manner as ψ^*. The following result is useful. If $\mu_{a\ldots b}$ is a covariant tensor field on $\mathscr{I}$ such that (i) $\mathbf{n}^a\mu_{a\ldots b} = 0,\ldots,\mathbf{n}^b\mu_{a\ldots b} = 0$ and (ii) $\mathscr{L}_{\mathbf{n}}\mu_{a\ldots b} = 0$, then there is a unique covariant tensor field $\hat{\mu}_{a\ldots b}$ on S^2 such that $\mu_{a\ldots b} = \pi^*(\hat{\mu}_{a\ldots b})$. Thus, for example, (5) shows that $\mathbf{g}_{ab}$ is the pullback of a metric on the two-sphere.

There is a freedom, which we shall call gauge, to pass from (M, g_{ab}, Ω) to an equivalent asymptotic description (M, g'_{ab}, Ω') by the transformations

$$g'_{ab} = \omega^2 g_{ab}, \quad \Omega' = \omega\Omega, \tag{9}$$

where ω is a positive scalar field on M. To preserve $\nabla_a n^a = 0$ at I in Definition 1, ω must satisfy $n^a\nabla_a\omega = 0$ at I. The gauge transformations change tensor fields on $\mathscr{I}$ (but leave invariant the field equations). For example,

$$\mathbf{g}'_{ab} = \omega^2\mathbf{g}_{ab}, \quad \mathbf{n}'^a = \omega^{-1}\mathbf{n}^a, \quad \epsilon'_{abc} = \omega^3\epsilon_{abc},$$
$$K'^{ab} = \omega^{-5}K^{ab}, \quad {}^*K'^{ab} = \omega^{-5}{}^*K^{ab}, \tag{10a}$$

$$S'_a{}^b = \omega^{-2}\mathbf{S}_a{}^b - 2\omega^{-3}D_a\omega^b + 4\omega^{-4}\omega^bD_a\omega - \omega^{-4}\delta_a{}^b\omega^mD_m\omega, \tag{10b}$$

where $\omega = \psi^*(\omega)$ and $\omega^a = \psi^*(\nabla^a\omega)$.

Physical quantities on $\mathscr{I}$ must be gauge invariant. One important gauge invariant tensor field is the Bondi news function which is essentially the gauge invariant part of $\mathbf{S}_a{}^b$. This is obtained by introducing an auxilliary symmetric tensor field ρ_{ab} which is *uniquely* defined by the requirements

$$\rho_{ab}\mathbf{n}^a = 0, \quad \rho_{ab}\mathbf{g}^{ab} = \mathscr{R}, D_{[a}\rho_{b]c} = 0, \tag{11}$$

where $\mathscr{R}$ is the scalar curvature of the two-sphere base space of $\mathscr{I}$ and $\mathbf{g}^{ab}$ is an "inverse" metric defined by $\mathbf{g}^{ab}\mathbf{g}_{ab}\mathbf{g}_{bn} = \mathbf{g}_{mn}\cdot\rho_{ab}$ has the same gauge behavior as $\mathbf{S}_{ab}$ and so the Bondi news N_{ab}, defined by

$$N_{ab} = \mathbf{S}_{ab} - \rho_{ab}, \tag{12}$$

is gauge invariant. From (8b) and (11)

$$D_{[a}N_{b]c} = \tfrac{1}{4}\epsilon_{abm}{}^*K^{mn}\mathbf{g}_{nc}. \tag{13}$$

Finally, there exists a group of diffeomorphisms of $\mathscr{I}$ that preserves its geometry.[19] In the case $\mathscr{I}$ has topology $S^2\times R$, this group, known as the Bondi–Metzner–Sachs (BMS)[6] group, turns out to be a semidirect product of an infinite dimensional abelian group (called supertranslations) and the Lorentz group. If $\mathscr{I}$ has different topology but remains a bundle over S^2, then the group has a covering space isomorphic to that of BMS. The infinitesimal supertranslations are of the form $\xi^a = \alpha\mathbf{n}^a$, where α is a scalar function on $\mathscr{I}$ such that $\mathscr{L}_{\mathbf{n}}\alpha = 0$. Note, under gauge transformation

$$\alpha' = \omega\alpha, \tag{14}$$

in order to have $\alpha'\mathbf{n}'^a = \alpha\mathbf{n}^a$. If, in addition, α satisfies $D_aD_b\alpha + \frac{1}{2}\alpha\rho_{ab} = \mu\mathbf{g}_{ab}$ for some function μ on $\mathscr{I}$, $\alpha\mathbf{n}^a$ is known as an infinitesimal *translation*. They form a four-dimensional abelian ideal of the BMS Lie algebra.

III. DUAL MASS

We recall first the description of mass at null infinity in an asymptotically flat space-time. Let 1_a be a covector on $\mathscr{I}$ such that $\mathbf{n}^a1_a = -1$ and fix an infinitesimal supertranslation $\alpha\mathbf{n}^a$. Define a vector

$$p^a = \tfrac{1}{4}\alpha K^{am}1_m + (\alpha D_m1_n + 1_mD_n\alpha)\mathbf{g}^{np}N_{pq}g^{q[m}\mathbf{n}^{a]}. \tag{15}$$

Then the number M defined by

$$M = \frac{1}{8\pi}\int_C\epsilon_{abc}p^cdS^{ab} \tag{16}$$

is gauge invariant, independent of the choice of 1_a and can be interpreted[4] as the total energy-momentum of the system associated with the supertranslation $\alpha\mathbf{n}^a$ at the instant represented by the cross section[20] C of $\mathscr{I}$. If C' is a cross section to the future of C and M' the number associated with it, then

$$M' - M = \frac{1}{8\pi}\int_A(D_ap^a)dV \tag{17}$$

gives the (α-weighted) flux of energy-momentum over the intervening region A. For an infinitesimal translation $\alpha\mathbf{n}^a$

$$D_ap^a = -\tfrac{1}{4}\alpha N_{mn}N_{pq}\mathbf{g}^{mp}\mathbf{g}^{nq}. \tag{18}$$

From (17) and (18) we see that the news is an amplitude for the gravitational radiation flux. Finally, we note that the second term in (15) involving the news N_{ab}, together with Eq. (13), are crucial in giving p^a the "right" properties that we just described.

To obtain a quantity representing "dual" mass, we seek a vector[21] ${}^*p^a$, analogous to (15) but involving ${}^*K^{ab}$, such that the number N defined by

$$N = \frac{1}{8\pi}\int_C \epsilon_{abc}\,{}^*p^c dS^{ab} \tag{19}$$

is gauge invariant and independent of the choice of 1_a.

An exact counterpart of (15) is impossible to obtain because of the asymmetry between the Einstein's equation and Bianchi identity. Specifically, there is no counterpart of Eq. (13) involving K^{ab} instead of ${}^*K^{ab}$. So there is no ("dual news") function akin to the news that can be used in obtaining an expression for ${}^*p^a$ analogous to p^a. However, we may consider the following natural candidate for ${}^*p^a$:

$${}^*p^a = \tfrac{1}{4}\alpha\,{}^*K^{am}1_m. \tag{20}$$

Since under gauge transformation $\mathbf{g}'_{ab} = \omega\mathbf{g}_{ab}, \epsilon'_{abc} = \omega^3\epsilon_{abc}, \alpha' = \omega\alpha, {}^*K'^{ab} = \omega^{-5}{}^*K^{ab}$, and $1'_a = \omega 1_a$ (since $\mathbf{n}^a 1'_a = \mathbf{n}^a 1_a = -1$) [see (10) and (14)], the quantity

$$N = \frac{1}{32\pi}\int_C \alpha\epsilon_{abc}\,{}^*K^{cm}1_m dS^{ab} \tag{21}$$

obtained by substituting (20) in (19) is gauge invariant. However, it is not independent of the choice of 1_a in general.

Theorem 1: N defined by (21) is independent of 1_a if and only if the news vanishes, i.e., $N_{ab} = 0$.

Proof: Let 1_a and $\hat{1}_a$ be two covectors such that $\mathbf{n}^a 1_a = -1$ and $\mathbf{n}^a\hat{1}_a = -1$. Then $\mathbf{n}^a(1_a - \hat{1}_a) = 0$ whence $1_a - \hat{1}_a = \mathbf{g}_{am}v^m$ for some vector field v^m on $\mathscr{I}$. If $N(1_a)$ is independent of 1_a, then $N(1_a) - N(\hat{1}_a) = 0$ gives

$$\int_C \alpha\epsilon_{abc}\,{}^*K^{cm}\mathbf{g}_{mn}v^n dS^{ab} = 0 \quad \forall v^n, \alpha, C.$$

So $\epsilon_{abc}\,{}^*K^{cm}\mathbf{g}_{mn} = 0$. Then from (13)

$$D_{[a}N_{b]c} = 0. \tag{22}$$

Since N_{ab} satisfies $\mathbf{g}^{ab}N_{ab} = 0$, $\mathbf{n}^a N_{ab} = 0$,[4] (22) shows that if ρ_{ab} is a solution of (11) then $\rho_{ab} + N_{ab}$ is also a solution. But the solution of (11) is unique. So $N_{ab} = 0$. The converse is trivial.

We shall henceforth restrict ourselves to the case when $N_{ab} = 0$ and explore the consequences of dual-mass defined by

$$N = \frac{1}{32\pi}\int_C \alpha\epsilon_{abc}\,{}^*K^{cm}1_m dS^{ab}. \tag{21}$$

The mass in this case is given by

$$M = \frac{1}{32\pi}\int_C \alpha\epsilon_{abc}K^{cm}1_m dS^{ab} \tag{23}$$

from which the duality is manifest.

Theorem 2: If $N_{ab} = 0$, then there exists smooth functions μ and λ on $\mathscr{I}$ such that

$$K^{ab} = \mu\mathbf{n}^a\mathbf{n}^b, \quad {}^*K^{ab} = \lambda\mathbf{n}^a\mathbf{n}^b, \quad \mathscr{L}_{\mathbf{n}}\mu = 0, \quad \mathscr{L}_{\mathbf{n}}\lambda = 0.$$

Proof: If $N_{ab} = 0$, then there is a choice of gauge for which $\mathbf{S}_{ab} = \mathbf{g}_{ab}$.[4] Then $\mathbf{S}_a{}^b = \delta_a{}^b - v_a\mathbf{n}^b$ for some covector field v_a on $\mathscr{I}$. From (8.2)

$$\tfrac{1}{4}\epsilon_{abm}\,{}^*K^{mc} = D_{[a}S_{b]}{}^c = -D_{[a}v_{b]}\mathbf{n}^c, \tag{24}$$

whence ${}^*K^{mc} = -2\epsilon^{mab}(D_a v_b)\mathbf{n}^c = x^m\mathbf{n}^c$. Since ${}^*K^{ab}$ is symmetric, $x^m = \lambda\mathbf{n}^m$ for some function λ on $\mathscr{I}$. The equation $D_a{}^*K^{ab} = 0$ (8c) implies $\mathbf{n}^a D_a\lambda = 0$, i.e., $\mathscr{L}_{\mathbf{n}}\lambda = 0$. Next, the form ${}^*K^{ab} = \lambda\mathbf{n}^a\mathbf{n}^b$ and (8d) lead to $\mathbf{g}_{am}K^{mb} = 0$ and $\epsilon_{amp}\mathbf{n}^p K^{mb} = 0$, which in turn implies $K_{ab} = \mu\mathbf{n}^a\mathbf{n}^b$ for some function μ on $\mathscr{I}$. Again, $\mathscr{L}_{\mathbf{n}}\mu = 0$ follows from (8c).

Corollary: If $N_{ab} = 0$, $P^a = \mu\mathbf{n}^a$ and ${}^*P^a = \lambda\mathbf{n}^a$, satisfying $D_a P^a = 0$ and $D_a{}^*P^a = 0$. Hence (21) and (23) are independent of the choice of cross setion C. Further, since $\mathscr{L}_{\mathbf{n}}\mu = 0$ and $\mathscr{L}_{\mathbf{n}}\lambda = 0$, there exists functions $\hat{\mu}$ and $\hat{\lambda}$ on the base space (S^2) of $\mathscr{I}$ such that $\pi^*(\hat{\mu}) = \mu$ and $\pi^*(\hat{\lambda}) = \lambda$. M and N can then be expressed by

$$M = -\frac{1}{32\pi}\int_{S^2}\hat{\alpha}\hat{\mu}\epsilon_{ab}dS^{ab}, \tag{25a}$$

$$N = -\frac{1}{32\pi}\int_{S^2}\hat{\alpha}\hat{\lambda}\epsilon_{ab}dS^{ab} \tag{25b}$$

as integrals over the base space, where $\hat{\alpha}$ is defined by $\pi^*(\hat{\alpha}) = \alpha$ (since for supertranslation $\alpha\mathbf{n}^a$, $\mathscr{L}_{\mathbf{n}}\alpha = 0$) and $\epsilon_{ab}(\equiv\epsilon_{abc}n^c)$ is the alternating tensor on the base space.

We now consider the properties of dual-mass associated with the infinitesimal *translations*. In this case, it is appropriate to refer to (20) and (21) as dual-Bondi-momentum and dual-Bondi-mass (henceforth denoted by N_0), respectively. We shall later argue that in fact the expression (21) is reasonable only when restricted to the infinitesimal translations.

Lemma 1: If $\alpha\mathbf{n}^a$ is an infinitesimal translation such that α vanishes nowhere on $\mathscr{I}$, i.e., $\alpha\mathbf{n}^a$ is an infinitesimal "time" translation, then there is a choice of gauge such that $\alpha =$ const and the scalar curvature of the base space (S^2) $\mathscr{R} =$ const.

Proof: For an infinitesimal translation $\alpha\mathbf{n}^a, \alpha$ satisfies the gauge invariant equation

$$g^{mn}D_m\alpha D_n\alpha - g^{mn}D_m D_n\alpha - \tfrac{1}{2}\mathscr{R}\alpha^2 = \text{const} \tag{26}$$

[Eq. (37) in Ref. 4]. If α is everywhere positive (say) then one can choose a gauge such that $\alpha =$ const because under gauge transformation $\alpha \to \alpha' = \omega\alpha$. Then by (26) $\mathscr{R} =$ const.

Theorem 3: If $N_{ab} = 0$ and $\mathscr{I}$ admits a two-sphere cross section, then dual-Bondi-mass $N_0 = 0$.

Proof: Fix a "time" translation $\alpha\mathbf{n}^a$ and a gauge such that $\alpha =$ const. Then from Lemma 1, $\mathscr{R} =$ const and if $N_{ab} = 0$, $S_{ab} = \rho_{ab}$ [from (12)] with $\rho_{ab} = (\mathscr{R}/2)g_{ab}$. Without loss of generality, set $\mathscr{R} = 2$. Then $\mathbf{S}_{ab} = \mathbf{g}_{ab}$ and, from (24), ${}^*K^{mc} = -2\epsilon^{mab}(D_a v_b)\mathbf{n}^c$. Therefore

$$N_0(\alpha) = \frac{1}{32\pi}\int_C \alpha\epsilon_{abc}\,{}^*K^{cm}1_m dS^{ab} = \frac{1}{8\pi}\int_C \alpha D_{[a}v_{b]}dS^{ab}. \tag{27}$$

But since in this choice of gauge $\alpha =$ const, N_0 is an integral of a curl of a smooth convector v_a over a two-sphere cross section. Hence, $N_0(\alpha) = 0$. Choosing four linearly indepen-

dent "time" translations (that span the space of infinitesimal translations) we find that the dual-mass associated with each is zero. The $N_0 = 0$ because N_0 is linear in the translations.

Theorem 4: If $N_{ab} = 0$ and $N_0 \neq 0$, then $\mathscr{I}$ is a principal S^1 fiber bundle over S^2 [$P(S^2, S^1)$][22] such that the dual-mass N_0 is proportional to the number of twists, n, in the bundle. Further, $\mathscr{I}$ has the topology of a lens space $L(n,1)$.[7] Conversely, if $\mathscr{I}$ has the structure $P(S^2, S^1)$ then $N_{ab} = 0$ and there exists some infinitesimal translation such that the dual-mass associated with it is nonzero.

Proof: Fix an infinitesimal "time" translation $\alpha \mathbf{n}^a$ and a gauge such that $\alpha = \text{const}$. If $N_{ab} = 0$ then from (25.2)

$$N_0(\alpha) = \frac{\hat{\alpha}}{8\pi} \int_{S^2} \hat{F}_{ab} dS^{ab}, \tag{28}$$

where $\hat{F}_{ab} = -(\hat{\lambda}/4)\epsilon_{ab}$. Also, from (24) and Theorem 2, $F_{ab} = D_{[a} v_{b]} = -(\lambda/4)\epsilon_{abc}\mathbf{n}^c$, where $F_{ab} = \pi^*(\hat{F}_{ab})$. Note that

$$(1)\ \mathbf{n}^a F_{ab} = 0 \text{ and } (2)\ \mathscr{L}_{\mathbf{n}} F_{ab} = 0. \tag{29}$$

From the identity

$$\mathbf{n}^a D_{[a} v_{b]} = \tfrac{1}{2}\mathscr{L}_{\mathbf{n}} v_b - \tfrac{1}{2} D_b(v_m \mathbf{n}^m)$$

and (29.1),

$$\mathscr{L}_{\mathbf{n}} v_b = D_b(v_m \mathbf{n}^m). \tag{30}$$

Since v_b is arbitrary up to a gradient, we can always choose

$$\mathscr{L}_{\mathbf{n}} v_b = 0. \tag{31}$$

Then from(30), $D_a(v_m \mathbf{n}^m) = 0$. So

$$v_m \mathbf{n}^m = K \tag{32}$$

for some constant K (we shall see that if $N_0 \neq 0$ then $K \neq 0$). Next, $\mathscr{I}$ can be viewed as a principal fiber bundle over S^2 with the group generated by the infinitesimal translation $\alpha \mathbf{n}^a$ acting on it. Then (31) and (32) show that v_a is in fact a *connection* on this bundle.[23] We now establish[24] that with $N_0 \neq 0$ $\mathscr{I}$ is in fact a S^1 bundle over S^2. Regard the base space two-sphere as a loop of loops $\hat{\gamma}_s(t)$, $0 \leqslant s,t \leqslant 1$, all with the same base point p [$\hat{\gamma}_s(0) = \hat{\gamma}_s(1) = p$] and beginning and ending with the trivial loop at p [$\hat{\gamma}_0(t) = \hat{\gamma}_1(t) = p$]. Fix a point P in $\mathscr{I}$ on the fiber over p [i.e., $\pi(P) = p$] and let $\gamma_s(t)$ to be the unique horizontal lift (defined by the connection v_a) of the loop $\hat{\gamma}_s$ such that $\gamma_s(0) = P$. Let P_s denote the point $\gamma_s(1)$ which necessarily lies on the fiber (containing P) over p. (Note: P_s need not coincide with P.) The surface Σ_s swept out by $\gamma_r, 0 \leqslant r \leqslant s$, is bounded by a closed curve consisting of γ_s from P to P_s and the segment σ_s from P_s to P along the fiber containing them. Using Stokes' theorem then

$$\int_{\Sigma_s} F_{ab} dS^{ab} = \int_{\gamma_s + \sigma_s} v_a dS^a = \int_{\sigma_s} v_a dS^a, \tag{33}$$

where in the last equality we have used the fact that γ_s is a horizontal lift. On the other hand the left side of (33) is $\int_{\hat{\Sigma}_s} \hat{F}_{ab} dS^{ab}$, where $\hat{\Sigma}_s$, the projection of Σ_s on the base space, is the surface swept out by $\gamma_r, 0 \leqslant r \leqslant s$. Hence, as s increases to 1, we have

$$N_0(\alpha) = \frac{\hat{\alpha}}{8\pi} \int_{S^2} \hat{F}_{ab} dS^{ab} = \frac{\alpha}{8\pi} \oint_{\sigma} v_a dS^a, \tag{34}$$

where σ is a closed path along the fiber over p, starting and ending at the point P. Using (32), one has

$$N_0(\alpha) = \frac{\alpha}{8\pi} K \oint_{\sigma} ds. \tag{35}$$

Note that K cannot vanish if $N_0 \neq 0$. Further, if $N_0 \neq 0$, then clearly σ must traverse the entire fiber (if it did not, then there would be a point $P_0 \neq P$ where σ would turn and retrace its path back to P, giving $N_0 = 0$). Hence the fiber must be a closed loop and

$$N_0(\alpha) = n\,(\alpha K L/8\pi), \tag{36}$$

where L is the period of the fiber [in the terms of the affine parameter s, $[\mathbf{n}^a = (\partial/\partial s)^a$ and n is an integer, called the twist, which counts the number of times σ winds around the fiber. The twist depends only on the underlying bundle structure (and not on the connection v_a, as might have been suggested by the above construction).[23] Since the above argument does not depend on the base point p, $\mathscr{I}$ is clearly a S^1 fiber bundle over S^2. Now, S^1 bundles over S^2 are $S^1 \times S^2$ and the lens spaces $L(n,1)$[7] for each integer n. In particular, the lens space for $n = 1$ is just the three-sphere (S^3). The product space $S^1 \times S^2$ admits two-sphere cross sections and is therefore incompatible with $N_0 \neq 0$ by Theorem 3. Hence $\mathscr{I}$ has the topology of $L(n,1)$. Finally, repeating the above construction for four linearly independent "time" translations, we conclude that the dual-mass is proportional to the number of twist in the bundle.

To establish the converse we merely note that if $\mathscr{I}$ is compact, orientable and without boundary, then $N_{ab} = 0$. Consider an infinitesimal translation with $\alpha = \text{const}$. Then integrating Eq. (18) over $\mathscr{I}$,

$$-\frac{\alpha}{4} \int_{\mathscr{I}} N_{mn} N_{pq} \mathbf{g}^{mp} \mathbf{g}^{nq} dv = \int_{\mathscr{I}} D_a p^a dv = 0. \tag{37}$$

Since the integrand in the left side of (37) is positive definite, $N_{ab} = 0$. The existence of a nonzero N_0 is easy to establish by retracing the construction presented above. This completes the proof.

It should be noted that the S^1 bundle structure of $\mathscr{I}$ implies that space-times with dual-mass N_0 are not stably causal.[8]

Theorems 3 and 4 bring out the interplay between the topology of null infinity and the existence of dual-Bondi-mass. The fact that it vanishes for $\mathscr{I}$ with topology $S^2 \times R$ (e.g., in asymptotically Minkowskian space-times) speaks in favor of the expression for dual-Bondi-mass, i.e., (21) restricted to infinitesimal *translations*. However, if (21) is generalized to include all supertranslations then the supertranslation-dual-mass need not vanish for an $S^2 \times R$ $\mathscr{I}$ [consider (27) with $\alpha \neq \text{const}$]. This generalization therefore seems to be an unreasonable one.

We end this section with an outline of the main results in the Newman–Penrose formalism. (For a description of asymptotically flat spacetimes in this framework see Ref. 25.)

Let (M, g_{ab}) be the conformal completion of an asymptotically flat space-time (Definition 1) with a null boundary I which is a line bundle over a two-sphere. At each point of I, the space tangent to I can be spanned by n^a, the tangent to the

null generators of I, and two complex vectors m^a and (its complex conjugate) $\bar{m}^a$ such that

$$n^a m_a = 0 = n^a \bar{m}_a,\ m^a m_a = 0 = \bar{m}^a \bar{m}_a,\ m^a \bar{m}_a = 1.$$

Given a neighborhood S on I, the coordinates u along the null generators of I in S is specified by *local* $u = \text{const}$ surfaces obtained by the intersection of S with null hypersurfaces in N. If these hypersurfaces are generated by the (null) vector field l^a in N, then one can arrange to have

$$l^a n_a = 1, \quad l^a m_a = 0 = l^a \bar{m}_a.$$

Note that $l_a = \nabla_a u$. Thus at points of $S \subset I$ one obtains a tetrad $(l^a, n^a, m^a, \bar{m}^a)$. Asymptotic tensor fields at I can be *locally* specified by their tetrad components (in S). The ten components of the asymptotic Weyl tensor is given by five complex functions $\Psi_0, \Psi_1, \Psi_2, \Psi_3, \Psi_4$. Also of interest is the quantity $\sigma := m^a m^b \nabla_a l_b$ at I which measures the *shear* of the $u = \text{const}$ surfaces in S. Defining: $\dot{} = n^a D_a$, $\eth := (\sqrt{2}) m^a D_a, \bar{\eth} := (\sqrt{2}) \bar{m}^a D_a$, the fundamental field equations (in S) are[25]

$$\dot{\psi}_0 = -\eth \psi_1 + 3\sigma \psi_2,$$
$$\dot{\psi}_1 = -\eth \psi_2 + 2\sigma \psi_3,$$
$$\dot{\psi}_2 = -\eth \psi_3 + \sigma \psi_4$$

and

$$\psi_4 = -\ddot{\bar{\sigma}},$$
$$\psi_3 = +\eth \dot{\bar{\sigma}},$$
$$\psi_2 - \bar{\psi}_2 = \bar{\eth}^2 \sigma - \eth^2 \bar{\sigma} + \bar{\sigma} \dot{\sigma} - \sigma \dot{\bar{\sigma}}.$$

Information about the *news* is captured by $\dot{\sigma}$,

$$\dot{\sigma} = m^a m^b N_{ab}.$$

In the usual description of asymptotically flat space-times (with $\mathscr{I}$ of topology $S^2 \times R$), the $u = \text{const}$ surfaces are two-sphere cross sections of I and so a global system of tetrads can be defined on I. Here we differ only in that the description is *local*, allowing the possibility that I need not admit a two-sphere cross section.

When the news vanishes, $\dot{\sigma} = 0$, the field equations show that $\dot{\psi}_2 = 0$, i.e., ψ_2 is a (complex) function on the base space (S^2) of I. In fact, if λ and μ are functions defined in Theorem 2, then ψ_2 is proportional to $\mu + i\lambda$. Then the dual-mass for a "time" translation n^a (setting $\alpha = 1$) is given by

$$N_0 = k \int_{S^2} \text{Im} \psi_2 d\Omega$$

where k is a numerical factor and S^2 is the base space.

Suppose I admits two-sphere cross sections and that σ is everywhere smooth on I. Then since $\text{Im}\psi_2 = \text{Im}\bar{\eth}^2 \sigma$ ($\dot{\sigma} = 0$) on I, one can pull back the right hand side on the base space by a cross section to obtain

$$N_0 = k \int_{S^2} \text{Im} \bar{\eth}^2 \sigma \, d\Omega = 0.$$

(Note: $\eth f$ is essentially a curl.) This is the content of Theorem 3. Consequently if $N_0 \neq 0$, then I cannot possibly admit a cross section. The assertion that N_0 must be proportional to the twist of the bundle (Theorem 4) is difficult to see in this framework.

Throughout we have assumed that all fields (and in particular σ) are smooth at I. An alternative stand would be to consider I with topology $S^2 \times R$ and allow $\bar{\eth}\sigma$ to be singular at some point (s) of a cross section. In that case it is possible to obtain a nonzero value for the dual-mass. Thus, in this picture, dual-mass reflects a discontinuity (in a rough sense) of the shear of any given cross section of I.

IV. AN EXAMPLE

We present here the NUT space-time as an example of the considerations in the previous section.

In a coordinate chart the NUT metric is given by[3,9]

$$ds^2 = \tilde{g}_{ab} dx^a dx^b = -U(dt + A\, d\phi)^2 + U^{-1} dr^2 + (r^2 + l^2)(d\theta^2 + \sin^2\theta\, d\phi^2),$$

where $U = 1 - 2(mr + 1^2)/(r^2 + 1^2)$, $A = 2l(1 - \cos\theta)$. The constants m and l are, respectively, the mass and the NUT parameter. The metric in the given form has a coordinate singularity at $\theta = \pi$ which is removable by replacing t by, for instance, $t' = t + 4l\phi$. Then the metric has the same form with A replaced by $A' = -2l(1 + \cos\theta)$ and the t' coordinate is periodic with period $8\pi l$. With this interpretation of the coordinate singularity the NUT space-time acquires the topology $S^3 \times R$ with the vector field $\tilde{\eta}^a = (\partial/\partial t)^a$ generating the S^1 (Hopf) fibers of S^3. The Euler coordinates of S^3 are $\psi = t/2l$, θ, and ϕ. The space-time admits Killing vector fields $\tilde{\eta}^a$ and $\tilde{\xi}^a_i, i = 1,2,3$, satisfying $\mathscr{L}_{\tilde{\xi}_1} \tilde{\xi}^a_2 = \tilde{\xi}^a_3$, etc., and $\mathscr{L}_{\tilde{\xi}_i} \tilde{\eta}^a = 0$, which act transitively on $r = \text{const}$ three-spheres. There are closed timelike curves (that cannot be removed by going to a covering space since NUT space is simply connected) and no reasonable spacelike surfaces in this space-time. Thus, a notion of "spatial" infinity cannot be introduced although, as we show below, the notion of "null infinity" exists.

Define new coordinates $u = t - \int U^{-1} dr$, $x = r^{-1}$. Then u has period $8\pi l$, and

$$\tilde{g}_{ab} dx^a dx^b = -U[du^2 - 2U^{-1}x^{-2} du\, dx + 2A(du - U^{-1}x^{-2}dx)\, d\phi] + x^{-2}[1 + l^2 x^2]\, d\theta^2 + [x^{-2}(1 + l^2 x^2)\sin^2\theta - UA^2]\, d\phi^2.$$

With $\Omega = x$, $g_{ab} = \Omega^2 \tilde{g}_{ab}$ is given by

$$g_{ab} dx^a dx^b = -U[x^2 du^2 - 2U^{-1} du\, dx + 2A(x^2 du - U^{-1} dx) d\phi] + (1 + l^2 x^2) d\theta^2 + [(1 + l^2 x^2)\sin^2\theta - UA^2 x^2] d\phi^2.$$

Extend the coordinate patch to include $x = 0$ (i.e., $r \to \infty$ points at infinity) for all u, θ, ϕ. This amounts to adding a three-sphere at "$r \to \infty$" infinity. Thus one has a smooth manifold $M = \tilde{M} \cup I$ with a three-sphere boundary I (at $x = 0$). The induced metric on I is $g_{ab} dx^a dx^b$ evaluated at $x = 0$, $dx = 0$, given by

$$\mathbf{g}_{ab} dx^a dx^b = d\theta^2 + \sin^2\theta\, d\phi^2,$$

which is clearly degenerate, with signature $(0,+,+)$. The metric annihilates all vectors parallel to $\mathbf{n}^a = (\partial/\partial u)^a$. The integral curves of this vector field are the Hopf fibers of the S^3 boundary, so the manifold B of orbits of $\mathbf{n}^a$ is a topological two-sphere. Since in our choice of conformal frame the induced metric on B is $h_{ab}dx^a dx^b = d\theta^2 + \sin\theta d\phi^2$, B is a metric two-sphere. It is straightforward to check that (M, g_{ab}, Ω) described above satisfies the conditions of Defintion 1. In this sense the NUT space-time is asymptotically flat at null infinity. Note that $\mathscr{I}$ in this case is to be viewed as I detached from M, i.e., a null three-sphere coordinatized by u, θ, ϕ and with the metric $\mathbf{g}_{ab}$ and vector field $\mathbf{n}^a$ there on.

Finally, we compute the mass and dual-mass associated with the time-like translation Killing field $\tilde{\eta}^a$, which on $\mathscr{I}$ is given by $\mathbf{n}^a$ in our choice of gauge. Note that since $\mathscr{I}$ is a three-sphere, the news vanishes from general considerations; thus the expression for dual-mass is well defined. We know K^{ab} and ${}^*K^{ab}$ must be multiples of $\mathbf{n}^a\mathbf{n}^b$. The multiples are computed as follows. First, define fields

$$\tilde{E}_{ab} = \tilde{C}_{ambn}\tilde{\eta}^m\tilde{\eta}^n, \quad \tilde{B}_{ab} = {}^*\tilde{C}_{ambn}\tilde{\eta}^m\tilde{\eta}^n, \tag{38}$$

where $\tilde{C}_{ambn}$ and ${}^*\tilde{C}_{ambn}$ are the Weyl tensor and its dual, and $\tilde{\eta}^a$ is the timelike Killing vector field $(\equiv(\partial/\partial t)^a)$ in the NUT space-time. In an orthonormal frame

$$\omega^0 = U^{1/2}(dt + A\,d\phi),$$
$$\omega^1 = U^{-1/2}dr,$$
$$\omega^2 = (r^2 + l^2)d\theta,$$
$$\omega^3 = (r^2 + l^2)^{1/2}\sin\theta\, d\phi,$$

we find, for large r,

$$\tilde{E}_{11} = -2m/r^3, \quad \tilde{E}_{22} = \tilde{E}_{33} = m/r^3,$$
$$\tilde{B}_{11} = 2l/r^3, \quad \tilde{B}_{22} = \tilde{B}_{33} = -l/r^3, \tag{39}$$

and all other components zero. Next, under conformal transformation

$$g_{ab} = \Omega^2\bar{g}_{ab}, \quad \tilde{C}_{abcd} = \Omega^{-2}C_{abcd}, \quad \bar{\eta}^a = \eta^a; \text{ so}$$
$$\tilde{E}_{ab} = \Omega^{-2}C_{ambn}\eta^m\eta^n, \quad \tilde{B}_{ab} = \Omega^{-2}{}^*C_{ambn}\eta^m\eta^n.$$

Now, (by definition)

$$K^{ab} = \lim_{\Omega\to 0}\epsilon^{aumn}\eta_u\epsilon^{bvpq}\eta_v\Omega^{-1}C_{mnpq}$$
$$= \lim_{\Omega\to 0}\eta_u\eta_v 4\Omega^{-1}C^{aubv}$$
$$= \lim_{\Omega\to 0}4\Omega^{-3}(\tilde{E}_{pq}\tilde{g}^{ap}\tilde{g}^{bq}).$$

Since only the $\mathbf{n}^a\mathbf{n}^b$ part survives, we have, using $\Omega = x = 1/r$ and (39),

$$K^{ab} = \lim_{r\to\infty}(4r^3\tilde{E}_{11})\mathbf{n}^a\mathbf{n}^b = -8m\mathbf{n}^a\mathbf{n}^b.$$

Similarly

$${}^*K^{ab} = 8l\mathbf{n}^a\mathbf{n}^b.$$

Equations (25) then show that $M = m$ and $N = -$

V. CONCLUSION

To summarize, we have formulated a general notion of dual-mass in asymptotically flat space-times in which the Bondi news vanishes. Within this framework, dual-mass is expressed as a real valued linear function on the Lie algebra of infinitesimal translations. The appeal of this characterization is that it depends only on the asymptotic properties of the gravitational field and not on the interior space-time. In particular, no assumption about the existence of isometries (Killing fields) in the space-time need be made.

With the structure of $\mathscr{I}$ considered here, the restriction to zero news is in a sense forced by the field equations. Note that Eq. (4), obtained by taking the curl of the Einstein equations (1), leads to (13) involving the news. To obtain a similar equation involving a "dual news" one would have to start with the Bianchi identity. One may check that no such equation exists. This absence of dual news leaves (21) as the only viable candidate for dual-mass. Then the field equation (13) forces the news to vanish (Theorem 1). To incorporate news in a description of dual-mass, one would have to consider a different structure on $\mathscr{I}$. This generalization will be discussed in a later paper.

The central result of this paper is that space-times with dual-Bondi-mass are characterized by an asymptotic regime with a compact null boundary having the structure of a S^1 fiber bundle over S^2. The topology of such a space is that of a lens space $L(n,1)$[7], i.e., S^3 (for $n = 1$) or S^3 with certain identifications made along its S^1 (Hopf) fibers. Such space-times may be called Asymptotically-NUT if $\mathscr{I}$ has topology S^3; otherwise Asymptotically-locally-NUT. The local asymptotic structure of these space-times is the same as that of asymptotically Minkowskian space-times but they differ globally. Because of the topology of $\mathscr{I}$, the asymptotic symmetry group (which one may call the NUT BMS) is different from the BMS. For example, the BMS group has a SO(3) subgroup (although not a unique one) but the three-sphere admits an action of SU(2) and not of SO(3).

An immediate consequence of the existence of dual-Bondi-mass is that the space-time is acausal. Since the covering space of $L(n,1)$ is S^3, the acausality cannot be removed by going to a covering space. The physical relevance of such space-times is therefore obscure. Nevertheless, it is of interest to question its role in quantum gravity. For instance, consider spin two quantum fluctuations about such a space-time. If they are to preserve the asymptotic structure they must preserve the periodicity of the time coordinate. Alternatively, the three-sphere $\mathscr{I}$ may be regarded as being unstable to quantum effects. In either case, it is difficult to formulate a reasonable quantum theory of spin two fluctuations in the presence of dual-mass. On the other hand, the dual-mass vacuum solutions are stationary points of the gravitational action functional and they play an important role in Euclidean quantum gravity. How should one interpret these features of dual-mass? More generally, the notion of dual-mass can be extended to include news. Then the asymptotic quantization scheme, proposed by Ashtekar,[26] can be used to study its significance in an S-matrix description of quantum gravity. What new quantum effects emerge due to the presence of dual-mass? These issues will be investigated in a later paper.

ACKNOWLEDGMENTS

We thank Abhay Ashtekar and Bob Geroch for valuable discussions. We also thank Professor P. G. O. Freund

for conversations during the early stages of this work. Supported in part by NSF contract PHY 79-06940 A01 with the University of Maryland.

APPENDIX

We summarize here a description of dual-mass, based on known results, for stationary, vacuum solutions of Einstein's equations.

Consider a stationary space-time (M, g_{ab}, ξ^a) with the stationary Killing vector field ξ^a and let $\pi: M \to \Sigma$ denote the projection map from M to the three-dimensional manifold Σ of orbits of ξ^a. There exists on M a curlfree skew tensor

$$F_{ab} := \nabla_{[a}\lambda^{-1}\xi_{b]}$$

(where $\lambda = \xi^a\xi_a$) which satisfies

$$\mathcal{L}_\xi F_{ab} = 0 \text{ and } \xi^a F_{ab} = 0.$$

Consequently, there exists a unique curlfree skew tensor $\hat{F}_{ab}$ on Σ such that $F_{ab} = \pi^*(\hat{F}_{ab})$ (π^* is the pullback map). $\hat{F}_{ab}$ is given by

$$\hat{F}_{ab} = -\lambda^{-3/2}\hat{\omega}^m\epsilon_{mab},$$

where $\hat{\omega}^a$ is a (twist) covector on Σ whose pullback on M is $\omega_a = \epsilon_{abcd}\xi^b\nabla^c\xi^d$. ϵ_{mab} is alternating tensor on Σ defined by the metric h_{ab} induced on Σ by g_{ab}. The number

$$N = \frac{1}{8\pi}\int_{S^2}\hat{F}_{ab}dS^{ab}$$

is idependent of the choice of two-sphere and gives the dual-mass associated with the stationary Killing field ξ^a. Note that it is necessary that the second homology group of Σ be nontrivial in order to obtain a nonzero N. For example, Σ has topology $S^2 \times R$ for the NUT space-time. If the stationary space-time can be complexified such that there exists an Euclidean section, then N gives precisely the NUT charge defined by Gibbons and Hawking[27] for the gravitational instantons with symmetries.

The Hansen potentials[10] are defined by

$$\phi_M = (1/4)(\lambda^2 + \omega^2 - 1),$$

$$\phi_J = \omega/2\lambda,$$

where ω is a smooth function on Σ satisfying $\nabla_a\omega = \omega_a$ on account of vacuum Einstein's equations. If Σ is asymptotically flat,[10] then the monopole moments corresponding to these potentials, evaluated at infinity give, respectively, the mass and dual-mass (≡ angular momentum monopole).

The duality is exhibited as follows. Given a stationary, vacuum solution described by $\tau = \omega + i\lambda$, another solution $\tau' = \omega' + i\lambda'$ can be generated by the (Geroch) transformation[14]

$$\tau' = \frac{\tau\cos\theta + \sin\theta}{-\tau\sin\theta + \cos\theta}.$$

This induces the following (duality) rotation in the (ϕ_M, ϕ_J) plane:

$$\phi'_M = \phi_M\cos 2\theta + \phi_J\sin 2\theta,$$

$$\phi'_J = -\phi_M\sin 2\theta + \phi_J\cos 2\theta.$$

As an example, consider Schwarzschild space-time given by $\phi_M = (1/4\lambda)(\lambda^2 - 1)$, $\phi_J = 0$, where $\lambda = -(1 - 2M/r)$. Here, the mass monopole is nonzero ($= M$) and dual-mass vanishes. Under a rotation by $\theta = \pi/4$, one obtains a new solution with potentials $\phi'_M = 0$ and $\phi'_J = -\phi_M$. Here, the mass vanishes while the dual-mass is nonzero ($= -M$). This new solution is precisely the NUT metric with zero mass and NUT parameter, $l = -M$. We note that the NUT space-time is asymptotically flat in the sense of Hansen.[10]

[1] For a review, see P. Goddard and D. Olive, Rep. Prog. Phys. **41**, 1357 (1978).
[2] M. Demianski and E. T. Newman, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys. **14**, 653 (1966); W. B. Bonnor, Proc. Cambridge Phil. Soc. **66**, 145 (1969); J. S. Dowker, Gen. Rel. Grav. **5**, 603 (1974); R. A. Kobiske and L. Parker, Phys. Rev. D **16**, 3355 (1977).
[3] E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[4] R. Geroch, *Asymptotic Structure of Spacetime*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).
[5] First introduced by Penrose in R. Penrose, Phys. Rev. Lett. **10**, 66 (1963); Proc. R. Soc. London Ser. A **284**, 159 (1965).
[6] See, for example, B. Schmidt, M. Walker, and P. Sommers, Gen. Rel. Grav. **6**, 489 (1975); R. Penrose, in *Group Theory in Non-Linear Problems*, edited by A. O. Barut (Reidel, New York, 1974).
[7] The lens space $L(n,1)$ is defined as follows. Regard S^3 as the space of all pairs (z_1,z_2) of complex numbers z_1 and z_2 such that $|z_1|^2 + |z_2|^2 = 1$. Then the map $(z_1,z_2) \to \exp 2\pi i/n(z_1,z_2)$ defines a fixed point free action of Z_n on the three-sphere. $L(n,1) = S^3/Z_n$.
[8] For the definition of stable causality, see for example, S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, England, 1976).
[9] Here we consider the space-time with a smooth NUT metric and having topology $S^2 \times R$ as described by C. W. Misner, J. Math. Phys. **4**, 924 (1963).
[10] R. Hansen, J. Math. Phys. **15**, 46 (1974).
[11] P. Kundu, University of Chicago, preprint, 1980.
[12] P. Kundu, J. Math. Phys. **22**, 1236 (1981).
[13] The angular momentum monopole moment vanishes identically if the second homology group of Σ is trivial. This is not the case for the NUT spacetime.
[14] R. Geroch, J. Math. Phys. **12**, 918 (1971).
[15] We shall not consider a description of "dual" mass at "spatial infinity". The absence of any reasonable spacelike surfaces in the NUT space-time suggests that such a formulation would be unnatural.
[16] By a space-time we shall mean a smooth, connected, four-dimensional manifold M with a smooth, time-oriented metric of Lorentz signature.
[17] See, for example, R. Geroch and G. Horowitz, Phys. Rev. Lett. **40**, 203 (1978); A. Ashtekar, in *General Relativity and Gravitation*, Vol. 2, edited by A. Held (Plenum, New York, 1980).
[18] Our convention for the curvature tensors are

$$[\nabla_a\nabla_b - \nabla_b\nabla_a]v_c = R_{abc}{}^d v_d; R_{ab} = R_{amb}{}^m; R = R_{ab}g^{ab}.$$

[19] Strictly speaking, the diffeomorphism preserves the conformally invariant tensor field $\Gamma_{ab}{}^{cd} = \mathbf{g}_{ab}\mathbf{n}^c\mathbf{n}^d$.
[20] By a cross section we mean a smooth mapping K from the base space S^2 to $\mathcal{S}$ such that $\pi \circ K$ is the identity on S^2. Since we are contemplating arbitrary bundles over S^2, a cross section may not always exist. We shall see, however, that in that case one can effectively express the relevant integrals as integrals over the base space.
[21] Alternatively, one can ask for a second rank skew tensor F_{ab} and set

$$^*p^a = \epsilon^{abc}F_{bc}.$$

[22] For the definition of a principal fiber bundle see, for example, S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. I (Interscience, New York, 1963).
[23] See, for example, J. Friedman, and R. Sorkin, Phys. Rev. D **20**, 2511 (1979).
[24] We adapt here a construction given in Ref. 23.
[25] E. T. Newman and K. P. Tod, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[26] A. Ashtekar, Phys. Rev. Lett. **46**, 573 (1981).
[27] G. Gibbons and S. Hawking, Commun. Math. Phys. **66**,291 (1979).

The classification of the Ricci and Plebański tensors in general relativity using Newman–Penrose formalism

C. B. G. McIntosh
Mathematics Department, Monash University, Clayton, Victoria, 3168, Australia

J. M. Foyster
Victorian Institute of Secondary Education, 582 St Kilda Road, Melbourne, Victoria, 3004, Australia

A. W. -C. Lun
Department of Mathematical Studies, Hong Kong Polytechnic, Kowloon, Hong Kong

(Received 15 July 1980; accepted for publication 5 September 1980)

A list is given of a canonical set of the Newman–Penrose quantities Φ_{AB}, the tetrad components of the trace-free Ricci tensor, for each Plebański class according to Plebański's classification of this tensor. This comparative list can easily be extended to cover the classification in tetrad language of any second-order, trace-free, symmetric tensor in a space-time. A fourth-order tensor which is the product of two such tensors was defined by Plebański and used in his classification. This has the same symmetries as the Weyl tensor. The Petrov classification of this tensor, here called the Plebański tensor, is discussed along with the classification of the Ricci tensor. The use of the Plebański tensor in a couple of areas of general relativity is also briefly discussed.

PACS numbers: 04.20.Cv, 04.02.Jb, 02.40.Ky,

1. INTRODUCTION

Many authors have discussed the algebraic classification of the Ricci tensor in general relativity; see, for example, Refs. 1–9. Most of these authors have been concerned with obtaining canonical classifications of the trace-free Ricci tensor with components

$$S_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{4} g_{\mu\nu} R \tag{1.1}$$

(or, equivalently, any symmetric, rank-2, trace-free tensor in space-time). The canonical classification scheme of Plebański[2] will be used as the basic scheme in this paper.

There are two purposes in writing another paper in this area. The first one is to give a canonical list of the trace-free Ricci tensor components for each Plebański class in terms of the tetrad formalism of Newman and Penrose[10] (hereafter called NP), i.e., in terms of the NP quantities Φ_{AB} $(A = 0\text{–}2)$. Such a list does not seem to have been published before, although Hall[5] has an equivalent list in which he compares the Plebański scheme with a classification of the tetrad components of the trace-free Ricci tensor for a null basis. His scheme is thus equivalent to, but notationally different from, one using NP language. Ludwig and Scanlon[4] also list canonical types, but in a comparison with their classification scheme, which is very different from that of Plebański.

The second purpose is to emphasize a second-order tensor which Plebański makes some use of in his classification scheme but which is hardly used elsewhere. We call this the Plebański tensor. It is constructed from the product of two trace-free Ricci tensors and has the same symmetries as the Weyl tensor. This last fact is very important because it means that the Plebański tensor can be classified in the same way as the Weyl tensor; i.e., according to the usual Petrov (or Penrose–Petrov) scheme. Its use in a couple of areas of general relativity will be referred to later.

2. THE PLEBAŃSKI TENSOR

From the trace-free Ricci tensor S with components as in (1.1), form the fourth-order tensor with components

$$P^{\alpha\beta}{}_{\gamma\sigma} = S^{[\alpha}{}_{[\gamma} S^{\beta]}{}_{\sigma]} + \delta^{[\alpha}{}_{[\gamma} S_{\sigma]\lambda} S^{\beta]\lambda} - \tfrac{1}{6}\delta^{[\alpha}{}_{[\gamma}\delta^{\beta]}{}_{\sigma]} S_{\mu\nu} S^{\mu\nu}. \tag{2.1}$$

This is trace-free, i.e.,

$$P^{\alpha\beta}{}_{\alpha\sigma} = 0. \tag{2.2}$$

and has the same symmetries as the Weyl tensor. We shall call it the Plebański tensor. Its spinor equivalent is

$$\chi_{ABCD} = \tfrac{1}{4}\Phi_{(AB}{}^{E'F'}\Phi_{CD)E'F'}, \tag{2.3}$$

where the $\Phi_{ABC'D'}$ are the spinor equivalents of the $S_{\mu\nu}$ as defined by NP. χ_{ABCD} is the same as Plebański's V_{ABCD} [Ref. 2, Eq. (6.2)]. Then, by using the definitions $\Phi_{00} = \Phi_{000'0'}$, etc. as given by NP, (2.3) gives

$$\chi_0 \equiv \chi_{0000} = \tfrac{1}{2}(\Phi_{00}\Phi_{02} - \Phi_{01}{}^2), \tag{2.4a}$$

$$\chi_1 \equiv \chi_{0001} = \tfrac{1}{4}(\Phi_{00}\Phi_{12} + \Phi_{10}\Phi_{02} - 2\Phi_{01}\Phi_{11}), \tag{2.4b}$$

$$\chi_2 \equiv \chi_{0011} = \tfrac{1}{12}(\Phi_{00}\Phi_{22} - 4\Phi_{11}{}^2 + \Phi_{02}\Phi_{20} + 4\Phi_{10}\Phi_{12} - 2\Phi_{21}\Phi_{01}), \tag{2.4c}$$

$$\chi_3 \equiv \chi_{0111} = \tfrac{1}{4}(\Phi_{22}\Phi_{10} + \Phi_{12}\Phi_{20} - 2\Phi_{21}\Phi_{11}), \tag{2.4d}$$

$$\chi_4 \equiv \chi_{1111} = \tfrac{1}{2}(\Phi_{22}\Phi_{20} - \Phi_{21}{}^2). \tag{2.4e}$$

These χ_A now transform under the action of the homogeneous Lorentz group at a point in the same way as the tetrad components of the Weyl tensor [the NP Ψ_A $(A = 0\text{–}4)$]. For example, where $\mathbf{l}$, $\mathbf{n}$, $\mathbf{m}$, $\bar{\mathbf{m}}$ are the four basis null vectors used by NP to span a given space-time (with $\mathbf{l}$ and $\mathbf{n}$ real and $\mathbf{m}$ complex), under the two-parameter group of null rotations leaving $\mathbf{l}$ fixed:

$$\begin{aligned} \tilde{l}_\mu &= l, \\ \tilde{n}_\mu &= z\bar{z}l_\mu + n_\mu + z\bar{m}_\mu + \bar{z}m_\mu, \\ \tilde{m}_\mu &= \bar{z}l_\mu + m_\mu, \end{aligned} \tag{2.5}$$

where z is an arbitrary complex scalar, the χ_A transform in an equivalent way to the Ψ_A:

 0022-2488/81/112620-04$01.00

TABLE. I Comparison of different classification schemes of the trace-free Ricci tensor, S. The table gives a canonical set of nonzero NP quantities Φ_{AB} which are the tetrad components of S for each Plebański class in the classification scheme in Ref. 2. Also given are corresponding Segré characteristics and the Petrov classification of the Plebański tensor constructed from S and defined in the text.

Plebański class	Segré characteristic	Canonical set of nonzero NP Φ_{AB}	Petrov classification of Plebański tensor	
$[T\text{-}S_1\text{-}S_2\text{-}S_3]_4$	[1111]	$\Phi_{00}=\Phi_{22},\Phi_{11},\Phi_{02}=\Phi_{20}$	I	I_a
$[2T\text{-}S_1\text{-}S_2]_3$	[(11)11]	$\Phi_{11},\Phi_{02}=\Phi_{20}$	D	D_{a1}
$[T\text{-}2S_1\text{-}S_2]_3$	[1(11)1]	$\Phi_{00}=\Phi_{22},\Phi_{11}$	D	D_{a2}
$[2T\text{-}2S]_2$	[(11)(11)]	Φ_{11}	D	D_{a3}
$[3T\text{-}S]_2$	[(111)1]	$2\Phi_{11}=\Phi_{02}$	O	O_{a1}
$[T\text{-}3S]_2$	[1(111)]	$\Phi_{00}=\Phi_{22}=2\Phi_{11}$	O	O_{a2}
$[4T]_1$	[(1111)]		O	O_{a3}
$[Z\text{-}\bar{Z}\text{-}S_1\text{-}S_2]_4$	$[Z\bar{Z}11]$	$\Phi_{00}=-\Phi_{22},\Phi_{11},\Phi_{02}=\Phi_{20}$	I	I_b
$[Z\text{-}\bar{Z}\text{-}2S]_3$	$[Z\bar{Z}(11)]$	$\Phi_{00}=-\Phi_{22},\Phi_{11}$	D	D_b
$[2N\text{-}S_1\text{-}S_2]_4$	[211]	$\Phi_{11},\Phi_{22},\Phi_{02}=\Phi_{20}$	II	II
$[2N\text{-}2S]_{(2-1)}$	[2(11)]	Φ_{11},Φ_{22}	D	D_2
$[3N\text{-}S]_3$	[(12)1]	$2\Phi_{11}=\Phi_{02},\Phi_{22}$	N	N_2
$[4N]_2$	[(112)]	Φ_{22}	O	O_2
$[3N\text{-}S]_4$	[31]	$2\Phi_{11}=\Phi_{02},\Phi_{01}\neq\Phi_{10}$	III	III
$[4N]_3$	[(13)]	Φ_{01}	N	N_3

$$\tilde{\chi}_0=\chi_0,$$
$$\tilde{\chi}_1=z\chi_0+\chi_1,$$
$$\tilde{\chi}_2=z^2\chi_0+2z\chi_1+\chi_2,$$
$$\tilde{\chi}_3=z^3\chi_0+3z^2\chi_1+3z\chi_2+\chi_3,$$
$$\tilde{\chi}_4=z^4\chi_0+4z^3\chi_1+6z^2\chi_2+4z\chi_3+\chi_4. \tag{2.6}$$

For other null notations, such as boosts in the **l**-**n** plane and rotations in the **m**-$\bar{\mathbf{m}}$ plane, the χ's transform as do the Ψ's and direct replacements of the Ψ's can be made in tables of transformations of the Ψ's, e.g., in the Tables E-3, 4, 5 in Ref. 11. These transformations can be used to find canonical forms of the χ's as for the Ψ's; e.g., χ is of Petrov (or Penrose–Petrov) type D iff rotations of the null vectors can be made such that χ_2 is the only nonzero χ_A.

The Φ_{AB}, of course, also change under the action of this group. A study of the changes of the quantity $\Phi_{22}\Phi_{20}-\Phi_{21}^2$ as in (2.4e) under (2.5) will lead to the change of χ_4 as in (2.6) and similarly with the other combinations. It was noticing this behavior that led one of us (A.W.-C.L.) to construct (2.4). MacCallum pointed out to one of us in a private communication that (2.4) was equivalent to (2.3) (and hence equivalent to Plebański's V_{ABCD}).

The relationship between $P_{\alpha\beta\gamma\delta}$ and the χ_A is

$$P_{\alpha\beta\gamma\delta}=16(Q_{\alpha\beta\gamma\delta}+\bar{Q}_{\alpha\beta\gamma\delta}), \tag{2.7}$$
$$Q_{\alpha\beta\gamma\delta}=\chi_0 Z^1{}_{\alpha\beta}Z^1{}_{\gamma\delta}+\chi_1(Z^1{}_{\alpha\beta}Z^2{}_{\gamma\delta}+Z^2{}_{\alpha\beta}Z^1{}_{\gamma\delta})$$
$$+\chi_2(Z^2{}_{\alpha\beta}Z^2{}_{\gamma\delta}+Z^1{}_{\alpha\beta}Z^3{}_{\gamma\delta}+Z^3{}_{\alpha\beta}Z^1{}_{\gamma\delta})$$
$$-\chi_3(Z^3{}_{\alpha\beta}Z^2{}_{\gamma\delta}+Z^2{}_{\alpha\beta}Z^3{}_{\gamma\delta})+\chi_4 Z^3{}_{\alpha\beta}Z^3{}_{\gamma\delta},$$

where the Z^i are defined as[12]

$$Z^1\equiv Z^1{}_{\alpha\beta}dx^\alpha\wedge dx^\beta=\bar{\mathbf{m}}\wedge\mathbf{n},$$
$$Z^2\equiv Z^2{}_{\alpha\beta}dx^\alpha\wedge dx^\beta=\mathbf{n}\wedge\mathbf{l}-\bar{\mathbf{m}}\wedge\mathbf{m}, \tag{2.8}$$
$$Z^3\equiv Z^3{}_{\alpha\beta}dx^\alpha\wedge dx^\beta=\mathbf{l}\wedge\mathbf{m}.$$

3. CLASSIFICATION OF THE RICCI AND PLEBAŃSKI TENSORS

We present in Table I without derivation a list of a canonical set of nonzero NP quantities Φ_{AB} for each Plebański type of the trace-free Ricci tensor. Also in this table are corresponding lists of basic Petrov types [I (a or b), II, and III], Segré characteristics, and the Petrov classification of the corresponding Plebański tensor (2.1) [or spinor equivalent (2.3)]. A diagramatic representation of the degeneracies in the four main classes in Table I is given in Fig. 1.

A number of comments need to be made about the table. The entries of the nonzero Φ_{AB} are modulo a change **l**↔**n**. For example, in $[4N]_2$ the entry of Φ_{22} is equivalent to an entry of Φ_{00}. A given set of Φ_{AB} from some metric can be

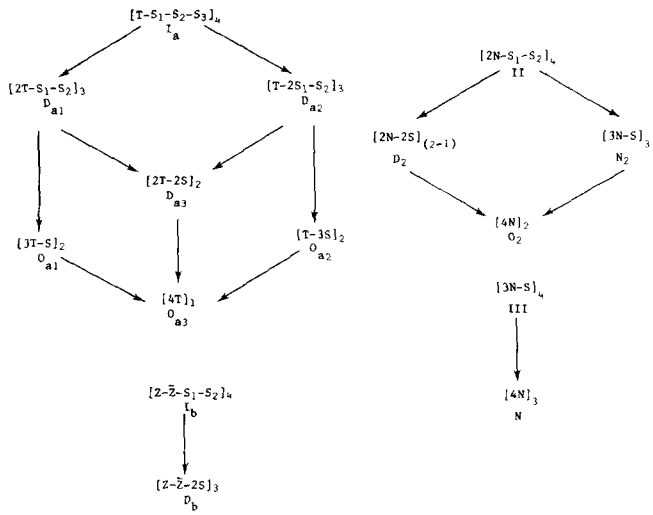


FIG. 1. Diagramatic representation of the degeneracies in the four main classes in the classification scheme in Table I. The Plebański class is given together with the Petrov classification of the Plebański tensor for each class.

put into the form of one of the cases listed in the table by means of the rotations of the base vectors of the type (2.5) or others discussed, for example in Ref. 11 (Tables E-3,4,5). The notation of the Plebański classes follows that in Ref. 2, e.g., $[T\text{-}S_1\text{-}S_2\text{-}S_3]_4$ becomes $[T\text{-}S_1\text{-}S_2\text{-}S_3]_{[1\text{-}1\text{-}1\text{-}1]}$ in Ref. 3. The classification of the tensor depends on the eigenvectors and character of the eigenspaces of that tensor, The symbol T, N, or S is used if the eigenspace of an eigenvalue contains a timelike, no timelike but a null eigenvector, or only spacelike eigenvectors, respectively, Z and $\bar{Z}$ refer to the eigenspace of a pair of complex conjugate eigenvalues. The numbers in round brackets specify the indices of nilpotency in the order in which the eigenvalues appear in the square bracket. In all but one case this is replaced by a single figure which is the order of the minimal equation. This together with the numbers of the kinds of eigenvectors determines uniquely the distribution of the nilpotent indices. (This would not be the case if $[2N\text{-}2S]_3$ was used). The Segré characteristic list should be self-evident. It is described by Hall[5] and used by a number of authors.

The classification of the Plebański tensor formed from the nonzero Φ_{AB} in each case is equivalent to the standard Penrose–Petrov classification of the Weyl tensor. Two columns are used in the table; one with the symbols I, II, III, D, N, and O which is all that could be used if the χ_A alone were given, and I_a,D_{a1} ... which can be used to distinguish between the types if more information, such as all the Φ_{AB}, is given. Note that the types D, N, and O here do not overlap with the types D, N and O as given by Goenner and Stachel[13] as they are classifying different quantities.

The classification used in Table I can be used to classify any trace-free symmetric rank-two tensor in a space-time; not just the Ricci one. It, is in fact, used in this way in a related paper.[14]

Other types of classification schemes as proposed by various authors could have been added to Table I, but were not in order to keep the table from becoming too cluttered.

4. CLASSIFICATION OF MORE GENERAL TENSORS

The classification of a second-order trace-free symmetric tensor such as the trace-free Ricci tensor **S** is straightforward and discussed in many places. Here the canonical forms of the NP quantities Φ_{AB} equivalent to the $S_{\mu\nu}$ are given for each Plebański class. However an entirely equivalent list holds for the tetrad components of any second-order symmetric tensor $X_{\mu\nu}$. Then X_{AB} ($A = 1$–4) can be defined as

$$\begin{aligned} X_{11} &= X_{\mu\nu}l^\mu l^\nu, & X_{22} &= X_{\mu\nu}n^\mu n^\mu, \\ X_{33} &= X_{\mu\nu}m^\mu m^\nu, & X_{44} &= X_{\mu\nu}\bar{m}^\nu\bar{m}^\nu, \\ X_{13} &= X_{\mu\nu}l^\mu m^\nu, & X_{14} &= X_{\mu\nu}l^\mu\bar{m}^\nu, \\ X_{23} &= X_{\mu\nu}n^\mu m^\nu, & X_{24} &= X_{\mu\nu}n^\mu\bar{m}^\nu, \\ X_{1234} &\equiv X_{12} + X_{34} = X_{\mu\nu}(l^\mu n^\nu + m^\mu\bar{m}^\nu). \end{aligned} \tag{4.1}$$

The trace of **X** is

$$X = X_{\mu\nu}g^{\mu\nu} = 2(X_{12} - X_{34}) = 2X_{\mu\nu}(l^\mu n^\nu - m^\mu\bar{m}^\nu). \tag{4.2}$$

If $X_{\mu\nu} = S_{\mu\nu}$, then

$$\begin{aligned} \Phi_{00} &= -\tfrac{1}{2}X_{11}, & \Phi_{22} &= -\tfrac{1}{2}X_{22}, \\ \Phi_{01} &= -\tfrac{1}{2}X_{13}, & \Phi_{02} &= -\tfrac{1}{2}X_{33}, \\ \Phi_{12} &= -\tfrac{1}{2}X_{23}, & \Phi_{11} &= -\tfrac{1}{4}X_{1234}, \end{aligned} \tag{4.3}$$

together with $X_{.4} = \bar{X}_{.3}$. The pattern of the indices A and B in the Φ_{AB} comes from that in their spinor equivalents: $\Phi_{01} = \Phi_{000'1'}$ etc., while in the X_{AB} it comes from vectors with which the inner products of the $X_{\mu\nu}$ are taken to form the X_{AB}.

In Table I, the column which contains the canonical set of Φ_{AB} could be replaced easily by one containing a canonical set of X_{AB} by using (4.3).

5. DISCUSSION

As stated in the Introduction, one purpose of this paper is to highlight the tensor **P** defined by (2.1) and which we have called the Plebański tensor. This tensor is little used in the literature but it seems to us that it should be used more often in the classification of second-order tensors. One of its main properties is that is has the same symmetries as the Weyl tensor.

McIntosh and Halford[14] (see also McIntosh[15] and McIntosh and van Leeuwen[16]) discuss solutions $x_{\mu\nu}$ of the equation

$$x_{\mu\nu}R^{\mu}{}_{\lambda\alpha\beta} + x_{\mu\lambda}R^{\mu}{}_{\nu\alpha\beta} = 0, \tag{5.1}$$

where $x_{\mu\nu}$ in not proportional to the $g_{\mu\nu}$, the components of the metric tensor g from which the $R^{\mu}{}_{\nu\alpha\beta}$ are constructed. This equation arises in particular in the study of curvature collineations and of the holonomy group in general relativity. It is shown there that the Petrov type of the Plebański tensor formed from such a nontrivial solution x of (5.1) is the same as the Petrov type of the Weyl tensor of the metric tensor for which such a solution exists. It is also almost always the same as the Petrov type of the Plebański tensor formed from the Ricci tensor for such a **g**. Likewise in a paper by MacCallum[17] on locally isotropic spacetimes with non-null homogeneous hypersurfaes, an examination of the Segré types (or, equivalently, Plebański classes) of the Ricci tensor of all those metrics as listed by MacCallum shows that once again the Petrov types of the Weyl and Plebański tensors are always the same. In both these cases it is not just that the Petrov types are the same but also that the repeated principal null directions of the Weyl and Plebański tensors align. MacCallum says in a private communication that these results are expected in his case because of the restrictions on the Weyl tensor and, in the same way, on the Plebański tensor of space-times which can admit isotropy groups (see, e.g., Theorem 2-2.6 by Ehlers and Kundt in Ref. 18). These properties will be discussed by one of us (C.McI) elsewhere.

Note added in proof: The components (2.4) of the Plebański tensor were published by Collinson and Shaw, Int. J. Theor. Phys. **6**, 347 (1972), in a discussion on the Rainich conditions for neutrino fields. They called the tensor (2.1) the Weyl Square. Another algebraic classification of the Ricci tensor, or, with the Einstein field equations holding, the matter tensor, in terms of NP quantities was published by Dozmorov, Sov. Phys. J. **16**, 1708 (1973)—English translation. This account, however, does not take into account all the degeneracies of the various classes in any sufficient way.

ACKNOWLEDGMENTS

We wish to thank Malcolm MacCallum and Graham Hall for useful correspondence in this area and for sending us preprints of their work.

[1]R. V. Churchill, Trans. Am. Math. Soc. **34**, 784 (1932).
[2]J. Plebański, Acta Phys. Polon. **26**, 963 (1964).
[3]J. Plebański and J. Stachel, J. Math. Phys. **9**, 269 (1968).
[4]G. Ludwig and G. Scanlon, Commun. Math. Phys. **20**, 291 (1971).
[5]G. S. Hall, J. Phys. A **9**, 541 (1976).
[6]G. S. Hall, Z. Naturforsch **33a**, 559 (1978).
[7]G. S. Hall, Proceedings of the Stefan Banach International Mathematical Centre Semester on Differential Geometry, Warsaw (to appear).
[8]W. J. Cormack and G. S. Hall, J. Phys. A **12**, 55 (1979).
[9]W. J. Cormack and G. S. Hall, Int. J. Theor. Phys. **18**, 279 (1979).
[10]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[11]M. Carmeli, *Group Theory and Relativity* (McGraw–Hill, New York, 1977).
[12]We follow here the conventions used by R. Debever, R. G. McLenaghan, and N. Tariq, Gen. Rel. Grav. **10**, 853 (1979).
[13]H. Goenner and J. Stachel, J. Math. Phys. **11**, 3358 (1970).
[14]C. B. G. McIntosh and W. D. Halford, "The Riemann tensor, the metric tensor and curvature collineations in general relativity," J. Math. Phys. (to appear).
[15]C. B. G. McIntosh, "Curvature collineations in general relativity are almost always special conformal motions" (Preprint).
[16]C. B. G. McIntosh and E. H. van Leeuwen, "Spacetimes admitting a vector field whose inner product with the Riemann tensor is zero," J. Math. Phys. (to appear).
[17]M. A. H. MacCallum, "Locally isotropic spacetimes with non-null homogeneous hypersurfaces" (Preprint).
[18]J. Ehlers and W. Kundt, of *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 2.

Bäcklund transformations in the Hauser–Ernst formalism for stationary axisymmetric spacetimes [a)]

Christopher M. Cosgrove [b)]
W. K. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125

(Received 30 March 1981; accepted for publication 26 June 1981)

It is shown that Harrison's Bäcklund transformation for the Ernst equation of general relativity is a two-parameter subset (not subgroup) of the infinite-dimensional Geroch group **K**. We exhibit the specific matrix $u(t)$ appearing in the Hauser–Ernst representation of **K** for vacuum spacetimes which gives the Harrison transformation. Harrison transformations are found to be associated with quadratic branch points of $u(t)$ in the complex t plane. The coalescence of two such branch points to form a simple pole exhibits in a simple way the known factorization of the (null generalized) HKX transformation into two Harrison transformations. We also show how finite (i.e., already exponentiated) transformations in the **B** group and nonnull groups of Kinnersley and Chitre can be constructed out of Harrison and/or HKX transformations. Similar considerations can be applied to electrovac spacetimes to provide hitherto unknown Bäcklund transformations. As an example, we construct a six-parameter transformation which reduces to the double Harrison transformation when restricted to vacuum and which generates Kerr–Newman–NUT space from flat space.

PACS numbers: 04.20.Cv, 04.20.Jb, 04.40.c

1. INTRODUCTION

Since the pioneering work of Geroch,[1,2] it has been known that the partial differential equations governing the metric of the stationary axisymmetric vacuum gravitational field admit an infinite-dimensional internal symmetry group of transformations. This large internal symmetry group (the Geroch group **K**) has encouraged many authors to hope that the complete class of solutions could some day be generated systematically from a particular solution, such as flat space, or from the important subclass of Weyl static solutions. In fact, a number of special transformations, some contained in **K**[3,4] and some Bäcklund transformations known or presumed to be outside **K**,[5–11] are now known which, when iterated, generate asymptotically flat solutions with an arbitrarily large number of parameters. A detailed study of the mathematical interrelationships between these various transformations has been undertaken by the author.[12]

A workable and very fruitful representation in terms of infinitesimal generators for the Geroch group **K** and its electrovac extension **K'** has been provided by Kinnersley and Chitre[13–16] (KC). Possibly, the most important discoveries arising from their formalism are the **B** group[16] which, among other things, generates the Kerr solution[17] from Schwarzschild, and the HKX transformation.[3] More recently, Hauser and Ernst[18–20] (HE) have deduced (initially from the KC formalism and later by a direct method) a qualitatively different representation which exploits the theory of functions of a complex variable. The HE formalism has the distinct advantage that elements of **K** and **K'** appear already exponentiated and each may be specified unambiguously by a matrix function $u(t)$ of a complex variable t satisfying certain conditions, a closed contour L in the complex t plane, and a choice of gauge for the $F(t)$ matrix potential[21] of the solution to be transformed. Composition of several transformations is conveniently represented by multiplication of corresponding $u(t)$ matrices. The generation of new solutions from old in their formalism can be accomplished by either solving a linear integral equation[18] or a homogeneous Hilbert problem.[19] A remarkable result of Hauser and Ernst's work is a formula for all $u(t)$ matrices which transform a given initial solution into a given final solution.[20] This provides a quantitative settlement of one form of a well-known conjecture of Geroch.[22]

The present paper addresses certain problems for which the Hauser–Ernst formalism is particularly well suited. First, we study two large classes of $u(t)$ matrices for which the homogeneous Hilbert problem (HHP) can be solved by elementary methods. It is already known[18] that when $u(t)$ has only simple poles in L_+ (the interior of the contour L), the solution of the corresponding HHP is a product of null generalized HKX transformations[3,4,12] and nonsimple poles correspond to confluent forms such as the rank-N transformation.[3]

On the other hand, for certain classes of $u(t)$ matrices for vacuum spacetimes which have quadratic branch points and branch cuts in L_+, we find that the solutions of the HHP's are products of Harrison's Bäcklund transformations.[8] In particular, the Harrison transformation, which was discovered by methods inspired by soliton theory and quite remote from the Geroch group, is here shown to be a two-parameter subset (not a subgroup) of **K**. This result should not be entirely unexpected according to a suggestion of Kinnersley[23] and the fact that arguments given in Ref. 12 to rule out certain other Bäcklund transformations being in **K**[24] were inconclusive for the Harrison transformation.

Conversely, the formula for the Harrison transform of the $F(t)$ potential given by Eqs. (4.44a) and (4.44b) of Ref. 12 can be substituted directly into the HE formula for $u(t)$ in terms of the initial and final solutions[20] [see Eq. (2.25) below]

[a)]Supported in part by the National Science Foundation (AST79-22012).
[b)]Richard Chace Tolman Research Fellow.

 0022-2488/81/112624-16$01.00

to give, explicitly,

$$u(t)=\left(1-\frac{s}{t}\right)^{-1/2}\begin{pmatrix}1 & -cst^{-1}\\ -c^{-1} & 1\end{pmatrix}, \tag{1.1}$$

s and c real constants. The product of several Harrison transformations is represented by the product of corresponding $u(t)$ matrices [the individual s and c parameters of such a product may be complex such that the product $u(t)$ is real for real t]. In Ref. 12, we proved that the product of two Harrison transformations with same s parameters is a null generalized HKX transformation by directly composing the transformation laws for $F(t)$. In the HE formalism, the proof of this theorem reduces to matrix multiplication.

As yet, finite algebraic transformations have not been written down for the **B** group[16] whose generators are[25]

$$\gamma_{11}^{(k+1)}+\gamma_{22}^{(k-1)},\quad k=0,1,2,..., \tag{1.2}$$

and the nonnull KC group, whose generators are[25]

$$q^{XY}\gamma_{XY}^{(k)},\quad k=0,1,2,...,\ q^{XY}q_{XY}\neq 0, \tag{1.3}$$

q^{XY} being a symmetric constant SL(2, R) tensor, independent of k, which includes the nonnull HKX transformations.[12] We find that, although we cannot exponentiate the infinitesimal transformations, we can express a large number of finite transformations in these groups as products of Harrison and/or HKX transformations.

Finally, in Sec. 5, we attempt to generalize the Harrison transformation to electrovac spacetimes by considering 3×3 $u(t)$ matrices with *cubic* branch points. Surprisingly, we manage to find a Bäcklund transformation with six parameters which reduces to the *double* Harrison transformation with two complex conjugate s parameters when restricted to vacuum, while the method fails to yield an electrovac version of the *single* Harrison transformation. The new transformation maps flat space to Kerr–Newman–NUT space, which is a satisfying result. However, in the conclusion, we present an argument based on analogy with vacuum that the single Harrison transformation should exist in the electrovac case and have perhaps four parameters. A possible reason for the failure of the HE formalism to account for the electrovac analog of the single Harrison transformation is that the latter is not expected to preserve the reality of the metric and electromagnetic potentials and so an extension to two complex dimensions is indicated.

2. THE HOMOGENEOUS HILBERT PROBLEM OF HAUSER AND ERNST

In this section, we include enough details on the $F(t)$ potential, the matrix representation $u(t)$ for **K**, and the homogeneous Hilbert problem (HHP) for later use. We wish to follow the SL(2) tensor notation of Kinnersley[13] (see also Appendix A of Ref. 12) in which we would identify $F(t)=F_{AB}(t)$, $u(t)=u^{A}{}_{B}(t)$, $A, B=1, 2$.[26] The electrovac case will be postponed till Sec. 5.

The metric of stationary axisymmetric spacetime can be written

$$ds^2=f_{AB}dx^A dx^B-f^{-1}e^{2\gamma}(d\rho^2+dz^2), \tag{2.1}$$

where $f_{AB}, f=f_{11}$, and γ are functions of cylindrical coordinates ρ and z only; x^1 is time, x^2 is azimuthal angle. We use the parametrization

$$f_{11}=f,\quad f_{12}=f_{21}=-f\omega,\quad f_{22}=f\omega^2-\rho^2 f^{-1}. \tag{2.2}$$

As is well known, the vacuum field equations are the integrability conditions for further potentials. The Ernst potential $\mathscr{E}$ and its tensor generalization H_{AB} are defined by[27]

$$\mathscr{E}=H_{11}=f+i\psi, \tag{2.3a}$$

$$\nabla\psi=-\rho^{-1}f^2\tilde{\nabla}\omega, \tag{2.3b}$$

$$H_{AB}=f_{AB}+i\psi_{AB}, \tag{2.4a}$$

$$\nabla\psi_{AB}=-\rho^{-1}f_A{}^X\tilde{\nabla}f_{XB}, \tag{2.4b}$$

and satisfy

$$f\nabla_3{}^2\mathscr{E}=\nabla\mathscr{E}\cdot\nabla\mathscr{E}, \tag{2.5a}$$

$$\nabla H_{AB}=-i\rho^{-1}f_A{}^X\tilde{\nabla}H_{XB}, \tag{2.5b}$$

$$H_{AB}-H_{BA}=2iz\epsilon_{AB}, \tag{2.6a}$$

$$f_{XA}f^X{}_B=-\rho^2\epsilon_{AB}, \tag{2.6b}$$

where $\nabla=(\partial/\partial\rho, \partial/\partial z)$, $\tilde{\nabla}=(\partial/\partial z, -\partial/\partial\rho)$, $\nabla_3{}^2=\partial^2/\partial\rho^2+\partial^2/\partial z^2+\rho^{-1}\partial/\partial\rho$. An asterisk (*) will denote complex conjugation, e.g., $H^*{}_{AB}=f_{AB}-i\psi_{AB}$.

When H_{AB} is known, a potential $F_{AB}(t)$ $[=F_{AB}(\rho, z, t)]$ which is a function of a complex variable t, as well as ρ and z, can be constructed from the linear differential equation,[28]

$$\nabla F_{AB}(t)=itS^{-2}(t)[(1-2tz)\nabla H_{AX}-2t\rho\tilde{\nabla}H_{AX}]F^X{}_B(t), \tag{2.7}$$

where

$$S(t)=[(1-2tz)^2+4t^2\rho^2]^{1/2},\quad S(0)=1, \tag{2.8}$$

subject to

$$F_{AB}(0)=i\epsilon_{AB}, \tag{2.9a}$$

$$\dot{F}_{AB}(0)=H_{AB}, \tag{2.9b}$$

$\dot{F}(t)=\partial F(t)/\partial t$. Two important first integrals of Eq. (2.7) are

$$F_{XA}(t)F^X{}_B(t)=-S^{-1}(t)\epsilon_{AB}\quad\text{or}\quad \det F(t)=-S^{-1}(t), \tag{2.10}$$

$$S(t)F^*{}_{AB}(t)=2itf_{AX}F^X{}_B(t)-(1-2tz)F_{AB}(t), \tag{2.11}$$

where $F^*{}_{AB}(t)$ is to be understood as the complex conjugate of $F_{AB}(t^*)$.

The differential equation (2.7) and initial conditions (2.9a) and (2.9b) define $F(t)$ up to a gauge change,

$$F_{AB}(t)\to F_A{}^X(t)g_{XB}(t)\quad\text{or}\quad F(t)\to F(t)g(t), \tag{2.12}$$

where $g(t)=-g^A{}_B(t)$ depends on t only. Equations (2.9)–(2.11) imply

$$g_{AB}(0)=\epsilon_{AB}\ \text{ or }\ g(0)=I,\quad g^*(t)=g(t),\quad \det g(t)=1, \tag{2.13}$$

I being the unit matrix. This gauge freedom may be used to minimize the singularities of $F(t)$ in the complex t plane.[19,20] In all cases, $F(t)$ is analytic at and in a neighborhood of $t=0$. Also, it is always possible to choose gauge so that $F(t)$ is analytic at and near $t=\infty$. Hauser and Ernst (HE) have imposed the slightly stronger condition

$$F(t)\begin{pmatrix}1 & 0\\ 0 & t\end{pmatrix}\ \text{analytic at } t=\infty\quad(\text{"HE gauge"}), \tag{2.14}$$

which can be brought about by a translation, $\omega \to \omega +$ constant. Later, in SL(2)-covariant applications, we shall wish to permit an arbitrary additive constant in ω, so we relax this condition to

$$F(t) \text{ analytic at } t = \infty \quad (\text{"modified HE gauge"}). \tag{2.14'}$$

Next, HE[19,20] have shown that, in a (ρ, z) domain covering at least one point of the z axis in which $\mathscr{E}$ is analytic[29] and $f \neq 0$, gauge can be chosen so that the only singularities of $F(t)$ in the t plane (including $t = \infty$) are quadratic branch points with index $-\frac{1}{2}$ at

$$t_+ = \frac{z + i\rho}{2r^2} \text{ and } t_- = \frac{z - i\rho}{2r^2} \quad (r^2 = \rho^2 + z^2), \tag{2.15}$$

i.e., the zeros of $S(t)$, and the cut is a finite arc from $t = t_+$ to $t = t_-$ (not through $t = 0$). On the z axis, where $\rho = 0$, the branch points and cut degenerate to a simple pole at $t = (2z)^{-1}$. This very special gauge will be called "special HE gauge" if condition (2.14) is also imposed; otherwise [condition (2.14')] we shall call it "modified special HE gauge."

If $F(t)$ is in (modified) special HE gauge, then a change of gauge will automatically introduce (ρ, z)-independent singularities in the finite t plane and/or at $t = \infty$. Furthermore, analytic continuation of $F(t)$ across the cut will reveal, in general, (ρ, z)-independent singularities of various types (including at $t = 0$) on the second Riemann sheet. If $\mathscr{E}$ is not analytic anywhere on the z axis, then special HE gauge may not exist. Nevertheless, in this case, HE[19] have proved that $F(t)$ can be chosen to be analytic in the whole t plane except for four quadratic branch points of index $-\frac{1}{2}$ at $t = t_\pm$ and at $t = t_{0\pm}$, the latter being complex conjugate points independent of ρ and z, each pair being joined by a cut.

In Ref. 20, HE have demonstrated that Eqs. (2.3)–(2.11) are easily solved on the z axis ($\rho = 0$), thereby providing a remarkably simple and convenient characterization of (modified) special HE gauge. First, if $\mathscr{E}$ is analytic and $f \neq 0$ on an open interval $\mathscr{I}$ of the z axis, then $\partial H_{AB}/\partial\rho$, $\partial F_{AB}(t)/\partial\rho$, and all derivatives of odd order with respect to ρ vanish on $\mathscr{I}$. Then, $\omega =$ constant $= \omega_0$, say, on $\mathscr{I}$. In special HE gauge, $\omega = 0$ on $\mathscr{I}$. Hence, integrating Eqs. (2.5b) and (2.7) along $\mathscr{I}$, we find

$$H_{AB}(0, z) = \begin{pmatrix} \mathscr{E}(0, z) & 2iz \\ 0 & 0 \end{pmatrix}, \tag{2.16a}$$

$$F_{AB}(0, z, t) = \begin{pmatrix} \dfrac{t\mathscr{E}(0, z)}{1 - 2tz} & \dfrac{i}{1 - 2tz} \\ -i & 0 \end{pmatrix}, \tag{2.16b}$$

in special HE gauge. In modified special HE gauge, where $\omega = \omega_0$ on $\mathscr{I}$,

$$H_{AB}(0, z) = \begin{pmatrix} \mathscr{E} & 2iz - \omega_0\mathscr{E} \\ -\omega_0\mathscr{E} & -2iz\omega_0 + \omega_0^2\mathscr{E} \end{pmatrix}, \tag{2.17a}$$

$$F_{AB}(0, z, t) = \begin{pmatrix} \dfrac{t\mathscr{E}}{1 - 2tz} & \dfrac{i - \omega_0 t\mathscr{E}}{1 - 2tz} \\ -i - \dfrac{\omega_0 t\mathscr{E}}{1 - 2tz} & \dfrac{-2itz\omega_0 + \omega_0^2 t\mathscr{E}}{1 - 2tz} \end{pmatrix}, \tag{2.17b}$$

where $\mathscr{E} = \mathscr{E}(0, z)$. Clearly, the only singularity in the t plane is a simple pole where t_+ and t_- coalesce at $t = (2z)^{-1}$.

Let us now turn to the description of the matrix element $u(t) = u^A{}_B(t)$ which represents an element of $\mathbf{K}$ in a manner depending on the choice of gauge for $F(t)$ and a contour L in the complex t plane.[19,20] For convenience of expression, we shall often denote an element of $\mathbf{K}$ by its corresponding $u(t)$ matrix if the gauge and choice of L is clear from the context. First, since $F(t)$ is always analytic in an open region containing $t = 0$, we can draw a simple closed curve L surrounding $t = 0$, symmetric about the real axis, whose interior we denote L_+, exterior L_-, such that $F(t)$ is analytic in L_+ and on L. In particular, the points $t = t_\pm$ and the cut joining them must be in L_-. The interior L_+ can be made as large as desired by putting $F(t)$ in (modified) special HE gauge and considering points (ρ, z) close to (0, 0). The matrix $u(t)$ must be analytic at least in an open annulus containing L. Further, it satisfies the algebraic constraints

$$\det u(t) = 1 \quad \text{or} \quad u_{XA}(t)u^X{}_B(t) = \epsilon_{AB}, \tag{2.18a}$$

$$u^*(t) = u(t). \tag{2.18b}$$

In addition, HE impose a boundary condition at $t = \infty$ [Eq. (2.19) below] which is not merely a restriction on the gauge, but actually excludes a significant portion of the group $\mathbf{K}$. This topic will be discussed in a sequel to the present paper. Here, we shall require that $F(t)$ be put in either HE gauge or modified HE gauge and that $u(t)$ be analytic in a neighborhood of $t = \infty$ and

(i) $u^1{}_1(t)$, $tu^1{}_2(t)$, $t^{-1}u^2{}_1(t)$, $u^2{}_2(t)$ are analytic at $t = \infty$ whenever $F(t)$ is in HE gauge; (2.19)

(ii) $u(t)$ is analytic at $t = \infty$ whenever $F(t)$ in in modified HE gauge (2.19')

[cf. Eqs. (2.14) and (2.14')]. When combining or multiplying transformations in $\mathbf{K}$ [represented by multiplication of corresponding $u(t)$ matrices], it is important not to mix the two types (2.19) and (2.19').

When $F(t)$, $u(t)$, and L are given, the $F(t)$-potential $F'(t)$ of a new solution $\mathscr{E}'$ of the field equations may be found by solving the matrix homogeneous Hilbert problem (HHP),

$$X_-(t) = X_+(t)G(t), \tag{2.20}$$

where

$$X_+(t) = F'(t)F(t)^{-1}, \tag{2.21a}$$

$$G(t) = F(t)u(t)F(t)^{-1}. \tag{2.21b}$$

In Eq. (2.20), $G(t)$ is a given matrix analytic on L, and the unknowns X_- and X_+ are required to satisfy

$$X_-(t) \text{ analytic in } L + L_- \text{ and at } t = \infty, \tag{2.22a}$$

$$X_+(t) \text{ analytic in } L + L_+, \quad X_+(0) = I. \tag{2.22b}$$

The HHP can also be written

$$X_-(t) = F'(t)u(t)F(t)^{-1}, \tag{2.23a}$$

in which X_- and F' are the unknowns, $F'_{AB}(0) = i\epsilon_{AB}$, or, in tensor notation,

$$X_-{}_A{}^B(t) = -S(t)F'_{AX}(t)u^X{}_Y(t)F^{BY}(t). \tag{2.23b}$$

With the boundary conditions at $t = 0$ and $t = \infty$, the solution of the HHP is unique if it exists.[19] The new metric tensor

f'_{AB} and Ernst potential $\mathscr{E}'$ follow from

$$H'_{AB} = \dot{F}'_{AB}(0), \quad f'_{AB} = \mathrm{Re}H'_{AB}, \quad \mathscr{E}' = H'_{11}. \tag{2.24}$$

There is no loss of generality in requiring $u(t)$ to be analytic throughout L_- as singularities can be absorbed by gauge changes. First, suppose $F(t)$ is in (modified) special HE gauge. Then Eq. (2.23) shows that $F'(t)$ is in the same special gauge if and only if $u(t)$ is analytic throughout L_-. If $u(t)$ is not analytic in L_-, then perform the factorization $u(t) = u_+(t)u_-(t)$, where u_+ and u_- satisfy the conditions for a $u(t)$ matrix and, in addition, u_+ is analytic in L_+, u_- in L_-, and $u_+(0) = I$ (this is an HHP). It follows that $F'(t)u_+(t)$ is an $F(t)$ potential in (modified) special HE gauge while $u_+(t)$ merely effects a gauge change [cf. Eqs. (2.12) and (2.13)]. Similarly, if $F(t)$ for the original solution is not in (modified) special HE gauge but still satisfies (2.14) or (2.14'), then write it as $F(t) = F_{sp}(t)g(t)$ where $F_{sp}(t)$ is in the special gauge and $g(t)$ satisfies Eqs. (2.13). Then perform the factorization $u(t)g(t)^{-1} = u_+(t)u_-(t)$, as before, and the HHP will take the form

$$X_-(t) = F'_{sp}(t)u_-(t)F_{sp}(t)^{-1},$$

where $F'_{sp}(t) = F'(t)u_+(t)$. Except where otherwise stated, we shall henceforth assume that $u(t)$ is analytic throughout L_-. [Notice that the full group of gauge transformations cannot be handled by the HE formalism because $F'(t)$ is required to be analytic in L_+. In some cases where $F'(t)$ is not analytic in L_+ it may be possible to deform the contour L without crossing singularities of $u_-(t)$, but this is clearly impossible when $F'(t)$ and $u_-(t)$ have coincident singularities in L_+. An example of such a coincidence is the "extended" HKX transformation.[12]]

The Hauser–Ernst formula, mentioned in Sec. 1, which gives all $u(t)$ which map a given initial solution $\mathscr{E}$ to a given final solution $\mathscr{E}'$ is derived as follows. Suppose that $\mathscr{E}$ and $\mathscr{E}'$ are analytic on an open interval $\mathscr{I}$ of the z axis containing $(\rho, z) = (0, 0)$ and put $F(t)$ and $F'(t)$ in special HE gauge. Then, on $\mathscr{I}$, $F(t)$ and $F'(t)$ are given by Eq. (2.16b). Substitute into the HHP Eq. (2.23a) and observe that the left-hand side is analytic in L_- whereas the right-hand side apparently has a simple pole at $t = (2z)^{-1}$ in L_-. Setting the residue to be zero gives

$$tu^1{}_2(t)\mathscr{E}\mathscr{E}' + t^{-1}u^2{}_1(t) - iu^1{}_1(t)\mathscr{E}' + iu^2{}_2(t)\mathscr{E} = 0, \tag{2.25}$$

where $\mathscr{E}$ and $\mathscr{E}'$ are to be evaluated at $\rho = 0, z = (2t)^{-1}$. The real and imaginary parts of Eq. (2.25), together with Eq. (2.18a), provide three equations for the four components of $u^A{}_B(t)$. If, instead, we put $F(t)$ and $F'(t)$ in modified special HE gauge, with $\omega = \omega_0$, $\omega' = \omega'_0$ on $\mathscr{I}$, then

$$u(t) = \begin{pmatrix} 1 & \omega'_0 \\ 0 & 1 \end{pmatrix} u_0(t) \begin{pmatrix} 1 & -\omega_0 \\ 0 & 1 \end{pmatrix}, \tag{2.26}$$

where $u_0(t)$ satisfies Eq. (2.25). This matrix product expresses the composition of three transformations: (i) $\omega \to \omega - \omega_0$, (ii) $\mathscr{E} \to \mathscr{E}'$ preserving special HE gauge, and (iii) $\omega' \to \omega' + \omega'_0$.

The Ehlers group **P** is given by

$$\mathscr{E}' = (P)_\alpha \mathscr{E} = \frac{\alpha_4\mathscr{E} + i\alpha_3}{\alpha_1 - i\alpha_2\mathscr{E}}, \quad \alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix}, \tag{2.27}$$

$\det\alpha = 1$, $(P)_\alpha \in \mathbf{P}$. The Matzner–Misner group **L** (rotation of Killing vectors) is given by

$$f'_{AB} = (L)_\beta f_{AB} = b_A{}^C b_B{}^D f_{CD}, \quad \beta = b_A{}^B = \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{pmatrix}, \tag{2.28}$$

$\det\beta = 1$, $(L)_\beta \in \mathbf{L}$. Both groups are one-to-two homomorphic to SL(2, R). The $(L)_\beta$ transformation is represented by

$$u^A{}_B(t) = -b^A{}_B = (\beta^T)^{-1} = \begin{pmatrix} \beta_4 & -\beta_3 \\ -\beta_2 & \beta_1 \end{pmatrix}, \tag{2.29}$$

^{T} denoting transpose, and the solution of the HHP (trivial in this case) is

$$F'_{AB}(t) = b_A{}^C b_B{}^D F_{CD}(t) \quad \text{or} \quad F'(t) = \beta F(t)\beta^T. \tag{2.30}$$

Note that $u(t)$ obeys condition (2.19') and preserves modified special HE gauge. The $(P)_\alpha$ transformation can be substituted into the HE formula (2.25) to yield

$$u(t) = \begin{pmatrix} \alpha_1 & -\alpha_2 t^{-1} \\ -\alpha_3 t & \alpha_4 \end{pmatrix}, \tag{2.31}$$

obeying condition (2.19). The solution of the HHP in this case can be obtained by a straightforward application of methods outlined in Secs. 3 and 5 below. The result is

$$F'(t) = (A + t^{-1}B)F(t)u(t)^{-1}, \tag{2.32}$$

where

$$A = \begin{pmatrix} \alpha_4 + i\alpha_2\mathscr{E}' & 0 \\ i\alpha_2(H_{12} + H'_{12} - 2iz) & \alpha_1 - i\alpha_2\mathscr{E} \end{pmatrix}, \tag{2.33a}$$

$$B = \begin{pmatrix} 0 & 0 \\ \alpha_2 & 0 \end{pmatrix}. \tag{2.33b}$$

Here, $\mathscr{E}'$ is given by Eq. (2.27) and H'_{12} by

$$H'_{12} = \frac{\alpha_1 H_{12} + \alpha_2 H^{(2)}_{11}}{\alpha_1 - i\alpha_2\mathscr{E}}, \tag{2.34}$$

where $H^{(2)}_{AB} = \frac{1}{2}\ddot{F}_{AB}(0)$. Eq. (2.32) preserves special HE gauge.

3. QUADRATIC BRANCH POINTS AND THE HARRISON TRANSFORMATION

The homogeneous Hilbert problem can be solved by elementary methods in the case when $u(t)$ has only poles in L_+ and for a large class of cases where $u(t)$ has also quadratic branch points and cuts contained in L_+. The case of N simple poles (at points $t \neq 0$ in L_+) has been treated adequately by Hauser and Ernst[18] using their integral equation. They have identified the transformation for which

$$u(t) = I + \sum_{i=1}^{N} \frac{\alpha_i t}{t - s_i} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{3.1}$$

α_i, s_i real constants, with the product of N HKX transformations.[3] Allowing all four components of $u(t)$ to have simple poles at $t = s_1, \ldots, s_N \neq 0$ leads to a product of null generalized HKX transformations.[12] HE have also obtained the electrovac counterparts of the latter transformations.[18] The case of nonsimple poles can be treated directly by the same methods or by regarding a pole of multiplicity N as a coales-

cence of N simple poles. The corresponding limit of the product of N HKX transformations is a combined HKX transformation of ranks 0, 1, ..., $N-1$.[3]

It is instructive to see the null generalized HKX transformation derived from the HHP in an SL(2)-covariant manner. Consider

$$u^A{}_B(t) = -\epsilon^A{}_B + \frac{\alpha t}{t-s} q^A{}_B, \tag{3.2}$$

where s, α, and $q^A{}_B$ are constants and q_{AB} is symmetric and null (recall $-\epsilon^A{}_B = \epsilon_B{}^A = \delta^A_B$). The contour L is drawn to enclose $t=0$ and the pole at $t=s$. The HHP is

$$X_{-A}{}^B(t) = -S(t)F'_{AX}(t)\left[-\epsilon^X{}_Y + \frac{\alpha t}{t-s} q^X{}_Y\right]F^{BY}(t). \tag{3.3}$$

The left-hand side is analytic in $L_+ + L_-$ and at $t=\infty$. The right-hand side is analytic in $L_+ + L_+$ except for a simple pole at $t=s$ and has the value $\epsilon_A{}^B$ at $t=0$. Hence

$$-S(t)F'_{AX}(t)\left[-\epsilon^X{}_Y + \frac{\alpha t}{t-s} q^X{}_Y\right]F^{BY}(t) = \epsilon_A{}^B + \frac{t}{t-s} R_A{}^B, \tag{3.4}$$

where $R_A{}^B$ is a constant (in t) tensor to be determined. Expressing $F'_{AB}(t)$ as the subject, we have

$$F'_{AB}(t) = \left[F_{AY}(t) + \frac{t}{t-s} R_A{}^X F_{XY}(t)\right]\left[\epsilon_B{}^Y - \frac{\alpha t}{t-s} q_B{}^Y\right]. \tag{3.5}$$

Comparing residues at $t=s$ of both sides of Eq. (3.4), we find

$$R_A{}^B = -\alpha S(s)F'_{AX}(s)q^X{}_Y F^{BY}(s). \tag{3.6}$$

Since q_{AB} is null and symmetric, it admits the factorization $q_{AB} = q_A q_B$. An expression for the vector $F'_{AX}(s)q^X$ and hence for $R_A{}^B$ can be obtained by multiplying both sides of Eq. (3.5) by q^B and taking the limit as $t \to s$. The results are

$$F'_{AX}(s)q^X = F_{AX}(s)q^X\left[1 + \alpha s S(s) q^{CD} F_{ZC}(s)\dot{F}^Z{}_D(s)\right]^{-1}, \tag{3.7}$$

$$R_A{}^B = -\alpha S(s)F_{AX}(s)q^X{}_Y F^{BY}(s) \times \left[1 + \alpha s S(s) q^{CD} F_{ZC}(s)\dot{F}^Z{}_D(s)\right]^{-1}. \tag{3.8}$$

The expression for the transformed $F(t)$ potential simplifies if we introduce the generating function $G_{AB}(s,t)$ of Kinnersley and Chitre[16] which is given in terms of $F_{AB}(t)$ by[3]

$$G_{AB}(s,t) = \frac{s}{s-t}\epsilon_{AB} + \frac{tS(s)}{s-t} F_{XA}(s)F^X{}_B(t), \tag{3.9}$$

with a suitable limit for $G_{AB}(s,s)$. Then Eqs. (3.5), (3.8), and (3.9) give

$$F'_{AB}(t) = \left\{F_A{}^X(t) + \frac{\alpha q^{CD} F_{AC}(s)\left[G_D{}^X(s,t) - s(s-t)^{-1}\epsilon_D{}^X\right]}{1 - \alpha q^{EF} G_{EF}(s,s)}\right\} \times \left[\epsilon_{XB} - \frac{\alpha t}{s-t} q_{XB}\right]. \tag{3.10}$$

This is precisely the formula given in Ref. 12 for the generalized HKX transformation, whose infinitesimal form is defined by[25]

$$\sum_{k=0}^{\infty} \alpha s^k q^{XY} \gamma^{(k)}_{XY}, \tag{3.11}$$

for the case of null q^{XY}. It is easy to show that the apparent pole at $t=s$ in the right-hand side of Eq. (3.10) is absent.

The second factor in Eq. (3.10) is a gauge function. The first factor, with the index X lowered by ϵ_{XB}, is the transform of $F_{AB}(t)$ under the "extended" HKX transformation,[12] defined by (3.11) with the sum taken from $k=-\infty$ to $k=\infty$. In Ref. 12, the extended HKX transformation was derived as the limit of the double soliton transformation of Belinsky and Zakharov[9] when the two simple poles of their matrix $\chi(\lambda)$ coalesce to form a double pole.

Transformations corresponding to a simple or nonsimple pole at $t=0$ have a longer history. These can be described as or represented by: (i) products of a finite number of $(P)_\alpha$ and $(L)_\beta$ transformations[1,2]; (ii) a broken null curve in the Geroch representation[2]; (iii) repeated applications of the Lagrangian invariance transformations of Hoenselaers[30]; (iv) transformations $q^{XY}\gamma^{(k)}_{XY}$, q^{XY} null, exponentiated by KC[15] and products thereof. The failure of these early transformations to preserve asymptotic flatness has a rational explanation in terms of the HE representation. Eq. (2.25) shows that $u(t)$ will preserve asymptotic flatness up to a NUT parameter whenever it is asymptotic to the right-hand side of (2.31) as $t \to 0$ (not necessarily with all components of α nonvanishing). Thus asymptotic flatness preservation is not sensitive to the types of singularities of $u(t)$ occurring at points $t \neq 0$, showing in a sense that the "majority" of elements of **K** actually preserve asymptotic flatness, but singularities at $t=0$ [except for a simple pole in $u^1{}_2(t)$] are not allowed.

Matrices $u(t)$ which have quadratic branch points arise naturally on account of the determinant condition, $\det u(t) = 1$. Let $v(t)$ be a matrix function of t analytic and possessing an inverse in $L_+ + L_-$, real for real t, and obeying either condition (2.19) or (2.19'). Then

$$u(t) = [\det v(t)]^{-1/2} v(t) \tag{3.12}$$

obeys all the conditions for a $u(t)$ matrix and is analytic in L_-, provided the branch cuts joining the zeros of $\det v(t)$ can be contained in L_+. This can be done if $t=\infty$ is not also a branch point. This situation can be excluded by requiring that $\det v(t) \to 1$ (or any positive real constant) as $t \to \infty$.

If $v(t)$ is a given rational function of t, then the HHP can be solved by essentially the same method as for $u(t)$ rational. The HHP can be rewritten

$$[\det v(t)]^{1/2} X_-(t) = F'(t)v(t)F(t)^{-1}. \tag{3.13}$$

The left-hand side is analytic in L_- and at $t=\infty$ while the right-hand side has only poles in L_+. The branch point singularities have been absorbed in $X_-(t)$. Thus either side of Eq. (3.13) defines a rational function of t and the problem reduces to the determination of the unknown coefficients in the numerator. The transformation represented by such a $u(t)$ will be seen to be a product of Harrison's Bäcklund transformations[8] (branch points at $t \neq 0$), null HKX transformations[3,12] (poles at $t \neq 0$), and null KC transformations[2,15,30] (either pole or branch point at $t=0$, depending on whether number of accompanying Harrison transformations is even

or odd, respectively). Further, when two Harrison transformations with same s parameters are multiplied, the corresponding product $u(t)$ matrix has only a pole at $t=s$ and so represents a null HKX transformation.[12]

The simplest SL(2)-covariant example is

$$v^A{}_B(t) = -\epsilon^A{}_B + t^{-1}h^A{}_B, \quad \det v(t) = 1 - t^{-1}h_X{}^X, \tag{3.14}$$

where h_{AB} is null and nonsymmetric. The explicit solution of the HHP is

$$F'_{AB}(t) = \left[s^{-1}h_A{}^C - \frac{t(h_{ZA} + ih_{AX}h_{ZY}H^{XY})F^{CZ}(s)}{(s-t)h_{MN}F^{NM}(s)}\right] \times F_{CD}(t)(\epsilon_B{}^D - t^{-1}h_B{}^D), \tag{3.15}$$

where $s = h_X{}^X = -h^X{}_X$. The more general case where h_{AB} is nonnull, so that $\det v(t) = 1 - st^{-1} + t^{-2}(\det h)$, is easily shown to factorize into two of the above transformations. The proof of Eq. (3.15) involves elegant and pleasing tensor manipulations, which we leave as an exercise for the reader. Instead, we shall give the derivation for the closely related, though non-SL(2)-covariant, Harrison transformation, for which

$$u(t) = \left(1 - \frac{s}{t}\right)^{-1/2}\begin{pmatrix} 1 & -cst^{-1} \\ -c^{-1} & 1 \end{pmatrix}. \tag{3.16}$$

The $u(t)$ matrix given by (3.12) and (3.14) can be factorized into matrices of the forms (3.16), (2.29), and (2.31) in at least three ways.

The transform of $F(t)$ under the Harrison (H) transformation was calculated in Ref. 12 by first deriving the transforms of $F(t)$ under the groups **P**, **L**, **Q**,[5-7] and $\tilde{\mathbf{Q}}$,[6,7] and then calculating the products,

$$H = (L)_\beta(\tilde{Q})_{-4s}(Q)_{4s}(P)_\alpha = (P)_\alpha(Q)_{-4s}(\tilde{Q})_{4s}(L)_\beta, \tag{3.17a}$$

$$\alpha = \begin{pmatrix} 1 & cs \\ 0 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & c^{-1} \\ 0 & 1 \end{pmatrix}, \tag{3.17b}$$

$(P)_\alpha \in \mathbf{P}$, $(L)_\beta \in \mathbf{L}$, $(Q)_{\pm 4s} \in \mathbf{Q}$ (commutes with **P**), $(\tilde{Q})_{\pm 4s} \in \tilde{\mathbf{Q}}$ (commutes with **L**). The resulting formula [Eqs. (4.44a) and (4.44b) of Ref. 12] preserves special HE gauge, and so can be substituted directly into Eq. (2.25) here to give the representing matrix (3.16). It is, of course, more instructive to see Eqs. (4.44a) and (4.44b) of Ref. 12 derived from (3.16) and the HHP. The insights gained will allow us to construct new Bäcklund transformations for electrovac spacetimes in Sec. 5.

The HHP takes the form,

$$\left(1 - \frac{s}{t}\right)^{1/2} X_-(t) = F'(t)\begin{pmatrix} 1 & -cst^{-1} \\ -c^{-1} & 1 \end{pmatrix}F(t)^{-1}. \tag{3.18}$$

The left-hand side is analytic in $L_+ + L_-$ and at $t=\infty$ [the branch cut in $(1-s/t)^{1/2}$ joins $t=0$ to $t=s$ in L_+]. The right-hand side is analytic in $L + L_+$ except for a pole at $t=0$. It follows that

$$F'(t)\begin{pmatrix} 1 & -cst^{-1} \\ -c^{-1} & 1 \end{pmatrix}F(t)^{-1} = cst^{-1}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 + icsH'_{11} & c^{-1} \\ ics(H'_{21} + H_{12}) & 1 - icsH_{11} \end{pmatrix},$$

where Eqs. (2.9a) and (2.9b) have been used. Hence

$$F'(t) = \frac{t}{t-s}\begin{pmatrix} 1 + icsH'_{11} & c^{-1} \\ cst^{-1} + ics(H'_{21} + H_{12}) & 1 - icsH_{11} \end{pmatrix} \times F(t)\begin{pmatrix} 1 & cst^{-1} \\ c^{-1} & 1 \end{pmatrix}.$$

The pole at $t=s$ is absent if

$$1 + icsH'_{11} = -c^{-1}T, \tag{3.19a}$$

$$ics(H'_{21} + H_{12}) = -c - (1 - icsH_{11})T, \tag{3.19b}$$

where

$$T = \frac{F_{22}(s) + cF_{21}(s)}{F_{12}(s) + cF_{11}(s)}. \tag{3.20}$$

Hence the final transformation formula is

$$F'(t) = \frac{t}{t-s}\begin{pmatrix} -c^{-1}T & c^{-1} \\ c(s-t)t^{-1} - (1 - ics\mathscr{E})T & 1 - ics\mathscr{E} \end{pmatrix} \times F(t)\begin{pmatrix} 1 & cst^{-1} \\ c^{-1} & 1 \end{pmatrix}. \tag{3.21}$$

When F_{21} and F_{22} are eliminated using Eq. (2.11), we get Eqs. (4.44a) and (4.44b) of Ref. 12. [T is a pseudopotential for the Bäcklund transformation and is a fractional linear function of the pseudopotential q used by Harrison,[7] the q used by Cosgrove,[6,12] and the two α's used by Neugebauer.[11] From Eq. (2.7), a total Riccati equation can be written for T.]

The decomposition of the Harrison transformation into factors **PQ** and $\mathbf{L}\tilde{\mathbf{Q}}$ (respectively, I_1 and I_2 of Neugebauer[11]) is not unique and Eq. (3.17b) presents only one possible parametrization. It has the disadvantage that it breaks down for the important cases $c=\infty$ and $c=0$, which map Weyl solutions to Weyl solutions. These may be accommodated by trivial rescalings, $f \to (\text{const})\times f$, and translations, either $\omega \to \omega + \text{const}$ or $\psi \to \psi + \text{const}$. In terms of representing matrices the products are, suppressing the factor $(1-s/t)^{-1/2}$,

$$v(t) = \lim_{c\to\infty}\begin{pmatrix} (cs)^{-1} & 0 \\ t & cs \end{pmatrix}\begin{pmatrix} 1 & -cst^{-1} \\ -c^{-1} & 1 \end{pmatrix} = \begin{pmatrix} 0 & -t^{-1} \\ t-s & 0 \end{pmatrix}; \tag{3.22}$$

$$v(t) = \lim_{c\to 0}\begin{pmatrix} c^{-1} & 1 \\ 0 & c \end{pmatrix}\begin{pmatrix} 1 & -cst^{-1} \\ -c^{-1} & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 - st^{-1} \\ -1 & 0 \end{pmatrix}. \tag{3.23}$$

The solutions of the corresponding HHP's are, respectively,

$$F'(t) = \begin{pmatrix} -sT & s \\ \dfrac{s-t}{st} + i\mathscr{E}T & -i\mathscr{E} \end{pmatrix}F(t)\begin{pmatrix} 0 & \dfrac{1}{t-s} \\ -t & 0 \end{pmatrix}, \quad T = \frac{F_{21}(s)}{F_{11}(s)}; \tag{3.24}$$

$$F'(t) = \begin{pmatrix} -T & 1 \\ \dfrac{s-t}{t} + is\mathscr{E}T & -is\mathscr{E} \end{pmatrix}F(t)\begin{pmatrix} 0 & -1 \\ \dfrac{t}{t-s} & 0 \end{pmatrix}, \quad T = \frac{F_{22}(s)}{F_{12}(s)}. \tag{3.25}$$

The former preserves special HE gauge, as does (3.21). The

latter preserves modified special HE gauge. The corresponding decompositions into **PQ** and $\mathbf{L}\tilde{\mathbf{Q}}$ transformations are, respectively,

$$H = (\tilde{Q})_{-4s}(Q)_{4s}(P)_{\alpha_1} = (P)_{\alpha_1}(Q)_{-4s}(\tilde{Q})_{4s}; \tag{3.26}$$

$$H = (L)_{\alpha_1}(\tilde{Q})_{-4s}(Q)_{4s} = (Q)_{-4s}(\tilde{Q})_{4s}(L)_{\alpha_1}; \tag{3.27}$$

where, in both cases, $\alpha_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

The factorization of the null HKX transformation (3.10) into two H transformations with same s parameters is expressed by the matrix product

$$\begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} 1 - \dfrac{\alpha t q^{12}}{t-s} & \dfrac{\alpha t q^{11}}{t-s} \\ -\dfrac{\alpha t q^{22}}{t-s} & 1 + \dfrac{\alpha t q^{12}}{t-s} \end{pmatrix}$$

$$= \frac{t}{t-s} \begin{pmatrix} 1 & -c_2 s t^{-1} \\ -c_2^{-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & -c_1 s t^{-1} \\ -c_1^{-1} & 1 \end{pmatrix}, \tag{3.28}$$

where

$$c_1 = q^{11}/q^{12}, \quad c_2 = -(q^{11}/q^{12})(1 + \alpha q^{12}), \tag{3.29a}$$

$$\lambda = 1 + \alpha q^{12}, \quad \mu = -\alpha q^{11}. \tag{3.29b}$$

the relations between the parameters here is in agreement with those given in Eqs. (6.56)–(6.58) of Ref. 12.

The $u(t)$ matrix (3.16) for the H transformation has quadratic branch points at $t = 0$ and $t = s$, joined by a cut in L_+. The product of two such matrices with $s = s_1$ and $s = s_2$ has branch points at $t = s_1$ and $t = s_2$, but $t = 0$ is an ordinary point [cf. right-hand side of Eq. (3.28)]. The branch cut can be taken from $t = s_1$ to $t = s_2$ without passing through $t = 0$. This shows incidentally that the double Harrison transformation preserves asymptotic flatness.

4. NONNULL KC TRANSFORMATIONS AND THE B GROUP

In Refs. 3, 12, and 15, methods are given which solve the problem of exponentiating particular infinitesimal *null* KC transformations.[14] None of these methods have been found to work in the nonnull case (except for the $s = \infty$ limit of the nonnull HKX[12]). We wish to study the **B** group,[16] whose infinitesimal generators are[25]

$$\beta^{(k-1)} = \gamma_{11}^{(k+1)} + \gamma_{22}^{(k-1)}, \quad k = 0, 1, 2, \ldots, \tag{4.1}$$

and the nonnull KC group, whose generators are

$$q^{XY}\gamma_{XY}^{(k)}, \quad k = 0, 1, 2, \ldots, \tag{4.2}$$

$q^{XY}q_{XY} = 2 \det q \neq 0$. We shall not attempt to exponentiate any combinations of the infinitesimal generators (the nonnull HKX transformation[12] being an example), but, nevertheless, we shall show how *finite* transformations can be calculated with Harrison and/or null HKX transformations. There are enough of these finite transformations that iteration of them until closure occurs will yield the full infinite-dimensional groups.

First, the $u(t)$ matrix representing an infinitesimal nonnull HKX transformation (α small) is identical to the null case. From Eq. (3.2), we have

$$u^A{}_B(t) = -\epsilon^A{}_B + \frac{\alpha t}{t-s} q^A{}_B + O(\alpha^2). \tag{4.3}$$

[When q_{AB} is null, the $O(\alpha^2)$ term is identically zero and the remainder of the equation is exact for finite α.] The $u(t)$ matrix representing the individual infinitesimal $\alpha q^{XY}\gamma_{XY}^{(k)}$ transformation is obtainable from the coefficient of s^k in the Taylor series expansion of the right-hand side of Eq. (4.3). An arbitrary infinitesimal generator of the nonnull KC group is expressed by the sum

$$\sum_{k=0}^{\infty} \alpha^{(k)} q^{XY} \gamma_{XY}^{(k)}, \tag{4.4}$$

$\alpha^{(k)}$ real constants. Define the generating function

$$\alpha(t) = \sum_{k=0}^{\infty} \alpha^{(k)} t^{-k}, \tag{4.5}$$

analytic at and near $t = \infty$. We may require that $\alpha(t)$ be analytic in $L + L_-$. Then the representing $u(t)$ matrix is

$$u^A{}_B(t) = -\epsilon^A{}_B + \alpha(t) q^A{}_B + O(\alpha^2). \tag{4.6}$$

Finally, a straightforward exponentiation of this matrix yields

$$u^A{}_B(t) = -[\cosh(q\alpha(t))]\epsilon^A{}_B + q^{-1}[\sinh(q\alpha(t))]q^A{}_B, \tag{4.7}$$

where the real or pure-imaginary constant q is defined by

$$q_{XA}q^X{}_B = -q^2\epsilon_{AB} \quad \text{or} \quad q^2 = (q^{12})^2 - q^{11}q^{22}. \tag{4.8}$$

When $\alpha(t)$ is a rational function, as in the case of the nonnull HKX transformation for which $\alpha(t) = \alpha t/(t-s)$, the $u(t)$ matrix has essential singularities in L_+ at the poles of $\alpha(t)$. The methods of this paper are not strong enough to handle essential singularities in $u(t)$. However, finite transformations can be written down when $\tanh(q\alpha(t))$ is a rational function of t, say

$$\tanh(q\alpha(t)) = R(t) \tag{4.9}$$

(note that $qR(t)$ is real for real t). Then

$$u^A{}_B(t) = [-\epsilon^A{}_B + q^{-1}R(t)q^A{}_B][1 - R^2(t)]^{-1/2}. \tag{4.10}$$

This is of the form (3.12) and so represents a product of a finite number of Harrison and $(L)_B$ transformations. When q is pure-imaginary, the s parameters of the H transformations form complex conjugate pairs such that the product transformation obeys condition (2.18b) and maps real solutions (as determined by f_{AB}) to real solutions.

Consider the simplest case, where

$$R(t) = \frac{at+b}{ct+d}, \quad c > a > 0, \tag{4.11}$$

a, b, c, d real constants. A short calculation reveals the factorization,

$$u(t) = \begin{pmatrix} \beta_4 & -\beta_3 \\ -\beta_2 & \beta_1 \end{pmatrix} \left(1 - \frac{s_2}{t}\right)^{-1/2} \left(1 - \frac{s_1}{t}\right)^{-1/2}$$

$$\times \begin{pmatrix} 1 & -c_2 s_2 t^{-1} \\ -c_2^{-1} & 1 \end{pmatrix} \begin{pmatrix} 1 & -c_1 s_1 t^{-1} \\ -c_1^{-1} & 1 \end{pmatrix} \tag{4.12}$$

[cf. Eqs. (2.29) and (3.16) for the $(L)_\beta$ and H transformations, respectively], where

$$\beta_4 = \Delta^{-1}(c^2 - a^2)^{1/2}(d - bq^{12}/q), \tag{4.13a}$$

$$\beta_3 = -(c^2 - a^2)^{-1/2}aq^{11}/q, \tag{4.13b}$$

$$\beta_2 = \Delta^{-1}(c^2 - a^2)^{1/2}bq^{22}/q, \tag{4.13c}$$

$$\beta_1 = (c^2 - a^2)^{-1/2}(c + aq^{12}/q); \tag{4.13d}$$

$$s_1 = -\frac{d+b}{c+a}, \quad s_2 = -\frac{d-b}{c-a}; \tag{4.14}$$

$$c_1 = \frac{q^{11}}{q^{12}+q}, \quad c_2 = -\frac{\Delta}{(c+a)(d-b)}\frac{q^{11}}{q^{12}+q}; \tag{4.15}$$

$$\Delta = cd - ab + (ad - bc)q^{12}/q. \tag{4.16}$$

There is a second factorization with same β matrix and

$$s_1 = -\frac{d-b}{c-a}, \quad s_2 = -\frac{d+b}{c+a}; \tag{4.14'}$$

$$c_1 = \frac{q^{11}}{q^{12}-q}, \quad c_2 = -\frac{\Delta}{(c-a)(d+b)}\frac{q^{11}}{q^{12}-q}. \tag{4.15'}$$

A matrix of the form (4.10) will represent a product of null HKX transformations alone when $1 - R^2(t)$ is a perfect square. This is the case whenever

$$R(t) = \frac{m^2(t) - n^2(t)}{m^2(t) + n^2(t)} \tag{4.17a}$$

or

$$R(t) = \frac{2m(t)n(t)}{m^2(t) + n^2(t)}, \tag{4.17b}$$

where $m(t)$ and $n(t)$ are polynomials in t.

The same considerations apply to the **B** group,[16] for which

$$u(t) = \begin{pmatrix} \cos\alpha(t) & t^{-1}\sin\alpha(t) \\ -t\sin\alpha(t) & \cos\alpha(t) \end{pmatrix}, \tag{4.18}$$

$\alpha(t)$ analytic in $L + L_-$ and at $t = \infty$. This is the group which maps flat space to itself in special HE gauge, as can be seen by putting $\mathscr{E} = 1 = \mathscr{E}'$ in Eq. (2.25). It is closely related to the nonnull KC group [see Eqs. (7.30)–(7.33b) of Ref. 12[31]]. Finite transformations in the **B** group can be constructed out of Harrison and $(P)_\alpha$ transformations by putting

$$\tan\alpha(t) = R(t), \tag{4.19}$$

a rational function of t. Then

$$u(t) = [1 + R^2(t)]^{-1/2}\begin{pmatrix} 1 & t^{-1}R(t) \\ -tR(t) & 1 \end{pmatrix}. \tag{4.20}$$

In this case, the s parameters form complex conjugate pairs. Also, with the choice,

$$R(t) = \frac{2m(t)n(t)}{m^2(t) - n^2(t)}, \tag{4.21}$$

where $m(t)$ and $n(t)$ are polynomials in t, the $u(t)$ matrix is itself rational and represents a product of null HKX transformations.

5. BÄCKLUND TRANSFORMATIONS FOR ELECTROVAC FIELDS

Hauser and Ernst have extended their complex-variable techniques to the enlarged Geroch group **K'** (due to Kinnersley[13]) for electrovac fields.[18–20] Their representation employs 3×3 matrices H_{ab}, $F_{ab}(t)$, and $u_{ab}(t)$ which correspond to the 2×2 matrices H_{AB}, $F_{AB}(t)$, and $u^A{}_B(t)$, respectively, of vacuum. The SL(2) tensor formalism for the potentials has been developed by Kinnersley and Chitre[13–16] and Jones.[32] However, since each 3×3 matrix equation conveys the information of four tensor equations, we shall use the matrix notation in this section.

With the hindsight gained from the vacuum case, it is natural to study electrovac transformations represented by $u(t)$ matrices which have only poles in the complex t plane and those which have algebraic singularities (cubic branch points in this case) for which the HHP can be solved by the methods of Sec. 3. Hauser and Ernst[18] have already treated the case where $u(t)$ has only poles using their integral equation and have obtained the electrovac version of the null HKX transformation and products thereof. The same results could equally well be obtained from the HHP by following the steps which led to Eq. (3.10) here.

In this section we study $u(t)$ matrices of the form

$$u(t) = [\det v(t)]^{-1/3}v(t) \tag{5.1}$$

[cf. Eq. (3.12)], in an attempt to generalize the Harrison transformation to electrovac spacetimes. We succeed in obtaining an electrovac counterpart to the *double* Harrison transformation which maps flat space to Kerr–Newman–NUT space and presumably an n-fold iteration of which maps flat space to the nonlinear superposition of n Kerr–Newman–NUT particles on the z axis. Nevertheless, in Sec. 6, we present an argument which suggests that this transformation may not be as large as it could be and attempt to explain why we were not able to derive the *single* electrovac Harrison transformation (presumed to exist) from the HE formalism.

The electrovac H potential has components

$$H_{ab} = \begin{pmatrix} H_{AB} & \varphi_A \\ 2iL^{(1,1)}_B & 2iK^{(1,1)} \end{pmatrix}, \quad a, b = 1, 2, 3, \quad A, B = 1, 2. \tag{5.2}$$

The definitions of the potentials on the right-hand side are to be found in Refs. 13 and 14. The metric is recoverable from the relation

$$f_{AB} = \tfrac{1}{2}(H_{AB} + H^*{}_{BA}) + \varphi_A\varphi^*{}_B - iz\epsilon_{AB}, \tag{5.3}$$

and we again use the parametrization (2.2). To specify an electrovac solution it is sufficient to know only the components (Ernst potentials)

$$\mathscr{E} = H_{11} = f_{11} - \varphi_1\varphi^*{}_1 + i\Omega_{11}, \quad \Phi = \varphi_1 = H_{13}, \tag{5.4}$$

which satisfy simple field equations given by Ernst[33] (see also Ref. 13[34]).

The $F(t)$ potential has components

$$F_{ab}(t) = \begin{pmatrix} F_{AB}(t) & D_A(t) \\ 2iS_B(t) & 2iQ(t) \end{pmatrix}, \tag{5.5}$$

where the entries on the right-hand side are generating functions whose definitions and direct methods of calculation are found in Ref. 32. (When comparing with the 3×3 matrix equations of Hauser and Ernst,[18–20] one should interchange the first two rows and first two columns as in Ref. 26.) At

$t = 0$,

$$F(0) = i\mathfrak{E}(2), \tag{5.6a}$$
$$\dot{F}(0) = H, \tag{5.6b}$$

where $\mathfrak{E}(2)$ is the value at $t = 2$ of

$$\mathfrak{E}(t) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2}it \end{pmatrix}. \tag{5.7}$$

The $F(t)$ potential satisfies a differential equation and algebraic relations corresponding to Eqs. (2.7), (2.10), and (2.11) which are given in Refs. 18, 19, and 32. We need write down only the determinant condition

$$\det F(t) = -S^{-1}(t). \tag{5.8}$$

The gauge [see Eq. (5.22) below] chosen by Hauser and Ernst ("HE gauge") requires that $F(t)$ be analytic near $t = \infty$ and also that F_{a1}, tF_{a2}, and F_{a3}, $a = 1, 2, 3$, be analytic at $t = \infty$. As in vacuum, there are special gauges such that the only singularities of $F(t)$, in a suitable (ρ, z) domain containing an open interval of the z axis, are quadratic branch points at the zeros t_+ and t_- of $S(t)$. If, in addition, the additive constants in $\omega = -f_{12}/f_{11}$ and φ_2 are chosen so that these latter potentials vanish on the z axis, then the special gauge is unique ("special HE gauge"). The explicit forms of H and $F(t)$ on the z axis in special HE gauge are given by[20]

$$H = \begin{pmatrix} \mathscr{E} & 2iz & \Phi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.9a}$$

$$F(t) = \begin{pmatrix} \dfrac{t\mathscr{E}}{1-2tz} & \dfrac{i}{1-2tz} & \dfrac{t\Phi}{1-2tz} \\ -i & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.9b}$$

The homogeneous Hilbert problem is to find $X_+(t)$ analytic in $L + L_+$ (drawn as in Sec. 2) and $X_-(t)$ analytic in $L + L_-$ and at $t = \infty$ such that, on L,

$$X_-(t) = X_+(t)G(t), \tag{5.10}$$

where

$$G(t) = F(t)u(t)F(t)^{-1}, \tag{5.11a}$$
$$X_+(0) = I. \tag{5.11b}$$

The matrix $u(t)$, depending only on t, is chosen subject to the constraints

$$\det u(t) = 1, \tag{5.12a}$$
$$u^\dagger(t)\mathfrak{E}(t)u(t) = \mathfrak{E}(t), \tag{5.12b}$$
$$u(t) \text{ analytic in } L + L_-, \tag{5.12c}$$
$$\begin{pmatrix} u_{11} & tu_{12} & u_{13} \\ t^{-1}u_{21} & u_{22} & t^{-1}u_{23} \\ u_{31} & tu_{32} & u_{33} \end{pmatrix} \text{ analytic at } t = \infty. \tag{5.12d}$$

[$u^\dagger(t)$ is the Hermitian conjugate of $u(t^*)$, i.e., $u^\dagger(t) = u^*(t)^T$.] When $X_+(t)$ is known, a new electrovac solution can be constructed from

$$F'(t) = X_+(t)F(t), \tag{5.13a}$$
$$H' = \dot{F}'(0). \tag{5.13b}$$

Hence an alternative form for the HHP is

$$X_-(t) = F'(t)u(t)F(t)^{-1}, \tag{5.14}$$

where $X_-(t)$ and $u(t)$ are analytic in $L + L_-$, $F'(t)$ and $F(t)$ are analytic in $L + L_+$. As before, a composition of several transformations is represented by the product of corresponding $u(t)$ matrices. Notice that $X_+(t)$ and hence $F'(t)$ is not sensitive to the replacement $u(t) \to \zeta u(t)$, where ζ is a complex cube root of unity.

A simple application of the HHP is to derive an equation for all $u(t)$ which transform a given initial solution $(\mathscr{E}, \Phi)$ into a given final solution $(\mathscr{E}', \Phi')$ in special HE gauge.[20] Substitute Eq. (5.9b) for $F(t)$ and $F'(t)$ into the right-hand side of Eq. (5.14). The condition that the pole at $t = (2z)^{-1}$ be absent is

$$(t\mathscr{E}', i, t\Phi')u(t)\begin{pmatrix} i & 0 \\ -t\mathscr{E} & it\Phi \\ 0 & 1 \end{pmatrix} = 0, \tag{5.15}$$

where $\mathscr{E}$, Φ, $\mathscr{E}'$, and Φ' are to be evaluated at $\rho = 0$, $z = (2t)^{-1}$.

If $u(t)$ is a rational function of t, then the method of solution of the HHP is exactly the same here as in the vacuum case. Similarly, if $v(t)$ is a rational matrix function of t and if the unimodular matrix

$$u(t) = [\det v(t)]^{-1/3}v(t) \tag{5.16}$$

also satisfies conditions (5.12b)–(5.12d), then the solution of the HHP proceeds along exactly the same lines as for a finite product of Harrison transformations in vacuum. Such transformations with $\det v(t)$ not a perfect cube will be new as none of the finite electrovac transformations discussed previously take the form (5.16).[35]

Despite these obvious similarities with vacuum, the single Harrison transformation with $\det v(t) = 1 - st^{-1}$ does not automatically generalize to electrovac. The simplest transformation of the form (5.16) that we find has two s parameters, complex conjugates of each other, and so more closely resembles the *double* Harrison transformation. This unexpected complication can be traced to condition (5.12b) being quadratic in $u(t)$. [This condition is needed to guarantee the algebraic relation between $F(t)$ and $F^*(t)$ given in Refs. 18, 19, and 32.] Thus, while we want $\det v(t)$ to be *not* a perfect cube, the product $[\det v(t)]\,[\det v^*(t)]$ must be a perfect cube. The simplest choices subject to these constraints take such (essentially equivalent) forms as

$$\det v(t) = \frac{t - s^*}{t - s} \tag{5.17a}$$

or

$$\det v(t) = \frac{s(t - s^*)}{s^*(t - s)}, \tag{5.17b}$$

or

$$\det v(t) = (1 - s/t)^2(1 - s^*/t), \tag{5.17c}$$

s being a complex constant. With $\det v(t)$ given by Eq.

(5.17b), the simplest possible choice for $u(t)$ is

$$u(t)=\left(\frac{s^*(t-s)}{s(t-s^*)}\right)^{1/3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&\dfrac{s(t-s^*)}{s^*(t-s)}\end{pmatrix}. \tag{5.18}$$

This matrix function of t has a branch cut from $t=s$ to $t=s^*$ inside L_+ and appears to represent a very special case of the electrovac version of the double Harrison transformation with no free c parameters. An instructive way to proceed to a more general transformation with four c parameters is to solve the HHP for this special $u(t)$ first, and then recognize that the solution can be generalized in a nontrivial way by exploiting the gauge freedom in $F(t)$.

The solution of the HHP for the $u(t)$ of Eq. (5.18) is

$$F'(t)=\left(F(t)+\frac{t(s-s^*)}{s^*(t-s)}\frac{1}{\Delta}\begin{pmatrix}F_{13}(s^*)\\F_{23}(s^*)\\F_{33}(s^*)\end{pmatrix}(m_1(t),m_2(t),m_3(t))\right)\times\begin{pmatrix}1&0&0\\0&1&0\\0&0&\dfrac{s^*(t-s)}{s(t-s^*)}\end{pmatrix}, \tag{5.19}$$

where

$$m_b(t)=\begin{vmatrix}F_{11}(s)&F_{12}(s)&F_{1b}(t)\\F_{21}(s)&F_{22}(s)&F_{2b}(t)\\F_{31}(s)&F_{32}(s)&F_{3b}(t)\end{vmatrix}, \tag{5.20a}$$

$$\Delta=m_3(s^*). \tag{5.20b}$$

Details of the method of solution of this HHP will be supplied later for a more general $u(t)$. [*Important note*: To calculate $F^{*\prime}(t)$, replace s, s^*, $F(t)$, $F(s)$, and $F(s^*)$ by s^*, s, $F^*(t)$, $F^*(s^*)$, and $F^*(s)$, respectively, in the right-hand sides of Eqs. (5.19), (5.20a), and (5.20b). In particular, $m^*_b(t)$ and Δ^* are the complex conjugates of $m_b(t^*)$ and Δ, respectively.] The components of the first row of $H'=\dot F'(0)$ are

$$\mathscr{E}'=H'_{11}=H_{11}-\frac{i(s-s^*)}{ss^*\Delta}F_{13}(s^*)\begin{vmatrix}F_{11}(s)&F_{12}(s)\\F_{31}(s)&F_{32}(s)\end{vmatrix}, \tag{5.21a}$$

$$H'_{12}=H_{12}-\frac{i(s-s^*)}{ss^*\Delta}F_{13}(s^*)\begin{vmatrix}F_{21}(s)&F_{22}(s)\\F_{31}(s)&F_{32}(s)\end{vmatrix}, \tag{5.21b}$$

$$\Phi'=H'_{13}=H_{13}-\frac{s-s^*}{ss^*\Delta}F_{13}(s^*)\begin{vmatrix}F_{11}(s)&F_{12}(s)\\F_{21}(s)&F_{22}(s)\end{vmatrix}. \tag{5.21c}$$

If the initial solution is a vacuum solution, for which $F_{a3}(t)=0=F_{3b}(t)$, $a,b=1,2$, $F_{33}(t)=1$, then the above transformation reduces to the identity. On the other hand, if the solution has a nonvanishing electromagnetic field, then Eqs. (5.21a)–(5.21c) define a nontrivial new transformation.

Four additional parameters can be incorporated into the above transformation by simply changing the gauge of the $F(t)$ potential in the right-hand sides of Eqs. (5.21a)–(5.21c) or by making a similar substitution for $F(s)$ and $F(s^*)$, but not for $F(t)$, in the right-hand side of Eq. (5.19). The differential equations and algebraic relations[18,19,32] which govern $F(t)$ are also satisfied after a substitution

$$F(t)\to F(t)g(t), \tag{5.22}$$

where $g(t)$ is a complex-valued function of t only subject to

$$g(0)=I,\quad \det g(t)=1,\quad g^\dagger(t)\mathfrak{S}(t)g(t)=\mathfrak{S}(t), \tag{5.23}$$

I denoting the 3×3 unit matrix. Since we wish to obtain formulas preserving special HE gauge, we shall make this substitution for $F(s)$ and $F(s^*)$ in Eqs. (5.21a) and (5.21c) and then deduce the new $F'(t)$ from the HHP, rather than Eq. (5.19).

Not all eighteen components of $g(s)$ and $g(s^*)$ will appear in the right-hand sides of Eqs. (5.21a) and (5.21c). The latter depend on only the four independent ratios h_a/h_b and g_a/g_b, $a,b=1,2,3$, of the components of the following vectors:

$$\mathbf{h}=\begin{pmatrix}h_1\\h_2\\h_3\end{pmatrix}=g(s^*)\begin{pmatrix}0\\0\\1\end{pmatrix}, \tag{5.24a}$$

$$\mathbf{g}=(g_1,g_2,g_3)=(0,0,1)g(s)^{-1}. \tag{5.24b}$$

The algebraic conditions (5.23) imply

$$h^*_1=-\tfrac{1}{2}isg_2,\quad h^*_2=\tfrac{1}{2}isg_1,\quad h^*_3=g_3. \tag{5.25}$$

Since the vectors h_a and g_b may be arbitrarily rescaled, there is no loss of generality in putting, for example, $g_3=1=h_3$. Inspection of later formulas suggests, however, that a more convenient normalization would be to take the quantity

$$\begin{aligned}E&\equiv g_1h_1+g_3h_3+(s/s^*)g_2h_2\\&=g_3g^*_3+\tfrac{1}{2}is^*g_1g^*_2-\tfrac{1}{2}isg_2g^*_1=E^*\end{aligned} \tag{5.26}$$

to be unity.

When the substitution (5.22) is made for $F(s^*)$ and $F(s)$ in the right-hand sides of Eqs. (5.21a) and (5.21c), the transformed Ernst potentials become

$$\mathscr{E}'=H'_{11}=H_{11}+[i(s-s^*)/ss^*\Delta]h_bF_{1b}(s^*)k_2, \tag{5.27a}$$

$$\Phi'=H'_{13}=H_{13}-[(s-s^*)/ss^*\Delta]h_bF_{1b}(s^*)k_3, \tag{5.27b}$$

where $\mathbf{k}=(k_1,k_2,k_3)$ is the row vector

$$\mathbf{k}=-S(s)\mathbf{g}F(s)^{-1}, \tag{5.28}$$

whose components are

$$k_1=\begin{vmatrix}g_1&g_2&g_3\\F_{21}(s)&F_{22}(s)&F_{23}(s)\\F_{31}(s)&F_{32}(s)&F_{33}(s)\end{vmatrix}, \tag{5.29a}$$

$$k_2=-\begin{vmatrix}g_1&g_2&g_3\\F_{11}(s)&F_{12}(s)&F_{13}(s)\\F_{31}(s)&F_{32}(s)&F_{33}(s)\end{vmatrix}, \tag{5.29b}$$

$$k_3=\begin{vmatrix}g_1&g_2&g_3\\F_{11}(s)&F_{12}(s)&F_{13}(s)\\F_{21}(s)&F_{22}(s)&F_{23}(s)\end{vmatrix}, \tag{5.29c}$$

and

$$\Delta=\mathbf{k}F(s^*)\mathbf{h}=\begin{vmatrix}g_1&g_2&g_3&0\\F_{11}(s)&F_{12}(s)&F_{13}(s)&h_bF_{1b}(s^*)\\F_{21}(s)&F_{22}(s)&F_{23}(s)&h_bF_{2b}(s^*)\\F_{31}(s)&F_{32}(s)&F_{33}(s)&h_bF_{3b}(s^*)\end{vmatrix}, \tag{5.30}$$

and the summation convention applies to the repeated index b.

Equations (5.27a) and (5.27b) as they stand are sufficient to determine the metric and electromagnetic potential of the transformed solution, but the transformed generating function $F'(t)$ and/or representing matrix $u(t)$ is also needed if we wish to iterate these transformations. If we restrict Eqs. (5.27a) and (5.27b) to the z axis, using Eqs. (5.9a) and (5.9b), and then substitute into Eq. (5.15), we find that $u(t)$ is uniquely determined up to a scalar multiplicative factor. The latter is then uniquely determined (up to a multiplicative cube root of unity) by the unit determinant condition (5.12a). The result of this calculation is

$$u(t) = (s^*(t-s)/s(t-s^*))^{1/3} w(t), \tag{5.31}$$

where

$$w(t) = I + \frac{s(s-s^*)}{s^* E(t-s)} \begin{pmatrix} h_1 \\ h_2 t/s^* \\ h_3 \end{pmatrix} (g_1 t/s, g_2, g_3 t/s), \tag{5.32}$$

E given by Eq. (5.26). The earlier $u(t)$ given by Eq. (5.18) can be recovered by putting $g_1 = g_2 = 0 = h_1 = h_2$. When the parameters satisfy Eq. (5.25), the second condition (5.12b) is satisfied automatically and provides a quick method of calculation of the inverse:

$$w(t)^{-1} = I - \frac{s-s^*}{E(t-s^*)} \begin{pmatrix} h_1 \\ h_2 t/s^* \\ h_3 \end{pmatrix} (g_1 t/s, g_2, g_3 t/s). \tag{5.33}$$

By absorbing the cubic surd in Eq. (5.31) into $X_-(t)$, the HHP implies that $F'(t)w(t)F(t)^{-1}$ is analytic everywhere (including at $t = \infty$) except for a simple pole at $t = s$. Hence, we can write

$$F'(t)w(t)F(t)^{-1} = A(I + [t/(t-s)]B), \tag{5.34}$$

where A and B are constant matrices to be determined. Condition (5.6a) at $t = 0$ gives immediately

$$A = I + \frac{s-s^*}{s^* E} \begin{pmatrix} 0 \\ h_1 \\ ih_3 \end{pmatrix} (g_2, 0, 0). \tag{5.35}$$

Solving Eq. (5.34) for $F'(t)$, we have

$$F'(t) = A\left(I + \frac{t}{t-s} B\right) F(t) w(t)^{-1}. \tag{5.36}$$

The conditions that the poles at $t = s$ and $t = s^*$ on the right-hand side of Eq. (5.36) be absent are, respectively,

$$BF(s)\left\{I - \frac{1}{E}\begin{pmatrix} h_1 \\ h_2 s/s^* \\ h_3 \end{pmatrix}\mathbf{g}\right\} = 0, \tag{5.37a}$$

$$\left(I - \frac{s^*}{s-s^*} B\right) F(s^*)\mathbf{h} = 0. \tag{5.37b}$$

Equations (5.37a) and (5.29) show that B is expressible as the outer product of a column vector and a row vector, the latter being proportional to $\mathbf{k}$. Then Eq. (5.37b) shows that the column vector in question is proportional to $F(s^*)\mathbf{h}$. Hence we find

$$B = [(s-s^*)/s^*\Delta\,] F(s^*)\mathbf{hk}, \tag{5.38}$$

which completes the solution of the HHP. An alternative form of the solution (5.36) is

$$F'(t) = A\left[F(t) + \frac{t(s-s^*)}{s^*(t-s)} \frac{1}{\Delta} F(s^*)\mathbf{h}\mathbf{m}(t)\right] w(t)^{-1}, \tag{5.39}$$

with A, Δ, and $w(t)^{-1}$ given by Eqs. (5.35), (5.30), and (5.33), respectively, and the components of $\mathbf{m}(t) = (m_1(t), m_2(t), m_3(t))$ given by

$$m_b(t) = \begin{vmatrix} g_1 & g_2 & g_3 & 0 \\ F_{11}(s) & F_{12}(s) & F_{13}(s) & F_{1b}(t) \\ F_{21}(s) & F_{22}(s) & F_{23}(s) & F_{2b}(t) \\ F_{31}(s) & F_{32}(s) & F_{33}(s) & F_{3b}(t) \end{vmatrix}. \tag{5.40}$$

Observe that

$$\mathbf{m}(0) = i\mathbf{k}\mathfrak{E}(2), \tag{5.41a}$$

$$\mathbf{m}(s) = -\mathbf{g}/S(s), \tag{5.41b}$$

$$\mathbf{m}(s^*)\cdot\mathbf{h} = \Delta. \tag{5.41c}$$

From Eqs. (5.6a) and (5.6b), the transformed H potential is found to be

$$H' = A\left[iH\mathfrak{E}(2) - \frac{s-s^*}{ss^*\Delta} F(s^*)\mathbf{hk}\right] A^{-1} i\mathfrak{E}(2) + \frac{s-s^*}{ss^*E} \begin{pmatrix} 0 & ig_2h_2 s/s^* & 0 \\ -ig_1h_1 & -ig_2h_1 \dfrac{s}{s^*E}\mathbf{g}\cdot\mathbf{h} & -ig_3h_1 \\ g_1h_3 & g_2h_3 \dfrac{s}{s^*E}\mathbf{g}\cdot\mathbf{h} & g_3h_3 \end{pmatrix}. \tag{5.42}$$

The transformation equations (5.39) and (5.42) preserve special HE gauge and the 11 and 13 components of Eq. (5.42) reduce to Eqs. (5.27a) and (5.27b). The 12 component of the transformed H potential is

$$H'_{12} = H_{12} - i\frac{s-s^*}{ss^*\Delta} h_b F_{1b}(s^*)k_1 + \frac{s-s^*}{s^*E} g_2(h_1 H'_{11} + h_3 H'_{13}) + \frac{i(s-s^*)}{s^{*2}E} g_2 h_2. \tag{5.43}$$

Now, the transformation formula,

$$F'(t) = \left[F(t) + \frac{t(s-s^*)}{s^*(t-s)} \frac{1}{\Delta} F(s^*)\mathbf{h}\mathbf{m}(t)\right] \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{s^*(t-s)}{s(t-s^*)} \end{pmatrix}, \tag{5.44}$$

Δ given by Eq. (5.30), $\mathbf{m}(t)$ by Eq. (5.40), is the result of making the substitution (5.22) for $F(s)$ and $F(s^*)$ but not for $F(t)$ in the formula (5.19). This $F'(t)$ is also a bona fide $F(t)$-potential in HE gauge satisfying all of the defining equations,[18,19,32] but is not in special HE gauge. It gives the same Ernst potentials as Eqs. (5.27a) and (5.27b), and H'_{12} is given by the first two terms on the right-hand side of Eq. (5.43). One should

expect to find a simple relationship between the transformations (5.44) and (5.39). Besides the gauge transformations of the form (5.22), we also need the trivial translations, $\omega \to \omega +$ real constant, $\varphi_2 \to \varphi_2 +$ complex constant, which are outside the gauge group. These are the $\gamma_{22}^{(0)}$ and $c_2^{(0)}$ transformations of Ref. 14 (or $\gamma_{11}^{(0)}$ and $c_1^{(0)}$ according to our own convention[25]) and are easily exponentiated to give

$$F'(t) = \begin{pmatrix} 1 & 0 & 0 \\ -\gamma & 1 & 0 \\ 2c^* & 0 & 1 \end{pmatrix} F(t) \begin{pmatrix} 1 & -\gamma - icc^*t & ct \\ 0 & 1 & 0 \\ 0 & -2ic^* & 1 \end{pmatrix}, \tag{5.45}$$

γ real, c complex. The transformation law (5.39) preserving special HE gauge can be achieved by appplying first the transformation (5.44), then (5.45) with γ and c chosen so that H'_{2b} and H'_{3b} become constant on the z axis, and finally (5.22). A direct evaluation of γ, c, and $g(t)$ starting from Eq. (5.44) is rather tedious, but it is easy to calculate their values by working backwards from Eq. (5.39). The results are

$$\gamma = -\frac{s - s^*}{s^* E} g_2 h_1, \quad c = \frac{s - s^*}{s s^* E} g_3 h_1,$$
$$c^* = \tfrac{1}{2} i \frac{s - s^*}{s^* E} g_2 h_3; \tag{5.46}$$

$g(t)$

$$= \begin{pmatrix} 1 & \gamma - icc^*t & -ct \\ 0 & 1 & 0 \\ 0 & 2ic^* & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{s(t - s^*)}{s^*(t - s)} \end{pmatrix} w(t)^{-1}. \tag{5.47}$$

Note that γ is real and $g(t)$ satisfies the conditions (5.23).

Generation of Kerr–Newman–NUT space

Let us calculate the effect of the transformation (5.27a) and (5.27b) on flat space, $\mathscr{E} = 1$, $\Phi = 0$. The $F(t)$ potential for flat space is[16]

$$F(t) = \begin{pmatrix} \dfrac{t}{S(t)} & \dfrac{i}{S(t)} & 0 \\ \dfrac{1 - 2tz + S(t)}{2iS(t)} & \dfrac{1 - 2tz - S(t)}{2tS(t)} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.48}$$

It is convenient to use oblate spheroidal coordinates $(\bar{x}, y)$ defined by

$$\rho = \bar{\kappa}(\bar{x}^2 + 1)^{1/2}(1 - y^2)^{1/2}, \quad z = \bar{\kappa}\bar{x}y + z_0, \tag{5.49a}$$
$$z_0 - i\bar{\kappa} = (2s)^{-1}, \quad z_0 + i\bar{\kappa} = (2s^*)^{-1}, \quad \mathrm{Im}\, s > 0. \tag{5.49b}$$

Then

$$S(s) = -\frac{\bar{\kappa}(\bar{x} + iy)}{z_0 - i\bar{\kappa}}, \quad S(s^*) = -\frac{\bar{\kappa}(\bar{x} - iy)}{z_0 + i\bar{\kappa}}, \tag{5.50}$$

where the signs have been chosen so that $F(s)$ and $F(s^*)$ reduce to the form (5.9b) on the branch $y = +1$ of the z axis. A straightforward substitution into Eqs. (5.27a) and (5.27b) gives

$$\mathscr{E}' = \frac{g_1 h_1(\bar{x} + i) + g^*{}_1 h^*{}_1(\bar{x} - i) + \tfrac{1}{2} i g_1 g^*{}_1(1 + y) - 2ih_1 h^*{}_1(1 - y) + g_3 g^*{}_3(\bar{x} - iy)}{g_1 h_1(\bar{x} - i) + g^*{}_1 h^*{}_1(\bar{x} + i) - \tfrac{1}{2} i g_1 g^*{}_1(1 - y) + 2ih_1 h^*{}_1(1 + y) + g_3 g^*{}_3(\bar{x} - iy)}, \tag{5.51a}$$

$$\Phi' = \frac{ig_3(g^*{}_1 + 2h_1)}{g_1 h_1(\bar{x} - i) + g^*{}_1 h^*{}_1(\bar{x} + i) - \tfrac{1}{2} i g_1 g^*{}_1(1 - y) + 2ih_1 h^*{}_1(1 + y) + g_3 g^*{}_3(\bar{x} - iy)}, \tag{5.51b}$$

where we have used Eqs. (5.25) to eliminate g_2, h_2, and h_3.

The solution (5.51a) and (5.51b) is recognizable as the Kerr–Newman–NUT space with magnetic charge. It can be generated from the vacuum Kerr solution by means of the SU(2, 1) Kinnersley group.[36] The appearance of *oblate* spheroidal coordinates indicates that the Kerr solution in question is without horizons, i.e., beyond the extreme ($|a| > m$ in a familiar parametrization[37]). A standard form[38,39] for the vacuum Kerr solution[17] is

$$\xi = px + iqy, \quad \kappa = mp, \quad a = mq, \tag{5.52}$$

where $\xi = (1 + \mathscr{E})/(1 - \mathscr{E})$,[27] $p^2 + q^2 = 1$, and (x, y) are *prolate* spheroidal coordinates with foci at $z = z_0 \pm \kappa$. This solution can be obtained from flat space by applying two Harrison transformations with real s parameters.[6,40] The beyond the extreme case, $|q| > 1$, where the s parameters form a complex conjugate pair, can be expressed in *oblate* spheroidal coordinates $(\bar{x}, y)$ defined by Eq. (5.49a) and obtainable from the prolate case by the substitutions, $x = i\bar{x}, \kappa = -i\bar{\kappa}, p = -i\bar{p}, \bar{p}^2 = q^2 - 1$. Thus the Ernst potentials are

$$\mathscr{E} = \frac{\bar{p}\,\bar{x} + iqy - 1}{\bar{p}\,\bar{x} + iqy + 1}, \quad \Phi = 0. \tag{5.53}$$

Now apply the Ehlers transformation,[13,36]

$$\mathscr{E}' = \frac{\mathscr{E} + i\lambda}{1 + i\lambda\mathscr{E}}, \quad \Phi' = \frac{(1 + i\lambda)\Phi}{1 + i\lambda\mathscr{E}}, \tag{5.54}$$

λ real, which introduces a NUT parameter, and the charging transformation of Harrison[41] and Ernst,[33]

$$\mathscr{E}' = \frac{\mathscr{E} + 2b^*\Phi - bb^*}{1 - 2b^*\Phi - bb^*\mathscr{E}},$$
$$\Phi' = \frac{b(\mathscr{E} - 1) + (1 + bb^*)\Phi}{1 - 2b^*\Phi - bb^*\mathscr{E}}, \tag{5.55}$$

b complex, in either order to the Kerr solution (5.53). The result is[42]

$$\mathscr{E}' = \frac{(1 - i\lambda bb^*)(\bar{p}\bar{x} + iqy - 1) + (i\lambda - bb^*)(\bar{p}\bar{x} + iqy + 1)}{(1 - i\lambda bb^*)(\bar{p}\bar{x} + iqy + 1) + (i\lambda - bb^*)(\bar{p}\bar{x} + iqy - 1)}, \tag{5.56a}$$

$$\Phi' = -\frac{2b(1 + i\lambda)}{(1 - i\lambda bb^*)(\bar{p}\bar{x} + iqy + 1) + (i\lambda - bb^*)(\bar{p}\bar{x} + iqy - 1)}. \tag{5.56b}$$

It is easy now to identify the solution (5.51a) and (5.51b) with (5.56a) and (5.56b) although the relations connecting the parameters are rather messy. Also, having obtained the solution (5.51a) and (5.51b) in oblate spheroidal coordinates ($|q| > 1$), it is a trivial matter to analtyically continue to the prolate case ($|q| < 1$). The six parameters, z_0, κ (or $\bar{\kappa}$), q, λ, Re b, Im b, determine, respectively, the following physical characteristics: position on z axis, mass, angular momentum, NUT parameter, electric charge, and magnetic charge.

It is to be expected that n successive applications of the new transformation of this section to flat space would give the nonlinear superposition of n Kerr–Newman–NUT particles on the z axis. This would generalize the known results for vacuum Kerr–NUT particles (Kramer and Neugebauer[43]) and electrically poised Kerr–Newman–NUT particles (Kobiske and Parker[44]). From Kramer and Neugebauer's work, it is clear that it is not necessary that Coulomb repulsion balance gravitational attraction for the conical stresses to be absent between the particles (assuming NUT singularities have already been removed) as a spin-induced "magnetic-type" gravitational force is also present and is able to balance the more familiar "electric-type" gravity even in the case of uncharged Kerr particles.

At present, the only known examples of two or more *black holes*, for which all spacetime singularities are enclosed by nonsingular even horizons, in equilibrium under mutual gravitational and electromagnetic interactions is the superposition of n static *extreme* Reissner–Nordström holes.[45] The two cases mentioned in the preceding paragraph are beyond the extreme (no horizons) when axial stresses are absent (except for the static limit of the Kobiske–Parker solution which consists of Reissner–Nordstrom holes). Preliminary calculations, however, suggest that it may be possible to balance two nonextreme Kerr–Newman black holes for certain range of the charges, masses, angular momenta, and spatial separation. Further work need to be done to clarify this interesting problem.

6. CONCLUSION

In Secs. 3 and 4 of this paper, we have explored the infinite-dimensional Geroch group **K** of transformations for vacuum spacetimes in the representation of Hauser and Ernst.[18–20] The main result of those sections is that Harrison's Bäcklund transformation[8] is in **K** (though not a subgroup) and that elements of **K** represented by $u(t)$ matrices of the form

$$u(t) = [\det v(t)]^{-1/2} v(t), \tag{6.1}$$

$v(t)$ being a rational matrix function of t, are in fact products of a finite number of Harrison transformations. Each quadratic branch point of $u(t)$ of index $\pm\frac{1}{2}$ in L_+ not at the origin corresponds to an individual Harrison transformation ($t = 0$ is also a branch point when there are an odd number of the latter). This result in a sense generalizes the earlier result of Hauser and Ernst[18] that simple poles of $u(t)$ correspond to null HKX transformations.[3,4] The coalescence of two quadratic branch points to form a simple pole at $t = s$ provides an easy proof of the theorem[12] that the null HKX transformation factorizes into two Harrison transformations with same s parameters.

We have chosen the Hauser–Ernst (HE) representation for **K** instead of the earlier representations of Geroch[2] and Kinnersley and Chitre[13–16] because we needed the transformations to be already exponentiated. Since the Harrison transformations are not subgroups of **K**, it would be rather difficult to give an adequate description of them in terms of infinitesimal generators of **K**. Furthermore, we have found the complex-variable techniques of Hauser and Ernst relatively easy to handle and often lead to significantly simpler computations. In Sec. 4, we exploited the already exponentiated property to exhibit finite transformations in the **B** group and nonnull groups of Kinnersley and Chitre in terms of products of Harrison and/or HKX transformations.

The reader interested in stationary axisymmetric gravitational fields is now confronted with a considerable variety of solution-generating techniques. One impression that the author hopes has been gained from Refs. 6 and 12 is that it is advantageous to be familiar with all of the available methods and their interrelationships as there are situations in which each is best suited. With regard to the HE formalism, for example, one can think of situations where the formalism is: (i) the obvious or only one to consider (e.g., applications in the present paper and Ref. 20); (ii) is more manageable than its competitors (e.g., most applications in Refs. 18 and 19; composition of several transformations in **K**); (iii) is less manageable than its competitors (e.g., SL(2)-covariant manipulations of Belinsky–Zakharov transformations[9] as in Secs. 5 and 6 of Ref. 12); or (iv) is not an applicable method (e.g., applications of the **Q** and $\tilde{\mathbf{Q}}$ groups[5–7,12] or, equivalently, Neugebauer's Bäcklund transformations,[11,40,43] which are

not in **K**).

Although the HE formalism does not pretend to represent transformations outside **K**, it can nevertheless be used to provide more satisfying proofs that such transformations are indeed outside **K**. For example, the **Q** group preserves asymptotic flatness and the transform of the $F(t)$ potential given by Eqs. (2.30a)–(2.31b) of Ref. 12 is easily shown to preserve special HE gauge.[46] It follows that there is an infinity of elements of **K** which transform a given solution $\mathscr{E}(\rho, z)$ to $\mathscr{E}'(\rho', z') = (Q)_{4s}\mathscr{E}(\rho, z)$, $(Q)_{4s} \in \mathbf{Q}$. Restricting Eqs. (2.30a)–(2.31b) and (2.2a) and (2.2b) of Ref. 12 to the symmetry axis ($\rho = 0 = \rho'$), we find

$$\mathscr{E}'(0, z') = \mathscr{E}(0, z), \quad z' = z/(1 - 2sz). \tag{6.2}$$

Similarly for $(\tilde{Q})_{4s} \in \tilde{\mathbf{Q}}$, we find[12]

$$\mathscr{E}'(0, z') = (1 - 2sz)^{-1}\mathscr{E}(0, z), \quad z' = z/(1 - 2sz). \tag{6.3}$$

When these values are substituted into Eq. (2.25) here, it is clearly impossible to choose a matrix $u(t)$ which is not explicitly dependent on the initial solution. Thus neither $(Q)_{4s}$ nor $(\tilde{Q})_{4s}$ can be identified with any of the elements of **K** which map $\mathscr{E}$ to $(Q)_{4s}\mathscr{E}$ or to $(\tilde{Q})_{4s}\mathscr{E}$, respectively. The same comments apply to other subgroups of **Q** and $\tilde{\mathbf{Q}}$, in particular, the trivial rescaling $(\rho, z) \to (k\rho, kz)$ and translation $(\rho, z) \to (\rho, z - z_0)$. One corollary of the result that the Harrison transformation is in **K** and factorizes into **Q** and $\tilde{\mathbf{Q}}$ transformations[12] is that any product of elements of **Q**, $\tilde{\mathbf{Q}}$, and **K** is contained in **K** if and only if the combined transformation leaves (ρ, z) fixed.

A reasonable question to ask is whether all known transformations which leave (ρ, z) fixed are members of **K**. The trivial reflection $\mathscr{E} \to \mathscr{E}^*$, which reverses the sense of rotation, is easily shown to be not in **K** by the same argument as in the previous paragraph. The Kramer–Neugebauer mapping[47] [see Ref. 12, Sec. 3, for the transform of $F(t)$] and the Belinsky–Zakharov single-soliton transformation[9,12] create curvature singularities along the whole z axis and generate complex-valued metrics from real-valued and so are necessarily outside the framework of all existing representations of **K**. However, preliminary calculations suggest that these transformations may be obtainable as limits of sequences of bona fide elements of **K** and should therefore be accepted as being in **K** themselves. These results will be included in a separate paper when details are finalized. A simpler example of such a limiting class of group elements is the $s = \infty$ limit of the HKX transformation.[12]

In Sec. 5, we attempted to generalize Harrison's Bäcklund transformation to electrovac spacetimes by studying elements of **K**′ for which

$$u(t) = [\det v(t)]^{-1/3} v(t) \tag{6.4}$$

in the HE representation, where $v(t)$ is a 3×3 matrix whose entries are rational functions of t. On comparison with Eq. (1.1) for the vacuum Harrison transformation, it is natural to expect that a choice of $v(t)$ such that $u(t)$ has cubic branch points at $t = 0$ and one other point, say $t = s$, in L_+ joined by a cut is a reasonable candidate for the electrovac Harrison transformation. Unfortunately, no such choice is compatible with condition (5.12b) which is quadratic in $u(t)$. The simplest choices for $v(t)$ compatible with conditions (5.12a)–(5.12d) have determinants of the forms (5.17a)–(5.17c) or similar forms with the result that $u(t)$ has cubic branch points in L_+ at $t = s$ and $t = s^*$, s complex, joined by a cut. The Bäcklund transformation deduced therefrom [see Eq. (5.39)] is obviously an electrovac enlargement of the *double* Harrison transformation with complex conjugate s parameters. (When $g_3 = 0 = h_3$, the transformation (5.39) maps vacuum to vacuum and is precisely the double Harrison transformation.) We proved that this transformation maps flat space to the full family of Kerr–Newman–NUT spacetimes with six parameters: mass, angular momentum, NUT parameter, electric charge, magnetic charge, and position on z axis.

The s parameters being complex conjugates in Eq. (5.39) gave us the Kerr–Newman–NUT solution in oblate spheroidal coordinates, i.e., beyond the extreme ($a^2 + e^2 > m^2$,[37] $|q| > 1$[39]). The prolate case ($a^2 + e^2 < m^2$, $|q| < 1$), which has horizons at $x = \pm 1$, can be obtained by a trivial analytic continuation of the parameters. One is tempted to consider an analytic continuation of the parameters in Eq. (5.39) in order to define a corresponding transformation with two real s parameters (e.g., by formally introducing a second imaginary unit, j say, subject to $i^2 = j^2 = -1$, $i^* = -i$, $j^* = j$), but it is not obvious how this may be achieved.

Of course, every vacuum-to-vacuum transformation in **K** is the restriction to vacuum of an infinite number of elements of **K**′. Suppose $F_{\text{vac}}(t)$ is a given vacuum $F(t)$ potential and $u_{\text{vac}}(t)$ a given 2×2 matrix obeying conditions (2.18a), (2.18b), and (2.19). Then the 3×3 $F(t)$ potential and corresponding elements of **K**′ are given by

$$F(t) = \begin{pmatrix} F_{\text{vac}}(t) & 0 \\ 0 & 1 \end{pmatrix}, \tag{6.5a}$$

$$u(t) = \begin{pmatrix} e^{-i\theta(t)}u_{\text{vac}}(t) & 0 \\ 0 & e^{2i\theta(t)} \end{pmatrix}, \tag{6.5b}$$

where $\theta(t) = \theta^*(t)$ is an arbitrary function of t, analytic on L, throughout L_-, and at $t = \infty$. If $u_{\text{vac}}(t)$ is chosen as the representing matrix of the single Harrison transformation [Eq. (1.1)], then it is still not possible to choose $\theta(t)$ such that Eq. (6.5b) takes the form (6.1) with algebraic branch points only at $t = 0$ and $t = s$. On the other hand, if we simply put $\theta(t) = 0$, say, the HHP cannot be solved by the methods of this paper when the given electrovac solution has a nontrivial electromagnetic field.

Perhaps the most positive evidence for the existence of an electrovac counterpart to the *single* Harrison transformation can be gleaned from recent work of Kinnersley and Lemley[48] on the electrovac counterparts of the **Q** and $\tilde{\mathbf{Q}}$ groups[5–7,12] (or equivalently, I_1 and I_2 Bäcklund transformations,[11] respectively). The arguments which follow are of a speculative nature and are based on analogy with known results in vacuum: the quantitative details should be filled in in the near future. First, note that in Refs. 6 and 12, the vacuum Harrison (H) transformation has been factorized into the forms

$$H = I_2 I_1 = (L)_\beta (\tilde{Q})_{-4s} (P)_\alpha (Q)_{4s}, \tag{6.6a}$$

$$H = I_1 I_2 = (P)_{\bar{\alpha}} (Q)_{-4s} (L)_{\bar{\beta}} (\tilde{Q})_{4s}; \tag{6.6b}$$

$(P)_\alpha \in \mathbf{P}$, $(L)_\beta \in \mathbf{L}$ (see Sec. 2); $(Q)_{\pm 4s} \in \mathbf{Q}$, $(\tilde{Q})_{\pm 4s} \in \tilde{\mathbf{Q}}$;[46]

$\alpha, \beta, \delta \in SL(2, R)$; $I_1 = (P)_\alpha (Q)_\delta = (Q)_\delta (P)_\alpha$, $I_2 = (L)_\beta (\tilde{Q})_\delta = (\tilde{Q})_\delta (L)_\beta$. The four groups, **P**, **L**, **Q**, and $\tilde{\mathbf{Q}}$, each locally isomorphic to SL(2, R), each contain one nontrivial parameter and two gauge parameters. This fact, together with the algebraic relations between α and β and between $\bar{\alpha}$ and $\bar{\beta}$,[12] limits the vacuum H transformation to two nontrivial parameters.

Let us now try to interpret Eqs. (6.6a) and (6.6b) in the electrovac context. The electrovac counterpart of the **P** group is the well-known SU(2, 1) Kinnersley group[13,36] (denoted **H'** in Ref. 13) with eight real parameters. Thus α would now be a 3×3 pseudounitary matrix with unit determinant. This group contains three nontrivial parameters [Ehlers transformation (5.54) and Ernst–Harrison charging transformation (5.55)] and five gauge parameters. [If applied to a vacuum solution, the SU(2, 1) group **P** only provides two nontrivial parameters as the phase of b in Eq. (5.55) gives rise to an electromagnetic duality rotation, which is already one of the gauge transformations in **P**.] The full eight-parameter electrovac counterpart of the **L** group is not so well known as it appears that no eight generators in **K'**, three being the $\gamma^{(0)}_{AB}$ which generate a unimodular linear transformation of the Killing vectors,[14] will close to form a representation of the Lie algebra of SU(2, 1). Recently, Kinnersley (private communication) has found an SU(2, 1) group by dropping the requirement that the generators preserve the reality of the metric f_{AB} and the electromagnetic potential $A_A = \frac{1}{2}(\varphi_A + \varphi^*_A)$, and other real potentials. So the **L** group is available if we allow such a complex extension. Next, the electrovac counterparts of **Q** and $\tilde{\mathbf{Q}}$, presumed to exist, should be locally isomorphic to SL(2, R) as they would necessarily transform the coordinates (ρ, z) exactly according to Eqs. (2.2a) and (2.2b) of Ref. 12 [transformations which leave (ρ, z) invariant are not considered to belong to **Q** or $\tilde{\mathbf{Q}}$ unless, for convenience, we use the latter symbols for the larger groups, $I_1 = \mathbf{PQ}$, $I_2 = \mathbf{L}\tilde{\mathbf{Q}}$, respectively]. Since $\tilde{\mathbf{Q}}$ is presumed to commute with **L**, it would not be surprising if $\tilde{\mathbf{Q}}$ did not preserve the reality of f_{AB}, A_A, and other real potentials.

The above comments imply that there should exist for electrovac spacetimes two eleven-dimensional groups, $I_1 = \mathbf{PQ}$ and $I_2 = \mathbf{L}\tilde{\mathbf{Q}}$, each containing four nontrivial parameters and seven gauge parameters. (When applied to a vacuum solution, there would only be three nontrivial parameters, the magnetic charge not being counted.) According to a further analogy with vacuum, the expected algebraic relation between the SU(2, 1) matrices α and β, or between $\bar{\alpha}$ and $\bar{\beta}$, in Eqs. (6.6a) and (6.6b) would restrict the H transformation to at most four parameters (possibly only three). It is certain that this electrovac H transformation, if it exists, would generate complex-valued solutions from real-valued. Thus, it would be necessary to extend the complex-variable formalism of Hauser and Ernst to two complex dimensions (C^2) in order to incorporate the H transformation into their representation for **K'**. Such an analytic continuation to C^2 is straightforward in the Kinnersley–Chitre representation[13–16] (e.g., introduce an imaginary unit j and let the real and imaginary parts of the i-complex potentials be j-complex, $j^* = j$) but will probably require more serious thought for the HE formalism.

If the electrovac version of the single Harrison transformation contains the maximum four parameters then the double transformation must contain eight (seven if applied to a vacuum solution), two more than in our formula (5.39). In that case, it is difficult to imagine what the full transform of flat space would be. The only known solution containing Kerr–Newman–NUT space is the stationary charged C metric[42] with one extra parameter (acceleration), a cosmological constant being inadmissable here. However, this solution must be ruled out because the stationary *uncharged* C metric is the transform of flat space under *three* Harrison transformations, as can be seen by direct substitution into Eq. (2.25).

ACKNOWLEDGMENTS

This research has benefitted from discussions with William Kinnersley and Terry Lemley. I also wish to thank Isidore Hauser and Frederick J. Ernst for making their results available prior to publication.

[1] R. Geroch, J. Math. Phys. **12**, 918 (1971).
[2] R. Geroch, J. Math. Phys. **13**, 394 (1972).
[3] C. Hoenselaers, W. Kinnersley, and B. C. Xanthopoulos, J. Math. Phys. **20**, 2530 (1979); Phys. Rev. Lett. **42**, 481 (1979); B. C. Xanthopoulos, J. Math. Phys **22**, 1254 (1981). The authors have shown how their transformation (denoted HKX here) may be used to construct exact solutions with preassigned multipoles.
[4] An SL(2) tensor generalization of the HKX transformation appears in Sec. 6 of Ref. 12 below.
[5] C. M. Cosgrove, in *Gravitational Radiation, Collapsed Objects, and Exact Solutions*, Proceedings of the Einstein Centenary Summer School, edited by C. Edwards, Lecture Notes in Physics **124** (Springer-Verlag, Berlin, 1980), pp. 444–453.
[6] C. M. Cosgrove, in Proceedings of the Second Marcel Grossmann Meeting on the Recent Developments of General Relativity, Trieste, Italy, July 5–11, 1979.
[7] C. M. Cosgrove, Ph.D. Thesis, University of Sydney (1979) (unpublished).
[8] B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978); Phys. Rev. D **21**, 1695 (1980).
[9] V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)]; Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].
[10] D. Maison, Phys. Rev. Lett. **41**, 521 (1978); J. Math. Phys. **20**, 871 (1979).
[11] G. Neugebauer, J. Phys. A: Math. Gen. (Letters) **12**, L67 (1979); **13**, L19 (1980).
[12] C. M. Cosgrove, J. Math. Phys. **21**, 2417 (1980).
[13] W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[14] W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977).
[15] W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 1926 (1978).
[16] W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978); Phys. Rev. Lett. **40**, 1608 (1978).
[17] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[18] I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 362 (1979) (vacuum); Phys. Rev. D **20**, 1783 (1979) (electrovac).
[19] I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980) (vacuum); J. Math. Phys. **21**, 1418 (1980) (electrovac).
[20] I. Hauser and F. J. Ernst, J. Math. Phys. **22**, 1051 (1981); Abstracts of Contributed Papers for the Discussion Groups, 9th International Conference on General Relativity and Gravitation, Jena, German Democratic Republic, July 14–19, 1980, Vol. 1, pp. 89–90 (vacuum); F. J. Ernst and I. Hauser, *ibid*, Vol. 1, pp. 84–85 (electrovac).
[21] Definitions and methods of calculation of the $F(t)$ potential for a given vacuum solution are given in Refs. 3, 12, 15, 16, and 19. For the electrovac case, methods are given in Ref. 32 below and Ref. 18. (See also Footnotes 26 and 28 below.) $F(t)$ and $u(t)$ are 2×2 matrices [or second-rank SL(2) tensors] in vacuum and 3×3 matrices in electrovac.
[22] Geroch's conjecture (Ref. 2) is that **K** acts transitively on all stationary axisymmetric vacuum solutions. The theorem of Hauser and Ernst (Ref. 20) states that a certain large subgroup of **K** acts multiply transitively on

all solutions which are analytic in a neighborhood of a given point of the z axis. Convincing though nonrigorous arguments for the more restricted class of asymptotically flat solutions have been given by several previous authors.[3,6,11] Crude arbitrary-function-counting arguments suggest that the Geroch conjecture should not be true for the larger class of solutions which have logarithmic singularities along the whole z axis but it may be possible to generate such solutions from flat space by taking suitable limiting transitions of Geroch group elements.

[23]W. Kinnersley, private communication (1980). Kinnersley's argument, based on Eqs. (2.24) and (4.41) of Ref. 12, was that Harrison transformations near the identity could be expanded in terms of infinitesimal KC transformations, at least to a few terms.

[24]The closely related Bäcklund transformations of Cosgrove, Maison, and Neugebauer involve the spacetime coordinates (ρ, z). In Sec. 6, we use the HE formalism to demonstrate why transformations with this property cannot be in **K**. The Belinsky–Zakharov $(2n+1)$-soliton transformation and Kramer–Neugebauer mapping were asserted in Ref. 12 to be outside **K** because they create curvature singularities along the whole z axis. However, there is a possibility that these transformations can be incorporated in **K** by taking suitable limits as mentioned in Ref. 22. The Belinsky–Zakharov $2n$-soliton transformation is identifiable with a product of $2n$ Harrison transformations (Ref. 12) and so is covered by the present paper.

[25]The notation $\gamma_{XY}^{(k)}$ is in accordance with Eqs. (2.22) and (2.23) of Ref. 12. The symbol $\gamma_{XY}^{(k)}$ of Eqs. (3.3) of Ref. 14 (restricted to vacuum) has the same meaning as q_{XY} here and in Ref. 12. [Note that, for any SL(2) tensor, $q^{11}=q_{22}, q^{12}=-q_{21}, q^{21}=-q_{12}, q^{22}=q_{11}$.]

[26]When comparing SL(2) tensor equations with matrix equations, it is useful to know that the matrix inverse of R_{AB} is $R^{BA}/(\det R)$, and the inverse of $R_A{}^B$ is $-R^B{}_A/(\det R)$; $\det R=\frac{1}{2}R_{XY}R^{XY}$ and, more generally, $R_{XA}R^X{}_B=R_{AX}R_B{}^X=(\det R)\epsilon_{AB}$. Indices are raised and lowered with $\epsilon_{AB}=\epsilon^{AB}=\begin{pmatrix}0&1\\-1&0\end{pmatrix}$ [e.g., $R_A{}^B=\epsilon^{BX}R_{AX}$, $R_{AB}=\epsilon_{XA}R^X{}_B$]. The indices 1 and 2 used here correspond to Hauser and Ernst's 4 and 3, respectively. Consequently, to convert one of the matrices in Hauser and Ernst's work to the corresponding matrix here, one must interchange the two rows with each other and similarly the two columns.

[27]The sign of ψ as defined by Eq. (2.3b) is in agreement with Ref. 13, whose conventions we follow. In Refs. 5–7, 12, and 39, the opposite sign was used and so $\mathscr{E}$ there is to be identified with $\mathscr{E}^*$ here.

[28]$F(t)$ can also be calculated by the method of characteristics (Ref. 16) or from the Riccati equations for pseudopotentials (see the last paragraph of Appendix A of Ref. 12).

[29]Analyticity of $\mathscr{E}$ in a real (ρ, z) domain means that it must be possible to define $\mathscr{E}(\rho, z)$, a function of two *complex* variables ρ and z, which is analytic on and near the appropriate parts of the real axes. Taylor series can be used to give an intrinsic definition.

[30]C. Hoenselaers, J. Math. Phys. **17**, 1264 (1976).

[31]Due to a printing error, Eq. (7.31b) of Ref. 12 was omitted. It should read:

$$(\text{ext.KC}: s, \alpha)=\exp\left\{\sum_{k=-\infty}^{\infty}\alpha s^k\beta^{(k-1)}\right\}.$$

[32]T. C. Jones, J. Math. Phys. **21**, 1790 (1980).

[33]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).

[34]The symbol Ω_{11} is taken from Ref. 13. Also $\mathscr{E}$ here is the same as the $\mathscr{E}$ of Ref. 13.

[35]If the close relationship (Ref. 12) between the Harrison transformation and the Belinsky–Zakharov single-soliton transformation in vacuum is any guide, the electrovac transformations discussed here should bear some close relationship to the soliton transformations found recently by G. A. Aleksejev, Abstracts of Contributed Papers for the Discussion Groups, 9th International Conference on General Relativity and Gravitation, Jena, German Democratic Republic, July 14–19, 1980, Vol. 1, pp. 1–2; Pis'ma Zh. Eksp. Teor. Fiz. **32**, 301 (1980) [Sov. Phys. JETP Letters **32**, 277 (1980)].

[36]W. Kinnersley, J. Math. Phys. **14**, 651 (1973); G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969).

[37]B. Carter, Phys. Rev. **174**, 1559 (1968); R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).

[38]F. J. Ernst, Phys. Rev. **167**, 1175 (1968); J. Math. Phys. **15**, 1409 (1974); Ref. 33.

[39]A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973); M. Yamazaki, Prog. Theor. Phys. **57**, 1951 (1977); C. M. Cosgrove, J. Phys. A: Math. Gen. **10**, 1481 (1977).

[40]G. Neugebauer and D. Kramer, Exp. Techn. Physik **28**, 3 (1980).

[41]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).

[42]An exhaustive bibliography of type-D electrovac solutions related to the Kerr solution is given in J. Plebański and M. Demiański, Ann. Phys. (USA) **98**, 98 (1976). In addition ot the five parameters (six if z_0 counted) in Eq. (5.56), the full Plebański–Demiański solution contains a cosmological constant and an acceleration parameter [as in the "stationary C metric": W. Kinnersley, J. Math. Phys. **10**, 1195 (1969)].

[43]D. Kramer and G. Neugebauer, Phys. Lett. A **75**, 259 (1980).

[44]R. A. Kobiske and L. Parker, Phys. Rev. D **10**, 2321 (1974), and references therein. The general theory of stationary electrically poised particles in three space dimensions is due to G. Neugebauer, Habilitationsschrift, Jena (1969); Z. Perjés, Phys. Rev. Lett. **27**, 1668 (1971); W. Israel and G. A. Wilson, J. Math. Phys. **13**, 865 (1972).

[45]A. Papapetrou, Proc. Roy. Irish Acad. A **51**, 191 (1947); S. D. Majumdar, Phys. Rev. **72**, 390 (1947). These papers treat the general theory of static electrically poised particles in three space dimensions.

[46]Eqs. (2.30a)–(2.31b) of Ref. 12 refer to the important subgroup $(Q)_{4s}$ of the full three-dimensional group **Q**, elements denoted $(Q)_\delta$, $\delta\in$SL(2, R), for which $\delta=\left(\begin{smallmatrix}1&0\\4s&1\end{smallmatrix}\right)$. These equations must be supplemented with Eqs. (2.2a) and (2.2b) of Ref. 12 which express the transform of the coordinates (ρ, z) under $(Q)_\delta$ and $(\tilde{Q})_\delta$.

[47]D. Kramer and G. Neugebauer, Commun. Math. Phys. **10**, 132 (1968), Eq. (13).

[48]W. Kinnersley and T. Lemley, private communications (1980).

General relativity: Dynamics without symmetry

Karel Kuchař
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 25 March 1981; accepted for publication 8 May 1981)

The concept of conditional symmetry is introduced for a parametrized relativistic particle model and generalized to geometrodynamics. Its role in maintaining a one-system interpretation of the quantized theory is emphasized. It is shown that geometrodynamics does not have any conditional symmetry: Such a symmetry should be generated by a dynamical variable $K[g_{ab}, p^{ab}]$ which is linear and homogeneous in the gravitational momentum p^{ab} and which has a weakly vanishing Poisson bracket with the super-Hamiltonian and supermomentum. The generators K fall into equivalence classes modulo the supermomentum constraint. It is shown that each equivalence class can be represented by a member which is a spatial invariant. The remaining weak equations are turned into strong equations by the method of Lagrange multipliers. The local structure of the super-Hamiltonian and supermomentum imposes locality restrictions on the multipliers. These restrictions imply that the generator must be weakly equivalent to a local generator. A recursive argument then shows that the local generator must actually be weakly ultralocal. This uniquely determines the generator as the conformal Killing (super)vector of the local supermetric. However, the curvature scalar in the super-Hamiltonian breaks the conditional symmetry of the supermetric term and turns geometrodynamics into a theory without any symmetry. This result is generalized to inhomogeneous generators.

PACS numbers: 04.20.Fy, 04.20.Me, 02.40. + m, 11.30.Na

I. MOTIVATION

Geometrodynamics describes how to build a Ricci-flat spacetime through the evolution of a spatial geometry. A positive definite metric $g_{ab}(x)$ and its conjugate momentum $p^{ab}(x)$ are assigned on a spacelike hypersurface $X^\alpha = X^\alpha(x)$ subject to the constraints[1]

$$H(x) = 0 = H_a(x). \tag{1.1}$$

The super-Hamiltonian

$$H(x) \equiv G_{ab\,cd}(x)\, p^{ab}(x)\, p^{cd}(x) - g^{1/2}(x) R(x), \tag{1.2}$$

$$G_{ab\,cd} \equiv \tfrac{1}{2} g^{-1/2}(g_{ac}\, g_{bd} + g_{ad}\, g_{bc} - g_{ab}\, g_{cd}) \tag{1.3}$$

and the supermomentum

$$H_a(x) \equiv -2\, p^b_{a|b}(x) \tag{1.4}$$

which constrain the canonical data also determine their development. If the original hypersurface is incorporated into a one-parameter family of hypersurfaces $X^\alpha = X^\alpha(x, t)$, $X^\alpha(x, 0) = X^\alpha(x)$, the canonical variables change in accordance with the Hamilton equations

$$\dot{g}_{ab}(x) = [\, g_{ab}(x), H_N + H_{\mathbf{N}}\,], \qquad \dot{p}^{ab}(x) = [\, p^{ab}(x), H_N + H_{\mathbf{N}}\,]. \tag{1.5}$$

Here, the lapse function $N(x)$ and the shift vector $\mathbf{N}(x)$ describe the deformation of the hypersurface in the normal $n^\alpha(x)$ and the tangential $X^\alpha_a(x) \equiv \partial_a X^\alpha$ directions,

$$\dot{X}^\alpha = N n^\alpha + N^a X^\alpha_a. \tag{1.6}$$

They smear the constraint functions (1.2) and (1.4),

$$H_N \equiv \int d^3x\, N(x) H(x), \tag{1.7}$$

$$H_{\mathbf{N}} \equiv \int d^3x\, N^a(x) H_a(x) = \int d^3x\, N_{(a|b)}\, p^{ab}, \tag{1.8}$$

and yield thereby the geometrodynamical Hamiltonian $H_N + H_{\mathbf{N}}$. The rate of change of an arbitrary dynamical variable $K[\, g_{ab}, p^{ab}\,]$ is given by its Poisson bracket with this Hamiltonian,

$$\dot{K} = [K, H_N + H_{\mathbf{N}}]. \tag{1.9}$$

Hamiltonian geometrodynamics closely resembles parametrized Hamiltonian dynamics of a relativistic particle which is moving in a given Riemannian spacetime $g_{\alpha\beta}(X)$. The 4-momentum P_α of such a particle is constrained onto a mass shell

$$H \equiv (1/2m)(\, g^{\alpha\beta}(X)\, P_\alpha\, P_\beta + m^2) = 0. \tag{1.10}$$

An even closer correspondence to geometrodynamics emerges if the rest mass m of the particle is allowed to depend on an external potential $\phi(X)$ so that the super-Hamiltonian constraint reads[2]

$$H \equiv (1/2m)(\, g^{\alpha\beta}(X)\, P_\alpha\, P_\beta + V(X)) = 0, \qquad V(X) \equiv (m e^{\phi(X)})^2. \tag{1.11}$$

The super-Hamiltonian again generates the motion. We parametrize the trajectory of the particle by an arbitrary label time t: $X^\alpha = X^\alpha(t)$, $P_\alpha = P_\alpha(t)$. We introduce the lapse function $N(t)$, which scales the super-Hamiltonian H into the Hamiltonian $H_N \equiv NH$. Then

$$\dot{X}^\alpha = [X^\alpha, H_N], \quad \dot{P}_\alpha = [P_\alpha, H_N], \tag{1.12}$$

and the rate of change of an arbitrary dynamical variable $K[X^\alpha, P_\alpha]$ is given by its Poisson bracket with H_N,

$$\dot{K} = [K, H_N]. \tag{1.13}$$

In the end, the equations of motion provide the interpretation of N. They imply the relation

$$N = e^{-\phi}\dot{\tau}, \tag{1.14}$$

 0022-2488/81/112640-15$01.00

where τ is the proper time determined by the background metric $g_{\alpha\beta}$.

The parallels between geometrodynamics and the relativistic particle model are best summarized in Table I. We did not pay much attention to an overall factor $(2m)^{-1}$ in the super-Hamiltonian (1.11), which was inserted there mainly for dimensional reasons.

Naturally, the analogy shown in Table I is not perfect in all details. While there is only one super-Hamiltonian, one metric, and one mass term in the particle model, there are infinitely many such expressions in geometrodynamics, one for each point $x \in \mathscr{M}^3$. While the mass potential term in the particle model is positive, the corresponding term in geometrodynamics is not. Finally, the supermomentum does not have a natural counterpart in this particular particle model. Fortunately, intuitive reasons for these differences are well understood.[3]

The analogy of Table I will help us to clarify the notion of symmetry in geometrodynamics. There is hardly any aspect of a dynamical system which is physically more important than the symmetry of the Hamiltonian. At the classical level, symmetries beget conserved quantities. Without symmetries, the transition to quantum theory becomes ambiguous. However, to say what is meant by symmetry in the presence of constraints requires some care. Symmetry may be conditioned by constraints. We shall see how this works in parametrized particle dynamics. The analogy with geometrodynamics will then lead us to the concept of conditional symmetry in geometrodynamics. But it will also lead us to the disturbing conclusion that geometrodynamics does not have any symmetry.

In parametrized particle dynamics, symmetries of the super-Hamiltonian are generated by conserved dynamical variables K which are linear in the 4-momentum P_α,

$$K = k^{\alpha}(X)P_{\alpha}. \tag{1.15}$$

To see that, we evaluate the Poisson bracket

$$[K, H] = -(1/2m)(L_k g^{\alpha\beta}(X) \cdot P_\alpha P_\beta + L_k V(X)). \tag{1.16}$$

When the Lie derivatives $L_k g^{\alpha\beta}$ and $L_k V$ vanish,

$$L_k g^{\alpha\beta} \equiv k^{\gamma}\partial_\gamma g^{\alpha\beta} - \partial^{(\alpha}k^{\beta)} = 0, \tag{1.17}$$

$$L_k V \equiv k^{\alpha}\partial_\alpha V = 0, \tag{1.18}$$

the Poisson bracket (1.16) vanishes,

$$[K, H] = 0. \tag{1.19}$$

Equation (1.13) then implies that $\dot{K} = 0$ for any N, so that K is conserved. Conditions (1.17) and (1.18) tell us that the metric and the potential are symmetric under the flow of k^{α}. According to Eq. (1.17), k^{α} is a Killing vector of the metric $g_{\alpha\beta}$. For $\phi = 0$, the mass term is constant, $V = m^2$, and Eq. (1.17) is automatically satisfied.

The conditions (1.17) and (1.18) are equivalent to the statement that the Poisson bracket (1.16) vanishes for all values of the canonical variables X^{α} and P_α. While this is a sufficient condition for K to be conserved, it is by no means necessary. $\dot{K}$ does not need to vanish for all values of the canonical variables, but only for such values which satisfy the constraint equation (1.11),

$$H = 0 \Rightarrow [K, H] = 0. \tag{1.20}$$

If a dynamical variable F vanishes only modulo the constraints, as in Eq. (1.20), we say with Dirac[4] that it vanishes weakly, and write $F \approx 0$. On the other hand, if a dynamical variable F vanishes identically in the canonical variables, as in Eq. (1.19), we say that it vanishes strongly, and write $F = 0$. In this notation, Eq. (1.20) reads

$$[K, H] \approx 0. \tag{1.21}$$

It is now not only a sufficient, but also a necessary condition for K to be conserved.

The strong equation (1.19) was equivalent to the symmetry conditions (1.17)–(1.18). The weak equation (1.21) leads to weaker conditions. To see what these conditions are, we first turn the weak equation (1.21) or (1.20) into a strong equation

$$[K, H] = \Lambda H \tag{1.22}$$

by the method of Lagrange multipliers. Both $[K, H]$ and H are quadratic functions of the 4-momentum P_α; therefore, the Lagrange multiplier Λ cannot depend on P_α, but only on the position X^{α}. If we compare the coefficients on both sides of Eq. (1.22), we get

$$L_k g^{\alpha\beta}(X) = -\Lambda(X) g^{\alpha\beta}(X) \tag{1.23}$$

and

$$L_k V(X) = -\Lambda(X) V(X). \tag{1.24}$$

Equation (1.23) is fulfilled if the spacetime has a conformal Killing vector k^{α}. Equation (1.24) means that the mass term $V(X)$ is conformally scaled along the flow of k^{α}. Such a symmetry is understandably weaker than that expressed by Eqs. (1.17)–(1.18). Avoiding the term "weak symmetry," we

TABLE I.

	Particle model	Geometrodynamics
Index	α (a discrete index)	$ab\,x^c$(a discrete-continuous index)
Canonical coordinate	X^{α}	$g_{ab}(x)$
Canonical momentum	P_α	$p^{ab}(x)$
Lapse	$N(t)$	$N(t; x)$
Hamiltonian	$H_N \equiv NH$	$H_N \equiv \int d^3x\, N(x)H(x)$
Super-Hamiltonian	H	$H(x)$
Metric	$g^{\alpha\beta}(X)$	$G_{ab\,cd}(x)[g_{mn}]$
Mass-Potential term	$V(X) = m^2 e^{2\phi(X)}$	$V(x)[g_{mn}] = -g^{1/2}R(x)[g_{mn}]$

shall speak instead of "conditional symmetry." We shall thus call a dynamical variable (1.16) which satisfies the weak equation (1.21) a generator of a conditional symmetry.

When both conditions (1.23) and (1.24) are satisfied, the rescaled metric

$$\tilde{g}^{\alpha\beta} \equiv e^{-2\phi} g^{\alpha\beta} \tag{1.25}$$

has a true Killing vector k^{α}, and the rescaled super-Hamiltonian

$$\tilde{H} \equiv e^{-2\phi} H = (1/2m)(\tilde{g}^{\alpha\beta} P_{\alpha} P_{\beta} + m^2) \tag{1.26}$$

has an unconditional symmetry

$$[\tilde{H}, K] = 0. \tag{1.27}$$

If we rescale the lapse function N by the factor $e^{2\phi}$,

$$\tilde{N} = N e^{2\phi}, \tag{1.28}$$

the Hamiltonian of the system remains the same,

$$\tilde{H}_{\tilde{N}} = H_N. \tag{1.29}$$

The rescaled expressions (1.26) and (1.28) thus lead to (weakly) equivalent equations of motion. If $X^{\alpha}(t)$, $P_{\alpha}(t)$ satisfy the old constraint (1.11) and the old equations of motion (1.12), they also satisfy the new constraint $\tilde{H} = 0$ and the new equations of motion (generated by $\tilde{H}_{\tilde{N}}$). The new equations imply that the rescaled lapse function is simply the rate of change of the proper time $\tilde{\tau}$ determined by the rescaled metric $\tilde{g}^{\alpha\beta}$,[5]

$$\tilde{N} = \dot{\tilde{\tau}}. \tag{1.30}$$

For $\phi = 0$, the weak equation (1.21) implies the strong equation (1.19). Indeed, from Eq. (1.24) it follows that the Lagrange multiplier Λ must vanish. The conditional symmetry in this case thus necessarily reduces to an unconditional symmetry. This is no longer true for $\phi \neq 0$.

If k^{α} is timelike, $K \equiv k^{\alpha}(X) P_{\alpha}$ can be interpreted as a conserved energy of the particle. Similarly, if k^{α} is spacelike, K can be interpreted as a conserved momentum. The variable K has the same meaning and the same value in the rescaled version of the theory as it has in the original version. The only difference is that $[K, \tilde{H}]$ vanishes strongly while $[K, H]$ vanishes weakly. In other words, K generates an unconditional symmetry in the rescaled version, while it generates a conditional symmetry in the original version.

The existence of a timelike Killing vector is all important when one tries to construct a one-particle Hilbert space from solutions of the Klein–Gordon equation corresponding to the super-Hamiltonian (1.10). In particular, such a Killing vector defines an unambiguous splitting of these solutions into positive and negative frequencies. In spacetimes which do not have a timelike Killing vector, the one-particle interpretation of the Klein–Gordon equation cannot be maintained. Particles, whichever way one tries to define them, are necessarily produced or destroyed by the changing background.

For $\phi = 0$, we have seen that the conditional symmetry of the super-Hamiltonian (1.10) leads naturally to a Killing vector of the original metric $g_{\alpha\beta}$. For $\phi \neq 0$, the conditional symmetry implies that the rescaled metric $\tilde{g}_{\alpha\beta}$ has a Killing vector. This enables us to construct a one-particle Hilbert space from solutions of the conformally covariant Klein–Gordon equation. To summarize, it is always the existence of a (timelike) *conditional* symmetry which enables us to construct a one-particle Hilbert space and maintain the one-particle interpretation of the Klein–Gordon equation.

In geometrodynamics, the role of the Klein–Gordon equation is played by the Wheeler–DeWitt equation for the state functional $\Psi[g_{ab}]$.[6] The existence of a conditional symmetry of the geometrodynamical Hamiltonian is even more crucial in geometrodynamics than in particle theory. When the particle Hamiltonian H_N does not have a conditional symmetry, one can simply say that particles are produced or destroyed by the changing background $g_{\alpha\beta}(X)$ or potential $\phi(X)$. In geometrodynamics, it is not at all clear how to interpret a formally constructed "many-geometries theory"[7] in terms of classical measurements.

The question whether the geometrodynamical Hamiltonian has a conditional symmetry has never been fully answered, mainly because it has never been properly formulated. All that could be established was that a straightforward "local" construction of such a symmetry fails.[8] We are now ready to ask the question whether the geometrodynamical Hamiltonian $H_N + H_{\mathbf{N}}$ has a conditional symmetry in its full generality. We are also able to answer this general question, and the answer is a straight NO. In our opinion, this negative answer implies that geometry cannot be quantized as a Klein–Gordon system while keeping the one-system interpretation. If so, one should either take the many-geometries formalism seriously or conclude (as we are inclined to do) that the metric representation of the gravitational field is inappropriate.

II. CONDITIONAL SYMMETRY IN GEOMETRODYNAMICS

We say that a dynamical variable $K[g_{ab}, p^{ab}]$ generates a conditional symmetry in geometrodynamics[9] if K is a linear homogeneous functional of the gravitational momentum p^{ab},

$$K = \int d^3x\, k_{ab}(x)[g_{mn}] p^{ab}(x), \tag{2.1}$$

which has a weakly vanishing Poisson bracket with the geometrodynamical Hamiltonian (1.7), (1.8):

$$[K, H_N + H_{\mathbf{N}}] \approx 0 \quad \forall N, \mathbf{N}. \tag{2.2}$$

Because the lapse function and the shift vector in Eq. (2.2) are arbitrary, Eq. (1.9) implies that K has the same value on every spacelike hypersurface, i.e., that it is conserved. The arbitrariness of N and $\mathbf{N}$ also enables us to split Eq. (2.2) into the lapse and the shift parts,

$$[K, H_N] \approx 0 \quad \forall N, \tag{2.3}$$

$$[K, H_{\mathbf{N}}] \approx 0 \quad \forall \mathbf{N}, \tag{2.4}$$

and to replace these by conditions on the super-Hamiltonian and the supermomentum,

$$[K, H(x)] \approx 0, \tag{2.5}$$

$$[K, H_a(x)] \approx 0. \tag{2.6}$$

By the method of Lagrange multipliers, Eqs. (2.3) and

(2.4) can be turned into strong equations

$$[K, H_N] = \int d^3x' \ (\Lambda_N(x')H(x') + \lambda_N{}^{a'}(x')H_a(x')), \tag{2.7}$$

$$[K, H_{\mathbf{N}}] = \int d^3x' \ (\Lambda_{\mathbf{N}}(x')H(x') + \lambda_{\mathbf{N}}{}^{a'}(x')H_{a'}(x')). \tag{2.8}$$

Because the constraints form a field on the space manifold, $x \in \mathscr{M}^3$, the same thing is true about the multipliers. Further, because H_N and $H_{\mathbf{N}}$ are the smeared forms of the super-Hamiltonian $H(x)$ and the supermomentum $H_a(x)$, Eqs. (1.7) and (1.8), the multipliers are to be considered as (bipoint) distributions acting on the test functions $N(x)$ and $\mathbf{N}(x)$:

$$\Lambda_N(x') = \int d^3x \, \Lambda(x', x)N(x), \tag{2.9}$$

$$\lambda_N{}^{a'}(x') = \int d^3x \, \lambda^{a'}(x', x)N(x), \tag{2.10}$$

with similar equations holding for the $\Lambda_{\mathbf{N}}(x')$ and $\lambda_{\mathbf{N}}{}^{a'}(x')$ multipliers. The condition (2.7) [and similarly the condition (2.8)] can then be written in terms of these distributions:

$$[K, H(x)] = \int d^3x' \ \Lambda(x', x)H(x') + \int d^3x' \ \lambda^{a'}(x', x)H_{a'}(x'). \tag{2.11}$$

Equation (2.11) is the strong form of the weak equation (2.5).

In principle, the Lagrange multipliers are some functionals of the canonical variables $g_{ab}(x)$ and $p^{ab}(x)$. Because we know that K and $H_{\mathbf{N}}$ are linear homogeneous functionals of the momentum, while H_N is a quadratic functional of the momentum without a linear term, we can immediately infer how the multipliers depend on $p^{ab}(x)$. The Poisson bracket $[K, H_N]$ is a quadratic functional of p^{ab} without a linear term [we shall write it in detail later, Eqs. (4.2), (4.3), and (4.8)]. Consequently, $\Lambda_N(x')$ cannot depend on the momentum,

$$\Lambda_N(x') = \Lambda_N(x')[\, g_{mn}\,], \tag{2.12}$$

while $\lambda_N{}^{a'}(x')$ must be a linear homogeneous functional of the momentum,

$$\lambda_N{}^{a'}(x') = \int d^3x'' \, \lambda_N{}^{a'}{}_{b''c''}(x', x'')[g_{mn}]p^{b''c''}(x''). \tag{2.13}$$

Similarly, $[K, H_{\mathbf{N}}]$ is a linear homogeneous functional of p^{ab}. Equation (2.8) then implies that $\Lambda_{\mathbf{N}}(x')$ must actually vanish,

$$\Lambda_{\mathbf{N}}(x') = 0, \tag{2.14}$$

while $\lambda_{\mathbf{N}}{}^{a'}(x')$ can depend only on the metric,

$$\lambda_{\mathbf{N}}{}^{a'}(x') = \lambda_{\mathbf{N}}{}^{a'}(x')[g_{mn}]. \tag{2.15}$$

Equation (2.8) thus reduces to a simpler statement

$$[K, H_{\mathbf{N}}] = \int d^3x' \, \lambda_{\mathbf{N}}{}^{a'}(x')H_{a'}(x') = \int d^3x' \, \lambda_{\mathbf{N}(a'|b')}(x')p^{a'b'}(x'). \tag{2.16}$$

Note that the multiplier $\lambda_{\mathbf{N}}{}^{a'}[g_{mn}]$ is not entirely fixed by Eq. (2.16). For a symmetric metric g_{mn}, a Killing vector field $\kappa_{\mathbf{N}}{}^{a'}(x')[g_{mn}]$ can be added to the multiplier $\lambda_{\mathbf{N}}{}^{a'}(x')$ without changing Eq. (2.16):

$$\lambda_{\mathbf{N}}{}^{a'} \longmapsto \lambda_{\mathbf{N}}{}^{a'} + \kappa_{\mathbf{N}}{}^{a'}, \qquad \kappa_{\mathbf{N}(a'|b')} = 0. \tag{2.17}$$

A similar argument can be made for Eq. (2.7) and the multiplier $\lambda_N{}^{a'}(x')$. Symmetric metrics, of course, are special points in superspace,[10] while the generic metric does not possess any symmetry.

Our main task in the following sections will be to find how the coefficients $\Lambda_N(x)[g_{mn}]$ and $\lambda_N{}^{a'}{}_{b''c''}(x', x'')[g_{mn}]$ depend on the metric. The coefficient $\lambda_{\mathbf{N}}{}^{a'}(x')[g_{mn}]$ can be removed by a gauge transformation which we are now going to introduce.

The generators K of conditional symmetries fall naturally into equivalence classes. First of all, any two dynamical variables $\bar{K}$ and K whose values coincide on the constraint surface,

$$\bar{K} \approx K, \tag{2.18}$$

are physically equivalent. Moreover, when K is conserved, Eq. (2.2), $\bar{K}$ is conserved as well. To see that, we replace Eq. (2.18) by the strong equation

$$\bar{K} = K + K_\mu + H_\mu = \int d^3x \, (\mu(x)H(x) + \mu^a(x)H_a(x)) \tag{2.19}$$

at the price of introducing the Lagrange multipliers $\mu(x)[g_{mn}, p^{mn}]$ and $\mu^a(x)[g_{mn}, p^{mn}]$. The constraint functions $H(x)$ and $H_a(x)$ satisfy the closure relations[11]

$$[H(x), H(x')] = H^a(x)\partial_a\delta(x, x') - (x \leftrightarrow x'), \tag{2.20}$$

$$[H_a(x), H(x')] = H(x)\partial_a\delta(x, x'), \tag{2.21}$$

$$[H_a(x), H_{b'}(x')] = H_b(x)\partial_a\delta(x, x') - (ax \leftrightarrow bx'). \tag{2.22}$$

When smeared by two different sets $\{M(x), M^a(x)\}$ and $\{N(x'), N^{a'}(x')\}$ of externally prescribed lapse and shift functions. Eqs. (2.20)–(2.22) read

$$[H_M, H_N] = H_{\mathbf{MN}}, \tag{2.23}$$

$$[H_{\mathbf{M}}, H_N] = H_{\mathbf{M}\cdot\partial N}, \tag{2.24}$$

$$[H_{\mathbf{M}}, H_{\mathbf{N}}] = H_{[\mathbf{M},\mathbf{N}]}, \tag{2.25}$$

where

$$(\mathbf{MN})^a \equiv M\partial^a N - N\partial^a M, \tag{2.26}$$

$$\mathbf{M}\cdot\partial N \equiv M^a\partial_a N, \tag{2.27}$$

$$[\mathbf{M}, \mathbf{N}] \equiv L_{\mathbf{M}}\mathbf{N} \tag{2.28}$$

define a composition of the smearing functions. When some of the smearing functions are themselves dynamical variables, e.g., when the set $\{M(x), M^a(x)\}$ is replaced by the set $\{\mu(x)[g_{mn}, p^{mn}], \mu^a(x)[g_{mn}, p^{mn}]\}$, Eqs. (2.23)–(2.25) acquire additional terms:

$$[H_\mu, H_N] = H_{[\mu,H_N]} + H_{\mu\mathbf{N}}, \tag{2.29}$$

$$[H_\mu, H_N] = H_{\mu\cdot\partial N} + H_{[\mu,H_N]}, \tag{2.30}$$

$$[H_\mu, H_{\mathbf{N}}] = H_{[\mu,H_{\mathbf{N}}]+[\mu,\mathbf{N}]}. \tag{2.31}$$

Still, the right-hand sides of Eqs. (2.29)–(2.31) are linear com-

binations of the constraints. Hence, $[H_\mu + H_\mu, H_N + H_N] \approx 0$, and if K is conserved, $\bar{K}$ is conserved.

For K to generate a conditional symmetry, it must be linear and homogeneous in the canonical momentum p^{ab} in addition to being conserved. To preserve the linearity, we must restrict the multipliers in Eq. (2.19) by the conditions

$$\mu(x) = 0, \quad \mu^a(x) = \mu^a(x)[g_{mn}]. \tag{2.32}$$

In other words, for generators of conditional symmetries the weak equation (2.18) is translated into the strong equation

$$\bar{K} = K + H_\mu = K + \int d^3x\, \mu^a(x)[g_{mn}] H_a(x). \tag{2.33}$$

The presence of supermomentum constraints (which are linear in p^{ab}) thus introduces a novel feature into the study of conditional symmetry. In our search for such a symmetry, we are interested only in the equivalence classes (2.18). If we know one representative, K, of an equivalence class, we get all the other members of that class from Eq. (2.33). We can view Eq. (2.33) as a gauge transformation generated by the gauge function(al) $\mu^a(x)[g_{mn}]$.

The conservation equations for K, Eqs. (2.3) and (2.4), were turned into strong equations (2.7) and (2.8) by the use of Lagrange multipliers. If we pass from K to $\bar{K}$ by the gauge transformation (2.33), the strong equations (2.7) and (2.8) still hold, but with changed multipliers. We identify the new multipliers by using Eqs. (2.29)–(2.31):

$$\bar{\Lambda}_N = \Lambda_N + \mu^a \partial_a N, \tag{2.34}$$

$$\bar{\lambda}_N{}^a(x) = \lambda_N{}^a(x) + [\mu^a(x), H_N], \tag{2.35}$$

$$\bar{\lambda}_N{}^a(x) = \lambda_N{}^a(x) + [\mu^a(x), H_N] - L_N \mu^a(x). \tag{2.36}$$

For symmetric metrics, we can still add a Killing vector field to the right-hand side of Eq. (2.36) or Eq. (2.35).

We are now going to prove that $\mu^a(x)[g_{mn}]$ can be chosen such that $\bar{\lambda}_N{}^a(x)[g_{mn}]$ vanishes.

III. SPATIAL INVARIANTS

The vector fields $\mathbf{N}(x)$ on $\mathscr{M}^3$ are closed under the Lie bracket operation [M, N] and they generate the Lie algebra of the group Diff($\mathscr{M}^3$). The smeared supermomentum H_N reproduces the Lie derivative of the canonical variables g_{ab} and p^{ab} along the shift vector $\mathbf{N}$ through the Poisson brackets,

$$\begin{aligned} L_N g_{ab}(x) &= [g_{ab}(x), H_N], \\ L_N p^{ab}(x) &= [p^{ab}(x), H_N]. \end{aligned} \tag{3.1}$$

Moreover, two smeared supermomenta, H_M and H_N, close according to the rule (2.25). This means that H_N represent the Lie algebra of Diff($\mathscr{M}^3$) on the geometrodynamical phase space. If a dynamical variable K has the vanishing Poisson bracket with all H_N,

$$[K, H_N] = 0 \quad \forall \mathbf{N}, \tag{3.2}$$

it is invariant under Diff($\mathscr{M}^3$).

Equation (2.16) tells us how the dynamical variable K transforms under spatial diffeomorphisms. In particular, it tells us that K is a spatial invariant on the constraint surface. If not only $\Lambda_N(x)$, but also $\lambda_N{}^a(x)$ were to vanish, K would be a spatial invariant also outside the constraint surface. It is much easier to work with expressions which are spatial invariants irrespective of the constraints. Let us thus try to learn what we can say about the multiplier $\lambda_N{}^a(x)[g_{mn}]$.

Our basic tool is the Jacobi identity

$$[[K, H_M], H_N] + [[H_N, K], H_M] + [[H_M, H_N], K] \equiv 0, \tag{3.3}$$

which, supplemented by the closure relation (2.25), can be written in the form

$$[[K, H_{[M}], H_{N]}] \equiv [K, H_{[M,N]}]. \tag{3.4}$$

If we replace the Poisson brackets of K in Eq. (3.4) by the expressions (2.16), we get a condition on the Lagrange multipliers,

$$\left[\int d^3x\, \lambda_{[M}{}^a(x) H_a(x), H_{N]}\right] = \int d^3x\, \lambda_{[M,N]}{}^a(x) H_a(x). \tag{3.5}$$

The Poisson bracket on the left-hand side of this condition can be rearranged by the rule (2.31), and so

$$\int d^3x\, ([\lambda_{[M}{}^a(x), H_{N]}] + L_{[M}\lambda_{N]}{}^a(x) - \lambda_{[M,N]}{}^a(x)) H_a(x) = 0. \tag{3.6}$$

The last equation has the form

$$\int d^3x\, l^a(x)[g_{mn}] H_a(x) = 0, \tag{3.7}$$

where $l^a(x)[g_{mn}]$ depends only on the metric g_{mn}, but not on the momentum p^{mn}. If we integrate Eq. (3.7) by parts, like Eq. (1.8), and recall that p^{ab} is arbitrary, we see that

$$l_{(a|b)}(x)[g_{mn}] = 0. \tag{3.8}$$

If g_{mn} is a generic metric without any symmetry, i.e., without a Killing vector, we are able to conclude that $l_a(x)[g_{mn}]$ must vanish. In other words,

$$[\lambda_{[M}{}^a(x), H_{N]}] + L_{[M}\lambda_{N]}{}^a(x) - \lambda_{[M,N]}{}^a(x) = 0 \tag{3.9}$$

except when the argument of $\lambda_M{}^a(x)[g_{mn}]$ happens to be a symmetric metric.

Equation (3.9) is not strong enough to imply that $\lambda_N{}^a(x)[g_{mn}]$ must vanish. However, it is strong enough to imply that $\lambda_N{}^a(x)[g_{mn}]$ can be removed by the gauge transformation (2.36), i.e., that $\mu^a(x)[g_{mn}]$ can be chosen so that

$$[H_N, \mu^a(x)] + L_N \mu^a(x) = \lambda_N{}^a(x). \tag{3.10}$$

To show that, notice what the character of Eq. (3.10) is. For a given $\mathbf{N}(x)$, the left-hand side of Eq. (3.10) is a linear combination of the unknown functional $\mu^a(x)[g_{mn}]$ and its first variational derivatives $\delta\mu^a(x)/\delta g_{mn}(y)$. The multiplier $\lambda_M{}^a(x)[g_{mn}]$ which we start from is a given functional of g_{mn}. Consequently, Eq. (3.10) can be considered as a infinite system [labeled by an "index" $\mathbf{N}(x)$] of first-order linear inhomogeneous partial variational differential equations for the functional $\mu^a(x)[g_{mn}]$. Relying on the analogy with the theory of finite systems of partial differential equations, the

integrability condition of Eq. (3.10) is the Jacobi identity

$$[H_{[M}, [H_{N]}, \mu^a(x)]] = [H_{[M,N]}, \mu^a(x)] \tag{3.11}$$

from which the Poisson brackets of $\mu^a(x)$ are eliminated by Eq. (3.10) itself. This procedure leads to the equation

$$[H_{[M}, \lambda_{N]}{}^a(x) - L_{N]}\mu^a(x)] = \lambda_{[M,N]}{}^a - L_{[M,N]}\mu^a(x) \tag{3.12}$$

as the desired integrability condition. It is easy to use Eq. (3.10) once more,

$$\begin{aligned}[H_{[M}, L_{N]}\mu^a(x)] &= L_{[N}[H_{M]}, \mu^a(x)] \\ &= L_{[N}(\lambda_{M]}{}^a(x) - L_{M]}\mu^a(x)) \\ &= L_{[N}\lambda_{M]}{}^a(x) - L_{[N,M]}\mu^a(x),\end{aligned} \tag{3.13}$$

and reduce thereby the integrability condition (3.12) to the form (3.9). To summarize, Eq. (3.9) is the integrability condition of Eq. (3.10).

In fact, it is unnecessary to exclude symmetric metrics. Strictly speaking, we do not need to satisfy Eq. (3.10), but rather the integrated equation

$$\int d^3x \{[H_N, \mu^a(x)] - L_N\mu^a(x) - \lambda_N{}^a(x)\}H_a(x) = 0. \tag{3.14}$$

The integrability condition of Eq. (3.14) is exactly the integrated equation (3.6).

This concludes our proof that the Lagrange multiplier $\lambda_N{}^a(x)[g_{mn}]$ can be gauged away. Therefore, for every K which satisfies the weak equation $[K, H_N] \approx 0$ there is an equivalent $\bar{K} \approx K$ which satisfies the strong equation

$$[\bar{K}, H_N] = 0 \quad \forall N \tag{3.15}$$

and which is thus invariant under Diff($\mathcal{M}^3$). Moreover, the new functional $\bar{K}$ satisfies Eq. (2.7) with the modified Lagrange multipliers (2.34) and (2.35).

From now on, we shall always represent an equivalence class $\bar{K} \approx K$ of generators by that member $\bar{K}$ which satisfies Eq. (3.15), and we shall simply omit the bar. In this manner, our problem of finding a conditional symmetry in geometrodynamics reduces to the problem of finding a functional K which has the form (2.1), is invariant under Diff($\mathcal{M}^3$), Eq. (3.2), and satisfies the strong equation (2.7) [or (2.11)] with the Lagrange multipliers (2.12) and (2.13).

The condition (3.2) applied to a functional of the form (2.1) implies that the coefficient $k_{ab}(x)[g_{mn}]$ must transform as a tensor under spatial transformations. This fact vastly simplifies our further calculations.

IV. EVALUATING $[K, H(x)]$. ULTRALOCAL GEOMETRODYNAMICS

We now turn our attention from Eq. (2.8) to Eq. (2.7)—or (2.11)—i.e., from the supermomentum to the super-Hamiltonian. Our first task is to evaluate the Poisson bracket $[K, H(x)]$ between the generator (2.1) and the super-Hamiltonian (1.2). We proceed in two steps, splitting the super-Hamiltonian into the "kinetic" and "potential" parts,

$$\begin{aligned}&H(x) = T(x) + V(x), \\ &T(x) \equiv G_{ab\,cd}(x)(g_{mn}(x))\,p^{ab}(x)\,p^{cd}(x), \\ &V(x) \equiv -g^{1/2}R(x).\end{aligned} \tag{4.1}$$

The computation of $[K, T(x)]$ is straightforward, requiring nothing else but the definition of the Poisson bracket. We get

$$[K, T(x)] = p^{ab}(x)h_{ab}(x), \tag{4.2}$$

where

$$h_{ab}(x) = \int d^3x'\, h_{ab\,c'd'}(x, x')p^{c'd'}(x'),$$

$$\begin{aligned}h_{ab\,c'd'}(x, x')[g_{mn}] &= 2G_{ab\,mn}(x)\,\delta k_{c'd'}(x')/\delta g_{mn}(x) \\ &\quad - \partial G_{ab\,cd}(x)/\partial g_{mn}(x) \cdot k_{mn}(x)\delta(x', x)\end{aligned} \tag{4.3}$$

is a linear homogeneous functional of the momentum with a kernel $h_{ab\,c'd'}(x, x')$ which is a bipoint tensor–tensor distribution. The Poisson bracket (4.2) is thus a quadratic form of the momenta. We decided not to symmetrize the coefficient $h_{ab\,c'd'}(x, x')$ in the pairs abx and $c'd'x'$, so that the privileged position of the point x remains clearly visible.

The evaluation of the Poisson bracket $[K, V(x)]$ is based on the variation of the Ricci scalar density,[12]

$$\delta(g^{1/2}R) = -g^{1/2}G^{ab}\,\delta g_{ab} + G^{ab\,cd}\,\delta g_{ab\,|cd}. \tag{4.4}$$

Here, G^{ab} is the spatial Einstein tensor and

$$G^{ab\,cd} \equiv \tfrac{1}{2}g^{1/2}(g^{ac}g^{bd} + g^{ad}g^{bc} - 2g^{ab}g^{cd}) \tag{4.5}$$

is the inverse of the "local supermetric" (1.3),

$$G^{ab\,mn}G_{mn\,cd} = \delta^{ab}_{cd} \equiv \tfrac{1}{2}\delta^a_{(c}\delta^b_{d)}. \tag{4.6}$$

Because

$$[K, V(x)] = \int d^3x'\, \delta(g^{1/2}R(x))/\delta g_{ab}(x') \cdot k_{ab}(x'), \tag{4.7}$$

$[K, V(x)]$ is nothing else but the variation of $g^{1/2}R(x)$ induced by the variation $\delta g_{ab} = k_{ab}$ of the metric. In other words,

$$[K, V(x)] = -g^{1/2}G^{ab}k_{ab} + G^{ab\,cd}k_{ab\,|cd}. \tag{4.8}$$

The expression (4.8) depends only on the metric g_{mn}, not on the momentum p^{mn}.

The Poisson bracket $[K, H(x)]$ is thus a quadratic functional of p^{mn} without the linear term. We already mentioned what this implies for the momentum dependence of the multipliers (2.12) and (2.13). When we return back to Eq. (2.11) and compare first the terms quadratic in the momenta and then the terms which depend only on the metric, we split Eq. (2.11) into two pieces:

$$\begin{aligned}p^{ab}(x)h_{ab}(x) &= \int d^3x'\, \Lambda(x', x)T(x') \\ &\quad + \int d^3x'\, \lambda^{a'}(x', x)H_{a'}(x'),\end{aligned} \tag{4.9}$$

$$\begin{aligned}&-g^{1/2}G^{ab}k_{ab} + G^{ab\,cd}k_{ab\,|cd} \\ &\quad = -\int d^3x'\, \Lambda(x', x)g^{1/2}(x')R(x').\end{aligned} \tag{4.10}$$

We shall first leave Eq. (4.10) aside and explore all the consequences of Eq. (4.9). Besides purely technical advantages, such a procedure has an additional reason. Isham[13] suggested studying the theory with $V(x) = 0$ as the starting point of a perturbative approach to quantum geometrodynamics. The kinetic super-Hamiltonian $T(x)$ does not contain any spatial derivatives of the canonical variables g_{ab} and p^{ab}. The truncated theory thus falls into the class of ultralocal theories, the quantization of which was studied by Klauder.[14]

The geometrical meaning of truncated geometrodynamics was clarified by Henneaux,[15] and its quantization by Klauder's techniques was discussed by Pilati.[16] Restricting ourselves first to Eq. (4.9) only, we are posing the question of what are the conditional symmetries of this ultralocal geometrodynamics. We shall see that ultralocal geometrodynamics has a single conditional symmetry and its generator is itself ultralocal. However, this ultralocal generator does not satisfy Eq. (4.10), so that the potential term $V(x)$ breaks the ultralocal conditional symmetry and turns full geometrodynamics into a theory without any symmetry.

V. LOCAL STRUCTURE OF LAGRANGE MULTIPLIERS

The first conclusion we are going to draw from Eq. (4.9) is that the distributions $\Lambda(x', x)$ and $\lambda^{a'}(x', x)$ have a local nature. As always, Eq. (4.9) is the strong version of the weak equation

$$T(x') = 0 = H_{a'}(x') \quad \forall x' \in \mathscr{M}^3 \Rightarrow p^{ab}(x)h_{ab}(x) = 0. \tag{5.1}$$

How can the constraints $H_{a'}(x') = 0$ and $T(x') = 0$ possibly conspire to make the expression $p^{ab}(x)h_{ab}(x)$ at a given point x vanish?

(I) It is clear that the supermomentum constraint can never be used to annihilate the first factor $p^{ab}(x)$ in $p^{ab}(x)h_{ab}(x)$. There is no integration over x in this expression, and the derivative $|b$ can thus never be transferred to $p^{ab}(x)$ to yield $p^{ab}{}_{|b}(x)$. Therefore, the only chance for $H_a = 0$ to convert the above expression to zero is to act on $h_{ab}(x)$ alone. In the simplest case, $H_a = 0$ makes $p^{ab}(x)h_{ab}(x)$ vanish because

$$H_{a'}(x') = 0 \quad \forall x' \in \mathscr{M} \Rightarrow h_{ab}(x) = 0, \tag{5.2}$$

i.e., because

$$\exists\, \lambda_{ab}{}^{c'}(x, x')[g_{mn}]: \int d^3x'\, h_{ab\,c'd'}(x, x')p^{c'd'}(x') = \int d^3x'\, \lambda_{ab}{}^{c'}(x, x')H_{c'}(x') = \int d^3x'\, \lambda_{ab\,(c'|d')}(x, x')p^{c'd'}(x'). \tag{5.3}$$

However, $p^{c'd'}(x')$ is arbitrary, and so

$$h_{ab\,c'd'} = \lambda_{ab\,(c'|d')}. \tag{5.4}$$

(II) While Eq. (5.4) is sufficient to annihilate the expression $p^{ab}(x)h_{ab}(x)$, it is far from being necessary. It may happen that under the supermomentum constraint $h_{ab}(x)$ does not vanish but becomes an expression which is "perpendicular" to $p^{ab}(x)$. This happens when

$$h_{ab\,c'd'} - \lambda_{ab\,(c'|d')} = M_{ab\,cd}(x)g^{-1/2}(x)\delta(x', x), \tag{5.5}$$

where $M_{ab\,cd}(x)[g_{mn}]$ is a functional of the metric with the symmetries

$$M_{ab\,cd} = M_{ba\,cd} = M_{ab\,dc} = -M_{cd\,ab}. \tag{5.6}$$

We have inserted $g^{-1/2}(x)$ into Eq. (5.5) to turn $M_{ab\,cd}(x)$ into a tensor.

(III) Still, even Eq. (5.5) is not strictly necessary to enforce Eq. (5.1). So far, we have used only the supermomentum constraint. It may happen, however, that $p^{ab}(x)h_{ab}(x)$ vanishes only by virtue of the (ultralocal) super-Hamiltonian constraint $T(x') = 0$. Assume that the canonical data are analytic, so that the constraint $T(x') = 0\ \forall x' \in \mathscr{M}^3$ can be replaced by an infinite sequence of constraints at x,

$$T(x) = 0, \quad T_{|i}(x) = 0, \quad T_{,i_1i_2}(x) = 0, \quad T_{,i_1i_2i_3}(x) = 0, \ldots . \tag{5.7}$$

We have symmetrized all the covariant derivatives,

$$T_{,i_1\cdots i_N}(x) \equiv \mathop{\mathrm{S}}_{i} T_{|i_1\cdots i_N}(x), \tag{5.8}$$

because any antisymmetric piece can be expressed as a combination of lower order constraints. Inspect now the expression $T(x)$ and its first and second derivatives

$$T_{|i}(x) = 2G_{ab\,cd}(x)p^{ab}(x)p^{cd}{}_{|i}(x), \tag{5.9}$$

$$T_{,ij}(x) = G_{ab\,cd}(x)p^{ab}(x)p^{cd}{}_{|(ij)}(x) + G_{ab\,cd}(x)p^{ab}{}_{|(i}(x)p^{cd}{}_{|j)}(x). \tag{5.10}$$

The constraints $T(x) = 0$ and $T_{|i}(x) = 0$ contain only the terms with $p^{ab}(x)$. On the other hand, the constraint $T_{,ij}(x) = 0$ also contains the term $G_{ab\,cd}(x)p^{ab}{}_{|(i}(x)p^{cd}{}_{|j)}(x)$ in which both momenta are differentiated. Obviously, no coefficient $\Lambda^{ij}(x)[g_{mn}]$ which multiplies the expression $T_{,ij}(x)$ can ever reduce this term to a combination of the $T(x) = 0$, $T_{|i}(x) = 0$, and $H_a(x) = 0$ constraints. Therefore, $p_{ab}(x)h_{ab}(x)$ can vanish when the equation

$$\int d^3x'\, \{h_{ab\,c'd'}(x, x') - \lambda_{ab\,(c'|d')}(x, x') - M_{ab\,cd}(x)g^{-1/2}(x)\delta(x', x)\}p^{c'd'}(x') = \Lambda(x)G_{ab\,cd}(x)p^{cd}(x) + 2\Lambda^{i}(x)G_{ab\,cd}(x)p^{cd}{}_{|i}(x) \tag{5.11}$$

makes the constraints $T(x) = 0$ and $T_{|i}(x) = 0$ effective, but it cannot vanish modulo the $T_{,ij}(x) = 0$ constraint. The higher order constraints cannot help us either, because no combination of such constraints,

$$\sum_{M=2}^{N} \Lambda^{i_1\cdots i_M}(x)[g_{mn}]T_{,i_1\cdots i_M}(x),$$

can eliminate all the terms in which both p^{ab} and p^{cd} are differentiated by reducing them to the constraints $T_{,i_1\cdots i_M}(x) = 0$, $M \leqslant N-1$, and $H_{a,i_1\cdots i_M}(x) = 0$, $M \leqslant N-2$. We thus conclude that Eq. (5.11) is a necessary and sufficient condition for the implication (5.1).

A comparison of Eq. (5.11) with the general formula (2.11) reveals the significance of our result. We have just proved that

$$\Lambda(x', x) = (\Lambda(x') - \Lambda^{i}{}_{|i'}(x'))\delta(x', x) - \Lambda^{i}(x')\partial_{i'}\delta(x', x) \tag{5.12}$$

and

$$\lambda^{c'}(x', x)[g_{mn}, p^{mn}] = p^{ab}(x)\lambda_{ab}{}^{c'}(x, x')[g_{mn}]. \tag{5.13}$$

Equation (5.12) tells us that the distribution $\Lambda(x', x)[g_{mn}]$ is local, being a linear combination of the delta function and its first derivatives. $\Lambda(x', x)[g_{mn}]$ is completely characterized by two sets of coefficients, $\Lambda(x')[g_{mn}]$ and $\Lambda^{i}(x')[g_{mn}]$, which are no longer distributions, but two fields, a scalar field and a vector field, constructed (possibly nonlocally)

from the metric. Similarly, Eq. (5.13) tells us that the multiplier $\lambda^{c}(x', x)[g_{mn}, p^{mn}]$ is a local functional of the momentum $p^{ab}(x)$. However, the coefficient $\lambda_{ab}{}^{c}(x, x')[g_{mn}]$ is still a general (possibly nonlocal) distribution. This fact considerably complicates the further argument. Once more we see that it is the presence of supermomentum constraints, rather than super-Hamiltonian constraints, which complicates the situation.

The locality restrictions (5.12) and (5.13) are themselves a direct consequence of the ultralocality of the truncated super-Hamiltonian $T(x)$. This fact—that $T(x)$ is a function of the undifferentiated canonical variables—led to the conclusion that the Poisson bracket (4.2) is ultralocal in at least one momentum variable, namely, in $p^{ab}(x)$. This conclusion became the cornerstone of our derivation of the local structure (5.12), (5.13) of the Lagrange multipliers.

At this stage, we have replaced Eq. (4.9) by Eq. (5.11), keeping $h_{ab\,c'd'}$ as an abbreviation for the expression (4.3). We now rewrite Eq. (5.11) in a form which is more suitable for our further considerations. Let $\delta g_{ab}(x)$ be an arbitrary variation of the metric and introduce the tensor density

$$\delta g^{ab}(x) \equiv G^{ab\,cd}(x)\delta g_{cd}(x) \tag{5.14}$$

[note that $\delta g^{ab}(x)$ is *not* the variation of the contravariant metric g^{ab}]. We multiply Eq. (5.11) by $\delta g^{ab}(x)$ and integrate it over x. For convenience, we rename the integration variables, $x \leftrightarrow x'$. Finally, we use the arbitrariness of $p^{cd}(x)$. This sequence of operations casts Eq. (5.11) into the new form

$$\begin{aligned} 2\,\delta k_{cd} = {} & G^{ab\,rs}\partial G_{rs\,cd}/\partial g_{mn}\cdot k_{mn}\delta g_{ab} \\ & + g^{-1/2}M^{ab}{}_{cd}\delta g_{ab} + \Lambda\delta g_{cd} - 2(\Lambda^{i}\delta g_{cd})_{|i} \\ & + \int d^{3}x'\,\lambda^{a'b'}{}_{(c|d)}(x', x)\,\delta g_{a'b'}(x'). \end{aligned} \tag{5.15}$$

All the terms in which we did not explicitly write the arguments are evaluated at the point x. The symbol δk_{cd} denotes the variation of the functional $k_{cd}(x)[g_{mn}]$ with respect to the metric. Note that the index pair ab in the expressions $M^{ab}{}_{cd}$ and $\lambda^{a'b'}{}_{(c|d)}$ was raised by the local supermetric (4.5). The rest of our derivation is based entirely on Eq. (5.15). We shall call it the principal equation.

VI. THE NONLOCAL PIECE OF THE GENERATOR WEAKLY VANISHES

The principal equation gives the variational derivative of k_{cd} as a function of the variables g_{ab} and k_{ab} and the still undetermined Lagrange multipliers $M^{ab}{}_{cd}$, Λ, Λ^{i}, and $\lambda^{a'b'}{}_{c}$. With the exception of the last term, all terms on the right-hand side of Eq. (5.15) are local in δg_{ab}. This helps us to determine the nonlocal part of the functional $k_{cd}(x)[g_{mn}]$, i.e., that part of $k_{cd}(x)[g_{mn}]$ which is not constructed from the metric $g_{mn}(x)$ and its derivatives up to a finite order at the point x.

Let x be the point at which the functional $k_{ij}(x)[g_{mn}]$ is taken and $x' \neq x$ a point at which g_{mn} is varied. The principal equation (5.15) then reduces to the statement

$$\frac{\delta k_{ij}(x)}{\delta g_{a'b'}(x')} = \lambda^{a'b'}{}_{(i|j)}(x', x). \tag{6.1}$$

Let x'' be another point different from x, $x'' \neq x$. By varying Eq. (6.1), we get

$$\frac{\delta^{2}k_{ij}(x)}{\delta g_{c''d''}(x'')\,\delta g_{a'b'}(x')} = \left(\frac{\delta\lambda^{a'b'}{}_{(i}(x', x)}{\delta g_{c''d''}(x'')}\right)_{|j)}. \tag{6.2}$$

We can now interchange the order of variations, $a'b'x' \leftrightarrow c''d''x''$, and obtain thus the integrability condition for the variational equation (6.1):

$$C^{a'b'\,c''d''}{}_{(i|j)} = 0. \tag{6.3}$$

Here,

$$\begin{aligned} & C^{a'b'\,c''d''}{}_{i}(x', x''; x) \\ & \equiv \frac{\delta\lambda^{a'b'}{}_{i}(x', x)}{\delta g_{c''d''}(x'')} - \frac{\delta\lambda^{c''d''}{}_{i}(x'', x)}{\delta g_{a'b'}(x')}. \end{aligned} \tag{6.4}$$

We again assume that the functional $k_{ij}(x)[g_{mn}]$ is evaluated at a generic metric g_{mn} which does not have any symmetry. Then, there are no Killing vectors, and Eq. (6.3) has only the trivial solution

$$C^{a'b'\,c''d''}{}_{i} = 0. \tag{6.5}$$

However, Eq. (6.5) is just the condition that $\lambda^{a'b'}{}_{i}$ be a functional gradient,

$$\exists\, k_{i}(x)[g_{mn}]:\ \lambda^{a'b'}{}_{i}(x', x) = \frac{\delta k_{i}(x)}{\delta g_{a'b'}(x')}. \tag{6.6}$$

We thus see that, for $x' \neq x$ and at a generic metric, the Lagrange multiplier necessarily has the form (6.6).

We now substitute the solution (6.6) back into the principal equation (6.1) and conclude that

$$\frac{\delta(k_{ij}(x) - k_{(i|j)}(x))}{\delta g_{a'b'}(x')} = 0 \quad \text{for } x' \neq x. \tag{6.7}$$

This means that

$$k_{ij}(x) = k_{(i|j)}(x) + \bar{k}_{ij}(x), \tag{6.8}$$

where the distribution $\delta\bar{k}_{ij}(x)/\delta g_{a'b'}(x')$ has its support on x. One knows[17] that such a distribution must be a combination of the delta function and its derivatives up to a finite order. This implies that $\bar{k}_{ij}(x)$ can depend only on the metric tensor and its derivatives up to a finite order at the point x. Equation (6.8) tells us that the functional $k_{ij}(x)[g_{mn}]$ splits into a piece $\bar{k}_{ij}(x)[g_{mn}]$ which is local in the metric g_{mn} and a piece $k_{(i|j)}(x)[g_{mn}]$ which is (in general) nonlocal. It also tells us that the nonlocal piece necessarily has a very special structure, namely, that it is the Killing form of a vector field $k_{i}(x)[g_{mn}]$.

The principal equation (5.15) is linear in k_{ij}, and we are thus able to fix the multipliers corresponding to the local piece of k_{ij} and to the Killing form piece of k_{ij} separately. We shall find the multipliers for the Killing form $k_{ij} = k_{(i|j)}$. We substitute the variation

$$\begin{aligned} \delta(k_{(c|d)}) &= (\delta k_{(c})_{|d)} - 2k_{m}\delta\Gamma^{m}{}_{cd} - 2k_{m}\delta\Gamma^{m}{}_{cd} \\ &= (k^{i}\delta^{ab}_{cd} - k^{a}\delta^{bi}_{cd} - k^{b}\delta^{ia}_{cd})\delta g_{ab|i} \end{aligned} \tag{6.9}$$

into the principal equation (5.15) and get

$$\int d^3x' \left(\lambda^{a'b'}{}_{(c}(x', x) - 2\frac{\delta k_{(c}(x)}{\delta g_{a'b')}(x')}\right)_{|d)} \delta g_{a'b'}(x')$$
$$+ (\partial G_{ab\,cd}/\partial g_{mn} \cdot k_{(m|n)} + g^{-1/2} M_{ab\,cd}$$
$$+ (\Lambda - 2\Lambda^{i}{}_{|i}) G_{ab\,cd}) \delta g^{ab}$$
$$+ 2(k^{a}\delta^{bi}_{cd} + k^{b}\delta^{ia}_{cd} - k^{i}\delta^{ab}_{cd} - \Lambda^{i}\delta^{ab}_{cd}) \delta g_{ab\,|i} = 0. \quad (6.10)$$

We put

$$\lambda^{a'b'}{}_{c}(x', x) = 2\frac{\delta k_{c}(x)}{\delta g_{a'b'}(x')} + \mu^{a'b'}{}_{c'}(x')\delta(x, x'), \quad (6.11)$$

where $\mu^{a'b'}{}_{c'}(x')[g_{mn}]$ is a tensor at x'. The last term in Eq. (6.11) is designed so that it does not yield higher derivatives of δg_{ab} than $\delta g_{ab\,|i}$ upon integration,

$$\int d^3x' \left(\mu^{a'b'}{}_{(c'}(x')\delta_{,d)}(x, x')\right.$$
$$\left. - 2\mu^{a'b'}{}_{m'}(x')\Gamma^{m'}{}_{c'd'}(x')\delta(x, x')\right) \delta g_{a'b'}(x')$$
$$= \mu^{ab}{}_{(c}\delta^{i}_{d)}\delta g_{ab\,|i} + \mu^{ab}{}_{(c|d)}\delta g_{ab}. \quad (6.12)$$

The arbitrariness of the variations $\delta g_{ab\,|i}$ and Eq. (5.14) lead to two sets of equations,

$$\mu^{ab}{}_{(c}\delta^{i}_{d)} + 2k^{a}\delta^{bi}_{cd} + k^{b}\delta^{ia}_{cd} - 2(k^{i} + \Lambda^{i})\delta^{ab}_{cd} = 0, \quad (6.13)$$

$$\mu_{ab\,(c|d)} + \partial G_{ab\,cd}/\partial g_{mn} \cdot k_{(m|n)}$$
$$+ (\Lambda - 2\Lambda^{i}{}_{|i}) G_{ab\,cd} + g^{-1/2} M_{ab\,cd} = 0, \quad (6.14)$$

for the remaining coefficients.

Equation (6.13) uniquely determines Λ^{i} and $\mu^{ab}{}_{c}$. To see that, we contract it in the indices cd and independently in the indices di. The resulting system of equations

$$\mu^{abi} + k^{(a}g^{b)i} - (k^{i} + \Lambda^{i})g^{ab} = 0,$$
$$\mu^{ab}{}_{c} + k^{(a}\delta^{b)}_{c} - \tfrac{1}{4}(k^{(a} + \Lambda^{(a})\delta^{b)}_{c} = 0 \quad (6.15)$$

has the solution

$$\Lambda^{i} = -k^{i}, \quad \mu^{ab}{}_{c} = -k^{(a}\delta^{b)}_{c}. \quad (6.16)$$

It is easy to check that the expressions (6.16) satisfy the original equation (6.13).

We now substitute the expressions (6.16) into Eq. (6.14), keeping in mind that $\mu_{ab\,(c|d)} = G_{ab\,mn}\mu^{mn}{}_{(c|d)}$. This yields the remaining multipliers

$$\Lambda = -k^{i}{}_{|i}, \quad (6.17)$$

$$M_{ab\,cd} = \tfrac{1}{2}(g_{cd}k_{(a|b)} - g_{ab}k_{(c|d)} + g_{ac}k_{[b|d]} + g_{bd}k_{[a|c]}$$
$$+ g_{ad}k_{[b|c)} + g_{bc}k_{[a|d]}). \quad (6.18)$$

The foregoing calculation shows that the Killing expression not only satisfies the principal equation for $x' \neq x$, but also satisfies, with a suitable choice of multipliers, the full principal equation (5.15). This result could have been foreseen from the very beginning. The generator $K = \int d^3x\, k_{(a|b)}p^{ab}$ must satisfy Eq. (2.30), with $\mu_a \longmapsto k_a$, and therefore it must also satisfy Eq. (2.11), with multipliers which are identified from Eqs. (2.34) and (2.35):

$$\Lambda(x', x) = k^{i'}(x')\partial_{i'}\delta(x', x),$$
$$\lambda^{a'}(x', x) = 2\frac{\delta k^{a'}(x')}{\delta g_{mn}(x)} G_{mn\,cd}(x)\, p^{cd}(x). \quad (6.19)$$

However, these are exactly the multipliers which we have recovered, by virtue of Eqs. (5.12)–(5.13), (6.11), (6.16)–(6.17), from our analysis of the principal equation.

Our discovery that the Killing form $k_{(a|b)}(x)[g_{mn}]$ satisfies the principal equation is thus not at all surprising. The important point of our proof is that the *general* solution (6.8) of the principal equation can differ from the Killing form only by a functional $\bar{k}_{ij}(x)[g_{mn}]$ which is *local* in the metric. In fact, we can disregard the Killing form altogether, because, by the argument of Sec. II, generators K and $\bar{K}$ whose coefficients differ by Eq. (6.8) are equivalent, $K \approx \bar{K}$. Expressed differently, the nonlocal piece of the generator weakly vanishes. Our task thereby reduces to that of finding the most general *local* solution $k_{ab}(x)[g_{mn}]$ of the principal equation. With this understanding, we shall once again omit the bar, $\bar{k}_{ab} \longmapsto k_{ab}$.

VII. THE GENERATOR MUST BE ULTRALOCAL

We thus assume that the functional $k_{cd}(x)[g_{mn}]$ is local in the metric, i.e., we assume that k_{cd} depends only on the metric and its derivatives up to a finite order N. Moreover, from Sec. III we already know that $k_{cd}(x)$ is a (symmetric) tensor under spatial transformations. In other words, $k_{cd}(x)[g_{mn}]$ is a tensor concomitant of the metric of order N.

It is well known that such a concomitant can depend on the derivatives of the metric only through the curvature tensor R_{abcd} and its covariant derivatives up to the order $\bar{N} = N - 2$. Furthermore, we can assume that the covariant derivatives are symmetrized, because any antisymmetric combination of the derivatives can be reduced by the commutation identity to terms containing only the derivatives of a lower order. As before, we indicate that the symmetrization has been performed by replacing the vertical stroke by the comma,

$$R_{abcd,m_1\cdots m_M} \equiv \mathop{\mathrm{S}}_{m} R_{abcd\,|m_1\cdots m_M}. \quad (7.1)$$

Similarly, we put

$$\delta g_{ab,m_1\cdots m_M} \equiv \mathop{\mathrm{S}}_{m} \delta g_{ab\,|m_1\cdots m_M}. \quad (7.2)$$

Using this convention, we can write

$$k_{cd} = k_{cd}(g_{ab}, R_{am_1m_2b}, R_{am_1m_2b,m_3}, \cdots, R_{am_1m_2b,m_3\cdots m_N}). \quad (7.3)$$

The variation of k_{cd} is thus expressed through the variation of the Riemann tensor,

$$\delta k_{cd} = k_{cd}{}^{ab}\delta g_{ab} + k_{cd}{}^{am_1m_2b}\delta R_{am_1m_2b}$$
$$+ \cdots + k_{cd}{}^{am_1m_2b,m_3\cdots m_N}\delta R_{am_1m_2b,m_3\cdots m_N}. \quad (7.4)$$

We have introduced the abbreviations

$$k_{cd}{}^{ab} \equiv \partial k_{cd}/\partial g_{ab},$$
$$k_{cd}{}^{am_1m_2b,m_3\cdots m_M} \equiv \partial k_{cd}/\partial R_{am_1m_2b,m_3\cdots m_M}. \quad (7.5)$$

If desired, the variation of the curvature tensor can be replaced by the variation of the metric,

$$\delta R_{am_1m_2b} = \Delta_{am_1m_2b}{}^{cdn_1n_2}\delta g_{cd,n_1n_2} + \Delta_{am_1m_2b}{}^{cd}\delta g_{cd}, \quad (7.6)$$

with

$$\Delta_{am_1m_2b}{}^{cdn_1n_2} \equiv \tfrac{1}{2}(\delta_{ab}^{cd}\delta_{m_1m_2}^{n_1n_2} + \delta_{m_1m_2}^{c\ d}\delta_{a\ b}^{n_1n_2} - \delta_{am_2}^{cd}\delta_{b\ m_1}^{n_1n_2} - \delta_{bm_1}^{cd}\delta_{a\ m_2}^{n_1n_2}) \tag{7.7}$$

and

$$\Delta_{am_1m_2b}{}^{cd} \equiv \tfrac{1}{8}(\delta_{(a}^{(c}R^{d)}{}_{m_1m_2b} + \delta_{m_2}^{(c}R^{d)}{}_{bam_1} - \delta_{m_1}^{(c}R^{d)}{}_{am_2b} - \delta_{b}^{(c}R^{d)}{}_{m_2am_1}). \tag{7.8}$$

Note that there is no first order term in Eq. (7.6).

When we take the covariant derivative of Eq. (7.6), $|n_3\cdots n_N$, and symmetrize it in $n_3\cdots n_N$, we obtain the variational formula for $\delta R_{am_1m_2b,\,n_3\cdots n_N}$. The leading term in this formula is

$$\delta_{(N)}R_{am_1m_2b,\,n_3\cdots n_N} = \Delta_{am_1m_2b}{}^{cdn_1n_2}\delta g_{cd,\,n_1\cdots n_N}. \tag{7.9}$$

In general, we shall use the symbol $\delta_{(M)}$ to denote the M th order term (proportional to $\delta g_{ab,m_1\cdots m_M}$) in a given variation. Because the symmetrization of $\delta g_{cd,\,n_1n_2|n_3\cdots n_N}$ into $\delta g_{cd,\,n_1\cdots n_N}$ brings the terms down only to the order $N-2$, and the differentiated equation does not contain any terms of order $N-1$, the terms of the order $N-1$ are missing in the variation $\delta R_{am_1m_2b,\,m_3\cdots m_N}$: $\delta_{(N-1)}R_{am_1m_2b,\,m_3\cdots m_N} = 0$.

This observation will become important in the subsequent argument.

The variation of k_{cd} contains the derivatives of δg_{ab} up to order N. Such terms should be compensated in the principal equation (5.15) by the multipliers. The compensating λ term

$$\int d^3x'\,\lambda^{a'b'}{}_c(x', x)\,\delta g_{a'b'}(x') \tag{7.10}$$

is a bipoint distribution acting on $\delta g_{a'b'}(x')$ as a test function. This action should yield the derivatives of δg_{ab} up to the order $N-1$; the Killing derivative $(c|d)$ of the term (7.10) in Eq. (5.15) then raises the order to N. We thus write

$$\int d^3x'\,\lambda^{a'b'}{}_c(x', x)\,\delta g_{a'b'}(x') = \lambda^{ab}{}_c(x)\delta g_{ab}(x) + \lambda^{ab}{}_c{}^{m_1}(x)\delta g_{ab,m_1}(x) + \cdots + \lambda^{ab}{}_c{}^{m_1\cdots m_N}(x)\delta g_{ab,m_1\cdots m_N}(x). \tag{7.11}$$

The coefficients $\lambda^{ab}{}_c{}^{m_1\cdots m_M}(x)$ cannot depend on $R_{an_1n_2b,\,n_3\cdots n_N}$; if they did, $\lambda^{ab}{}_{(c}{}^{m_1\cdots m_M}{}_{|d)}$ terms on the right-hand side of the principal equation would contain the $\bar{N} = N-1$ order derivatives $R_{an_1n_2b,\,n_3\cdots n_{N+1}}$ of the curvature tensor, which are absent in the variation (7.4), (7.5).

We want to prove that k_{cd} is weakly equivalent to a tensor k_{cd} which depends only on the metric, not on the curvature tensor or its derivatives. We use a recursive procedure, showing first that $\bar{k}_{cd}$ can be made independent of the highest order derivative of R_{abcd}, i.e., of the derivative of the order $\bar{N} = N-2$, and then repeat the argument until $\bar{N} = 0$. In the proof, it is sufficient to single out only the two highest order terms in the principal equation. This is possible, because the variation of the metric and its *symmetrized* covariant derivatives at the point x can be assigned arbitrarily and thus the individual terms in $\delta g_{ab,m_1\cdots m_N}$ must vanish independently of each other.

From Eq. (7.11) we see that

$$\delta_{(N)}\int d^3x'\,\lambda^{a'b'}{}_c(x', x)\delta g_{a'b'}(x') = \lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-1}}\delta_{d)}^{m_N}\delta g_{ab,m_1\cdots m_N} \tag{7.12}$$

and

$$\delta_{(N-1)}\int d^3x'\,\lambda^{a'b'}{}_c(x', x)\delta g_{a'b'}(x') = (\lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-2}}\delta_{d)}^{m_{N-1}} + \lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-1}}{}_{|d)})\delta g_{ab,m_1\cdots m_{N-1}}, \tag{7.13}$$

because the symmetrization of the derivative $\delta g_{ab,m_1\cdots m_{N-1}|m_N}$ does not contribute to the term of the order $N-1$. For $N\geqslant 2$, no other multipliers join the term (7.12) in the principal equation. Similarly, for $N\geqslant 3$, no other multipliers join the term (7.13). The two highest order terms in the variation δk_{cd} are

$$\delta_{(N)}k_{cd} = k_{cd}{}^{am_1m_2b,m_3\cdots m_N}\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N}, \tag{7.14}$$

$$\delta_{(N-1)}k_{cd} = k_{cd}{}^{am_1m_2b,m_3\cdots m_{N-1}}\delta_{(N-1)}R_{am_1m_2b,m_3\cdots m_{N-1}},$$

because $\delta_{(N-1)}R_{am_1m_2b,m_3\cdots m_N} = 0$. The two highest order terms of the principal equation thus yield

$$2k_{cd}{}^{am_1m_2b,m_3\cdots m_N}\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N} = \lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-1}}\delta_{d)}^{m_N}\delta g_{ab,m_3\cdots m_N} \quad \text{for } N\geqslant 2, \tag{7.15}$$

and

$$2k_{cd}{}^{am_1m_2b,m_3\cdots m_{N-1}}\delta_{(N-1)}R_{am_1m_2b,m_3\cdots m_{N-1}} = (\lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-2}}\delta_{d)}^{m_{N-1}} + \lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-1}}{}_{|d)})\delta g_{ab,m_1\cdots m_{N-1}} \quad \text{for } N\geqslant 3. \tag{7.16}$$

We shall first analyze the consequences of Eq. (7.15). If the $\delta g_{ab,m_1\cdots m_N}$ assume all possible values, Eq. (7.9) generates all the possible variations $\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N}$ which have the symmetries of the Riemann tensor and satisfy the differentiated Bianchi identities

$$\delta_{(N)}R_{am_1|m_2b,\,m_3|m_4\cdots m_N} \equiv 0. \tag{7.17}$$

However, there are

$$3\cdot\tfrac{1}{2}(N+2)(N+3) \tag{7.18}$$

more independent variations $\delta g_{ab,m_1\cdots m_N}$ of the metric than there are independent variations $\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N}$ of the curvature tensor.

To see that, note that a completely symmetric tensor of the rank N in a three-dimensional space has $\frac{1}{2}(N+1)(N+2)$ algebraically independent components. The variation $\delta g_{ab,m_1\cdots m_N}$ is symmetric in the pair ab and completely symmetric in the N indices $m_1\cdots m_N$. Consequently, it has

$$3(N+1)(N+2) \tag{7.19}$$

algebraically independent components. The spatial Riemann curvature tensor $R_{am_1m_2b}$ has six independent components. The variation $\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N}$ is completely symmetric in the $N-2$ indices $m_3\cdots m_N$ and so it has $3N(N-1)$ components. Not all of these components, however, are independent, because they must satisfy the differentiated Bianchi identities (7.17). The antisymmetric index pair am_1 in Eq. (7.17) ranges through three independent combinations, the

completely antisymmetric triplet $\{m_2 b, m_3\}$ has only one independent combination, while the $N-3$ completely symmetric indices $m_4 \cdots m_N$ can assume $\frac{1}{2}(N-2)(N-1)$ independent combinations. Altogether there are thus $\frac{3}{2}(N-2)(N-1)$ equations (7.17) and

$$3N(N-1) - \tfrac{3}{2}(N-2)(N-1) = \tfrac{3}{2}(N-1)(N+2) \quad (7.20)$$

algebraically independent components of $\delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N}$. The difference between the numbers (7.19) and (7.20) leads to the excess (7.18).

There must thus be $3 \cdot \frac{1}{2}(N+2)(N+3)$ variations $\delta g_{ab, m_1 \cdots m_N}$ which leave the derivatives of the curvature tensor unchanged,

$$\delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N} = 0. \quad (7.21)$$

It is easy to check that these are generated through the formula

$$\delta g_{ab, m_1 \cdots m_N} = \delta h_{(a,b) m_1 \cdots m_N} \quad (7.22)$$

by the $3 \cdot \frac{1}{2}(N+2)(N+3)$ tensorial quantities $\delta h_{a, bm_1 \cdots m_N}$ which are symmetric in the indices $bm_1 \cdots m_N$. Such a counting is well known from the theory of normal Riemann coordinates. The variation (7.22), (7.21) is one induced by an infinitesimal coordinate transformation.

We now use the variation (7.21), (7.22) in Eq. (7.15). This eliminates any reference to the generator k_{cd} and leaves us with the condition

$$\lambda^{ab}{}_{(c}{}^{m_1 \cdots m_{N-1}} \delta^{m_N}_{d)} \delta h_{a, bm_1 \cdots m_N} = 0 \quad (7.23)$$

on the multiplier. (Because λ is symmetric in the pair ab, we dropped the symmetrization bracket in $\delta h_{(a,b) m_1 \cdots m_N}$.)

Equation (7.23) implies that

$$\mathop{\mathrm{S}}_{b,m} \lambda^{ab}{}_{c}{}^{m_1 \cdots m_{N-1}} = 0. \quad (7.24)$$

The system (7.24) imposes $3 \cdot 3 \cdot \frac{1}{2}(N+1)(N+2)$ conditions on the $6 \cdot 3 \cdot \frac{1}{2} N(N+1)$ components $\lambda^{ab}{}_{c}{}^{m_1 \cdots m_{N-1}}$. It means that the multipliers can be expressed in terms of some $3 \cdot \frac{3}{2}(N+1)(N-2)$ independent quantities. A comparison with Eq. (7.20) indicates that it is convenient to choose these quantities as components of a tensor $K_c{}^{am_1 m_2 b, m_3 \cdots m_{N-1}}$ which has the symmetries of a differentiated Riemann tensor in the upper indices. Due to the symmetries of the coefficients λ and K, there is virtually only one way how to construct λ out of K. We formulate the connection as a lemma.

Lemma: Let $\lambda^{ab, m_1 \cdots m_N}$ be such that $\mathop{\mathrm{S}}_{b,m} \lambda^{ab, m_1 \cdots m_N} = 0$. Then there exists a tensor $K^{am_1 m_2 b, m_3 \cdots m_N}$ with the symmetries

$$K^{am_1 m_2 b, \cdots} = K^{m_2 b a m_1, \cdots} = -K^{m_1 a m_2 b, \cdots} = -K^{am_1 bm_2, \cdots},$$

$$K^{am_1 | m_2 b, m_3 | m_4 \cdots m_N} = 0, \quad (7.25)$$

which generates $\lambda^{ab, m_1 \cdots m_N}$ according to the rule

$$\lambda^{ab, m_1 \cdots m_N} = 4 \mathop{\mathrm{S}}_{m} K^{am_1 m_2 b, m_3 \cdots m_N}. \quad (7.26)$$

We prove the lemma by showing that

$$K^{am_1 m_2 b, m_3 \cdots m_N} \equiv \tfrac{1}{4}(N-1)(N+1)^{-1} (\lambda^{[ab, m_1] m_2 m_3 \cdots m_N} - \lambda^{[am_2, m_1] bm_3 \cdots m_N}) \quad (7.27)$$

satisfies its requirements. It is trivial to check that the expression (7.27) has the symmetries (7.25). Next, we symmetrize Eq. (7.27) and get

$$\mathop{\mathrm{S}}_{m} K^{am_1 m_2 b, m_3 \cdots m_N} = \tfrac{1}{4}(N-1)(N+1)^{-1} (\lambda^{ab, m_1 \cdots m_N} - \mathop{\mathrm{S}}_{m} \lambda^{bm_1, m_2 \cdots m_N a} - \mathop{\mathrm{S}}_{m} \lambda^{am_1, m_2 \cdots m_N b} + \mathop{\mathrm{S}}_{m} \lambda^{m_1 m_2, m_3 \cdots m_N ab}). \quad (7.28)$$

The first term on the right-hand side of Eq. (7.28) is already symmetric in m. The second and the third terms can be evaluated from the equation

$$\lambda^{b[m_1, m_2 \cdots m_N]a} + \lambda^{ba, m_1 \cdots m_N} = \lambda^{b[m_1, m_2 \cdots m_N a]} = 0, \quad (7.29)$$

which can be written in the form

$$\mathop{\mathrm{S}}_{m} \lambda^{bm_1, m_2 \cdots m_N a} = -N^{-1} \lambda^{ba, m_1 \cdots m_N} = \mathop{\mathrm{S}}_{m} \lambda^{am_1, m_2 \cdots m_N b}. \quad (7.30)$$

Similarly, the fourth term is obtained from the equation

$$\lambda^{m_1 [m_2, m_3 \cdots m_N] ab} + \lambda^{m_1 a, m_2 \cdots m_N b} + \lambda^{m_1 b, m_2 \cdots m_N a} = \lambda^{m_1 [m_2, m_3 \cdots m_N ab]} = 0. \quad (7.31)$$

After we symmetrize Eq. (7.31) in m,

$$(N-1) \mathop{\mathrm{S}}_{m} \lambda^{m_1 m_2, m_3 \cdots m_N ab} + \mathop{\mathrm{S}}_{m} \lambda^{am_1, m_2 \cdots m_N b} + \mathop{\mathrm{S}}_{m} \lambda^{bm_1, m_2 \cdots m_N a} = 0, \quad (7.32)$$

and use Eq. (7.30), we get

$$\mathop{\mathrm{S}}_{m} \lambda^{m_1 m_2, m_3 \cdots m_N ab} = \frac{2}{N(N-1)} \lambda^{ab, m_1 \cdots m_N}. \quad (7.33)$$

When we sutstitute the expressions (7.30) and (7.33) into Eq. (7.28), we recover Eq. (7.26).

We now apply the lemma to Eq. (7.24). We learn that there exists a tensor $K_c{}^{am_1 m_2 b, m_3 \cdots m_{N-1}}$ such that

$$\lambda^{ab}{}_{c}{}^{m_1 \cdots m_{N-1}} = 4 \mathop{\mathrm{S}}_{m} K_c{}^{am_1 m_2 b, m_3 \cdots m_{N-1}}. \quad (7.34)$$

This enables us to rearrange the right-hand side of Eq. (7.15) so that the variations $\delta g_{ab, m_1 \cdots m_N}$ are collected into $\delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N}$:

$$\begin{aligned}
&\lambda^{ab}{}_{(c}{}^{m_1 \cdots m_{N-1}} \delta^{m_N}_{d)} \delta g_{ab, m_1 \cdots m_N} \\
&\quad = 4 \mathop{\mathrm{S}}_{m_1 \cdots m_{N-1}} K_{(c}{}^{am_1 m_2 b, m_3 \cdots m_{N-1}} \delta^{m_N}_{d)} \delta g_{ab, m_1 \cdots m_N} \\
&\quad = 4 K_{(c}{}^{am_1 m_2 b, m_3 \cdots m_{N-1}} \delta^{m_N}_{d)} \delta g_{ab, m_1 \cdots m_N} \\
&\quad = K_{(c}{}^{am_1 m_2 b, m_3 \cdots m_{N-1}} \delta^{m_N}_{d)} (\delta g_{ab, m_1 m_2 m_3 \cdots m_N} + \delta g_{m_2 m_1, bam_3 \cdots m_N} \\
&\qquad - \delta g_{am_2, m_1 bm_3 \cdots m_N} - \delta g_{bm_1, m_2 am_3 \cdots m_N}) \\
&\quad = 2 K_{(c}{}^{am_1 m_2 b, m_3 \cdots m_{N-1}} \delta^{m_N}_{d)} \delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N}. \quad (7.35)
\end{aligned}$$

After this rearrangement, Eq. (7.15) reads

$$(k_{cd}{}^{am_1 m_2 b, m_3 \cdots m_N} - K_{(c}{}^{am_1 m_2 b, m_3 \cdots m_{N-1}} \delta^{m_N}_{d)}) \times \delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N} = 0, \quad (7.36)$$

where $\delta_{(N)} R_{am_1 m_2 b, m_3 \cdots m_N}$ is an arbitrary tensor with appropriate symmetries. In particular, Eq. (7.36) must hold for

$\delta_{(N)} R_{am_1m_2b,m_3\cdots m_N} = R_{am_1m_2b,m_3\cdots m_N}$:

$$\frac{\partial k_{cd}}{\partial R_{am_1m_2b,m_3\cdots m_N}} R_{am_1m_2b,m_3\cdots m_N} = K_{(c}{}^{am_1m_2b,m_3\cdots m_{N-1}} R_{am_1m_2b,m_3\cdots m_{N-1}d)}. \tag{7.37}$$

However, we have seen that the coefficients $\lambda^{ab}{}_{c}{}^{m_1\cdots m_M}$, and hence the coefficient $K_c{}^{am_1m_2b,m_3\cdots m_{N-1}}$, cannot depend on $R_{am_1m_2b,m_3\cdots m_N}$. Therefore, k_{cd} must be *linear* in the highest order derivative $R_{am_1m_2b,m_3\cdots m_N}$:

$$k_{cd} = K_{(c}{}^{am_1m_2b,m_3\cdots m_{N-1}} R_{am_1m_2b,m_3\cdots m_{N-1}d)} + \tilde{k}_{cd}(g_{ab}, R_{am_1m_2b,\cdots}, R_{am_1m_2b,m_3\cdots m_{N-1}}). \tag{7.38}$$

Our goal is to prove that k_{cd}, at least in the highest order, has the structure of a Killing form. In other words, we want to prove that

$$k_{cd} = K_{(c|d)} + \bar{k}_{cd}, \tag{7.39}$$

where K_c and $\bar{k}_{cd}$ are tensor concomitants of the order $N-1$. Expanding the Killing form $K_{(c|d)}$, we cast Eq. (7.39) into an equivalent statement

$$k_{cd} = \frac{\partial K_{(c}}{\partial R_{am_1m_2b,m_3\cdots m_{N-1}}} \delta^{m_N}_{d)} R_{am_1m_2b,m_3\cdots m_N} + \bar{k}_{cd}. \tag{7.40}$$

When we compare Eq. (7.38) with Eq. (7.40), we see that the only point which remains to be proved is that the term $K_{(c}{}^{am_1m_2b,\cdots m_{N-1}} R_{am_1m_2b,m_3\cdots m_{N-1}d)}$ is an R-gradient:

$$K_{(c}{}^{am_1m_2b,m_3\cdots m_{N-1}} R_{am_1m_2b,m_3\cdots m_{N-1}d)} = \frac{\partial K_{(c}}{\partial R_{am_1m_2b,m_3\cdots m_{N-1}}} R_{am_1m_2b,m_3\cdots m_{N-1}d)}. \tag{7.41}$$

The integrability condition of Eq. (7.41) can be written in the form

$$\left(\frac{\partial K_{(c}{}^{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}}}{\partial R_{am_1m_2b,m_3\cdots m_{N-1}}} - \frac{\partial K_{(c}{}^{am_1m_2b,m_3\cdots m_{N-1}}}{\partial R_{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}}}\right) R_{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{n-1}d)} \delta_{(N-1)} R_{am_1m_2b,m_3\cdots m_{N-1}} = 0. \tag{7.42}$$

We shall show that Eq. (7.42) is a consequence of the $(N-1)$-order principal equation (7.16). Indeed, Eq. (7.16) must hold as an identity in g_{ab}, $R_{am_1m_2b}$, ..., $R_{am_1m_2b,m_3\cdots m_N}$. Therefore, the terms linear in $R_{am_1m_2b,m_3\cdots m_N}$ must balance independently of the lower order terms. When selecting such terms, we can disregard $\lambda^{ab}{}_{c}{}^{m_1\cdots m_{N-1}} \doteq 0$, while Eq. (7.38) gives

$$k_{cd}{}^{am_1m_2b,m_3\cdots m_{N-1}} \doteq \frac{\partial K_{(c}{}^{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}}}{\partial R_{am_1m_2b,m_3\cdots m_{N-1}}} R_{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}}. \tag{7.43}$$

Relying on the rearrangement (7.35), we can write

$$\lambda^{ab}{}_{(c}{}^{m_1\cdots m_{N-1}} \delta g_{ab,m_1\cdots m_{N-1}|d)} \doteq 2 \frac{\partial K_{(c}{}^{am_1m_2b,m_3\cdots m_{N-1}}}{\partial R_{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}}} R_{\bar{a}\,\bar{m}_1\bar{m}_2\bar{b},\bar{m}_3\cdots\bar{m}_{N-1}d)} \times \delta_{(N-1)} R_{am_1m_2b,m_3\cdots m_{N-1}}. \tag{7.44}$$

The selected terms (7.43) and (7.44) in Eq. (7.16) give us exactly the integrability condition (7.42).

In this way, we have proved that k_{cd} has the structure (7.39). This means, however, that the original k_{cd} of order N is weakly equivalent to another tensor concomitant, $\bar{k}_{cd}$, of order $N-1$. The argument can be repeated for $N-1$ and then recursively down to the third order, in which Eq. (7.16) is still valid. In the second order, only Eq. (7.15) holds unchanged. From this equation alone, we conclude as in Eq. (7.24) that

$$\lambda^{a(b}{}_{c}{}^{m_1)} = 0. \tag{7.45}$$

We perform the cyclic permutation of the indices abm_1 and subtract the resulting equations from Eq. (7.45). In this way, we learn that $\lambda^{ab}{}_{c}{}^{m_1} = 0$. Equation (7.15) alone then implies that k_{cd} does not depend on $R_{am_1m_2b}$.

We can thus conclude that the original tensor concomitant of order N is weakly equivalent to an ultralocal concomitant k_{cd}, i.e., to a tensor which is constructed entirely from the metric. The only such tensor, however, is a multiple of the metric itself,

$$k_{cd} = \kappa g_{cd}, \quad \kappa \in \mathbb{R}. \tag{7.46}$$

It is straightforward to check that the metric tensor satisfies the full principal equation (5.15) when we assign to the remaining multipliers the values

$$\Lambda = -\tfrac{3}{2}\kappa, \quad M^{ab}{}_{cd} = 0, \quad \Lambda^{i} = 0, \quad \lambda^{ab}{}_{c} = 0. \tag{7.47}$$

Indeed, it is easy to show that this choice of multipliers is uniquely determined by the principal equation. For $k_{cd} = g_{cd}$ and all the multipliers $\lambda^{ab}{}_{c}{}^{m_1\cdots m_N}$, $N \geqslant 1$, transformed away, the principal equation takes the form

$$(\tfrac{3}{4}\delta^{ab}_{cd} + \tfrac{3}{2} g^{-1/2} M^{ab}{}_{cd} + (\tfrac{1}{2}\Lambda - \Lambda^{i}{}_{|i})\delta^{ab}_{cd} + \tfrac{1}{2}\lambda^{ab}{}_{(c|d)})\delta g_{ab} + (-\Lambda^{m}\delta^{ab}_{cd} + \tfrac{1}{2}\lambda^{ab}{}_{(c}\delta^{m}_{d)})\delta g_{ab|m} = 0. \tag{7.48}$$

The coefficients of δg_{ab} and $\delta g_{ab|m}$ must vanish separately. Take the equation

$$-\Lambda^{m}\delta^{ab}_{cd} + \tfrac{1}{2}\lambda^{ab}{}_{(c}\delta^{m}_{d)} = 0 \tag{7.49}$$

and contract it in the pair bd:

$$-4\Lambda^{m}\delta^{a}_{c} + \lambda^{am}{}_{c} + \lambda^{ab}{}_{b}\delta^{m}_{c} = 0. \tag{7.50}$$

Two further contractions, in the pair ac and in the pair mc, ensure that $\Lambda^{m} = 0 = \lambda^{ab}{}_{b}$. Equation (7.50) then implies that $\lambda^{ab}{}_{c} = 0$. Under these conditions, the coefficient of δg_{ab} reduces to $\frac{1}{2}(\frac{3}{2} + \Lambda)\delta^{ab}_{cd} + \frac{1}{2}M^{ab}{}_{cd}$. We lower the index pair ab by the supermetric and impose the condition that this coefficient vanishes:

$$(\tfrac{3}{2} + \Lambda)G_{ab\,cd} + g^{-1/2}M_{ab\,cd} = 0. \tag{7.51}$$

Because $G_{ab\,cd}$ is symmetric and $M_{ab\,cd}$ antisymmetric in the interchange of pairs $ab \leftrightarrow cd$, the two terms in Eq. (7.51) must vanish separately. This shows that the multipliers are uniquely fixed as in Eq. (7.47).

We now restate the main result of the whole argument. The ultralocal geometrodynamics, based on the kinetic super-Hamiltonian T, has a unique conditional symmetry (7.46), which is itself ultralocal. All Lagrange multipliers, with the exception of Λ, are gauged to zero by the proper choice (7.46) of the representative of the equivalence class (2.33). Consequently, only the super-Hamiltonian constraint, but not the supermomentum constraint, enforces the

conservation of the generator (7.46). It does so locally, Λ being a constant.

Due to the ultralocal character of the super-Hamiltonian T and of the generator (7.46), the tensor $k_{cd} = \kappa g_{cd}(x)$ can be interpreted point by point as a conformal Killing (super)vector of the local supermetric (4.5). One can scale the super-Hamiltonian T by an ultralocal factor $g^{1/2}(x)$ in such a way that the generator (7.46) becomes an unconditional symmetry of the scaled super-Hamiltonian

$$\tilde{T}(x) = g^{1/2}(x)T(x). \tag{7.52}$$

Expressed differently, the tensor $k_{cd} = \kappa g_{cd}(x)$ is point by point a true Killing (super)vector of the scaled local supermetric

$$\tilde{G}^{ab\,cd}(x) = g^{-1/2}(x)G^{ab\,cd}(x). \tag{7.53}$$

The constant κ can be fixed by the normalization condition:

$$\tilde{G}^{ab\,cd}k_{ab}k_{cd} = -1 \Rightarrow |\kappa| = 6^{-1/2}. \tag{7.54}$$

The minus sign in Eq. (7.54) indicates that k_{ab} has a "timelike" character.

The fact that $k_{cd} = \kappa g_{cd}$ is a conformal Killing (super)vector of the local supermetric (4.5) is, of course, well known.[18] What is new is the proof that, besides it, the ultralocal geometrodynamics has no other conditional symmetry.[19]

VIII. THE SCALAR CURVATURE BREAKS THE ULTRALOCAL SYMMETRY

Conditional symmetry in ultralocal geometrodynamics is determined by Eqs. (4.9) and (4.3). By analyzing these equations, we draw the conclusion that ultralocal geometrodynamics has a unique conditional symmetry (7.46). Full geometrodynamics differs from the ultralocal geometrodynamics by the curvature term $V(x) = -g^{1/2}R$ in the super-Hamiltonian. Its presence restricts the conditional symmetry by another condition, Eq. (4.10). From Eqs. (7.47) and (5.12) we learned that the multiplier $\Lambda(x', x)$ is ultralocal,

$$\Lambda(x', x) = \Lambda\delta(x', x), \quad \Lambda = -\tfrac{3}{2}\kappa. \tag{8.1}$$

The condition (4.10) thus takes the form

$$-g^{1/2}G^{ab}k_{ab} + G^{ab\,cd}k_{ab\,|cd} = -\Lambda g^{1/2}R. \tag{8.2}$$

The ultralocal generator $k_{ab} = \kappa g_{ab}$ satisfies Eq. (8.2) for $\Lambda = -\frac{1}{2}\kappa$, but it does not satisfy it for the required value $\Lambda = -\frac{3}{2}\kappa$. The scalar curvature term thus breaks the conditional symmetry of the ultralocal theory. Equations (4.9) and (4.10) for the generator coefficient k_{ab} are incompatible, but, strangely enough, the incompatibility boils down to a mere numerical difference of the conformal factor. The ultralocal generator

$$K \equiv \kappa \int d^3x\, p(x) \tag{8.3}$$

scales the kinetic term by the factor $\Lambda = -\frac{3}{2}\kappa$, while it scales the potential term by the factor $\Lambda = -\frac{1}{2}\kappa$:

$$\begin{aligned} [K, T(x)] &= -\tfrac{3}{2}\kappa T(x), \\ [K, V(x)] &= -\tfrac{1}{2}\kappa V(x). \end{aligned} \tag{8.4}$$

Small as this discrepancy may seem, an uncomfortable conclusion nevertheless follows: Geometrodynamics does not have any symmetry, not even a conditional one.

IX. INHOMOGENEOUS GENERATORS

We have seen that no dynamical variable which is linear and homogeneous in the momentum can be conditionally conserved. We can generalize this result to linear inhomogeneous functionals

$$K^* = K + K^0, \tag{9.1}$$

where K has the form (2.1) and K^0 is a functional which depends only on the metric.[20]

Write the conditions (2.3) and (2.4) that K^* is weakly conserved. Because the Poisson bracket $[K^0, H_N]$ does not depend on the momentum and $[K, H_N]$ is linear in the momentum, we can still conclude, as in Eqs. (2.14)–(2.15), that $\Lambda_N(x') = 0$ and $\lambda_N{}^{a'}(x')$ depend only on the metric. The strong equation (2.16) written for K^* then splits naturally into two pieces. The first piece collects those terms which do not contain the momentum,

$$[K^0, H_N] = 0. \tag{9.2}$$

It tells us that K^0 is a spatial invariant. The second piece collects the terms containing the momentum. It replicates Eq. (2.16) for the homogeneous part K of the generator K^*. The argument of Sec. III then leads to an equivalent functional $\bar{K} \approx K$, which is a spatial invariant. The functional $\bar{K}^* = \bar{K} + K^0$ is thus equivalent to K^* and is also a spatial invariant. We choose it to represent the equivalence class and omit the bar. As a result, we can assume that K^* satisfies the invariance condition (3.2).

Next, we evaluate the Poisson bracket $[K^*, H(x)]$. Because $[K^0, V(x)] = 0$, we get

$$[K^*, H(x)] = [K, T(x)] + [K, V(x)] + [K^0, T(x)]. \tag{9.3}$$

From Sec. IV we know that $[K, V(x)]$ depends only on the metric. The new term

$$[K^0, T(x)] = 2p^{ab}(x)G_{ab\,cd}(x)\delta K^0/\delta g_{cd}(x) \tag{9.4}$$

is linear in the momentum. We conclude, as before, that the multiplier $\Lambda_N(x')[g_{mn}]$ cannot depend on the momentum. The multiplier $\lambda_N{}^{a'}(x')$, however, can acquire an inhomogeneous piece $\lambda^0{}_N{}^{a'}(x')$:

$$\lambda^*{}_N{}^{a'}(x') = \lambda_N{}^{a'}(x') + \lambda^0{}_N{}^{a'}(x'), \tag{9.5}$$

where $\lambda_N{}^{a'}(x')$ is a linear homogeneous functional of the momentum and $\lambda^0{}_N{}^{a'}(x')$ depends only on the metric. Equation (2.11) written for K^* then splits into three pieces in the powers of the momentum. The quadratic piece and the absolute piece replicate our old equations (4.9)–(4.10) for the homogeneous generator K. Our old argument allows us to conclude that this generator must vanish weakly, $K \approx 0$. The linear piece of Eq. (2.11) is completely decoupled from the previous two pieces, and it reads

$$[K^0, T(x)] = \int d^3x'\, \lambda^{0\,a'}(x', x)H_{a'}(x'). \tag{9.6}$$

We use Eq. (9.4) and write Eq. (9.6) as a weak equation:

$$H_{a'}(x') = 0 \forall\, x'$$
$$\Rightarrow \quad p^{ab}(x)G_{ab\,cd}(x)\delta K^0/\delta g_{cd}(x) = 0. \tag{9.7}$$

The condition $H_{a'}(x') = 0$ limits only the derivatives of p^{ab} and leaves p^{ab} at the point x completely arbitrary. Therefore, $\delta K^0/\delta g_{cd}(x) = 0$. In other words, the functional K^0 cannot depend on the metric, and so it must reduce to a numerical constant. This shows that a linear inhomogeneous functional (9.1) which is conditionally conserved,

$$[K^*, H_N + H_N] \approx 0, \tag{9.8}$$

is weakly equivalent to a numerical constant.

X. A SUMMARY

The statement that geometrodynamics does not have a conditional symmetry has a simple meaning, but a rather complicated proof. To gain the *unity of apprehension*, we shall summarize the main line of argument in a decalogue of steps. The relevant equations have the boldface numbering in the text.

(I) We say that a dynamical variable $K\,[g_{ab}, p^{ab}]$ generates a conditional symmetry in geometrodynamics if K is a linear homogeneous functional of the gravitational momentum p^{ab} [Eq. (2.1)] which has a weakly vanishing Poisson bracket with the geometrodynamical Hamiltonian [Eq. (2.2)].

(II) The weak equation (2.2) is turned into the strong equations (2.11) and (2.16) by the method of Lagrange multipliers. The dependence of the multipliers on the gravitational momentum [Eqs. (2.12)–(2.15)] can be inferred from the momentum dependence of the constraints.

(III) The closure relations (2.20)–(2.22) of the constraints imply that the generators K fall into equivalence classes (2.33) modulo the supermomentum constraint. A transition within a single equivalence class induces the gauge transformation (2.34)–(2.36) of the multipliers.

(IV) By virtue of the Jacobi identity, the multiplier (2.15) is restricted by Eq. (3.9). This equation ensures that the multiplier (2.15) can be gauged away by the transformation (2.36). Each equivalence class is then represented by an unconditional spatial invariant (3.2).

(V) The super-Hamiltonian is split into the "kinetic" and "potential" parts (4.1) and its Poisson bracket with the generator K is evaluated [Eqs. (4.2)–(4.3), (4.8)]. The conditional symmetry (2.11) of the super-Hamiltonian is interpreted as the conditional symmetry (4.9), (4.10) of its kinetic and potential terms.

(VI) The ultralocality of the kinetic term imposes locality restrictions on the remaining Lagrange multipliers. These restrictions are summarized by the principal equation (5.15).

(VII) The principal equation leads to a variational integrability condition (6.3)–(6.4) which, for generic metrics, ensures that the generator is weakly equivalent to a local generator.

(VIII) The principal equation for the local generator is split into different orders M of the symmetrized variations $\delta g_{ab,\,m_1\cdots m_M}$. Only the two highest order equations, $M = N$ [Eq. (7.15)] and $M = N - 1$ [Eq. (7.16)], need to be considered. The N th-order variation $\delta g_{ab,m_1\cdots m_N}$ of the metric is decomposed into those variations which are induced by the variations $\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N}$ of the differentiated curvature tensor and those variations, Eq. (7.22), which leave $\delta_{(N)}R_{am_1m_2b,m_3\cdots m_N} = 0$. The highest order principal equation (7.15) restricted to the variations (7.22) implies that the highest order multiplier (7.34) is algebraically determined by coefficients with the symmetries of a differentiated curvature tensor. The highest order principal equation written down for the remaining variations then implies that the generator is linear in the highest order derivative of the curvature tensor, Eq. (7.38).

(IX) The lower order principal equation (7.16) ensures that the coefficient of this linear function has the structure of an R-gradient. Therefore, the generator K is weakly equivalent to a generator of a lower order (namely, the order $M = N - 1$) in the derivatives $R_{am_1m_2b,m_3\cdots m_N}$: Eq. (7.39). A recursive argument then shows that the generator must be weakly ultralocal.

(X) There is just one ultralocal generator which is a spatial invariant and satisfies the principal equation (5.15). Its coefficient (7.46) is a conformal Killing (super)vector of the local supermetric. This generator scales the kinetic term by the factor $-\frac{3}{2}\kappa$, while it scales the potential term by the factor $-\frac{1}{2}\kappa$, Eq. (8.4). The potential term thus breaks the conditional symmetry of the kinetic term. As a result, there is no conditional symmetry in geometrodynamics.

ACKNOWLEDGMENT

I want to thank Julian Barbour who, patient and slightly amused, kept pointing out my sins against English.

[1] We shall outline our notation. Spacetime points are capitalized: X^α, $\alpha = 0,1,2,3$, while points on a spacelike hypersurface are labeled by x^a, $a = 1,2,3$. Indices are suppressed in the arguments of functions. The signature of $g_{\alpha\beta}$ is $(-,+,+,+)$, with $\eta_{\alpha\beta} = \mathrm{diag}(-1,1,1,1)$. The determinant of the spatial metric g_{ab} is denoted by g. Spatial vectors are boldface, $\mathbf{N}$, if not written in terms of their components, N^a. We use ∂_a for partial derivatives within the hypersurface, the vertical stroke $|a$ for covariant derivatives with respect to the spatial metric g_{ab}, $L_{\mathbf{k}}$ for the Lie derivative along the spatial vector $\mathbf{k}$, and the dot $\cdot$ for the derivative with respect to a label time t. Expressions are understood to be symmetric in an uninterrupted sequence of indices following a comma. Thus, $\delta h_{a,bm_1\cdots m_N}$ is symmetric in the indices $bm_1\cdots m_M$. In particular, a comma denotes a completely symmetrized covariant derivative, as in $\delta g_{ab,m_1\cdots m_M} \equiv \underset{m}{\mathrm{S}}\, \delta g_{ab|m_1\cdots m_M} \equiv (M!)^{-1} \sum_{P(m)} \delta g_{ab|m_1\cdots m_M}$. Here, $\underset{m}{\mathrm{S}}$ denotes complete symmetrization in $m \equiv (m_1, m_2, \ldots, m_M)$ achieved by the permutations $P(m)$. Braces stand for cyclic permutation in the sequence of indices enclosed between them; thus, $k_{|ab}l_{c|} = k_{ab}l_c + k_{bc}l_a + k_{ca}l_b$. Parentheses mean anticommutation in a pair of indices: $k_{(a\cdots b)} \equiv k_{a\cdots b} + k_{b\cdots a}$. Square brackets fulfill a multitude of needs. When enclosing a pair of indices or a pair of smearing functions, they denote commutation. So, $k_{|a\cdots b|} \equiv k_{a\cdots b} - k_{b\cdots a}$ and $\lambda_{|\mathbf{M}}{}^a\mu_{\mathbf{N}|a} \equiv \lambda_{\mathbf{M}}{}^a\mu_{\mathbf{N}a} - \lambda_{\mathbf{N}}{}^a\mu_{\mathbf{M}a}$. When enclosing a pair of spatial vectors, they denote Lie brackets: $[\mathbf{M}, \mathbf{N}] \equiv L_{\mathbf{M}}\mathbf{N}$. When applied to two dynamical variables, as in $[K, H_N]$, they denote Poisson brackets. Finally, when enclosing arguments, they emphasize the functional dependence, while the parentheses imply a function dependence. So, $K\,[g_{ab}, p^{ab}]$ means that the dynamical variable K is a functional constructed from the canonical variables $g_{ab}(x)$ and $p^{ab}(x)$, while $g_{ab}(x)$ means that g_{ab} is a function of x^c. The mean-

ing of the square brackets is thus determined by the nature of the variables which enter into them and stand before them. The symbol $(x \leftrightarrow x')$ abbreviates the expression "the same term with the labels x and x' interchanged." The curvature tensor is fixed by the convention $k_{a[[bc]]} = k_d R^{d}{}_{abc}$. Our delta functions $\delta(x, x')$ are always scalars in the first and scalar densities in the second argument. The units are chosen such that $c = 1 = 16\pi G$.

[2]If $g^{\alpha\beta} = \eta^{\alpha\beta}$ and $\phi(X)$ is identified with the gravitational potential, Eqs. (1.11)–(1.14) describe the free fall of a particle in the Nordström theory of gravitation. See G. Nordström, Phys. Z. **13**, 1126 (1912); Ann. Phys. (Leipzig) **40**, 856 (1913); **42**, 533 (1913); **43**, 1101 (1914).

[3]K. Kuchař, J. Math. Phys. **13**, 768 (1972); K. Kuchař, in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, The Netherlands, 1973).

[4]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva U. P., New York, 1964).

[5]$\tilde{g}_{\alpha\beta}$ was identified with the physical metric and $\tilde{\tau}$ with the physical proper time in the Fokker–Einstein reinterpretation of the Nordström theory of gravitation. See A. Einstein and A. D. Fokker, Ann. Phys. (Leipzig) **44**, 321 (1914).

[6]J. A. Wheeler, in *Relativity, Groups and Topology, 1963 Les Houches Lectures*, edited by B. S. DeWitt and C. M. DeWitt (Gordon and Breach, New York, London, 1964); B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).

[7]C. Teitelboim, preprint, The Institute for Advanced Study, Princeton, March 1980.

[8]K. Kuchař, in *Quantum Gravity II: A Second Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. Sciama, to be published.

[9]The transition from a finite-dimensional particle model to infinitely dimensional geometrodynamics is unquestionably a vast generalization. In a forthcoming paper, we shall show that our definition of conditional symmetry in geometrodynamics has a natural counterpart in parametrized field theories, where it leads to an intuitively clear connection with the symmetry of the spacetime background.

[10]A. E. Fischer, in *Relativity*, edited by M. S. Carmeli, S. I. Fickler, and L. Witten (Plenum, New York, London 1970).

[11]See, e.g., K. Kuchař, J. Math. Phys. **18**, 1589 (1977), Sec. 4.

[12]Cf. the variation of the curvature tensor, Eqs. (7.6)–(7.8). See also B. S. DeWitt, Ref. 6, Eq. (4.25).

[13]C. J. Isham, Proc. R. Soc. London Ser. A **351**, 209 (1976).

[14]J. R. Klauder, Comm. Math. Phys. **18**, 307 (1970); Acta Phys. Austr. Suppl. **VIII**, 227 (1971); J. Math. Phys. **11**, 609 (1970); in *Lectures in Theoretical Physics*, edited by W. E. Brittin (Colorado Assoc. U. P., Boulder, 1973).

[15]M. Henneaux, Bull. Soc. Math. Belgique **31**, 77 (1979).

[16]M. Pilati, Ph.D. thesis, Princeton University, 1980.

[17]See., e.g., I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, London, 1968), Vol. II, §4.5.

[18]See C. W. Misner in *Magic without Magic* (J. A. Wheeler 60th Birthday Volume), edited by J. R. Klauder (Freeman, San Francisco, 1972) and K. Kuchař, Ref. 8.

[19]This may seem queer in view of DeWitt's analysis of the symmetry of the local supermetric. DeWitt has shown[6] that the local supermetric (4.5) has a "nested structure" related to a symmetric space of $SL(3, R)/SO(3)$. As noticed by Martin Pilati, this implies that the local supermetric (4.5) has other conformal (super)vectors in addition to the expression (7.46). This descrepancy is resolved by observing that these other (super)vectors do not transform as (symmetric) tensors under spatial transformations. DeWitt's ultralocal analysis does not pay attention to the interplay of the super-Hamiltonian and supermomentum constraints, which is vital in our concept of conditional symmetry.

[20]In a free particle model, conserved dynamical variables which are linear but inhomogeneous in the momentum generate the transition to the initial canonical data. A detailed paper on conditional symmetries in parametrized particle mechanics is in preparation.

All nontwisting N's with cosmological constant

A. García Díaz and J. F. Plebański[a)]
Centro de Investigación y de Estudios Avanzados del I.P.N., Departamento de Física, Apdo. Postal 14-740, México 14, D.F., Mexico

(Received 10 December 1980; accepted for publication 1 May 1981)

The set of *all* essentially different nontwisting type-N solutions with cosmological constant is presented in such a tetrad gauge, and coordinatization that the metric depends *linearly* on an arbitrary structural function. The special branches of the solutions of this type are shown to amount to the contractions of the most general branch.

PACS numbers: 04.20.Jb

I. INTRODUCTION

The aim of this work is to present *all* nontwisting N-type solutions with cosmological constant in such a representation that the metric depends linearly on the arbitrary structural complex function and to establish the interrelations among the different branches.

We employ the null tetrad formalism.[1,2] Working with the signature + 2, the metric is given by

$$g = 2e^1 \otimes e^2 + 2e^3 \otimes e^4, \qquad e^2 = \bar{e}^1, \quad e^3 = \bar{e}^3, \quad e^4 = \bar{e}^4. \tag{1.1}$$

The first Cartan structure equations

$$de^a = e^b \wedge \Gamma^a{}_b \tag{1.2}$$

define the connection 1-forms $\Gamma_{ab} = \Gamma_{[ab]}$. Whenever the direction of e^3 is oriented along the quadruple Debever–Penrose vector, the second structure equations, into which the Einstein equations ($G_{\mu\nu} = \lambda g_{\mu\nu}$) are built in, can be written, for the N-type fields, as

$$\begin{aligned} &d\Gamma_{42} + \Gamma_{42} \wedge (\Gamma_{12} + \Gamma_{34}) = \tfrac{1}{3}\lambda e^3 \wedge e^1, \\ &d(\Gamma_{12} + \Gamma_{34}) + 2\Gamma_{42} \wedge \Gamma_{31} = -\tfrac{1}{3}\lambda (e^1 \wedge e^2 + e^3 \wedge e^4), \\ &d\Gamma_{31} + (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{31} = \tfrac{1}{3}\lambda e^4 \wedge e^2 + \tfrac{1}{2}C^{(1)} e^3 \wedge e^1, \end{aligned} \tag{1.3}$$

where $C^{(1)}$ is the only nonvanishing conformal curvature coefficient.

In general, when e^3 is a distinguished direction, there arises a natural subclassification of the Riemannian structures invariant under σ and γ gauges,

$$\begin{aligned} &\sigma = \omega + i\phi\text{:} \quad e'^1 = e^{i\phi} e^1, \quad e'^2 = e^{-i\phi} e^2, \\ &e'^3 = e^{\omega} e^3, \quad e'^4 = e^{-\omega} e^4 \end{aligned} \tag{1.4}$$

$$\begin{aligned} &\gamma\text{:} \quad e'^1 = e^1 + \bar{\gamma} e^3, \quad e'^2 = e^2 + \gamma e^3, \\ &e'^3 = e^3, \quad e'^4 = e^4 - \gamma e^1 - \bar{\gamma} e^2 - \gamma\bar{\gamma} e^3, \end{aligned}$$

which maintain the direction e^3 unchanged. These gauges induce on the connections the transformations

$$\begin{aligned} &\Gamma'_{42} = e^{\sigma}\Gamma_{42}, \quad \Gamma'_{12} + \Gamma'_{34} = \Gamma_{12} + \Gamma_{34} + d\sigma, \\ &\Gamma'_{31} = e^{-\sigma}\Gamma_{31}, \end{aligned} \tag{1.5}$$

and

$$\Gamma'_{42} = \Gamma_{42}, \quad \Gamma'_{12} + \Gamma'_{34} = \Gamma_{12} + \Gamma_{34} + 2\gamma\Gamma_{42}, \quad \Gamma'_{31} = \Gamma_{31} + \gamma(\Gamma_{12} + \Gamma_{34}) + \gamma^2\Gamma_{42} + d\gamma, \tag{1.6}$$

respectively.

This subclassification is:

First, there are structures with e^3 twisting (T) and nontwisting (NT). The twist of e^3 is defined as the 3-form

$$\text{T:} = e^3 \wedge de^3; \tag{1.7}$$

hence

$$\begin{aligned} &\text{(T):} \quad e^3 \wedge de^3 \neq 0, \\ &\text{(NT):} \quad e^3 \wedge de^3 = 0, \end{aligned} \tag{1.8}$$

these conditions being manifestly invariant under (1.4)–(1.6).

Next, we consider the gauge invariant relative properties of Γ_{42} and e^3 under (1.4)–(1.6). One finds the following list of disjoint possibilities:

$$\begin{aligned} &\text{I:} \quad \Gamma_{42} = 0 \\ &\text{II:} \quad \Gamma_{42} \neq 0, \quad \Gamma_{42} \wedge \bar{\Gamma}_{42} = 0 \begin{cases} \text{II}_S\text{:} \ e^3 \wedge \Gamma_{42} = 0, \\ \text{II}_G\text{:} \ e^3 \wedge \Gamma_{42} \neq 0, \end{cases} \\ &\text{III:} \quad \Gamma_{42} \wedge \bar{\Gamma}_{42} \neq 0 \begin{cases} \text{III}_S\text{:} \ e^3 \wedge \Gamma_{42} \wedge \bar{\Gamma}_{42} = 0, \\ \text{III}_G\text{:} \ e^3 \wedge \Gamma_{42} \wedge \bar{\Gamma}_{42} \neq 0. \end{cases} \end{aligned} \tag{1.9}$$

(the subscripts S and G distinguish the special and general subcases). In the following sections we shall study type-N solutions for each of the three branches of (1.9).

II. COMPLETE SOLUTION FOR CASE I

This branch ($\Gamma_{42} = 0$) does not admit a solution with nonvanishing cosmological constant, and futhermore the twist here is zero. Consequently, the only solutions of this case are the well-known Robinson waves, R.[3–5]

The tangent null tetrad, $\partial_a = e^{\mu}{}_a \partial_{\mu}$, the curvature and the Γ_{42}-form which describe the R waves are

$$\begin{aligned} &\partial_1 = \partial_{\xi}, \quad \partial_2 = \partial_{\bar{\xi}}, \quad \partial_3 = \partial_t - (f + \bar{f})\partial_r, \quad \partial_4 = \partial_r, \\ &C^{(1)} = -2f_{\xi\xi}, \quad \Gamma_{42} = 0, \end{aligned} \tag{2.1}$$

where $f = f(\xi, t)$ is an entirely arbitrary analytic function in ξ.

The contravariant metrical density is given by

$$\begin{aligned} g[\phi,\phi] &:= eg^{ab}\phi_{,a}\phi_{,b} \\ &= -2\phi_{\xi}\phi_{\bar{\xi}} - 2\phi_{,r}[\phi_{,t} - (f + \bar{f})\phi_{,r}], \end{aligned} \tag{2.2}$$

where $e = \det(e^a{}_{\mu})$ and ϕ is an arbitrary function.

Because these R waves are prototypic for the other branches of the N problem studied, we would like to add some comments concerning their properties. Notice that the

[a)] On leave of absence from University of Warsaw, Warsaw, Poland.

 0022-2488/81/112655-04$01.00

tangent tetrad and the contravariant density are linear in the free structural function $f(\xi,t)$. The dependence of $f(\xi,t)$ on the variable t is completely free. This variable defines a set of characteristic surfaces (t = const) of the metric satisfying the eikonal equation $g[t,t] = 0$. With the dependence of f on t entirely free, the amplitude of the curvature $C^{(1)}$ is determined by an analytic function chosen arbitrarily along each of the characteristic surfaces (t = const). These properties are common to all the nontwisting N-type solutions, as we shall see later.

III. ALL SOLUTIONS OF CASE II

One can show that nontrivial solutions may exist only in the subbranch II_S. The direction e^3 is then geodesic, shearless, and free of complex expansion.

These solutions new to our knowledge, which we denote by $K(\lambda)$ waves, are the generalization of the Kundt solution[6] K and are given via:

(i) the tangent null tetrad:

$$\partial_1 = \cosh x\partial_\xi, \quad \partial_2 = \cosh x\partial_{\bar\xi}, \quad \partial_4 = \cosh x\partial_r,$$

$$\frac{1}{\sqrt{\mu}}\tanh x\partial_3 = \partial_t + r\cosh x(\partial_\xi + \partial_{\bar\xi} - \cosh x[(\partial_\xi + \partial_{\bar\xi} - 2(\sqrt{\mu})\tanh x)(f+\bar f) - (\sqrt{\mu})r^2\tanh x]\partial_r \tag{3.1}$$

where $\mu := \lambda/6$, $x := (\sqrt{\mu})(\xi + \bar\xi)$;

(ii) the conformal curvature:

$$C^{(1)} = -2(\sqrt{\lambda}/6)(\cosh^3 x/\sinh x)(\partial_\xi\partial_\xi - \tfrac{2}{3}\lambda)f_\xi; \tag{3.2}$$

and

(iii) the connection form Γ_{42}:

$$\Gamma_{42} = (\sqrt{\mu}/\sinh x)e^3. \tag{3.3}$$

The metrical density is then

$$-g[\phi,\phi] = (2/\sqrt{\mu})(\sinh x/\cosh^2 x) \times\{\phi_\xi\phi_{\bar\xi} + (\sqrt{\mu}/\sinh x)\phi_r[\phi_t + r\cosh x(\phi_\xi + \phi_{\bar\xi}) - \cosh x((\partial_\xi + \partial_{\bar\xi} - 2(\sqrt{\mu})\tanh x)(f+\bar f) - (\sqrt{\mu})r^2\tanh x)\phi_r]\}. \tag{3.4}$$

Here λ is assumed to be positive; for $\lambda < 0$, the hyperbolic functions are to be replaced by the corresponding trigonometric ones.

As in the case of the Robinson waves, the tetrad and the metrical density are linear in the free structural function $f(\xi,t)$. The variable t determines the characteristics of the metric, and the analytic function which determines the amplitude of the curvature can be arbitrarily chosen along each characteristic surface.

IV. ALL NONTWISTING SOLUTIONS OF CASE III

Within this class there exist solutions of the type N only in the subbranch III_G. We restrict ourselves here to study the diverging but free of rotation N-solutions. These were investigated by Leroy,[7] who obtained basic results following the theory of Robinson and Trautman.[8] However, in his treatment, the solution was given modulo some equations. Also, his choice for the tetrad did not exhibit the linear dependence on an arbitrary structural function. We have succeeded, we believe, in deriving a much more satisfactory description of the solutions of this type, which depend linearly on an arbitrary analytic function.

Exploiting the freedom of gauges, we succeeded to integrate equations (1.2) and (1.3) determining an "optimal" chart of coordinates, $\{\xi,\bar\xi,r,t\}$, and the corresponding tetrad gauge in which all nontwisting N waves with λ are given by

$$e^1 = rd\xi + (\psi_{\bar\xi} - rf)\,dt, \quad e^2 = (\bar e^1),$$

$$e^3 = \psi dt,$$

$$e^4 = dr + \left| -\psi_{\xi\bar\xi} + \tfrac{1}{2}r(f_\xi + \bar f_{\bar\xi}) + \tfrac{1}{6}\lambda r^2\psi\right| dt, \tag{4.1}$$

where $f = f(\xi,t)$ is an arbitrary complex function depending on ξ and t only.

The tangent tetrad associated with (4.1) is

$$r\partial_1 = \partial_\xi, \quad r\partial_2 = \partial_{\bar\xi}, \quad \partial_4 = \partial_r,$$

$$\psi\partial_3 = \partial_t - (1/r)[(\psi_{\bar\xi} - rf)\partial_\xi + (\psi_\xi - r\bar f)\partial_{\bar\xi}] - \left| -\psi_{\xi\bar\xi} + \tfrac{1}{2}r(f_\xi + \bar f_{\bar\xi}) + \tfrac{1}{6}\lambda\,\psi r^2\right|\partial_r. \tag{4.2}$$

By using appropriately the remaining coordinate freedom, the structural function ψ, without any loss of generality, can be always brought to the form

$$\psi = 1 + \epsilon\xi\bar\xi, \tag{4.3}$$

where the discrete parameter ϵ takes the values 1, 0, -1. Optionally, for $\epsilon = -1$, we represent ψ as

$$\psi = (1/\sqrt{a})\sinh x, \quad x := (\sqrt{a})(\xi + \bar\xi), \quad a = \text{const} \tag{4.4}$$

[here a is assumed to be positive; for a negative $\psi \to (1/\sqrt{a})\sin x$]. The connections accompanying our tetrad are

$$\Gamma_{42} = -(1/r)e^1 + (1/r\psi)\psi_{\bar\xi}e^3,$$

$$\Gamma_{12} + \Gamma_{34} = (\tfrac{1}{3}\lambda r + \psi^{-1}f_\xi)e^3,$$

$$\Gamma_{31} = (1/r\psi)\psi_{\xi\xi}e^1 + \tfrac{1}{6}\lambda\,re^2 + (1/r\psi)(-\psi_{\xi\xi\bar\xi} + \tfrac{1}{2}rf_{\xi\xi} + \tfrac{1}{6}\lambda r^2\psi_\xi)e^3. \tag{4.5}$$

In particular, for ψ given by (4.3) the connections above take the simple form

$$\Gamma_{42} = -(1/r)e^1 + (\epsilon/r\psi)\xi e^3,$$

$$\Gamma_{12} + \Gamma_{34} = (\tfrac{1}{3}\lambda\, r + \psi^{-1}f_\xi)e^3,$$

$$\Gamma_{31} = \tfrac{1}{6}\lambda\, re^2 + (1/2\psi)(f_{\xi\xi} + \tfrac{1}{3}\lambda\, r\epsilon\bar\xi)e^3. \tag{4.6}$$

The conformal curvatures, corresponding to the ψ's given by (4.3) and (4.4), are

$$C^{(1)} = -f_{\xi\xi\xi}/r(1 + \epsilon\xi\bar\xi) \neq 0 \tag{4.7}$$

and

$$C^{(1)} = -(1/r)[(\sqrt{a})/\sinh x](\partial_\xi\partial_\xi - 4a)f_\xi \neq 0, \tag{4.8}$$

respectively.

Note that the tetrad, from (4.1), is linear in the structural function $f(\xi,t)$; note also that the variable t defines the set of

characteristic surfaces of the metric, $t = \text{const}$.

In what follows we shall denote this class of solutions by NT(λ,Z,ϵ), where λ stands for the cosmological constant, Z represents the complex expansion $-\Gamma_{421}$, and the parameter ϵ takes the values 1, 0, -1, depending upon whether the "source" lines of the gravitational waves are respectively timelike, null or spacelike. The complex expansion Z in the studied problem is real and equal to $1/r$; therefore, it represents the divergence of the congruence e^3.

All vacuum nontwisting solutions of the type N, denoted as NT(0,Z,ϵ), are obtained from the NT(λ,Z,ϵ) by simply equating λ to zero in the expressions above. These solutions were obtained in Ref. 8; see also Refs. 9 and 10. Nevertheless, the tetrad gauge used there does not exhibit a linear dependence on an arbitrary structural function. The simple form of these solutions, given in this text, having a linear dependence on the structural function, facilitates, among other things, the limiting transitions to the subcases of Kundt and Robinson which depend linearly on the structural function.

V. CONTRACTIONS

The purpose of this section is to show that the subbranches $K(\lambda)$, K, R can be derived from the NT(λ,Z,ϵ) via corresponding limiting transitions.

From expressions (3.1)–(3.2), by letting λ go to zero and at the same time changing $f_{\xi} \to f$, we readily obtain the Kundt waves K

$$\partial_1 = \partial_{\xi}, \quad \partial_2 = \partial_{\bar{\xi}},$$
$$(\xi + \bar{\xi})\partial_3 = \partial_t + r(\partial_{\xi} + \partial_{\bar{\xi}}) - (f + \bar{f})\partial_r, \partial_4 = \partial_r,$$
$$C^{(1)} = -2f_{\xi\xi}/(\xi + \bar{\xi}). \tag{5.1}$$

Executing then the coordinate transformation

$$\xi = \xi' + \tfrac{1}{2}\varepsilon^{-1}, \quad \bar{\xi} = \bar{\xi}' + \tfrac{1}{2}\varepsilon^{-1}, \quad r = r', \quad t = \varepsilon t',$$

replacing the structural function f by $\varepsilon^{-1}f'(\xi',t')$ and letting ε go to zero we obtain easily the Robinson solution (2.1).

In order to show that the $K(\lambda)$ waves are limiting contractions of the NT(λ,Z,ϵ) solutions, we start from (4.2) with ψ taken in the form (4.3). Executing the σ and γ gauges with

$$\sigma = \ln \cosh x, \quad \gamma = -\nu(\sqrt{a})\, r\cdot\sinh x\cdot\cosh x,$$

where ν is an arbitrary constant, dropping primes, we arrive at

$$r\partial_1 = \partial_{\xi} - \nu(\sqrt{a})\, r^2 \sinh x \partial_r, \quad \partial_2 = (\bar{\partial}_1), \quad \cosh x \partial_4 = \partial_r,$$
$$\frac{1}{\sqrt{a}}\tanh x \partial_3 = \partial_t - \frac{1}{r}(\cosh x - rf - \nu r \sinh^2 x)\partial_{\xi} - \frac{1}{r}(\cosh x - r\bar{f} - \nu r \sinh^2 x)\partial_{\bar{\xi}}$$
$$- \left[-(\sqrt{a})\sinh x + \frac{r}{2}(f_{\xi} + \bar{f}_{\bar{\xi}}) + \frac{\lambda}{6}\frac{r^2}{\sqrt{a}}\sinh x + \nu^2(\sqrt{a})\, r^2 \sinh^3 x \right]\partial_r, \tag{5.2}$$
$$C^{(1)} = -\frac{\cosh^2 x}{r}\frac{\sqrt{a}}{\sinh x}(\partial_{\xi}\partial_{\xi} - 4a)f_{\xi}, \quad x := (\sqrt{a})(\xi + \bar{\xi}).$$

We change now the coordinates and other quantities which appear above according to

$$t = \varepsilon^{-1}t', \quad \xi = \varepsilon\xi', \quad \bar{\xi} = \varepsilon\bar{\xi}',$$
$$r = \frac{1}{\cosh x}\left(\frac{r'}{\cosh x} + \varepsilon^{-1}\right), \tag{5.3}$$
$$\nu = \varepsilon, \quad a = \tfrac{1}{6}\lambda\varepsilon^{-2}, \quad f = \varepsilon(2\varepsilon^2 f' + 1),$$

where the function $f'(\xi',t')$ is assumed to be independent of the parameter ε. Making ε tend to zero, we obtain precisely the $K(\lambda)$ solutions given by (3.1)–(3.2).

We now consider the pertinent contractions of the NT(0,Z,ϵ) solutions. The Kundt metrics happen to be the contraction of the NT(0,Z, -1) waves. Indeed, setting $\epsilon = -1$ in (4.3) and $\lambda = 0$ in (4.2) and transforming the coordinates and the structural function according to

$$t = \varepsilon^{-1}t', \quad \xi = \varepsilon\xi' - 1, \quad \bar{\xi} = \varepsilon\bar{\xi}' - 1, \quad r = r' + \varepsilon^{-1},$$
$$f = 2\varepsilon^3\int^{\xi'} f'(\xi',t')\, d\xi' - \varepsilon^2\xi' + \varepsilon, \tag{5.4}$$

by letting ε go to zero, we arrive at a slightly modified version of the Kundt waves

$$\partial_1 = \partial_{\xi}, \quad \partial_2 = \partial_{\bar{\xi}}, \quad \partial_4 = \partial_r,$$
$$(\xi + \bar{\xi})\partial_3 = \partial_t + r(\partial_{\xi} + \partial_{\bar{\xi}}) + (r - f - \bar{f})\partial_r,$$
$$C^{(1)} = -2f_{\xi\xi}/(\xi + \bar{\xi}). \tag{5.5}$$

To recover the expressions (5.1), it is enough to execute in (5.5) a σ-gauge, with $\sigma = t$, and then to change coordinates and redefine the structural function according to

$$t' = e^t, \quad r' = e^{-t}r, \quad \xi' = \xi, \quad \bar{\xi}' = \bar{\xi}, \quad f' = e^{-2t}f.$$

Finally, to establish that NT(0,Z,ϵ) $\to R$, we start from (4.2) with $\lambda = 0$ and ψ taken from (4.3). Performing there the coordinate transformation

$$t = t', \quad \xi = \varepsilon\xi', \quad \bar{\xi} = \varepsilon\bar{\xi}', \quad r = r' + \varepsilon^{-1}, \tag{5.6}$$

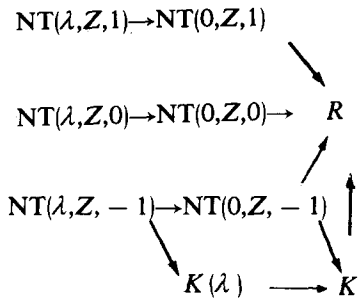
accompanied by

$$f = \varepsilon^2\left[\epsilon\xi' + 2\int^{\xi'} f'(\xi',t')d\xi'\right] \tag{5.7}$$

and taking ε to zero we arrive at (3.1), i.e., the Robinson solution.

VI. CONCLUSION

We have found that *all* geometrically different nontwisting N-type solutions with λ are interrelated, by contractions, according to the diagram given below.



Therefore, the set of NT (λ,Z,ϵ) solutions completely exhausts all N-type nontwisting solutions.

[1]J. F. Plebański; *Spinors, Tetrads and Forms*, a monograph of Centro de Investigación y de Estudios Avanzados del IPN, 297 pages (1974).
[2]G. Debney, R. P. Kerr, and A. Schild; J. Math. Phys. **10**, 1942 (1969).
[3]I. Robinson, Lecture at King's College, London (1956). Report to the Rayaumont Conference (1959).
[4]H. W. Brinkmann; Proc. Nat. Acad. Sci. USA **9**, 1 (1923).
[5]J. Helý; C. R. Acad. Sci. Paris **249**, 1867 (1959).
[6]W. Kundt; Z. Physik **163**, 77 (1961).
[7]J. Leroy; Bull. Acad. Roy. Belg. Cl. Sci. **62**, 259 (1976).
[8]I. Robinson and A. Trautman; Phys. Rev. Lett. **4**, 431 (1960).
[9]J. Foster and E. T. Newman; J. Math. Phys. **8**, 189 (1967).
[10]D. Kramer, H. Stephani, M. Mac Callum, and E. Herlt; *Exact Solutions of Einstein's Field Equations* (VEB Deutsch. Verlag Wiss. Berlin, 1980), p. 261.

On the existence of simultaneous synchronous coordinates in spacetimes with spacelike singularities

Robert M. Wald and Ping Yip
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637

(Received 13 February 1981; accepted for publication 18 May 1981)

We investigate the issue of when a synchronous coordinate system exists which entirely covers a singularity simultaneously. For a spacetime with a spacelike singularity (or, more precisely, with a spacelike causal boundary) we prove that such coordinates exist if and only if the "maximum lifetime function" satisfies certain boundedness, differentiability, and limit properties. Our results lend plausibility to the validity of the use of simultaneous synchronous coordinates to analyze physically realistic spacelike singularities. Nevertheless, we construct a number of examples which do not satisfy our conditions and thus do not admit such coordinates.

PACS numbers: 04.20.Jb

I. INTRODUCTION

The singularity theorems of general relativity have established the inevitability of spacetime singularities in cosmology and gravitational collapse under a wide range of circumstances. However, while these theorems prove the existence of singularities under very general assumptions, they tell us very little about the nature of these singularities. Thus, an important goal of general relativity is to obtain a precise description of the spacetime structure near singularities.

The most straightforward approach toward this goal is to solve Einstein's equation in a neighborhood of a singularity. This formidable task has been undertaken in a series of papers by Belinskii, Lifschitz, and Khalatnikov (BLK). (See Ref. 1 for a review.) These authors used a local approach and employed a number of approximations, which have been criticized recently by Barrow and Tipler.[2] However, a crucial first step of BLK which has not been analyzed satisfactorily is the assumption that a synchronous coordinate system can be found which reaches the singularity simultaneously and covers the entire singularity. (These notions will be defined precisely in Sec II.) We will refer to such a coordinate system as *simultaneous synchronous coordinates* (SSC).

The use of SSC greatly simplifies the analysis of solutions of Einstein's equation near the singularity, but by assuming the existence of these coordinates some important assumptions are being made about the nature of the singularity. For example, if one creates an artificial singularity in Minkowski spacetime by removing the axis, $x = y = z = 0$ for $t \geqslant 0$, it is not difficult to see that no synchronous coordinate system can cover the entire singularity simultaneously. About the best one can do is reach the "bottom point" of the singularity simultaneously by synchronous coordinates, as illustrated in Fig. 1. More generally, it is easy to construct a wide variety of examples with timelike and null singularities such that SSC cannot be found (Precise definition of "timelike", "null", and "spacelike" causal boundaries of spacetimes will be given in Sec. II.)

Thus, it seems clear that the use of synchronous coordinates cannot be justified for investigations of timelike or null singularities. Of course, we do not yet know the true structure of singularities which occur in nature; that is what the analyses of BLK and others seek to determine. However, the possibility that physically realistic singularities are spacelike appears to be well worth pursuing. After all, the singularities of the Robertson–Walker and Schwarzschild solutions are spacelike. Furthermore, Penrose[3] recently has given arguments in favor of the idea that singularities in a physically reasonable spacetime cannot be timelike. He has shown that this condition on the singularities—which he calls "strong cosmic censorship"—is equivalent to the condition that the spacetime be globally hyperbolic. Thus, the condition that singularities be spacelike may be viewed as a further strengthening of this hypothesis of Penrose. In any case, an investigation of the general nature of spacelike singularities would be of considerable interest in its own right for what it would tell us about general relativity.

However, even in the case of a spacelike singularity, it is far from obvious that SSC exist. Indeed, Misner[4] has emphasized that no demonstration has been given for the existence of SSC, and has suggested that irregularities in the curvature should almost always cause the geodesics of the SSC to cross (resulting in the breakdown of the SSC) before the singularity is reached. Furthermore, even if the SSC do not break down, it is not at all clear that the entire singularity can be covered (i.e., the phenomenon of Fig. 1 may be unavoidable even for spacelike singularities), in which case an analysis of solutions in SSC would yield misleading results on the nature of the singularity. Thus, the validity of the first step of the analysis of BLK is questionable even for spacelike singularities.

The aim of this paper is to investigate the existence of SSC in arbitrary spacetimes whose future causal boundary is spacelike. We shall prove that existence of SSC is equivalent to certain differentiability, boundedness, and limit properties holding for the "maximum lifetime function" of the spacetime. Thus, we obtain precise criteria for the existence of SSC.

In Sec. II we will prove that in any spacetime whose future causal boundary is spacelike, the geodesics normal to any Cauchy surface reach every "singular point." The neces-

 0022-2488/81/112659-07$01.00

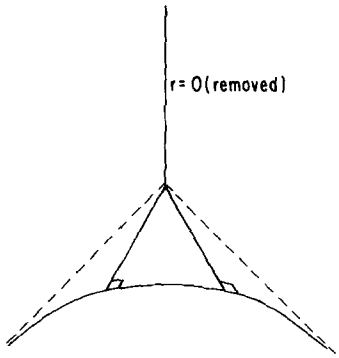


FIG. 1. Minkowski spacetime with the upper part of the axis $r = 0$ removed.

sary and sufficient conditions for the existence of SSC in terms of the maximum lifetime function are derived in Sec. III. Finally, in Sec. IV we present examples illustrating the failure of SSC in cases where the necessary conditions on the maximum lifetime function are not satisfied. Our definitions and notation throughout the paper follow the conventions of Hawking and Ellis[5] except that we define a Cauchy surface to be a closed, achronal set whose domain of dependence is the entire spacetime, and hypersurfaces will not be assumed to be C^2 unless explicitly stated. All spacetimes (M, g_{ab}) considered in the paper will be assumed to be C^∞.

II. SPACELIKE CAUSAL BOUNDARIES AND SIMULTANEOUS SYNCHRONOUS COORDINATES: DEFINITIONS AND IMPLICATIONS

Our first task is to adopt a precise notion of singularities which allows us to discuss their causal structure. We shall use the approach of Geroch, Kronheimer, and Penrose,[6] who define an indecomposable past set W as a subset of the spacetime (M, g_{ab}) which is an open, past set (i.e., $I^-(W) = W$ where $I^-(W)$ denotes the chronological past of W) and cannot be expressed as the union of two proper subsets of this form. The past $I^-(p)$ of a point $p \in M$ is an indecomposable past set, and is called a proper indecomposable past set (PIP). The indecomposable past sets which are not proper are called terminal indecomposable past sets (TIP's). The TIP's comprise the *future causal boundary* of a spacetime. This causal boundary need not represent a singularity, as, for example, the TIP's of Minkowski spacetime generate future null infinity and future timelike infinity. However, in the case of an all encompassing singularity as envisioned, for example, in cosmology, each TIP may be viewed as a "point" of the future singularity. The past causal boundary of a spacetime is defined analogously.

It is not difficult to show that the past $I^-(\gamma)$ of any future inextendible causal curve γ is a TIP. Furthermore every TIP can be shown to be the past of some future inextendible causal (in fact, timelike) curve.[6] Thus, TIP's could be defined alternatively as the equivalence classes of future inextendible timelike curves, where timelike curves γ and γ' are considered equivalent if $I^-(\gamma) = I^-(\gamma')$.

The causal relation between points on the causal boundary now may be defined. A TIP V is said to lie to the chronological past of a TIP W (i.e., V and W are *timelike related*) if there exists a point $p \in W$ such that $V \subset I^-(p)$. V and W are said to be *null related* if they are not timelike related, but $V \subset W$ (or $W \subset V$). Finally, V and W are said to be *spacelike related* if they are neither timelike nor null related. The future causal boundary is said to be *spacelike* if all pairs of TIP's are spacelike related. In other words, the future causal boundary is spacelike if and only if no TIP contains another TIP as a proper subset. Strictly speaking, the terminology of "acausal" would be more accurate than "spacelike," but the confusion created by statements such as "the causal boundary is acausal" lead us to prefer the terminology "timelike," "null," and "spacelike" to describe the causal relationship of points on the causal boundary. The relationship between points of the past causal boundary are defined similarly.

As discussed in the introduction, our interest is in spacetimes whose future (or past) causal boundary is spacelike. Penrose[3] has shown that the future (or past) causal boundary will be achronal (i.e., spacelike or null) if and only if the spacetime (M, g_{ab}) is globally hyperbolic. Thus, our restriction to spacetimes with spacelike singularities implies that we shall consider only globally hyperbolic spacetimes.

Let us now define with more precision the requirements we impose for "simultaneous synchronous coordinates" (SSC). A *synchronous* (or *Gaussian normal*) coordinate system is one constructed by choosing a C^1 spacelike hypersurface Σ and using (arbitrarily chosen) coordinates on Σ and the proper time τ along geodesics normal to Σ to label events. Since we are dealing with globally hyperbolic spacetimes, we will require that Σ be a Cauchy surface; otherwise, the study of evolution via Einstein's equation from initial data on Σ will not even determine the solution on all of (M, g_{ab}) and thus cannot be expected to determine the complete structure of the singularity. If Σ is C^2, the construction of synchronous coordinates always works in an open neighborhood of Σ, and it is not difficult to show that all the hypersurfaces of constant τ remain orthogonal to the geodesics used in the construction. However, these geodesics may eventually cross, in which case the labeling of events becomes nonunique, and the coordinate system "breaks down." Therefore, in order to use synchronous coordinates to analyze the structure of the future (or past) singularity of a spacetime in the manner of BLK, it is essential that the coordinate system does not break down before reaching the singularity, i.e., that the future (or, respectively, past) directed normal geodesics emanating from Σ never cross each other. Furthermore, we want the coordinate system to "cover" the entire future singularity. This requirement can be formulated precisely by demanding that all TIP's be generated by geodesics of the coordinate system, i.e., that for each TIP, W, there exists at least one geodesics γ normal to Σ with $W = I^-(\gamma)$. Finally, we wish the synchronous coordinates to "reach the singularity" simultaneously, i.e., that each of the normal geodesics terminates at the same length τ_0. (In general τ_0, of course, may be infinite, but our main interest is the case of an "all encompassing singularity" with finite τ_0.) Thus, we may formalize these requirements by the following definition:

Definition: A globally hyperbolic spacetime (M, g_{ab}) is said to admit *simultaneous synchronous coordinates* (SSC) covering the future singularity if there exists a congruence of (nowhere crossing) hypersurface orthogonal, future inexten-

dible, timelike geodesics such that (1) the orthogonal hypersurfaces Σ_τ are C^1 Cauchy surfaces, (2) the geodesics all have equal length, τ_0, measured from one of the Cauchy surfaces Σ_0, and (3) every TIP is generated by (at least) one of the geodesics. (The requirement that Σ_τ be C^1 is needed for the normal vectors to Σ_τ to be defined. If one wishes to write down Einstein's equation in this coordinate system, the Σ_τ must be C^3.)

The main result of this section is that requirement (3) is unnecessary in the case of a spacelike singularity; it follows automatically from requirement (1). We begin by establishing two lemmas.

Lemma 1: Let (M, g_{ab}) be a globally hyperbolic spacetime, let W be a TIP and let Σ be a Cauchy surface. Then $W \cap \Sigma$ is connected.

Proof: First we note that W itself must be connected. If not, we would have $W = W_1 \cup W_2$, with W_1 and W_2 disjoint open sets. It is not difficult to show that W_1 and W_2 each must be past sets, and thus we would obtain a decomposition of W in violation of the definition of TIP's given at the beginning of this section. Similar arguments show that $W \cap I^+(\Sigma)$ must be connected.

To show that $W \cap \Sigma$ is connected, we note that every past directed causal curve from a point in W must remain in W, and that every inextendible, past directed causal curve from $I^+(\Sigma)$ must cross Σ, since Σ is a
Cauchy surface for (M, g_{ab}). Hence, in the spacetime (W, g_{ab}), every inextendible, past directed causal curve from $W \cap I^+(\Sigma)$ intersects $W \cap \Sigma$. Thus, in the spacetime (W, g_{ab}) we have $\text{int} D^+(W \cap \Sigma) = W \cap I^+(\Sigma)$, i.e., $W \cap \Sigma$ is a future Cauchy surface for W. Consequently $W \cap I^+(\Sigma)$ has the topology $\mathbb{R} \times (W \cap \Sigma)$. Thus, $W \cap \Sigma$ is disconnected if and only if $W \cap I^+(\Sigma)$ is disconnected. Since $W \cap I^+(\Sigma)$ is connected, we obtain the desired result.

Lemma 2: Let (M, g_{ab}) be a spacetime with a spacelike future causal boundary. (As remarked above, this implies that (M, g_{ab}) is globally hyperbolic.) Let W be a TIP and let Σ be a Cauchy surface. Then $\bar{W} \cap \Sigma$ is compact, where $\bar{W}$ denotes the closure of W.

Proof: Since (M, g_{ab}) is globally hyperbolic, we may choose a continuous global time function t throughout the future of Σ such that the surfaces of constant t, which we denote by Σ_t, are Cauchy surfaces. For convenience, we shall let t have the range $[0,1)$, with $t = 0$ corresponding to Σ. The first and most complicated step of the proof is to show that for any point $p \in W$, we can find a t such that $(J^+(p) \cap \Sigma_t) \supset W \cap \Sigma_t$ (see Fig. 2), where $J^+(p)$ denotes the causal future of p. Suppose this were false. Then for all t we must have $(\tilde{J}^+(p) \cap \Sigma_t) \cap W \neq \emptyset$, where $\tilde{J}^+(p)$ denotes the complement of $J^+(p)$ and is an open set since $J^+(p)$ is always a closed set in a globally hyperbolic spacetime. On the other hand, for all $t > t(p)$ we have $(\text{int} J^+(p)) \cap \Sigma_t \cap W \neq \emptyset$, where $\text{int} J^+(p) = I^+(p)$ denotes the interior of $J^+(p)$. Therefore, since by Lemma 1 $\Sigma_t \cap W$ is connected, for all $t > t(p)$ we must have $\dot{J}^+(p) \cap \Sigma_t \cap W \neq \emptyset$ where $\dot{J}^+(p)$ denotes the boundary of $J^+(p)$, as otherwise $\Sigma_t \cap W$ would be the disjoint union of two nonempty open sets. Thus, if we let $\{t_n\}$ be a sequence of numbers in $(t(p),1)$ with $\lim_{n \to \infty} t_n = 1$, then for each n there exists a point $q_n \in \dot{J}^+(p) \cap \Sigma_{t_n} \cap W$. Now any point in $\dot{J}^+(p)$ lies in $J^+(p)$ [since $J^+(p)$ is closed] but not in $I^+(p)$ [since $I^+(p)$ is open] and thus must be connected to p by a null geodesics. Let λ_n denote the future inextendible null geodesic which joins p to q_n. The segment of λ_n between p and q_n must be contained within W, since $q_n \in W$ and every past directed causal curve from a point in W must lie within W. Now, the $\{\lambda_n\}$ form a sequence of future inextendible nonspacelike curves, each of which pass through p. Hence, by Lemma 6.2.1 of Ref. 5, there exists a future inextendible nonspacelike limit curve λ passing through p. Let $r \in \lambda$. We now choose N sufficiently large that for all $n > N$ we have $t_n > t(r)$ and thus each λ_n for $n > N$ will be a null geodesic which, for $t \leqslant t(r)$, lies on $\dot{J}^+(p)$ and is contained within $\bar{W}$. Since, by definition of limit curve, every neighborhood or r intersects infinitely many λ_n this implies $r \in \overline{\dot{J}^+(p)} = \dot{J}^+(p)$ and $r \in \bar{W}$. Thus, λ is a future inextendible null geodesics which lies on $\dot{J}^+(p)$ and is contained within $\bar{W}$. Let V denote the TIP $I^-(\lambda)$. Then $V \subset W$ since $I^-(\lambda) \subset I^-(\bar{W}) = W$. (The relation $I^-(\bar{W}) = I^-(W)$ holds for an arbitrary set.) On the other hand, $p \notin I^-(\lambda)$ since $\lambda \subset \dot{J}^+(p)$, so $V \neq W$. Thus, we have constructed a TIP V which is a proper subset of W. This contradicts the hypothesis that the causal boundary of (M, g_{ab}) is spacelike. This contradiction proves the existence of a Σ_t such that $(J^+(p) \cap \Sigma_t) \supset W \cap \Sigma_t$.

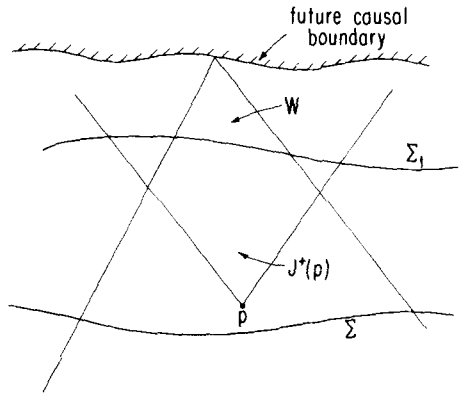


FIG. 2. Diagram for Lemma 2.

The remainder of the proof of the lemma is relatively simple. Taking the closure of the above result, we have $J^+(p) \cap \Sigma_t \supset \bar{W} \cap \Sigma_t$. By Proposition 6.6.6 of Ref. 5, $J^+(p) \cap J^-(\Sigma_t)$ is compact. Thus $\bar{W} \cap \Sigma_t$ is a closed subset of the compact set $J^+(p) \cap J^-(\Sigma_t)$ and thus is compact. To show that $\bar{W} \cap \Sigma$ is compact, we note that $\bar{W} \cap \Sigma \subset J^-(\bar{W} \cap \Sigma_t) \cap J^+(\Sigma)$. However, a simple extension of Proposition 6.6.6 of Ref. 5 shows that $J^-(K) \cap J^+(\Sigma)$ is compact for any compact set K. [To prove this, we cover K with sets of the form $I^-(q)$ where q is a point, take a finite subcover, and therby show that $J^-(K) \cap J^+(\Sigma)$, which is a closed set, is contained within the finite union of compact sets of the form $J^-(q_i) \cap J^+(\Sigma)$]. Thus, $\bar{W} \cap \Sigma$ is also a closed subset of a compact set [namely, $J^-(\bar{W} \cap \Sigma_t) \cap J^+(\Sigma)$] and hence is compact. This completes the proof of Lemma 2.

Using the fact that $\bar{W} \cap \Sigma$ is compact, we now prove that the normal geodesics of any Cauchy surface generate all TIP's.

Theorem 1: Let (M, g_{ab}) be a spacetime with a spacelike future causal boundary. Let W be a TIP and Σ be a C^1

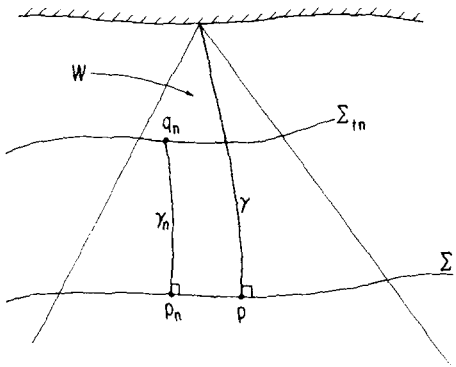


FIG. 3. Diagram for Theorem 1.

Cauchy surface. Then there exists a geodesic γ normal to Σ such that $W = I^-(\gamma)$.

Proof: Let the Cauchy surfaces $\{\Sigma_t\}$ with $t\in[0,1)$ foliate $J^+(\Sigma)$ as in Lemma 2. Let $\{t_n\}$ be a sequence in $[0,1)$ with $\lim_{n\to\infty} t_n = 1$ and let $q_n \in \Sigma_{t_n} \cap W$. Then by the corollary stated on p. 217 of Ref. 5, there exists a timelike geodesic γ_n orthogonal to Σ which passes through q_n, as illustrated in Fig. 3. (The C^2 and acausal properties of Σ assumed in the hypothesis of this corollary are not needed for this result.) Clearly the segment of γ_n between q_n and Σ lies within W. Let p_n be the intersection point of γ_n with Σ. Then $\{p_n\}$ is a sequence contained within the compact set $\bar{W}\cap\Sigma$ and thus has an accumulation point $p\in\bar{W}\cap\Sigma$. Let γ denote the future inextendible timelike geodesic orthogonal to Σ which passes through p. Since geodesics depend continuously on their initial point and initial tangent, it follows that γ will be a limit curve of the sequence $\{\gamma_n\}$. Hence, by the same argument used in the proof of Lemma 2, we have $\gamma\in\bar{W}$ and thus $I^-(\gamma)\subset W$. Since the future causal boundary is spacelike, the TIP $I^-(\gamma)$ cannot be a proper subset of W, so we must have $I^-(\gamma) = W$. This completes the proof.

III. EXISTENCE CRITERIA FOR SSC

In this section we shall derive necessary and sufficient conditions for the existence of SSC which terminate after a finite time τ_0 for a spacetime (M, g_{ab}) whose future causal boundary is spacelike. We begin by defining the "maximum lifetime function" $f: M\to\mathbb{R}$ for an arbitrary spacetime. For a causal curve λ, we let $T[\lambda]$ denote the length (i.e., proper time) of the curve

$$T[\lambda] = \int (-t^a t_a)^{1/2}\, dt, \tag{3.1}$$

where t^a is the tangent to λ and t is the curve parameter. For each $p\in M$ we define,

$$f(p) = \text{l.u.b.}\,\{T[\lambda]\}, \tag{3.2}$$

where λ ranges over all future directed causal curves which begin at p. We define $f(p) = \infty$ if $T[\lambda]$ is not bounded. The function $f(p)$ is denoted as $d(p, J^+(p))$ by Barrow and Tipler.[2]

Several important properties of f follow immediately from its definition. Clearly, if $q\in J^+(p)$ and λ is a causal curve connecting p with q, we have

$$f(p)\geqslant f(q) + T[\lambda]. \tag{3.3}$$

Hence f cannot increase along any future directed causal curve and, indeed, must strictly decrease along causal curves with the possible exception of null geodesics. Furthermore, Eq. (3.3) shows that if f is differentiable we must have $t^a\nabla_a f < 0$ for any timelike vector t^a, and thus $\nabla^a f$ is non-vanishing and is either timelike or null.

Our first important result is the following lemma:

Lemma 3: Suppose (M, g_{ab}) is a spacetime such that SSC exist, beginning at the Cauchy surface Σ_0 and terminating at the future causal boundary at time $\tau_0 < \infty$. Then for $p\in J^+(\Sigma_0)$, we have $f(p) = \tau_0 - \tau(p)$, where $\tau(p)$ denotes the SSC time of p. Thus, the geodesics of the SSC are curves of maximum length.

Proof: We know that $f(p)\geqslant\tau_0 - \tau(p)$ because the future directed geodesic through p of the SSC has length $\tau_0 - \tau(p)$. Suppose $f(p) > \tau_0 - \tau(p)$, i.e., suppose $f(p) = \tau_0 - \tau(p) + \epsilon$ with $\epsilon > 0$. Then there exists a point $q\in I^+(p)$ and a causal curve λ from p to q such that the length of λ is $T[\lambda] = \tau_0 - \tau(p) + \epsilon/2$. Let Σ denote the SSC Cauchy surface on which p lies. According to the corollary on p.217 of Ref. 5 there exists a curve γ of maximal length between q and Σ, which must be a geodesic orthogonal to Σ, and thus is a geodesic of the SSC. Thus, we have $T[\gamma]\geqslant T[\lambda] > \tau_0 - \tau(p)$ since γ is of maximal length. On the other hand, $T[\gamma] < \tau_0 - \tau(p)$ since it is a SSC geodesic. This contradiction proves that $f(p) = \tau_0 - \tau(p)$.

Lemma 3 shows that if SSC exist, then the Cauchy hypersurfaces Σ of the SSC are surfaces of constant f. Thus, if SSC exist, they are unique.[2] Since these surfaces are required to be C^1, it follows that f must be C^1. Furthermore, since for all $\tau < \tau_0$ the surface $f = \tau$ is a surface of the SSC and thus a Cauchy surface, it follows that f must approach zero along every future inextendible causal curve, since each such curve must cross the surface $f = \epsilon$ for arbitrarily small ϵ. The next theorem proves that these properties of f (or slight weakenings of them) are also sufficient to guarantee existence of SSC.

Theorem 2: Let (M, g_{ab}) be a spacetime with a spacelike future causal boundary. Then SSC exist which terminate in a finite time if and only if the maximum lifetime function f of the spacetime satisfies the following three properties: (i) There exists a Cauchy surface Σ and a $\tau_0 > 0$ such that $\tau_0 < f(p) < \infty$ for all $p\in\Sigma$; (ii) f is C^1 in $D^+(\Sigma)$; and (iii) f approaches zero along every future inextendible causal curve λ.

Proof: That the three conditions are necessary follows immediately from the remarks following Lemma 3. To prove sufficiency, we shall show that for all $\tau\leqslant\tau_0$ the surfaces S_τ defined by equation $f = \tau$ are hypersurfaces Σ_τ of SSC. According to the definition of SSC given in Sec. II, this entails showing that (a) each surface S_τ is a C^1 Cauchy surface, and (b) the normals of $\{S_\tau\}$ are tangent to a congruence of timelike geodesics, each of whose length measured starting from S_{τ_0} is τ_0. [The fact that the geodesics are normal to all the $\{S_\tau\}$ implies that they cannot cross. That the requirement (3) of SSC is satisfied follows, of course, from Theorem 1 of Sec. II.]

To show that the surface S_τ defined by $f = \tau$ is a Cauchy surface for $\tau < \tau_0$, we need to show that S_τ is achronal and that each inextendible causal curve λ intersects S_τ.

That S_τ is achronal follows immediately from Eq. (3.3). To show that every causal curve λ must cross S_τ we note that since Σ is a Cauchy surface, λ must cross Σ. Let p denote this intersection point. By property (i) we have $f(p) \geqslant \tau_0$. By property (iii), f must approach zero along λ, and by property (ii), f is continuous along λ. Hence, f must obtain the value τ along λ, i.e., λ does cross S_τ. This proves that S_τ is a Cauchy surface for all $\tau < \tau_0$. That S_τ is C^1 follows immediately from property (ii).

To show that (b) holds, we let $p \in S_{\tau_0}$ and consider an arbitrary surface S_τ for $\tau < \tau_0$. By the same corollary on p.217 of Ref. 5 which we have used several times before, there exists a maximum length curve γ from p to S_τ which must be a timelike geodesic orthogonal to S_τ. Let $q \in S_\tau$ denote the intersection of this curve with S_τ. If the geodesic γ had length greater than $\tau_0 - \tau$, we could construct a curve of length greater than τ_0 from p by joining γ to a curve from q of length arbitrarily close to τ. If γ had length less than $\tau_0 - \tau$, then since each timelike curve λ from p must cross S_τ, its length would be bounded by $T[\gamma] + \tau < \tau_0$. Since $f(p) = \tau_0$, neither of these possibilities can occur, and thus the length of γ must be $\tau - \tau_0$. Furthermore, γ must be orthogonal to S_{τ_0}; if not, we could find a curve connecting S_τ and S_{τ_0} with length greater than $\tau_0 - \tau$ and obtain a contradiction as above. Thus, we have proven that passing through each $p \in S_{\tau_0}$, there is a timelike geodesic γ which is orthogonal to S_{τ_0} and is also orthogonal to every other S_τ for $\tau < \tau_0$. Furthermore, the length of γ between S_{τ_0} and S_τ is simply $\tau_0 - \tau$, from which it follows that the total length of γ is τ_0. This proves property (b) and thus completes the proof of the theorem.

We note, that in Theorem 2 if Σ is compact and $f(p) < \infty$ for some point p, then property (i) follows automatically from the continuity of f implied by property (ii). Thus, in closed universes we need only check properties (ii) and (iii). The most difficult property to check is property (ii) since in spacetimes without special symmetries the maximum lifetime function is very difficult to compute explicitly. Nevertheless, properties (i)–(iii) provide a useful criterion for the existence of SSC and lend considerable plausibility to their existence in physically realistic situations. In the next section we shall attempt to gain further insight into the existence of SSC by constructing examples where, according to the criteria of Theorem 2, SSC cannot exist.

IV. EXAMPLES

In this section we shall present examples of spacetimes which have a spacelike future causal boundary where properties (i), (ii), or (iii) of Theorem 2 are not satisfied and thus SSC do not exist. By doing so, we will demonstrate explicitly the types of things that can go wrong in constructing SSC. In addition, our examples will show that properties (i), (ii), and (iii), of Theorem 2 are independent conditions [although we have not fully verified the validity of our proposed example of a spacetime which satisfies (i) and (ii) but not (iii)].

First, we shall present an example where properties (ii) and (iii) are satisfied, but not property (i). Let M be two-dimensional Minkowski spacetime between the surfaces $t = 0$ and $t = 1$ and let the metric be

$$ds^2 = \Omega^2(x)(-dt^2 + dx^2), \tag{4.1}$$

where Ω is smooth and $\Omega(x) \to 0$ as $x \to \infty$. (This spacetime is extendible, but by multiplying the metric by a suitable function of t we could construct an inextendible spacetime with similar properties.) Then (M, g_{ab}) will have a spacelike future causal boundary and properties (ii) and (iii) will be satisfied. However, as illustrated in Fig. 4, on any Cauchy surface Σ, we have $f \to 0$ as $x \to \infty$, so property (i) is not satisfied. However, the failure of SSC is not too serious in this case. The only thing that goes wrong is that the surfaces of constant f are not Cauchy hypersurfaces and not all of the singularity will be reached by constructing synchronous coordinates from a surface of constant f, although one can reach as much of the singularity as one wishes by starting the coordinate system on a surface of sufficiently small f.

A much more serious failure occurs if f is not C^1. First, we present an example where f is not even continuous. (This provides a counterexample to the claim of Barrow and Tipler[2] that f must be at least C^{1-}.) Let M be two-dimensional Minkowski spacetime for $t < 0$ and let the metric be

$$ds^2 = \Omega^2(x,t)(-dt^2 + dx^2) \tag{4.2}$$

with

$$\Omega(t,x) = 1 + h(\theta)/r, \tag{4.3}$$

where $r = (t^2 + x^2)^{1/2}$, $\tan\theta = -t/x$, and $h(\theta)$ is smooth, nonnegative, vanishes outside the spacelike lines $\theta = \theta_1$ and $\theta = \theta_2$, and rises to a maximum of 1 on an open interval in θ contained between θ_1 and θ_2, as illustrated in Fig. 5. Let λ be a causal curve with tangent t^a which passes through the wedge bounded by $\theta = \theta_1$ and $\theta = \theta_2$. Then we have

$$T[\lambda] = \int (-g_{ab} t^a t^b)^{1/2}\, dt \tag{4.4}$$

$$= \int \Omega (-\eta_{ab} t^a t^b)^{1/2}\, dt, \tag{4.5}$$

where η_{ab} denotes the flat metric $-dt^2 + dx^2$. Hence, we have

$$T_M[\lambda] + (1/r_{\max}) T'_M[\lambda] \leqslant T[\lambda]$$

$$\leqslant T_M[\lambda] + \frac{1}{r_{\min}} T''_M[\lambda], \tag{4.6}$$

where T_M denotes the total length of λ in the flat metric η_{ab}, $T'_M[\lambda]$ denotes its length in the flat metric in the portion of

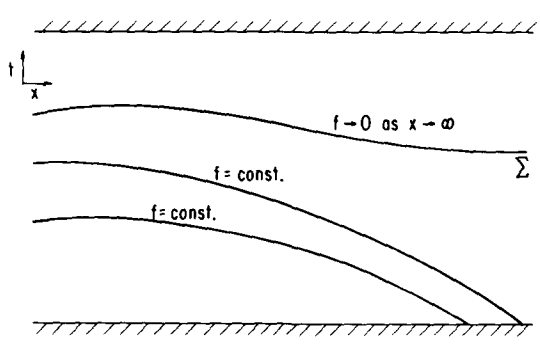


FIG. 4. An example which satisfies conditions (ii) and (iii), but violates (i).

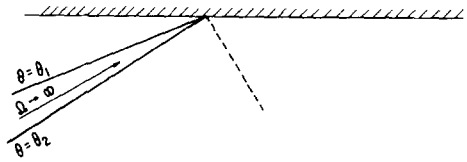


FIG. 5. An example satisfying (i), but not (ii) and (iii).

the wedge where $h = 1$, $T''_M[\lambda]$ denotes its flat metric length in the full wedge, and r_{max} and r_{min} denote the maximum and minimum value of r achieved by the curve in the dimensional Minkowski spacetime between the surfaces $t = 0$ and $t = 1$ and let the metric be

$$ds^2 = \Omega^2(x)(dt^2 - dx^2), \tag{4.1}$$

where Ω is smooth and $\Omega(x) \to 0$ as $x \to \infty$. (This spacetime is extendible, but by multiplying the metric by a suitable function of t we could construct an inextendible spacetime with similar properties.) Then (M, g_{ab}) will have a spacelike future causal boundary and properties (ii) and (iii) will be satisfied. However, as illustrated in Fig. 5, on any Cauchy surface Σ, we have $f \to 0$ as $x \to \infty$, so property (i) is not satisfied. However, the failure of SSC is not too serious in this case. The only thing that goes wrong is that the surfaces of constant f are not Cauchy hypersurfaces and not all of the singularity will be reached by constructing synchronous coordinates from a surface of constant f, although one can reach as much of the singularity as one wishes by starting the coordinate system on a surface of sufficiently small f.

A much more serious failure occurs if f is not C^2. First, we present an example where f is not even continuous. (This provides a counterexample to the claim of Barrow and Tipler[2] that f must be at least C^{1-}.) Let M be two-dimensional Minkowski spacetime for $t < 0$ and let the metric be

$$ds^2 = \Omega^2(x,t)(-dt^2 + dx^2) \tag{4.2}$$

with

$$\Omega(t,x) = 1 + h(\theta)/r, \tag{4.3}$$

where $r = (t^2 + x^2)^{1/2}$, $\tan\theta = -t/x$, and $h(\theta)$ is smooth, nonnegative, vanishes outside the spacelike lines $\theta = \theta_1$ and $\theta = \theta_2$, and rises to a maximum of 1 on an open interval in θ contained between θ_1 and θ_2, as illustrated in Fig. 6. Let λ be a causal curve with tangent t^a which passes through the wedge bounded by $\theta = \theta_1$ and $\theta = \theta_2$. Then we have

$$T[\lambda] = \int (-g_{ab}t^a t^b)^{1/2}\, dt \tag{4.4}$$

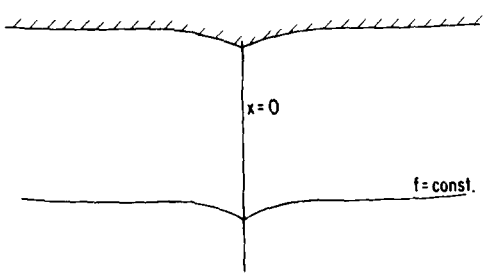


FIG. 6. Minkowski spacetime with a "sharp corner." Properties (i) and (iii), but not (ii), are satisfied.

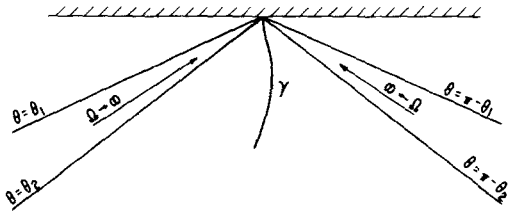


FIG. 7. A two-dimensional spacetime with two "wedges". Property (i) is satisfied and we believe that property (ii) can be satisfied, but (iii) is violated.

$$= \int \Omega (-\eta_{ab}t^a t^b)^{1/2}\, dt, \tag{4.5}$$

where η_{ab} denotes the flat metric $-dt^2 + dx^2$. Hence, we have

$$T_M[\lambda] + (1/r_{max})T'_M[\lambda] \leqslant T[\lambda] \leqslant T_M[\lambda] + \frac{1}{r_{min}}T''_M[\lambda], \tag{4.6}$$

where T_M denotes the total length of λ in the flat metric η_{ab}, $T'_M[\lambda]$ denotes its length in the flat metric in the portion of the wedge where $h = 1$, $T''_M[\lambda]$ denotes its flat metric length in the full wedge, and r_{max} and r_{min} denote the maximum and minimum value of r achieved by the curve in the full wedge. Furthermore, the Minkowski length of the curve in the full wedge is bounded by the size of the wedge, which is proportional to the "radius", i.e.,

$$T''_M[\lambda] \leqslant Cr_{min} \tag{4.7}$$

On the other hand, there exists a constant C' such that for any given entry point in the wedge in the wedge we can find a curve σ such that

$$T'_M[\sigma] \geqslant C' r_{max}. \tag{4.8}$$

Eqs. (4.6)–(4.8) demonstrate that the maximum lifetime function will get a finite "kick" at points which have access to the wedge. Consequently, f will be discontinuous along the null line shown in Fig. 5 for $t > -C'$. Thus, property (ii) fails badly and SSC do not exist. Property (iii) also fails, as f does not approach zero along the causal curves which generate the TIP represented by the point $x = 0$, $t = 0$. However, property (i) is satisfied in this example.

If f is continuous, then the surface of constant f will be achronal boundaries[5] and hence f must be at least C^1. However, it is easy to find examples where f is not C^1. Let M be two-dimensional Minkowski spacetime below a surface with a downward pointing sharp corner at $x = 0$, as illustrated in Fig. 6. We let the metric be the flat Minkowski metric. Then the surfaces of constant f will have similar sharp corners and f will fail to be C^1 along the $x = 0$ axis, so SSC will not exist, although properties (i) and (iii) are satisfied. Note that if the corner were rounded off, f would still fail to be C^1 on the axis sufficiently far below the causal boundary, but near the causal boundary, f would become C^∞ and SSC would exist.

The spacetime in the above example is, of course, extendible and the example is far from physically realistic. However, analogous behavior occurs in some inextendible spacetimes representing spherically symmetric collapse. We have verified that for the Vaidya metric describing the col-

lapse of a null fluid,

$$ds^2 = -(1-g)\,dv^2 + 2\,dv\,dr + r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \tag{4.9}$$

with

$$g = \begin{cases} 0 & v \leqslant 0 \\ v/r & v > 0 \end{cases} \tag{4.10}$$

f fails to be C^1 along the "axis" $r = 0$ in a manner similar to the behavior of f along the $x = 0$ axis of the above example.[7] In this example, the spacetime metric itself is not C^1 but we believe that it is possible to smooth out the spacetime metric and preserve the nondifferentiable property of f while maintaining the nonnegativity of energy density. However, we expect that if the collapse were made less violent or less highly focused the effect would be analogous to "rounding off the corner" in the previous example, and f would become smooth near the singularity. The question of whether f fails to be C^1 in physically realistic examples of gravitational collapse is the most important outstanding issue concerning the existence of SSC.

Finally, we propose an example which satisfies (i) and (ii) but not (iii). We modify the example of Fig. 5 by adding a second "conformal factor wedge" as illustrated in Fig. 7. Then property (i) is satisfied and we believe that by a suitable choice of Ω we can make f be smooth in a neighborhood of the future causal boundary, although we have not proved this, However, property (iii) will not be satisfied as f will not approach zero along the timelike curve γ shown. Thus SSC will not exist in this example.

ACKNOWLEDGMENT

This research was supported in part by NSF grant PHY-78-24275 to the University of Chicago.

[1]V. A. Belinskii, J. M. Khalatnikov, and E. M. Lifschitz, Adv. Phys. **19**, 525–573 (1970).
[2]J. D. Barrow and F. J. Tipler, Phys. Rep. **56**, 371–402 (1979).
[3]R. Penrose, in *General Relativity: An Einstein Centennary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, England, 1979).
[4]C. W. Misner, Phys. Rev. **186**, 1328–1333 (1969).
[5]S. W. Hawking and G. R. F. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, England, 1973).
[6]R. P. Geroch, E. H. Kronheimer, and R. Penrose, Proc. Roy. Soc. London Ser. A **327**, 545–567 (1972).
[7]We are indebted to D. Eardley for providing us with this example.

The equivalence of electromagnetic fields and viscous fluids in general relativity

B. O. J. Tupper
Department of Mathematics and Statistics, University of New Brunswick, Fredericton, New Brunswick, E3B 5A3, Canada

(Received 22 December 1980; accepted for publication 30 January 1981)

It is shown that spacetimes satisfying the Einstein–Maxwell equations, either *in vacuo* or coupled with a perfect fluid, may also satisfy the field equations for a viscous fluid. The necessary conditions for this alternative interpretation are found and examples of Einstein–Maxwell solutions admitting, and also not admitting, the viscous fluid interpretation are given.

PACS numbers: 0.4.40. + c, 04.20.Jb

I. INTRODUCTION

Synge[1] has commented to the effect that one approach to the field equations of general relativity, viz.,

$$G_{\mu\nu} = T_{\mu\nu}, \tag{1.1}$$

where $T_{\mu\nu}$ is the stress–energy tensor of the material distribution, is to regard these not as a set of partial differential equations for the metric tensor $g_{\mu\nu}$ for a given $T_{\mu\nu}$, but as a set of equations which determine $T_{\mu\nu}$ for any given metric tensor. In other words, given a spacetime with metric tensor $g_{\mu\nu}$, we substitute these functions into the left-hand side of Eq. (1.1) and read off the components of $T_{\mu\nu}$. Of course, this approach will, in most cases, lead to a $T_{\mu\nu}$ which is physically unreasonable and which will not satisfy the necessary energy conditions,[2] so many spacetimes will have to be rejected.

One consequence of this approach that was not mentioned by Synge is that one is forced to pose the question: Given a metric tensor that does, in fact, lead to a $T_{\mu\nu}$ satisfying the energy conditions, how do we know what type of field this energy tensor represents? Put another way, is it possible that $T_{\mu\nu}$ is *nonuniquely* a viable energy tensor, i.e., can the energy tensors of two apparently different fields be identical in that they have precisely the same components?

The answer to this is in the affirmative. It is already known[3] that, in certain circumstances, the energy tensor of the neutrino field can be identically zero, so that the corresponding spacetime is also a vacuum solution. This is the case of the so-called "ghost-neutrinos"; considered as neutrino solutions they are pathological. The neutrino energy tensor can also be identical to the energy tensor of a null electromagnetic field provided that certain conditions hold. However, the neutrino field is not a classical field, and relating a nonclassical field to a classical field seems a somewhat dubious procedure. In this article we shall show that two classical fields, namely an electromagnetic field and an imperfect fluid distribution, may have identical energy tensors. This identity can also exist when the electromagnetic field is generalized to include perfect fluid source terms.

The problem of the identification of different energy tensors can be discussed in the realm of special relativity with a background of Minkowski spacetime. However, we are interested in energy tensors which are not only identical, but which each form the right-hand sides of a set of field equations of the form (1.1) with the identical Einstein tensor $G_{\mu\nu}$ on the left-hand side, and so we shall confine our attention to the general relativistic problem.

II. THE ELECTROVAC CASE

We consider the following pair of field equations:

$$G_{\mu\nu} = E_{\mu\nu}, \tag{2.1}$$

where $E_{\mu\nu}$ is the stress–energy tensor of the electromagnetic field and is given by

$$E_{\mu\nu} = F_{\mu\alpha}F_{\nu}{}^{\alpha} - \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}, \tag{2.2}$$

where $F_{\mu\nu}$ is the Maxwell tensor, and

$$G_{\mu\nu} = M_{\mu\nu}, \tag{2.3}$$

where $M_{\mu\nu}$ is the stress–energy tensor of a viscous fluid and is given by

$$M_{\mu\nu} = (\rho + p^*)u_\mu u_\nu + p^* g_{\mu\nu} - 2\eta\sigma_{\mu\nu}, \tag{2.4}$$

where ρ is the density, $p^* = p - \zeta\theta$ is the kinetic pressure, p is the thermodynamic pressure, θ is the expansion, $\sigma_{\mu\nu}$ is the shear tensor, $\zeta\ (\geqslant 0)$ is the bulk viscosity, and $\eta\ (\geqslant 0)$ is the shear viscosity. For the moment we shall omit heat conduction terms from $M_{\mu\nu}$.

It is our contention that the two energy tensors $E_{\mu\nu}$ and $M_{\mu\nu}$ can be identical so that the same spacetime can satisfy the field equations (2.1) and also (2.3). Equating $E_{\mu\nu}$ and $M_{\mu\nu}$ and using the fact that $E_{\mu\nu}$ satisfies the Rainich conditions, viz.,

$$E^{\mu}{}_{\mu} = 0, \quad E_{\mu\alpha}E^{\nu\alpha} = \tfrac{1}{4}\delta_{\mu}{}^{\nu}E_{\alpha\beta}E^{\alpha\beta}, \tag{2.5}$$

we obtain, from (2.4), the following conditions:

$$\rho = 3p^* \tag{2.6}$$

and

$$(4u_\mu u^\nu + \delta_\mu{}^\nu)p^{*2} + 2\eta p^*\sigma_\mu{}^\nu + \eta^2(\sigma^2\delta_\mu{}^\nu - 2\sigma_{\mu\alpha}\sigma^{\nu\alpha}) = 0, \tag{2.7}$$

where $\sigma^2 = \tfrac{1}{2}\sigma_{\alpha\beta}\sigma^{\alpha\beta}$. Contracting Eq. (2.7) with $u^\mu u^\nu$ and $\sigma^{\mu\nu}$ yields the following relations:

$$\eta^2\sigma^2 = 3p^{*2}, \tag{2.8}$$

$$\eta\Sigma = 2p^*\sigma^2, \tag{2.9}$$

where $\Sigma = \sigma_\mu{}^\alpha\sigma_\alpha{}^\nu\sigma_\nu{}^\mu$. Using these two relations, Eq. (2.7) can be written in the form

$$(\Sigma\sigma^{-2}h_\mu{}^\alpha - \sigma_\mu{}^\alpha)(\Sigma\sigma^{-2}h_\alpha{}^\nu + 2\sigma_\alpha{}^\nu) = 0, \tag{2.10}$$

where $h_\mu{}^\nu = \delta_\mu{}^\nu + u_\mu u^\nu$ is the projection tensor. Equation

 0022-2488/81/112666-08$01.00

(2.10) is the equation that must be satisfied by u^μ in order for the identification of $E_{\mu\nu}$ with $M_{\mu\nu}$ to be possible. For a spacetime with a diagonal metric and with $u^\mu = (1,0,0,0)$, Eq. (2.10) would be satisfied, for example, by $\sigma_1{}^1 = \Sigma\sigma^{-2}$, $\sigma_2{}^2 = \sigma_3{}^3 = -\frac{1}{2}\Sigma\sigma^{-2}$, since $\sigma_\alpha{}^\alpha = 0$.

From Eq. (2.6), p^* is necessarily nonzero and positive. Using the relations (2.6) and (2.8), we find that the invariant $E_{\alpha\beta}E^{\alpha\beta} = M_{\alpha\beta}M^{\alpha\beta}$ has value $36p^{*2}$, so it follows that $E_{\alpha\beta}E^{\alpha\beta}$ cannot be zero, i.e., the electromagnetic field must be nonnull. Hence, a spacetime whose metric satisfies the field equations (2.1) for a nonnull electrovac spacetime also satisfies the field equations (2.3) for a viscous fluid spacetime (without heat conduction) provided that we can find a 4-velocity u^μ which satisfies Eq. (2.10) and also the condition $\Sigma > 0$, so that, from Eq. (2.9), $\eta > 0$.

We shall now give a number of examples of known nonnull electrovac solutions for which a vector u^μ satisfying Eq. (2.10) can be found so that the solution may be interpreted as a viscous fluid solution. Since, in each case, the necessary energy conditions are satisfied by the electromagnetic energy tensor, they will be satisfied by the energy tensor of the viscous fluid.

(i) The solution found by Tariq and Tupper,[4] viz.,

$$ds^2 = -dt^2 + \tfrac{4}{3}t^2\,dx^2 + te^{-2x}\,dy^2 + te^{2x}\,dz^2, \quad (2.11)$$

$$F_{01} = (2/\sqrt{3})\cos a\,, \quad F_{23} = \sin a,$$

where a is a constant. The metric (2.11) is also a solution of the field equations (2.4) with

$$u^\mu = (1,0,0,0), \quad \rho = 3p^* = \tfrac{1}{2}t^{-2}, \quad \eta = t^{-1}. \quad (2.12)$$

(ii) The Kerr–Newman[5] solution. In Boyer–Lindquist coordinates the metric is

$$ds^2 = -\Delta R^{-2}(dt - a\sin^2\theta\,d\phi)^2 + R^{-2}\sin^2\theta\,[(r^2+a^2)\,d\phi - a\,dt]^2 + R^2\Delta^{-1}\,dr^2 + R^2\,d\theta^2, \quad (2.13)$$

where $R^2 = r^2 + a^2\cos^2\theta$ and $\Delta = r^2 + a^2 + Q^2 - 2Mr$. This satisfies the field equations (2.4) if u^μ has the form

$$u^\mu = R^{-1}\Delta^{-1/2}[(r^2+a^2)\alpha,\ \Delta\beta,\ 0,\ a\alpha], \quad (2.14)$$

where $\alpha^2 - \beta^2 = 1$, in which case

$$\rho = 3p^* = Q^2R^{-4}, \quad \eta = Q^2R^{-1}\Delta^{1/2}[R^2\Delta\beta' + (R^2r - R^2M - 2r\Delta)\beta]^{-1}, \quad (2.15)$$

and β is nonzero and must be chosen so that $\eta > 0$. This requirement is easy to satisfy; for example, the choice $\beta = R^3\Delta^{-1/2}$ leads to $\eta = Q^2R^{-4}r^{-1}$. This freedom of choice for the function β illustrates a curious feature of the viscous fluid field equations, namely, the same energy tensor components can result from different choices of the 4-velocity.

If we simplify the Kerr–Newman solution by putting $a = 0$ and $Q = M$, we obtain the critical Reissner–Nordstrom solution

$$ds^2 = -(1 - Q/r)^2dt^2 + (1 - Q/r)^{-2}dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \quad (2.16)$$

with a singular sphere at $r = Q$. If we make the choice $\beta = r^3(r-e)^{-1}$, corresponding to the choice made above for the Kerr–Newman solution, we obtain $\eta = Q^2r^{-5}$, which is valid for all values of $r > 0$. However, if we choose $\beta = kr^{-2}$, where k is a constant, we obtain $\eta = Q^2k^{-1}(4Q - 3r)^{-1}$. In this case, if $k > 0$ ($k < 0$), i.e., if the radial velocity is outward (inward), then the model is confined to the region $r < \frac{4}{3}Q$ ($r > \frac{4}{3}Q$). For the Kerr–Newman solution, also, there exist choices of β for which the viscous fluid interpretation, with $\eta > 0$, is valid only in certain regions of spacetime.

(iii) The Bertotti–Robinson[6,7] solution with metric

$$ds^2 = -e^{-2r}dt^2 + a^2\,dr^2 + a^2(d\theta^2 + \sin^2\theta\,d\phi^2), \quad (2.17)$$

is a solution of Eq. (2.4) with

$$\rho = 3p^* = a^{-2}, \quad \eta = a^{-1}(\beta' - \beta)^{-1}, \quad u^\mu = (\alpha e^r,\ \beta a^{-1},\ 0,\ 0), \quad (2.18)$$

where $\alpha^2 - \beta^2 = 1$. Again, β must be chosen so that $\eta > 0$.

The question of whether or not every nonnull electrovac spacetime can admit the viscous fluid interpretation has been answered by Raychaudhuri and Saha,[8] who showed that there are electrovac solutions which cannot admit the second interpretation. The crux of their argument can be restated as follows: The field equations (2.3) can be rewritten in the form

$$R_\mu{}^\nu = \tfrac{1}{3}\rho(4u_\mu u^\mu + \delta_\mu{}^\nu) - 2\eta\sigma_\mu{}^\nu, \quad (2.19)$$

so that

$$R_\mu{}^\nu u_\nu = -\rho u_\mu, \quad (2.20)$$

i.e., u_μ is an eigenvector of $R_\mu{}^\nu$ with eigenvalue $-\rho$. Now we can always find an orthonormal tetrad $e_i{}^\mu$, where the Latin suffixes are tetrad suffixes which are raised and lowered by the Minkowski metric tensor $\eta^{ij} = \eta_{ij} = \text{diag}(-1,1,1,1)$, such that the tetrad components of the electromagnetic tensor $E_{ij} = e_i{}^\mu e_j{}^\nu E_{\mu\nu}$ have the canonical form[2]

$$E_{ij} = \begin{pmatrix} X^2 & 0 & 0 & 0 \\ 0 & -X^2 & 0 & 0 \\ 0 & 0 & X^2 & 0 \\ 0 & 0 & 0 & X^2 \end{pmatrix}, \quad (2.21)$$

with a suitable labelling of the coordinates. Substituting these values into the tetrad form of Eq. (2.20), we obtain $u_{2'} = u_{3'} = 0$, where the prime indicates a tetrad component. From the tetrad form of Eq. (2.19), this implies that

$$\sigma_{2'}{}^{2'} = \sigma_{3'}{}^{3'}. \quad (2.22)$$

Hence, for the viscous fluid interpretation to be possible, the 4-velocity u_μ must have tetrad components of the form $u_i = (u_{0'}, u_{1'}, 0, 0)$, and the corresponding shear tensor must have tetrad components satisfying Eq. (2.22).

Raychaudhuri and Saha gave an example of a known nonnull electrovac solution[9,10] which does not satisfy Eq. (2.22). The metric is

$$ds^2 = -A^{-1}dt^2 + A\,dx^2 + x^2A\,dy^2 + A\,dz^2,$$

where $A = [\ln ax^m]^2$, a and m being constants. We have found another such example, namely the electrovac solution with metric[11,12]

$$ds^2 = -(dt - 2z\,d\phi)^2 + r^2\,d\phi^2 + a^2r^{-2}(dr^2 + dz^2).$$

Labelling the coordinates $(t, r, z, \phi) = (x^0, x^1, x^2, x^3)$, we find that the nonzero components of the tetrad and its inverse which diagonalize the Ricci tensor are

$$\begin{aligned} &e^{0'}{}_0 = 1, \quad e^{0'}{}_3 = -2z, \quad e^{1'}{}_1 = ar^{-1}, \\ &\quad e^{2'}{}_2 = ar^{-1}, \quad e^{3'}{}_3 = r, \\ &e_{0'}{}^0 = 1, \quad e_{1'}{}^1 = a^{-1}r, \quad e_{2'}{}^2 = a^{-1}r, \\ &\quad e_{3'}{}^0 = 2zr^{-1}, \quad e_{3'}{}^3 = r^{-1}. \end{aligned}$$

The 4-velocity has tetrad components $u_i = (u_{0'}, u_{1'}, 0, 0)$, where $u_{0'}$, $u_{1'}$ are arbitrary functions of the coordinates satisfying $u_{0'}{}^2 - u_{1'}{}^2 = 1$, so that $u_\mu = (u_{0'}, ar^{-1}u_{1'}, 0, -2zu_{0'})$. Equation (2.22) has the coordinate form $\sigma_2{}^2 = 2z\sigma_0{}^3 + \sigma_3{}^3$, and it is easy to show that this condition cannot be satisfied by u_μ for any choice of $u_{0'}$ and $u_{1'}$.

III. HEAT CONDUCTION TERMS

We now extend the investigation of the previous section by including heat conduction terms in the viscous fluid energy tensor, i.e., we consider the following situation

$$G_{\mu\nu} = E_{\mu\nu} = (\rho + p^*)u_\mu u_\nu + p^* g_{\mu\nu} - 2\eta\sigma_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu, \tag{3.1}$$

where q_μ is the heat conduction vector satisfying $q_\mu u^\mu = 0$. Equation (2.6) still holds, so Eq. (3.1) can be rewritten in the form

$$R_{\mu\nu} = E_{\mu\nu} = \tfrac{1}{3}\rho(4u_\mu u_\nu + g_{\mu\nu}) - 2\eta\sigma_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu. \tag{3.2}$$

The Rainich conditions for $E_{\mu\nu}$, on contraction with u^ν, yield

$$4p^2 = 3Q^2 + 12\eta^2\sigma^2, \tag{3.3}$$

$$\rho q_\mu + 3\eta\sigma_{\mu\nu}q^\nu = 0, \tag{3.4}$$

where $Q^2 = q_\mu q^\mu$. The Rainich conditions can be simplified to

$$4(\rho^2 - 9\eta^2\sigma^2)h_\mu{}^\nu - 12\rho\eta\sigma_\mu{}^\nu + 36\eta^2\sigma_{\mu\alpha}\sigma^{\alpha\nu} - 9q_\mu q^\nu = 0. \tag{3.5}$$

Equation (3.2) leads to

$$R_\mu{}^\nu u_\nu = -\rho u_\mu - q_\mu, \tag{3.6}$$

so that u_μ is not an eigenvector of $R_\mu{}^\nu$. Assuming that the electromagnetic field is nonnull, we transform Eq. (3.6) into tetrad components and use the canonical form (2.21) to obtain

$$q_{0'} = (X^2 - \rho)u_{0'}, \quad q_{1'} = (X^2 - \rho)u_{1'}, \quad q_{2'} = -(X^2 + \rho)u_{2'}, \quad q_{3'} = -(X^2 + \rho)u_{3'}. \tag{3.7}$$

From $q_i u^i = 0$ and Eq. (3.7) we obtain

$$\rho = X^2(2u_{0'}{}^2 - 2u_{1'}{}^2 - 1) = X^2(2u_{2'}{}^2 + 2u_{3'}{}^2 + 1), \tag{3.8}$$

since $u_{0'}{}^2 - u_{1'}{}^2 - u_{2'}{}^2 - u_{3'}{}^2 = 1$, so that (3.7) becomes

$$q_{0'} = -2X^2(u_{2'}{}^2 + u_{3'}{}^2)u_{0'}, \quad q_{1'} = -2X^2(u_{2'}{}^2 + u_{3'}{}^2)u_{1'}, \quad q_{2'} = -2X^2(u_{0'}{}^2 - u_{1'}{}^2)u_{2'}, \quad q_{3'} = -2X^2(u_{0'}{}^2 - u_{1'}{}^2)u_{3'}, \tag{3.9}$$

and

$$Q^2 = 4X^4(u_{0'}{}^2 - u_{1'}{}^2)(u_{2'}{}^2 + u_{3'}{}^2). \tag{3.10}$$

Equations (3.3), (3.8), and (3.10) lead to

$$12\eta^2\sigma^2 = 4X^4(2u_{2'}{}^2 + 2u_{3'}{}^2 + 1)^2 - 12X^4(1 + u_{2'}{}^2 + u_{3'}{}^2) \times (u_{2'}{}^2 + u_{3'}{}^2),$$

i.e.,

$$3\eta^2\sigma^2 = X^4(\beta^4 + \beta^2 + 1), \tag{3.11}$$

where $\beta^2 = u_{2'}{}^2 + u_{3'}{}^2$. From the tetrad form of Eq. (3.2) and Eqs. (3.8), (3.10), and (3.11), we obtain

$$\left.\begin{aligned} 2\eta\sigma_{0'2'} &= -\tfrac{2}{3}X^2(2\beta^2 + 1)u_{0'}u_{2'}, \\ 2\eta\sigma_{1'2'} &= -\tfrac{2}{3}X^2(2\beta^2 + 1)u_{1'}u_{2'}, \\ 2\eta\sigma_{2'2'} &= \tfrac{2}{3}X^2(-1 + \beta^2 - 4u_{2'}{}^2 - 2\beta^2u_{2'}{}^2), \\ 2\eta\sigma_{2'3'} &= -\tfrac{4}{3}X^2(\beta^2 + 2)u_{2'}u_{3'}, \\ 2\eta\sigma_{0'3'} &= -\tfrac{2}{3}X^2(2\beta^2 + 1)u_{0'}u_{3'}, \\ 2\eta\sigma_{1'3'} &= -\tfrac{2}{3}X^2(2\beta^2 + 1)u_{1'}u_{3'}, \\ 2\eta\sigma_{3'3'} &= -\tfrac{2}{3}X^2(-1 + \beta^2 - 4u_{3'}{}^2 - 2\beta^2u_{3'}{}^2). \end{aligned}\right\} \tag{3.12}$$

Using Eqs. (3.8), (3.9), (3.11), and (3.12), we find that the (2', 2') and (3', 3') tetrad components of Eq. (3.5) lead to

$$\beta^2(2\beta^2 + 1)^2u_{2'}{}^2 = 0, \quad \beta^2(2\beta^2 + 1)^2u_{3'}{}^2 = 0,$$

which imply that $u_{2'} = u_{3'} = 0$. Hence, from Eq. (3.9), we see that $q_\mu = 0$, and so a neccessary condition for the equivalence between the energy tensors of a nonull electrovac field and a viscous fluid distribution is that the viscous fluid must have zero heat conduction, as described in Sec. II.

From Eq. (3.2) we find that

$$E_{\mu\nu}E^{\mu\nu} = \tfrac{4}{3}\rho^2 + 8\eta^2\sigma^2 - 2Q^2, \tag{3.13}$$

so that, in the presence of heat conduction terms it is possible for $E_{\mu\nu}E^{\mu\nu} = 0$,i.e., for the electromagnetic field to be null, in which case

$$2\rho^2 + 12\eta^2\sigma^2 - 3Q^2 = 0. \tag{3.14}$$

On contracting Eq. (3.5) with u^ν and using Eq. (3.14), we obtain

$$\rho^2 = Q^2 = 12\eta^2\sigma^2, \tag{3.15}$$

for a null electromagnetic field, in which case Eq. (3.5) can be written in the form

$$L_{\mu\alpha}L^{\nu\alpha} = 0, \tag{3.16}$$

where $L_{\mu\nu}$ is a symmetric tensor given by

$$L_{\mu\nu} = -\tfrac{1}{3}Q^2 h_{\mu\nu} + 2|Q|\eta\sigma_{\mu\nu} + q_\mu q_\nu. \tag{3.17}$$

Since $L_\mu{}^\mu = 0$, Eq. (3.16) shows that $L_{\mu\nu}$ can be expressed in the form

$$L_{\mu\nu} = L_\mu L_\nu, \tag{3.18}$$

where L_μ is a null vector. Now, for a null electromagnetic field, the energy tensor $E_{\mu\nu}$ has the form

$$E_{\mu\nu} = M_\mu M_\nu, \tag{3.19}$$

where M_μ is a null vector. Eliminating $\eta\sigma_{\mu\nu}$ between Eqs. (3.2) and (3.17) and using Eqs. (3.18) and (3.19), we find that L_μ and M_μ are each multiples of the null vector

$$N_\mu = u_\mu + |Q|^{-1}q_\mu, \tag{3.20}$$

i.e., N_μ is the null eigenvector of $E_{\mu\nu}$.

The canonical tetrad components of the null electromagnetic energy tensor are of the form[2]

$$E_{ij} = \begin{pmatrix} A^2 & A^2 & 0 & 0 \\ A^2 & A^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.21}$$

Using this, together with the tetrad form of Eq. (3.6), we obtain

$$\begin{aligned} q_{0'} &= A^2(u_{0'} - u_{1'})^2 u_{1'} - A^2(u_{0'} - u_{1'})(u_{2'}{}^2 + u_{3'}{}^2), \\ q_{1'} &= A^2(u_{0'} - u_{1'})^2 u_{0'} - A^2(u_{0'} - u_{1'})(u_{2'}{}^2 + u_{3'}{}^2), \\ q_{2'} &= -A^2(u_{0'} - u_{1'})^2 u_{2'}, \quad q_{3'} = -A^2(u_{0'} - u_{1'})^2 u_{3'}, \end{aligned} \tag{3.22}$$

and

$$\rho = |Q| = A^2(u_{0'} - u_{1'})^2. \tag{3.23}$$

Thus, we have shown that a necessary condition for the equivalence of the energy tensors of a null electrovac field and a viscous fluid distribution is that the viscous fluid must have nonzero heat conduction. Note that, in contrast to the nonnull case,it is possible for $u_{2'} = u_{3'} = 0$, since this does not imply that $q_\mu = 0$. However, if we follow the same procedure as used in the nonnull case, we do not find that $u_{2'} = u_{3'} = 0$ necessarily since, in this case, the equations are all satisfied identically.

The energy tensor of the electromagnetic field can be written in the form[13]

$$\begin{aligned} E_{\mu\nu} = {} & (\tfrac{1}{2}g_{\mu\nu} + v_\mu v_\nu)(E_\alpha E^\alpha + B_\alpha B^\alpha) \\ & - (E_\mu E_\nu + B_\mu B_\nu) + (v_\mu S_\nu + v_\nu S_\mu), \end{aligned} \tag{3.24}$$

where E_μ, B_μ, and S_μ are, respectively, the electric field, the magnetic field, and the Poynting vector, as measured by a comoving observer and defined by

$$E_\mu = F_{\mu\nu}v^\nu, \quad B_\mu = \tfrac{1}{2}\eta_{\mu\nu\alpha\beta}v^\nu F^{\alpha\beta}, \quad S_\mu = \eta_{\mu\alpha\beta\gamma}E^\alpha B^\beta v^\gamma. \tag{3.25}$$

It follows that $E_\mu v^\mu = B_\mu v^\mu = S_\mu v^\mu = E_\mu S^\mu = B_\mu S^\mu = 0$ and provided that E_μ and B_μ are not parallel nor one of them zero, i.e., $S_\mu \neq 0$, it follows that v_μ, E_μ, B_μ, and S_μ form a set of basis vectors at any point in spacetime.

If we make the assumption that the timelike 4-vector v_μ associated with the electromagnetic field in the expression (3.24) for $E_{\mu\nu}$ is identical with the 4-velocity u_μ appearing in the viscous fluid energy tensor (3.1), then, on writing $v_\mu \equiv u_\mu$ and contracting the expression (3.25) with u^ν, we obtain

$$E_\mu{}^\nu u_\nu = -\tfrac{1}{2}(E^2 + B^2)u_\mu - S_\mu, \tag{3.26}$$

where $E^2 = E_\alpha E^\alpha$ and $B^2 = B_\alpha B^\alpha$. Comparing Eqs. (3.6) and (3.26), noting that both q_μ and S_μ are orthogonal to u_μ, we find that

$$\rho = \tfrac{1}{2}(E^2 + B^2), \tag{3.27}$$

$$q_\mu = S_\mu. \tag{3.28}$$

The last expression shows that it is necessary to have a nonzero Poynting vector in order to include heat conduction. Alternatively, this expression shows that, in the absence of heat conduction, as discussed in Sec. II, u_μ must be chosen so that $S_\mu = 0$.

The expressions (3.24)–(3.28) hold for any electromagnetic field. If we specify a null field, then

$$E^2 = B^2, \quad E_\alpha B^\alpha = 0, \tag{3.29}$$

$$S^2 = S_\alpha S^\alpha = E^4, \tag{3.30}$$

so that S_μ cannot be zero and v_μ, E_μ, B_μ, and S_μ always constitute a set of mutally orthogonal basis vectors. Introducing the unit basis vectors

$$e_\mu = |E|^{-1}E_\mu, \quad b_\mu = |B|^{-1}B_\mu, \quad s_\mu = |S|^{-1}S_\mu, \tag{3.31}$$

we find that $(v_\mu + s_\mu)$ is the repeated null eigenvector of $E_{\mu\nu}$ given by (3.24), so that we can write

$$E_{\mu\nu} = E^2(v_\mu + s_\mu)(v_\nu + s_\nu), \tag{3.32}$$

where the factor E^2 is determined by comparing the results of contracting each of the expressions (3.24) and (3.32) with v^ν. Alternatively, the expression (3.32) can be obtained from Eq. (3.24) by noting that, in terms of the unit mutually orthogonal basis vectors, the metric tensor takes the form

$$g_{\mu\nu} = -v_\mu v_\nu + e_\mu e_\nu + b_\mu b_\nu + s_\mu s_\nu. \tag{3.33}$$

The null eigenvector $v_\mu + s_\mu$ must be proportional to N_μ given by expression (3.20), i.e.,

$$v_\mu + |S|^{-1}S_\mu = k(u_\mu + |Q|^{-1}q_\mu). \tag{3.34}$$

If, as before, we assume $v_\mu \equiv u_\mu$, then $k = 1$, and we again obtain expression (3.28), so that $|Q| = |S| = E^2$. In this case, eliminating $\eta\sigma_{\mu\nu}$ between Eqs. (3.2) and (3.17), and using Eq. (3.32), we find that $L_{\mu\nu}$ is identically zero, so that Eq. (3.17) leads to the following expression for the shear viscosity term:

$$2\eta\sigma_{\mu\nu} = \tfrac{1}{3}E^2(e_\mu e_\nu + b_\mu b_\nu - 2s_\mu s_\nu). \tag{3.35}$$

Note that the fact that $L_{\mu\nu} = 0$ implies that $L_\mu = 0$. However, this does not imply that $N_\mu = 0$, but merely that L_μ is a zero multiple of N_μ.

In searching for an example of a null electrovac spacetime which satisfies the field equations (3.1), we note that the algebraic expression (3.35) for $\sigma_{\mu\nu}$ [or, if we do not assume $v_\mu \equiv u_\mu$, Eq. (3.2) rewritten as an expression for $\sigma_{\mu\nu}$] must agree with the expression for $\sigma_{\mu\nu}$ obtained from the covariant derivative of u_μ. We have found one null electrovac solution that satisfies all of these requirements, namely the most generally conformally flat null electrovac spacetime with metric[14]

$$ds^2 = -f^2(u)(x^2 + y^2)du^2 - 2\,dudr + dx^2 + dy^2, \tag{3.36}$$

where $f(u)$ is an arbitrary function of u. This spacetime satis-

fies the field equations (3.1) with

$$\rho = 3p^* = 3p + 3\zeta\dot{f}f^{-2}(x^2+y^2)^{-1/2} = 2(x^2+y^2)^{-1}, \quad (3.37)$$

$$\eta = \dot{f}^{-1}f^2(x^2+y^2)^{-1/2}, \quad (3.38)$$

$$u^\mu = [f^{-1}(x^2+y^2)^{-1/2}, 0, 0, 0], \quad (3.39)$$

$$q_\mu = [0, 2f^{-1}(x^2+y^2)^{-3/2}, 0, 0], \quad (3.40)$$

where the dot denotes differentiation with respect to u and the coordinates (u, r, x, y) are labelled (x^0, x^1, x^2, x^3), respectively. From the expression (3.38) we see that $f(u)$ is restricted by the condition $\dot{f} > 0$ for a viable viscous fluid interpretation.

The heat conduction vector is given in terms of the thermal conductivity κ and the temperature T by the expression[15]

$$q_\mu = -\kappa h_\mu{}^\nu(T_{,\nu} + Ta_\nu), \quad (3.41)$$

where a_ν is the acceleration vector defined by $a_\nu = u_{\nu;\alpha}u^\alpha$. Using the expressions (3.39) and (3.40), Eq. (3.41) leads to the two equations

$$\kappa(f^{-1}T)^{\cdot} = 2(x^2+y^2)^{-1/2}, \quad (3.42)$$

$$T = (x^2+y^2)^{-1/2}F(u), \quad (3.43)$$

where $F(u)$ is an arbitrary function of u. In order to proceed further, we need some knowledge of the composition of the viscous fluid. For example, if the viscous fluid is composed of a material medium with very short mean free times, interacting with radiation quanta with a finite mean free time, then κ, η, and T are related by[15]

$$\kappa T = 5\eta. \quad (3.44)$$

In this case Eqs. (3.42) and (3.43) lead to the expressions

$$T = T_0 f^{7/5}(x^2+y^2)^{-1/2}, \quad (3.45)$$

$$\kappa = 5T_0{}^{-1}\dot{f}^{-1}f^{3/5}, \quad (3.46)$$

where T_0 is a positive constant. Note that, since $\dot{f} > 0$, we have $\kappa > 0$, as required.

Thus the spacetime (3.36), with $\dot{f} > 0$, appears to represent a model of a heat-conducting viscous fluid in which the necessary physical requirement of positivity is satisfied by the energy and by the coefficients of viscosity and conductivity.

IV. PERFECT FLUID SOURCES

We now consider the situation in which the electromagnetic field is associated with a perfect fluid, i.e., the field equations (2.1) are replaced by

$$G_{\mu\nu} = E_{\mu\nu} + (\bar{\rho} + \bar{p})v_\mu v_\nu + \bar{p}g_{\mu\nu}, \quad (4.1)$$

and we investigate the case in which a spacetime satisfying these equations also satisfies the field equations (2.3), i.e.,

$$G_{\mu\nu} = (\rho + p^*)u_\mu u_\nu + p^* g_{\mu\nu} - 2\eta\sigma_{\mu\nu}. \quad (4.2)$$

This problem could be generalized by the addition of heat conduction terms on the right-hand side of Eq. (4.1), or of Eq. (4.2),or both. However, such generalizations lead to complications which we shall not consider here.

Combining Eqs. (4.1) and (4.2), we obtain

$$E_{\mu\nu} = (\rho + p^*)u_\mu u_\nu - (\bar{\rho} + \bar{p})v_\mu v_\nu + (p^* - \bar{p})g_{\mu\nu} - 2\eta\sigma_{\mu\nu}, \quad (4.3)$$

and the Rainich conditions (2.5) imply that

$$\rho - 3p^* = \bar{\rho} - 3\bar{p}, \quad (4.4)$$

and

$$\begin{aligned} &-(\rho+p)(\rho - p + 2\bar{p})u_\mu u_\nu - (\bar{\rho}+\bar{p})(\bar{\rho} - \bar{p} + 2p)v_\mu v_\nu \\ &-(\rho+p)(\bar{\rho}+\bar{p})u_\alpha v^\alpha(u_\mu v_\nu + u_\nu v_\mu) \\ &-4(p-\bar{p})\eta\sigma_{\mu\nu} + 2(\bar{\rho}+\bar{p})\eta v_\alpha(v_\mu\sigma^\alpha{}_\nu + v_\nu\sigma^\alpha{}_\mu) \\ &\quad + 4\eta^2\sigma_{\mu\alpha}\sigma^\alpha{}_\nu \\ &= \tfrac{1}{4}g_{\mu\nu}[(\rho+p)(\rho - p + 2\bar{p}) + (\bar{\rho}+\bar{p})(\bar{\rho} - \bar{p} + 2p) \\ &-2(\rho+p)(\bar{\rho}+\bar{p})(u_\alpha v^\alpha)^2 + 4(\bar{\rho}+\bar{p})\eta v_\alpha v_\beta\sigma^{\alpha\beta} + 8\eta^2\sigma^2]. \end{aligned} \quad (4.5)$$

We now introduce the following notation:

$$\rho + p^* = M, \quad \bar{\rho} + \bar{p} = N, \quad u_\mu v^\mu = -\alpha, \quad (4.6)$$

where, from energy consideration, $M > 0$, $N > 0$. In addition, $\alpha \geqslant 1$, the value $\alpha = 1$ implying that $v_\mu \equiv u_\mu$. Using this notation, contracting Eq. (4.5) with $u^\mu u^\nu$, and using the result to simplify Eq. (4.5), we obtain

$$\begin{aligned} &M(M+N)u_\mu u_\nu + N(M+N)v_\mu v_\nu \\ &\quad - 2MN\alpha(u_\mu u_\nu + v_\mu v_\nu) + 2(M-N)\eta\sigma_{\mu\nu} \\ &\quad - 4N\eta v_\alpha(v_\mu\sigma^\alpha{}_\nu + v_\nu\sigma^\alpha{}_\mu) - 8\eta^2\sigma_{\mu\alpha}\sigma^{\alpha\nu} \\ &\quad + g_{\mu\nu}[M(M+N) - N(3M-N)\alpha^2] = 0. \end{aligned} \quad (4.7)$$

We now assume that the timelike 4-vector v_μ in the expression (3.24) is identical with the 4-velocity v_μ in the field equations (4.1), and, from Eqs. (3.24) and (4.3), we obtain

$$\begin{aligned} &Mu_\mu u_\nu - Nv_\mu v_\nu + \tfrac{1}{4}(M-N)g_{\mu\nu} - 2\eta\sigma_{\mu\nu} \\ &\quad = (\tfrac{1}{2}g_{\mu\nu} + v_\mu v_\nu)(E^2 + B^2) - (E_\mu E_\nu + B_\mu B_\nu) \\ &\quad\quad + (v_\mu S_\nu + v_\nu S_\mu). \end{aligned}$$

Contracting this with v^ν leads to

$$\begin{aligned} 2\eta\sigma_{\mu\nu}v^\nu = &-M\alpha u_\mu + [\tfrac{1}{2}(E^2+B^2) + \tfrac{1}{4}(M+3N)]v_\mu \\ &+ S_\mu. \end{aligned} \quad (4.8)$$

On the other hand, contracting Eq. (4.5) with u^ν leads to

$$2\eta\sigma_{\mu\nu}v^\nu = \tfrac{1}{2}(M-N)(\alpha u_\mu - v_\mu), \quad (4.9)$$

so that, from Eq. (4.8) and (4.9), we have

$$S_\mu = \tfrac{1}{2}(3M-N)\alpha u_\mu - [\tfrac{1}{2}(E^2+B^2) + \tfrac{1}{4}(3M+N)]v_\mu. \quad (4.10)$$

Contracting Eq. (4.10) with v^μ leads to

$$E^2 + B^2 = (3M-N)\alpha^2 - \tfrac{1}{2}(3M+N), \quad (4.11)$$

which shows that $3M - N > 0$, and the expression for S_μ becomes

$$S_\mu = \tfrac{1}{2}(3M-N)\alpha(u_\mu - \alpha v_\mu). \quad (4.12)$$

Note that if $\alpha = 1$, i.e., $u_\mu \equiv v_\mu$, then $S_\mu = 0$.

Contracting Eq. (4.7) yields

$$16\eta^2\sigma^2 = 3M^2 + N^2 - 2N(3M-N)\alpha^2,$$

which implies that

$$3M^2 + N^2 > 2N(3M-N)\alpha^2. \quad (4.13)$$

Equation (4.11) implies that $(3M - N)\alpha^2 > \frac{1}{2}(3M + N)$, so that Eq. (4.13) leads to

$$M - N > 0, \tag{4.14}$$

i.e., $\rho + p^* > \bar{\rho} + \bar{p}$. Combining this with Eq. (4.4) yields

$$\rho > \bar{\rho}, \quad p^* > \bar{p}. \tag{4.15}$$

Now it has been shown[16] that if the field equations (4.1) hold with the perfect fluid energy condition and the electromagnetic energy condition each satisfied, then, in the tetrad frame in which the total energy tensor takes its canonical form, viz., diag(μ, p_1, p_2, p_3), the components of v_i have, at most, one spacelike component. In other words, in the tetrad frame in which the energy tensor, and hence the Einstein tensor, is diagonal, v_i must have one of the following forms:

$$v_i = (-1, 0, 0, 0), \tag{4.16}$$

$$v_i = (v_{0'}, v_{1'}, 0, 0), \tag{4.17}$$

with a suitable labelling of coordinates where $v_{1'}$ is nonzero and $v_{0'}{}^2 - v_{1'}{}^2 = 1$. In the first case the electromagnetic field is nonnull and its energy tensor has the tetrad components (2.21), while in the second case its tetrad components are

$$E_{ij} = \begin{pmatrix} \frac{1}{2}(A^2 + B^2) & AB & 0 & 0 \\ AB & \frac{1}{2}(A^2 + B^2) & 0 & 0 \\ 0 & 0 & \frac{1}{2}(A^2 - B^2) & 0 \\ 0 & 0 & 0 & \frac{1}{2}(A^2 - B^2) \end{pmatrix}. \tag{4.18}$$

The electromagnetic field is null if, and only if, $A^2 = B^2$, in which case the form (4.18) becomes identical to the form (3.21).

Contracting Eq. (3.24) with u^ν and using Eqs. (4.11) and (4.12) leads to

$$E_\mu{}^\nu u_\nu = -\tfrac{1}{4}(3M + N)u_\mu + N\alpha v_\mu. \tag{4.19}$$

Similarly, contraction with v^ν gives

$$E_\mu{}^\nu v_\nu = -\tfrac{1}{2}(3M - N)\alpha u_\mu + (3M + N)v_\mu. \tag{4.20}$$

Consider first the case given by Eqs. (4.16) and (2.21). The tetrad components of u_μ are $u_i = (-\alpha, u_{1'}, u_{2'}, u_{3'})$. Substituting in the tetrad form of Eq. (4.19) yields

$$\begin{aligned} X^2\alpha &= \tfrac{1}{4}(3M + N)\alpha - N\alpha, \\ X^2 u_{1'} &= \tfrac{1}{4}(3M + N)u_{1'}, \\ X^2 u_{2'} &= -\tfrac{1}{4}(3M + N)u_{2'}, \\ X^2 u_{3'} &= -\tfrac{1}{4}(3M + N)u_{3'}, \end{aligned} \tag{4.21}$$

and these equations lead to

$$X^2 = \tfrac{3}{4}(M - N), \tag{4.22}$$

and $u_{1'} = u_{2'} = u_{3'} = 0$, which implies that $\alpha = 1$ and $u_\mu \equiv v_\mu$.

Turning now to the case given by Eqs. (4.17) and (4.18), the 2′ and 3′ components of the tetrad form of (4.20) lead to $u_{2'} = u_{3'} = 0$, since $3M - N > 0$. Hence, u_μ has tetrad components of the form $u_i = (u_{0'}, u_{1'}, 0, 0)$, where $u_{0'}{}^2 - u_{1'}{}^2 = 1$, and also

$$u_{0'}v_{0'} - u_{1'}v_{1'} = \alpha. \tag{4.23}$$

The tetrad components of Eq. (4.9) are

$$-2\eta\sigma_{0'0'}v_{0'} + 2\eta\sigma_{0'1'}v_{1'} = \tfrac{1}{2}(M - N)(\alpha u_{0'} - v_{0'}), \tag{4.24}$$

$$-2\eta\sigma_{1'0'}v_{0'} + 2\eta\sigma_{1'1'}v_{1'} = \tfrac{1}{2}(M - N)(\alpha u_{1'} - v_{1'}), \tag{4.25}$$

and the orthogonality relation $\sigma_{\mu\nu}u^\nu = 0$ becomes

$$-\sigma_{0'0'}u_{0'} + \sigma_{0'1'}u_{1'} = 0, \tag{4.26}$$

$$-\sigma_{1'0'}u_{0'} + \sigma_{1'1'}u_{1'} = 0. \tag{4.27}$$

Eliminating $\sigma_{0'0'}$ from Eqs. (4.24) and (4.26), we obtain

$$2\eta\sigma_{0'1'}(v_{1'}u_{0'} - u_{1'}v_{0'}) = \tfrac{1}{2}(M - N)(\alpha u_{0'} - v_{0'})u_{0'}. \tag{4.28}$$

From Eq. (4.23) we find that $\alpha u_{0'} - v_{0'} = u_{1'}(u_{1'}v_{0'} - u_{0'}v_{1'})$, so that Eq. (4.28) yields either

$$u_{1'}v_{0'} - u_{0'}v_{1'} = 0, \tag{4.29}$$

which implies that $u_\mu \equiv v_\mu$, or

$$2\eta\sigma_{0'1'} = -\tfrac{1}{2}(M - N)u_{0'}u_{1'}. \tag{4.30}$$

Now, since the tetrad components of $G_{\mu\nu}$ in Eq. (4.1) are diagonal, it follows that in Eq. (4.2) they are also diagonal, so that the (0′, 1′) tetrad component of these field equations gives

$$2\eta\sigma_{0'1'} = Mu_{0'}u_{1'}. \tag{4.31}$$

Assuming that Eq. (4.30) holds, then, from Eqs. (4.30) and (4.31), we obtain $(3M - N)u_{0'}u_{1'} = 0$, which implies that $u_{1'} = 0$, so that $u_{0'} = -1$ and $v_{0'} = -\alpha$. In this case we find that $\sigma_{0'0'} = \sigma_{0'1'} = 0$ and

$$2\eta\sigma_{1'1'} = -\tfrac{1}{2}(M - N). \tag{4.32}$$

Hence, if the field equations (4.1) and (4.2) are to be satisfied by the same spacetime, then, in the case in which the local canonical forms of the electromagnetic energy tensor and the velocity 4-vector are given by Eqs. (2.21) and (4.16), we have $u_\mu \equiv v_\mu$ necessarily. In the case in which the local canonical forms are given by (4.18) and (4.17), then either $u_\mu \equiv v_\mu$ or u_μ is locally comoving, i.e., $u_i = (-1, 0, 0, 0)$. These are the only possibilities.

Given a spacetime satisfying the field equations (4.1) for an electromagnetic field in the presence of a prefect fluid, there are, essentially, four possible conclusions resulting from an investigation into the possibility of the same spacetime satisfying the field equations (4.2) for a viscous fluid, namely:

(a) The spacetime satisfies Eq. (4.2) with the 4-velocity u_μ identical to v_μ of Eq. (4.1).

(b) The spacetime satisfies Eq. (4.2) with the 4-velocity u_μ being locally comoving, but v_μ is locally tilting in one spacelike tetrad direction.

(c) The spacetime does not satisfy Eq. (4.2) because the equations imply the unphysical situation $\eta < 0$.

(d) The spacetime does not satisfy Eq. (4.2) because the components of the shear tensor $\sigma_{\mu\nu}$ given by Eq. (4.2) do not

agree with the components of $\sigma_{\mu\nu}$ calculated from the covariant derivative of u_μ.

We shall now give examples illustrating each of these four possibilities.

(1) Some members of the class of type VI_0 cosmological models discussed by Dunn and Tupper[17] satisfy both field equations. The metric is of the form

$$ds^2 = -dt^2 + (m-n)^{-2}t^2\,dx^2 + t^{-2(m+n)} \times (e^{-2x}\,dy^2 + e^{2x}\,dz^2), \tag{4.33}$$

where m and n are arbitrary constant parameters, and is a generalization of the electrovac solution (2.11). This spacetime represents a magnetohydrodynamic model; the 4-current J^μ is spacelike and the conductivity is finite.[18] The metric satisfies Eq. (4.1), together with the electromagnetic energy condition and the perfect fluid energy condition, for a certain set of values of m and n. These values are illustrated by a diagram in Ref. 18. It has been shown[19] that the metric (4.33) also satisfies the viscous fluid equations (4.2) for certain values of m and n and these values, for various equations of state, are illustrated in Ref. 19. It will be noticed that there are values of m and n which satisfy both sets of equations. Rather than repeat the diagrams of Refs. 18 and 19, we demonstrate the dual interpretation by choosing particular values for m and n, namely $m = -\frac{1}{2}$, $n = \frac{1}{8}$, so that the metric (4.33) becomes

$$ds^2 = -dt^2 + \tfrac{64}{25}t^2\,dx^2 + t^{3/4}(e^{-2x}\,dy^2 + e^{2x}\,dz^2). \tag{4.34}$$

As a solution of Eq. (4.1), this metric leads to

$$\bar{\rho} = \tfrac{11}{32}t^{-2}, \quad \bar{p} = \tfrac{3}{32}t^{-2}, \tag{4.35}$$

$$F_{01} = \tfrac{8}{5}(\tfrac{5}{16} - k^2 t^{1/2})^{1/2}, \quad F_{23} = k, \tag{4.36}$$

$$J^1 = \tfrac{25}{512}(\tfrac{5}{16} - k^2 t^{1/2})^{-1/2}t^{-3}, \tag{4.37}$$

$$\lambda = \tfrac{5}{64}(\tfrac{5}{16} - k^2 t)^{-1}t^{-1}, \tag{4.38}$$

where λ is the conductivity and k is an arbitrary constant. The components of $E_\mu{}^\nu$ are $E_0{}^0 = E_1{}^1 = -E_2{}^2 = -E_3{}^3 = \frac{5}{32}t^{-2}$, and $v^\mu = (1,0,0,0)$. This implies that, in the viscous fluid interpretation, $u^\mu = (1, 0, 0, 0)$, and we find that the metric (4.34) satisfies the field equations (4.2) with

$$\rho = \tfrac{1}{2}t^{-2}, \quad p^* = p - \tfrac{7}{45}\zeta t^{-1} = \tfrac{7}{48}t^{-2}, \tag{4.39}$$

$$\eta = \tfrac{1}{4}t^{-1}, \tag{4.40}$$

$$\sigma_\mu{}^\nu = \operatorname{diag}\left[0, \tfrac{5}{12}t^{-1}, -\tfrac{5}{24}t^{-1}, -\tfrac{5}{24}t^{-1}\right]. \tag{4.41}$$

One feature of the solution given by Eqs. (4.35)–(4.38) is that the spacetime geometry has no singularities apart from the initial singularity at $t = 0$, but, provided that the magnetic field is nonzero, i.e., $k \neq 0$, the electric field vanishes at time t_0 given by

$$t_0 = \tfrac{25}{256}k^{-4}, \tag{4.42}$$

and thereafter becomes imaginary, and the 4-current and conductivity become infinite at that time. The singular point arises from the Maxwell equations and not from the field equations (4.1), and has the effect of imposing a finite time span on the model on physical grounds, whereas the geometry allows an infinite time span. However, in the alternative viscous fluid interpretation given by Eqs. (4.39)–(4.41), there is no such physical shortening of the time span. All physical and kinematical quantities have infinite values at $t = 0$ and proceed smoothly to zero as $t \to \infty$.

(2) The class of type II cosmological models with metric[16]

$$ds^2 = -dt^2 + k^{-2}t^{4/3}\,dx^2 + t^{2/3}(dy + x\,dz)^2 + t^{4/3}\,dz^2, \tag{4.43}$$

where k is an arbitrary constant parameter, satisfies the field equations (4.1), the electromagnetic field being null. If, as in Ref. 16, the coordinates are labelled $(t, x, y, z) = (x^0, x^1, x^2, x^3)$, then the tetrad components of v_μ are $(v_{0'}, 0, v_{2'}, 0)$, necessitating a labelling change in the previous discussion. The canonical form of the electromagnetic energy tensor corresponds to the form (3.19) with this labelling change, where $A^2 = (\frac{20}{9} - k^2)(k^2 - \frac{2}{9})(\frac{16}{9} + k^2)^{-1}t^{-2} = h^2 t^{-2}$. The Maxwell tensor components are

$$\begin{aligned} &F_{01} = -hk^{-1}t^{-1/3}\sin\theta, \quad F_{23} = hx\cos\theta, \\ &F_{02} = 0, \quad F_{31} = -hk^{-1}x\sin\theta, \\ &F_{03} = ht^{-1/3}\cos\theta, \quad F_{12} = hk^{-1}\sin\theta, \end{aligned} \tag{4.44}$$

where θ is a constant, and the 4-current $J^\mu = 0$, so that the model represents an uncharged fluid in the presence of a null electromagnetic field. The pressure and density of the fluid are given by

$$\bar{\rho} = 5\bar{p} = \tfrac{5}{4}(\tfrac{8}{9} - k^2)t^{-2}. \tag{4.45}$$

The dominant energy condition and other necessary conditions are satisfied if, and only if,

$$\tfrac{2}{9} < k^2 < \tfrac{8}{9}. \tag{4.46}$$

The components of the 4-velocity are given by

$$v_\mu = (-m, 0, nt^{1/3}, xnt^{1/3}), \tag{4.47}$$

where m and n are positive constants satisfying $m^2 = n^2 + 1$ and given by

$$\begin{aligned} m &= (20 - 9k^2)[3(16 + 9k^2)(8 - 9k^2)]^{-1/2}, \\ n &= 2(9k^2 - 2)[3(16 + 9k^2)(8 - 9k^2)]^{-1/2}. \end{aligned} \tag{4.48}$$

This 4-velocity field has nonzero expansion, shear, and twist.

For the viscous fluid interpretation, u_μ must be either identical to v_μ or given by $u_\mu = (-1, 0, 0, 0)$. Using the form (4.47) for u_μ does not give a viable viscous model since it leads to $\eta < 0$. However, the comoving form does lead to a viable model with

$$\rho = \tfrac{1}{4}(\tfrac{32}{9} - k^2)t^{-2}, \quad p^* = p - \tfrac{5}{3}\zeta t^{-1} = \tfrac{1}{12}(k^2 + \tfrac{16}{9})t^{-2}, \tag{4.49}$$

$$\eta = \tfrac{3}{2}(k^2 - \tfrac{2}{9})t^{-1}. \tag{4.50}$$

In this case the 4-velocity u_μ has associated expansion and shear, but no twist. From Eq. (4.50), this solution requires $k^2 > \frac{2}{9}$, as previously, but k^2 can range beyond the limit of $\frac{8}{9}$ required by the interval (4.46), so that the class of viscous fluid solutions is larger than the class of electromagnetic field plus perfect fluid solutions. From Eq. (4.49), it follows that p is always positive, and the condition $\rho \geqslant p$ leads to

$$\tfrac{2}{9} < k^2 < -5\zeta t + \tfrac{20}{9}, \tag{4.51}$$

so that $\zeta < \frac{2}{5}t^{-1}$. If we take $\zeta = 0$, so that there is no bulk viscosity, then $p = \frac{1}{12}(k^2 + \frac{16}{9})t^{-2}$ and

$$\tfrac{2}{9} < k^2 < \tfrac{20}{9}. \tag{4.52}$$

For the range of values given by (4.46), we find that

$$\tfrac{5}{6}t^{-2} > \rho > \tfrac{2}{3}t^{-2}, \quad \tfrac{1}{6}t^{-2} < p^* < \tfrac{2}{9}t^{-2}, \tag{4.53}$$

$$0 < \eta < t^{-1}.$$

(3) The type I cosmological model with metric[16]

$$ds^2 = -dt^2 + t^{2a}dx^2 + t^{2b}(dy^2 + dz^2), \tag{4.54}$$

satisfies the field equations (4.1) for a nonnull electromagnetic field, with $v_\mu = (-1, 0, 0, 0)$, provided that the parameters a and b satisfy $a - b < 0$, $a - 2b + 1 < 0$, $b \leqslant \frac{1}{2}$. The viscous fluid field equations are not satisfied since they lead to $\eta < 0$.

The metric (4.54) also satisfies Eq. (4.1) for a null electromagnetic field. In this case $a + b = 1$ and the 4-velocity $v_\mu = (8a)^{-1/2}[-(2a+1), (2a-1)t^a, 0, 0]$. Whether we choose u_μ to be identical to v_μ, or to be comoving, the viscous fluid interpretation leads to $\eta < 0$.

(4) The solution found by Wainwright and Yaremovich[20], namely

$$ds^2 = -(dt + az\,dy)^2 + dx^2 + dy^2 + dz^2, \tag{4.55}$$

satisfies Eqs. (4.1) for a nonnull electromagnetic field with $E_0{}^0 = E_1{}^1 = -E_2{}^2 = -E_3{}^3 = -\frac{1}{4}a^2$. The density, pressure, and 4-velocity are $\rho = \frac{1}{2}a^2$, $p = 0$, and $v_\mu = (-1, 0, 0, 0)$. For the viscous fluid interpretation we must have $u_\mu = (-1, 0, 0, 0)$, but this leads to $\sigma_{\mu\nu} = 0$, so the model cannot represent a viscous fluid.

V. COMMENTS

In looking for examples of this dual interpretation we have concentrated our attention on the task of determining whether or not known Einstein–Maxwell solutions satisfy the viscous fluid field equations, rather than determining whether or not known viscous fluid solutions satisfy the Einstein–Maxwell equations. The reason for this bias is the abundance of known Einstein–Maxwell solutions and the paucity of known viscous fluid solutions. However, it should be noted that, given a known viscous fluid solution, if it is found to satisfy also the field equations (2.1) or (4.1), we must determine whether, in addition, it satisfies the Rainich–Misner–Wheeler[21] conditions, i.e., the vanishing of the curl of the complexion vector, before it can be declared a viable Einstein–Maxwell solution. This will impose certain conditions on the various quantities appearing in the viscous fluid energy tensor.

One question that arises from this work is: Given a spacetime that satisfies both sets of field equations, what is the "correct" interpretation, an electromagnetic field or a viscous fluid? For example, does the Kerr–Newman solution really represent the field of a spinning charged particle, or does it represent a viscous fluid distribution, or can it represent both of these? An attempt will be made to answer this question in a future article.

ACKNOWLEDGMENTS

I wish to express my gratitude to the many people who have shown interest in this problem through comments, discussions, and correspondence, with particular thanks to W. B. Bonnor, C. B. Collins, K. A. Dunn, E. C. Ihrig, M. A. H. MacCallum, J. Safko, A. H. Taub, J. Wainwright, and L. Witten. Part of the work described here was presented at the GR8 Conference in Waterloo, 1977. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada through an operating grant to the author.

[1] J. L. Synge, in *Perspectives in Geometry and Relativity*, edited by B. Hoffmann (Indiana U. P., Bloomington, 1966).
[2] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[3] J. B. Griffiths, Comm. Math. Phys. **28**, 295 (1972).
[4] N. Tariq and B. O. J. Tupper, Gen. Rel. Grav. **6**, 345 (1975).
[5] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[6] B. Bertotti, Phys. Rev. **116**, 1331 (1969).
[7] I. Robinson, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astr. Phys. **7**, 351 (1959).
[8] A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. (to appear).
[9] W. B. Bonnor, Proc. Phys. Soc. London Sect. A **66**, 145 (1953).
[10] A. K. Raychaudhuri, Ann. Phys. (N.Y.)**11**, 501 (1960).
[11] R. G. McLenaghan and N. Tariq, J. Math. Phys. **16**, 2306 (1975).
[12] B. O. J. Tupper, Gen. Rel. Grav. **7**, 479 (1976).
[13] A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
[14] R. G. McLenaghan, N. Tariq, and B. O. J. Tupper, J. Math. Phys. **16**, 829 (1975).
[15] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[16] K. A. Dunn and B. O. J. Tupper, Astrophys. J. **235**, 307 (1980).
[17] K. A. Dunn and B. O. J. Tupper, Astrophys. J. **204**, 322 (1976).
[18] B. O. J. Tupper, Astrophys. J. **216**, 192 (1977).
[19] K. A. Dunn and B. O. J. Tupper, Astrophys. J. **222**, 405 (1978).
[20] J. Wainwright and P. E. A. Yaremovich, Gen. Rel. Grav. **7**, 345 (1976).
[21] C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) **2**, 525 (1957).

A generalization of the concept of spacetime

Carl Schaer
Exxon Production Research Company, P.O. Box 2189, Houston, Texas 77001

(Received 29 April 1981; accepted for publication 12 June 1981)

It is argued that a physically relevant generalization of spacetime should maintain the line element of general relativity. A limited number of spaces can be constructed which satisfy this condition. In addition to the symmetric metric field, three antisymmetric fields naturally arise, and it is shown that, if one of these is interpreted as the electromagnetic field, the other two fields have equal but opposite charges.

PACS numbers: 04.50. + h

I. INTRODUCTION

Since general relativity is one of the most successful theories of modern physics, it is natural to ask how it could be generalized to include interactions other than gravity. This means that fields other than a symmetric metric field $g_{\mu\nu}$ must be included in the theory. One attempt by Einstein to do this was to include an antisymmetric part in $g_{\mu\nu}$.[1]

$$g_{\mu\nu} = g_{(\mu\nu)} + g_{[\mu\nu]}. \tag{1.1}$$

The antisymmetric part was then related to the electromagnetic field. The line element, though, is still given by the symmetric part.

$$ds^2 = dx^\mu g_{\mu\nu} dx^\nu = dx^\mu g_{(\mu\nu)} dx^\nu. \tag{1.2}$$

General relativity is a theory based on a space parameterized by four real coordinates. Another way of generalizing general relativity is to generalize the space on which the theory is based. Kaluza and Klein developed such a theory on a five-dimensional space.[2,3] The components $g_{\mu 4}$ of the five-dimensional metric allowed them to include fields describing the four potential of electromagnetism. However, this poses an immediate difficulty. Four of the parameters can be associated with time and the three spatial dimensions, but the fifth parameter has no known direct physical interpretation. One has to either given an interpretation to this parameter or argue that it is unobservable. An interesting approach to this problem was proposed by Chodos and Detweiler.[4] Using the Kasner metric in five dimensions they argue that one of the dimensions could have contracted to the order of 10^{-31} cm, while the other four dimensions would be expanding.

A more recent example of the extension of spacetime is the superspaces of supergravity. These spaces are parametrized by the coordinates

$$(z^\mu, z^i), \quad \mu = 0,1,2,3, \tag{1.3}$$

where the z^μ are even elements of a Grassman algebra, and the z^i are odd elements. The line element can be expanded as[5]

$$ds^2 = dz^\mu g_{\mu\nu} dz^\nu + 2dz^\alpha g_{\alpha i} dz^i + dz^i g_{ij} dz^j. \tag{1.4}$$

Again, the last two terms have no known physical interpretation. The first term also cannot be considered as the line element of physical spacetime, since it is in general Grassman valued. Only the real part of this term has a direct physical significance.

Once the decision has been made to extend the mathematical space on which general relativity is formulated, we are faced with an infinite number of possibilities, two of which have been mentioned above. In order to select the appropriate space, or at least to restrict the number of possibilities, some physical principle based on our observations of physical reality should be found. In this paper we will propose such a principle and exhibit spaces which satisfy it.

One of the most fundamental observations of reality is that spacetime is four dimensional. This means that the relationship between a collection of events can be described by assigning four numbers to each event, three for the spatial dimensions, and one for time. So far no event has been observed which would require more than four parameters to describe its relationship to neighboring events. The exact assignment of parameters is somewhat arbitrary, but the infinitesimal line element of general relativity

$$ds^2 = dx^\mu g_{\mu\nu} dx^\nu, \quad \mu,\nu = 0,1,2,3 \tag{1.5}$$

between two neighboring events is an invariant quantity. It is proposed that, since this line element is a fundamental invariant of physics, any physically relevant generalization of spacetime should also leave this line element invariant. In other words, all terms in the new line element involving any extra parameters should drop out. The alternative would be to give a physical interpretation to the new terms appearing in the line element.

II. ALGEBRAIC CONDITION ON THE SPACE

As was shown in the example above, adding more coordinates to the four real coordinates describing spacetime leads to terms in the line element which have no known physical interpretation. Therefore, in order to circumvent this problem, instead of directly adding more coordinates to spacetime, we will generalize the algebra from which the coordinates take their values from real numbers to hypercomplex numbers.[6] The space is then parameterized as follows

$$z^\mu = (z^0, z^1, z^2, z^3). \tag{2.1}$$

The question which faces us now is, can we find some algebra of hypercomplex numbers for which the line element of the space parameterized by (2.1) reduces to the line element of general relativity?

A hypercomplex number can be expanded as a linear combination of hypercomplex units e_I.

$$z = x^I e_I, \quad I = 0,\ldots,n. \tag{2.2a}$$

 0022-2488/81/112674-04$01.00

Here the x^I are real, and e_0 denotes the identity. It will later be convenient to separate the real term and write

$$z = x + y^i e_i, \quad i = 1,...,n. \tag{2.2b}$$

The algebra is specified by stating the multiplication rules for the e_I.

$$e_I e_J = r_{IJK} e_K. \tag{2.3}$$

It will also be assumed that this multiplication is associative.

The line element of spacetime is real valued. In order to obtain real numbers from the product of hypercomplex numbers, one introduces the concept of conjugates. A conjugate of a hypercomplex number can be obtained as follows.

Each element z of the algebra (2.2) can be associated with an $(n+1)\times(n+1)$ matrix M_{IJ}.

$$ze_I = x^K r_{KIJ} e_J = M_{JI} e_J. \tag{2.4}$$

Because of the associativity of hypercomplex multiplication this correspondence is an isomorphism. Now, the Cayley-Hamilton theorem says that every matrix satisfies its own characteristic equation. In other words, if

$$|M_{IJ} - \lambda I| = 0 \tag{2.5}$$

or

$$\lambda^{n+1} + c_n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c = 0 \tag{2.6}$$

is the characteristic equation for a matrix M, then substituting M for the real parameter λ gives an identity.

$$M^{n+1} + c_n M^n + c_{n-1} M^{n-1} + \cdots + c = 0. \tag{2.7}$$

Since the set of matrices M is isomorphic to the hypercomplex numbers z, we also have

$$z^{n+1} + c_n z^n + c_{n-1} z^{n-1} + \cdots + c = 0. \tag{2.8}$$

Thus

$$z(-z^n - c_n z^{n-1} - \cdots - c_1) = c = \text{real} \tag{2.9}$$

and we can define a conjugate $\bar{z}$

$$\bar{z} = -z^n - c_n z^{n-1} - \cdots - c_1 \tag{2.10}$$

such that

$$z\bar{z} = \text{real}. \tag{2.11}$$

It often happens that the characteristic polynomial (2.8) can be factored into a polynomial of lower degree. This polynomial can then be used to define a conjugate $\bar{z}$ using a lower power of z.

This conjugate is more general than the conjugates of complex and quaternionic numbers. In particular, the conjugate defined by (2.10) is nonlinear in general. For the following, though, we will need the usual properties of conjugates listed below.

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \tag{2.12}$$

$$\overline{z_1 z_2} = \bar{z}_2 \bar{z}_1, \tag{2.13}$$

$$\bar{\bar{z}} = z, \tag{2.14}$$

$$z = \bar{z} \Leftrightarrow \text{is real}. \tag{2.15}$$

We are now in a position to derive conditions on the metric and the hypercomplex algebra such that the line element of the space (z^0,z^1,z^2,z^3) reduces to the line element of general relativity. The metric $G_{\mu\nu}$ on this space will also be hypercomplex valued.

$$G_{\mu\nu} = g_{\mu\nu} + g^i_{\mu\nu} e_i. \tag{2.16}$$

The line element will then be

$$ds^2 = \overline{dz^\mu} G_{\mu\nu} dz^\nu. \tag{2.17}$$

We will first show that, for ds^2 to be real, $G_{\mu\nu}$ must satisfy

$$G_{\mu\nu} = \overline{G_{\nu\mu}}. \tag{2.18}$$

To this end we take the conjugate of ds^2.

$$\overline{ds^2} = \overline{\overline{dz^\mu} G_{\mu\nu} dz^\nu}$$
$$= \overline{dz^\nu}\ \overline{G_{\mu\nu}}\, dz^\mu$$

or

$$\overline{ds^2} = \overline{dz^\mu}\ \overline{G_{\nu\mu}}\, dz^\nu. \tag{2.19}$$

Then, for an arbitrary dz^μ,ds^2 will be equal to its conjugate, and thus real by (2.15), if and only if (2.18) is true.

We next derive a condition on the algebra such that, not only is ds^2 real, but it also reduces to the line element of general relativity

$$ds^2 = dx^\mu g_{\mu\nu} dx^\nu, \tag{2.20}$$

where dx^μ is the real part of dz^μ. In particular, if we choose

$$G_{\mu\nu} = \eta_{\mu\nu} \tag{2.21}$$

we must have

$$\overline{dz^\mu} dz_\mu = dx^\mu dx_\mu. \tag{2.22}$$

This means that the algebra must satisfy

$$\bar{z}z = x^2. \tag{2.23}$$

Since we have choosen e_0 to be the identity, the matrix M associated with z has the form

$$M = \begin{pmatrix} x & & & & \\ & x+\_\_ & & & \\ & & x+\_\_ & \_\_\_\_ & \\ & & & \ddots & \\ \_\_\_\_ & & & & x+\_\_ \end{pmatrix}, \tag{2.24}$$

where the blanks denote arbitrary numbers. The point is that x appears only in the diagonal terms. According to (2.9), c, the constant term in the secular equation, is a power of the norm of z, which by (2.23) is x^2. On the other hand, c is the determinant of the matrix M so we have

$$|M| = x^{n+1}. \tag{2.25}$$

Since this is true for all z, it is true for all x. In particular, the determinant $|M - \lambda I|$ is just the determinant of a matrix M' obtained from M by replacing x by $(x - \lambda)$. This means that the secular equation is

$$|M - \lambda I| = (x - \lambda)^{n+1} = 0. \tag{2.26}$$

z itself satisfies the secular equation, so we have

$$(x - z)^{n+1} = 0. \tag{2.27}$$

As shown in (2.9) and (2.10), this equation can be used to define a conjugate. However, in taking the conjugate of ds^2, we required that the conjugate be linear. Therefore, we must require not only that $(x-z)^{n+1}$ vanishes, but also that

$$(x-z)^2=0. \tag{2.28}$$

Expanding this,

$$z(2x-z)=x^2 \tag{2.29}$$

and the conjugate is

$$\bar{z}=2x-z=x-y^i e_i. \tag{2.30}$$

It can easily be verified that this conjugate satisfies the conditions (2.13)–(2.15).

Equation (2.28) also implies that

$$y^i y^j e_i e_j=0. \tag{2.31}$$

Since $y^i y^j$ is symmetric in i and j, $e_i e_j$ must be antisymmetric for (2.31) to hold.

$$e_i e_j=-e_j e_i. \tag{2.32}$$

This is a necessary condition on the hypercomplex algebra. In particular, this implies that the square of any hypercomplex unit is zero.

$$e_i e_i=0. \tag{2.33}$$

We now show that the conditions (2.18) on the metric and (2.32) on the algebra are not only necessary conditions, but are also sufficient conditions for the line element (2.17) to reduce to the line element of general relativity. First, we show that the product $e_i e_j$ cannot be a nonzero real number, for if

$$e_i e_j=r=\text{real} \tag{2.34}$$

multiplying on the left by e_i shows that

$$0=e_i e_i e_j=e_i r, \tag{2.35}$$

which implies that r is zero. Since the product of the hypercomplex units cannot give a nonzero real number, then the real part of a product of hypercomplex numbers satisfying (2.32) must be the product of the real parts of each hypercomplex number. This means that

$$\begin{aligned} ds^2 &= \overline{dz^\mu}\, G_{\mu\nu} dz^\nu \\ &= \mathrm{Re}(\overline{dz^\mu}\, G_{\mu\nu} dz^\nu) \quad (\text{since } ds^2 \text{ is real}) \\ &= (\mathrm{Re}\ dz^\mu)(\mathrm{Re} G_{\mu\nu})(\mathrm{Re} dz^\nu) \end{aligned} \tag{2.36}$$

or

$$ds^2=dx^\mu g_{\mu\nu} dx^\nu, \tag{2.37}$$

which was what we had required.

III. SPECIFIC ALGEBRAS

The condition (2.32) that all hypercomplex units e_i anticommute immediately suggests Grassman algebras. A Grassman algebra is generated by the units $\theta_1,\ldots,\theta_n$, with the property

$$\theta_i\theta_j=-\theta_j\theta_i=\theta_{ij}, \tag{3.1}$$

where θ_{ij} is a unit distinct from the n generators θ_k. This multiplication can be continued until units with n indices are formed.

$$\begin{aligned} &\theta_i\theta_j\theta_k=\theta_{ijk}, \\ &\theta_i\theta_j\theta_k\theta_l=\theta_{ijkl}, \\ &\theta_1\theta_2\cdots\theta_n=\theta_{12\cdots n}. \end{aligned} \tag{3.2}$$

The units with an even number of indices in this general Grassman algebra commute with the other units, though, and therefore do not satisfy the anticommutivity condition. The largest Grassman algebra that does satisfy (2.32) is thus

$$z=x+y^1\theta_1+y^2\theta_2+y^{12}\theta_{12}. \tag{3.3}$$

In this case,

$$\theta_1\theta_{12}=\theta_2\theta_{12}=0 \tag{3.4}$$

and θ_{12}, having trivial multiplication laws, still satisfies (2.32). The problem with Grassman algebras with more than two generators θ_i is fundamentally that a linear conjugate cannot be formed. A more general algebra satisfying the anticommutivity requirement is

$$z=x+y^{i1}g_i\theta_1+y^{i2}g_i\theta_2+y^{i,12}g_i\theta_{12}, \tag{3.5}$$

where the g_i are hypercomplex units satisfying

$$g_i g_j=g_j g_i. \tag{3.6}$$

This can also be written as

$$z=x+z^1\theta_1+z^2\theta_2+z^{12}\theta_{12}, \tag{3.7}$$

where z^1, z^2, z^{12} are elements of a commuting hypercomplex algebra. In particular, they could be complex numbers. However, x must remain real to insure that the line element is determined by four real displacements dx^μ. It seems unnatural, though, to generalize the y^i of (3.3) to a larger algebra z^i while keeping x real. Therefore, we propose to use the Grassman algebra (3.3) with two generators to generalize the coordinates of spacetime.

IV. SPECULATIONS ON A PHYSICAL INTERPRETATION

The metric on our space is now

$$G_{\mu\nu}=g_{\mu\nu}+g^i_{\mu\nu}\theta_i+g^{12}_{\mu\nu}\theta_{12}, \quad i=1,2. \tag{4.1}$$

According to the second section, this metric must satisfy

$$G_{\mu\nu}=\overline{G_{\nu\mu}}, \tag{4.2}$$

which, using (2.30), can be written as

$$g_{\mu\nu}+g^i_{\mu\nu}\theta_i+g^{12}_{\mu\nu}\theta_{12}=g_{\nu\mu}-g^i_{\nu\mu}\theta_i-g^{12}_{\nu\mu}\theta_{12}. \tag{4.3}$$

This means that $g_{\mu\nu}$ is symmetric, while the other fields are antisymmetric. Since the line element is

$$ds^2=dx^\mu g_{\mu\nu} dx^\nu. \tag{4.4}$$

$g_{\mu\nu}$ can be interpreted as the metric of spacetime.

Since the other fields are antisymmetric, we postulate that they may be field strengths, such as the electromagnetic field. We will try to interpret $g^{12}_{\mu\nu}$ as the electromagnetic field. Under the transformation

$$\theta_1\to\theta_2, \tag{4.5a}$$

$$\theta_2\to\theta_1, \tag{4.5b}$$

θ_{12} transforms as

$$\theta_{12}\to-\theta_{12}. \tag{4.5c}$$

Since θ_1 and θ_2 are arbitrary labels for the units, any theory

based on this space should be invariant under this transformation. This means that, if $(g^{12}_{\mu\nu}, g^{1}_{\mu\nu}, g^{2}_{\mu\nu})$ satisfy the field equations of the theory, $(-g^{12}_{\mu\nu}, g^{2}_{\mu\nu}, g^{1}_{\mu\nu})$ must also satisfy the equations. Since $g^{12}_{\mu\nu}$ has changed signs, if $g^{12}_{\mu\nu}$ is interpreted as the electromagnetic field, the two solutions must describe physical situations with opposite charges. Therefore, $g^{1}_{\mu\nu}$ and $g^{2}_{\mu\nu}$, being interchanged in the two solutions, have equal but opposite charges.

This raises the possibility that $g^{1}_{\mu\nu}$ and $g^{2}_{\mu\nu}$ could describe the charged weak interactions. The neutral weak interaction is missing, though. This could possibly arise as some combination of the three fields. Another possibility would be to use the algebra

$$z = x + y^{i}\theta_{i} + y^{12}\theta_{12} + y^{3}\phi, \quad i = 1,2, \tag{4.6}$$

where ϕ has trivial multiplication laws with the θ^{i}. The metric would be

$$G_{\mu\nu} = g_{\mu\nu} + g^{i}_{\mu\nu}\theta_{i} + g^{12}_{\mu\nu}\theta_{12} + g^{3}_{\mu\nu}\phi. \tag{4.7}$$

When θ^{1} and θ^{2} are interchanged, ϕ transforms into itself, and $g^{3}_{\mu\nu}$ must be neutral. Since ϕ has trivial multiplication laws, though, this algebra does not seem very attractive.

We thus have enough fields to describe the gravitational, electromagnetic, and charged weak interactions, but we do not have fields representing the nuclear interaction or particles. Since the size of the Grassman algebra is restricted by the condition we imposed, the number of fields is also restricted. The elementary particles would therefore have to arise as solitonlike solutions to field equations of the theory. The strong interaction would then have to be an effective short distance strong interaction between the solitons.

ACKNOWLEDGMENTS

I would like to thank Ralph McNutt, Jr., for his encouragement of this work. I am also indebted to Freydoon Mansouri, Alan Kostolecky, and Anthony Longhitano for helpful discussions.

[1]A. Einstein, *The Meaning of Relativity* (Princeton U. P., Princeton, N.J., 1950).
[2]Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. LIV 966 (1921).
[3]O. Klein, Z. Phys. **37**, 875 (1926).
[4]A. Chodos and S. Detweiler, Phys. Rev. D **21**, 2167 (1980).
[5]P. Nath and R. Arnowitt, Phys. Lett. B **56**, 177 (1975).
[6]The exposition of hypercomplex numbers given here follows closely that of R. Feuter in "Funktionen Theorie in Hyperkomplexen," a series of lectures given at the University of Zurich (1948/49) and transcribed by E. Bareiss.

Transformation modified singular spherical integrals for an arbitrarily time-dependent charged particle in a crystal

Lim Chee-Seng [a)]
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México 20, D. F. México

(Received 18 February 1981; accepted for publication 8 May 1981)

A charged particle activates inside a crystalline medium and, thereafter, acts with an arbitrary time-dependent amplitude factor. Both electric and magnetic anisotropies coexist. Corresponding sets of permittivity and permeability principal axes are unaligned. A scalar field is subsequently formulated as the difference between two quadruple Fourier integrals. Each of these is first subjected to a separate transformation of the wavevector variable. It is then reduced to a singular spherical integral. This leads to the establishment of (i) a deactivation by-product of a dark core, (ii) induction of in-phase modes. Other, more general, singular spherical integrals are next deduced. Applications are then made to the compounded uniaxial medium; results are physically and geometrically interpreted.

PACS numbers: 41.10.Hv

1. INTRODUCTION

Within a crystalline medium, the electric and magnetic fields **E** and **H** associated with a current of density **J** and a charge of density ρ are governed by

$$\nabla\times\mathbf{E}+\mathbf{H},\mu=0, \tag{1.1}$$

$$\nabla\times\mathbf{H}-\mathbf{E},\epsilon=\mathbf{J}, \tag{1.2}$$

$$\nabla\cdot(\mathbf{E}\,\epsilon)=\rho, \tag{1.3}$$

relative to the $\mathbf{x}=(x_1,x_2,x_3)$ frame, say. Here ϵ and μ are the constant permittivity and permeability matrices, these being real, symmetric and positive definite. It is assumed throughout that the medium is anisotropic with regard to both ϵ and μ. Furthermore, corresponding sets of principal axes are generally unaligned.

A recent paper[1] deals with the construction of some spherical integral results for a suddenly *activated* line source current conducted at speeds comparable to the wave speeds of the crystal. Following *activation*, the source current is *time-independent* between its entry point and its forward tip. The spherical integrals do not incorporate any transformation. This omission makes them somewhat intractable under certain circumstances. For this reason, the formulation leading to these spherical integrals, rather than their actual final forms, were employed, instead, in a companion paper[2] dealing with a *compounded uniaxial* case.

In the present paper, we consider the problem for an *arbitrarily time-dependent* charged particle:

$$\rho=\delta(\mathbf{x})\mathscr{A}(t) \tag{1.4}$$

that activates at instant $t=0$:

$$\mathscr{A}(t)=\mathscr{B}(t)\,H(t). \tag{1.5}$$

Here $\delta(\mathbf{x})$ and $H(t)$ are, respectively, the Dirac delta function of **x** and Heaviside unit function of t. A general solution will be established in terms of spherical integrals again. However, each spherical integral will incorporate a separate linear transformation. Actually, to achieve the desired effect, such a transformation should not normally be directly applied to the particular spherical integral, because it can be partially deformative. It should be applied instead to a volumetric integral over a wavevector space. In this way, a four-dimensional Fourier integral over an unbounded complex frequency-wavevector product space becomes linearly transformed. It can then be systematically reduced to a spherical integral which is two-dimensional in character and is generally singular. To enhance its applicability, such an integral may be converted into other singular spherical integrals via differential operations together with **x**-dependent orthogonal transformations.

Derivation of singular spherical integrals accomplishes only a partial objective in this paper. The remaining objectives are: (i) to extract from them physical information such as that pertaining to charge deactivation, a dark interior and in-phase modes; (ii) to study, in closer detail, their physical and geometric features for the compounded uniaxial case through closed-form evaluations as well as further reductions.

Other integral forms, viz. over the Minkowski four-space, have been secured by Johannsen[3] (and Handelsman and Lewis[4]) and applied to uniaxial media as well as an ionized gas. An extensive list of related works is cited in Chee-Seng's paper.[2]

2. MATRIX AND SCALAR OPERATORS

It is known[5] that there exists a real symmetric nonsingular matrix $\mu^{1/2}$ satisfying $\mu^{1/2}\,\mu^{1/2}=\mu$. Note that the determinant $\det\mu>0$, and that the inverse $(\mu^{1/2})^{-1}=\mu^{-1/2}$ is symmetric. Consider the real matrix

$$\tau=\mu^{-1/2}\epsilon\,\mu^{-1/2}, \tag{2.1}$$

which is known to be symmetric and positive definite. Let the row vector

$$\mathbf{F}=(\det\mu)^{-1/2}\,\mathbf{E}\,\mu^{1/2} \tag{2.2}$$

and the modified gradient operator

[a)]Permanent address: Department of Mathematics, National University of Singapore, Singapore 0511.

 0022-2488/81/112678-11$01.00

$$\nabla_y = \left(\frac{\partial}{\partial y_1}, \frac{\partial}{\partial y_2}, \frac{\partial}{\partial y_3}\right) = (\det\boldsymbol{\mu})^{-1/2}\,\nabla\boldsymbol{\mu}^{1/2}, \tag{2.3}$$

which exists in y-space with

$$\mathbf{y} = (y_1, y_2, y_3) = (\det\boldsymbol{\mu})^{1/2}\,\mathbf{x}\boldsymbol{\mu}^{-1/2}. \tag{2.4}$$

Suppose the 3×3 matrix operator

$$\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) = \mathbf{I}\frac{\partial^2}{\partial t^2} + \mathbf{P}(\nabla_y), \tag{2.5}$$

where $\mathbf{I}$ is the identity matrix and

$$\mathbf{P}(\nabla_y) = \boldsymbol{\tau}^{-1}(\nabla_y^T\nabla_y - \mathbf{I}\nabla_y\nabla_y^T), \tag{2.6}$$

with superscript T denoting the transpose. Then the following have been established (Ref. 1, Sec. 2)

$$\mathbf{F}_{tt}\boldsymbol{\tau} + \nabla_y\times(\nabla_y\times\mathbf{F}) = -(\det\boldsymbol{\mu})^{-1/2}\,\mathbf{J}_t\boldsymbol{\mu}^{-1/2}, \tag{2.7}$$

$$[\nabla_y\times(\nabla_y\times\mathbf{F})]^T = \boldsymbol{\tau}\,\mathbf{P}(\nabla_y)\mathbf{F}^T, \tag{2.8}$$

$$\det\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right) = L\left(\frac{\partial}{\partial t}, \nabla_y\right)\frac{\partial^2}{\partial t^2}, \tag{2.9}$$

where the scalar operator L is defined by

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right) = \frac{\partial^4}{\partial t^4} + \frac{\partial^2}{\partial t^2}\mathrm{tr}\mathbf{P}(\nabla_y) + \tfrac{1}{2}[\mathrm{tr}\mathbf{P}(\nabla_y)]^2 - \tfrac{1}{2}\mathrm{tr}\mathbf{P}^2(\nabla_y), \tag{2.10}$$

with tr denoting the trace; furthermore, for any constant X, we have the homogeneity relations

$$\mathrm{tr}\mathbf{P}(X\nabla_y) = X^2\,\mathrm{tr}\mathbf{P}(\nabla_y),$$
$$\mathrm{tr}\mathbf{P}^2(X\nabla_y) = X^4\,\mathrm{tr}\mathbf{P}^2(\nabla_y), \tag{2.11}$$

so that

$$L\left(X\frac{\partial}{\partial t}, X\nabla_y\right) = X^4 L\left(\frac{\partial}{\partial t}, \nabla_y\right). \tag{2.12}$$

Equation (2.7) can be derived from (1.1) and (1.2).

From (2.5), (2.7) and (2.8), we get

$$\mathbf{Q}\left(\frac{\partial}{\partial t}, \nabla_y\right)\mathbf{F}^T = -(\det\boldsymbol{\mu})^{-1/2}\,\boldsymbol{\tau}^{-1}\boldsymbol{\mu}^{-1/2}\,\mathbf{J}_t^T, \tag{2.13}$$

so that if, accounting for (2.9), the row vector $\mathbf{W}$ satisfies

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right)\mathbf{W} = -\mathbf{J}(\det\boldsymbol{\mu})^{-1}, \tag{2.14}$$

then

$$\mathbf{E}_t^T = (\det\boldsymbol{\mu})\boldsymbol{\mu}^{-1/2}\,\mathrm{adj}\mathbf{Q}\left(\frac{\partial}{\partial t}, (\det\boldsymbol{\mu})^{-1/2}\nabla\boldsymbol{\mu}^{1/2}\right)\times\boldsymbol{\tau}^{-1}\boldsymbol{\mu}^{-1/2}\,\mathbf{W}^T. \tag{2.15}$$

Now, the continuity condition arising from (1.2) and (1.3) is

$$\nabla\cdot\mathbf{J} = -\rho_t. \tag{2.16}$$

Consequently,

$$L\left(\frac{\partial}{\partial t}, \nabla_y\right)\nabla\cdot\mathbf{W} = \rho_t(\det\boldsymbol{\mu})^{-1}, \tag{2.17}$$

which together with (2.14) represent ultimate forms acquired through reduction of (1.1)–(1.3). In this paper, we focus our interests on the quantity $\nabla\cdot\mathbf{W}$ which satisfies (2.17).

3. FOURIER FORMULATION

We seek, ultimately, a general solution corresponding to the charged particle defined by (1.4) and (1.5). First we express the amplitude factor $\mathscr{A}(t)$ as a Fourier integral over a horizontal complex path

$$\mathscr{A}(t) = \int_{-\infty+ic}^{\infty+ic} \mathring{\mathscr{A}}(\sigma)\exp(-i\sigma t)\,d\sigma. \tag{3.1}$$

Then

$$\rho = (2\pi)^{-3}\det\boldsymbol{\mu}\iiint_{R_3}\exp(i\boldsymbol{\alpha}\cdot\mathbf{y})\,d\boldsymbol{\alpha} \times\int_{-\infty+ic}^{\infty+ic}\mathring{\mathscr{A}}(\sigma)\exp(-i\sigma t)\,d\sigma, \tag{3.2}$$

where the outer triple integral ranges with the wavevector position $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ over the infinite three-dimensional space R_3: $-\infty < \alpha_j < \infty\,(j = 1,2,3)$. Let us impose a stability rule that all singularities of $\mathring{\mathscr{A}}(\sigma)$ confine themselves to the region $\mathrm{Im}\sigma \leqslant 0$, and that $\mathring{\mathscr{A}}(\sigma)\to 0$ uniformly over $0 \leqslant \arg(\sigma - ic) < 2\pi$ as $|\sigma - ic|\to\infty$. We now stipulate that $c > 0$. In this case, $\mathring{\mathscr{A}}(\sigma)$ is analytic throughout $\mathrm{Im}\sigma \geqslant c$. If, for $t < 0$, the path $(-\infty + ic, \infty + ic)$ is closed by an infinite semicircle drawn into $\mathrm{Im}\sigma > c$, conditions of Jordan's lemma are satisfied so that, by the Cauchy–Goursat theorem, $\mathscr{A}(t)\equiv 0 (t < 0)$, which is consistent with (1.5). There is, consequently, no preexisting induced field, i.e.,

$$\nabla\cdot\mathbf{W}\equiv 0 \quad \text{during } t < 0. \tag{3.3}$$

To accommodate this behavior, we again employ the same complex path $(-\infty + ic, \infty + ic)$ in the Fourier representation

$$\nabla\cdot\mathbf{W} = \iiint_{R_3}\exp(i\boldsymbol{\alpha}\cdot\mathbf{y})d\boldsymbol{\alpha} \times\int_{-\infty+ic}^{\infty+ic}\mathrm{FT}[\nabla\cdot\mathbf{W}]\exp(-i\sigma t)\,d\sigma. \tag{3.4}$$

It follows from (2.17), (3.2), and (3.4) together with the homogeneity law (2.12) that the Fourier transform

$$\mathrm{FT}[\nabla\cdot\mathbf{W}] = -\frac{i\sigma\mathring{\mathscr{A}}(\sigma)}{(2\pi)^3 L(-\sigma,\boldsymbol{\alpha})}. \tag{3.5}$$

Now (Ref. 1, Sec. 3)

$$L(-\sigma,\boldsymbol{\alpha}) = [\sigma^2 - \sigma_+^2(\boldsymbol{\alpha})][\sigma^2 - \sigma_-^2(\boldsymbol{\alpha})], \tag{3.6}$$

where

$$\sigma_\pm(\boldsymbol{\alpha}) = \{-\tfrac{1}{2}\mathrm{tr}\mathbf{P}(\boldsymbol{\alpha}) \pm \tfrac{1}{2}[2\,\mathrm{tr}\mathbf{P}^2(\boldsymbol{\alpha}) - (\mathrm{tr}\mathbf{P}(\boldsymbol{\alpha}))^2]^{1/2}\}^{1/2}. \tag{3.7}$$

Furthermore, the positive definite symmetric matrix $\boldsymbol{\tau}$ has positive eigenvalues $\lambda_j^{-2}\,(j = 1,2,3)$, say. There is no loss of generality in assuming

$$\lambda_1 \geqslant \lambda_2 > \lambda_3 > 0, \tag{3.8}$$

as well as the diagonal form

$$\boldsymbol{\tau} = \begin{pmatrix} \lambda_1^{-2} & 0 & 0 \\ 0 & \lambda_2^{-2} & 0 \\ 0 & 0 & \lambda_3^{-2} \end{pmatrix} \tag{3.9}$$

due to the principal axes theorem. Whereupon, ϵ and μ should generally be accepted as being nondiagonal. One can now express

$$-\mathrm{tr}\mathbf{P}(\boldsymbol{\alpha}) = \lambda_1^2(\alpha_2^2+\alpha_3^2)+\lambda_2^2(\alpha_3^2+\alpha_1^2)+\lambda_3^2(\alpha_1^2+\alpha_2^2), \tag{3.10}$$

$$2\,\mathrm{tr}\mathbf{P}^2(\boldsymbol{\alpha}) - [\,\mathrm{tr}\mathbf{P}(\boldsymbol{\alpha})]^2 = [(|\alpha_1||\lambda_2^2-\lambda_3^2|^{1/2} - |\alpha_3||\lambda_1^2-\lambda_2^2|^{1/2})^2 + \alpha_2^2(\lambda_1^2-\lambda_3^2)] \times [(|\alpha_1||\lambda_2^2-\lambda_3^2|^{1/2} + |\alpha_3||\lambda_1^2-\lambda_2^2|^{1/2})^2 + \alpha_2^2(\lambda_1^2-\lambda_3^2)], \tag{3.11}$$

and then go on to verify that

$$\sigma_\pm(\mathbf{0}) = 0, \quad \text{while } \sigma_\pm(\boldsymbol{\alpha}) > 0 \ \forall \boldsymbol{\alpha} \neq \mathbf{0}. \tag{3.12}$$

Thus by (3.5) and (3.6), FT[$\nabla\cdot\mathbf{W}$] satisfies the same conditions within the σ-plane as those imposed on $\mathring{\mathscr{A}}(\sigma)$. With reference to (3.4), then, extension of the path $(-\infty+ic, \infty+ic)$ into the halfplane $\mathrm{Im}\sigma > c$ immediately confirms satisfaction of (3.3).

Henceforth, unless specified otherwise, we assume that $t > 0$. If the σ-integral in (3.4) is to be evaluated first, further properties of $\mathring{\mathscr{A}}(\sigma)$ over $\mathrm{Im}\sigma < c$ must be available. In particular, all its singularities inside that region must be classified. All this may entail some relaxation of generality. The difficulty can be avoided by initially treating the triple integral over R_3 as the inner integral. To adopt this approach, we first observe from (2.10) and (2.11) that

$$L(-\sigma, -\boldsymbol{\alpha}) = L(-\sigma, \boldsymbol{\alpha}). \tag{3.13}$$

Whence, from (3.4) and (3.5),

$$\nabla\cdot\mathbf{W} = \int_{-\infty+ic}^{\infty+ic} \exp(-i\sigma t)\,d\sigma \times \iiint_{\mathrm{R}_3} \mathrm{FT}\,[\nabla\cdot\mathbf{W}]\cos(\boldsymbol{\alpha}\cdot\mathbf{y})d\boldsymbol{\alpha}. \tag{3.14}$$

4. GENERAL SOLUTION

From (3.5), (3.6), and (3.14), we can express

$$\nabla\cdot\mathbf{W} = I_+ - I_-, \tag{4.1}$$

where

$$I_\nu = i(2\pi)^{-3}\int_{-\infty+ic}^{\infty+ic} \sigma\mathring{\mathscr{A}}(\sigma)\exp(-i\sigma t)d\sigma \times \iiint_{\mathrm{R}_3} \frac{\cos(\boldsymbol{\alpha}\cdot\mathbf{y})\,d\boldsymbol{\alpha}}{[\sigma_-^2(\boldsymbol{\alpha}) - \sigma_+^2(\boldsymbol{\alpha})][\sigma^2 - \sigma_\nu^2(\boldsymbol{\alpha})]} \tag{4.2}$$

with $\nu = \pm$. We shall next attempt to reduce I_ν to a spherical integral. For purposes of application, it is, as indicated in Sec. 1, desirable to incorporate linear transformations, one for each I_ν. Such transformations, which can be partially deformative as well as partially orthogonal, should be implemented at this stage rather than on attainment of spherical integral forms. Thus we propose that

$$\boldsymbol{\alpha} = \boldsymbol{\beta}\mathbf{A}_\nu \quad (\nu = \pm), \tag{4.3}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ and $\mathbf{A}_\nu$ is a $\boldsymbol{\beta}$-independent, real, nonsingular matrix. Any choice for $\mathbf{A}_\nu$ would depend on the particular type of crystal as well as, probably, the directional bearing of the recording point.

Owing to (3.7) and both homogeneity relations of (2.11),

$$\sigma_\pm(\boldsymbol{\beta}\mathbf{A}_\nu) = |\boldsymbol{\beta}|\sigma_\pm(\boldsymbol{\zeta}\mathbf{A}_\nu), \tag{4.4}$$

wherein the unit position vector $\boldsymbol{\zeta} = \boldsymbol{\beta}|\boldsymbol{\beta}|^{-1} \in \Omega_3$, the three-dimensional unit spherical surface. Whence, in (4.2), we have under (4.3),

$$\iiint_{\mathrm{R}_3} \ldots d\boldsymbol{\alpha} = |\det\mathbf{A}_\nu| \iiint_{\mathrm{R}_3} \ldots d\boldsymbol{\beta} = |\det\mathbf{A}_\nu| \iint_{\Omega_3} d\Omega \int_0^\infty \ldots |\boldsymbol{\beta}|^2 d|\boldsymbol{\beta}|, \tag{4.5}$$

$$= \tfrac{1}{2}|\det\mathbf{A}_\nu| \iint_{\Omega_3} \frac{d\Omega}{\sigma_-^2(\boldsymbol{\zeta}\mathbf{A}_\nu) - \sigma_+^2(\boldsymbol{\zeta}\mathbf{A}_\nu)} \times \int_{-\infty}^{\infty} \frac{\exp(i\beta\,\boldsymbol{\zeta}\mathbf{A}_\nu\mathbf{y}^T)}{\sigma^2 - \beta^2\sigma_\nu^2(\boldsymbol{\zeta}\mathbf{A}_\nu)}\,d\beta, \tag{4.6}$$

$d\Omega$ being a surface element of Ω_3. Now, $(\boldsymbol{\zeta}\mathbf{A}_\nu)\mathbf{A}_\nu^{-1} = \boldsymbol{\zeta} \neq \mathbf{0}$, implying $\boldsymbol{\zeta}\mathbf{A}_\nu \neq \mathbf{0}$ so that, by (3.12), $\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu) > 0$. Hence, for each $\sigma \in (-\infty+ic, \infty+ic)$ and each $\boldsymbol{\zeta} \in \Omega_3$, the β integrand in (4.6) possesses a symmetric pair of complex poles at

$$\beta = \pm\,\sigma/\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu). \tag{4.7}$$

These poles never meet as they are permanently restricted to the respective halfplanes $\mathrm{Im}\beta \gtrless 0$.

If the scalar $\boldsymbol{\zeta}\mathbf{A}_\nu\mathbf{y}^T \gtrless 0$, the β integral in (4.6) may be tackled by contour extension into $\mathrm{Im}\beta \gtrless 0$ and by invoking Jordan's lemma coupled to the residue theory to yield

$$\int_{-\infty}^{\infty} \ldots d\beta = -\frac{\pi i}{\sigma\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu)} \exp[i\sigma|\boldsymbol{\zeta}\mathbf{A}_\nu\mathbf{y}^T|/\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu)]. \tag{4.8}$$

Here, $\mathrm{Im}\sigma = c$. We must also admit the situation wherein $\boldsymbol{\zeta}\mathbf{A}_\nu\mathbf{y}^T = 0$. Let us assume, hereafter, that the recording point $\mathbf{x}$ avoids the charged particle. Then, via (2.4),

$$\boldsymbol{\mu}^{1/2}\mathbf{A}_\nu^{-1}(\mathbf{y}\mathbf{A}_\nu^T)^T(\det\boldsymbol{\mu})^{-1/2} = \mathbf{x}^T \neq \mathbf{0}^T (\text{zero column}), \tag{4.9}$$

so that, in particular, $\mathbf{y}\mathbf{A}_\nu^T \neq \mathbf{0}$. Thus, $\boldsymbol{\zeta}\mathbf{A}_\nu\mathbf{y}^T = 0$ if and only if the unit vector $\boldsymbol{\zeta}$ is orthogonal to $\mathbf{y}\mathbf{A}_\nu^T$. For every such $\boldsymbol{\zeta}$ direction, appropriate contour extensions are still covered by zero convergence conditions, and the result (4.8) remains valid. Whereupon, from (4.2), (4.6), (4.8), (3.1), and (2.4), we arrive at

$$I_\nu = \frac{|\det\mathbf{A}_\nu|}{16\pi^2} \times \iint_{\Omega_3} \frac{\mathscr{A}[t - (\det\boldsymbol{\mu})^{1/2}|\boldsymbol{\zeta}\mathbf{A}_\nu\boldsymbol{\mu}^{-1/2}\mathbf{x}^T|/\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu)]}{\sigma_\nu(\boldsymbol{\zeta}\mathbf{A}_\nu)[\sigma_-^2(\boldsymbol{\zeta}\mathbf{A}_\nu) - \sigma_+^2(\boldsymbol{\zeta}\mathbf{A}_\nu)]}\,d\Omega \tag{4.10}$$

within the original $(\mathbf{x},t)$ hyperspace. The charge amplitude factor $\mathscr{A}(t)$ is implicated in (4.10) which provides the spherical integral sought. The general solution is now given by (4.1) accompanied by (4.10).

The functions $\sigma_+(\boldsymbol{\zeta}\mathbf{A}_\nu)$ and $\sigma_-(\boldsymbol{\zeta}\mathbf{A}_\nu)$ are distinct on Ω_3 except at four separate points where they concur (cf Ref. 5). It can be verified from (3.7) and (3.11) that, under the present circumstances, the four points of concurrence along Ω_3 oc-

cur at

$$\zeta = \pm \mathbf{l}_+ \mathbf{A}_\nu^{-1} |\mathbf{l}_+ \mathbf{A}_\nu^{-1}|^{-1},$$
$$\zeta = \pm \mathbf{l}_- \mathbf{A}_\nu^{-1} |\mathbf{l}_- \mathbf{A}_\nu^{-1}|^{-1}, \tag{4.11}$$

where we can take

$$\mathbf{l}_\pm = (|\lambda_1^2 - \lambda_2^2|^{1/2}, 0, \pm |\lambda_2^2 - \lambda_3^2|^{1/2}). \tag{4.12}$$

Amongst these four points, those that satisfy, in view of (1.5), the constraint relation

$$(\det\boldsymbol{\mu})^{1/2} |\zeta \mathbf{A}_\nu \boldsymbol{\mu}^{-1/2} \mathbf{x}^T| < t\sigma_\nu(\zeta \mathbf{A}_\nu) \tag{4.13}$$

lie along the effective integral range contained by Ω_3 and constitute integrand singularities. The latter must, in practice, be accounted for in (4.10) via, say, a Cauchy principal value interpretation of the singular spherical integral.

5. DEACTIVATION, DARK INTERIOR, IN-PHASE MODES

We shall next examine certain aspects of the crystal's response if the *activated* charge *deactivates* at some positive time t_0:

$$\mathscr{B}(t) = \mathscr{C}(t)\, H(t_0 - t). \tag{5.1}$$

In such an event, the one-dimensional Fourier transform

$$\mathring{\mathscr{A}}(\sigma) = (2\pi)^{-1} \int_0^\infty \mathscr{B}(t) \exp(i\sigma t)\, dt$$
$$= \mathring{\mathscr{A}}_1(\sigma) - \mathring{\mathscr{A}}_2(\sigma) \exp(i\sigma t_0). \tag{5.2}$$

Here,

$$\mathring{\mathscr{A}}_1(\sigma) = (2\pi)^{-1} \int_0^\infty \mathscr{C}(t) \exp(i\sigma t)\, dt, \tag{5.3}$$

$$\mathring{\mathscr{A}}_2(\sigma) = (2\pi)^{-1} \int_0^\infty \mathscr{C}(t + t_0) \exp(i\sigma t)\, dt, \tag{5.4}$$

which are assumed to satisfy the same criteria as those stipulated in Sec. 3 for the previous $\mathring{\mathscr{A}}(\sigma)$. In this case, however, the present $\mathring{\mathscr{A}}(\sigma)$ fails to satisfy the uniform convergence-to-zero criterion within $\pi < \arg(\sigma - ic) < 2\pi$ because of the odd factor $\exp(i\sigma t_0)$. Nevertheless, it must be emphasized that that criterion is essentially a sufficiency criterion. We only need to ensure that (3.3) holds under present circumstances. First, we express, via (5.2) and (3.5), the σ-integral in (3.4) as

$$\int_{-\infty + ic}^{\infty + ic} = \frac{1}{(2\pi)^3 i} \int_{-\infty + ic}^{\infty + ic} \frac{\sigma \mathring{\mathscr{A}}_1(\sigma)}{L(-\sigma, \boldsymbol{\alpha})} \exp(-i\sigma t)\, d\sigma$$
$$- \frac{1}{(2\pi)^3 i} \int_{-\infty + ic}^{\infty + ic} \frac{\sigma \mathring{\mathscr{A}}_2(\sigma)}{L(-\sigma, \boldsymbol{\alpha})} \exp[-i\sigma(t - t_0)]\, d\sigma. \tag{5.5}$$

Appealing to the same contour integration principles as those previously applied, directly after (3.12), to FT[$\nabla \cdot \mathbf{W}$], we deduce: on the right side of (5.5), the first integral involving $\mathring{\mathscr{A}}_1(\sigma)$ vanishes identically throughout $t < 0$, while the second integral involving $\mathring{\mathscr{A}}_2(\sigma)$ vanishes identically throughout $t < t_0$ and, therefore, over $t < 0$ as well. Thus, returning to (3.4), we confirm that (3.3) again holds. Hence (5.1) is compatible.

Consider (4.10). The $\mathscr{A}$-factor involved there is, presently, by (1.5) and (5.1),

$$\mathscr{A}(t - \chi_\nu) = \mathscr{C}(t - \chi_\nu) H(t - \chi_\nu) H(t_0 - t + \chi_\nu), \tag{5.6}$$

where

$$\chi_\nu = (\det\boldsymbol{\mu})^{1/2} |\zeta \mathbf{A}_\nu \boldsymbol{\mu}^{-1/2} \mathbf{x}^T| / \sigma_\nu(\zeta \mathbf{A}_\nu), \tag{5.7}$$

$$= \frac{(\det\boldsymbol{\mu})^{1/2} |(\zeta \mathbf{A}_\nu) \cdot (\mathbf{x} \boldsymbol{\mu}^{-1/2})|}{|\zeta \mathbf{A}_\nu| \sigma_\nu(\zeta \mathbf{A}_\nu | \zeta \mathbf{A}_\nu|^{-1})} \tag{5.8}$$

after accounting for the linear homogeneity rule of (4.4) with $\mathbf{I}$ replacing $\mathbf{A}_\nu$ and $\zeta \mathbf{A}_\nu$ replacing $\boldsymbol{\beta}$. Now, since $\sigma_-(\zeta) > 0$, it achieves, on Ω_3, a positive minimum σ_0, say, in which case we have, via (3.7), (3.8), (3.11), and (3.12),

$$\sigma_+(\zeta) \geqslant \sigma_-(\zeta) \geqslant \sigma_0 \text{ over } \Omega_3. \tag{5.9}$$

Hence

$$0 \leqslant \chi_\nu \leqslant Z \text{ on } \Omega_3, \tag{5.10}$$

with

$$Z = \sigma_0^{-1} (\det\boldsymbol{\mu})^{1/2} |\mathbf{x}\, \boldsymbol{\mu}^{-1/2}|. \tag{5.11}$$

N.B. $Z > 0$ under our ruling assumption: $\mathbf{x} \neq \mathbf{0}$. It is easily verified[1] that there exists an orthogonal matrix $\mathbf{M}$ such that if

$$\mathbf{x}\, \mathbf{M} = \mathbf{X} = (X_1, X_2, X_3), \tag{5.12}$$

then

$$Z^2 = \sigma_0^{-2} \mathbf{x}\, \boldsymbol{\mu}^{-1} \mathbf{x}^T \det\boldsymbol{\mu} = \sigma_0^{-2} (\det\boldsymbol{\mu}) \Sigma_{j=1,2,3} \mu_j^{-1} X_j^2, \tag{5.13}$$

$\mu_j (j = 1,2,3)$ being the eigenvalues, all necessarily positive, of the real symmetric positive definite matrix $\boldsymbol{\mu}$. The x_j axes and the X_j axes are obviously separated by a pure rotation. Consequently, for some fixed number $s \in [0,t)$, the equation $Z = t - s$ describes an ellipsoid $\xi(s)$ that expands about the charge point from which it originates at instant s. Furthermore, $Z < t - s$ (or $> t - s$) if and only if $\mathbf{x} \in \text{int}\xi(s)$ (or $\text{ext}\xi(s)$); here, $\text{int}\xi(s)$ and $\text{ext}\xi(s)$ denote, respectively, the interior and exterior to $\xi(s)$.

The ellipsoid $\xi(t_0)$ originates on *deactivation* and, thereafter, trails behind and concentrically with the ellipsoid $\xi(0)$ that has originated during activation. Let us focus on the period *after deactivation*: $t > t_0$. Take any $\mathbf{x} \in \text{int}\xi(t_0)$. Then $\chi_\nu \leqslant Z < t - t_0$, and so (5.6) gives $\mathscr{A}(t - \chi_\nu) \equiv 0$. Therefore, by (4.1) and (4.10),

$$\nabla \cdot \mathbf{W} \equiv 0 \qquad \forall \mathbf{x} \in \text{int}\xi(t_0) \text{ but } \neq \mathbf{0}, \tag{5.14}$$

so that total darkness is experienced around the particle charge and inside the expanding $\xi(t_0)$. Actually, $\text{int}\xi(t_0)$ may be envisaged as a subregion of an expanding interior core of darkness. The formation of such a dark core should not be surprising because of the radiation cutoff due to deactivation. The radiation field precedes the dark core and recedes as the core expands so that, ultimately, an overall state of darkness prevails at infinite time. In Sec. 8, a dark core together with other neighbouring dark interiors will be explicitly constructed for the compounded uniaxial medium.

Suppose, between activation and deactivation, the charge is frequency modulated with or without damping, viz.

$$\mathscr{C}(t) = \exp(-i\omega t), \tag{5.15}$$

where $\text{Im}\omega \leqslant 0$. Then (5.3) and (5.4) yield

$$\mathring{\mathscr{A}}_1(\sigma) = (2\pi i)^{-1} (\omega - \sigma)^{-1},$$
$$\mathring{\mathscr{A}}_2(\sigma) = (2\pi i)^{-1} (\omega - \sigma)^{-1} \exp(-i\omega t_0) \tag{5.16}$$

along our prescribed σ path $(-\infty + ic, \infty + ic)$. On continuation beyond this path, both $\mathring{\mathscr{A}}_1(\sigma)$ and $\mathring{\mathscr{A}}_2(\sigma)$ obviously satisfy, as postulated, the same conditions as those imposed, after (3.1), on the previous $\mathring{\mathscr{A}}(\sigma)$. Now, in a harmonic state problem without deactivation, one employs purely pulsatory sources or boundary values without damping and seeks field quantities that pulsate in phase at some definite time. It is normally implicit that activation occurs at some negative infinite time. Hence, between activation and reception, there is actually an infinite time lapse. This is often more than adequate to allow developments to settle down everywhere to a steady harmonic state with in-phase modes. In the present investigation, the time lapse is finite. In this case, corollary results that will be secured in Secs. 7 and 8 suggest that in-phase modes do not exist everywhere but are confined to evolving illuminated regions. These are normally bounded by progressive wavefronts that convey singular modes.

We shall now derive, for the general crystalline medium, just one in-phase mode associated with (5.15). Let us pick any instant t before deactivation and any position $\mathbf{x}$ inside the expanding ellipsoid $\xi(0)$. Then, via (5.10), $t_0 - t + \chi_\nu > 0$ while $t - \chi_\nu \geqslant t - Z > 0$, so that by (5.6) and (5.15),

$$\mathscr{A}(t-\chi_\nu) = \mathscr{C}(t-\chi_\nu) = \exp(-i\omega t)\exp(i\omega\chi_\nu) \tag{5.17}$$

on Ω_3. Hence (4.1), (4.10), and (5.7) lead to

$$\nabla\cdot\mathbf{W} = (I^*_+ - I^*_-)\exp(-i\omega t) \qquad \forall \mathbf{x} \in \text{int}\xi(0) \text{ but } \neq \mathbf{0}, \tag{5.18}$$

wherein the time-independent factor

$$I^*_\nu = \frac{|\det \mathbf{A}_\nu|}{16\pi^2} \times \iint_{\Omega_3} \frac{\exp[i\omega(\det\boldsymbol{\mu})^{1/2}|\boldsymbol{\xi}\mathbf{A}_\nu\boldsymbol{\mu}^{-1/2}\mathbf{x}^T|/\sigma_\nu(\boldsymbol{\xi}\mathbf{A}_\nu)]}{\sigma_\nu(\boldsymbol{\xi}\mathbf{A}_\nu)[\sigma^2_-(\boldsymbol{\xi}\mathbf{A}_\nu) - \sigma^2_+(\boldsymbol{\xi}\mathbf{A}_\nu)]}\, d\Omega. \tag{5.19}$$

Thus, about the charged particle, an in-phase mode originates with $\xi(0)$ on activation and is subsequently sustained over some expanding region containing int$\xi(0)$. Without deactivation: $t_0 = \infty$, such an in-phase mode would advance indefinitely and, ultimately, permeate the entire crystal. As Ω_3 is the effective integral range in (5.19), the integrand is presently singular at all four points identified in (4.11); an appropriate Cauchy principal value should again be taken.

Suppose the charge is time-independent between activation and deactivation. i.e., with reference to (5.15), $\omega = 0$. Then selecting $\mathbf{A}_+ = \mathbf{A}_- = \mathbf{I}$, we deduce from (5.18) and (5.19) that before deactivation

$$\nabla\cdot\mathbf{W} = \frac{1}{16\pi^2}\iint_{\Omega_3} \frac{d\Omega}{\sigma_+(\boldsymbol{\xi})\sigma_-(\boldsymbol{\xi})[\sigma_+(\boldsymbol{\xi}) + \sigma_-(\boldsymbol{\xi})]}, \tag{5.20}$$

a $(\mathbf{x},t)$ − independent solution which holds throughout an expanding region that includes int$\xi(0)$ but excludes the charge point. The particular integral representation is obviously nonsingular. A Cauchy principal value interpretation is therefore not required.

6. DIFFERENTIAL OPERATION AND COMPLEMENTARY ORTHOGONAL TRANSFORMATION

A differential operation on $\nabla\cdot\mathbf{W}$ invariably alters its physical characteristics. Geometric structures may also be altered. Suitable combinations of differential operations with linear transformations can reduce the complexities in the spherical integrals through the introduction of symmetries. An example of this will be demonstrated for the compounded uniaxial problem (Sec. 7) where, in particular, a selected operation eliminates any integrand singularity derived from (4.11)–(4.13) and facilitates the construction of closed form results. Furthermore, certain geometric features of the wave configuration change once the operation is withdrawn; this case will be treated in Sec. 9.

If the transformation matrix is $\mathbf{x}$-dependent, it will be affected by any spatial operation. Should one attempt to enforce a differential operation at the stage corresponding to (4.2), before the implementation of any $\mathbf{x}$-dependent transformation, then instead of the convergence situation leading to (4.8), one could eventually be confronted by a divergent β-integral if the operator involved has an order exceeding one in the spatial terms. Such a divergent integral would require a different method of approach. The difficulty is surmountable if the $\mathbf{x}$-dependence is completely embodied within an orthogonal matrix factor. However, any accompanying $(\mathbf{x},t)$-independent matrix factor, if nonorthogonal, should first be incorporated as in the previous manner with $\mathbf{A}_\nu$. A selected spatially influential differential operation may next be applied to each derived spherical integral. This can be subjected finally to the complementary orthogonal transformation which may be $\mathbf{x}$-dependent.

For all purposes, we confine our interests to second-order linear operations. The technique can be extended to higher-order linear operations. A second-order linear differential scalar operator with, say, constant coefficients is generally expressible as

$$\mathscr{L}\left(\frac{\partial}{\partial t}, \nabla_y\right) = \sum_{n=0,1,2} \mathscr{L}_n\left(\frac{\partial}{\partial t}, \nabla_y\right), \tag{6.1}$$

where

$$\mathscr{L}_0\left(\frac{\partial}{\partial t}, \nabla_y\right) = \mathscr{L}_0 \text{ (constant number)} \tag{6.2}$$

and, for any constant X,

$$\mathscr{L}_n\left(X\frac{\partial}{\partial t}, X\nabla_y\right) = X^n\mathscr{L}\left(\frac{\partial}{\partial t}, \nabla_y\right) \quad (n = 1,2), \tag{6.3}$$

which defines an nth degree homogeneity of $\mathscr{L}_n$. Thus, $\mathscr{L}_1$ is a linear combination of $\partial/\partial t$ and $\partial/\partial y_j$, while $\mathscr{L}_2$ is a linear combination of $\partial^2/\partial t^2$, $\partial^2/\partial t\partial y_j$ and $\partial^2/\partial y_j\partial y_k$; here, $1 \leqslant j,k \leqslant 3$.

Suppose $\mathbf{A}_\nu$ is $(\mathbf{x},t)$-independent. Then putting

$$\mathbf{y}\mathbf{A}^T_\nu = \mathbf{Y}_\nu, \tag{6.4}$$

it can be shown that

$$\mathscr{L}_n\left(\frac{\partial}{\partial t},\nabla_y\right)\mathscr{A}\left[t-\frac{|\beta\mathbf{Y}_\nu^T|}{\sigma_\nu(\beta\,\mathbf{A}_\nu)}\right]$$

$$=\begin{cases}\mathscr{L}_1\left[1,-\beta\mathbf{A}_\nu\dfrac{\operatorname{sgn}(\beta\mathbf{Y}_\nu^T)}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right]\mathscr{A}'\left[t-\dfrac{|\beta\mathbf{Y}_\nu^T|}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right] & (n=1), \quad (6.5)\\ \mathscr{L}_2\left[1,-\beta\mathbf{A}_\nu\dfrac{\operatorname{sgn}(\beta\mathbf{Y}_\nu^T)}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right]\mathscr{A}''\left[t-\dfrac{|\beta\mathbf{Y}_\nu^T|}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right] & \\ \quad -2\mathscr{A}'(t)\mathscr{L}_2(0,\beta\mathbf{A}_\nu)\dfrac{\delta(\beta\mathbf{Y}_\nu^T)}{\sigma_\nu(\beta\mathbf{A}_\nu)} & (n=2), \quad (6.6)\end{cases}$$

the row vector β being real and $(\mathbf{x},t)$-independent. The derivatives of $\mathscr{A}$ must be accepted within the context of generalized functions. Note, specifically with regard to (6.6), that if

$$\mathscr{L}_2\left(\frac{\partial}{\partial t},\nabla_y\right)=\mathscr{L}_2(\nabla_y), \quad (6.7)$$

which never involves $\partial/\partial t$, then

$$\mathscr{L}_2(1,\beta)=\mathscr{L}_2(0,\beta)=\mathscr{L}_2(\beta)\ \forall\beta\in\mathrm{R}_3. \quad (6.8)$$

On applying (6.1)–(6.6) to (4.10), we obtain

$$\mathscr{L}I_\nu=\mathscr{L}\left(\frac{\partial}{\partial t},\nabla_y\right)I_\nu=(4\pi)^{-2}|\det\mathbf{A}_\nu|\,(K_{1\nu}-K_{2\nu}), \quad (6.9)$$

where

$$K_{j\nu}=\iint_{\Omega_3}\mathscr{F}_j(\xi\mathbf{A}_\nu,\xi\mathbf{Y}_\nu^T,t)\,d\Omega\ (j=1,2), \quad (6.10)$$

with

$$\mathscr{F}_1(\beta\mathbf{A}_\nu,\beta\mathbf{Y}_\nu^T,t)$$

$$=\sum_{n=0,1,2}\mathscr{L}_n\left[1,-\beta\mathbf{A}_\nu\frac{\operatorname{sgn}(\beta\mathbf{Y}_\nu^T)}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right]\mathscr{A}^{(n)}\left[t-\frac{|\beta\mathbf{Y}_\nu^T|}{\sigma_\nu(\beta\mathbf{A}_\nu)}\right]$$

$$\times\{\sigma_\nu(\beta\mathbf{A}_\nu)[\sigma_-^2(\beta\mathbf{A}_\nu)-\sigma_+^2(\beta\mathbf{A}_\nu)]\}^{-1}, \quad (6.11)$$

$$\mathscr{F}_2(\beta\mathbf{A}_\nu,\beta\mathbf{Y}_\nu^T,t)$$

$$=\frac{2\mathscr{A}'(t)\mathscr{L}_2(0,\beta\mathbf{A}_\nu)\delta(\beta\mathbf{Y}_\nu^T)}{\sigma_\nu^2(\beta\mathbf{A}_\nu)[\sigma_-^2(\beta\mathbf{A}_\nu)-\sigma_+^2(\beta\mathbf{A}_\nu)]}. \quad (6.12)$$

For the first term in (6.11) $\mathscr{L}_0(1,\beta)=\mathscr{L}_0$. From (4.1), the corresponding operation on $\nabla\cdot\mathbf{W}$ yields, in terms of $\mathscr{L}I_\nu$:

$$\mathscr{L}\nabla\cdot\mathbf{W}=\mathscr{L}I_+-\mathscr{L}I_-. \quad (6.13)$$

To accomplish our objective, (6.10) must be modified to accommodate an orthogonal transformation. First, we define the related integral

$$f(\alpha)=\iint_{\Omega_3}\mathscr{F}_j(\alpha\xi\mathbf{A}_\nu,\alpha\xi\mathbf{Y}_\nu^T,t)\,d\Omega \quad (6.14)$$

over $0\leqslant\alpha<\infty$. Let a,b denote any two real numbers satisfying $b>a>0$. Then

$$\int_a^b f(\alpha)\alpha^2\,d\alpha$$

$$=\iiint_{\mathrm{R}_3}\mathscr{F}_j(\alpha\mathbf{A}_\nu,\alpha\mathbf{Y}_\nu^T,t)H(b-|\alpha|)H(|\alpha|-a)\,d\alpha, \quad (6.15)$$

where $\alpha=|\alpha|\xi\in\mathrm{R}_3$ with $\xi\in\Omega_3$ and $|\alpha|\in[0,\infty)$. We now implement the transformation: $\alpha\Rightarrow\alpha\mathbf{B}_\nu$ governed by the $\mathbf{x}$-dependent (α-independent) orthogonal matrix

$$\mathbf{B}_\nu=\mathbf{B}_\nu(\mathbf{Y}_\nu)=(\mathbf{B}_\nu^{-1})^T=\begin{pmatrix}\mathbf{b}_{1\nu}\\ \mathbf{b}_{2\nu}\\ \mathbf{b}_{3\nu}\end{pmatrix}, \quad (6.16)$$

$\mathbf{b}_{1\nu},\mathbf{b}_{2\nu},\mathbf{b}_{3\nu}$ being a right-handed set of mutually orthogonal unit row vectors. *N.B.*

$$\det\mathbf{B}_\nu=\mathbf{b}_{1\nu}\cdot(\mathbf{b}_{2\nu}\times\mathbf{b}_{3\nu})=1,\ |\alpha\mathbf{B}_\nu|=(\alpha\mathbf{B}_\nu\mathbf{B}_\nu^T\alpha^T)^{1/2}=|\alpha|.$$

Whereupon, (6.15) leads to

$$\int_a^b\left[f(\alpha)-\iint_{\Omega_3}\mathscr{F}_j(\alpha\xi\mathbf{B}_\nu\mathbf{A}_\nu,(\alpha\xi\mathbf{B}_\nu\mathbf{Y}_\nu^T,t)d\Omega\right]\alpha^2\,d\alpha$$

$$=0. \quad (6.17)$$

As $[a,b]$ represents an arbitrary subinterval of $(0,\infty)$, the square-bracketed integrand factor must be indentically zero. In particular, taking $\alpha=1$, we deduce that

$$K_{j\nu}=\iint_{\Omega_3}\mathscr{F}_j(\xi\mathbf{B}_\nu\mathbf{A}_\nu,\xi\mathbf{B}_\nu\mathbf{Y}_\nu^T,t)\,d\Omega, \quad (6.18)$$

the transformed version of (6.10). The $\mathscr{F}_j$-integrand is normally singular at corresponding images of the earlier singularities determined through (4.11)–(4.13). Hence $K_{j\nu}$ must also be accepted in the sense of a Cauchy principal value.

We shall next propose two examples for the $\mathbf{x}$-dependent matrix $\mathbf{B}_\nu$. Each example will be applicable to a separate situation covering the compounded uniaxial system.

In the first example we pick, for the third row of $\mathbf{B}_\nu$,

$$\mathbf{b}_{3\nu}=\mathbf{Y}_\nu|\mathbf{Y}_\nu|^{-1}. \quad (6.19)$$

Then

$$\mathbf{Y}_\nu\mathbf{B}_\nu^T=|\mathbf{Y}_\nu|\mathbf{b}_{3\nu}(\mathbf{b}_{1\nu}^T,\mathbf{b}_{2\nu}^T,\mathbf{b}_{3\nu}^T)=|\mathbf{Y}_\nu|(0,0,1). \quad (6.20)$$

So

$$\xi\mathbf{B}_\nu\mathbf{Y}_\nu^T=\xi\cdot(\mathbf{Y}_\nu\mathbf{B}_\nu^T)=|\mathbf{Y}_\nu|\cos\theta. \quad (6.21)$$

On adopting spherical polars,

$$\xi=(\sin\theta\cos\psi,\sin\theta\sin\psi,\cos\theta), \quad (6.22)$$

with $0\leqslant\theta\leqslant\pi$ and $0\leqslant\psi<2\pi$. Thus (6.18) becomes

$$K_{j\nu}=\int_0^{2\pi}d\psi\int_0^{\pi}\mathscr{F}_j(\xi\mathbf{B}_\nu\mathbf{A}_\nu,|\mathbf{Y}_\nu|\cos\theta,t)\sin\theta\,d\theta, \quad (6.23)$$

which displays an explicit dependence on $|\mathbf{Y}_\nu|$. However, $K_{j\nu}$ is also, in view of (6.19), implicitly dependent on the direction of $\mathbf{Y}_\nu$ through $\mathbf{B}_\nu$. Formula (6.1) or (6.12) should be used to compute $K_{j\nu}$ from (6.23). In particular, the double integral form for $K_{2\nu}$ reduces to

$$K_{2\nu}=2\frac{\mathscr{A}'(t)}{|\mathbf{Y}_\nu|}\times\int_0^{2\pi}\frac{\mathscr{L}_2(0,\xi^0\mathbf{B}_\nu\mathbf{A}_\nu)\,d\psi}{\sigma_\nu^2(\xi^0\mathbf{B}_\nu\mathbf{A}_\nu)[\sigma_-^2(\xi^0\mathbf{B}_\nu\mathbf{A}_\nu)-\sigma_+^2(\xi^0\mathbf{B}_\nu\mathbf{A}_\nu)]}, \quad (6.24)$$

where $\xi^0=(\cos\psi,\sin\psi,0)\in\Omega_3$.

Relative to a Cartesian frame,

$$\mathbf{Y}_\nu=(Y_{1\nu},Y_{2\nu},Y_{3\nu})=|\mathbf{Y}_\nu|(\sin\Theta_\nu\cos\Psi_\nu,\sin\Theta_\nu\sin\Psi_\nu,\cos\Theta_\nu), \quad (6.25)$$

where $0 \leqslant \Theta_\nu \leqslant \pi$, $0 \leqslant \Psi_\nu < 2\pi$. For the second example we choose

$$\mathbf{B}_\nu = \begin{pmatrix} \cos\Psi_\nu & \sin\Psi_\nu & 0 \\ -\sin\Psi_\nu & \cos\Psi_\nu & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6.26}$$

whose rows do constitute a right-handed set of mutually orthogonal unit vectors. It is then seen that

$$\mathbf{Y}_\nu \mathbf{B}_\nu^T = (\mathscr{R}_\nu, 0, Y_{3\nu}), \tag{6.27}$$

where

$$\mathscr{R}_\nu = |\mathbf{Y}_\nu| \sin\Theta_\nu = (Y_{1\nu}^2 + Y_{2\nu}^2)^{1/2}. \tag{6.28}$$

Whereupon, we derive from (6.18),

$$K_{j\nu} = \int_0^{2\pi} d\psi \int_0^{\pi} \mathscr{F}_j(\boldsymbol{\xi}\mathbf{B}_\nu\mathbf{A}_\nu, \mathscr{R}_\nu \sin\theta\cos\psi + Y_{3\nu}\cos\theta, t) \times \sin\theta\, d\theta, \tag{6.29}$$

which makes explicit the dependence on $\mathscr{R}_\nu$ and $Y_{3\nu}$. Moreover, $K_{j\nu}$ is additionally dependent on the azimuthal angle Ψ_ν through the choice of (6.26) for $\mathbf{B}_\nu$. Thus (6.29) is a cylindrical polar representation of $K_{j\nu}$ in $\mathscr{R}_\nu$, $Y_{3\nu}$ and Ψ_ν coordinates; observe that the variable

$$\boldsymbol{\xi}\mathbf{B}_\nu\mathbf{A}_\nu = (\sin\theta\cos(\Psi_\nu + \psi), \sin\theta\sin(\Psi_\nu + \psi), \cos\theta)\mathbf{A}_\nu. \tag{6.30}$$

In contrast, (6.23) is, on taking (6.19) and (6.25) into account, a spherical polar representation of $K_{j\nu}$ in $|\mathbf{Y}_\nu|, \Theta_\nu, \Psi_\nu$ coordinates.

7. THE COMPOUNDED UNIAXIAL PROBLEM

The crystalline medium employed by Chee-Seng[2] possesses a compounded uniaxiality on the basis that $\boldsymbol{\tau}$ has a double eigenvalue, i.e.,

$$\lambda_1 = \lambda_2 = \lambda_+, \text{ say, with } \lambda_+ > \lambda_- = \lambda_3 > 0. \tag{7.1}$$

Such a medium, whose corresponding permittivity and permeability principal axes remain unaligned, is uniaxial, in the conventional sense, only within the y-frame; within the actual x-space of observations, none of the six principal axes serves as a uniaxis. Media under this category include [refer to Ref. 2, (7.61)–(7.63)] the class of crystals dealt with by Besieris,[6] Majumdar and Pal,[7] and Lewandowski[8] (cf. also Felsen[9,10]); a typical crystal belonging to this particular class has aligned pairs of permittivity and permeability principal axes and is governed by a certain "generalized uniaxiality" which, in turn, covers ordinary uniaxiality under single anisotropy in either $\boldsymbol{\epsilon}$ or $\boldsymbol{\mu}$ (Lee and Lo,[11] Kong,[12] Mei,[13] Lu and Mei[14]).

On adopting (7.1), expressions (3.7), (3.10), and (3.11) yield

$$\sigma_+(\boldsymbol{\alpha}) = \lambda_+ |\boldsymbol{\alpha}|, \tag{7.2}$$

$$\sigma_-(\boldsymbol{\alpha}) = [\lambda_-^2 (\alpha_1^2 + \alpha_2^2) + \lambda_+^2 \alpha_3^2]^{1/2}. \tag{7.3}$$

The four possible singularities represented by (4.11) and (4.12) presently decrease to only two. In particular, the $\mathscr{F}_j$ integrand in (6.18) is potentially singular under ordinary circumstances at the two positions

$$\boldsymbol{\zeta} = \pm \frac{(0,0,1)(\mathbf{B}_\nu\mathbf{A}_\nu)^{-1}}{|(0,0,1)(\mathbf{B}_\nu\mathbf{A}_\nu)^{-1}|} \in \Omega_3. \tag{7.4}$$

Consider the spatial operator

$$\mathscr{L} = (\lambda_+^2 - \lambda_-^2)\left(\frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial y_2^2}\right), \tag{7.5}$$

a scalar multiple of the two-dimensional Laplacian within the y-frame. $\mathscr{L}$ obviously falls under category (6.1):

$$\mathscr{L}_1\left(\frac{\partial}{\partial t}, \nabla_y\right) \equiv \mathscr{L}_0 \equiv 0, \quad \mathscr{L} = \mathscr{L}_2(\nabla_y). \tag{7.6}$$

It then follows from (6.7), (6.8), (7.2), (7.3), (7.5), and (7.6) that, in the notations of Sec. 6,

$$\mathscr{L}_2(1, X\boldsymbol{\alpha}) = \mathscr{L}_2(0, X\boldsymbol{\alpha}) = \mathscr{L}_2(X\boldsymbol{\alpha}) = X^2[\sigma_+^2(\boldsymbol{\alpha}) - \sigma_-^2(\boldsymbol{\alpha})] \tag{7.7}$$

for any scalar X and any $\boldsymbol{\alpha} \in \mathrm{R}_3$. Whereupon,, if the particular $\mathscr{L}$-operator of (7.5) acts on I_ν, then by (6.11) and (6.12), neither of the two singularities expressed by (7.4) can appear as these would be cancelled out.

Accounting for (3.9) and (7.1), one may express $\boldsymbol{\tau}^{-1} = \boldsymbol{\tau}^{-1/2}\boldsymbol{\tau}^{-1/2}$ with

$$\boldsymbol{\tau}^{-1/2} = \begin{pmatrix} \lambda_+ & 0 & 0 \\ 0 & \lambda_+ & 0 \\ 0 & 0 & \lambda_- \end{pmatrix}. \tag{7.8}$$

Let us now select, for each $(\mathbf{x}, t)$-independent $\mathbf{A}_\nu$,

$$\mathbf{A}_+ = \mathbf{I}\lambda_+^{-1}, \quad \mathbf{A}_- = \boldsymbol{\tau}^{-1/2}\lambda_+^{-1}\lambda_-^{-1}. \tag{7.9}$$

Thus, if $\mathbf{B}_\nu$ is orthogonal, (7.2) and (7.3) yield

$$\sigma_\nu(\boldsymbol{\xi}\mathbf{B}_\nu\mathbf{A}_\nu) = 1 \quad \forall \boldsymbol{\xi} \in \Omega_3 \quad (\nu = \pm). \tag{7.10}$$

We then deduce from (6.11), (7.6), (7.7), and (7.10) that

$$\mathscr{F}_1(\boldsymbol{\xi}\mathbf{B}_\nu\mathbf{A}_\nu, \boldsymbol{\xi}\mathbf{B}_\nu\mathbf{Y}_\nu^T, t) = -\mathscr{A}''(t - |\boldsymbol{\xi}\mathbf{B}_\nu\mathbf{Y}_\nu^T|), \tag{7.11}$$

which, applied to (6.23), gives

$$K_{1\nu} = -\int_0^{2\pi} d\psi \int_0^{\pi} \mathscr{A}''(t - |\mathbf{Y}_\nu||\cos\theta|)\sin\theta\, d\theta, \tag{7.12}$$

$$= \frac{4\pi}{|\mathbf{Y}_\nu|}\int_t^{t-|\mathbf{Y}_\nu|} \mathscr{A}''(\xi)\, d\xi$$

$$= \frac{4\pi}{|\mathbf{Y}_\nu|}[\mathscr{A}'(t - |\mathbf{Y}_\nu|) - \mathscr{A}'(t)]. \tag{7.13}$$

Likewise, by (6.24), (7.7) and (7.10),

$$K_{2\nu} = -\frac{2\mathscr{A}'(t)}{|\mathbf{Y}_\nu|}\int_0^{2\pi} d\psi = -\frac{4\pi}{|\mathbf{Y}_\nu|}\mathscr{A}'(t). \tag{7.14}$$

Whereupon, on accommodating (1.5), (6.9) simplifies into

$$\mathscr{L}I_\nu = \frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|} \times [\mathscr{B}'(t - |\mathbf{Y}_\nu|)H(t - |\mathbf{Y}_\nu|) + \mathscr{B}(0)\delta(t - |\mathbf{Y}_\nu|)], \tag{7.15}$$

the $\mathscr{L}$-operator being given by (7.5).

We shall next provide interpretations over the original x-space. First, by (2.1), (3.9), and (7.1),

$$\det\boldsymbol{\mu} = (\det\boldsymbol{\tau})^{-1}\det\boldsymbol{\epsilon} = \lambda_+^4\lambda_-^2\det\boldsymbol{\epsilon}, \tag{7.16}$$

$$\boldsymbol{\mu}^{-1/2}\boldsymbol{\tau}^{-1}\boldsymbol{\mu}^{-1/2} = \boldsymbol{\epsilon}^{-1}. \tag{7.17}$$

Thus we secure from (6.4), (2.4), (7.8), and (7.9),

$$|\mathbf{Y}_+| = \lambda_+^{-1}(\mathbf{x}\boldsymbol{\mu}^{-1}\mathbf{x}^T \det\boldsymbol{\mu})^{1/2}, \tag{7.18}$$

$$|\mathbf{Y}_-| = \lambda_+(\mathbf{x}\boldsymbol{\epsilon}^{-1}\mathbf{x}^T \det\boldsymbol{\epsilon})^{1/2}. \tag{7.19}$$

Comparison of (7.18) with (5.13) indicates

$$|\mathbf{Y}_+|^2 = \lambda_+^{-2}(\det\boldsymbol{\mu}) \sum_{j=1,2,3} \mu_j^{-1} X_j^2. \tag{7.20}$$

Likewise, it can be verified (cf. Chee-Seng Ref. 2, Sec. 7) that (7.19) leads to

$$|\mathbf{Y}_-|^2 = \lambda_+^2 (\det\boldsymbol{\epsilon}) \sum_{j=1,2,3} \epsilon_j^{-1} X_{j-}^2 \tag{7.21}$$

where each $\epsilon_j > 0 (j = 1,2,3)$ and denotes an eigenvalue of $\boldsymbol{\epsilon}$; also,

$$(X_{1-}, X_{2-}, X_{3-}) = \mathbf{X}_- = \mathbf{x}\,\mathbf{M}_-, \tag{7.22}$$

which derives from $\mathbf{x}$ via a pure rotation with orthogonal matrix $\mathbf{M}_-$. The equation $|\mathbf{Y}_\nu| = t$ therefore represents an expanding ellipsoid ξ_ν that originates on activation. In particular, ξ_+ is a scaled version of the ellipsoid $\xi(0)$ encountered previously in Sec. 5. The two ellipsoids ξ_+ and ξ_- are concentered at the charged particle. However, their principal axes are not generally pairwise aligned. One instance where they become pairwise aligned is when $\mathbf{M} = \mathbf{M}_- = \mathbf{I}$; this leads to the "generalized uniaxial" criterion $\epsilon_1\epsilon_2^{-1} = \mu_1\mu_2^{-1}$ exploited by Besieris (1969), Majumdar and Pal (1970) and Lewandowski (1971). It has been proved (Chee-Seng Ref. 2, Sec. 7) that $\text{int}\xi_- \subset \text{int}\xi_+$; nevertheless, ξ_+ and ξ_- do, in fact, meet tangentially at two diametral retreating points at

$$\mathbf{x} = \pm\lambda_+ t(\det\boldsymbol{\mu})^{-1/2}(0,0,1)\boldsymbol{\mu}^{1/2}. \tag{7.23}$$

Elsewhere, ξ_+ leads ξ_-. Consequently, from (6.13) and (7.15),

$$\mathscr{L}\nabla\cdot\mathbf{W} = \begin{cases} 0 \quad \forall \mathbf{x} \in \text{ext}\xi_+, & (7.24) \\ \dfrac{|\det\mathbf{A}_+|}{4\pi|\mathbf{Y}_+|}\mathscr{B}'(t - |\mathbf{Y}_+|) \\ \forall \mathbf{x} \in \text{int}\xi_+ \cap \text{ext}\xi_-, & (7.25) \\ \dfrac{|\det\mathbf{A}_+|}{4\pi|\mathbf{Y}_+|}\mathscr{B}'(t - |Y_+|) - \dfrac{|\det\mathbf{A}_-|}{4\pi|\mathbf{Y}_-|} \\ \times\mathscr{B}'(t - |\mathbf{Y}_-|) \quad \forall \mathbf{x} \in \text{int}\xi_- \text{ but} \neq \mathbf{0}. & (7.26) \end{cases}$$

Additionally, a singular distribution

$$\pm\frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|}\mathscr{B}(0)\delta(t - |\mathbf{Y}_\nu|) \quad (\nu = \pm) \tag{7.27}$$

is convected along with each ξ_ν. There are clearly two unsteady wavefields, each progressing behind an ellipsoidal wavefront ξ_ν, to which an accompanying singular field adheres. The leading wavefront ξ_+ expands into a dark exterior wherein $\mathscr{L}\nabla\cdot\mathbf{W} \equiv 0$, consistent with Sommerfeld's radiation principle. Each wavefield suffers an ellipsoidal attenuation like $|\mathbf{Y}_\nu|^{-1}$.

Suppose $\{\mathscr{B}'(t_j)\}$ denotes the sequence of charge amplitude gradients over a time sequence $\{t_j\}$ satisfying $0 < t_j < t$. Then the constituent $\mathscr{L}I_\nu$-wavefield behind ξ_ν at instant t can be spatially plotted as an associated sequence

$$\left\{\frac{|\det\mathbf{A}_\nu|}{4\pi(t - t_j)}\mathscr{B}'(t_j)\right\} \tag{7.28}$$

of constants over a family $\{\xi_{j\nu}\}$ of concentric ellipsoids inside ξ_ν. Each member ellipsoid $\xi_{j\nu}$ has the equation $|\mathbf{Y}_\nu| = t - t_j$. It therefore originates at the instant t_j. Like ξ_+ and ξ_-, $\{\xi_{j_+}\}$ and $\{\xi_{j_-}\}$ are obliquely oriented with respect to one another. Thus we say that the system supports a pair of mutually oblique ellipsoidal symmetries.

8. EXPLICIT DEACTIVATION EFFECTS

In the event of a subsequent deactivation represented by (5.1), (7.15) becomes

$$\mathscr{L}I_\nu = \phi_\nu H(t - |\mathbf{Y}_\nu|)H(t_0 - t + |\mathbf{Y}_\nu|) + \frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|}[\mathscr{C}(0)\delta(t - |\mathbf{Y}_\nu|) - \mathscr{C}(t_0)\delta(t_0 - t + |\mathbf{Y}_\nu|)], \tag{8.1}$$

where

$$\phi_\nu = \frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|}\mathscr{C}'(t - |\mathbf{Y}_\nu|). \tag{8.2}$$

By substituting $\mathscr{C}$ for $\mathscr{B}$ into (7.15) and (7.24)–(7.27), we obtain corresponding results that hold before deactivation.

Consider the period after deactivation. The equation $|\mathbf{Y}_\nu| = t - t_0$ describes an ellipsoid ξ_ν^0 that originates on deactivation; it then expands behind and concentrically with ξ_ν. By similarity deduction, $\text{int}\xi_-^0 \subset \text{int}\xi_+^0$, while ξ_-^0 touches ξ_+^0 twice along the same diametric line as the two points of (7.23). Suppose the active interval t_0 is short enough to allow ξ_+^0 to intersect ξ_-. It then follows from (6.13) and (8.1) that

$$\mathscr{L}\nabla\cdot\mathbf{W} = \begin{cases} 0 & \forall \mathbf{x} \in \text{ext}\xi_+, & (8.3) \\ \phi_+ & \forall \mathbf{x} \in \text{int}\xi_+ \cap \text{ext}\xi_+^0 \cap \text{ext}\xi_-, & (8.4) \\ \phi_+ - \phi & \forall \mathbf{x} \in \text{int}\xi_- \cap \text{ext}\xi_+^0, & (8.5) \\ -\phi_- & \forall \mathbf{x} \in \text{int}\xi_- \cap \text{ext}\xi_-^0 \cap \text{int}\xi_+^0, & (8.6) \\ 0 & \forall \mathbf{x} \in (\text{int}\xi_+^0 \cap \text{ext}\xi_-) \cup \text{int}\xi_-^0 & \\ & \text{but} \neq \mathbf{0}. & (8.7) \end{cases}$$

Fig. 1 depicts the propagation pattern. The $\mathscr{L}\nabla\cdot\mathbf{W}$-field is partitioned by the four ellipsoidal wavefronts ξ_+, ξ_-, ξ_+^0, and ξ_-^0. The foremost among these, viz. ξ_+, expands into a dark exterior, as in the previous case represented by (7.24). Consistent with our earlier prediction via (5.14), an expanding dark core occupies $\text{int}\xi_-^0$ but avoids the particle charge.

This core occurs between but remains isolated from two other dark interiors jointly denoted by $\text{int}\xi_+^0 \cap \text{ext}\xi_-$. Illuminated constituents of the field are expressed by (8.4)–(8.6). The dividing wavefronts are singularly illuminated by the following superimposed distributions:

$$\pm\frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|}\mathscr{C}(0)\delta(t - |\mathbf{Y}_\nu|) \quad \text{along } \xi_\nu, \tag{8.8}$$

$$\mp\frac{|\det\mathbf{A}_\nu|}{4\pi|\mathbf{Y}_\nu|}\mathscr{C}(t_0)\delta(t_0 - t + |\mathbf{Y}_\nu|) \quad \text{along } \xi_\nu^0. \tag{8.9}$$

The situation wherein ξ^0_+ cannot intersect ξ_- can be similarly analysed.

For the *pulsatory* charge defined through (5.15), (8.4)–(8.6) provide representations of in-phase modes that pulsate, after deactivation, with the same real or complex source frequency ω. These modes do not, therefore, occupy the entire crystalline medium at finite time, but are confined within definite evolving domains illustrated in Fig. 1. On the other hand, in-phase modes before deactivation occur behind ξ_- as well as between ξ_+ and ξ_-.

If the charge is time-independent between activation and deactivation, then away from it, $\mathscr{L}\nabla\cdot\mathbf{W}$ vanishes everywhere except on ξ_+ and ξ_-, as well as on ξ^0_+ and ξ^0_- after deactivation. Precisely, $\mathscr{L}\nabla\cdot\mathbf{W}$ is singularly propagated along with each ellipsoidal wavefront in accordance with (8.8) or (8.9).

9. REDUCED SPHERICAL INTEGRALS

So far, results secured for the compounded uniaxial medium pertain to $\mathscr{L}I_\nu$ and $\mathscr{L}\nabla\cdot\mathbf{W}$ with $\mathscr{L}$ given by (7.5). Physical and geometric characteristics of both these quantities differ from those of I_ν and $\nabla\cdot\mathbf{W}$, which may be handled through (4.1) and (4.10). An alternative procedure, which we shall adopt, is to employ (6.9) with (6.10) by taking, with reference to (6.1) and (6.2),

$$\mathscr{L}_1\left(\frac{\partial}{\partial t}, \nabla_y\right)\equiv\mathscr{L}_2\left(\frac{\partial}{\partial t}, \nabla_y\right)\equiv 0,$$
$$\mathscr{L}_0 = 1\equiv\mathscr{L}\left(\frac{\partial}{\partial t}, \nabla_y\right). \tag{9.1}$$

In particular, then, (6.12) discloses

$$\mathscr{F}_2(\beta\mathbf{A}_\nu, \beta\mathbf{Y}^T_\nu, t)\equiv 0. \tag{9.2}$$

Hence

$$K_{2\nu} = 0. \tag{9.3}$$

To consider $K_{1\nu}$, it is presently appropriate to rely on (6.29). First, we derive from (6.30), (7.2), (7.3), (7.8), and (7.9),

$$\sigma^2_-(\xi\mathbf{B}_\nu\mathbf{A}_\nu) - \sigma^2_+(\xi\mathbf{B}_\nu\mathbf{A}_\nu) = \lambda_\nu^{-2}(\lambda^2_- - \lambda^2_+)\sin^2\theta. \tag{9.4}$$

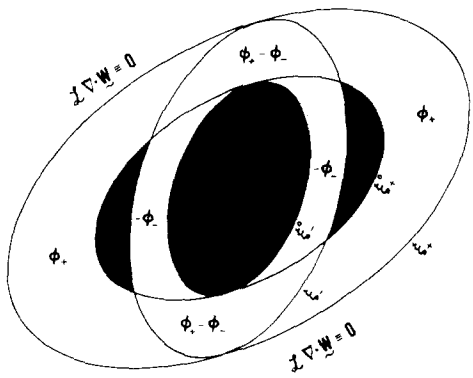


FIG. 1. The composite $\mathscr{L}\nabla\cdot\mathbf{W}$-field after deactivation, with ξ^0_+ intersecting ξ_-. The darkness representation: $\mathscr{L}\nabla\cdot\mathbf{W}\equiv 0$ holds throughout $\mathrm{ext}\,\xi_+$, the two darkened side regions between ξ^0_+ and ξ_-, as well as the darkened core region inside ξ^0_+ and surrounding the particle charge. Illuminated constituents are ϕ_+, $\phi_+ - \phi_-$, $-\phi_-$. Every ellipsoidal wavefront is singularly illuminated.

Then by (6.11), (7.10), (9.1), and (9.4),

$$\mathscr{F}_1(\xi\mathbf{B}_\nu\mathbf{A}_\nu, \xi\mathbf{B}_\nu\mathbf{Y}^T_\nu, t) = \frac{\lambda_\nu^2\mathscr{A}(t - |\xi\mathbf{B}_\nu\mathbf{Y}^T_\nu|)}{(\lambda^2_- - \lambda^2_+)\sin^2\theta}. \tag{9.5}$$

Observe that

$$|\xi\mathbf{B}_\nu\mathbf{Y}^T_\nu| = |\xi\cdot(\mathbf{Y}_\nu\mathbf{B}^T_\nu)| = |\mathscr{R}_\nu\sin\theta\cos\psi + Y_{3\nu}\cos\theta| \leqslant |\mathbf{Y}_\nu|, \tag{9.6}$$

which follows from (6.22), (6.27), and the orthogonality of $\mathbf{B}_\nu$.

Let us examine developments inside the ellipsoid ξ_ν. Accordingly, we choose any $\mathbf{x}\in\mathrm{int}\xi_\nu$ but $\neq\mathbf{0}$, so that $|\xi\mathbf{B}_\nu\mathbf{Y}^T_\nu| < t$. Whereupon, (1.5), (6.9), (6.29), (7.8), (7.9), (9.1), (9.3), and (9.5) lead to the reduced spherical integral representation

$$I_\nu = D\int_0^{2\pi} d\psi\int_0^{\pi}\mathscr{B}(t - |\mathscr{R}_\nu\sin\theta\cos\psi + Y_{3\nu}\cos\theta|)\times\csc\theta\, d\theta, \tag{9.7}$$

with

$$D = (16\pi^2\lambda_+)^{-1}(\lambda^2_- - \lambda^2_+)^{-1}. \tag{9.8}$$

Consequently,

$$\nabla\cdot\mathbf{W} = D\int_0^{2\pi} d\psi\int_0^{\pi}[\mathscr{B}(t - |\mathscr{R}_+\sin\theta\cos\psi + Y_{3+}\cos\theta|) - \mathscr{B}(t - |\mathscr{R}_-\sin\theta\cos\psi + Y_{3-}\cos\theta|)]\csc\theta\, d\theta, \tag{9.9}$$

another reduced spherical integral which holds inside ξ_- and about the charged particle. This region becomes a dark core: $\nabla\cdot\mathbf{W}\equiv 0$ should $\mathscr{B}(t)$ remain constant.

The integrand factor $\csc\theta$ involved in (9.7) and (9.9) is singular at the two inner integral limits $\theta = 0,\pi$. This agrees with the fact that both possible singularities indicated by

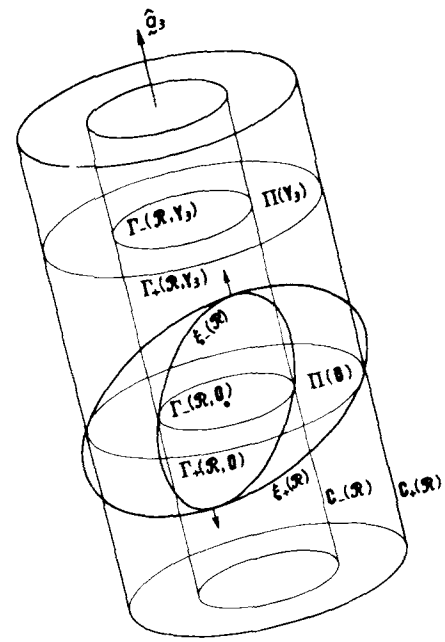


FIG. 2. Elliptical coaxial symmetries. The normals to both tangential contacts between $\xi_+(\mathbb{R})$ and $\xi_-(\mathbb{R})$ are parallel to the coaxial direction $\hat{\mathbf{a}}_3$.

(7.4) are actually

$$\pm \frac{(0,0,1)\mathbf{A}_\nu^{-1}\mathbf{B}_\nu^T}{|(0,0,1)\mathbf{A}_\nu^{-1}\mathbf{B}_\nu^T|} = \pm(0,0,1), \tag{9.10}$$

after accounting for the orthogonality of $\mathbf{B}_\nu$ together with (7.8), (7.9), and the choice of (6.26) which forms the basis of (6.29) and, hence, of (9.7) and (9.9).

Let us focus again on the region inside ξ_ν and about the charge. Here, according to (9.7), I_ν depends on $\mathscr{R}_\nu, Y_{3\nu}$, and t. The spatial dependence can be expounded in geometric terms. Let $\mathbf{a}_j$ denote the jth row (vector) of the symmetric matrix $\boldsymbol{\mu}^{-1/2}$, i.e.,

$$\boldsymbol{\mu}^{-1/2} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix} = (\mathbf{a}_1^T, \mathbf{a}_2^T, \mathbf{a}_3^T). \tag{9.11}$$

N. B. $\mathbf{a}_j \neq \mathbf{0} (j = 1,2,3)$. Then from (2.4), (6.4), (7.8), and (7.9), we have

$$\mathbf{Y}_\nu = (Y_{1\nu}, Y_{2\nu}, Y_{3\nu}) = (\det\boldsymbol{\mu})^{1/2}\left(\frac{\mathbf{x}\mathbf{a}_1^T}{\lambda_\nu}, \frac{\mathbf{x}\mathbf{a}_2^T}{\lambda_\nu}, \frac{\mathbf{x}\mathbf{a}_3^T}{\lambda_+}\right). \tag{9.12}$$

Hence

$$I_\nu = I_\nu\{((\mathbf{x}\mathbf{a}_1^T)^2 + (\mathbf{x}\mathbf{a}_2^T)^2)^{1/2}\lambda_\nu^{-1}, \mathbf{x}\mathbf{a}_3^T\lambda_+^{-1}, t\}. \tag{9.13}$$

Consider a parametric family $\{\xi_\nu(\mathbb{R})\}$ of permanently stationary ellipsoids that are distributed concentrically with and inside ξ_ν; as ξ_ν expands, such a family is free to grow in membership. A typical ellipsoidal member $\xi_\nu(\mathbb{R})$ of the family may be described by the equation

$$|\mathbf{Y}_\nu| = \mathbb{R}(\det\boldsymbol{\mu})^{1/2} (0 < \mathbb{R}(\det\boldsymbol{\mu})^{1/2} < t). \tag{9.14}$$

Next, the equation

$$\mathbf{x}\mathbf{a}_3^T = \lambda_+ Y_3 \tag{9.15}$$

describes a plane $\Pi(Y_3)$ having unit normal $\hat{\mathbf{a}}_3 = \mathbf{a}_3|\mathbf{a}_3|^{-1}$ and situated at a scaled distance $|Y_3|$ from the particle charge. Consider, inside ξ_ν, a uniform cylindrical surface $C_\nu(\mathbb{R})$ whose normal intersection with $\Pi(Y_3)$ is a closed circuit $\Gamma_\nu(\mathbb{R}, Y_3)$ governed by the equations

$$(\mathbf{x}\mathbf{a}_1^T)^2 + (\mathbf{x}\mathbf{a}_2^T)^2 = \lambda_\nu^2\mathbb{R}^2 : \mathbf{x}\,\mathbf{a}_3^T = \lambda_+ Y_3. \tag{9.16}$$

In particular

$$\Gamma_\nu(\mathbb{R},0) = \xi_\nu(\mathbb{R}) \cap \Pi(0), \tag{9.17}$$

which is an ellipse. The cylindrical surface $C_\nu(\mathbb{R})$ is therefore elliptic. Evidently, $C_+(\mathbb{R})$ and $C_-(\mathbb{R})$ share the same axis which passes through the charged particle on $\Pi(0)$ and is aligned with $\hat{\mathbf{a}}_3$. The geometric scheme is portrayed in Fig. 2. Along the normal elliptic cross section $\Gamma_\nu(\mathbb{R}, Y_3)$:

$$I_\nu = I_\nu(\mathbb{R}, Y_3, t). \tag{9.18}$$

Thus, along a specified elliptic cylinder $C_\nu(\mathbb{R})$ (derived from a specified ellipsoidal member $\xi_\nu(\mathbb{R})$), $\mathbf{I}_\nu$ varies only with t and the scaled distance Y_3 [determined by the variable distance plane $\Pi(Y_3)$ and] measured along the coaxial direction $\hat{\mathbf{a}}_3$. On the other hand, for a specified Y_3 (corresponding to a specified $\Pi(Y_3)$), I_ν varies only with t and $\mathbb{R}$; here, $\mathbb{R}$ ranges over the family $\{\xi_\nu(\mathbb{R})\}$; of elliptic cylinders derived from the ellipsoidal family $\{\xi_\nu(\mathbb{R})\}$; $\hat{\mathbf{a}}_3$ is the invariant coaxial direction for the two families $\{C_+(\mathbb{R})\}$ and $\{C_-(\mathbb{R})\}$. We infer: I_ν possesses, inside ξ_ν, an elliptical axial symmetry about the direction $\hat{\mathbf{a}}_3$. Moreover, inside ξ_- there exists, about the direction $\hat{\mathbf{a}}_3$, a pair of elliptical coaxial symmetries associated with I_+ and I_-. Observe that at any instant,

$$I_+ \text{ along } \Gamma_+(\mathbb{R}, Y_3) = I_- \text{ along } \Gamma_-(\mathbb{R}, Y_3). \tag{9.19}$$

It now appears certain that the previous availability (Secs. 7 and 8) of closed-form results with strikingly different features, e.g., the oblique ellipsoidal symmetries, is largely due to an application of the particular $\mathscr{L}$-operator of (7.5).

It is easily verified from (9.12) and (9.14) that $\xi_+(\mathbb{R})$ meets $\xi_-(\mathbb{R})$ if and only if

$$\mathbf{x}\mathbf{a}_1^T = 0 = \mathbf{x}\mathbf{a}_2^T, \quad \mathbf{x}\mathbf{a}_3^T = \pm\lambda_+\mathbb{R}, \tag{9.20}$$

precisely, at the two radially opposite stationary points

$$\mathbf{x} = \pm\lambda_+\mathbb{R}(0,0,1)\boldsymbol{\mu}^{1/2} = \pm\lambda_+\mathbb{R}\mathbf{b}_3. \tag{9.21}$$

Here, $\mathbf{b}_j$ denotes the jth row of $\boldsymbol{\mu}^{1/2}$. These two meetings occur between and along the same diametric line as the two retreating contacts expressed by (7.23). To achieve a meeting, the vector $\mathbf{x}$ must obviously become parallel to $\mathbf{b}_3$ but orthogonal to both $\mathbf{a}_1$ and $\mathbf{a}_2$. This is, of course, compatible with the fact that $\mathbf{b}_j\mathbf{a}_k^T = \delta_{jk}$ (the Kronecker delta). Now,

$$\nabla|\mathbf{Y}_\nu|^2 = 2(\det\boldsymbol{\mu}) \times [\mathbf{a}_1(\mathbf{x}\mathbf{a}_1^T)\lambda_\nu^{-2} + \mathbf{a}_2(\mathbf{x}\mathbf{a}_2^T)\lambda_\nu^{-2} + \mathbf{a}_3(\mathbf{x}\mathbf{a}_3^T)\lambda_+^{-2}]. \tag{9.22}$$

Consequently, at each of the meetings represented by (9.20) or (9.21) as well as by (7.23), $\xi_+(\mathbb{R})$, $\xi_-(\mathbb{R})$, ξ_+ and ξ_- possess the common unit normal $\pm\hat{\mathbf{a}}_3$ which is parallel to the invariant coaxis of $\{C_+(\mathbb{R})\}$ and $\{C_-(\mathbb{R})\}$ (see Fig. 2). Specifically, as in the case of ξ_+ and ξ_-, each meeting between $\xi_+(\mathbb{R})$ and $\xi_-(\mathbb{R})$ is tangential.

We shall conclude with some brief remarks on deactivation. From (5.1), (9.6), (9.7), and the fact that the deactivation-induced ellipsoid ξ_ν^0 expands inside ξ_ν, we deduce

$$I_\nu \equiv 0 \ \forall \mathbf{x} \in \text{int}\xi_\nu^0 \text{ but } \neq \mathbf{0} \tag{9.23}$$

after deactivation. Throughout this period, then, there is again, inside ξ_-^0 and about the particle charge, a dark core in which $\nabla\cdot\mathbf{W} \equiv 0$. Likewise, in view of (5.1) and (9.9), the pulsatory charge corresponding to (5.15) generates, before deactivation, in-phase modes with real or complex frequency ω inside ξ_-.

ACKNOWLEDGMENTS

It is a pleasure to thank Professors Jorge Ize, Antonmaria Minzoni, and Carlos Vargas for the very kind hospitality extended to me during the brief tenure of a visiting appointment at the IIMAS, Universidad Nacional Autónoma de México. I am also indebted to Professor Tomás Garza, Director: IIMAS-UNAM for approving my visit.

[1] L. Chee-Seng, J. Math. Phys. **21**, 189 (1980).
[2] L. Chee-Seng, J. Math. Phys. **21**, 195 (1980).
[3] G. Johannsen, J. Math. Phys. **13**, 78 (1972).

[4]R. Handelsman and R. M. Lewis, J. Math. Phys. **7**, 1982 (1966).
[5]M. Kline and I. W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965).
[6]I. M. Besieris J. Math. Phys. **10**, 1156 (1969).
[7]S. D. Majumdar and R. Pal. Proc. Roy. Soc. Lond. A **316**, 525 (1970).
[8]S. J. Lewandowski, J. Phys. A: Gen. Phys. **4**, 197 (1971).
[9]L. B. Felsen, Proc. IEEE **111**, 445 (1964).
[10]L. B. Felsen, Proc. IEEE **111**, 454 (1964).
[11]S. W. Lee and Y. T. Lo, J. Appl. Phys. **38**, 870 (1967).
[12]J. A. Kong, J. Appl. Phys. **41**, 554 (1970).
[13]K. K. Mei, IEEE Trans. Antennas Propagation **19**, 665 (1971).
[14]H. S. Lu and K. K. Mei, Alta Freq. N. 4 Vol. **XLI, 290** (1972).

Analytical solution for the change of Raman transitions quantum statistics

P. F. González-Díaz
Instituto de Optica "Daza de Valdés," C.S.I.C., Serrano, 121, Madrid-6, Spain

(Received 17 March 1981; accepted for publication 29 May 1981)

By using the Laplace transform method we give exact analytical solutions for the time development of the diagonal and off-diagonal elements of the photon density matrix in Raman transitions.

PACS numbers: 42.50. + q,33.20.Fb,03.20. + i

In recent publications[1-3] the change of photon statistics in multiphoton absorption has been considered. Two different approaches have been thus proposed to obtain an exact solution of the corresponding master equation. Zubairy and Yeh[1] and Zubairy[2] used a density matrix procedure while Voigt, Bandilla, and Ritze[3] used a Laplace transform method.

The photon statistics of the stimulated and hyper Raman effects was dealt with early by Simaan.[4] The purpose of this paper is to give an exact analytical solution for the time development of the diagonal and off-diagonal elements of the radiation density matrix for Raman transitions (simultaneous absorption and emission) between localized states. Here, one photon is absorbed while the other is emitted. The case for antistokes luminescence (i.e., two photons being absorbed and one emitted) will be studied in a following paper.

The interaction Hamiltonian for Raman transitions is[5]

$$H_I = \sum_i [d_R c_{2i}^+ c_{i1} E_k^-(r_i) E_q^+(r_i) + \text{adj}], \tag{1}$$

where c_{2i}^+ and c_{1i} are creation and annihilation operators for the ith atom in states 2 and 1, respectively, d_R is the electric dipole matrix element for Raman transition, and $E_k^-(r_i)$ and $E_q^+(r_i)$ are the negative- and positive-frequency parts of, respectively, the input and output fields

$$E_k^-(r_i) = -i(2\pi\hbar\omega_k)^{1/2} u_k(r_i) a_k,$$
$$E_q^+(r_i) = i(2\pi\hbar\omega_q)^{1/2} u_q^*(r_i) a_q^+.$$

Here, ω is frequency, and a_k and a_q^+ are the creation and annihilation operators for the k th and qth modes, respectively. The spatial functions $u_k(r_i)$ and $u_q(r_i)$ are normalized functions for the respective mode in the ith atom.

In the single mode case, the time evolution of the photon density matrix becomes[5]

$$\dot{\rho}_{ph} = -\beta_R [(a_k^+ a_q a_k a_q^+ \rho_{ph} - 2a_k a_q^+ \rho_{ph} a_k^+ a_q + \rho_{ph} a_k^+ a_q a_k a_q^+) p_{1A}^\circ + (a_k a_q^+ a_k^+ a_q \rho_{ph} - 2a_k^+ a_q \rho_{ph} a_k a_q^+ + \rho_{ph} a_k a_q^+ a_k^+ a_q) p_{2A}^\circ], \tag{2}$$

where p_{1A}° and p_{2A}° are the thermal populations for the ground and first-excited states and the remaining symbols have their conventional meaning.[5]

The diagonal matrix elements of the photon density operator in the Fock representation is immediately obtained from (2):

$$\dot{P}_{l,m}(T) = -P_{l,m}(T)[l(m+1)p_{1A}^\circ + m(l+1)p_{2A}^\circ] + P_{l+1,m-1}(T) m(l+1) p_{1A}^\circ + P_{l-1,m+1}(T) l(m+1) p_{2A}^\circ, \tag{3}$$

where $T = 2\beta_R t$ and $P_{l,m}(T) = \langle l,m|\rho_{ph}|l,m\rangle$ is the probability to simultaneously find l photons in the input mode k and m photons in the Stokes output mode q. Then, according to the method given in Ref. 3, we obtain

$$\mathcal{L}_{l,m}(s) = \int_0^\infty P_{l,m}(T) \exp(-sT)\, dT = \frac{p_{1A}^\circ m(l+1)\mathcal{L}_{l+1,m-1}(s) + p_{2A}^\circ l(m+1)\mathcal{L}_{l-1,m+1}(s) + P_{l,m}(0)}{s + p_{1A}^\circ l(m+1) + p_{2A}^\circ m(l+1)}. \tag{4}$$

If we choose as initial conditions

$$\sum_{l=0}^\infty \sum_{m=0}^\infty P_{l,m}(0) = 1,$$

and $P_{r,z}(0) = 0$ for either $r > l, z > m$, or both (where $m > 0$ only for the case in which $P_{2A}^\circ \neq 0$, so that there may be spontaneous emission of $q$the mode photons), we find

$$\mathcal{L}_{n,t}(s) = \frac{P_{n,t}(0)}{s + p_{1A}^\circ n(t+1) + p_{2A}^\circ t(n+1)} \tag{5}$$

so that, for single Raman transitions

 0022-2488/81/112689-03$01.00

$$\left[\frac{p_{1A}^{\cdot} n(t+1)P_{n,t}(0)}{s+p_{1A}^{\cdot} n(t+1)+p_{2A}^{\cdot} t(n+1)}+\frac{p_{2A}^{\cdot}(n-1)(t+2)P_{n-2,t+2}(0)}{s+p_{1A}^{\cdot}(n-2)(t+3)+p_{2A}^{\cdot}(n-1)(t+2)}+P_{n-1,t+1}(0)\right]$$
$$\times\left[s+p_{1A}^{\cdot}(n-1)(t+2)+p_{2A}^{\cdot} n(t+1)\right]^{-1}=\mathscr{L}_{n-1,t+1}(s), \tag{6}$$

which can be eventually generalized into

$$\mathscr{L}_{n-j,t+k}(s)=\sum_{r=0}^{j}\sum_{v=0}^{k}P_{n-j+r,t+k-v}(0)\times\prod_{a=0}^{r}\prod_{b=0}^{v}\frac{p_{1A}^{\cdot}(n-j+a)(t+k+1-b)}{p_{1A}^{\cdot}(n-j)(t+k+1)+p_{2A}^{\cdot}(t+k)(n-j+1)}$$
$$\times\left[s+p_{1A}^{\cdot}(n-j+a)(t+k+1-b)+p_{2A}^{\cdot}(n-j+1+a)(t+k-b)\right]^{-1}$$
$$+\sum_{r=-j}^{0}\sum_{v=-k}^{0}P_{n-j+r,t+k-v}(0)\prod_{a=-r}^{0}\prod_{b=-v}^{0}\frac{p_{2A}^{\cdot}(n-j+a+1)(t+k-b)}{p_{1A}^{\cdot}(n-j)(t+k+1)+p_{2A}^{\cdot}(t+k)(n-j+1)}$$
$$\times\left[s+p_{1A}^{\cdot}(n-j+a)(t+k+1-b)+p_{2A}^{\cdot}(n-j+1+a)(t+k-b)\right]^{-1}, \tag{7}$$

where it has been considered the most general case in which the number of emitted Stokes photons may not be the same as the number of absorbed input photons due to a variety of possible phenomena such as self-induced stimulated emission, nonradiative relaxation, etc. If all of these additional effects are disregarded, then we should make $k=j$, $r=v$, and $a=b$, obtaining the pure Raman transition equation.

Now, following the method of Ref. 3, we transform Eq. (7) back and obtain,[6] for the most general case,

$$P_{n-j,t+k}(T)=\sum_{r=0}^{j}\sum_{v=0}^{k}\prod_{a=0}^{r}\prod_{b=0}^{v}p_{1A}^{\cdot}(n-j+a)(t+k+1-b)P_{n-j+r,t+k-v}(0)$$
$$\times\sum_{\alpha=0}^{r}\sum_{\beta=0}^{v}\left[p_{1A}^{\cdot}(n-j)(t+k-1)+p_{2A}^{\cdot}(t+k)(n-j+1)\right]^{-1}$$
$$\times\exp\{-\left[p_{1A}^{\cdot}(n-j+\alpha)(t+k+1-\beta)+p_{2A}^{\cdot}(n-j+\alpha+1)(t+k-\beta)\right]T\}$$
$$\times\left\{\Pi_{\substack{\nu=0\\ \nu\neq\alpha}}^{r}\Pi_{\substack{\omega=0\\ \omega\neq\beta}}^{v}\left[p_{1A}^{\cdot}(n-j+\nu)(t+k+1-\omega)\right.\right.$$
$$\left.\left.+p_{2A}^{\cdot}(n-j+\nu+1)(t+k-\omega)-p_{1A}^{\cdot}(n-j+\alpha)(t+k+1-\beta)-p_{2A}^{\cdot}(n-j+1+\alpha)(t+k-\beta)\right]\right\}^{-1}$$
$$+\sum_{r=-j}^{0}\sum_{v=-k}^{0}\prod_{a=-r}^{0}\prod_{b=-v}^{0}p_{2A}^{\cdot}(n-j+a+1)(t+k-b)P_{n-j+r,t+k-v}(0)$$
$$\times\sum_{\alpha=-r}^{0}\sum_{\beta=-v}^{0}\left[p_{1A}^{\cdot}(n-j)(t+k-1)+p_{2A}^{\cdot}(t+k)(n-j+1)\right]^{-1}$$
$$\times\exp\{-\left[p_{1A}^{\cdot}(n-j+\alpha)(t+k+1-\beta)+p_{2A}^{\cdot}(n-j+\alpha+1)(t+k-\beta)\right]T\}$$
$$\times\left\{\prod_{\substack{\nu=-r\\ \nu\neq\alpha}}^{0}\prod_{\substack{\omega=-v\\ \omega\neq\beta}}^{0}\left[p_{1A}^{\cdot}(n-j+\nu)(t+k+1-\omega)+p_{2A}^{\cdot}(n-j+\nu+1)(t+k-\omega)\right.\right.$$
$$\left.\left.-p_{1A}^{\cdot}(n-j+\alpha)(t+k+1-\beta)-p_{2A}^{\cdot}(n-j+1+\alpha)(t+k-\beta)\right]\right\}^{-1}, \tag{8}$$

where we also have $j=k$, $r=v$, $a=b$, $\alpha=\beta$, and $\nu=\omega$ for pure Raman transitions.

Substituting first $n-j=n'$, $t+k=t'$, making then $n\to\infty$, $t\to\infty$, and omitting finally the primes, one gets

$$P_{n,t}(T)=\sum_{r=0}^{\infty}\sum_{v=-\infty}^{0}P_{n+r,t-v}(0)\sum_{\alpha=0}^{r}\sum_{\beta=v}^{0}p_{1A}^{\cdot}(n+\alpha)(t-\beta+1)\frac{V(n,t,\alpha,\beta,T)}{\Pi_{\substack{\nu=0\\ \nu\neq\alpha}}^{r}\Pi_{\substack{\omega=v\\ \omega\neq\beta}}^{0}W(n,t,\alpha,\beta,\nu,\omega)}$$
$$+\sum_{r=-\infty}^{0}\sum_{v=0}^{\infty}P_{n+r,t-v}(0)\sum_{\alpha=r}^{0}\sum_{\beta=0}^{v}p_{2A}^{\cdot}(n+\alpha+1)(t-\beta)\frac{V(n,t,\alpha,\beta,T)}{\Pi_{\substack{\nu=0\\ \nu\neq\alpha}}^{0}\Pi_{\substack{\omega=v\\ \omega\neq\beta}}^{v}W(n,t,\alpha,\beta,\nu,\omega)}, \tag{9}$$

$$V(n,t,\alpha,\beta,T)=\frac{\exp\{-\left[p_{1A}^{\cdot}(n+\alpha)(t-\beta+1)+p_{2A}^{\cdot}(n+\alpha+1)(t-\beta)\right]T\}}{p_{1A}^{\cdot} n(t+1)+p_{2A}^{\cdot} t(n+1)},$$
$$W(n,t,\alpha,\beta,\nu,\omega)=p_{1A}^{\cdot}\left[(n+\nu)(t+1-\omega)-(n+\alpha)(t+1-\omega)\right]+p_{2A}^{\cdot}\left[(n+\nu+1)(t-\omega)-(n+\alpha+1)(t-\beta)\right].$$

For calculating the off-diagonal density matrix elements in the Fock representation, we operate from the master equation (2) and obtain

$$\dot{P}_{l,m;l',m'}(T) = \langle l,m|\dot{\rho}_{\rm ph}|l',m'\rangle$$
$$= -P_{l,m;l',m'}(T)\{p_{1A}^{\cdot}[l(m+1)+l'(m'+1)]+p_{2A}^{\cdot}[m(l+1)+m'(l'+1)]\}$$
$$+P_{l+1,m-1;l'+1,m'-1}(T)[p_{1A}^{\cdot}m(l+1)+p_{2A}^{\cdot}l'(m'+1)]$$
$$+P_{l-1,m+1;l'-1,m'+1}(T)[p_{1A}^{\cdot}m'(l'+1)+p_{2A}^{\cdot}l(m+1)] \tag{10}$$

and

$$\mathscr{L}_{l,m;l',m'}(s) = [p_{1A}^{\cdot}m(l+1)+p_{2A}^{\cdot}l'(m'+1)]\mathscr{L}_{l+1,m-1;l'+1,m'-1}(s)$$
$$+[p_{1A}^{\cdot}m'(l'+1)+p_{2A}^{\cdot}l(m+1)]\mathscr{L}_{l-1,m+1;l'-1,m'+1}(s)+P_{l,m;l',m'}(0)$$
$$\times\{s+p_{1A}^{\cdot}[l(m+1)+l'(m'+1)]+p_{2A}^{\cdot}[m(l+1)+m'(l'+1)]\}^{-1} \tag{11}$$

Imposing now initial conditions similar to those for the diagonal matrix elements, we get

$$\mathscr{L}_{n-j,t+k;n'-j',t'+k'}(s) = \sum_{r=0}^{j}\sum_{v=0}^{k}\sum_{r'=0}^{j'}\sum_{v'=0}^{k'}P_{n-j+r,t+k-v;n'-j'+r',t'+k'-v'}(0)$$
$$\times\prod_{a=0}^{r}\prod_{b=0}^{v}\prod_{a'=0}^{r'}\prod_{b'=0}^{v'}[p_{1A}^{\cdot}(n-j+a)(t+k+1-b)+p_{2A}^{\cdot}(n'-j'+1+a')(t'+k'-b')]R(a,b;a',b')^{-1}$$
$$+\sum_{r=-j}^{0}\sum_{v=-k}^{0}\sum_{r'=-j'}^{0}\sum_{v'=-k'}^{0}P_{n-j+r,t+k-v;n'-j'+r',t'+k'-v'}(0)$$
$$\times\prod_{a=-r}^{0}\prod_{b=-v}^{0}\prod_{a'=-r'}^{0}\prod_{b'=-v'}^{0}[p_{1A}^{\cdot}(n'-j'+a')(t'+k'+1-b')+p_{2A}^{\cdot}(n-j+1+a)(t+k-b)]R(a,b;a',b')^{-1}, \tag{12}$$

where

$$R(a,b;a',b') = p_{1A}^{\cdot}[(n-j)(t+k-1)+(n'-j')(t'+k'-1)]+p_{2A}^{\cdot}[(n-j+1)(t+k)+(n'-j'+1)(t'+k')]$$
$$\times\{s+p_{1A}^{\cdot}[(n-j+a)(t+k+1-b)+(n'-j'+a')(t'+k'+1-b')]$$
$$+p_{2A}^{\cdot}[(n-j+1+a)(t+k-b)+(n'-j'+1+a')(t'+k'-b')]\}^{-1}.$$

Following the same procedure as for the diagonal matrix elements, one finally obtains

$$P_{n,t;n',t'}(T) = \sum_{r=0}^{\infty}\sum_{v=-\infty}^{0}\sum_{r'=0}^{\infty}\sum_{v'=-\infty}^{0}P_{n+r,t-v;n'+r',t'-v'}(0)$$
$$\times\sum_{\alpha=0}^{r}\sum_{\beta=v}^{0}\sum_{\alpha'=0}^{r'}\sum_{\beta'=v'}^{0}\frac{[p_{1A}^{\cdot}(n+\alpha)(t+1-\beta)+p_{2A}^{\cdot}(n'+1+\alpha')(t'-\beta')]\exp[-B(\alpha,\beta;\alpha',\beta')T]}{A(n,t;n',t')\prod_{\nu\neq\alpha}^{r}\prod_{\omega\neq\beta}^{0}\prod_{\nu'\neq\alpha'}^{r'}\prod_{\omega'\neq\beta'}^{0}C(\alpha,\beta,\nu,\omega;\alpha',\beta',\nu',\omega')}$$
$$+\sum_{r=-\infty}^{0}\sum_{v=0}^{\infty}\sum_{r'=-\infty}^{0}\sum_{v'=0}^{\infty}P_{n+r,t-v;n'+r',t'-v'}(0)$$
$$\times\sum_{\alpha=r}^{0}\sum_{\beta=0}^{v}\sum_{\alpha'=r'}^{0}\sum_{\beta'=0}^{v'}\frac{[p_{1A}^{\cdot}(n'+\alpha')(t'+1-\beta')+p_{2A}^{\cdot}(n+1+\alpha)(t-\beta)]\exp[-B(\alpha,\beta;\alpha',\beta')T]}{A(n,t;n',t')\prod_{\nu\neq\alpha}^{0}\prod_{\omega\neq\beta}^{v}\prod_{\nu'\neq\alpha'}^{0}\prod_{\omega'\neq\beta'}^{v'}C(\alpha,\beta,\nu,\omega;\alpha',\beta',\nu',\omega')}, \tag{13}$$

where

$$A(n,t;n't') = p_{1A}^{\cdot}[n(t+1)+n'(t'+1)]+p_{2A}^{\cdot}[t(n+1)+t'(n'+1)],$$
$$B(\alpha,\beta;\alpha',\beta') = p_{1A}^{\cdot}[(n+\alpha)(t+1-\beta)+(n'+\alpha')(t'+1-\beta')]+p_{2A}^{\cdot}[(n+1+\alpha)(t-\beta)+(n'+1+\alpha')(t'-\beta')],$$
$$C(\alpha,\beta,\nu,\omega;\alpha',\beta',\nu',\omega') = B(\nu,\omega;\nu',\omega')-B(\alpha,\beta;\alpha',\beta').$$

Expressions (9) and (13) are our most general analytical solutions. The pure Raman transition case may be straightforwardly obtained by making in these expressions $r=v$, $a=b$, $\alpha=\beta$, $\nu=\omega$, and likewise for the primed indices in the off-diagonal case. Spontaneous emission of photons could be omitted if it is assumed that all the two-level systems are maintained in their ground states by some external influence, so that $P_{2A}^{\cdot}=0$. The well-known[7] dependence of the yield of the Raman process on the statistical properties of the input field may be shown from Eqs. (9) and (13).

[1]M. S. Zubairy and J. J. Yeh, Phys. Rev. A **21**, 1624 (1980).
[2]M. S. Zubairy, J. Math. Phys. **21**, 2690 (1980).
[3]H. Voigt, A. Bandilla, and H.-H. Ritze, Z. Phys. B **36**, 295 (1980).
[4]H. D. Simaan, J. Phys. A **8**, 1620 (1975).
[5]Y. R. Shen, Phys. Rev. **155**, 921 (1967).
[6]*Tables of Integral Transforms*, edited by A. Erdélyi (McGraw–Hill, New York, 1954).
[7]R. W. Hellwarth, Phys. Rev. **130**, 1850 (1963).

Linear wave conversion in a warm, unmagnetized, collisionless plasma

Gregory A. Kriegsmann
Department of Engineering Sciences and Applied Mathematics, The Technological Institute, Northwestern University, Evanston, Illinois 60201

(Received 11 December 1980; accepted for publication 10 April 1981)

In this paper we apply the method of matched asymptotic expansions to study the problem of linear mode conversion in a warm, unmagnetized, collisionless plasma. The plasma is assumed to fill the half space $x > 0$. At the boundary, $x = 0$, the transition from free space to plasma is very smooth. Two linear second-order ordinary differential equations govern the electric field within the plasma. Since a very small parameter (ϵ^2) multiplies the highest derivative in one equation, the mathematical problem is singular and not unlike the Orr–Sommerfeld equation. The method of multiple scales decouples the equations which allows us to determine the transverse and longitudinal waves. These results break down in a small neighborhood of the critical point where we develop new expansions. The composite results are computed to second order in ϵ.

PACS numbers: 52.35.Hr, 52.40.Db

1. INTRODUCTION

The propagation of electromagnetic waves through a warm plasma has received a great deal of study during the past twenty years. In his classic work, Denisov[1] formed a model of this important problem which is still being used today. Originally researchers were interested in the propagation of radio waves through the ionosphere.[2] More recently, his model has been used to describe resonant absorption or linear wave conversion in connection with laser produced plasmas.[3,4]

Previous studies of Denisov's model have always assumed the plasma to be infinite in extent[3–5] and excited by a plane wave at infinity. This circumnavigates the problem of a plasma-free space boundary. In these studies assumptions are made regarding the structure of the field components near the critical surface where the highly oscillatory plasma waves are excited. They are roughly equivalent to a "boundary layer analysis" of the governing equations near the critical point (see Sec. 4). The results of this analysis are asymptotically evaluated near the critical surface where they yield a simple physical interpretation: the field is the sum of longitudinal (plasma) and transverse (electromagnetic) waves. Although the region of validity is small, the incident wave is "matched up" with the transverse wave. This determines the amplitude of the plasma oscillations and the absorption coefficient. The structure of the fields away from the critical surface is undetermined.

In this paper we are concerned with a slightly different physical situation and problem. We assume that the warm plasma fills the half space $x \geqslant 0$. The plasma density is taken to be smooth across the plasma-free space boundary $x = 0$. This model is a first cut at understanding the scattering and linear wave conversion process in a warm plasma pellet. In this spirit we take $k = k'L$ to be order one in size, where $x = L$ is the critical plane and k' is the wave number of the incident radiation. This corresponds to a plasma with a very steep density profile.[6] Finally, we assume that the parameter $\epsilon^2 = 3T_e/mv^2$ is very small, where T_e is the electron temperature, m is the electronic mass, and v is the speed of light in vacuo.

When these assumptions are inserted into Denisov's model we are forced into solving two coupled second-order ordinary differential equations. The small parameter ϵ^2 multiplies the second derivative of a field component in one equation yeilding a singular perturbation problem. If these equations were decoupled, the ensuing fourth-order equation would be similar to the Orr–Sommerfeld equation.[7] However, it contains singular coefficients. If these were not present, the results of Meksyn[8] could be applied directly. None the less, the technique used by Meksyn (the WKB[10] method) works nicely on the problem at hand. In this paper we shall use the equivalent method of multiple scales and matched asymptotic expansions to study Denisov's equations.

We now present a brief outline of this paper. In Sec. 1 the mathematical problem is carefully derived and stated. In Sec. 2 the outer solutions (valid away from the critical surface $x = L$) are obtained to second order and compared with the cold-plasma theory ($\epsilon^2 = 0$). In Sec. 3 a boundary layer analysis is performed at the critical surface and the local equations are derived. They are identical to those given in Refs. 3 and 4 when the angle of incidence isn't too small. In Sec. 5 the boundary layer solutions are matched to the outer solutions. This process shows that the plasma wave is $O(1/\sqrt{\epsilon})$ in amplitude which indicates that the electrons are driven near resonance by the incident electromagnetic field. The largeness of this field may be important in producing parametric instabilities.[4]

Another type of resonance occurs at discrete values of the scaled wave number k. At these values k_N, the amplitude of the plasma wave becomes singular. These are the same eigenvalues that occur in the separated problem when $\alpha = 0$ (normal incidence). We observe the interesting fact that the k_N are spaced closely together. This resonance is not present in an infinite plasma; it is a direct consequence of the free space-plasma boundary.

Finally, we discuss the limit $\alpha \to 0$ for a fixed ϵ. Our asymptotic solution possesses a singularity in this limit. We describe the cause of this difficulty and conjecture about its removal.

 0022-2488/81/112692-07$01.00

2. FORMULATION

A plane electromagnetic wave impinges upon a warm overdense plasma slab and scatters from it. The plasma fills the half space $x>0$ where the unperturbed electron density $N_0(x)$ and its derivatives depend continuously upon x. These assumptions lead naturally to the following hypotheses regarding the index of refraction $n(x)$:

$$\text{(H1) } n(xL)=1,\quad \frac{d^r}{dx^r}n(xL)=0,\quad x\leqslant 0,\quad r=1,2,3,\ldots,$$

$$\text{(H2) } 0\leqslant n(xL)\leqslant 1,\quad 0\leqslant x\leqslant 1,\quad n(L)=0,$$

$$\text{(H3) } n(xL)\leqslant 0,\quad x\geqslant 1.$$

The argument xL implies that the independent spatial variables were scaled with respect to the distance L. Conditions H1–H3 depend upon the implicit assumption that the incident plane wave's frequency (ω) is large enough to neglect ionic motion.

According to the classic work of Denisov,[1] the electric field $\mathbf{E}e^{-i\omega t}$ (in Cartesian coordinates) satisfies

$$\nabla^2\mathbf{E}+k^2\mu(x)\mathbf{E}=[1-\epsilon^2H(x)]\nabla(\nabla\cdot\mathbf{E}), \tag{2.1}$$

where $k^2=\omega^2L^2/v^2=O(1)$, $\mu(x)=n^2(xL)$, $H(x)$ is the Heaviside function, $\epsilon^2\ll 1$, and ω is the frequency of the incident plane wave. Exterior to the plasma ($x<0$) the field takes the form

$$\mathbf{E}=e^{ik(x\cos\alpha+y\sin\alpha)}(-\sin\alpha\hat{i}+\cos\alpha\hat{j}) + e^{ik(-x\cos\alpha+y\sin\alpha)}(A_x\hat{i}+A_y\hat{j}), \tag{2.2}$$

which is the sum of an obliquely incident and reflected wave. The unknown constants A_x and A_y determine the amplitude of the scattered field. Within the plasma ($x>0$) the *ansatz*

$$\mathbf{E}=[u(x)\hat{i}+v(x)\hat{j}]e^{iky\sin\alpha} \tag{2.3}$$

describes the electric field when $u(x)$ and $v(x)$ satisfy

$$\epsilon^2u_{xx}+k^2(\mu-\sin^2\alpha)u=ik(1-\epsilon^2)\sin\alpha v_x,$$

$$v_{xx}+k^2\mu v=ik(1-\epsilon^2)\sin\alpha u_x+\epsilon^2k^2\sin^2\alpha v. \tag{2.4}$$

The subscripts denote differentiation with respect to x.

Since the system (2.4) is fourth order, it requires four pieces of boundary data to determine a unique solution. The assumption of a bounded electric field as $x\to\infty$ requires [see Sec. 3, Eqs. (3.37) and (3.38)] that

$$\lim_{x\to\infty}u(x)=\lim_{x\to\infty}v(x)=0. \tag{2.5}$$

The derivation of the remaining two conditions necessitates rewriting (2.1) as

$$\nabla\times\mathbf{E}=k\mathbf{B},$$

$$\nabla\times\mathbf{B}=k\mu\mathbf{E}+(\epsilon^2/k)H(x)\nabla(\nabla\cdot\mathbf{E}), \tag{2.6}$$

where $\mathbf{B}$ is the magnetic field. It then follows by familiar arguments[9] that the tangential components of the electric and magnetic fields are continuous across $x=0$. Consequently the following relations hold:

$$v(0)=\cos\alpha+A_y, \tag{2.7}$$

$$v_x(0)-ik\sin\alpha u(0)=ik(1-\cos\alpha A_y-\sin\alpha A_x). \tag{2.8}$$

Taking the divergence of $\nabla\times\mathbf{B}$ in (2.6) gives

$$\rho=0,\quad x<0, \tag{2.9}$$

$$k^2(\nabla\cdot\mu\mathbf{E})+\epsilon^2\nabla^2\rho=0,\quad x>0, \tag{2.10}$$

where $\rho=\nabla\cdot\mathbf{E}$ represents the time harmonic perturbed plasma density. Combining (2.2) and (2.9) yields the familiar result

$$\cos\alpha A_x=\sin\alpha A_y, \tag{2.11}$$

which states that the scattered electric field is perpendicular to the propagation direction, $(-\cos\alpha,\sin\alpha)$, of the reflected wave. [This result also follows from integrating (2.10) over a small cylinder perpendicular to the plane $x=0$.] Finally Eqs. (2.7), (2.8), and (2.11) simplify to

$$v_x(0)+ik\sec\alpha v(0)-ik\sin\alpha u(0)=2ik. \tag{2.12}$$

The boundary conditions (2.5) and (2.12) give only three out of four constraints needed to uniquely solve Eq. (2.4). Since all possible conditions on u and v at $x=0$ have been extracted from Maxwell's equations, the hypothesis

$$\text{(H4) } \lim_{x\to 0}\rho(x)=0$$

must be made to close the problem. This states that the perturbed electron density is continuous at $x=0$ and translates into the final boundary condition

$$u_x(0)+ik\sin\alpha v(0)=0. \tag{2.13}$$

This final hypothesis is consistent with the assumption of Denisov: $N_0(x)\gg\rho(x)$.

The mathematical problem of determining $\mathbf{E}$ for $|x|<\infty$ now becomes precise and well posed. The solution of Eqs. (2.4), (2.5), (2.12), and (2.13) determines the electric field within the slab ($x\geqslant 0$). Knowing the value of $v(x)$ at $x=0$ gives A_y from (2.7); the constant A_x then follows from (2.11). Thus Eq. (2.2) gives the scattered electric field in free space, $x<0$.

3. THE OUTER REGIONS

We shall now exploit the smallness of ϵ and seek an asymptotic approximation to the solutions of (2.4) as $\epsilon\to 0$. This is a natural setting for the methods of multiple scales and matched asymptotic expansions.[10] According to this procedure one assumes that

$$u(x,\epsilon)=U(\bar{x},x,\epsilon),$$

$$v(x,\epsilon)=V(\bar{x},x,\epsilon), \tag{3.1}$$

where the fast variable $\bar{x}$ is defined as

$$\bar{x}=\frac{1}{\epsilon}\int_0^x[\mu(s)]^{1/2}ds\quad 0<x<1. \tag{3.2}$$

This particular choice of a fast scale alleviates considerable algebraic difficulty in the ensuing calculations. As shall be seen, this particular dependence of u and v on x becomes invalid in a small neighborhood of $x=1$ where a different expansion is required (see Sec. 4). Moreover, a slightly different fast variable is required in the region $x>1$.

When Eqs. (3.1) and (3.2) are inserted into (2.4) we find that U and V satisfy

$$LU = \frac{ic}{\epsilon f} V_{\bar{x}} + \frac{ic}{\mu} V_x - \epsilon\left(\frac{2U_{\bar{x}x}}{f} + \frac{f'}{\mu} U_{\bar{x}} + \frac{ic}{f} V_{\bar{x}} \right) - \epsilon^2 (U_{xx} + icV_x)\mu^{-1}, \tag{3.3}$$

$$V_{\bar{x}\bar{x}} = \epsilon\left(\frac{ic}{f} U_{\bar{x}} - \frac{2}{f} V_{x\bar{x}} - \frac{f'}{\mu} V_{\bar{x}} \right) + \epsilon^2 [icU_x - G(V)]\mu^{-1} - ic\epsilon^3 f U_{\bar{x}} + O(\epsilon^4), \tag{3.4}$$

where $f = \sqrt{\mu}$, $c = k\sin\alpha$, $L = \partial^2/\partial\bar{x}^2 + k^2(\mu - \sin^2\alpha)/\mu$, $G = \partial^2/\partial x^2 + k^2\mu$, and the subscripts denote partial differentiation. Substituting (3.1) and (3.2) into the boundary conditions (2.5), (2.12), and (2.13) and noting that $\mu(0) = n(0) = 1$, the boundary conditions transform into

$$\lim_{x,\bar{x} \to \infty} U(\bar{x},x) = \lim_{x,\bar{x} \to \infty} V(\bar{x},x) = 0, \tag{3.5}$$

$$U_{\bar{x}}(0,0) = -\epsilon[U_x(0,0) + ik\sin\alpha V(0,0)], \tag{3.6}$$

$$V_{\bar{x}}(0,0) = \epsilon[-BV + ik\sin\alpha U(0,0) + 2ik], \tag{3.7}$$

where

$$BV = V_x(0,0) + ik\sec\alpha V(0,0).$$

We next assume that $U(\bar{x},x,\epsilon)$ and $V(\bar{x},x,\epsilon)$ have the asymptotic expansions

$$U(\bar{x},x,\epsilon) \sim \sum_{n=-1}^{\infty} \epsilon^{n/2} U_n(\bar{x},x), \tag{3.8}$$

$$V(\bar{x},x,\epsilon) \sim \sum_{n=0}^{\infty} \epsilon^{n/2} V_n(\bar{x},x), \tag{3.9}$$

as $\epsilon \to 0$. (These asymptotic forms are necessary for matching; see Sec. 5.) When (3.8) and (3.9) are inserted into (2.4) a sequence of coupled second-order differential equations is generated by equating to zero the coefficients of the powers of ϵ. A similar process yields a sequence of boundary conditions when (3.8) and (3.9) are substituted into (2.12) and (2.13). Since our primary concern is to describe the behavior of the leading terms, we only list the first five sets of equations and boundary conditions

$$V_{0,\bar{x}\bar{x}} = V_{0,\bar{x}} = 0,\ V_{0,\bar{x}}(0,0) = 0, \tag{3.10}$$

$$LU_{-1} = M(V_1,V_{-1}),\ U_{-1,\bar{x}}(0,0) = 0,$$

$$V_{1,\bar{x}\bar{x}} = M(U_{-1},U_{-3}),\ V_{1,\bar{x}}(0,0) = icU_{-1}(0,0), \tag{3.11}$$

$$LU_0 = M(V_2,V_0),\ U_{0,\bar{x}}(0,0) = 0,$$

$$V_{2,\bar{x}\bar{x}} = M(U_0,U_{-2}),$$

$$V_{2,\bar{x}}(0,0) = 2ik - BV_0 + icU_0(0,0), \tag{3.12}$$

$$LU_1 = M(V_3,V_1) - P(U_{-1}),\ U_{1,\bar{x}}(0,0) = -U_{-1,x}(0,0),$$

$$V_{3,\bar{x}\bar{x}} = M(U_1,U_{-1}) - P(V_1),$$

$$V_{3,\bar{x}}(0,0) = -BV_1 + icU_1(0,0), \tag{3.13}$$

$$LU_2 = M(V_4,V_2) - P(U_0),$$

$$U_{2,\bar{x}}(0,0) = -U_{0,x}(0,0) - icV_0(0,0),$$

$$V_{4,\bar{x}\bar{x}} = M(U_2,U_0) - P(V_2) - (1/\mu)G(V_0),$$

$$V_{4,\bar{x}}(0,0) = -BV_2 + icU_2(0,0), \tag{3.14}$$

where $U_{-3} = U_{-2} = V_{-1} = 0$, $M(\phi,\psi) = ic(\phi_{\bar{x}}/f + \psi_x/\mu)$, and $P(\phi) = (2/f)\phi_{\bar{x}x} + (f'/\mu)\phi_{\bar{x}}$. In solving Eqs. (3.10)–(3.14) secular terms of the form $\bar{x}$ and $\bar{x}e^{ik\bar{x}}$ are generated. These terms cause nonuniformities in (3.8) and (3.9); they are systematically removed by choosing the proper functional dependence of U_n and V_n on the slow variable x.

Equation (3.10) trivially gives the result

$$V_0 = V_0(x). \tag{3.15}$$

The system of differential equations given in (3.11) is ordinary with constant coefficients (x is a parameter). The characteristic equation for the system is

$$\lambda^2(\lambda^2 + k^2) = 0, \tag{3.16}$$

which yields solutions of the form $A(x)$, $\bar{x}B(x)$, and $C(x)e^{\pm ik\bar{x}}$. The solution $\bar{x}B(x)$ is secular and is suppressed. Thus a slight calculation yields

$$U_{-1} = A_{-1}(x)e^{ik\bar{x}} + B_{-1}(x)e^{-ik\bar{x}}, \tag{3.17}$$

$$V_1 = \frac{\sin\alpha}{f}[A_{-1}(x)e^{ik\bar{x}} - B_{-1}(x)e^{-ik\bar{x}}] + D_1(x). \tag{3.18}$$

The first boundary condition yields

$$A_{-1}(0) = B_{-1}(0), \tag{3.19}$$

while the second is satisfied identically [$f(0) = (\mu(0))^{1/2} = 1$]. Taking care to suppress the linear secular terms we obtain from the system (3.12) the solutions

$$U_0(\bar{x},x) = A_0(x)e^{ik\bar{x}} + B_0(x)e^{-ik\bar{x}} + \frac{i\sin\alpha}{k(\mu - \sin^2\alpha)} V_0'(x), \tag{3.20}$$

$$V_2(\bar{x},x) = \frac{\sin\alpha}{f}\{A_0(x)e^{ik\bar{x}} - B_0(x)e^{-ik\bar{x}}\} + D_2(x). \tag{3.21}$$

The first boundary condition again yields

$$A_0(0) = B_0(0), \tag{3.22}$$

while the second simplifies to

$$V_0'(0) + ik\cos\alpha V_0(0) = 2ik\cos^2\alpha. \tag{3.23}$$

Equations (3.13) take a little more work to solve. Differentiating the first equation by $\bar{x}$ and substituting for $V_{3\bar{x}\bar{x}}$ from the second yields

$$U_1''' + k^2 U_1' = (2k^2/f)[Q(A_{-1})e^{ik\bar{x}} + Q(B_{-1})e^{-ik\bar{x}}], \tag{3.24}$$

where $Q(\phi) = d\phi/dx + \frac{1}{2}(f'/f)\phi$. The term on the right in (3.24) induces secular terms in U_1 and therefore must be set to zero. The functions A_{-1} and B_{-1} are given by $A_{-1}(x) = B_{-1}(x) = a_{-1}/2(\mu)^{1/4}$ in view of (3.19). Inserting these results into (3.17) and (3.18) gives

$$U_{-1}(\bar{x},x) = a_{-1}(\mu)^{-1/4}\cos k\bar{x}, \tag{3.25}$$

$$V_1(\bar{x},x) = i\sin\alpha a_{-1}(\mu)^{-3/4}\sin k\bar{x} + D_1(x). \tag{3.26}$$

Continuing to solve (3.13) in a straightforward manner yields

$$U_1(\bar{x},x) = A_1(x)e^{ik\bar{x}} + B_1(x)e^{-ik\bar{x}} + \frac{i\sin\alpha D_1'(x)}{k(\mu - \sin^2\alpha)}, \tag{3.27}$$

$$V_3(\bar{x},x) = \frac{\sin\alpha}{f}\,[A_1 e^{ik\bar{x}} - B_1 e^{-ik\bar{x}}] + D_3(x) - \frac{3}{4}\,\frac{\sin\alpha\mu'}{k(\mu)^{9/4}}\,a_{-1}\cos k\bar{x}. \tag{3.28}$$

Since $(d^r\mu/dx^r)(0) = 0$ for $r \geqslant 1$, the boundary conditions (3.13) state

$$A_1(0) = B_1(0), \tag{3.29}$$

$$D_1'(0) + ik\cos\alpha\, D_1(0) = 0. \tag{3.30}$$

The system (3.14) is now solved in the same manner as (3.13). The equation for U_2 becomes slightly more complex

$$U_2''' + k^2 U_2' = (2k^2/f)[Q(A_0)e^{ik\bar{x}} + Q(B_0)e^{-ik\bar{x}}] - \frac{ic}{f(\mu - \sin^2\alpha)}\,L_0 V_0, \tag{3.31}$$

where

$$L_0 = \frac{d^2}{dx^2} - \frac{\sin^2\alpha\mu'}{\mu(\mu - \sin^2\alpha)}\,\frac{d}{dx} + k^2(\mu - \sin^2\alpha).$$

Suppressing the secular terms generated by the functions $e^{\pm ik\bar{x}}$ gives

$$A_0(x) = B_0(x) = \tfrac{1}{2}a_0(\mu)^{-1/4} \tag{3.32}$$

in view of (3.22). Since the L_0V_0 term on the right-hand side of (3.31) generates linearly secular terms, it too is set to zero

$$\frac{d^2V_0}{dx^2} - \frac{\sin^2\alpha\mu'}{\mu(\mu - \sin^2\alpha)}\,\frac{dV_0}{dx} + k^2(\mu - \sin^2\alpha)V_0 = 0. \tag{3.33}$$

Sparing the reader the details of yet another calculation we generate a new system of U_3 and V_5. The solution of these equations generates the solvability condition

$$L_0 D_1 = 0. \tag{3.34}$$

We now summarize the results up to this point

$$U \sim \frac{1}{\sqrt{\epsilon}}\,\frac{a_{-1}}{(\mu)^{1/4}}\cos k\bar{x} + \left(\frac{a_0}{\mu^{1/4}}\cos k\bar{x} + \frac{i\sin\alpha V_0'(x)}{k(\mu - \sin^2\alpha)}\right) + O(\epsilon), \tag{3.35}$$

$$v \sim V_0(x) + \sqrt{\epsilon}\left(\frac{i\sin\alpha a_{-1}}{\mu^{3/4}}\sin k\bar{x} + D_1(x)\right) + O(\epsilon), \tag{3.36}$$

where

$$\bar{x} = \frac{1}{\epsilon}\int_0^x (\sqrt{\mu})dx',\ L_0V_0 = L_0D_1 = 0, V_0' + ik\cos\alpha V_0(0) = 2ik\cos^2\alpha,$$

and $D_1'(0) + ik\cos\alpha D_1(0) = 0$.

There are several things worth noting about these asymptotic solutions. First, the equation and boundary condition for V_0 can be obtained directly from (2.4), (2.12), and (2.13) by setting $\epsilon = 0$. This procedure gives u as the term with V_0' as a factor in (3.35). These results duplicate the "cold" plasma equations.[2,3] It is well known that V_0 and $V_0'/(\mu - \sin^2\alpha)$ are well behaved at $x = x_\alpha$ where $\mu(x_\alpha) = \sin^2\alpha$. This is not the case at $x = 1$ where $\mu = 0$. Near this point V_0 becomes logarithmically singular while V_0' behaves like $(1 - x)^{-1}$. Moreover, the rapidly varying components of u and v (the plasma waves) are algebraically singular at $x = 1$. Thus the asymptotic solution given in (3.35) and (3.36) is invalid in a small region centered about $x - 1$. Finally, the *ansatz* (3.1) has decoupled the plasma and the electromagnetic waves.

Before analyzing the solutions in this "boundary layer region," we state the asymptotic approximations to u and v for $x > 1$. These results are found in exactly the same way as before except that the boundary conditions (3.5) must be applied. They are

$$u \sim \frac{1}{\sqrt{\epsilon}}\,\tilde{a}_{-1}\,\frac{e^{-k\xi}}{(q)^{1/4}} + \left(\frac{\tilde{a}_0}{q^{1/4}}\,e^{-k\xi} + \frac{i\sin\alpha\tilde{V}_0'}{k(\mu - \sin^2\alpha)}\right) + O(\sqrt{\epsilon}), \tag{3.37}$$

$$v \sim \tilde{V}_0(x) + \sqrt{\epsilon}\left(\frac{i\sin\alpha}{\mu^{3/4}}\,\tilde{a}_{-1}e^{-k\xi} + \tilde{D}_1(x)\right) + O(\epsilon), \tag{3.38}$$

where $\xi = \frac{1}{\epsilon}\int_1^x(\sqrt{q})dx'$, $q = -\mu$, $L_0\tilde{V}_0 = L_0\tilde{D}_1 = 0$,

$$\lim_{x,\bar{x}\to\infty}(\tilde{V}_0, \tilde{D}_1) = 0.$$

Exponentially growing terms were also obtained. However, the asumption that u and v are bounded as $x \to \infty$ dictated their removal. Thus, the boundary conditions (2.5) are justified.

4. THE BOUNDARY LAYER EQUATIONS

We now investigate the solutions of (2.4) in a small region centered about the critical surface $x = 1$. To this end we define the boundary layer variable

$$\eta = \Omega(1 - x)/\epsilon^{2/3}, \tag{4.1}$$

where $\Omega^3 = k^2|\mu'(1)|$ and $\eta = O(1)$ as $\epsilon \to 0$. We also assume that u and v have the following expansions:

$$u \sim u_{-1}(\eta)/\epsilon^{2/3} + u_0(\eta) + \epsilon^{2/3}u_1(\eta) + O(\epsilon^{4/3}\ln\epsilon), \tag{4.2}$$

$$v \sim C_{-1}\ln\epsilon + C_0 + v_0(\eta) + \epsilon^{2/3}v_1(\eta) + \epsilon^{4/3}v_2(\eta) + O(\epsilon^2) \tag{4.3}$$

as $\epsilon \to 0$, where C_{-1} and C_0 are constants. Inserting these expansions and (4.1) into (2.4) and equating to zero the coefficients of the various powers of ϵ yields an infinite system of equations. We state here the first six which are necessary to determine u and v to lowest order

$$P(u_{-1}, v_0) = 0, \tag{4.4}$$

$$\Omega^2 A(u_{-1}) = P(u_0, v_1), \tag{4.5}$$

$$\Omega^2 A(u_0) = P(u_1, v_2) + ic\Omega v_0'', \tag{4.6}$$

$$v_K'' = -\frac{ic}{\Omega}\,u_{K-1}', \quad K = 0,1,2, \tag{4.7}$$

where $P(u,v) = c^2u - ic\Omega v'$, $A(u) = u'' + \eta u$, and the primes denote differentiation with respect to η. These equations are valid as long as $\sin^2\alpha \gg \gamma\epsilon^{2/3}$, where $\gamma = |\mu'(1)|/\Omega$.

Equations (4.4) and (4.7) for $K = 0$ contain the same information. However, when (4.5) is combined with (4.7) for $K = 1$ we find that

$$u_{-1}'' + \eta u_{-1} = k_1/\pi, \tag{4.8}$$

where k_1 is a constant. This is the same equation obtained by

Pilyia.[5] Differentiating (4.6) and using (4.7) with $K = 2$ yields

$$u_0'' + \eta u_0 = k_2/\pi + (c^2/\Omega^2)u'_{-1} \tag{4.9}$$

after integration. The parameter k_2 is another constant of integration. The functions v_0 and v_1 are obtained from (4.7) by integration.

The solution of (4.8) is given by

$$u_{-1}(\eta) = k_1\psi_0(\eta), \tag{4.10}$$

where $\psi_0(\eta)$ is given by

$$\psi_0(\eta) = C_3A_i(-\eta) - I_2B_i(-\eta) + \int_\eta^\infty [B_i(-\eta)A_i(-t) - A_i(-\eta)B_i(-t)]\,dt. \tag{4.11}$$

The parameter C_3 is a constant, $I_2 = \int_{-\infty}^{\infty} A_i(-t)\,dt$, and the functions $A_i(-\xi)$, $B_i(-\xi)$ are the Airy functions. The choice of the constant I_2 implies

$$\psi_0 \sim \frac{1}{\pi\eta} + \frac{(C_3 - I_1)}{2(\sqrt{\pi})(-\eta)^{1/4}} - e^{-(2/3)\phi} \tag{4.12}$$

as $\eta \to \infty$, where $I_1 = \int_0^\infty B_i(-t)\,dt$ and $\phi = (-\eta)^{3/2}$. This behavior is needed for matching into the region $x > 1$ where exponentially small plasma waves are present. On the other hand as $\eta \to \infty$ we obtain

$$\psi_0 \sim \frac{1}{\pi\eta} + \frac{C_3}{(\sqrt{\pi})(\eta)^{1/4}}\sin\theta - \frac{I_2}{(\sqrt{\pi})(\eta)^{1/4}}\cos\theta, \tag{4.13}$$

where $\theta = \frac{2}{3}\eta^{3/2} + \pi/4$. Both (4.12) and (4.13) are obtained using the asymptotic expansions of the Airy functions.

The functions $v_0(\eta)$ is found from (4.7) with $K = 0$ and (4.10). It is given by

$$v_0 = \frac{-ic}{\Omega}\left\{\frac{k_1}{\pi}\ln\eta - k_1\int_\eta^\infty [\psi_0(t) - 1/\pi t]\,dt\right\}, \quad \eta > 0. \tag{4.14}$$

The integral converges at $t = \infty$ because of (4.13). The function is regular at $\eta = 0$ because the logarithms cancel out. When η is negative, the formula

$$v_0 = \frac{-ic}{\Omega}\left\{\frac{k_1}{\pi}\ln(-\eta) + k_1\int_{-\infty}^{\eta}\left[\psi_0(t) - \frac{1}{\pi t}\right]dt - k_1 J\right\}, \tag{4.15}$$

is more convenient to use. The constant J is given by

$$J = \int_1^\infty\left(\psi_0 - \frac{1}{\pi t}\right)dt + \int_{-\infty}^{-1}\left(\psi_0 - \frac{1}{\pi t}\right)dt + \int_{-1}^{1}\psi_0\,dt. \tag{4.16}$$

The solution of (4.9) is found by judicious guessing to be

$$u_0 = k_2\psi_0 + (k_1c^2/2\Omega^2)(\eta\psi_0 - 1/\pi). \tag{4.17}$$

When (4.10), (4.11), (4.17), and (4.14) are inserted into (4.2) and (4.3) the functional dependence of u and v on η is determined to $O(\epsilon^{2/3})$. Only the constants C_0, C_{-1}, k_1, k_2, and C_3 need to be determined. These constants along with those appearing in the "outer solutions," (3.35)–(3.38), will be determined in Sec. 5. There we shall need the following asymptotic expression of u and v as $\eta \to \pm\infty$:

$$u \sim (k_1\epsilon^{-2/3} + k_2)\left(\frac{1}{\pi\eta} + \frac{C_3\sin\theta - I_2\cos\theta}{\sqrt{\pi}(\eta)^{1/4}}\right) + \cdots \text{ as } \eta \to \infty, \tag{4.18}$$

$$v \sim \frac{-ic}{\Omega}(C_0 + C_{-1}\ln\epsilon) - \frac{ick_1}{\Omega} \times\left(\frac{1}{\pi}\ln\eta - \frac{(C_3\cos\theta + I_2\sin\theta)}{\sqrt{\pi}(\eta)^{3/4}}\right) + \cdots \text{ as } \eta \to \infty, \tag{4.19}$$

$$u \sim (k_1\epsilon^{-2/3} + k_2)\left(\frac{1}{\pi\eta} + \frac{(C_3 - I_1)e^{(-2/3)\phi}}{2(\sqrt{\pi})(-\eta)^{1/4}}\right) + \cdots \text{ as } \eta \to \infty, \tag{4.20}$$

$$v \sim \frac{-ic}{\Omega}(C_0 + C_{-1}\ln\epsilon) - \frac{ick_1}{\Omega}\left(\frac{1}{\pi}\ln(-\eta) - J + \int_{-\infty}^{\eta}\left[\psi_0 - \frac{1}{\pi t}\right]dt\right) + \cdots \text{ as } \eta \to -\infty. \tag{4.21}$$

5. MATCHING

The matching procedure is based on the assumption that the expansions (4.18) and (4.19) agree with the expansions of (3.35) and (3.36) as $x \to 1^-$ and the expansions (4.20) and (4.21) agree with the expansions of (3.37) and (3.38) as $x \to 1^+$. This can only happen when the constants a_{-1}, a_0, $\tilde{a}_{-1}$, $\tilde{a}_0$, ..., C_1, C_3, k_1, k_2 are interrelated. Indeed these relationships uniquely determine the constants.

To obtain these relationships we must first find two independent solutions of (3.33). This is impossible to do analytically except possibly for simple μ. However, we can use the method of Frobenius to find them. Let

$$\phi(x) = (1 + (c^2/4)\xi^2 + \cdots)\ln|\xi| - \delta\xi + \cdots, \tag{5.1}$$

$$\phi_1(x) = 1 + c^2\xi^2/4 + \cdots \tag{5.2}$$

be two such solutions where $\xi = 1 - x$ and $\delta = \csc^2\alpha|\mu'(1)| + \mu''(1)/2|\mu'(1)|$. The Wronskian of these functions is given by

$$W = A(\mu - \sin^2\alpha)/\mu, \tag{5.3}$$

where $A = |\mu'(1)|/\sin^2\alpha$. The special value

$$W(0) = |\mu'(1)|\cot^2\alpha \tag{5.4}$$

will be needed shortly.

In the region $0 < x < 1$ we set

$$V_0(x) = a\phi_0(x) + b\phi_1(x). \tag{5.5}$$

Since V_0 must satisfy the boundary condition (3.23) we have the linear relationship between a and b

$$a\hat{B}\phi_0 + b\hat{B}\phi_1 = 2ik\cos^2\alpha, \tag{5.6}$$

where $\hat{B}\psi = \psi'(0) + ik\cos\alpha\psi(0)$. In the region $x > 1$ we take

$$\tilde{V}_0(x) = e\phi_0(x) + f\phi_1(x), \tag{5.7}$$

where e and f are linearly related since $\tilde{V}_0 \to 0$ as $x \to \infty$. In fact both ϕ_0 and ϕ_1 have exponentially growing and decaying parts as $x \to \infty$. Suppressing the component which diverges gives

$$f = \chi e. \tag{5.8}$$

The parameter χ is real because ϕ_0 and ϕ_1 are real.

We now expand (3.35) and (3.36) using (5.1), (5.2), and

(5.5) as $x\to 1^-$. We find that

$$u\sim\gamma^{-1/4}\left(\frac{a_{-1}}{\epsilon^{2/3}}+\frac{a_0}{\epsilon^{1/6}}\right)\frac{\cos(\theta+\psi)}{\eta^{1/4}}+\frac{ia\Omega}{c}\frac{\epsilon^{-2/3}}{\eta}+\cdots, \tag{5.9}$$

$$v\sim\tfrac{2}{3}a\ln\epsilon+(b-a\ln\Omega)+a\ln\eta+\frac{i\gamma^{-3/4}\sin\alpha}{\eta^{3/4}}\times a_{-1}\sin(\theta+\psi)+\cdots, \tag{5.10}$$

where $\psi=-\frac{\pi}{4}+\frac{k}{\epsilon}\int_0^1(\sqrt{\mu})(s)\,ds$. (These hold only when $\sin^2\alpha\gg\gamma\epsilon^{2/3}$.) When (5.9) and (5.10) are compared with (4.18) and (4.19) we find that

$$k_1=(i\pi\Omega/c)a, \tag{5.11}$$

$$k_2=a_0=0, \tag{5.12}$$

$$C_3=I_2\tan\psi, \tag{5.13}$$

$$a_{-1}=\left(\frac{-i\gamma^{1/4}(\sqrt{\pi})\Omega I_2}{c\cos\psi}\right)a, \tag{5.14}$$

$$C_{-1}=(2i\Omega/3c)a, \tag{5.15}$$

$$C_0=(i\Omega/c)(b-a\ln\Omega). \tag{5.16}$$

Finally we expand (3.37) and (3.38), using (5.1) and (5.2) as $x\to 1^+$, and obtain

$$u\sim\gamma^{-1/4}\left(\frac{\tilde a_{-1}}{\epsilon^{2/3}}+\frac{\tilde a_0}{\epsilon^{1/6}}\right)\frac{e^{-(2/3)\phi}}{(-\eta)^{1/4}}+\frac{ie\Omega}{c}\frac{\epsilon^{-2/3}}{\eta}+\cdots, \tag{5.17}$$

$$v\sim\tfrac{2}{3}e\ln\epsilon+(f-e\ln|\Omega|)+\frac{i\gamma^{-3/4}\sin\alpha}{(-\eta)^{3/4}}\tilde a_{-1}e^{-(2/3)\phi}+e\ln|\eta|+\cdots. \tag{5.18}$$

These results are compared with (4.20) and (4.21). They require that

$$k_1=(i\pi\Omega/c)e, \tag{5.19}$$

$$k_2=\tilde a_0=0, \tag{5.20}$$

$$\tilde a_{-1}=[\gamma^{1/4}(C_3-I_1)k_1]/(2\sqrt{\pi}), \tag{5.21}$$

$$C_{-1}=(2i\Omega/3c)e, \tag{5.22}$$

$$C_0=(i\Omega/c)[f-e\ln\Omega]-k_1J. \tag{5.23}$$

Equations (5.11), (5.19) and (5.15), (5.22) imply

$$e=a. \tag{5.24}$$

Combining Eqs. (5.16), (5.23), (5.24), and (5.11) we find

$$f=b+\pi Ja. \tag{5.25}$$

This expression simplifies to

$$(\chi-\pi J)a-b=0, \tag{5.26}$$

when (5.8) and (5.24) are combined.

Equations (5.6) and (5.26) allow us to solve for a and b. The determinant of this system cannot vanish. For if it did, we would have

$$(\chi-\pi J)\phi_1'(0)=-\phi_0'(0), \tag{5.27}$$

$$(\chi-\pi J)\phi_1(0)=-\phi_0(0), \tag{5.28}$$

because ϕ_1 and ϕ_0 are real. Multiplying the first equation by ϕ_0, and the second by $\phi_0'(0)$ and subtracting the result gives

$$(\chi-\pi J)W(0)=0. \tag{5.29}$$

By (5.4), $W(0)\neq 0$. If $\chi=\pi J$, then $\phi_0(0)=0$ by (5.27) and (5.28). Since $x=0$ is an ordinary point of (3.33), this implies $\phi_0\equiv 0$, which is nonsense. Thus Eqs. (5.6) and (5.26) have the unique solution

$$a=2ik\cos^2\alpha/[\hat B\phi_0+(\chi-\pi J)\hat B\phi_1], \tag{5.30}$$

$$b=2ik(\chi-\pi J)\cos^2\alpha/[\hat B\phi_0+(\chi-\pi J)\hat B\phi_1]. \tag{5.31}$$

The remaining constants are given by (5.8), (5.24), (5.14), (5.15), (5.16), (5.21), and (5.25).

The reader will note that our expansions of (3.36) as $x\to 1^-$ and of (3.37) as $x\to 1^+$ did not contain the functions $D_1(x)$ and $\tilde D_1(x)$. These functions are identically zero as can be seen by the following brief argument. The term $D_1(x)$ would induce an additional boundary layer series in powers of $\sqrt{\epsilon}$. These would be uncoupled from the series in powers of $\epsilon^{2/3}$ at least to several orders. The same matching procedure would follow as above. However, D_1 satisfies the homogeneous boundary condition, $\hat B D_1=0$. Thus "a and b" for D_1 would vanish.

6. DISCUSSIONS

We restate here for the reader's convenience our leading order results

$$u\sim\epsilon^{-1/2}\left(\frac{-i\gamma^{1/4}(\sqrt{\pi})\Omega I_2}{c\cos\psi}\right)\times\frac{a}{\mu^{1/4}}\cos k\bar x+\frac{i\sin\alpha(a\phi_0'+b\phi_1')}{k(\mu-\sin^2\alpha)}+O(\sqrt{\epsilon}), \tag{6.1}$$

$$v\sim a\phi_0+b\phi_1+\sqrt{\epsilon}\left[\frac{\gamma(\sqrt{\pi})\Omega I_2}{k\cos\psi}\right]\frac{a\sin k\bar x}{\mu^{3/4}}+O(\epsilon) \tag{6.2}$$

for $x<1$ and

$$u\sim\epsilon^{-1/2}\left(\frac{i\gamma^{1/4}(\sqrt{\pi})\Omega(C_3-I_1)}{2c}\right)\frac{ae^{-k\xi}}{(-\mu)^{1/4}}+\frac{i\sin\alpha[a\phi_0'+(b+\pi Ja)\phi_1']}{k(\mu-\sin^2\alpha)}+O(\sqrt{\epsilon}), \tag{6.3}$$

$$v\sim a\phi_0+(b+\pi Ja)\phi_1+\left(\frac{-\gamma^{1/4}(\sqrt{\pi})\Omega(C_3-I_1)}{2k}\right)\times\frac{ae^{-k\xi}}{(-\mu)^{3/4}}+O(\epsilon) \tag{6.4}$$

for $x>1$. The constants a and b are given by (5.30) and (5.31). In the boundary layer we have

$$u\sim\epsilon^{-2/3}(i\pi\Omega/c)a\psi_0+O(1), \tag{6.5}$$

$$v\sim\ln\epsilon(2i\Omega/3c)a+(i\Omega/c)(b-a\ln\Omega)+a\left(\ln\eta-\int_\eta^\infty\left[\pi\psi_0(t)-\frac{1}{t}\right]dt\right)+O(\epsilon^{2/3}). \tag{6.6}$$

There are several interesting features in our asymptotic solution. First, the amplitude of the leading term in u is $O(1/\sqrt{\epsilon})$ and thus is quite large. This points out that the electrons are driven near resonance by the incident electric field near $x=1$. The leading term is of course the "plasma wave."

Secondly, another type of resonance occurs when

$\cos\psi = 0$ or when

$$k = k_n = \epsilon\pi(N + \tfrac{3}{4})\int_0^1 (\sqrt{\mu})(s)\,ds. \tag{6.7}$$

Since our analysis assumed $k = O(1)$ as $\epsilon \to 0$, (6.7) is satisfied for $N = O(1/\epsilon)$. These values of k are the eigenvalues of the homogeneous problem. The singularity of the plasma wave implies that the electrons are being pumped at just the right wavelength. From (6.7) we see that

$$k_{N+1} - k_N = O(\epsilon). \tag{6.8}$$

That is the eigenvalues are spaced close together. Of course, these peculiarities can be removed by making μ slightly complex at $x = 1$.

Thirdly, the amplitude of the scattered field

$$R = |A_x|^2 + |A_y|^2 \tag{6.9}$$

is found by inserting (2.7) and (2.11) into this expression and using (5.30), (5.31), and (6.2). We find that

$$R = 1 + O(\epsilon), \tag{6.10}$$

which indicates that no energy is absorbed. This is consistent with the asumption that μ is real. Thus, although the plasma wave is huge, no energy is transferred from the incident electromagnetic wave to the lossless plasma.

Finally, there is a singularity in the leading term of (6.5) as $c = k\sin\alpha$ approaches zero. This nonuniformity occurs when $\sin\alpha = O(\epsilon^{1/3})$. It is a direct consequence of our asymptotic method and does not reflect reality. For when $\alpha = 0$, the original problem becomes

$$\epsilon^2 u'' + k^2\mu u = 0, \tag{6.11}$$

$$v'' + k^2\mu v = 0, \tag{6.12}$$

$$\lim_{x\to\infty}(u,v) = (0,0), \tag{6.13}$$

$$u'(0) = 0, \tag{6.14}$$

$$v'(0) + ikv(0) = 2ih, \tag{6.15}$$

and no singularity is present. The singularity arises from the fact that the parameter $c = k\sin\alpha$ in (3.3) and (3.4) is small and the expansions (3.8) and (3.9) are no longer valid. A new asymptotic series is required in the outer and inner regions. We are presently working on this problem.

We conjecture that $a \to 0$ as $\alpha \to 0$. Our reasoning is based on the observation that $\sin^2\alpha$ enters the full problem in a regular way. The vanishing of a implies that no plasma waves are present when the incident wave impinges normally on the slab. A standard WKB[10] analysis of (6.11), (6.13) and (6.14) yields $u \equiv 0$ except when $\cos\psi = 0$. Thus no plasma waves are present except when $k = k_N$ [see (6.7)].

The plausibility of our conjecture is further strengthened by the observation

$$\phi_0 \sim \frac{-|\mu'(1)|}{\sin^2\alpha}(1-x) + \ln|1-x| + O(\alpha), \tag{6.16}$$

$$\phi_1 \sim 1 + \frac{k^2\sin^2\alpha}{4}(x-1)^2 + O(1) \tag{6.17}$$

as $\alpha \to 0$. Only the first three terms of each series was computed. If these expressions are inserted into (5.6) and (5.26), we deduce that $a = O(\sin^2\alpha)$.

ACKNOWLEDGMENT

The author wishes to thank Professor W. H. Reid of the University of Chicago for demonstrating $I_1 = 0$, $I_2 = 1$, and $J = C_3$.

[1] N. G. Denisov, Zh. Eksp. Teor. Fiz. **31**, 609 (1957) [Sov. Phys. JETP **4**, 544 (1957)].
[2] K. G. Budden, *Radio Waves in the Ionosphere* (Cambridge U.P., Cambridge, England, 1961).
[3] T. Speziale and P. J. Catto, Phys. Fluids **20**, 990 (1977).
[4] D. K. Kelly and A. Baños, University of California at Los Angeles Plasma Physics Group Report PPG-170 (1974).
[5] A. D. Piliya, Zh. Tek. Fiz. **36**,818 (1966) [Sov. Phys. Tech. Phys. **11**, 609 (1966)].
[6] K. A. Bruecker and S. Jorna, Rev. Mod. Phys. **46**, 325 (1974).
[7] J. A. McHugh, Arch. Hist. Exact Sci. **7**, 279 (1971).
[8] D. Meksyn, Proc. London Math. Soc. **49** (2), 436 (1944).
[9] D. S. Jones, *The Theory of Electromagnetism* (Pergamon, New York, 1964).
[10] J. D. Cole, *Perturbation Methods in Applied Mathematics* (Blaisdell, Waltham, 1968).

Time-dependent, finite, rotating universes

Marcelo J. Rebouças and J. Ademir S. de Lima [a)]
Centro Brasilerio de Pesquisas Físicas/CNPq, Av. Wenceslau Braz, 71, 22.290-Rio de Janeiro-RJ-Brazil

(Received 30 September 1980; accepted for publication 20 July 1981)

We present three new classes of finite, homogeneous and nonsingular solutions of Einstein's equations which have time-dependent expansion, shear, and rotation. The t-constant sections are of Bianchi type IX. The source of these geometries is a fluid which has not been thermalized.

PACS numbers: 98.80.Dr, 04.20.Jb, 04.20. – q, 98.80.Pp

I. INTRODUCTION

The problem of obtaining solutions to Einstein's gravitational field equations in which the source of curvature of space–time possesses rotation is fairly old. It was Bach[1] who first treated the field produced by a slowly rotating sphere of matter and found an approximate solution which contains the Schwarzschild solution as its zeroth-order approximation. Two years later Lanczos[2] considered a special case of stationary fields and obtained a nonhomogeneous solution to Einstein's equations. Afterwards Akeley[3] and Andress[4] treated more general cases of gravitational fields produced by matter in stationary rotation. They both found approximate solutions to the problem they studied. In 1932 Lewis[5] determined an exact solution whose source is rotating matter and which contains the approximate solutions due to Bach and Andress as a special case. Five years later Stockum[6] discovered a new nonhomogeneous exact solution for matter rotating with constant angular velocity about its axis of symmetry. He showed that his solution contains that of Lewis as a particular case depending on the boundary conditions. However, Gödel[7] was the first to exhibit an exact, *rotating, cosmological solution* to the Einstein field equations. Gödel's model may be characterized as a solution to Einstein's equations for incoherent matter which is homogeneous in space–time with nonvanishing density and equidistant world lines of matter (no expansion and no shear), which have a nonvanishing and constant angular velocity relative to the compass of inertia. In spite of its numerous pathologies (violation of causality, absence of a global Cauchy surface, nonorientability of time, etc.[8]) the new example of a cosmological solution of the field equations of general relativity discussed by Gödel constituted (and constitutes) a considerable stimulus to the investigation of solutions more complex than those studied until then. The literature on cosmological solutions of Einstein's equations in which the material content of the models has rotation is rather large today.[9–17] All of them, however, share a common feature: The rotation is constant with respect to time.[18] Such a characteristic has given rise among some cosmologists to a feeling that the structure of Einstein's equations does not allow a rotation varying with time. In this paper we show that this feeling is incorrect by exhibiting three classes of exact cosmological solutions of the Einstein equations in which the rotation is time-dependent. Our universes are, as far as we know, the first cosmological models which are spatially homogeneous (the t-constant sections are of Bianchi type IX), finite, and with rotation, expansion, and shear which are time-dependent. The source of our geometries is a fluid which has not been thermalized, so that there is heat exchange between parts of it.

II. THE METRIC

Let us start by discussing the group of proper motions and the topology of the t-constant sections of our geometry. The calculation methods to be employed in this first part have been inspired by the work of Ozsvath and Schuking,[11] which will be presented here concisely to make our text as clear and self-contained as possible.

Consider the Euclidean space E_4 and let q^μ and $\mathbf{e}_\mu (\mu = 0,1,2,3)$ be the Cartesian coordinates and the unit vectors along the coordinate axes. The vector space over the real numbers spanned by these vectors can be converted into an algebra by defining

$$\begin{aligned} &\mathbf{e}_0\mathbf{e}_\mu = \mathbf{e}_\mu\mathbf{e}_0 = \mathbf{e}_\mu, \\ &\mathbf{e}_i\mathbf{e}_i = -\mathbf{e}_0 \quad \text{(no sum over } i\text{)}, \\ &\mathbf{e}_i\mathbf{e}_j = \epsilon_{ijk}\mathbf{e}_k \quad (i \neq j), \end{aligned} \tag{2.1}$$

where here and in the following, Greek indices vary from 0 to 3, Latin indices vary from 1 to 3, and there is always a sum over repeated indices, unless otherwise stated. With the multiplication law (2.1) E_4 becomes an algebra known by the name of quaternion algebra, and the vectors

$$\mathbf{q} = q^\mu\mathbf{e}_\mu = q^0\mathbf{e}_0 + \sum_{i=1}^{3} q^i\mathbf{e}_i \tag{2.2}$$

are called Hamiltonian quaternions or simply quaternions.

Given quaternion $\mathbf{q}$, its conjugate $\mathbf{q}^*$ and its norm $\mathrm{N}(\mathbf{q})$ are defined respectively by

$$\mathbf{q}^* = q^0\mathbf{e}_0 - \sum_i q^i\mathbf{e}_i, \tag{2.3}$$

and

$$\mathrm{N}(\mathbf{q}) = \mathbf{q}\,\mathbf{q}^* = (q^0)^2 + \sum_{i=1}^{3} (q^i)^2, \tag{2.4}$$

where in the equality (2.4) we identify the subfield $q^0\mathbf{e}_0$ with the field of real numbers in virtue of the isomorphism between them.

Let us mention now, for future use, some properties of the associative but noncommutative field of quaternions that can be easily verified. Given two quaternions $\mathbf{p}$ and $\mathbf{q}$

[a)] On leave of absence from Departamento de Física, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil. Partially supported by CAPES.

 0022-2488/81/112699-05$01.00

then

(i) $N(\mathbf{q}) = \sum_{\mu} (q^{\mu})^2 \geqslant 0$ and $N(\mathbf{q}) = 0 \Leftrightarrow \mathbf{q} = 0$;

(ii) If $\mathbf{q} \neq 0$ there exists the inverse quaternion $\mathbf{q}^{-1}$ such that $\mathbf{q}^{-1}\mathbf{q} = \mathbf{q}\,\mathbf{q}^{-1} = 1$, and it is given by $\mathbf{q}^{-1} = \mathbf{q}^*/N(\mathbf{q})$;

(iii) $(\mathbf{p}\,\mathbf{q})^* = \mathbf{q}^*\,\mathbf{p}^*$;

(iv) $(\alpha_1\mathbf{p} + \alpha_2\mathbf{q})^* = \alpha_1\mathbf{p}^* + \alpha_2\mathbf{q}^*$, where α_1, α_2 belong to the field of real numbers.

The set of all quaternions that satisfy the equation

$$N(\mathbf{q}) = \mathbf{q}\,\mathbf{q}^* = (q^0)^2 + \sum_{i=1}^{3} (q_i)^2 = 1 \tag{2.5}$$

obviously makes up a group with respect to the multiplication law defined by Eq. (2.1). On the other hand, Eq. (2.5) above is the equation for the points of a sphere S^3, with center at the origin, embedded in E_4.

The group of motions of the three-dimensional spherical space S^3 may be easily expressed by the left quaternion product. In fact let $\mathbf{p}$ be any quaternion of $S^3(\mathbf{p}\,\mathbf{p}^* = 1)$. A left translation is expressed by

$$\mathbf{q}' = \mathbf{p}\,\mathbf{q}. \tag{2.6}$$

By making use of property (iii) we find

$$\mathbf{q}'\mathbf{q}'^* = 1. \tag{2.7}$$

Moreover

$$\mathbf{p}^*\mathbf{q}' = \mathbf{p}^*\mathbf{p}\,\mathbf{q} = \mathbf{q}. \tag{2.8}$$

Thus, a left translation transforms a point $\mathbf{q}$ of S^3 into another point $\mathbf{q}'$ of the spherical space, and more than this, by Eq. (2.8) there exists an unique translation (2.6) which transforms a given $\mathbf{q} \neq 0$ into a given $\mathbf{q}'$ with the same norm. In other words, the translations turn S^3 into a simply transitive Lie group. In this way we may identify S^3 with the motion group of S^3, with S^3 acting upon itself by left translations.

There is a one-to-one correspondence between the points of S^3 and the elements of the group SU_2. So we can replace every normed quaternion $\mathbf{q} = q^{\mu}\mathbf{e}_{\mu}$ by the unitary matrix

$$A = \begin{vmatrix} q^0 + iq^3 & q^1 + iq^2 \\ -q^1 + iq^2 & q^0 - iq^3 \end{vmatrix}, \quad \det A = N(\mathbf{q}) = 1, \tag{2.9}$$

and the quaternion multiplication goes over to the matrix multiplication.

We introduce the Eulerian coordinates (x^1,x^2,x^3) on S^3 by

$$q^0 = \cos\frac{x^1}{2}\cos\frac{x^2+x^3}{2}, \quad q^1 = -\sin\frac{x^1}{2}\sin\frac{x^2-x^3}{2},$$
$$q^2 = \sin\frac{x^1}{2}\cos\frac{x^2-x^3}{2}, \quad q^3 = \cos\frac{x^1}{2}\sin\frac{x^2+x^3}{2}, \tag{2.10}$$

where $0 \leqslant x^1 \leqslant \pi$ and $0 \leqslant x^2, x^3 \leqslant 2\pi$. One knows that the left invariant 1-forms of a matrix group whose general element is given by the matrix A can be obtained by computing

$$\omega = A^{-1}dA \tag{2.11}$$

Making use of the above equation we have

$$\omega^1 = \cos x^3\,dx^1 + \sin x^1 \sin x^3\,dx^2,$$
$$\omega^2 = -\sin x^3\,dx^1 + \sin x^1 \cos x^3\,dx^2, \tag{2.12}$$
$$\omega^3 = \cos x^1\,dx^2 + dx^3,$$

with correspondent left invariant vector field given by

$$X_1 = \cos x^3\frac{\partial}{\partial x^1} + \frac{\sin x^3}{\sin x^1}\frac{\partial}{\partial x^2} - \cot x^1 \sin x^3\frac{\partial}{\partial x^3},$$
$$X_2 = -\sin x^3\frac{\partial}{\partial x^1} + \frac{\cos x^3}{\sin x^1}\frac{\partial}{\partial x^2} - \cot x^1 \cos x^3\frac{\partial}{\partial x^3},$$
$$X_3 = \frac{\partial}{\partial x^3}. \tag{2.13}$$

Clearly

$$d\omega^1 = -\omega^2\Lambda\omega^3, \quad d\omega^2 = -\omega^3\Lambda\omega^1, \quad d\omega^3 = -\omega^1\Lambda\omega^2. \tag{2.14}$$

Therefore the algebra of the invariant vector fields on the sphere S^3 coincides with the algebra of the invariant vector fields of the type Bianchi IX.[19]

We shall construct the metric of our space–time manifold by defining the 1-forms

$$\Theta^0 = dt - \alpha A(t)\omega^3,$$
$$\Theta^1 = B_0\omega^1 \quad (\alpha, B_0 = \text{const}, \ \alpha^2 \neq 1), \tag{2.15}$$
$$\Theta^2 = B_0\omega^2,$$
$$\Theta^3 = A(t)\omega^3,$$

and requiring the local Lorentzian character, namely,

$$ds^2 = \eta_{AB}\Theta^A\Theta^B \quad (A,B = 0,1,2,3), \tag{2.16}$$

where

$$\eta_{AB} = \text{diag}(+1, -1, -1, -1). \tag{2.17}$$

Thus our $t = \text{const}$ sections are of Bianchi type IX which have the topological structure of a sphere S^3. Indeed, the construction (2.16) may be understood as corresponding to a deformation of S^3 into the ellipsoid ("bag") $\bar{S}^3$, and therefore our universes will be finite.

III. THE MATTER CONTENT

Friedmann's homogeneous and isotropic solutions are those which best fit the observational data that we have at our disposal about the universe in its present state. However all of these solutions of the Friedmann type, as well as most relativistic cosmological models, treat the matter content of the universe as a perfect fluid of density ρ and pressure p. Such a hypothesis, in spite of greatly simplifying the description of the Universe, has led us systematically to a cosmos with singularities. Investigations about the galactic fluid in its primeval phase (near the singularity) have induced some authors to incorporate, through phenomenological relations borrowed from fluid mechanics, dissipative terms to the energy–momentum tensor of the matter content.

The discovery of the isotropy of the 3 °K blackbody radiation had spurred Misner[20] to use, in examining the causes for this isotropy, a relation between anisotropic pressure and shear ($\pi_{\mu\nu} = \lambda\sigma_{\mu\nu}$) with the introduction of a viscosity coefficient. Murphy[21] and Klimek,[22] when dealing with homogeneous and isotropic models without singularities, employed a second viscosity coefficient. Recently, Be-

linskii and Khalatnikov[23] investigated the influence of viscosity upon the character of the cosmological evolution in Bianchi type I homogeneous models by making use of expressions which add to the pressure p a term proportional to the expansion by means of viscosity coefficients. Phenomenological relations of the type[24] $q_\alpha = -\chi h_\alpha{}^\lambda (T_{|\lambda} + T\dot{v}_\lambda)$, which relate the heat flux q_α with the acceleration $\dot{v}^\lambda$ through a temperature function and the heat-conduction coefficient χ, have been proposed. Even more recently,[14] configurations in which the galactic fluid is not in thermal equilibrium have been investigated by introducing a heat flux q^α between parts of the matter content by means of a kinematic basis erected with the velocity, acceleration and rotation vectors $[q^\alpha = \zeta(x^\mu)a^\alpha + \beta(x^\mu)\omega^\alpha + \phi(x^\mu)\eta^\alpha_{\epsilon\lambda\rho}\, a^\epsilon \omega^\lambda v^\rho]$. Such a construction, however, is possible whenever the vorticity vector generates an isometry. In the same vein, the matter content of our model is a fluid in which there is heat exchange between parts of it:

$$T_{\alpha\beta} = (\rho + p)v_\alpha v_\beta - p g_{\alpha\beta} + 2q_{(\alpha} v_{\beta)}, \tag{3.1}$$

where

$$q_\alpha v^\alpha = 0. \tag{3.2}$$

Since the rotation of our geometry is not a Killing vector, we shall follow a different path to introduce the flux of heat. We shall use a basis of vectors e^A_α defined by $\Theta^A = e^A_\alpha\, dx^\alpha$ (Eq. 2.26) in order to express our fourvector q^A as

$$q^A = e^A_\alpha \psi^\alpha(x^\mu) \tag{3.3}$$

or, equivalently,

$$q_{(A)} = e^\alpha_A \psi_\alpha(x^\mu), \tag{3.4}$$

where $e^A_\alpha e^\alpha_B = \delta^A_B$ and $\psi_\alpha(x^\mu)$ are scalar fields. The above decomposition has the advantages of not taking for granted any *a priori* phenomenological relations [such as $q_\alpha = -\chi h^\lambda_\alpha (T_{|\lambda} + T\dot{v}_\lambda)$], of allowing us to impose additional restrictions upon q^α, and of being entirely general.

In the comoving coordinate system, $v^\alpha = \delta^\alpha{}_0$ implies $v^A = e^A{}_0$ and condition (3.2) leads to $q_{(0)} = 0$. As matter of fact the decomposition (3.4) is equivalent to using the invariant vector fields Y_A, associated with the invariant 1-forms of the $t = \text{const}$ section, and given by

$$Y_1 = (1/B_0)X_1, \quad Y_2 = (1/B_0)X_2, \quad Y_3 = (1/A)X_3. \tag{3.5}$$

In the inertial frame determined by Eqs. (2.15) and (2.16) our energy–momentum tensor takes the form

$$T^{AB} = \rho e^A_0 e^B_0 - p h^{AB} + 2q^{(A} v^{B)},$$

where q^A is given by (3.4), $h_{AB} = v_A v_B - \eta_{AB}$, and $(\,)$ means symmetrization.

IV. THE EINSTEIN EQUATIONS

A straightforward calculation furnishes the curvature tensor in the local inertial frame defined by (2.15). Explictly

$$\begin{aligned}
&e^{(0)}_0 = 1, \quad e^{(0)}_2 = -\alpha A \cos x^1, \quad e^{(0)}_3 = -\alpha A,\\
&e^{(1)}_1 = B_0 \cos x^3, \quad e^{(1)}_2 = B_0 \sin x^1 \sin x^3,\\
&e^{(2)}_1 = -B_0 \sin x^3, \quad e^{(2)}_2 = B_0 \sin x^1 \cos x^3,\\
&e^{(3)}_2 = A \cos x^1, \quad e^{(3)}_3 = A,
\end{aligned} \tag{4.1}$$

where $e^{(A)}_\alpha$ is such that A = tetrad index and α = coordinate index. We learn that the only nonvanishing components of the Ricci tensor $R_{AB} = \eta^{CD} R_{DACB}$ are given by

$$\begin{aligned}
R_{00} &= (\alpha^2 - 1)\frac{\ddot{A}}{A} + \frac{\alpha^2}{B_0^4}\frac{A^2}{2},\\
R_{03} &= \frac{\alpha}{B_0^4}\frac{A^2}{2},\\
R_{11} &= R_{22} = \frac{\alpha^2 - 1}{B_0^4}\frac{A^2}{2} + \frac{1}{B_0^2},\\
R_{33} &= \frac{1}{B_0^4}\frac{A^2}{2} - (\alpha^2 - 1)\frac{\ddot{A}}{A},
\end{aligned} \tag{4.2}$$

where the dot denotes derivative with respect to the time.

The Einstein equations with a cosmological constant assume the form

$$R_{AB} = kT_{AB} - \tfrac{1}{2}kT\eta_{AB} + \Lambda\eta_{AB},$$

where, due to our conventions in the calculation of the Riemann tensor, $k > 0$.

The fact that $R_{01} = R_{02} = 0$ implies $T_{01} = T_{02} = 0$ and, as a consequence,

$$\begin{aligned}
&\psi_1 = 0,\\
&\psi_2 = \cos x^1 \psi_3.
\end{aligned} \tag{4.3}$$

So in the coordinate basis we have

$$q_\alpha = (0,0,\cos x^1, 1)\psi_3. \tag{4.4}$$

The remaining Einstein equations are

$$\begin{aligned}
&R_{00} = \tfrac{1}{2}k(\rho + p) + kp + \Lambda,\\
&k\psi_3 = (\alpha/B_0^4)\tfrac{1}{2}A^3,\\
&R_{11} = R_{22} = \tfrac{1}{2}k(\rho - p) - \Lambda,\\
&R_{33} = \tfrac{1}{2}k(\rho + p) - kp - \Lambda.
\end{aligned} \tag{4.5}$$

Setting $\Lambda = -1/B_0^2$ and assuming an equation of state like

$$p = \lambda\rho, \quad 0 \leqslant (\lambda = \text{const}) < 1, \tag{4.6}$$

the system of differential equations (4.5) reduces to just one equation

$$\frac{\ddot{A}}{A} + \frac{1}{B_0^4}\left(\frac{\alpha^2 - 2}{\alpha^2 - 1}\right)\frac{A^2}{2} = \frac{\Lambda}{\alpha^2 - 1}, \tag{4.7}$$

with

$$\frac{k}{2}\rho = \frac{1}{1 - \lambda}\left(\frac{\alpha^2 - 1}{B_0^4}\right)\frac{A^2}{2}, \tag{4.8}$$

where

$$\alpha^2 = \frac{\lambda + 3}{3\lambda + 1}, \quad 1 < \alpha^2 \leqslant 3. \tag{4.9}$$

This last condition guarantees the compatibility of Eq. (4.5) and the positivity of energy and pressure.

In the integration of Eq. (4.7) we must distinguish three different classes of solutions. The first of them occurs when $\alpha^2 = 2$, which corresponds to an equation of state $p = \rho/5$. In this case Eq. (4.7) reduces to the differential equation of a simple harmonic oscillator, the solution of which is

$$A = A_0 \sin\left[B_0^{-1}(t - t_0)\right], \tag{4.10}$$

where A_0 and t_0 are constants.

The second occurs when $1<\alpha^2<2$ ($p=\lambda\rho, \frac{1}{5}<\lambda<1$) and Eq. (4.7) is formally identical to the differential equation of a simple pendulum which may be integrated in terms of the Jacobi elliptic functions,[25] that is,

$$A = J_0 \,\mathrm{sn}[\delta(t-t_0), k\,], \tag{4.11}$$

where δ and t_0 are arbitrary constants, and where k^2 and J_0^2 are determined by

$$k^2 = \frac{B_0^2\delta^2(1-\alpha^2)+1}{\delta^2 B_0^2(\alpha^2-1)}, \quad J_0^2 = \frac{4\delta^2 B_0^4(1-\alpha^2)+4B_0^2}{2-\alpha^2}. \tag{4.12}$$

The third class occurs when $2<\alpha^2\leqslant 3$ ($p=\lambda\rho$, $0\leqslant\lambda<\frac{1}{5}$) and again the integral of (4.5) is given in terms of elliptic functions

$$A = C_0 \,\mathrm{dn}[\eta(t-t_0), k\,], \tag{4.13}$$

where t_0 and η are arbitrary constants, and where k^2 and C_0^2 are determined as

$$k^2 = \frac{2\eta^2 B_0^2(\alpha^2-1)+1}{\eta^2(\alpha^2-1)B_0^2}, \quad C_0^2 = \frac{4\eta^2(\alpha^2-1)B_0^4}{\alpha^2-2}. \tag{4.14}$$

Finally, let us note that in the case $\lambda = 1$ the Einstein equations cannot be made compatible.

V. MISCELLANEOUS RESULTS AND FINAL REMARKS

In the tetrad frame defined by $\theta^A = e_\alpha^{(A)}\,dx^\alpha$ the acceleration a^A, the expansion Θ, the shear σ_{AB}, and the rotation of the congruence of the world lines of matter are given by

$$a^A = V^A{}_{;C}V^C + \gamma^A{}_{BC}V^BV^C, \tag{5.1}$$

$$\Theta = V_{A;C}\eta^{AC} - \gamma^A{}_{BC}\eta^{BC}V_A, \tag{5.2}$$

$$\sigma_{AB} = \tfrac{1}{2}h_A{}^M h_B{}^N[V_{M;N}+V_{N;M}+(\gamma_{MPN}+\gamma_{NPM})V^P] - \tfrac{1}{3}\Theta h_{AB}, \tag{5.3}$$

$$\omega_{AB} = \tfrac{1}{2}h_A{}^M h_B{}^N[V_{M;N}-V_{N;M}+(\gamma_{MPN}-\gamma_{NPM})V^P], \tag{5.4}$$

where $V^A{}_{;C} = e^\alpha_{(C)}(\partial V^A/\partial x^\alpha)$ and the Ricci rotation coefficient are defined by $\gamma^A{}_{BC} = -e^{(A)}_{\alpha\|\beta}\,e^\alpha_{(B)}\,e^\beta_{(C)}$ (double bar means covariant derivative). In the inertial frame determined by (4.1) an observer moving along with the fluid has a 4-velocity $V^A = \delta^A_0$, and Ricci's coefficients have as nonvanishing components

$$\gamma^0_{12} = -\gamma^0_{21} = \gamma^1_{20} = \alpha A/2B_0^2,$$
$$\gamma^0_{30} = \alpha\dot{A}/A, \quad \gamma^0_{33} = \dot{A}/A, \tag{5.5}$$
$$\gamma^1_{23} = -\gamma^2_{31} = \gamma^1_{32} = A/2B_0^2.$$

Thus, making use of the vector $e_\alpha^{(A)}$ and relation (5.5) we obtain

$$a^A = (0,0,0,\alpha\dot{A}/A), \tag{5.6}$$

$$\Theta = \dot{A}/A, \tag{5.7}$$

and the nonvanishing components

$$\sigma_{11} = \sigma_{22} = -\tfrac{1}{2}\sigma_{33} = \tfrac{1}{3}\dot{A}/A, \tag{5.8}$$

$$\omega_{12} = -(\alpha/B_0^2)^{\frac{1}{2}}A. \tag{5.9}$$

The rotation vector $\omega^A = \frac{1}{2}\epsilon^{ABCD}\omega_{BC}V_D$ is given by

$$\omega^A = (0,0,0,\alpha A/2B_0^2). \tag{5.10}$$

So, all kinematical quantities associated with the congruence of the fluid are time-dependent as we mentioned before.

We will calculate now the components of the Weyl tensor field. The nonvanishing components are (in the tetrad frame) given by

$$\begin{aligned} C_{0101} &= \tfrac{1}{2}(\alpha^2-1)\frac{\ddot{A}}{A} + \tfrac{1}{3}\frac{(1-\alpha^2)}{B_0^4}\frac{A}{2} - \frac{1}{6B_0^2} \\ &= C_{0202} = -C_{1313} = -C_{2323}, \\ C_{0303} &= \tfrac{1}{3}\frac{(\alpha^2-1)}{B_0^4}A^2 + \frac{1}{3B_0^2} = C_{1212}, \\ C_{0213} &= \frac{(1-\alpha^2)}{B_0^2}\dot{A} = \tfrac{1}{2}C_{0312} = -C_{0123}. \end{aligned} \tag{5.11}$$

This shows that our geometry is not conformally flat.

Since the right-invariant vector fields K_i on S^3 satisfy by construction $[K_i, X_j] = 0, i,j = 1,2,3$, the line element (2.16) will have K_i as Killing vectors, which in Eulerian coordinates are expressed as

$$\begin{aligned} K_1 &= \frac{\partial}{\partial x^2}, \\ K_2 &= \cos x^2\frac{\partial}{\partial x^1} - \cot x^1 \sin x^2\frac{\partial}{\partial x^2} + \frac{\sin x^2}{\sin x^1}\frac{\partial}{\partial x^3}, \\ K_3 &= -\sin x^2\frac{\partial}{\partial x^1} - \cot x^1\cos x^2\frac{\partial}{\partial x^2} + \frac{\cos x^2}{\sin x^1}\frac{\partial}{\partial x^3}. \end{aligned} \tag{5.12}$$

Another interesting characteristic of the cosmological solutions (4.10), (4.11), and (4.13) is that such geometries have no singularities—the space–time curvature and the fluid density are finite for all proper time of any observer. This does not contradict the general result found by Ryan[26–27] which assures that in Bianchi IX universes with expansion, shear, and rotation, the character of the singularity cannot be drastically altered by the rotation. The reason why there is no contradiction lies in the fact that Ryan's analysis is restricted to pure fluid energy-momentum tensors, which is not our case. Thus the lack of singularity in our universes may be due to the presence of rotation and/or to the existence of dissipative processes (heat exchange) as those that occur in the models of Murphy[21] and Klimek.[22]

As a remark about our solutions, let us mention the fact that according to Eq. (4.4) our models are also interesting examples that satisfy the group criterion of homogeneity but are not homogeneous according to the physical criterion, just like what happens to the solutions due to Demianski and Grishchuk.[13]

It is important to show the relation between the metric (2.16) and the Taub model, and point out how the energy-momentum tensor and the cosmological constant allowed the solutions we have found. If we perform the following

coordinate transformation,

$$t' = t/(1-\alpha^2)^{1/2},$$
$$x'^1 = x^1, \tag{5.13}$$
$$x'^2 = x^2,$$
$$x'^3 = x^3 + g(t),$$

where $g(t) = \int [\alpha/(1-\alpha^2)A\,]\,dt$, our line element (2.16) becomes

$$ds^2 = (dt')^2 - B_0^2[(\omega'^1)^2 + (\omega'^2)^2] - (1-\alpha^2)A^2(t')(\omega'^3)^2, \tag{5.14}$$

where

$$\omega'^1 = \cos g\omega^1 + \sin g\omega^2,$$
$$\omega'^2 = -\sin g\omega^1 + \cos g\omega^2, \tag{5.15}$$
$$\omega'^3 = \omega^3 + [\alpha/(1-\alpha^2)A\,]\,dt,$$

which has the same form as the Taub metric.[28,29] On the other hand, if we try to introduce as before a heat flux q^α in the Taub model we discover, via Einstein's field equations,[30] that this is impossible. The differences between our models and that due to Taub, such as the presence of the cosmological constant and the fact that our universes are tilted models,[31] allowed the heat exchange between parts of the galactic fluid and consequently the peculiar solution we have found.

ACKNOWLEDGMENTS

The authors wish to thank Professor Mário J. D. Assad (President of PRI) for many valuable discussions during the course of this work, and Nivaldo A. Lemos (Secretary-general of PRI) without whose help in the English transcription this paper could not have been written in its present form.

[1]R. Bach, Mat. Zeitsnrift. **13**, 134 (1922).
[2]K. Lanczos, Z. Physik **21**, 73 (1924).
[3]E. Akeley, Proc. R. Soc. London, Ser. A **126**, 592 (1930).
[4]W. R. Andress, Philos. Mag. **11**, 322 (1931).
[5]T. Lewis, Proc. R. Soc. London, Ser. A **136**, 175 (1932).
[6]W. J. van Stockum, Proc. R. Soc. Edinburgh **57**, 135 (1937).
[7]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[8]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).
[9]J. P. Wright, J. Math. Phys. **6**, 103 (1965).
[10]S. C. Maitra, J. Math. Phys. **7**, 1025 (1966).
[11]I. Ozsvath and E. L. Schucking, Ann. Phys. (NY) **55**, 166 (1969).
[12]I. Ozsvath, J. Math. Phys. **11**, 2871 (1970).
[13]M. Demianski and L. P. Grishchuk, Commun. Math. Phys. **25**, 233 (1972).
[14]M. Novello and M. J. Rebouças, Astrophys. J. **225**, 719 (1978).
[15]M. J. Rebouças, Phys. Lett. A **70**, 161 (1979).
[16]M. Novello and M. J. Rebouças, Phys. Rev. D **19**, 2850 (1979).
[17]I. D. Soares, J. Math. Phys. **21**, 521 (1980).
[18]In the solutions due to Maitra, Wright, and Demianski *et al.* the rotating depends on one space coordinate.
[19]A. H. Taub, Ann. Math. **53**, 472 (1951).
[20]C. W. Misner, Astrophys. J. **151**, 431 (1968).
[21]G. Murphy, Phys. Rev. D **8**, 4231 (1973).
[22]Z. Klimek, Acta Cosmologica **2**, 49 (1973).
[23]V. A. Belinskii and I. M. Khalatnikov, Zh. Eksp. Teor. Fiz. **69**, 401 (1975) [Sov. Phys. JEPT **42**, 205 (1976)].
[24]G. F. R. Ellis, in *Rendiconti della Scuola Internazionale di Fisca Enrico Fermi*, Varema 1971 (Academic, New York ,1971).
[25]H. T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover, New York, 1962).
[26]M. P. Ryan, Jr., Ann. Phys. (NY) **65**, 506 (1971).
[27]M. P. Ryan, Jr., Ann. Phys. (NY) **68**, 541 (1971).
[28]We acknowledge the referee for calling our attention to this point.
[29]We remark that in our geometries $\alpha^2 > 1$, that is, the transformation $t' = t/(1-\alpha^2)^{1/2}$ is imaginary. Therefore the metric (5.14) is invalid in our case.
[30]See, for example, R. P. Ryan, Jr. and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, New Jersey, 1975), where the Ricci tensor for Taub's model can be found.
[31]A. R. King and G. F. Ellis, Commun. Math. Phys. **31**, 209 (1973).

Nonperturbative confinement in quantum chromodynamics. I. Study of an approximate equation of Mandelstam

D. Atkinson, J. K. Drohm, P. W. Johnson, [a] and K. Stam
Institute for Theoretical Physics, University of Groningen, P. O. Box 800, 9700 AV Groningen, The Netherlands

(Received 10 April 1981; accepted for publication 12 June 1981)

An approximated form of the Dyson–Schwinger equation for the gluon propagator in quarkless QCD is subjected to nonlinear functional and numerical analysis. It is found that solutions exist, and that these have a double pole at the origin of the square of the propagator momentum, together with an accumulation of soft branch points. This analytic structure is strongly suggestive of confinement by infrared slavery.

PACS numbers: 11.10.Np

I. INTRODUCTION

Of the various ways that quarks and gluons might automatically be confined in quantum chromodynamics, the hypothesis that the gluon propagator has a strong singularity at the origin of the k^2 plane, where k is the gluon four-momentum, is especially attractive. Recently Mandelstam,[1] working in the Landau gauge, has approximated the Dyson equation for the gluon propagator and claims that the latter probably behaves like k^{-4} as $k \to 0$. Similarly, Baker *et al.*,[2] using an axial gauge, make a different sequence of approximations, obtaining a much more complicated equation. However, again they claim the same infrared behavior for the gluon propagator.

We propose to study these equations by functional and numerical methods in order to elucidate the infrared properties of QCD. In this initial paper, we consider Mandelstam's equation in its simplest form and prove that it indeed possesses a solution with the k^{-4} infrared behavior. Further, we study the analytical structure of the propagator, away from the origin, by numerical means.

Let us first clarify the sense in which a gluon propagator that behaves like k^{-4}, as $k \to 0$, corresponds to a linearly rising potential,

$$V(|\mathbf{x}|) \sim |\mathbf{x}| \text{ as } |\mathbf{x}| \to \infty, \tag{1.1}$$

acting between a quark and an antiquark in their center-of-mass system, effectively confining them. In nonrelativistic Born approximation, the scattering amplitude is given in terms of the potential by

$$A(\mathbf{k}) = -\frac{1}{4\pi}\int d^3x\, V(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{1.2}$$

where $\mathbf{k}$ is the three-momentum transfer between initial and final quark. For a linear potential, the amplitude is thus proportional to

$$-\frac{1}{4\pi}\int d^3x |\mathbf{x}| e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{1.3}$$

which is well defined as a tempered distribution, as we now show. For let $\Omega(\mathbf{k})$ be any infinitely differentiable test function that vanishes faster than any inverse power of $|\mathbf{k}|$, as $|\mathbf{k}| \to \infty$, and which satisfies in addition

$$\Omega(0) = \Omega'(0) = \Omega''(0) = \Omega'''(0) = 0. \tag{1.4}$$

We show in Appendix A that

$$-\frac{1}{4\pi}\int d^3x |\mathbf{x}| \int d^3k\, \Omega(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} = 2\int d^3k\, \Omega(\mathbf{k}) |\mathbf{k}|^{-4}, \tag{1.5}$$

i.e., the amplitude is proportional to $|\mathbf{k}|^{-4}$, as a distribution on our space of test functions. In the center-of-mass system, the square of the four momentum transfer, k^2, is equal to $-|\mathbf{k}|^2$, so we see that a linear potential indeed corresponds to a propagator that behaves like k^{-4} as $k \to 0$.

In Sec. 2, we discuss the rather drastic approximations that Mandelstam makes; and this culminates in the derivation of a deceptively simple nonlinear Volterra equation. The rest of the present work is devoted to a study of this equation; but it is appropriate to ask at this point how much of QCD really has survived. The infrared behavior may be such a survivor, since the approximations are less severe for small momenta. Moreover, there are indications that the confining singularity may be specific to the vector theory, for in massless QED, if one approximates the Dyson equation for the electron propagator by replacing the full vertex and the full photon propagator by their bare values, one finds in general that there is no singularity of the solution at $k^2 = 0$: the chiral symmetry has been broken.[3] Unfortunately, the branch point nearest to the origin, which should correspond to the point $k^2 = m_e^2$, is complex,[4] and moreover its position is gauge-dependent.[5] Because of certain algebraic differences, the massless vector case (QCD) is quite different; and we find that a solution exists in which the propagator has a double pole at the origin, in the variable $x = -k^2$ (implying confinement), and also a branch point at $x = 0$, corresponding to soft gluon effects. Unfortunately, these desirable properties of the solution are marred by the fact that, as in the spinor case, we also find unwanted complex branch points.

In Sec. 2, we summarize Mandelstam's method of obtaining his gluon equation, and we convert it into a nonlinear Volterra equation. This equation is not suited to an application of the Banach theorem, because of cancellations for small k that are hard to handle, so in Sec. 3 we deduce a new integral equation, by way of a nonlinear differential equa-

[a] Permanent address: Illinois Institute of Technology, Chicago, IL 60616.

 0022-2488/81/112704-09$01.00

tion, that is so suited; and in Sec. 4 we elaborate a proof of the local uniqueness of an analytic solution of the propagator equation in a cardioid region around the origin of the k^2 plane. In Sec. 5, we outline a program for the numerical integration, by a fourth-order Runge–Kutta routine, of the nonlinear differential equation of Sec. 3, which has been used to continue out of the domain of the complex plane for which the existence proof is valid. Six complex branch points have been found close to the origin, and there may be more. In Appendices A and B we collect some technical details, while Appendix C is devoted to a further approximation of the Mandelstam equation, in which the kernel is replaced by its average value. This is a gross simplification; but the advantage is that the averaged equation is explicitly soluble, and this is of great help in testing the computer programs.

II. MANDELSTAM'S METHOD

In Ref. 1, Mandelstam considers a Dyson equation for the gluon propagator, in QCD without quarks, in which the four-gluon vertices are thrown away. Further, the three-gluon vertex is replaced by its bare value; and simultaneously one, but not both, of the internal gluon propagators is replaced by its bare value. This rather arbitrary procedure is justified on the grounds that if the propagator behaves like k^{-4}, as hoped, then the full three-gluon vertex behaves like k^2, and therefore the replacement of the full by the bare vertex should be matched by the softening of the k^{-4} behavior of one full propagator to the k^{-2} behavior of a bare propagator. Clearly this simplification is drastic, and it is uncertain whether the physical import of the Dyson equation has been fundamentally altered. The contribution of the ghost field is expected to be fairly small, and is in the first instance neglected, although it can be included without too much trouble.

The mutilated Dyson equation is depicted in Fig. 1. It is assumed that the full gluon propagator can be written in the form

$$D'^{ab}_{F\mu\nu}(k) = F(-k^2) D^{ab}_{F\mu\nu}(k), \tag{2.1}$$

where F is an unknown function of the scalar $-k^2$, and where

$$D^{ab}_{F\mu\nu}(k) = i[-g_{\mu\nu} + k_\mu k_\nu / k^2] \frac{\delta^{ab}}{k^2 + i\epsilon} \tag{2.2}$$

is the bare propagator in the Landau gauge. Here μ, ν are Lorentz indices and a, b are SU(3) color superscripts. The Dyson equation is then

$$D'^{-1}_F(k) = D^{-1}_F(k) - \frac{g^2}{(2\pi)^4} \times \int d^4k' \, \Gamma(k, k'-k, -k') D'_F(k') \times \Gamma(k', k-k', -k) D_F(k'-k), \tag{2.3}$$

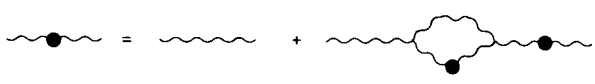


FIG. 1. Approximate Dyson equation for the gluon propagator. The wavy lines with a black dot are full, the others bare gluon propagators.

where Lorentz and color indices have been omitted, where Γ is the bare three-gluon vertex, and where g is the SU(3) coupling constant. After a Wick rotation has been performed, the angular integrations can be done, and Mandelstam finds

$$[F(x)]^{-1} = 1 - \frac{G}{x} \int_0^\infty dy \, K_1(x,y) F(y), \tag{2.4}$$

where $x = -k^2, y = -k'^2$, and where G is proportional to g^2, the constant of proportionality being unimportant. In (2.4) the kernel is

$$K_1(x,y) = 18 + \left(25\frac{y}{x} - \frac{7}{2}\frac{y^2}{x^2}\right)\theta(x-y) + \left(25\frac{x}{y} - \frac{7}{2}\frac{x^2}{y^2}\right)\theta(y-x). \tag{2.5}$$

One can rewrite (2.4) in the form

$$[F(x)]^{-1} = 1 + \frac{B}{x} - \frac{G}{x}\int_0^x dy\left(25\frac{y}{x} - \frac{7}{2}\frac{y^2}{x^2}\right)F(y) - \frac{G}{x}\int_x^\infty dy\left(25\frac{x}{y} - \frac{7}{2}\frac{x^2}{y^2}\right)F(y), \tag{2.6}$$

where

$$B = -18G\int_0^\infty dy \, F(y). \tag{2.7}$$

This integral needs both infrared and ultraviolet cutoffs, which we do not write explicitly. Now is it possible that the propagator (2.1) behaves like k^{-4} as $k \to 0$, i.e., that $F(x)$ behaves like x^{-1} as $x \to 0$? Let us write, following Mandelstam,

$$F(x) = A/x + F_1(x), \tag{2.8}$$

where $F_1(x)$ is to be less singular than a pole as $x \to 0$ (in fact, it will turn out to be bounded). We substitute (2.8) into (2.6) and perform the integrals over the pole term:

$$\frac{x}{A + xF_1(x)} = 1 + \frac{1}{x}\left[B - \frac{93}{2}AG\right] - \frac{G}{x}\int_0^x dy\left(25\frac{y}{x} - \frac{7}{2}\frac{y^2}{x^2}\right)F_1(y) - \frac{G}{x}\int_x^\infty dy\left(25\frac{x}{y} - \frac{7}{2}\frac{x^2}{y^2}\right)F_1(y). \tag{2.9}$$

Now if $F_1(x)$ is well behaved at the origin, the pole term on the right hand side of (2.9) must vanish, i.e., A is given by

$$A = 2B/93G.$$

A constant behavior of $F_1(x)$ as $x \to 0$ would be inconsistent with (2.9), since the first term under the second integral would yield $\ln x$, which could not be cancelled; and likewise a linear behavior, $F_1(x) \sim x$, is inconsistent, for now the second term under the second integral gives an insupportable $\ln x$. Mandelstam shows that these inconsistencies are removed if $F_1(x) \sim x^\alpha$, where $\alpha = -1 + (31/6)^{1/2} \approx 1.273$. He further suggests, as a reasonable first approximation, the dropping of the "$\frac{7}{2}$ terms" in (2.9). The equation then becomes

$$\frac{x}{A + xF_1(x)} = 1 - 25\frac{G}{x}\int_0^x dy \, \frac{y}{x}F_1(y) - 25\frac{G}{x}\int_x^\infty dy \, \frac{x}{y}F_1(y), \tag{2.10}$$

and now a behavior $F_1(x) \sim x$ as $x \to 0$ is no longer impossible, since the offending second term under the second integral in (2.9) has been thrown away. In this paper, we shall study the approximate equation (2.10) exclusively: it is hoped that the behavior x instead of $x^{1.273}$ is not too serious an error. In a future work, we propose to return to the full equation (2.9): the analysis can be completed, but it is appreciably more complicated.

The approximate equation (2.10) can be rewritten

$$\frac{x}{A + xF_1(x)} = 1 - C + 25\frac{G}{x}\int_0^x dy\left(\frac{x}{y} - \frac{y}{x}\right)F_1(y), \tag{2.11}$$

where

$$C = 25G\int_0^\infty \frac{dy}{y}F_1(y). \tag{2.12}$$

For consistency as $x \to 0$, we must impose $C = 1$; but since the integral in (2.12) needs an ultraviolet cutoff, this imposition can be regarded as a renormalization condition. Further, if we make the scaling transformations

$$x \to 5AG^{1/2}x, \quad y \to 5AG^{1/2}y, \quad F_1(x) \to \tfrac{1}{5}G^{-1/2}F_1(x), \tag{2.13}$$

then Eq. (2.11) takes on the pleasing form

$$\frac{x^2}{1 + xF_1(x)} = \int_0^x dy\left(\frac{x}{y} - \frac{y}{x}\right)F_1(y), \tag{2.14}$$

in which there are no divergences left, in which the unknown constant A, and even the coupling G, have disappeared.

III. APPROXIMATE GLUON PROPAGATOR

We first rewrite the approximate equation (2.14) in terms of the new unknown function,

$$G(x) = F_1(x)/[x + x^2F_1(x)], \tag{3.1}$$

so that

$$G(x) = \frac{1}{x^2} - \frac{1}{x^3}\int_0^x dy\left(1 - \frac{y^2}{x^2}\right)\frac{G(y)}{1 - y^2G(y)}. \tag{3.2}$$

Unfortunately, this equation is poorly suited either to numerical iteration or to an existence proof via the Banach theorem. For example, $G(0)$ is finite, but this is a result of delicate cancellation:

$$G(x) \underset{x\to 0}{\sim} \frac{1}{x^2} - \frac{1}{x^3}\int_0^x dy\left(1 - \frac{y^2}{x^2}\right)[G(0) + O(y)]$$
$$= \frac{1}{x^2}\left(1 - \tfrac{2}{3}G(0)\right) + O\left(\frac{1}{x}\right), \tag{3.3}$$

so that we must require

$$G(0) = \tfrac{3}{2}. \tag{3.4}$$

By pushing this analysis further, one can show that $G'(0)$ must vanish, that $G''(0) = -27$, and so on. In fact, in Sec. 5, we shall obtain an asymptotic (but divergent) series for $G(x)$ in the variable x^2.

To construct an integral equation that does not involve cancellations, we differentiate $x^3G(x)$, and then $x^3[x^3G(x)]'$, in order to obtain, from (3.2), the nonlinear differential equation

$$x^4G'' + 9x^3G' + (15x^2 + 2)G = 3 - 2x^2G^2/(1 - x^2G). \tag{3.5}$$

It is possible to resolve this equation in terms of the solution of the homogeneous linear differential equation [i.e., the left-hand side of (3.5) equal to zero]; but this route involves Bessel functions and their estimation in terms of simpler functions. It is more straightforward to observe that the functions

$$g_\pm(x) = x^{-7/2}\exp(\pm i2^{1/2}/x) \tag{3.6}$$

solve the homogeneous equation

$$x^4g_\pm'' + 9x^3g_\pm' + (\tfrac{63}{4}x^2 + 2)g_\pm = 0. \tag{3.7}$$

This is almost like the left-hand side of (3.5), excepting only that the coefficient of $x^2g_\pm$ is not quite correct! However, by adding $\frac{3}{4}x^2G$ to both sides of (3.5), we obtain

$$x^4G'' + 9x^3G' + (\tfrac{63}{4}x^2 + 2)G$$
$$= 3 + \tfrac{3}{4}x^2G - 2x^2G^2/(1 - x^2G). \tag{3.8}$$

This equation can be resolved in terms of $g_\pm$ by the method of variation of parameters. The linear term $\frac{3}{4}x^2G$ on the right-hand side of (3.8) will not give trouble for small x, thanks to the factor x^2. We find

$$G(x) = 2^{-1/2}x^{-7/2}\int_0^x dy\,\sin\left[2^{1/2}\left(\frac{1}{x} - \frac{1}{y}\right)\right]$$
$$\times\left\{-3y^{3/2} - \tfrac{3}{4}y^{7/2}G(y) + \frac{2y^{7/2}G^2(y)}{1 - y^2G(y)}\right\}. \tag{3.9}$$

It can easily be checked that this is the correct solution of the differential equation: additional multiples of g_+ or g_- would be inconsistent with (3.4), and thus with (3.2).

The form (3.9) is suitable for an existence proof; but the argument of the sine factor makes matters somewhat awkward. To simplify matters, define

$$\xi = 2^{1/2}/x, \quad \tilde{G}(\xi) = G(x) \tag{3.10}$$

and

$$\zeta = 2^{1/2}/y - 2^{1/2}/x, \tag{3.11}$$

so that

$$\tilde{G}(\xi) = f(\xi) + \xi^{7/2}\int_0^\infty d\zeta\,\frac{\Sigma(\zeta + \xi)}{(\zeta + \xi)^{11/2}}\sin\zeta, \tag{3.12}$$

where

$$f(\xi) = \tfrac{3}{2}\xi^{7/2}\int_0^\infty \frac{d\zeta}{(\zeta + \xi)^{7/2}}\sin\zeta, \tag{3.13}$$

and

$$\Sigma(\zeta) = \tfrac{3}{4}\tilde{G}(\zeta) - \frac{2\tilde{G}^2(\zeta)}{1 - 2\tilde{G}(\zeta)/\zeta^2}. \tag{3.14}$$

This equation will now be used for an existence proof for $|\xi|$ large enough, i.e., $|x|$ small enough, for all $|\arg x| \leqslant \pi - \epsilon$, $\epsilon > 0$.

IV. PROOF OF EXISTENCE

Let $D(\rho,\epsilon)$ be the following domain in the complex ξ plane:

$$D(\rho,\epsilon) = \{\xi\,|\mathrm{Re}\xi \geqslant 0, |\xi| > \rho^{-1};$$
$$(\mathrm{Re}\xi < 0, (|\mathrm{Im}\xi| - \rho^{-1})/|\mathrm{Re}\xi| > \tan\epsilon\}. \tag{4.1}$$

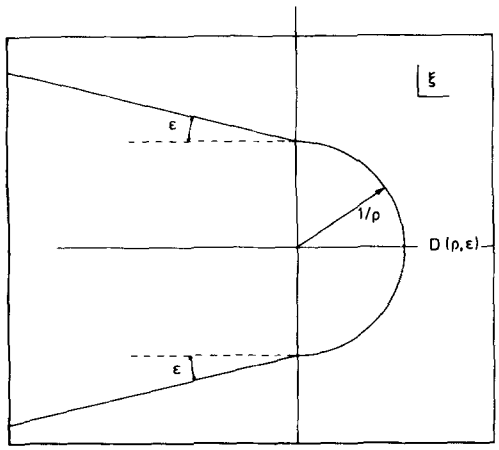


FIG. 2. The domain $D(\rho,\epsilon)$ lies *outside* the comet-shaped region depicted.

In words, it is the region outside a semicircle in the right half-plane and above or below lines inclined at an angle ϵ to the real axis (see Fig. 2). Let B be the Banach space of functions, $f(\xi)$, analytic in D, with norm

$$||f|| = \sup_{\xi\in D(\rho,\epsilon)} |f(\xi)|. \tag{4.2}$$

Let P be the mapping

$$P(\tilde{G};\xi) = f(\xi) + \xi^{7/2}\int_0^\infty d\zeta \frac{\Sigma(\zeta+\xi)}{(\zeta+\xi)^{11/2}}\sin\zeta, \tag{4.3}$$

where f and Σ were defined in (3.13) and (3.14). It is clear that, if $\xi\in D(\rho,\epsilon)$, for a given $\rho>0$ and $0<\epsilon<\pi/2$, then $\zeta+\xi\in D(\rho,\epsilon)$ for all $0<\zeta<\infty$. Let $\tilde{G}$ lie in a ball defined by

$$\{\tilde{G}\,|\tilde{G}\in B, ||\tilde{G}||\leqslant b\}. \tag{4.4}$$

Then $\Sigma(\zeta+\xi)$ is analytic in ξ, for any fixed ζ in $(0,\infty)$, on condition that the denominator in (3.14) does not vanish. This can be prevented for any $\tilde{G}$ satisfying (4.4) by restricting ρ as follows:

$$\rho^2 < 1/2b. \tag{4.5}$$

Under this condition, the integral term in (4.3) is analytic for $\xi\in D(\rho,\epsilon)$, since the integrand is so analytic, and the integral converges uniformly. Clearly $f(\xi)$, the known function (3.13), is also analytic in this domain: in fact it is analytic in the plane, cut $-\infty<\xi<0$, as we show in Appendix B. Thus $P(\tilde{G};\xi)$ is analytic for $\xi\in D(\rho,\epsilon)$.

We shall now show that, if one imposes further constraints on ρ and b, P maps the ball (4.4) into itself contractively. The Banach theorem then applies, and so we assert the existence and uniqueness of a solution of the equation

$$\tilde{G}(\xi) = P(\tilde{G};\xi) \tag{4.6}$$

in the ball (4.4).

When $\tilde{G}$ satisfies (4.6) and ρ is restricted by (4.5), we see from (3.14) that

$$|\Sigma(\zeta)| \leqslant \frac{3b}{4} + \frac{2b^2}{1-2b\rho^2} \tag{4.7}$$

for all $\zeta\in D(\rho,\epsilon)$. Hence

$$|P(\tilde{G};\xi)| \leqslant C_\epsilon + \left(\frac{3b}{4} + \frac{2b^2}{1-2b\rho^2}\right)|\xi|^{7/2} \times\int_0^\infty \frac{d\zeta}{|\zeta+\xi|^{11/2}}, \tag{4.8}$$

where C_ϵ is the bound (B4) on $|f(\xi)|$. We change the integration variable to $\omega=\zeta/|\xi|$ and define

$$D_\epsilon = \int_0^\infty \frac{d\omega}{|\omega - e^{i\epsilon}|^{11/2}}. \tag{4.9}$$

Then

$$|P(\tilde{G};\xi)| \leqslant C_\epsilon + \rho\left(\frac{3b}{4} + \frac{2b^2}{1-2b\rho^2}\right)D_\epsilon. \tag{4.10}$$

If

$$b > C_\epsilon \tag{4.11}$$

and

$$\rho \leqslant (b - C_\epsilon)\left(\frac{3b}{4} + \frac{2b^2}{1-2b\rho^2}\right)D_\epsilon, \tag{4.12}$$

then the right-hand side of (4.10) is not greater than b: the ball (4.4) has been mapped into itself.

To demonstrate the contractivity, we differentiate (3.14) with respect to $\tilde{G}(\zeta)$:

$$\frac{d\Sigma(\zeta)}{d\tilde{G}(\zeta)} = \frac{3}{4} - \frac{4\tilde{G}(\zeta)[1-\zeta^{-2}\tilde{G}(\zeta)]}{[1-2\zeta^{-2}\tilde{G}(\zeta)]^2}. \tag{4.13}$$

Since (4.5) implies that

$$|\zeta^{-2}\tilde{G}(\zeta)| \leqslant b\rho^2 < \tfrac{1}{2}, \tag{4.14}$$

it follows that

$$|1-\zeta^{-2}\tilde{G}(\zeta)| \leqslant \tfrac{3}{2}. \tag{4.15}$$

Hence

$$\left|\frac{d\Sigma(\zeta)}{d\tilde{G}(\zeta)}\right| \leqslant \frac{3}{4} + \frac{6b}{[1-2b\rho^2]^2}. \tag{4.16}$$

Now let $\tilde{G}_1(\zeta)$ and $\tilde{G}_2(\xi)$ be any two functions in the ball (4.4). The mean value theorem implies

$$|\Sigma_1(\zeta) - \Sigma_2(\zeta)| \leqslant \sup_{0<\mu<1}\left|\frac{d\Sigma(\zeta)}{d\tilde{G}(\zeta)}\right|_{\tilde{G}=\tilde{G}_\mu} |\tilde{G}_1(\zeta) - \tilde{G}_2(\zeta)|, \tag{4.17}$$

where $\tilde{G}_\mu = \mu\tilde{G}_1 + (1-\mu)\tilde{G}_2$. Now since the ball (4.4) is quintessentially convex, $\tilde{G}_\mu$ belongs to it, and we may use the bound (4.16) for the supremum in (4.17). By the same analysis as in Eq. (4.8) *et seq.*, it follows that

$$||P(\tilde{G}_1) - P(\tilde{G}_2)|| \leqslant \rho\left[\tfrac{3}{4} + 6b/(1-2b\rho^2)^2\right]D_\epsilon\,||\tilde{G}_1 - \tilde{G}_2||. \tag{4.18}$$

The mapping is contractive if this Lipschitz coefficient is strictly less than unity, i.e.,

$$\rho < \left|\left[\tfrac{3}{4} + 6b/(1-2b\rho^2)^2\right]D_\epsilon\right|^{-1}. \tag{4.19}$$

The conditions for a contraction mapping are (4.5), (4.11), (4.12), and (4.19). They are clearly consistent, since if b is not too small, then ρ can certainly be made so small that (4.5), (4.11), and (4.12) are all satisfied. Since C_ϵ and D_ϵ [Eqs. (B4) and (4.9)] tend to infinity as $\epsilon\to 0$, it follows that the permissible values of ρ tend to zero in the same limit. Accordingly, one can imagine repeating the proof for various

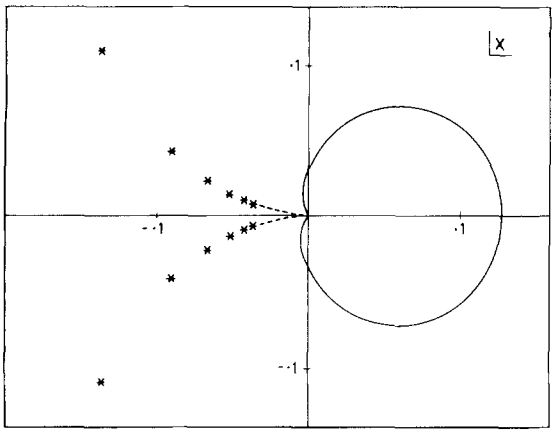


FIG. 3. Cardioid region in which the Banach theorem applies. The stars outside this region are the locations of branch points, as determined by the numerical method of Sec. 5.

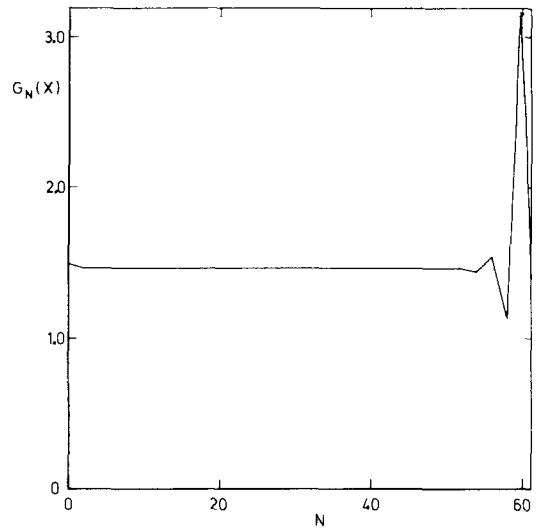


FIG. 4. The partial sums $G_N(x)$ for $x = 0.05$ and N values up to 60 [see Eq. (5.4a)].

values of ϵ, starting with $\epsilon = \pi/2$ and gradually allowing ϵ to decrease to zero. For ever-decreasing values of ϵ, one proves existence and uniqueness of an analytic solution in the corresponding domain $D(\rho,\epsilon)$ of Fig. 2, but with ever-increasing values of the radius $1/\rho$. This means that, in the original variable, $x = 2^{1/2}/\xi$, of Eq. (3.9), one has a proof of existence in the union of the corresponding domains, which has been plotted in Fig. 3. This curve has been obtained by computing C_ϵ and D_ϵ by a numerical integration routine, and then by finding the largest value of ρ consistent with (4.12) and (4.19), as b varies between C_ϵ and infinity. This was done in practice by rewriting the inequality (4.12) as a negative cubic form, and by determining numerically, for increasing values of b, larger than C_ϵ, the real positive root. We then determined the largest value of ρ that was consistent with (4.19).

The angle made by the locus of Fig. 3 to the real axis at the origin is zero. Equation (3.9), and thus also the original equation (3.2), has an analytic solution inside the cardioid of Fig. 3. Of course, the solution has a continuation outside the cardioid; but it will in general have singularities, the location of which we shall ascertain numerically.

V. NUMERICAL ANALYSIS

In order to set up a computer program to effect the continuation out of the contractive region of Sec. 4, we return to the differential equation (3.5), which can be rewritten in the form

$$(1 - x^2G)(x^4G'' + 9x^3G' + 15x^2G) = 3 - (2 + 3x^2)G. \quad (5.1)$$

On substituting a formal power series solution,

$$G(x) = \sum_{n=0}^{\infty} G_n x^n, \quad (5.2)$$

we find that the odd terms vanish, $G_0 = \frac{3}{2}$, $G_2 = -\frac{27}{2}$ and, for $n \geqslant 4$,

$$G_n = \frac{1}{2}\sum_{m=2}^{n-2}(m^2 + 4m + 3)G_{m-2}G_{n-m-2} - \tfrac{1}{2}(n^2 + 4n + 6)G_{n-2}. \quad (5.3)$$

This relation allows G_n to be determined recursively in terms of $G_m, m < n$. However, it is easy to show by estimates that $(G_n)^{-1/n} \to 0$ as $n \to \infty$. This means that the series (5.2) has a zero radius of convergence, and therefore that $G(x)$ has a singularity at the origin. This is the expected branch point caused by soft-gluon effects. Although (5.2) is divergent for any x, it can be used as an asymptotic series for small x. This is important, since the Runge–Kutta routine cannot be used at the origin, due to the existence of the singularity.

In Fig. 4, we show how the partial sum,

$$G_N(x) = \sum_{n=0}^{N} G_n x^n, \quad (5.4a)$$

depends on N for the typical value, $x = 0.05$. Any value of N between 15 and 52 gives a stable value of $G_N(x)$ to five decimal places, although the series blows up as N is further increased (see Fig. 4). The derivative of (5.4a),

$$G_N'(x) = \sum_{n=1}^{N} nG_n x^{n-1}, \quad (5.4b)$$

was also computed. At $x = 0.05$, this gave four-decimal stability for N between 15 and 35. In fact, we finally used the program for the still smaller value, $x = 0.01$, since this gives G_N and G_N', stable to more than ten significant figures, for N up to and beyond 50. We find

$$G(0.01) = 1.498\,652\,724, \quad (5.5a)$$

$$G'(0.01) = -0.268\,912\,765, \quad (5.5b)$$

and this was used as the initial point for the production runs of the Runge–Kutta routine.

As a check on the reliability of the method, we have recalculated $G(x)$ from the formula

$$G(x) = \frac{3}{2} + 2^{-1/2}x^{-7/2}\int_0^x dy\, y^{7/2}\sin\left[2^{1/2}\left(\frac{1}{x} - \frac{1}{y}\right)\right] \times\left(\frac{189}{8} - \frac{3}{4}G(y) + 2\frac{G^2(y)}{1 - y^2G(y)}\right), \quad (5.6)$$

which is obtained from (3.9) by two partial integrations. The expression (5.6) was iterated 10 times, the initial value for

$G(y)$ on the right-hand side being 3/2. The integral was performed at 5000 equally spaced points between 0 and x; and this then allowed the integrand to be determined at the same points for the next iterate. A four-point finite-difference formula was used to effect the quadrature. Convergence of the iteration was rapid, and agreement with the result of the asymptotic series, for small x, was good. For example, the routine yielded

$$G(0.01) = 1.498\,658\,026, \tag{5.7}$$

which agrees to five places of decimals with (5.5a). The value (5.5a), from the series, was stable to beyond the ninth decimal, and is the more accurate value.

In order to apply the Runge–Kutta integration procedure, the differential equation (3.5) was rewritten in the form

$$G''(x) = f(x,G,G') \equiv \frac{3}{x^4} - \frac{15x^2+2}{x^4}G(x) - \frac{2G^2(x)}{x^2[1-x^2G(x)]} - \frac{9}{x}G'(x). \tag{5.8}$$

Given $G(x)$ and $G'(x)$ at $x = x_n$, a standard formula[6] estimates these functions at the next point, $x = x_{n+1}$. It was found that satisfactory accuracy was obtained (to at least 7 significant figures) if 5000 steps were taken between points of interest. Close to the origin, large cancellations occur, due to the denominators in (5.8); and so the first few steps must be cautious. In practice, we found that 5000 steps from $x = 0.01$, as given by (5.5), to $x = 0.1$, followed by 5000 to $x = 1.0$, preserved good accuracy. Complex excursions could then be made without danger from rounding errors; and in particular, a subroutine called LOOP continued $G(x)$ around a rectangle. The initial and final values of $G(x)$ and $G'(x)$ were the same, typically to an accuracy of about 10^{-10}, unless a branch point was enclosed in the rectangle in question, in which case the mismatch was of order 1 or more. Thus it was easy to distinguish between cases in which a branch point was, or was not, enclosed within a given rectangle; and an automatic routine, in which a rectangle containing a branch point was systematically halved and subjected to the test, enabled branch points to be located to good accuracy. The method had earlier been developed[4] to handle a similar problem in QED.

As a check on the reliability of this program, we repeated the procedure with different real and complex values of x in the asymptotic series (5.4), which defined the starting point for the Runge–Kutta routine. For stability, the starting point could not be too far from the origin, for then the asymptotic series was useless, nor too close to the origin, for then the cancellations between the various terms in (5.8) were too fierce. Fortunately, an intermediate region exists, in which stable results were obtained. An oval region around the origin, defined roughly by ± 0.03 in the real, and ± 0.06 in the imaginary direction, is inaccessible to the Runge–Kutta routine, because of large cancellations.

In Table I, and in Fig. 3, we show the locations of six branch points that we have discovered in the second quadrant. We expect that an infinite number of them exists, accumulating at the origin; but we were unable to approach the origin more closely, to pick up more branch points, because of numerical instabilities. Since the cardioid domain, in which the existence theorem of Sec. 4 works, has a horizontal tangent at the origin it follows that, if an infinite number of branch points do accumulate at the origin, they must approach it along a curve that is asymptotic to the negative real axis (see Fig. 3).

TABLE I. Location of the first six branch points of $G(x)$.

n	Re x	Im x
1	−0.1355	0.1110
2	−0.0903	0.0428
3	−0.0658	0.0226
4	−0.0514	0.0141
5	−0.0421	0.0097
6	−0.0356	0.0071

It is interesting to compare the above results with those of the averaged case that is treated in Appendix C. There, it is proved that an infinite number of branch points also accumulate at the origin, but along the lines $\arg x = \pm 3\pi/4$. Apparently the averaging procedure has aggravated the situation, for first-sheet complex singularities should not be present in a healthy theory. They are inconsistent with causality, and they spoil the Wick rotation that relates the Lorentz and Euclidean equations. Their occurrence must be regarded as a sickness of the Mandelstam approximation; and it is of importance to improve the treatment of the Dyson–Schwinger equation, in order to cure the disease. A hopeful sign is the fact that the equation of Sec. 3 appears to be less sick than that of Appendix C. Perhaps a relatively minor modification of the Mandelstam approximation will cause the complex branch points to be pushed on to a secondary Riemann sheet, where they can be tolerated.

ACKNOWLEDGMENTS

We should like to thank M. Baker, H. Boelens, K. Dietz, S. Mandelstam, and I. S. Stefanescu for helpful remarks.

APPENDIX A

In this appendix, we shall prove Eq. (1.5). In polar coordinates, the left-hand side of (1.5) can be written

$$-\frac{1}{4\pi}\lim_{\lambda\to\infty}\int_0^\lambda r^3dr\int d^3k\,\Omega(\mathbf{k})\int_{-1}^1\cos\theta\int_0^{2\pi}d\phi\,e^{i|\mathbf{k}|r\cos\theta}$$
$$=\lim_{\lambda\to\infty}\int d^3k\,\Omega(\mathbf{k})\left(\frac{2}{k^4} - \frac{2\cos k\lambda}{k^4} - \frac{2\lambda\sin k\lambda}{k^3} + \frac{\lambda^2\cos k\lambda}{k^2}\right), \tag{A1}$$

where $k = |\mathbf{k}|$, and where the r, θ, and ϕ integrals have been performed inside the k integral, this being permissible for finite λ. To show that the last three terms vanish in the limit $\lambda\to\infty$, we use polar coordinates in the k variable and perform integrations by parts: three for the fourth term, two for the third term, and one for the second term. For example,

$$\int d\Omega_k\int_0^\infty dk\,\Omega(\mathbf{k})\frac{\cos k\lambda}{k^2} = \frac{1}{\lambda}\int d\Omega_k\int_0^\infty dk\,\sin k\lambda\,\frac{\partial}{\partial k}\left(\frac{\Omega(\mathbf{k})}{k^2}\right). \tag{A2}$$

The vanishing of $\Omega(\mathbf{k})$ and its first three derivatives at the origin is essential.

APPENDIX B

In this appendix we shall study the function $f(\xi)$, defined in Eq. (3.13). We shall show that, for any ϵ lying in $(0,\pi)$, $|f(\xi)|$ is bounded by a constant C_ϵ when $|\arg\xi| \leqslant \pi - \epsilon$. However, as $\epsilon \to 0$, so $C_\epsilon \to \infty$. Thus $|f(\xi)|$ is bounded in $D(\rho,\epsilon)$, for any nonzero ϵ.

Integrating (3.13) by parts, we find

$$f(\xi) = \frac{3}{2} - \frac{21}{4}\xi^{7/2}\int_0^\infty \frac{d\zeta}{(\zeta+\xi)^{9/2}}\cos\zeta. \tag{B1}$$

We change the integration variable to

$$\omega = \zeta/|\xi|, \tag{B2}$$

so that

$$f(\xi) = \frac{3}{2} - \frac{21}{4}e^{7i\theta/2}\int_0^\infty \frac{d\omega}{(\omega+e^{i\theta})^{9/2}}\cos(\omega|\xi|), \tag{B3}$$

where $\theta = \arg\xi$. Hence, if $|\theta| \leqslant \pi - \epsilon$,

$$|f(\xi)| \leqslant \frac{3}{2} + \frac{21}{4}\int_0^\infty \frac{d\omega}{|\omega - e^{i\epsilon}|^{9/2}} \equiv C_\epsilon, \tag{B4}$$

which is finite if $\epsilon > 0$. It may easily be shown that $C_\pi = 3$ and $C_{\pi/2} \leqslant \frac{3}{2} + 21\pi/16 \simeq 5.6$, i.e., $|f(\xi)|$ is bounded by 3 on the positive real axis and by 5.6 in the right half-plane.

In order to examine the behavior of $f(\xi)$ on the negative real axis, we need to modify the expression (3.13), since this diverges if ξ is real and negative, and it then no longer represents the continuation of $f(\xi)$. Integrate twice by parts the other way:

$$f(\xi) = -\frac{2}{5}\xi^2 + \frac{2}{5}\xi^{7/2}\int_0^\infty \frac{d\zeta}{(\zeta+\xi)^{3/2}}\sin\zeta. \tag{B5}$$

This still fails to exist for ξ real and negative; but we dare not integrate once more by parts, for this would cause the integral to diverge at infinity. Instead add and subtract $\sin\xi$ to $\sin\zeta$ in (B5). The result can be written

$$f(\xi) = -\tfrac{2}{5}\xi^2 - \tfrac{4}{5}\xi^3\sin\xi + \tfrac{2}{5}\xi^3 g(\xi), \tag{B6}$$

where

$$g(\xi) = \xi^{1/2}\int_0^\infty \frac{d\zeta}{(\zeta+\xi)^{3/2}}[\sin\zeta + \sin\xi]. \tag{B7}$$

The integral (B7) exists for ξ real and negative, and it effects the analytic continuation of $f(\xi)$. Substitute $\eta = \zeta + \xi$, for the moment keeping ξ real and *positive*:

$$g(\xi) = \xi^{1/2}\int_\xi^\infty \frac{d\eta}{\eta^{3/2}}[\sin\xi(1-\cos\eta) + \cos\xi\sin\eta]. \tag{B8}$$

Now since

$$\int_0^\infty \frac{d\eta}{\eta^{3/2}}(1-\cos\eta) = (2\pi)^{1/2} = \int_0^\infty \frac{d\eta}{\eta^{3/2}}\sin\eta, \tag{B9}$$

we may write

$$g(\xi) = (2\pi)^{1/2}\xi^{1/2}[\sin\xi + \cos\xi] - g_1(\xi)\sin\xi - g_2(\xi)\cos\xi, \tag{B10}$$

where

$$g_1(\xi) = \xi^{1/2}\int_0^\xi \frac{d\eta}{\eta^{3/2}}[1-\cos\eta], \tag{B11a}$$

$$g_2(\xi) = \xi^{1/2}\int_0^\xi \frac{d\eta}{\eta^{3/2}}\sin\eta, \tag{B11b}$$

which are closely related to the Fresnel integrals. Scale η by setting $\omega = \eta/\xi$:

$$g_1(\xi) = \int_0^1 \frac{d\omega}{\omega^{3/2}}[1-\cos(\omega\xi)], \tag{B12a}$$

$$g_2(\xi) = \int_0^1 \frac{d\omega}{\omega^{3/2}}\sin(\omega\xi). \tag{B12b}$$

Now we continue ξ away from the real, positive axis, and discover that g_1 and g_2 are entire functions! This means that the only contribution to $\mathrm{Im}\, f(\xi)$ for ξ real and *negative* comes from the $\xi^{1/2}$ term in (B10). Thus $f(\xi)$ has a cut $-\infty < \xi < 0$, and on it

$$\mathrm{Im}\, f(\xi \pm i0) = \mp\tfrac{2}{5}(2\pi)^{1/2}|\xi|^{7/2}[\sin\xi + \cos\xi]. \tag{B13}$$

For ξ real and negative, it is easy to prove from (B12) that

$$|g_1(\xi)| \leqslant 6\xi^{1/2}, \tag{B14a}$$

$$|g_2(\xi)| \leqslant 4\xi^{1/2}. \tag{B14b}$$

Combining these results together, we conclude that, for ξ real and negative,

$$|\mathrm{Re}\, f(\xi)| \leqslant \tfrac{2}{5}|\xi|^2 + \tfrac{2}{5}|\xi|^3 + 4|\xi|^{7/2}, \tag{B15a}$$

$$|\mathrm{Im}\, f(\xi)| \leqslant \tfrac{4}{5}\pi^{1/2}|\xi|^{7/2}. \tag{B15b}$$

Hence there exists a constant C such that

$$|f(\xi)| \leqslant C|\xi|^{7/2}, \tag{B16}$$

when ξ is real and negative.

In (B4) we showed that $|f(\xi)|$ is bounded by a constant when $|\arg\xi| \leqslant \pi - \epsilon$; and in (B16) we have a power bound on the negative real axis. In fact we can use the Phrágmen–Lindelöf theorem to extend the bound (B16) to the wedge $\pi - \epsilon < |\arg\xi| \leqslant \pi$. In order to do this, we need to be sure that $|f(\xi)|$ does not behave too badly as $|\xi| \to \infty$ in this wedge. At first sight the trigonometric functions in (B6) and (B12) look sick for large, complex ξ, and indeed they are, but they must cancel, because if we go back to (B5) and substitute

$$\frac{1}{\zeta+\xi} = \frac{1}{\zeta+1} + \frac{1-\xi}{(\zeta+\xi)(\zeta+1)}, \tag{B17}$$

we find

$$f(\xi) = -\tfrac{2}{5}\xi^2 + \tfrac{4}{5}\xi^{7/2}\int_0^\infty \frac{d\zeta}{(\zeta+\xi)^{1/2}} \times\left[\frac{\sin\xi}{\zeta+1} + 2(1-\xi)\left(\frac{\cos\xi}{\zeta+1} - \frac{\sin\zeta}{(\zeta+1)^2}\right)\right], \tag{B18}$$

after an integration by parts. This integral exists in the whole cut ξ plane, and it is easy to obtain the crude upper bound

$$|f(\xi)| \leqslant C|\xi|^5, \tag{B19}$$

where C is some constant. This is a poor bound, and it is only needed to make sure that the Phrágmen–Lindelöf theorem

really applies in the wedge $\pi - \epsilon < |\arg\xi| \leqslant \pi$, so that actually (B19) can be tightened to (B16) in the wedge.

The above fulsome treatment of $f(\xi)$ was motivated by the fact that this function may be regarded as the zeroth approximation to $\tilde{G}(\xi)$, and thus to the Mandelstam gluon propagator itself.

APPENDIX C

Here we shall approximate the integral equation (3.2) still further by replacing the kernel $(1 - y^2/x^2)$ by its average value, namely $\frac{2}{3}$. By a simultaneous rescaling of x, y, and G, this multiplicative factor can be removed [cf. Eq. (2.13) *et seq.*]. The result is

$$G(x) = \frac{1}{x^2} - \frac{1}{x^3}\int_0^x dy \frac{G(y)}{1 - y^2 G(y)}, \tag{C1}$$

and it is interesting to study this averaged equation, since it turns out to be explicitly soluble. A similar averaging in the QED case[4] was shown to have minor quantitative, but not qualitative, effects on the solution.

From (C1) we deduce the differential equation

$$\frac{d}{dx}[x^3 G(x)] = 1 - \frac{G(x)}{1 - x^2 G(x)}, \tag{C2}$$

which is greatly simplified if we substitute

$$\psi(x) = x - 1/x - x^3 G(x), \tag{C3}$$

for then

$$\frac{dx}{d\psi} = 1 + x\psi, \tag{C4}$$

which is a linear equation for x as a function of ψ, with the solution

$$x(\psi) = \int_0^\infty d\omega \exp[\omega\psi - \tfrac{1}{2}\omega^2]. \tag{C5}$$

This is the relevant solution of (C4), since it corresponds to the boundary condition $x \to 0$ through positive values as $\psi \to -\infty$, which is consistent with (C3) and (C1). Evidently x is an entire function of ψ, and hence the only singularities of ψ as a function of x are the points where

$$\frac{dx}{d\psi} = 0. \tag{C6}$$

These points can be evaluated numerically, and the work has already been done, for

$$x(\psi) = (\pi/2)^{1/2} w(-i2^{-1/2}\psi), \tag{C7}$$

where

$$w(z) = e^{-z^2}\left(1 + 2\pi^{-1/2} i \int_0^z dt\, e^{t^2}\right), \tag{C8}$$

and its derivative have been tabulated in the literature.[7] Using these results, we find that $\psi(x)$, and therefore $G(x)$, have an infinite number of first-sheet branch points that accumulate at the origin along the asymptotes $\arg x = \pm 3\pi/4$ (see Fig. 5).

The differential equation (C2), together with an asymptotic series like that of Sec. 5, have been used for testing the Runge–Kutta program: the first few of the branch points shown in Fig. 5 are picked up without difficulty.[8]

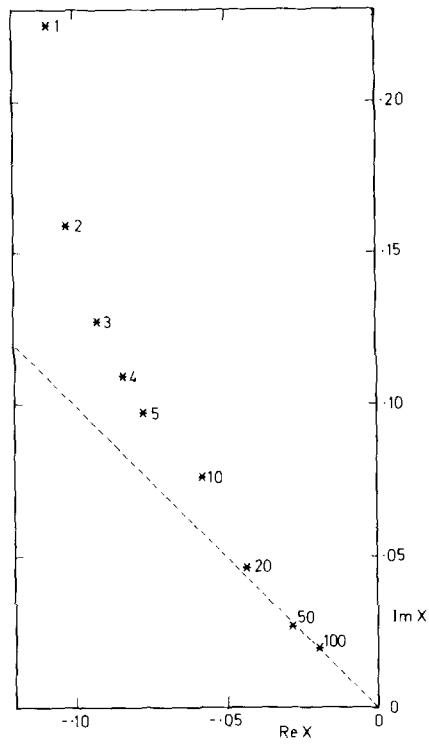


FIG. 5. Some of the branch points in the upper half-plane for the averaged equation (C1), which have been numbered, starting from the one most distant from the origin. The branch points accumulate at the origin along the line $\arg x = 3\pi/4$.

Although an existence proof along the lines of that given in Sec. 4 is strictly superfluous, since (C5) gives a representation of the solution, albeit in inverse form, it is nevertheless instructive. Without giving all the details,[9] we quote the analogs of Eqs. (3.12)–(3.14):

$$\tilde{G}(\xi) = f(\xi) - \xi^3 \int_0^\infty \frac{d\zeta}{(\zeta + \xi)^2} e^{-\xi\zeta - (1/2)\zeta^2} \Sigma(\zeta + \xi), \tag{C9}$$

where

$$f(\xi) = \xi^3 \int_0^\infty \frac{d\zeta}{(\zeta + \xi)^2} e^{-\xi\zeta - (1/2)\zeta^2}, \tag{C10}$$

and

$$\Sigma(\zeta) = \frac{\tilde{G}^2(\zeta)}{\zeta^2 - \tilde{G}(\zeta)}, \tag{C11}$$

where $\xi = 1/x$, etc. An existence proof can be completed for $|x|$ small (i.e., $|\xi|$ large) only if $|\arg x| \leqslant 3\pi/4 - \epsilon$, instead of $|\arg x| \leqslant \pi - \epsilon$, as was possible in Sec. 4. The reason is that $f(\xi)$ blows up as $|\xi| \to \infty$ when $\pi \geqslant |\arg\xi| > 3\pi/4$. In fact, one can show easily that $f(\xi)$ is bounded if $\mathrm{Re}\,\xi \geqslant 0$, and also that

$$f(-\xi) = f(\xi) - (\pi/2)^{1/2} \xi^3 e^{(1/2)\xi^2}, \tag{C12}$$

from which the results follow. The failure of the proof for $|\arg\xi| \geqslant 3\pi/4 - \epsilon$ is of course due to the existence of the branch points of Fig. 5.

It is interesting that Eq. (3.2) is "softer" than is Eq. (C1), in the sense that the accumulation of branch points has been pushed in the former, as compared with the latter equation, from $|\arg\xi| = 3\pi/4$ to $|\arg\xi| = \pi$.

[1]S. Mandelstam, Phys. Rev. D **20**, 3223 (1979).
[2]M. Baker, J. S. Ball, P. Lucht, and F. Zachariasen, Phys. Lett. B **89**, 211 (1980); S. K. Kim and M. Baker, Nucl. Phys. B **164**, 152 (1980); R. Anishetty, M. Baker, S. K. Kim, J. S. Ball, and F. Zachariasen, Phys. Lett. B **86**, 52 (1979); J. S. Baker and F. Zachariasen, Nucl. Phys. B **143**, 148 (1978).
[3]B. G. Dragović, D. P. Mavlo, and A. T. Filippov, Fizika (Yugoslavia) **10**, 51 (1978).
[4]D. Atkinson and D. W. E. Blatt, Nucl. Phys. B **151**, 342 (1979).
[5]D. Atkinson and M. P. Fry, Nucl. Phys. B **516**, 301 (1979); D. Atkinson, M. P. Fry, and E. J. Luit, Lett. Nuovo Cimento **26**, 413 (1979).
[6]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965), p. 897, formula 25.5.20.
[7]V. N. Faddeyeva and N. M. Terent'ev, *Mathematical Tables of the Values of the Function w(z)* (Pergamon, New York, 1961).
[8]See J. K. Drohm, Groningen University, Internal Report No. 166 (1981).
[9]See K. Stam, Groningen University, Internal Report No. 169 (1981).

Generalization of Stokes theorem and pure gauge fields

M. Iyanaga
Department of Applied Mathematics, Faculty of Engineering, Yamaguchi University, Tokiwadai, Ube 755 Japan

(Received 16 April 1981; accepted for publication 5 June 1981)

A generalization of the Stokes theorem is derived for the path-order phase factor in nonabelian gauge theories. It is proved that the curvature form vanishes identically if and only if the connection form is pure gauge.

PACS numbers: 11.10.Np, 11.30. – j, 02.20.Sv

In the case of nonabelian gauge fields, the Stokes theorem is formulated with the help of the differentiation of the path-ordered phase factor[1] with respect to path. It is shown that the usual Stokes theorem[2] is derived from this non-abelian Stokes theorem in case of abelian gauge fields. Furthermore we get from this theorem that the curvature form vanishes identically if and only if the connection form is pure gauge, and we see that this proposition becomes the Poincaré theorem[2] in case of abelian gauge fields. Finally it is noted that pure gauge connection form is also obtained by requiring the existence of a geodesic moving frame[3] for a neighborhood of each point of the manifold.

1. DIFFERENTIATION OF PATH-ORDERED PHASE FACTORS

Let M be the pseudo-Riemannian manifold and S be the set of smooth paths $\gamma: x = x(s)$, $s \in I = [0,1]$, in M. Suppose that two paths $\gamma_0: x_0 = x_0(s)$ and $\gamma_1: x_1 = x_1(s)$ belong to S. Then γ_0 and γ_1 are said to be homotopy equivalent if there exists a map x of class $C^{(r)}$, $x: I \times I \to M$, such that $x(s,0) = x_0(s)$ and $x(s,1) = x_1(s)$. A map $\gamma(t): I \to M$ such that $\gamma(t)(s) = x(s,t)$, belongs to S. Such a smooth path $\gamma(t)$ is called the path depending on a parameter t.

Now let us consider the path-ordered phase factors as follows:

$$U[\gamma(t)] = P \exp\left(\int_{\gamma(t)} A_\mu \, dx^\mu\right),$$

where A_μ are elements of the Lie algebra of the group SU(n) or the Lorentz group. We shall start from the calculation of $U[\gamma(t+\delta t)]$. Suppose that two smooth paths $\gamma(t)$, $\gamma(t+\delta t)$ are expressed by the equations $x(s,t) = (x^\mu(s,t))$, $x(s,t+\delta t) = (x^\mu(s,t+\delta t))$ in a local coordinate system and that we set $\Delta x_i^\mu = x^\mu(s_i,t) - x^\mu(s_{i-1},t)$, $\delta x_i^\mu = x^\mu(s_i, t+\delta t) - x^\mu(s_i,t)$, $\Delta \bar{x}_i^\mu = x^\mu(s_i, t+\delta t) - x^\mu(s_{i-1},t+\delta t)$, $\xi_i = x(\xi_i,t)$, $\bar{\xi}_i = x(\xi_i, t+\delta t)$, where s_i are the dividing points of $[0,1]$ as follows: $0 = s_0 < s_1 < s_2 < \cdots < s_{N-1} < s_N = 1$ and ξ_i is some point of the ith segment $[s_{i-1}, s_i]$. Then we have from the definition[3] of path-ordered phase factors

$$U[\gamma(t+\delta t)] = \lim_{\|\bar{\Delta}\| \to 0} \prod_{i=1}^{N} (I + A_\mu(\bar{\xi}_i)\Delta \bar{x}_i^\mu),$$

$$U[\gamma(t)] = \lim_{\|\bar{\Delta}\| \to 0} \prod_{i=1}^{N} (I + A_\mu(\xi_i)\Delta x_i^\mu),$$

where I is the unit matrix, $\|\Delta\| = \max_i\{\|\Delta x_i\|\}$ and $\|\bar{\Delta}\| = \max_i\{\|\Delta \bar{x}_i\|\}$. Setting $\bar{A}_i = I + A_\mu(\bar{\xi}_i)\Delta \bar{x}_i^\mu$, $A_i = I + A_\mu(\xi_i)\Delta x_i^\mu$, and $B_i = I + A_\mu(\xi_i)\delta x_i^\mu$, we first calculate $B_i \bar{A}_i B_{i+1}^{-1} A_i^{-1}$. As can be seen by a simple calculation, we get

$$\bar{A}_i = I + A_\mu(\xi_i)\Delta x_i^\mu + \partial_\mu A_\nu(\xi_i)\delta x_i^\mu \Delta x_i^\nu + O(\delta t)^2,$$
$$B_{i+1}^{-1} = I - A_\mu(\xi_i)\delta x_i^\mu - \partial_\nu A_\mu(\xi_i)\delta x_i^\mu \Delta x_i^\nu + O(\Delta s)^2,$$

where $\Delta s = \max_i\{(s_i - s_{i-1})\}$. Hence a straightforward calculation leads to the following result,

$$B_i \bar{A}_i B_{i+1}^{-1} A_i^{-1} = I + F_{\mu\nu}(\xi_i)\delta x_i^\mu \Delta x_i^\nu + O(\delta t)^2 + O(\Delta s)^2,$$

where $F_{\mu\nu}(\xi_i) = \partial_\mu A_\nu(\xi_i) - \partial_\nu A_\mu(\xi_i) + [A_\mu(\xi_i), A_\nu(\xi_i)]$. Thus it follows from the above equality that

$$\begin{aligned}\prod_{i=1}^{N}(I + A_\mu(\bar{\xi}_i)\Delta \bar{x}_i^\mu) &= \prod_{i=1}^{N} \bar{A}_i \\ &= B_1^{-1} \prod_{i=1}^{N}(I + F_{\mu\nu}(\xi_i)\delta x_i^\mu \Delta x_i^\nu) \\ &\quad \times (I + A_\mu(\xi_i)\Delta x_i^\mu) B_{N+1} \\ &\quad + O(\Delta s)^2 + O(\delta t)^2,\end{aligned}$$

where $B_{N+1} = I + A_\mu(x(1,t))\delta x^\mu(1,t)$. Furthermore performing the calculation successively, we obtain

$$\begin{aligned}\prod_{i=1}^{N}(I + A_\mu(\bar{\xi}_i)\Delta \bar{x}_i^\mu) &= \prod_{i=1}^{N}(I + A_\mu(\xi_i)\Delta x_i^\mu) \\ &\quad - A_\mu(\xi_1)\delta x_1^\mu \prod_{i=1}^{N}(I + A_\mu(\xi_i)\Delta x_i^\mu) \\ &\quad + \prod_{i=1}^{N}(I + A_\mu(\xi_i)\Delta x_i^\mu) A_\mu(x(1,t))\delta x^\mu(1,t) \\ &\quad + \sum_{j=1}^{N}\left\{\prod_{i=1}^{j-1}(I + A_\mu(\xi_i)\Delta x_i^\mu) F_{\lambda\rho}(\xi_j)\delta x_j^\lambda \Delta x_j^\rho \right. \\ &\quad \left. \times \prod_{i=j+1}^{N}(I + A_\mu(\xi_i)\Delta x_i^\mu)\right\} + O(\Delta s)^2 + O(\delta t)^2.\end{aligned}$$

Consequently we are led to the following conclusion:

$$\begin{aligned}U[\gamma(t+\delta t)] &= U[\gamma(t)] - A_\mu(x(0,t))\delta x^\mu(0,t) U[\gamma(t)] \\ &\quad + U[\gamma(t)] A_\mu(x(1,t))\delta x^\mu(1,t) \\ &\quad + \int_0^1 U[\gamma^s(t)] F_{\mu\nu}(x(s,t)) U[\gamma_s(t)] \\ &\quad \times \frac{\partial x^\nu(s,t)}{\partial s}\, \delta x^\mu(s,t)\, ds \\ &\quad + O(\delta t)^2, \qquad (1)\end{aligned}$$

 0022-2488/81/112713-03$01.00

where

$$U[\gamma^s(t)] = P\exp\left(\int_{\gamma^s(t)} A_\mu(x)dx^\mu\right)$$
$$= P\exp\left(\int_0^s A_\mu(x(s,t))\frac{\partial x^\mu(s,t)}{\partial s}ds\right),$$
$$U[\gamma_s(t)] = P\exp\left(\int_{\gamma_s(t)} A_\mu(x)dx^\mu\right)$$
$$= P\exp\left(\int_s^1 A_\mu(x(s,t))\frac{\partial x^\mu(s,t)}{\partial s}ds\right).$$

If γ is the loop, Eq. (1) becomes the complete variation.[1] Thus we have the following definition.

Definition 1: We call $\delta_x U[\gamma(t)]$ the directional derivative of the path-ordered phase factor $U[\gamma(t)]$ with respect to $\gamma(t)$ in the direction $X = (\partial x^\mu(s,t)/\partial t)$, which is given by the following formula

$$\delta_x U[\gamma(t)] = \int_0^1 \frac{\delta U[\gamma(t)]}{\delta x^\mu(s,t)}\delta x^\mu(s,t)ds, \tag{2}$$

where

$$\frac{\delta U[\gamma(t)]}{\delta x^\mu(s,t)}$$
$$= U[\gamma(t)]A_\mu(x(s,t))\delta(s-1) - \delta(s-0)A_\mu(x(s,t))U[\gamma(t)]$$
$$+ U[\gamma^s(t)]F_{\mu\nu}(x(s,t))U[\gamma_s(t)]\frac{\partial x^\nu(s,t)}{\partial s}. \tag{3}$$

And we call $\delta U[\gamma(t)]/\delta x^\mu(s,t)$ the partial derivative with respect to $x^\mu(s,t)$ at $\gamma(t)$.

Proposition 1: If $\gamma(t)$ is the loop such that $x(0,t) = x(1,t)$ for every $t \in [0,1]$, then we have

$$U[\gamma(0)] - U[\gamma(1)] = \int_0^1\int_0^1 U[\gamma^s(t)]\left\{\sum_{\mu<\nu}F_{\mu\nu}(x(s,t))\frac{\partial(x^\mu,x^\nu)}{\partial(s,t)}\right\}$$
$$\times U[\gamma_s(t)]ds\,dt. \tag{4}$$

Proof: It is clear that we have

$$U[\gamma(1)] - U[\gamma(0)] = \int_0^1 \delta_x U[\gamma(t)].$$

Since we see $\delta x^\mu(0,t) = \delta x^\mu(1,t) = 0$ from the condition $x(0,t) = x(1,t)$, $\delta_x U[\gamma(t)]$ becomes, from (2) and (3),

$$\delta_x U[\gamma(t)] = -\int_0^1 U[\gamma^s(t)]\left\{F_{\mu\nu}(x(s,t))\frac{\partial x^\mu(s,t)}{\partial s}\frac{\partial x^\nu(s,t)}{\partial t}\right\}$$
$$\times U[\gamma_s(t)]ds\,dt$$
$$= -\int_0^1 U[\gamma^s(t)]\left\{\sum_{\mu<\nu}F_{\mu\nu}(x(s,t))\frac{\partial(x^\mu,x^\nu)}{\partial(s,t)}\right\}$$
$$\times U[\gamma_s(t)]ds\,dt,$$

where we used $F_{\mu\nu} = -F_{\nu\mu}$. Thus we get immediately the formula (4).

2. NONABELIAN STOKES THEOREM

Theorem 1: Let S be a smooth surface with smooth boundary γ in M. Then we have the following formula:

$$\int_\gamma \mathscr{A}_\mu dx^\mu = \int_s \mathscr{F}_{\mu\nu}dx^\mu \wedge dx^\nu,$$

where

$$\int_\gamma \mathscr{A}_{\mu\nu}dx^\mu = U[\gamma] - I,$$
$$\int_s \mathscr{F}_{\mu\nu}dx^\mu \wedge dx^\nu = \frac{1}{2}\int_0^1\int_0^1 \mathscr{F}_{\mu\nu}(x(s,t))\frac{\partial(x^\mu,x^\nu)}{\partial(s,t)}ds\,dt,$$
$$\mathscr{F}_{\mu\nu}(x(s,t)) = U[\gamma^s(t)]F_{\mu\nu}(x(s,t))U[\gamma_s(t)],$$
$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + [A_\mu(x),A_\nu(x)],$$

and $\gamma(t):(x(s,t)) = (x^\mu(s,t)), s \in [0,1]$, is the smooth loop on the surface S such that $x(0,t) = x(1,t)$ for every $t \in [0,1]$, $\gamma(0) = \gamma$, $\gamma(1) = x(0,0)$ and when the parameter t changes from 0 to 1, this loop $\gamma(t)$ sweeps on S all over.

This theorem is obtained easily from Proposition 1 by noting $U[\gamma(1)] = I$.

As an example we take up the case of SU(1). In this case, the path-ordered phase factor $U[\gamma]$ and the field strength $F_{\mu\nu}(x)$ become

$$U[\gamma] = \exp\left(i\int_\gamma A_\mu dx^\mu\right), \quad F_{\mu\nu}(x) = i(\partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)),$$

where $A_\mu(x)$ are real functions. Thus it follows that

$$\mathscr{F}_{\mu\nu}(x(s,t))dx^\mu \wedge dx^\nu = \left(\exp i\int_0^s A_\mu(x(s,t))\frac{\partial x^\mu}{\partial s}ds\right)$$
$$\times (F_{\mu\nu}(x(s,t))dx^\mu \wedge dx^\nu)$$
$$\times\left(\exp i\int_s^1 A_\mu(x(s,t))\frac{\partial x^\mu}{\partial s}ds\right)$$
$$= (F_{\mu\nu}(x(s,t))dx^\mu \wedge dx^\nu)\exp i\int_{\gamma(t)} A_\mu dx^\mu,$$
$$\int_\gamma \mathscr{A}_\mu dx^\mu = \exp\left(i\int_\gamma A_\mu dx^\mu - 1\right).$$

Executing the Taylor expansion of exponential functions of the above equalities, we have

$$\mathscr{F}_{\mu\nu}(x(s,t))dx^\mu \wedge dx^\nu = (F_{\mu\nu}(x(s,t))dx^\mu \wedge dx^\nu)$$
$$\times\left\{\sum_{n=0}^\infty \frac{i^n}{n!}\left(\int_{\gamma(t)} A_\mu dx^\mu\right)^n\right\},$$
$$\int_\gamma \mathscr{A}_\mu dx^\mu = \sum_{n=1}^\infty \frac{i^n}{n!}\left(\int_\gamma A_\mu dx^\mu\right)^n.$$

Furthermore, replacing A_μ by λA_μ and using the non-abelian Stokes theorem, we are led to

$$\sum_{n=1}^\infty \frac{(\lambda i)^n}{n!}\left(\int_\gamma A_\mu dx^\mu\right)^n = \sum_{n=0}^\infty \frac{(\lambda i)^{n+1}}{2n!}\int_s\left(\int_{\gamma(t)} A_\mu dx^\mu\right)^n$$
$$\times(\partial_\mu A_\nu - \partial_\nu A_\mu)dx^\mu \wedge dx^\nu,$$

where λ is an arbitrary constant. By comparing the λ term of both sides of the above equality, we obtain the following formula:

$$\int_\gamma A_\mu dx^\mu = \int_s \sum_{\mu<\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu)dx^\mu \wedge dx^\nu.$$

This is the usual Stokes theorem.

Proposition 2: If the value of the path-ordered phase factor $U[\gamma] = P\exp(\int_\gamma A_\mu\, dx^\mu)$ is independent of the path γ of integration and depends only on the end points x_0 and x of the path γ, then there exists a matrix element $U(x)$ belonging to the group SU(n) or the Lorentz group such that

$A_\mu(x) = U^{-1}(x)\partial_\mu U(x)$.

Proof: We set

$$U(x) = P\exp\left(\int_\gamma A_\mu dx^\mu\right).$$

Now we join two points $x = (x^1, x^2, x^3, x^4)$ and $x + \Delta x = (x^1 + \Delta x^1, x^2, x^3, x^4)$ by the segment $\Delta\gamma$. Then it follows from the property[3] of path-ordered phase factors that

$$\begin{aligned} U(x+\Delta x) - U(x) &= P\exp\int_{\gamma+\Delta\gamma} A_\mu dx^\mu - P\exp\int_\gamma A_\mu dx^\mu \\ &= \left(P\exp\int_\gamma A_\mu dx^\mu\right)\left(P\exp\int_{\Delta\gamma} A_\mu dx^\mu\right) \\ &\quad - P\exp\int_\gamma A_\mu dx^\mu \\ &= U(x)\left(P\exp\int_{\Delta\gamma} A_1 dx^1 - I\right), \end{aligned}$$

therefore

$$\frac{1}{\Delta x^1}(U(x+\Delta x) - U(x)) = U(x)\frac{1}{\Delta x^1}\left(P\exp\int_{\Delta\gamma} A_1 dx^1 - I\right). \tag{5}$$

As $U(x)$ satisfies the differential equations[3] $\partial_\mu U(x) = U(x)A_\mu(x)$, we have the relation

$$\lim_{\Delta x^1 \to 0}\frac{1}{\Delta x^1}\left(P\exp\int_{\Delta\gamma} A_1 dx^1 - I\right) = A_1(x).$$

Thus it follows from the equality (5) that $\partial U(x)/\partial x^1 = U(x)A_1(x)$. In a similar manner we get the equalities $\partial U(x)/\partial x^\mu = U(x)A_\mu(x)$ ($\mu = 2, 3, 4$). It has been already verified that the path-ordered phase factor $U(x)$ belongs to the group SU(n) [the Lorentz group].[3]

Proposition3: Given the path-ordered phase factor $U[\gamma] = P\exp(\int_\gamma A_\mu dx^\mu)$, if there exists a matrix $U(x) \in \mathrm{SU}(n)$ [the Lorentz group] such that $U^{-1}(x)\partial_\mu U(x) = A_\mu(x)$, then the curvature $F_{\mu\nu}(x)$ vanishes identically.

Proof: This proposition can be seen immediately by inserting $A_\mu = U^{-1}\partial_\mu U$ into $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ and by using $(\partial_\mu U)U^{-1} = -U\partial_\mu U^{-1}$, $\partial_\mu\partial_\nu U = \partial_\nu\partial_\mu U$.

Theorem 2: The curvature $F_{\mu\nu}(x)$ vanishes identically if and only if there exists a matrix $U(x) \in \mathrm{SU}(n)$ [the Lorentz group] such that $A_\mu(x) = U^{-1}(x)\partial_\mu U(x)$.

Proof: First we show that the path-ordered phase factor is independent of the path if and only if $\int_\gamma \mathscr{A}_\mu dx^\mu = 0$, where γ is a smooth loop. Now let γ_1, γ_2 be the smooth paths joining two points x_0 and x, and γ be the loop which goes to x from x_0 along γ_1 and comes back to x_0 along γ_2. Set $U[\gamma] = P\exp\int_\gamma A_\mu dx^\mu$. By the properties[3] of path-ordered phase factors, we have the relation $\int_\gamma \mathscr{A}_\mu dx^\mu = U[\gamma] - I = U[\gamma_1 + (-\gamma_2)] - I = U[\gamma_1]U[-\gamma_2] - I = U[\gamma_1]U^{-1}[\gamma_2] - I$, namely,

$$\int_\gamma \mathscr{A}_\mu dx^\mu = U[\gamma_1]U^{-1}[\gamma_2] - I. \tag{6}$$

This relation (6) show that $\int_\gamma \mathscr{A}_\mu dx^\mu = 0$ if and only if the path-ordered phase factor $U[\gamma_1]$ is independent of the path γ_1 and depends only on the end points x_0, x of γ_1.

In view of Porpositions 2 and 3, it is thus sufficient for our purpose to show that if the curvature $F_{\mu\nu}(x)$ vanishes identically, then $\int_\gamma \mathscr{A}_\mu dx^\mu = 0$. This is obtained immediately from the non-abelian Stokes theorem.

Set $F(x) = \frac{1}{2}F_{\mu\nu}(x)dx^\mu \wedge dx^\nu$ and $A(x) = A_\mu(x)dx^\mu$. Then Theorem 2 is rewritten by means of the differential forms $F(x)$ and $A(x)$ as follows. The curvature form $F(x) = dA(x) + A(x)\wedge A(x)$ vanishes identically if and only if there exists a matrix $U(x) \in \mathrm{SU}(n)$ [the Lorentz group] such that the connection form $A(x)$ is given by the form $U^{-1}(x)dU(x)$.

Definition 2: We call $A(x)$ the pure gauge connection form which is given by the form $A(x) = U^{-1}(x)dU(x)$ and call $A_\mu(x) = U^{-1}(x)\partial_\mu U(x)$ the pure gauge field.[4]

As an example, we consider the case of SU(1). In this case, $A_\mu(x)$ and $U(x)$ are functions. Therefore $F(x)$ and $A(x)$ are the usual differential forms and we see $A(x)\wedge A(x) = 0$. Thus Theorem 2 becomes as follows. The differential 2-form $dA(x) = 0$ if and only if there exists a function $\phi(x)$ such that $A(x) = d\phi(x)$ where $\phi(x) = \ln U(x)$. This is the usual Poincaré theorem.

Finally we show that the pure gauge connection form is also obtained by setting up the following requirement.

(R) There exists a geodesic moving frame[3] $b_g(x)$ for the neighborhood U of each point of M such that $Db_g(x) = 0$ for every point $x \in U$, where D is the exterior covariant differentiation.

In fact, let $b(x)$ be an arbitrary moving frame for U, then we have the transformation $b(x) = b_g(x)U(x)$, where $U(x) \in \mathrm{SU}(n)$ [the Lorentz group]. Applying D to both sides of this transformation and using $Db_g(x) = 0$, $Db(x) = b(x)A(x)$, we have $b(x)A(x) = b_g(x)dU(x)$. Hence we get $A(x) = U^{-1}(x)dU(x)$.

ACKNOWLEDGMENT

The author would like to thank Professor T. Muta for valuable advice.

[1] E. Corrigan and B. Hasslacher, Phys. Lett. B **81**, 181 (1979).

[2] H. Cartan, *Differential Forms* (Hermann, Paris, 1970), p. 81, p.33.

[3] M. Iyanaga, "Path-Ordered Phase Factors as Generator of Gauge Fields" (to be published).

[4] W. Drechsler and M. E. Mayer, "Fiber Bundle Techniques in Gauge Theories," *Lecture Notes in Physics*, Vol. 67 (Springer, New York, 1977), p. 114.

Three-dimensional relativistic scattering of electromagnetic waves by an object in uniform translational motion

B. L. Michielsen, G. C. Herman, and A. T. de Hoop [a)]
Delft University of Technology, Department of Electrical Engineering, Laboratory of Electromagnetic Research, P.O. Box 5031, 2600 GA Delft, The Netherlands

D. de Zutter
State University of Ghent, Electromagnetism and Acoustics Laboratory, St. Pietersnieuwstraat 41, Ghent, Belgium

(Received 25 March 1981; accepted for publication 12 June 1981)

A general, relativistic formalism is developed for the three-dimensional scattering of electromagnetic waves by an object that is in uniform translational motion with respect to a source of electromagnetic radiation. The theory applies to objects of arbitrary size, shape, and physical composition. In particular, the temporal frequency spectrum of the field detected by a receiver that is stationary with respect to the source is determined. Numerical results pertaining to the scattering of a time-harmonic plane wave by a small, uniformly moving particle are presented.

PACS numbers: 11.80. — m, 41.10.Hv

1. INTRODUCTION

The relativistic theory of scattering of electromagnetic waves by obstacles in uniform translational motion has been the subject of investigation by several authors. For the scattering of two-dimensional waves by cylindrical obstacles,[1–3] as well as for three-dimensional scattering by obstacles of bounded extent,[4,5] expressions have been obtained for the scattered field as it is observed by a receiver that is stationary with respect to the source that illuminates the moving object. However, all these expressions apply to the case where the distance between the moving obstacle and the receiver is large at all times, in which case the far-field approximation for the scattered field can be used.

In the present paper, we develop the three-dimensional relativistic scattering of electromagnetic waves by a uniformly moving obstacle of bounded extent for any location of the receiver, stationary with respect to the source. As in Refs. 1–5, the Lorentz transformation is used to transform the field radiated by the source to the frame of reference in which the obstacle is at rest and the scattering problem is solved in this reference frame, after which the scattered field is transformed to the frame of reference in which the receiver is at rest. A Fourier transform of the latter result finally leads to the desired frequency content of the received signal (SI units are used throughout).

2. FORMULATION OF THE PROBLEM

We consider the scattering of electromagnetic waves by an obstacle that moves with uniform velocity $\mathbf{v}$ with respect to a source of electromagnetic radiation in free space. We adopt two inertial frames of reference K and K' (Fig. 1), where K is denoted as the laboratory frame and K' as the obstacle frame. The source is at rest with respect to K, while the obstacle is at rest with respect to K'. We assume that K and K' coincide at the instant $t = t' = 0$ and therefore, the space-time coordinates $\{\mathbf{r},t\}$ of K and $\{\mathbf{r}',t'\}$ of K' are mutually related by the Lorentz transformation[6]

$$\mathbf{r}' = (\mathbf{i}_{\|} \times \mathbf{r}) \times \mathbf{i}_{\|} + \gamma\{ -\mathbf{v}t + (\mathbf{i}_{\|} \cdot \mathbf{r})\mathbf{i}_{\|} \} \tag{1}$$

and

$$t' = \gamma\{t - \mathbf{v}\cdot\mathbf{r}/c_0^2\}, \tag{2}$$

where c_0 is the speed of light *in vacuo* and $\mathbf{i}_{\|}$ and γ are given by

$$\mathbf{i}_{\|} = \mathbf{v}/(\mathbf{v}\cdot\mathbf{v})^{1/2} \tag{3}$$

and

$$\gamma = [1 - (\mathbf{v}\cdot\mathbf{v})/c_0^2]^{-1/2}, \tag{4}$$

respectively.

For notational simplicity, the electric field strength $\mathbf{E}$ and the magnetic field strength $\mathbf{H}$ are grouped in the column matrix $[F]$ according to

$$[F] = \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix}. \tag{5}$$

The incident field $[F^i]$ in K is defined as the field that would be present in the absence of the obstacle. The scattered field $[F^s]$ in K is introduced as the difference between the total

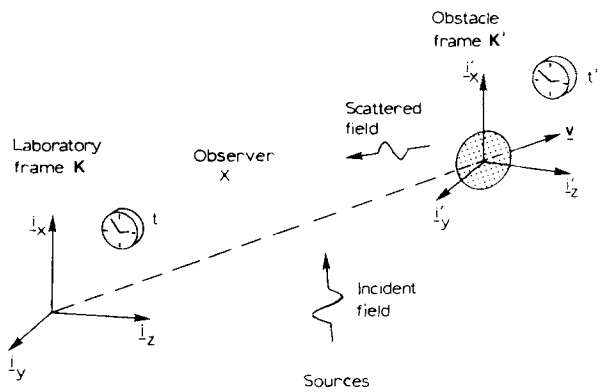


FIG. 1. Geometry of the scattering problem.

[a)]Address mail to A. T. de Hoop (see above).

 0022-2488/81/112716-07$01.00

field $[F]$ in K and $[F^i]$, i.e.,

$$[F^s] = [F] - [F^i]. \tag{6}$$

Similarly the incident, the total, and the scattered fields as observed in K' will be denoted by $[F^{i'}]$, $[F']$, and $[F^{s'}]$, respectively.

It is our purpose to arrive at an expression from which the scattered field $[F^s]$ in K can be calculated if the incident field $[F^i]$ (in K), the geometry and the electromagnetic properties of the obstacle (in K'), and its relative speed $\mathbf{v}$ with respect to K are known.

3. DERIVATION OF AN EXPRESSION FOR THE SCATTERED FIELD IN THE LABORATORY FRAME

In this section we derive a general expression for the scattered field in the laboratory frame that is suitable for our further considerations. To this aim we carry out a specific scheme of Lorentz transformations that is also at the root of the analysis in Refs. 1–5. The relevant steps (see also Fig. 2) are listed below.

A. The incident field $[F^i]$ is specified in the laboratory frame.

B. From $[F^i]$, the incident field $[F^{i'}]$ in the obstacle frame K' follows by the Lorentz transformation (see, e.g., Ref. 6),

$$[F^{i'}] = [\mathscr{L}(\mathbf{v})][F^i], \tag{7}$$

in which the square matrix $[\mathscr{L}(\mathbf{v})]$ can be written as

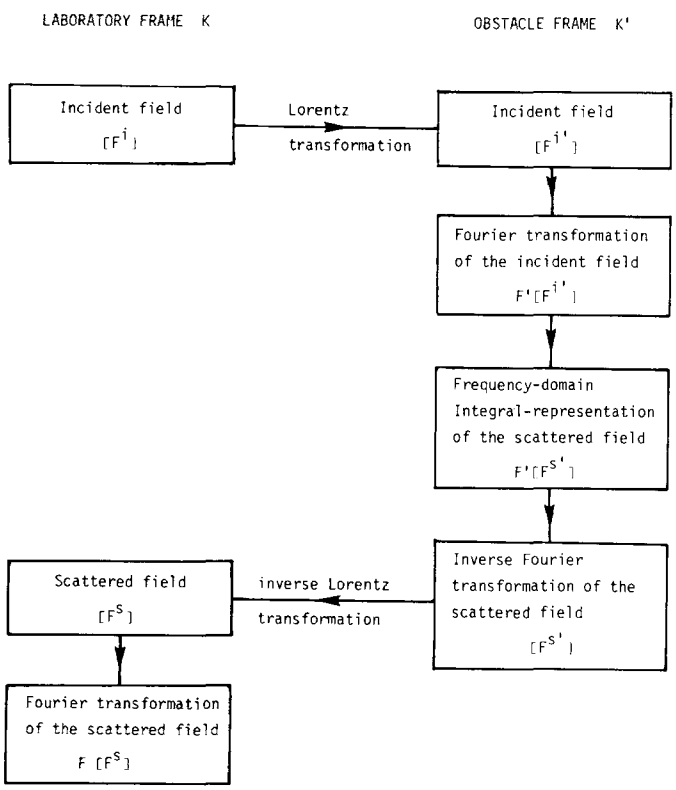


FIG. 2. Flow diagram illustrating the derivation of an expression for the scattered field in the laboratory frame.

$$[\mathscr{L}(\mathbf{v})] = \begin{bmatrix} [\mathscr{L}^{ee}(\mathbf{v})] & [\mathscr{L}^{em}(\mathbf{v})] \\ [\mathscr{L}^{me}(\mathbf{v})] & [\mathscr{L}^{mm}(\mathbf{v})] \end{bmatrix}, \tag{8}$$

with the submatrices defined through the relations

$$\begin{aligned} &[\mathscr{L}^{ee}(\mathbf{v})][\mathbf{E}^i] = (\mathbf{i}_\parallel \cdot \mathbf{E}^i)\mathbf{i}_\parallel + \gamma(\mathbf{i}_\parallel \times \mathbf{E}^i)\times \mathbf{i}_\parallel, \\ &[\mathscr{L}^{em}(\mathbf{v})][\mathbf{H}^i] = \mu_0 \gamma \mathbf{v} \times \mathbf{H}^i, \\ &[\mathscr{L}^{me}(\mathbf{v})][\mathbf{E}^i] = -\epsilon_0 \gamma \mathbf{v} \times \mathbf{E}^i, \\ &[\mathscr{L}^{mm}(\mathbf{v})][\mathbf{H}^i] = (\mathbf{i}_\parallel \cdot \mathbf{H}^i)\mathbf{i}_\parallel + \gamma(\mathbf{i}_\parallel \times \mathbf{H}^i)\times \mathbf{i}_\parallel, \end{aligned} \tag{9}$$

where $\mathbf{i}_\parallel$ is given by Eq. (3), γ is given by Eq. (4), and ϵ_0 and μ_0 are the permittivity and permeability of vacuum, respectively.

C. Since the scattering problem in K' will be formulated in the frequency domain, the next step is the application of the temporal Fourier transformation $\mathscr{F}'$ to $[F^{i'}]$; it is defined by the relation

$$\mathscr{F}'[F^{i'}] = \int_{-\infty}^{\infty} dt' [F^{i'}] \exp(i\omega' t'). \tag{10}$$

D. The scattered field in K' can be thought of as to be generated by contrast currents (polarization and magnetization currents) that are of the volume type for penetrable objects and of the surface type for impenetrable objects. The electromagnetic properties of the former are described in terms of constitutive relations (in the obstacle frame K') that express the contrast currents in terms of the field values in the obstacle, while for the latter the electromagnetic properties are described in terms of boundary conditions to be laid upon the limiting values of the field at the boundary surface of the obstacle (see, e.g., Ref. 7).

Let the domain occupied by the scattering object in K' be denoted by V' and let $\partial V'$ be the boundary surface of V'. For penetrable obstacles, we can now express the Fourier transform of the vector potential $[A']$, i.e., $\mathscr{F}'[A']$, in terms of contrast currents of the volume type $[J'_V]$ by

$$\mathscr{F}'[A'] = \int_{V'} dV(\mathbf{\rho}') G' \mathscr{F}'[J'_V], \tag{11}$$

in which $[A']$ and $[J'_V]$ can be written as

$$[A'] = \begin{bmatrix} \mathbf{A}^{e'} \\ \mathbf{A}^{m'} \end{bmatrix} \tag{12}$$

and

$$[J'_V] = \begin{bmatrix} \mathbf{J}^{e'}_V \\ \mathbf{J}^{m'}_V \end{bmatrix}, \tag{13}$$

where $\mathbf{A}^{e'}$ and $\mathbf{A}^{m'}$ are the volume vector potentials due to the polarization current of the volume type $\mathbf{J}^{e'}_V$ and the magnetization current of the volume type $\mathbf{J}^{m'}_V$, respectively. These currents are defined as

$$\begin{aligned} &\mathbf{J}^{e'}_V = \mathbf{J}' + \partial_{t'} \mathbf{P}' \\ &\text{and} \\ &\mathbf{J}^{m'}_V = \partial_{t'} \mu_0 \mathbf{M}', \end{aligned} \tag{14}$$

where $\mathbf{J}'$ is the current density, $\mathbf{P}'$ is the electric polarization, and $\mathbf{M}'$ is the magnetization inside V'. In (11), G' is the three-dimensional, free space Green's function, given by Ref. 7

$$G' = \exp[i\omega'|\mathbf{r}' - \mathbf{\rho}'|/c_0]/4\pi|\mathbf{r}' - \mathbf{\rho}'|. \tag{15}$$

Similarly, for impenetrable scatters, we have

$$\mathcal{F}'[A'] = \int_{\partial V'} dA(\rho')G'\mathcal{F}'[J'_S], \tag{16}$$

where the surface currents are defined as

$$\mathbf{J}_S^{e'} = -\mathbf{n}'\times\mathbf{H}' \tag{17}$$

and

$$\mathbf{J}_S^{m'} = \mathbf{n}'\times\mathbf{E}'.$$

The Fourier transform of the scattered field in K', $\mathcal{F}'[F^{s'}]$, can be expressed in terms of the vector potentials according to[7]

$$\mathcal{F}'[F^{s'}] = [D']\mathcal{F}'[A'], \tag{18}$$

where $\mathcal{F}'[A']$ is given by Eq. (11) for a penetrable scatterer and by Eq. (16) for an impenetrable scatterer and where the differential operator $[D']$ can be written as

$$[D'] = \begin{bmatrix} [D^{ee'}][D^{em'}] \\ [D^{me'}][D^{mm'}] \end{bmatrix}, \tag{19}$$

with the submatrices defined through the relations

$$\begin{aligned}
[D^{ee'}]\mathcal{F}'\mathbf{A}^{e'} &= i\omega'\mu_0\mathcal{F}'\mathbf{A}^{e'} \\
&\quad -(i\omega'\epsilon_0)^{-1}\nabla'\{\nabla'\cdot\mathcal{F}'\mathbf{A}^{e'}\}, \\
[D^{em'}]\mathcal{F}'\mathbf{A}^{m'} &= -\nabla'\times\mathcal{F}'\mathbf{A}^{m'}, \\
[D^{me'}]\mathcal{F}'\mathbf{A}^{e'} &= \nabla'\times\mathcal{F}'\mathbf{A}^{e'}, \\
[D^{mm'}]\mathcal{F}'\mathbf{A}^{m'} &= i\omega'\epsilon_0\mathcal{F}'\mathbf{A}^{m'} \\
&\quad -(i\omega'\mu_0)^{-1}\nabla'\{\nabla'\cdot\mathcal{F}'\mathbf{A}^{m'}\},
\end{aligned} \tag{20}$$

with $\nabla' = \mathbf{i}_{x'}\partial_{x'} + \mathbf{i}_{y'}\partial_{y'} + \mathbf{i}_{z'}\partial_{z'}$.

E. In order to calculate the scattered field $[F^s]$ in the laboratory frame K, we apply the inverse Lorentz transformation to $[F^{s'}]$. Now, $[F^{s'}]$ is recovered from $\mathcal{F}'[F^{s'}]$ by applying the inverse Fourier transformation $\mathcal{F}^{-1'}$:

$$[F^{s'}] = \mathcal{F}^{-1'}\{\mathcal{F}'[F^{s'}]\}, \tag{21}$$

which is defined by the relation

$$\begin{aligned}
&\mathcal{F}^{-1'}\{\mathcal{F}'[F^{s'}]\} \\
&\quad = (2\pi)^{-1}\int_{-\infty}^{\infty} d\omega'\mathcal{F}'[F^{s'}]\exp(-i\omega' t').
\end{aligned} \tag{22}$$

F. Now, $[F^s]$ follows from $[F^{s'}]$ by applying the inverse Lorentz transformation

$$[F^s] = [\mathcal{L}^{-1}(\mathbf{v})][F^{s'}], \tag{23}$$

where $[\mathcal{L}^{-1}(\mathbf{v})]$ follows from (8) and (9) through

$$[\mathcal{L}^{-1}(\mathbf{v})] = [\mathcal{L}(-\mathbf{v})]. \tag{24}$$

G. The temporal frequency spectrum of the scattered field in K is finally found from the Fourier transformation

$$\mathcal{F}[F^s] = \int_{-\infty}^{\infty} dt\,[F^s]\exp(i\omega t). \tag{25}$$

From Eqs. (23)–(25), it follows that the temporal frequency spectrum of the scattered field in K can be written as

$$\mathcal{F}[F^s] = \mathcal{F}[\mathcal{L}^{-1}][F^{s'}]. \tag{26}$$

The transformations occurring in (26) are linear and we now use this property to exhibit the structure of $\mathcal{F}[F^s]$ more explicitly. In principle, the approach presented here runs along the same lines as followed in Ref. 5, be it that we have not restricted our considerations to the far-field approximation of the scattered field in K'. Of course, the latter approximation follows as a special case from our expressions.

Since the transformations $\mathcal{F}$ and $[\mathcal{L}^{-1}]$ are linear and $[\mathcal{L}^{-1}]$ does not depend on the time coordinate t, they can be interchanged. To perform the Fourier transformation $\mathcal{F}$ of the scattered field $[F^{s'}]$, this field has to be specified as a function of the time coordinate t. This is accomplished by first expressing, with the aid of (21) and (18), $[F^{s'}]$ in terms of the space-time coordinates $\{\mathbf{r}',t'\}$ in K', introducing the vector potentials $[A']$. Equation (26) can now be written as

$$\mathcal{F}[F^s] = [\mathcal{L}^{-1}]\mathcal{F}\mathcal{F}^{-1'}[D']\mathcal{F}'[A']. \tag{27}$$

Substitution of (18) and (22) in (27) then results in

$$\begin{aligned}
\mathcal{F}[F^s] &= (2\pi)^{-1}\int_{-\infty}^{\infty} d\omega'[\mathcal{L}^{-1}]\mathcal{F}[D'] \\
&\quad \times\exp(-i\omega' t')\mathcal{F}'[A'],
\end{aligned} \tag{28}$$

where the integration with respect to ω' has been interchanged with $\mathcal{F}$ and $[\mathcal{L}^{-1}]$. We have to keep in mind that, in order to perform the Fourier transformation $\mathcal{F}$, all space-time coordinates $\{\mathbf{r}',t'\}$ occurring in the terms to the right of the operator $\mathcal{F}$ in (28) have to be expressed in terms of the space-time cordinates $\{\mathbf{r},t\}$ by the Lorentz transformation (1) and (2).

It can be shown that $\mathcal{F}$ and $[D']$ can be interchanged, too, in the following way:

$$\begin{aligned}
&\mathcal{F}[D']\exp(-i\omega' t')\mathcal{F}'[A'] \\
&\quad = [\mathcal{D}]\mathcal{F}\exp(-i\omega' t')\mathcal{F}'[A'],
\end{aligned} \tag{29}$$

where $[\mathcal{D}]$ follows from $[D']$ given in (19) and (20) by replacing ∇' by $(\omega'/c_0)\mathbf{d}$, where $\mathbf{d}$ is given by

$$\mathbf{d} = (c_0/\omega')\nabla_\perp + is\mathbf{i}_\parallel, \tag{30}$$

with

$$\nabla_\perp = -\mathbf{i}_\parallel\times(\mathbf{i}_\parallel\times\nabla) = \nabla - \mathbf{i}_\parallel(\mathbf{i}_\parallel\cdot\nabla) \tag{31}$$

and

$$s = (\omega - \gamma\omega')/\beta\gamma\omega', \tag{32}$$

where γ is given by (4) and β is given by

$$\beta = |\mathbf{v}|/c_0. \tag{33}$$

Since in Eq. (29) $G'\exp(-i\omega' t')$ is the only term to the right of $\mathcal{F}$ that needs expression in terms of $\{\mathbf{r},t\}$ [cf. (11)], we only have to determine

$$\mathcal{G} = \mathcal{F}\{G'\exp(-i\omega' t')\}, \tag{34}$$

where G' is given by (15). The relevant function $\mathcal{G}$ is obtained as[10]

$$\begin{aligned}
\mathcal{G} &= (-i/4\gamma|\mathbf{v}|)\exp(-i\omega' s\rho'_\parallel/c_0)\exp[i\omega'\gamma(\beta+s)r_\parallel/c_0] \\
&\quad\times\begin{bmatrix} H_0^{(1)}[\phi(1-s^2)^{1/2}] & (|s|<1) \\ (-2i/\pi)K_0[\phi(s^2-1)^{1/2}] & (|s|>1) \end{bmatrix}
\end{aligned} \tag{35}$$

[for this purpose, 10.1.1 from Ref. 9 has to be used in (34), after which the result can be found on p. 56 of Ref. 10.] where $H_0^{(1)}$ is the Hankel function of the first kind and order zero, K_0 is a modified Bessel function of the second kind and order zero, while

$$\rho'_{\|} = \mathbf{i}_{\|}\cdot\boldsymbol{\rho}',$$
$$r_{\|} = \mathbf{i}_{\|}\cdot\mathbf{r},$$
$$\boldsymbol{\rho}'_{\perp} = \boldsymbol{\rho}' - \rho'_{\|}\mathbf{i}_{\|}, \tag{36}$$
$$\mathbf{r}_{\perp} = \mathbf{r} - r_{\|}\mathbf{i}_{\|},$$
$$\phi = \omega'|\mathbf{r}_{\perp} - \boldsymbol{\rho}'_{\perp}|/c_0.$$

Finally, combining (26)–(36), the temporal frequency spectrum of the scattered field in K, $\mathcal{F}[F^s]$, can be written as

$$\mathcal{F}[F^s] = (2\pi)^{-1}\int_{-\infty}^{\infty} d\omega'[\mathcal{L}^{-1}][\mathcal{D}] \times \int_{V'} dV(\boldsymbol{\rho}')\mathcal{G}\mathcal{F}'[J'_V] \tag{37}$$

for penetrable scatterers and as

$$\mathcal{F}[F^s] = (2\pi)^{-1}\int_{-\infty}^{\infty} d\omega'[\mathcal{L}^{-1}][\mathcal{D}] \times \int_{\partial V'} dA(\boldsymbol{\rho}')\mathcal{G}\mathcal{F}'[J'_s] \tag{38}$$

for impenetrable scatterers.

Through (37) and (38) the temporal frequency spectrum of the scattered field in the laboratory frame is expressed in terms of volume or surface sources in the obstacle frame that are located either in the interior region or on the boundary of the obstacle, respectively. The strength of these sources can be determined by solving the scattering problem in the obstacle frame, i.e., for an obstacle at rest that is illuminated by $[F^i]$. Depending upon the geometry and the electromagnetic properties of the obstacle, computational as well as analytical methods can be used for this purpose.

The Green's function occurring in (37) and (38), as given by (35), has a singularity at $|s| = 1$. This kind of singularity has already been noted and discussed by de Zutter.[5] When the distance between observer and moving obstacle is large at all times ($|\mathbf{r}_{\perp}|\to\infty$), asymptotic expressions for $H_0^{(1)}$ and K_0 can be used in (35) and, in that case, Eqs. (37) and (38) can be reduced to an expression already presented in Ref. 5.

4. PLANE-WAVE SCATTERING BY A SMALL OBSTACLE IN UNIFORM, TRANSLATIONAL MOTION

We shall now apply the results from Sec. 3 to the scattering of a sinusoidal uniform plane wave with angular frequency Ω by a uniformly moving obstacle with spatial dimensions that are small compared to the wavelength of the incident field. The incident field in K is now given as

$$[F^i] = [f^i]\cos(\Omega t - \mathbf{k}\cdot\mathbf{r}), \tag{39}$$

where the wave vector $\mathbf{k}$ is given by

$$\mathbf{k} = \mathbf{u}\Omega/c_0 \tag{40}$$

($\mathbf{u}$ is the direction of propagation of the incident plane wave), and $[f^i]$ by

$$[f^i] = \begin{bmatrix}\mathbf{e}^i\\ \mathbf{h}^i\end{bmatrix}. \tag{41}$$

Here $\mathbf{e}^i$ and $\mathbf{h}^i$ are the plane-wave amplitudes of the electric and the magnetic field strengths, respectively. The incident field in K' is also a plane wave, written as

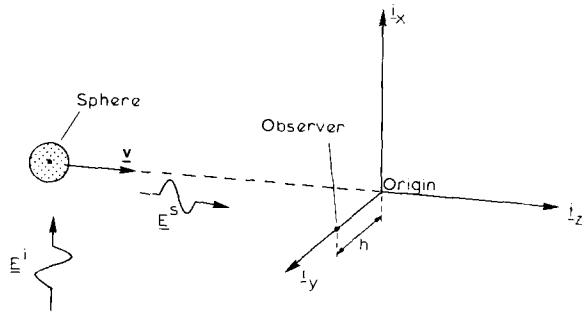


FIG. 3. Geometry of the scattering problem for which numerical results are presented in Figs. 4–6. The incident electric field is directed along the y axis.

$$[F^i] = [f^i]\cos(\Omega' t' - \mathbf{k}'\cdot\mathbf{r}'). \tag{42}$$

Expressions for $[f^i]$, $\mathbf{k}'$, and Ω' are given in the Appendix. Since further, the obstacle is assumed to be very small, we can in the limit $|\boldsymbol{\rho}'|\to 0$, use the dipole approximation (see, e.g., Ref. 8) in (37) and (38):

$$\left.\begin{aligned}&\int_{V'} dV(\boldsymbol{\rho}')\mathcal{G}\mathcal{F}'[J'_V]\\ &\int_{\partial V'} dV(\boldsymbol{\rho}')\mathcal{G}\mathcal{F}'[J'_s]\end{aligned}\right\} = -i\Omega'\begin{bmatrix}\mathbf{p}'\\ \mu_0\mathbf{m}'\end{bmatrix}\pi\{\delta(\omega'-\Omega') + \delta(\omega'+\Omega')\}\mathcal{G}_{\boldsymbol{\rho}'=\mathbf{0}}, \tag{43}$$

where $\mathcal{G}_{\boldsymbol{\rho}'=\mathbf{0}}$ follows from Eq. (35) by substituting $\boldsymbol{\rho}' = \mathbf{0}$. In (43), the electric dipole moment $\mathbf{p}'$ and the magnetic dipole

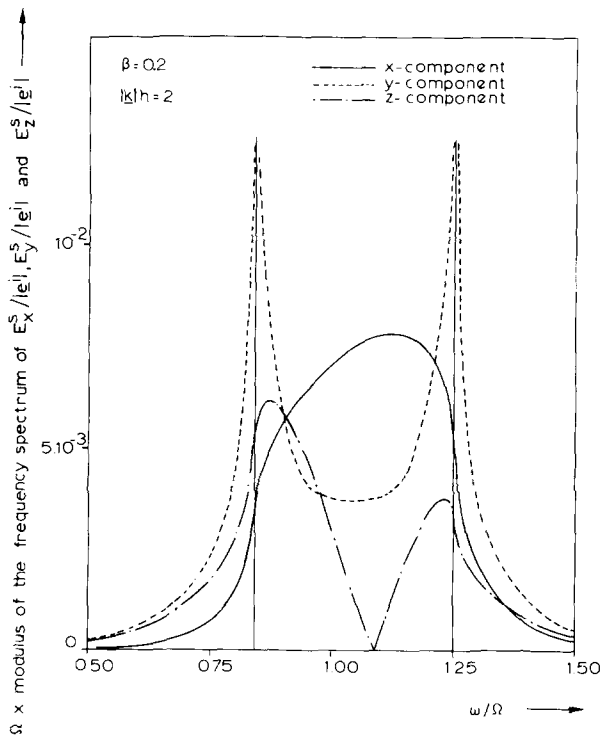


FIG. 4. The normalized x, y, and z component of the spectrum of the scattered electric field as a function of the normalized frequency for $\beta = 0.2$ and $\mathbf{k}|h| = 2$. The other parameters are $\Pi^e = \Pi^m = 10^{-2}|\mathbf{k}|^{-3}I$, $\mathbf{v} = 0.2c\mathbf{i}_z$, $\mathbf{u} = \mathbf{i}_x$, $\mathbf{e}^i = |\mathbf{e}^i|\mathbf{i}_y$.

moment $\mathbf{m}'$ are given by

$$\mathbf{p}' = \epsilon_0 \Pi^{e'} \cdot \mathbf{e}^{i'} \tag{44}$$

and

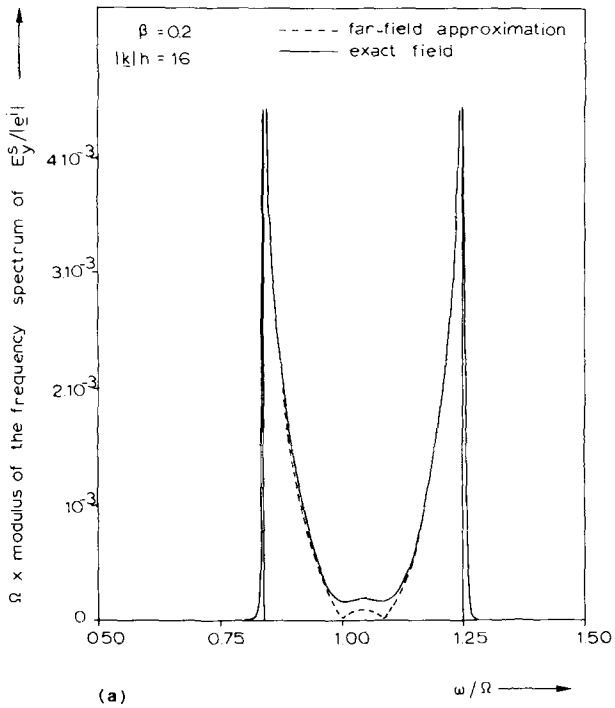
$\mathbf{m}' = \Pi^{m'} \cdot \mathbf{h}^{i'}$, respectively.

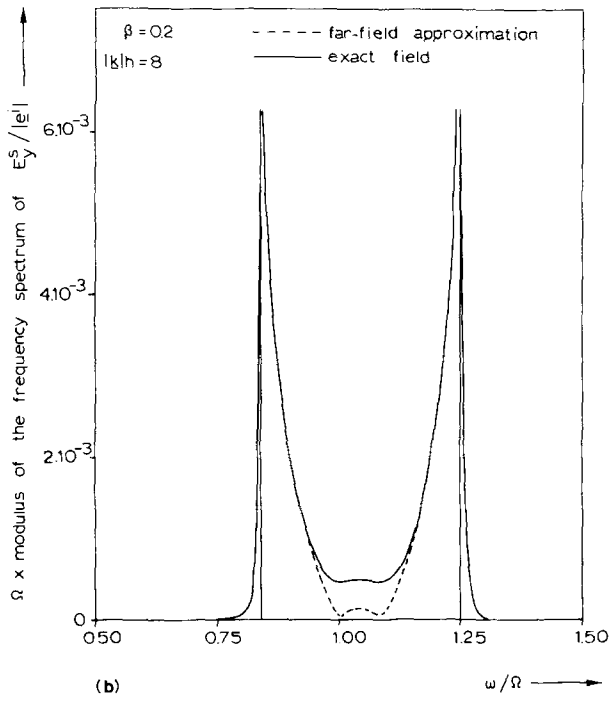
The electric polarizability tensor $\Pi^{e'}$ and the magnetic polarizability tensor $\Pi^{m'}$ reflect the geometry and the electromagnetic properties of the obstacle.

After substitution of Eq. (43) in (37) and (38), we now have an explicit expression from which the temporal frequency spectrum of $[F^s]$ can be calculated as soon as $[F^i]$ and the electric and magnetic polarizability tensors $\Pi^{e'}$ and $\Pi^{m'}$ are specified. The differentiations occurring in $[\mathscr{D}]$ can be performed explicitly with the aid of the well-known differentiation formulas for $H_0^{(1)}$ and K_0.[9]

5. NUMERICAL RESULTS

In this section we present some numerical results for the scattering of a time-harmonic, uniform plane wave with frequency Ω by a small, uniformly moving, homogeneous, and isotropic sphere. For a sphere Π^e and Π^m are given by Ref. 11 (p. 259)





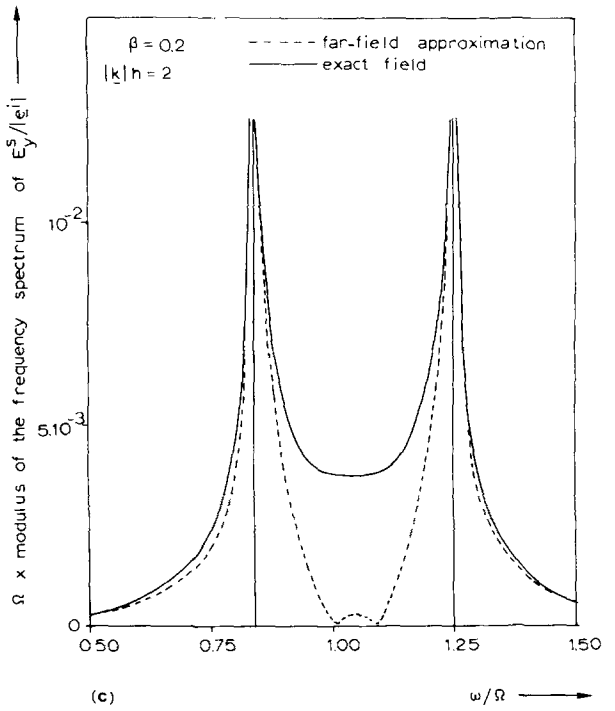


FIG. 5. Comparison between the y component of the scattered electric field calculated with the aid of the far-field approximation and calculated with the exact field expressions for $|\mathbf{k}|h = 16$ (a), $|\mathbf{k}|h = 8$ (b),and $|\mathbf{k}|h = 2$ (c). The other parameters are $\Pi^e = \Pi^m = 10^{-2}|\mathbf{k}|^{-3}I$, $\beta = 0.2$, $\mathbf{v} = 0.2c\mathbf{i}_z$, $\mathbf{u} = \mathbf{i}_x$, $\mathbf{e}^i = |\mathbf{e}^i|\mathbf{i}_y$.

$$\Pi^e = 4\pi a^3 \frac{\epsilon_r - 1}{\epsilon_r + 2} I \tag{45}$$

and

$$\Pi^m = 4\pi a^3 \frac{\mu_r - 1}{\mu_r + 2} I, \tag{46}$$

where I is the unit tensor of rank two, a is the radius of the sphere, and ϵ_r and μ_r are the relative permittivity and permeability of the sphere, respectively.

The numerical results presented in this section apply to the case $\Pi^e = \Pi^m = 10^{-2}|\mathbf{k}|^{-3} I$, where the wave vector of the incident wave, $\mathbf{k}$, is defined by Eq. (40). The sphere is moving uniformly in the direction of increasing z and therefore, we can write

$$\mathbf{v} = \beta c_0 \mathbf{i}_z. \tag{47}$$

The direction of propagation of the incident wave is assumed to be given by

$$\mathbf{u} = \mathbf{i}_x. \tag{48}$$

Furthermore, the incident electric field is taken to be directed along the y axis, i.e.,

$$\mathbf{e}^i = |\mathbf{e}^i|\mathbf{i}_y, \tag{49}$$

while the incident magnetic field is given by

$$\mathbf{h}^i = (\epsilon_0/\mu_0)^{1/2}\mathbf{u}\times\mathbf{e}^i. \tag{50}$$

The observer is located on the positive y axis of the laboratory frame at a distance h from the origin 0 (see Fig. 3).

In Fig. 4, the frequency spectrum of the scattered electric field is shown as a function of the normalized frequency ω/Ω for $|\mathbf{k}|h = 2$. It is observed that only the y component has a singularity at $|s| = \pm 1$ $[\omega/\Omega = (1 \pm \beta)\gamma^2]$.

Figure 5 shows for several values of $|\mathbf{k}|h$, the difference between the actual field values and the far-field approximation.[5] The dashed curve is obtained from the latter approximation, while the solid curve is obtained from (37), (43), and (44). It follows that the far-field approximation is quite good at $|\mathbf{k}|h = 8$ and $|\mathbf{k}|h = 16$, while significant deviations occur at $|\mathbf{k}|h = 2$. The largest differences occur around those frequencies where the major contribution to the spectrum originates from those time values during which the scatterer passes the observer. Since the distance between the scatterer and observer is small at that time, the far-field approximation is then inaccurate. It was already stated in Ref. 5 that, for larger values of $|\mathbf{k}|h$, the scattered field decays very rapidly for ω/Ω outside the region $((1-\beta)\gamma^2, (1+\beta)\gamma^2)$; Fig. 5(a) illustrates this. For small values of $|\mathbf{k}|h$, however, this is no longer true, as Fig. 5(c) ($|\mathbf{k}|h = 2$) and Fig. 6 ($|\mathbf{k}|h = 0.5$) show.

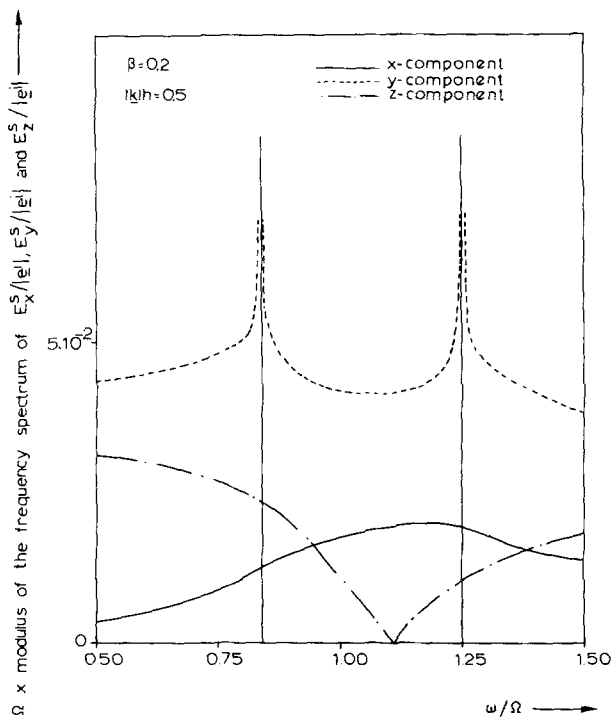


FIG. 6. The normalized x, y, and z components of the scattered electric field as a function of the normalized frequency for $\beta = 0.2$ and $|\mathbf{k}|h = 0.5$. The other parameters are $\Pi^e = \Pi^m = 10^{-2}|\mathbf{k}|^{-3} I$, $v = 0.2c\mathbf{i}_z$, $\mathbf{u} = \mathbf{i}_x$, $\mathbf{e}^i = |\mathbf{e}^i|\mathbf{i}_y$.

CONCLUSION

In this paper, an integral-equation formalism is developed for the three-dimensional, relativistic scattering of electromagnetic waves by an object that is in translational motion with respect to a source of electromagnetic radiation. The theory applies to objects or arbitrary size, shape, and physical composition and the field expressions that are obtained are valid for arbitrary positions of the observer with respect to the moving obstacle. In this respect, the formalism is more general than the ones published up to now. Also, an attempt has been made to make the presentation of the transformation schemes involved more transparent.

ACKNOWLEDGMENT

This work was carried out while G. C. Herman had the financial support of the Netherlands Organization for the Advancement of Pure Research (ZWO).

APPENDIX. LORENTZ TRANSFORMATION OF A UNIFORM PLANE WAVE WITH SINUSOIDAL TIME DEPENDENCE

A uniform plane wave with angular frequency Ω, wave vector $\mathbf{k}$, and plane-wave amplitude $[f^i]$ in the laboratory frame K, as given by (39), is also a uniform plane wave in the obstacle frame K', but now with angular frequency Ω', wave vector $\mathbf{k}'$, and plane-wave amplitude $[f'^i]$. The quantities Ω', $\mathbf{k}'$, Ω, and $\mathbf{k}$ are interrelated through[3]

$$\Omega' = \gamma(\Omega - \mathbf{v}\cdot\mathbf{k}) \tag{A1}$$

and

$$\mathbf{k}' = (\mathbf{i}_\parallel \times \mathbf{k})\times\mathbf{i}_\parallel + \gamma\{-\mathbf{v}\Omega/c_0 + (\mathbf{i}_\parallel\cdot\mathbf{k})\mathbf{i}_\parallel\}, \tag{A2}$$

where $\mathbf{i}_\parallel$ and γ are given by (3) and (4), respectively. The relations between $[f^i]$ and $[f'^i]$ can be obtained from (7)–(9) by replacing $[F^i]$ by $[f^i]$ and $[F'^i]$ by $[f'^i]$.

[1] A. W. Lee and R. Mittra, Can. J. Phys. **45**, 2999 (1967).
[2] T. Shiozawa, N. Kumagai, and T. Takenaka, Elect. Comun. Jpn. **51-B**, 75 (1968).

[3]D. de Zutter and J. van Bladel, IEE Journal on Microwaves, Optics and Acoustics **1**, 192 (1977).
[4]R. C. Restrick, Radio Sci. **3**, 1144 (1968).
[5]D. de Zutter, Appl. Sci. Res. **36**, 241–269 (1980).
[6]C. Møller, *The Theory of Relativity*, 2nd ed. (Oxford University, Oxford, 1972).
[7]A. T. de Hoop, Chap. 6 in *Modern Topics in Electromagnetics and Antennas* (PPL Conference publications 13, Stevenage, 1977).
[8]D. S. Jones, J. Inst. Math. Appl. **23**, 421 (1979).
[9]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[10]A. Erdelyi, *Tables of Integral Transforms*, Vol. 1 (McGraw–Hill, New York, 1954).
[11]J. van Bladel, *Electromagnetic Fields* (McGraw–Hill, New York, 1964).