JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 6, NUMBER 1 JANUARY 1965

Spherically Symmetric Gravitational Fields*

P. G. BERGMANN,† M. CAHEN‡ AND A. B. KOMAR
Belfer Graduate School of Science, Yeshiva University, New York, New York
(Received 7 July 1964)

It is shown that the Schwarzschild solution is the only spherically symmetric solution of the Einstein vacuum field equations, even when the differentiability of the metric is weakened to the extent of permitting solutions which are C^0, piecewise C^1. Petrov's purported counterexample is analyzed and shown to be essentially equivalent to Schwarzschild's example.

I. INTRODUCTION

In a recent paper, Petrov[1] has called attention to the relatively stringent assumptions under which Birkhoff's theorem[2] is usually derived. Because the transformation of a tensor involves derivatives of the coordinates, the assumption that the metric tensor of a particular manifold is to be of class C^k reduces admissible coordinate transformations at least to class C^{k+1}. Hence, the conventional proof of Birkhoff's theorem, requiring coordinate transformations merely to possess transformation coefficients [i.e., $x(x')$ to be C^1], may force on a given metric a coordinate system that artificially reduces its continuity properties to class C^0, whereas in a different coordinate system its differentiability may well be C^2 or even higher. To emphasize this general remark, Petrov presents what purports to be a class of spherically symmetric solutions of the vacuum field equations of Einstein that are not equivalent to Schwarzschild's solution.

This paper has led us to launch a somewhat more comprehensive investigation of solutions of the vacuum field equations that possess spherical symmetry but possibly a low-order differentiability. Our essential result is to confirm the uniqueness of Schwarzschild's solution, as long as spherical symmetry is preserved. Petrov's solution turns out to be equivalent to Schwarzschild's though Petrov's chosen coordinates have certain attractive features that we comment on. However, we examine the possible existence of non-Schwarzschild solutions more broadly than was done by Petrov, and indicate the degree of arbitrariness that results from dropping the requirement of strict spherical symmetry. This examination may be performed either on the basis of a four-dimensional approach, or in terms of propagation of Cauchy data off a spacelike three-dimensional hypersurface, e.g., in a Hamiltonian formalism elaborated by Dirac.[3] Section II of this paper analyzes Petrov's solution(s), whereas Secs. III and IV are concerned with the freedom of continuation of solutions in the more general contexts indicated above.

II. PETROV'S SOLUTION

Let us consider the following solution of the Einstein field equations for the vacuum as given by Petrov[1] (with minor changes in notation and several corrections of typographical errors).

* Supported in part by Aerospace Research Laboratories and Air Force Office of Scientific Research.
† Present address: Department of Physics, Syracuse University, Syracuse, New York.
‡ On leave of absence from Université Libre, Brussels; current address: Southwest Center for Advanced Studies, Dallas, Texas.
[1] A. Z. Petrov, Zh. Eksperim. i Teor. Fiz. 44, 1525 (1963); [English transl: Soviet Physics—JETP 17, 1026 (1963)].
[2] G. D. Birkhoff, *Relativity and Modern Physics* (Cambridge University Press, New York, 1923), p. 256.
[3] P. A. M. Dirac, Phys. Rev. 114, 924 (1959).

$$ds^2 = \psi^{-2}(\gamma^2\, dx^{02} - dx^{12} - d\theta^2 - \sin^2\theta\, d\phi^2), \quad (2.1)$$

where ψ and γ depend only on the variables x^1 and x^0 and are defined by the equations

$$\begin{aligned} \psi_{,0} &= \nu(x^0)\gamma \\ \psi_{,1}^2 &= C\psi^3 + \psi^2 + \nu^2 \\ \gamma_{,1}\psi_{,1} - \gamma\psi_{,11} &= \nu_{,0}, \end{aligned} \quad (2.2)$$

where ν is an arbitrary function of x^0 and C is a constant. If we define

$$\psi = (1/C)(4p - \tfrac{1}{3}), \quad (2.3)$$

the second of Eqs. (2.2) yields

$$p_{,1}^2 = 4p^3 + \tfrac{1}{12}p + (\tfrac{1}{16}C^2\nu^2 + \tfrac{1}{216}), \quad (2.4)$$

which is the defining equation of the elliptic Weierstrass function, $p(x^1 + \lambda(x^0))$ whose invariants, g_2 and g_3, are obtained from the coefficients of p^1 and p^0 respectively. [The function $\lambda(x^0)$ is arbitrary. It is of interest to note that the invariant g_3 contains $\nu(x^0)$ so that at different instants of "time" different elliptic functions are used.]

The above space–time admits a four-parameter group of motions whose Killing vectors are

$$\begin{aligned} {}_1\lambda^\alpha &= \delta_2^\alpha \cos\phi - \delta_3^\alpha \sin\phi \cot\theta, \\ {}_2\lambda^\alpha &= \delta_2^\alpha \sin\phi + \delta_3^\alpha \cos\phi \cot\theta, \\ {}_3\lambda^\alpha &= \delta_3^\alpha, \\ \xi^\alpha &= -(\psi_{,1}/\gamma)\delta_0^\alpha + \nu\delta_1^\alpha. \end{aligned} \quad (2.5)$$

The first three Killing vectors are evident from the manifest spherical symmetry of Eq. (2.1). We focus our attention on the fourth Killing vector, ξ^α. Employing Eqs. (2.1) and (2.2) we may easily compute its norm

$$\xi^2 = \xi^\alpha\xi_\alpha = 1 + C\psi. \quad (2.6)$$

From the first of Eqs. (2.2) we have

$$\xi^\alpha\psi_{,\alpha} = 0. \quad (2.7)$$

If we confine our attention to those domains where $\xi^2 \neq 0$, we can form the vector

$$T_\mu \equiv \xi^{-2}\xi_\mu. \quad (2.8)$$

It follows from the fact that ξ_μ is effectively a Killing vector in a two-dimensional manifold [its components lie entirely in the (x^1, x^0) surface and the coefficients depend only on x^1 and x^0] that T_μ is necessarily a gradient. Thus

$$T_\mu = T_{,\mu}. \quad (2.9)$$

In addition, from Eq. (2.2), we can deduce

$$(\psi^{-1})_{,\mu}(\psi^{-1})_{,\nu}g^{\mu\nu} = -(1 + C\psi). \quad (2.10)$$

It is evident from Eqs. (2.6) through (2.10) that, apart from the region where $\psi = 0$, the Jacobian J

$$J = \begin{vmatrix} T_{,1} & T_{,2} \\ (\psi^{-1})_{,1} & (\psi^{-1})_{,2} \end{vmatrix} \quad (2.11)$$

does not vanish.

Up to this point the assumed differentiability class of the arbitrary functions, λ and ν, did not enter into our considerations. We now wish to consider the following coordinate transformation

$$\begin{aligned} r = \bar{x}^1 = \psi^{-1}, \qquad \bar{\theta} = \theta, \\ t = \bar{x}^4 = T \qquad\quad \bar{\phi} = \phi. \end{aligned} \quad (2.12)$$

It is true that in general if ψ and T are of class C^1 the resulting metric may be of class C^0. But this does not distrub us, in view of the work of Papapetrou and Treder.[4] In point of fact, performing the indicated coordinate transformation (2.12) we see immediately from Eqs. (2.6) through (2.10) that we obtain the metric

$$ds^2 = \left(1 + \frac{C}{r}\right)dt^2 - \frac{dr^2}{1 + C/r} - r^2\, d\theta^2 - r^2\sin^2\theta\, d\phi^2,$$

which is not C^0, but rather analytic. (We recognize the constant, C, as the negative of the Schwarzschild radius.)

Alternatively, if we prefer not to perform the coordinate transformations of Eq. (2.12), we can use the additional geometric structure available to us by the existence of the Killing vector fields in order to obtain an invariant characterization of Petrov's space–time. Thus Eq. (2.6) provides us with an invariant specification of the function ψ, from which it follows that Eqs. (2.7) through (2.10) [taken together with the remaining Killing vector fields of Eq. (2.5)] yield an invariant determination of the essential geometric properties of the solution. That is, we see that apart from the capricious properties of the particular coordinate system we choose in order to exhibit the metric tensor, the solution of Petrov is *in every invariant sense* identical to that of Schwarzschild.

We should like to observe that Petrov's choice of coordinates has the virtue of permitting one to continue the solution through the Schwarzschild "singularity" where $\xi^2 = 0$ and the usual coordinates become singular. In this sense the Petrov solution is more nearly akin to that of Kruskal.[5] The poles of the elliptic function correspond to the true singularity at $r = 0$. The region where $\psi \to 0$, corresponds

[4] A. Papapetrou and H. Treder, Math. Nachr. **23**, 371 (1962).

[5] M. D. Kruskal, Phys. Rev. **119**, 1743 (1960).

to $r \to \infty$. The space–time is perfectly well behaved in that limit and in fact approaches flat Minkowski space. We see that it is merely Petrov's coordinate system (as well as Kruskal's) which becomes singular in that limit.

III. JUMP DISCONTINUITIES OF DERIVATIVES

Together with Petrov we consider a solution spherically symmetric if it possesses an isometry (sub)group that permits the introduction of spherical coordinates, so that under that group each orbit is the two-dimensional surface of a three-dimensional sphere. With this assumption, the line element may be written in the form

$$dt^2 = A(dx^0)^2 + 2B\,dx^0\,dx^1 - C(dx^1)^2 - N^2\,d\Omega^2, \qquad (3.1)$$

$$d\Omega^2 = d\theta^2 + \sin^2\theta\,d\varphi^2,$$

where A, B, C, and N are functions of x^0 and x^1 only. To eliminate the function B we should have to perform a coordinate transformation which in the (x^0, x^1) plane would render the curves $x^0 = \text{const}$ and $x^1 = \text{const}$ each others' orthogonal trajectories. If we denote the new coordinates by the symbols $\bar{t}$, $\bar{r}$, it is required that, with

$$AC + B^2 \neq 0, \qquad (3.2)$$

we have

$$C\bar{r}_{,0}\bar{t}_{,0} + B(\bar{r}_{,0}\bar{t}_{,1} + \bar{r}_{,1}\bar{t}_{,0}) + A\bar{r}_{,1}\bar{t}_{,1} = 0. \qquad (3.3)$$

In view of the fact that the partial derivatives of $\bar{r}$, $\bar{t}$ also enter the new expressions for A and C, these new quantities will be of the lower class of differentiability of the two quantities, the original metric tensor and the transformation coefficients. But as the transformation coefficients are restricted by the one Eq. (3.3), plus the integrability conditions, the transformation coefficients need not be of lower C class (at least locally, i.e., piecewise) than the original metric tensor. Let us confine our attention to a local domain; then we can assert that the construction of orthogonal trajectories need not introduce a lower piecewise C classification into the metric tensor components than the one they had to begin with.

As the next step we consider local coordinate transformations that leave the orthogonality condition unchanged. Such transformations will satisfy an equation of the form (3.3), except insofar that now the second term on the left vanishes. In particular, we may adopt as our spacelike coordinate r any function of $\bar{r}$, $\bar{t}$ whose gradient is spacelike. Such a function is N the so-called luminosity distance. Assuming now that N, as a component of the metric tensor, is C^k, its adoption as a coordinate may render some other components of the metric tensor C^{k-1}. Aside from this contingency, the adoption of N as a coordinate will be legitimate if its gradient exists (this condition will be satisfied if N is at least C^1) and is spacelike. Once we have come this far, Birkoff's theorem may be proved in two stages: (1) From one of the field equations we find that $\dot{C} = 0$, $\dot{A} = 0$. (2) One of the field equations contains only reference to A and is of the first differential order,

$$(d/dr)(r/C) = 1. \qquad (3.4)$$

Solutions are of the form $C = (1 - a/r)^{-1}$, where the constant of integration a is the only arbitrary quantity. Thus this form of solution *arises from purely local considerations and is quite independent of the adoption of any boundary conditions at infinity.* Positive, vanishing, and negative masses are all acceptable. Finally, A is obtained from a quadrature, solving again a first-order equation, which possesses as its only constant of integration a trivial scale factor of the time axis.

Hence, the only operation in which the differentiability of the metric tensor, and of the coordinate transformations, enters is the adoption of N as the coordinate r. For this operation to be meaningful it is necessary that N be piecewise C^1. If it is piecewise C^k, the metric field in the new coordinate system will be piecewise C^{k-1}, hence at worst C^0.

The field equations are of the second order, so one should ordinarily require that the solutions be C^2. However, it is well known that one can generalize the notion of a solution of differential equations to functions of lower differentiability when these are limits of sequences of solutions of class C^2. Papapetrou and Treder[4] have investigated jumps in the first partial derivatives of the metric tensor field on the assumption that a field that is C^0 and only piecewise C^1 is to be considered a solution of the (vacuum) field equations if it can be approximated by C^2 solutions. Jump discontinuities in any derivatives may come about through the choice or coordinate system, or they may arise in such a manner that they cannot be transformed away even by coordinate transformations that are themselves of low differentiability. Only the latter situation is of real interest. It was found that intrinsic jump discontinuities must lie on characteristic hypersurfaces, which are well defined because the metric itself is continuous across the hypersurface.

Accordingly, given a solution of the field equations that represents a Schwarzschild metric in some local

four-dimensional neighborhood which may or may not be in the vicinity of (spatial) infinity, we may consider cutting this solution off along some characteristic hypersurface (these are any null hypersurfaces) and continuing it beyond this null hypersurface in a manner consistent with the jump requirements. If the jump is not to destroy the assumed spherical symmetry, the jump hypersurface itself must be built up from orbits of the rotational isometry (sub)group, i.e., it must be a three-dimensional hypersurface generated by the multiple rotation of a null curve in the (x^0, x^1) plane. Such null curves will generally not be mapped into themselves by the timelike isometry of the Schwarzschild solution, but this is not a necessary requirement for our purposes. (The one null surface that is a complete orbit under all isometries is the Schwarzschild radius. In terms of Kruskal coordinates[5] this hypersurface, for finite coordinate times, is, however, not three- but only two-dimensional.)

The requirements on the jumps in the first derivatives of the metric,

$$[g_{\mu\nu,\rho}] = h_{\mu\nu}k_\rho, \tag{3.5}$$

(if the jump is to be intrinsic) are that the normal vector k_ρ be lightlike (this requirement is entirely equivalent to the previous statement that the jump hypersurface be characteristic, or null), and that the coefficients $h_{\mu\nu}$ satisfy the conditions

$$g^{\mu\nu}h_{\mu\nu} = 0, \qquad h_{\mu\rho}k^\rho = 0. \tag{3.6}$$

These relationships are the analogs of the conditions on gravitational waves that they be transverse–transverse and tracefree. The same conditions hold, incidentally, *mutatis mutandis*, for jump discontinuities of higher than the first derivatives. We now examine whether the jump conditions permit us to continue a Schwarzschild solution beyond a null surface in such a manner that it becomes inequivalent to the conventional solution beyond the null surface, but without violating the requirement of spherical symmetry.

To preserve spherical symmetry, we must, first of all, see to it that the discontinuity surface itself be composed of orbits corresponding to the spherical symmetry isometry, i.e., that it be composed of spherical surfaces. Beyond the jump the solution in question must maintain spherical symmetry as well. This requirement may be restated in terms of the line element. The line element will have to be of the form (3.1) on both sides of, as well as across the surface of discontinuity [which will be a null curve in the two-dimensional (r, t) manifold]. However, the functions A, B, C, and N need to be only of class C^0.

The normal vector of the null surface lies, of course, within the null surface itself. But in addition it must be perpendicular to all the directions tangential to the spherical isometry. Thus the vector must lie in the (r, t) plane. Within that plane it is the only direction tangential to the discontinuity surface. From the second of Eqs. (3.6), the tensor $h_{\mu\nu}$ is seen to have components only within the spherical symmetry orbit, i.e., it is in effect a two-dimensional tensor (with vanishing trace) defined on that surface. Hence $h_{\mu\nu}$ possesses only two nonvanishing eigenvalues, whose sum vanishes and whose eigenvectors both lie within the spherical surface. If $h_{\mu\nu}$ is nonzero, it is certainly nondegenerate within the spherical surfaces, and its eigenvectors distinguish a pair of mutually perpendicular directions on the sphere, *contrary to the assumption of spherical symmetry*. We conclude that intrinsic, spherically symmetric jumps cannot occur.

IV. DISCONTINUITIES IN TERMS OF CAUCHY DATA

It is instructive to review the problem of discontinuities in terms of Cauchy data. We shall adopt for this purpose the formalism of Dirac,[3] which permits us to characterize a solution in terms of the metric field, and of the second fundamental form, of a spacelike three-dimensional hypersurface. Though one Riemann–Einstein manifold may be represented in terms of more than one set of Cauchy data, a set of g_{mn}, p^{mn} uniquely determines at least a neighborhood of a Riemann–Einstein manifold; these Cauchy data themselves are not entirely arbitrary but must satisfy at each point of the hypersurface the four so-called Hamiltonian constraints,

$$p^{ms}{}_{/s} = 0 \tag{4.1}$$

and

$$p_{mn}p^{mn} - \tfrac{1}{2}p^2 + {}^{(3)}R = 0. \tag{4.2}$$

We shall ask what degree of arbitrariness we have in choosing spherically symmetric Cauchy data, given the assumption that these data agree at least outside some radius R with those typical for the Schwarzschild solution.

Our task is greatly simplified if we choose coordinates x^0, x^1 so that the coefficient B in the line element (3.1) vanishes, a choice which according to the discussion given in Sec. 3 does not place excessive restrictions on the differentiability class of the metric field. In accordance with the spherical symmetry of the field desired we give the canonical momentum

density field the form:

$$p^{11} = A^{\frac{1}{2}}N^2P \sin\theta, \quad p^{22} = A^{\frac{1}{2}}N^2Q \sin\theta, \quad p^{33} = \sin^{-2}\theta\, p^{22}, \tag{4.3}$$

where P, the longitudinal component, and Q, the transverse component of the canonical momentum density, are functions of x^1 only, i.e., independent of the angles; there is no dependence on x^0, as we define our Cauchy data on a x^0 = constant three-dimensional hypersurface. The evaluation of the three divergence constraints (4.1) then leads to two empty relations (those associated with the free indices 2 and 3, respectively) and one differential condition on the two functions P and Q,

$$\frac{dP}{dx^1} + 2\frac{d\ln N}{dx^1}P - \frac{1}{A}\frac{d(N^2)}{dx^1}Q = 0. \tag{4.4}$$

We now show that both P and Q can be made to vanish by an appropriate choice of hypersurface. We evaluate P on the assumption that $B = 0$, and find:

$$P = -\frac{2}{A^{\frac{1}{2}}}\frac{1}{C}\frac{\partial \ln N}{\partial x^0}. \tag{4.5}$$

It follows that P will vanish if we choose the coordinate x^1 to be some function of N, the luminosity distance, or if we choose our initial-value hypersurface so as to be perpendicular everywhere to the surfaces N = constant. As the hypersurface itself is only required to be spacelike (and even this requirement is perhaps not absolutely necessary), all that is required for our construction to succeed is that the metric field be piecewise C^0.

If $P = 0$, then Q must vanish as well, because of Eq. (4.4). The Hamiltonian constraints (4.1), (4.2) thereby reduce to the single condition ${}^{(3)}R = 0$. Evaluation of this condition leads to the result that with our assumption of spherical symmetry the curvature scalar can be made free of second derivatives of the field variables describing the metric field by adopting for the three-dimensional line element the form:

$$ds^2 = b^{-1}\,dN^2 + N^2\,d\Omega^2, \tag{4.6}$$

where b and N are to be functions of r only. N is the luminosity distance, as before, and both b and N are scalars with respect to transformations of the radial coordinate r. For the form (4.6) to be admissible, N must be at least piecewise C^1 with respect to r. If, in addition, the metric field itself is C^0, piecewise C^1 then b will be piecewise C^0. The curvature scalar, in terms of b and N, turns out to be the expression

$${}^{(3)}R = (2/N^2)[b - 1 + N(db/dN)]. \tag{4.7}$$

Requiring this scalar to vanish in some connected domain of the r coordinate yields in this domain

$$b = 1 - N_0/N, \tag{4.8}$$

N_0 being a constant of integration. This result evidently corresponds to the Schwarzschild solution. In the vicinity of $N = 0$ the variable b must approach unity if the hypersurface is not to be metrically singular at that point. Hence all solutions with $N_0 \neq 0$ have a true singularity at the center. With respect to the variables chosen, b and N, the point ($N = N_0$, $b = 0$) does not appear singular, but the metric determinant

$$({}^{(3)}g)^{\frac{1}{2}} = b^{-\frac{1}{2}}N^2(dN/dr)\sin\theta, \tag{4.9}$$

will vanish at that point unless the coordinate r is suitably chosen. This choice involves an asymptotic behavior at that point of the form

$$r = r_0 + \gamma^{-1}(N - N_0)^{\frac{1}{2}}, \quad \gamma \neq 0, \neq \infty. \tag{4.10}$$

For real values of γ, r will be real outside the luminosity distance N_0. In order to assure reality of f for smaller luminosity distances a different, imaginary γ would have to be chosen.

To obtain the parameter r customarily used in the conventional presentation of the Schwarzschild solution one would have to change the value of γ at N_0 in order to cover the entire range of luminosity distance $0 \leq N \leq \infty$. It is well known, and can be confirmed in terms of the formalism employed here, that inside N_0 the chosen hypersurface is no longer spacelike, but that the metric becomes indefinite. If we go through $N = N_0$ with a constant value of γ, we are led to a coordinate system that is equivalent to the real domain of Kruskal's coordinates.

Whatever topology we adopt, the condition (4.8) prevents us from piecing several inequivalent Schwarzschild solutions together in order to obtain a new solution that is inequivalent to any one Schwarzschild solution. On the contrary, any such spherically symmetric piecing is equivalent to the establishment of spherical surfaces with finite source (i.e., mass) surface density. Again we find that the differential equations lead uniquely to Schwarzschild's original solution with a single mass constant throughout the domain of validity of the field equations.

Note added in proof: A. Hamoui has very kindly called our attention to his paper, which appeared in June 1964, Compt. Rend. Acad. Sci. Paris **258**, 6085 (1964), where he obtained results identical with some of ours. Unfortunately, this paper had not been known to us at the time we submitted our work.

Cluster Properties of Multiparticle Systems*

W. HUNZIKER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 30 June 1964)

The object of investigation is a system of N particles in nonrelativistic quantum mechanics. The particles interact via two- or many-body potentials, for which a sufficient condition is that they be square integrable in the relative coordinates of the interacting particles. Cluster properties are derived for the time translation operator, for the wave operators, for the transition probabilities, and for the S operator.

INTRODUCTION

We consider a system of N particles in nonrelativistic quantum mechanics. The particles are assumed to interact via two- or many-body forces vanishing sufficiently fast for large separations of the particles. An intuitively evident property of such a system is the following one: If, at a given time, the particles form clusters far separated from each other, then the subsequent motion will be approximately the same as if no forces between the clusters were present. Our aim is to formulate this property mathematically, to prove it, and to study certain consequences for scattering theory.

CLUSTER PROPERTIES OF THE TIME TRANSLATION OPERATOR

Let $x_1 \cdots x_N$ be the Cartesian coordinates of particles $1 \cdots N$. Relative, or internal, coordinates of any subsystem of particles are always taken as linear combinations of $x_1 \cdots x_N$. For the time being, the potential describing any of the k-body forces acting in the system is assumed to be square integrable in the relative coordinates of the k interacting particles.[1] Once for all, we group the indices $1 \cdots N$ into n clusters $C_1 \cdots C_n$. The Hamiltonian of the system is

$$H = \sum_{k=1}^{n} H_k + V = H_C + V,$$

$$H_k = P_k^2/2M_k + h_k,$$

where V is the sum of all potentials linking particles in different clusters, H_k the Hamiltonian of the subsystem C_k, P_k and M_k the total momentum and the total mass of C_k, and h_k its internal energy. H, H_C, H_k, and h_k are self-adjoint operators on the corresponding Hilbert spaces.[1] For any n 3-vectors $a_1 \cdots a_n$ we define the spatial translation operator $T(a_1 \cdots a_n)$ on the Hilbert space $\mathfrak{C}$ of the N-particle system by

$$(T(a_1 \cdots a_n)f)(x_1 \cdots x_N) = f(x_1' \cdots x_N'),$$

where $x_i' = x_i + a_k$ for all $i \in C_k$, and we set $a = \min_{i,k} |a_i - a_k|$. Note that $T(a_1 \cdots a_n)$ commutes with H_C. The asymptotic independence of the motion of the n clusters can be stated as follows:

Theorem 1.

$$\lim_{a \to \infty} T(-a_1 \cdots -a_n) e^{iHt} T(a_1 \cdots a_n) = e^{iH_C t} = \prod_{k=1}^{n} e^{iH_k t},$$

in the strong sense on $\mathfrak{C}$, uniformly in $-\infty < t < +\infty$.

Before we turn to the proof, we note that the assumption of square integrability will be used only for the potentials in V. All we shall need about the operators h_k is their self-adjointness.

Proof. A dense linear set in $\mathfrak{C}$ is spanned by the functions

$$f(x_1 \cdots x_N) = \prod_{k=1}^{n} e^{-\frac{1}{2}(y_k - b_k)^2} f_k(z_k), \qquad (1)$$

where y_k is the position of the center of mass of C_k, f_k any square integrable function of the internal coordinates z_k of C_k and b_k an arbitrary 3-vector. Thus it is sufficient to prove Theorem 1 for states of the form (1), for which we have

$$(e^{iHt} - e^{iH_C t}) T(a_1 \cdots a_n) f = ie^{iHt} \int_0^t d\tau e^{-iH\tau} V T(a_1 \cdots a_n) e^{iH_C \tau} f. \qquad (2)$$

That the right-hand side of (2) is well defined be-

* Supported in part by the U. S. Air Force through the Air Force Office of Scientific Research.

[1] T. Kato, Trans. Am. Math. Soc. 70, 195 (1951). *Note added in proof:* For k-body forces with $k \geq 3$, square integrability of the potentials is not sufficient for the self-adjointness of H. Kato assumes these potentials to be bounded. I am grateful to Professor E. Nelson for pointing out this error to me.

comes clear in the following. By (2), it is enough to show that

$$\lim_{a\to\infty}\int_{-\infty}^{+\infty} d\tau\ ||VT(a_1\cdots a_n)e^{iH_C\tau}f|| = 0. \tag{3}$$

The y dependence of $(\exp iH_C\tau)f$ can be evaluated explicitly, yielding

$$|(e^{iH_C\tau}f)(x_1\cdots x_N,\tau)|^2 = \prod_{k=1}^{n}\mu_k^{\frac{3}{2}}e^{-\mu_k(y_k-b_k)^2}\,|f_k(z_k,\tau)|^2, \tag{4}$$

$$\mu_k = (1+\tau^2/M_k^2)^{-1},\quad f_k(\cdot,\tau) = e^{ih_k\tau}f_k.$$

The triangle inequality allows to prove (3) separately for each term in the sum V. In order to avoid purely formal complications, we treat a special case only, but the argument can be generalized easily. We assume $C_1 = (1, 2)$, $C_2 = (3)$, $C_3\cdots C_n$ arbitrary, and we estimate the term arising from a pair-interaction V_{23} between particles 2 and 3. Then we have

$$||V_{23}T(a_1\cdots a_n)e^{iH_C\tau}f||^2 = \int dx_1\cdots dx_N V_{23}^2\prod_{k=1}^{n}\mu_k^{\frac{3}{2}}e^{-\mu_k(y_k+a_k)^2}\,|f_k(z_k,\tau)|^2 \equiv N(a_1\cdots a_n,\tau). \tag{5}$$

Here we have put $b_k = 0$, that is, we have absorbed b_k in a_k. This is possible since

$$\min_{i,k}|a_i - b_i - a_k + b_k| \to \infty$$

for $a\to\infty$. The integrations over the coordinates not belonging to the clusters linked by V_{23} are trivial, they yield a factor 1 if the f_k are properly normalized. We are left with

$$N(a_1\cdots a_n,\tau) = \int dx_1\cdots dx_3 V_{23}^2(x_{23})\times|f_1(x_{12},\tau)|^2\,(\mu_1\mu_2)^{\frac{3}{2}}e^{-\mu_1(y_1+a_1)^2-\mu_3(z_3+a_3)^2}, \tag{6}$$

with $x_{ik} = x_i - x_k$. As integration variables we chose y_1, x_{12} and x_{23}. Then $x_3 = y_1 + \alpha x_{12} + x_{23}$, α being a constant with $|\alpha| < 1$, and the y_1 integration can be carried out:

$$N(a_1\cdots a_n,\tau) = \text{const}\times\int dx_{12}\,dx_{23}V_{23}^2(x_{23})\times|f_1(x_{12},\tau)|^2\,\mu^{\frac{3}{2}}e^{-\mu(a_1-a_3-\alpha x_{12}+x_{23})^2}, \tag{7}$$

$$\mu(\tau) = (2+\tau^2/M_1^2+\tau^2/M_2^2)^{-1}.$$

Since $||f_1(\cdot,\tau)||^2$ is independent of τ, we conclude that $(1+|\tau|^3)N(a_1\cdots a_n,\tau)$ is bounded uniformly in $a_1\cdots a_n$ and $-\infty < \tau < +\infty$. To prove (3), it is enough, therefore, to show that

$$\lim_{a\to\infty} N(a_1\cdots a_n,\tau) = 0. \tag{8}$$

uniformly in any finite τ interval. For this, we split the region of integration in (7) into two parts R_1, R_2:

R_1: $|x_{12}| \le \frac{1}{4}|a_1 - a_2|$ and $|x_{23}| \le \frac{1}{4}|a_1 - a_2|$,

R_2: $|x_{12}| > \frac{1}{4}|a_1 - a_2|$ or $|x_{23}| > \frac{1}{4}|a_1 - a_2|$.

In R_1 we have $|a_1 - a_2 - \alpha x_{12} + x_{23}| \ge \frac{1}{2}|a_1 - a_2|$. Using again that $||f_1(\cdot,\tau)||$ is τ independent, we find after some obvious steps:

$$N(a_1\cdots a_n,\tau) \le \text{const}\times\left(\int_{|x|>\frac{1}{2}|a_1-a_2|} dx(V_{23}^2(x) + |f_1(x,\tau)|^2) + e^{-\frac{1}{4}\mu(a_1-a_2)^2}\right). \tag{9}$$

According to (4), $f_1(\cdot,\tau)$ is strongly continuous in τ, so that the right-hand side of (9) is continuous in τ, and, for any fixed finite τ, converges monotonically decreasing to zero for $|a_1 - a_2|\to\infty$. By Dini's theorem, therefore, the convergence to zero is uniform in any finite τ interval.[2] This concludes the proof of (8) and Theorem 1.

APPLICATION TO SCATTERING THEORY

Cluster properties of the S matrix and related quantities have been the subjects of recent investigations, mainly within the framework of relativistic quantum field theories.[3,4] In this section, we want to start a similar discussion for the nonrelativistic case. The reader is assumed to be familiar, to some extent, with the concepts of time-dependent scattering theory as developed by Jauch.[5]

A channel α of the N-particle system is defined by specifying the fragments $F_1\cdots F_m$, and, in addition, a bound state for each composite fragment. The states in channel α are the states of the form

$$f = f(y_1\cdots y_m)\prod_{k=1}^{m} f_k(z_k), \tag{10}$$

where y_k is the position of the center of mass of F_k, $f_k(z_k)$ the bound-state wavefunction in terms of the internal coordinates of F_k ($f_k = 1$ if F_k is a single particle), and $f(y_1\cdots y_m)$ an arbitrary square integrable function of $y_1\cdots y_m$. The states in channel α form a subspace D_α of $\mathfrak{H}$. The channel Hamiltonian H_α is the full Hamiltonian minus all inter-

[2] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Springer-Verlag, Berlin, 1931), Vol. 1, Chap. 2, Sec. 2.

[3] E. H. Wichmann and J. H. Crichton, Phys. Rev. **132**, 2788 (1963).

[4] K. Hepp, Helv. Phys. Acta (to be published).

[5] J. M. Jauch, Helv. Phys. Acta **31**, 661 (1958).

actions between the fragments. The wave operators

$$\Omega_{\pm}^{\alpha} = \lim_{t\to\pm\infty} e^{iHt}e^{-iH_{\alpha}t}E_{\alpha} \tag{11}$$

exist as strong limits on $\mathfrak{H}$, E_α being the projection onto D_α. They are isometric on D_α and vanish on $D_\alpha^{\perp}$. The proof of (11), which is due to Hack,[6] has been the model for our proof of Theorem 1 and is implicitly contained in it. To show this, we take $m = n$, $C_k = F_k$, and channel states of the form (1), but with f_k as defined in (10). These states span a dense linear set in D_α, and, since $H_C = H_\alpha$, we have

$$e^{iHt}e^{-iH_{\alpha}t}Tf = Tf + i\int_0^t d\tau e^{iH\tau}VTe^{-iH_C\tau}f. \tag{12}$$

From (7), we have concluded that

$$(1 + |\tau|^s)\,||VTe^{-iH_C\tau}f||^2$$

is bounded uniformly in $a_1 \cdots a_n$, τ, and this is all we need for the convergence of (12) as $t \to \pm\infty$. To prove (11), we can of course take $T = 1$, but for later use we note here that the convergence of (12) for $t \to \pm\infty$ is uniform in $a_1 \cdots a_n$.

After these preliminaries, we return to the cluster decomposition $C_1 \cdots C_n$. A channel α is said to be consistent with this decomposition, if each of its fragments belongs to a definite cluster. The set of all consistent channels is denoted by C, it is the same as the set of all channels of the system characterized by the Hamiltonian H_C. If $\alpha \in C$, then the channel α reduces to a channel α_k for each one of the subsystems C_k, the channel Hamiltonians H_α, H_{α_k} and the channel spaces D_α, D_{α_k} being related by

$$H_\alpha = \sum_{k=1}^{n} H_{\alpha_k}, \qquad D_\alpha = \bigotimes_{k=1}^{n} D_{\alpha_k},$$

and

$$\Omega_{\pm}^{C,\alpha} = \lim_{t\to\pm\infty} e^{iH_C t}e^{-iH_\alpha t}E_\alpha,$$

$$\Omega_{\pm}^{\alpha_k} = \lim_{t\to\pm\infty} e^{iH_k t}e^{-iH_{\alpha_k}t}E_{\alpha_k},$$

in the strong sense on the corresponding Hilbert-spaces. Note that H_α and $\Omega_{\pm}^{C,\alpha}$ commute with $T(a_1 \cdots a_n)$ if $\alpha \in C$. The wave operators have the following cluster property:

Theorem 2.

$$\lim_{a\to\infty} T(-a_1 \cdots -a_n)\Omega_{\pm}^{\alpha}T(a_1 \cdots a_n)$$

$$= \begin{cases} \Omega_{\pm}^{C,\alpha} = \prod_{k=1}^{n} \Omega_{\pm}^{\alpha_k} & \text{if } \alpha \in C, \qquad (13a) \\ 0 & \text{if } \alpha \notin C, \qquad (13b) \end{cases}$$

in the strong sense on $\mathfrak{H}$.

[6] M. N. Hack, Nuovo Cimento 13, 231 (1959).

Perhaps more appealing to intuition are the consequences for the transition probabilities. Consider two channels $\alpha, \beta \in C$, and channel states f_α, f_β which are direct products of states f_{α_k}, f_{β_k} in the subchannels α_k, β_k. $P(f_\alpha \to f_\beta) = |(\Omega_-^{\alpha}f_\alpha, \Omega_+^{\beta}f_\beta)|^2$ is the probability for the transition $f_\alpha \to f_\beta$. Theorem 2 implies immediately

$$\lim_{a\to\infty} P(Tf_\alpha \to Tf_\beta) = \prod_{k=1}^{n} P(f_{\alpha_k} \to f_{\beta_k}),$$

that is, if we separate the clusters in the initial and final state in the same way, then the transition probability factorizes, the factors being the transition probabilities in the subsystems C_k. On the other hand, we have for $\alpha \in C$ and for arbitrary β, by Theorem 2,

$$\lim_{a\to\infty} P(Tf_\alpha \to f_\beta) = \lim_{a\to\infty} |(T\Omega_-^{C,\alpha}f_\alpha, \Omega_+^{\beta}f_\beta)|^2 = 0,$$

since $T(a_1 \cdots a_n) \to 0$ for $a \to \infty$, weakly on $\mathfrak{H}$ [see (14)], and similarly for $P(f_\beta \to Tf_\alpha)$. This means, for instance, that if the clusters are separated in the initial state, then the probability for the production of a fragment containing particles of different clusters vanishes asymptotically.

Proof of (13a). By (11), we have

$$\lim_{t\to\pm\infty} e^{iHt}e^{-iH_\alpha t}E_\alpha T(a_1 \cdots a_n) = \Omega_{\pm}^{\alpha}T(a_1 \cdots a_n),$$

strongly on $\mathfrak{H}$, and similarly for $\Omega_{\pm}^{C,\alpha}T$. But the essential point is that the convergence is uniform in $a_1 \cdots a_n$, as has been noted after (12). It is therefore sufficient to show that

$$\lim_{a\to\infty} (e^{iHt}e^{-iH_\alpha t} - e^{iH_C t}e^{-iH_\alpha t})E_\alpha T$$

$$= \lim_{a\to\infty} (e^{iHt} - e^{iH_C t})Te^{-iH_\alpha t}E_\alpha = 0,$$

strongly on $\mathfrak{H}$, for any fixed finite time t. But this is an immediate consequence of Theorem 1.

Proof of (13b). Part b of Theorem 2 is not related to the cluster property of $\exp(iHt)$ and will not be used in the following. It simply states that, for any $f \in \mathfrak{H}$, Tf becomes asymptotically orthogonal to D_α, in the sense that $||E_\alpha Tf|| \to 0$ for $a \to \infty$. This is a consequence of the following fact: Consider the translations $(T(a))(x) = f(x + a)$ on the space $\mathfrak{L}^2$ of square integrable functions of one variable. Then, in the weak sense on $\mathfrak{L}^2$,

$$\lim_{a\to\infty} T(a) = 0, \tag{14}$$

since $(g, T(a)f) = \int dp e^{ipa}\bar{F}(p)G(p) \to 0$ for $a \to \infty$ by the Riemann Lebesgue lemma; F, G being the

Fourier transforms of f, g. We leave the rest of the proof to the reader.

We prepare the discussion of the S operator by proving a lemma which follows from (13a). Let $R_\pm^\alpha$ be the range of $\Omega_\pm^\alpha$, $F_\pm^\alpha$ the projection onto $R_\pm^\alpha$, and similarly for $\Omega_\pm^{C,\alpha}$ and $\Omega_\pm^{\alpha_k}$. The orthogonality theorem states that $R_\pm^\alpha \perp R_\pm^\beta$ if $\alpha \neq \beta$.[5] With $R_\pm$, $R_\pm^C$ we denote the direct sum of all the mutually orthogonal subspaces $R_\pm^\alpha$, $R_\pm^{C,\alpha}$, respectively.

Lemma 1.

$$\lim_{a\to\infty} T(-a_1 \cdots -a_n)F_\pm^\alpha T(a_1 \cdots a_n) = \begin{cases} F_\pm^{C,\alpha} = \prod_{k=1}^{n} F_\pm^{\alpha_k} & \text{if } \alpha \in C, \\ 0 & \text{if } \alpha \notin C, \end{cases}$$

strongly on $R_\pm^C$.

Proof. It is sufficient to prove the lemma for states $f \in R_\pm^{C,\beta}$, β being any channel in C. We have to distinguish two cases:

(a) $\alpha = \beta$: This is possible only if $\alpha \in C$. Then $f = \Omega_\pm^{C,\beta} g$, $F_\pm^{C,\beta} f = f$ and

$$||T^* F_\pm^\alpha T f - F_\pm^{C,\alpha} f|| = ||(1 - F_\pm^\beta) T f||$$

$$= \text{shortest distance between } Tf \text{ and } R_\pm^\beta$$

$$\leq ||T\Omega_\pm^{C,\beta} g - \Omega_\pm^\beta T g|| \to 0 \quad \text{for} \quad a \to \infty,$$

by Theorem 2.

(b) $\alpha \neq \beta$: Then the right-hand side of Lemma 1 vanishes when applied to f, whether $\alpha \in C$ or not, and

$$||F_\pm^\alpha T f|| = ||F_\pm^\alpha (1 - F_\pm^\beta) T f||$$

$$\leq ||(1 - F_\pm^\beta) T f|| \to 0 \quad \text{for} \quad a \to \infty, \quad \text{as in Part a.}$$

By summing formally over all α in Lemma 1 we obtain

$$\lim_{a\to\infty} T(-a_1 \cdots -a_n)F_\pm T(a_1 \cdots a_n) = F_\pm^C, \qquad (15)$$

strongly on $R_\pm^C$, where $F_\pm$ and $F_\pm^C$ are the projections onto $R_\pm$ and $R_\pm^C$. The proof is analogous to the proof of Theorem 4.

Following Jauch,[5] we introduce partial S operators S^α and a total S operator S by

$$S^\alpha = \Omega_+^\alpha \Omega_-^{\alpha *}, \qquad S = \sum_\alpha S^\alpha$$

and similarly, for $\alpha \in C$, $S^{C,\alpha}$, S^C, S^{α_k}, S^k, the last being the total S operator of the subsystem C_k.

Theorem 3.

$$\lim_{a\to\infty} T(-a_1 \cdots -a_n)S^\alpha T(a_1 \cdots a_n) = \begin{cases} S^{C,\alpha} = \prod_{k=1}^{n} S^{\alpha_k} & \text{if } \alpha \in C, \\ 0 & \text{if } \alpha \notin C, \end{cases}$$

in the strong sense on R_-^C.

Proof.

(a) $\alpha \in C$: According to Theorem 2, $T^*\Omega_-^{\alpha *} T$ converges to $\Omega_-^{C,\alpha *}$ for $a \to \infty$, weakly on $\mathfrak{H}$. But on R_-^C, the convergence is strong, since, for any $f \in R_-^C$, $\lim_{a\to\infty} ||T^*\Omega_-^{\alpha *} T f|| = ||\Omega_-^{C,\alpha *} f||$. (This follows from $||T^*\Omega_-^{\alpha *} T f|| = ||T^* F_-^\alpha T f||$, $||\Omega_-^{C,\alpha *} f|| = ||F_-^{C,\alpha} f||$ and Lemma 1.) From this and Theorem 2 we conclude that $T^* S^\alpha T = T^*\Omega_+^\alpha T T^*\Omega_-^{\alpha *} T$ has the property stated in Theorem 3.

(b) $\alpha \notin C$: Then, by Lemma 1, $||\Omega_-^{\alpha *} T f|| = ||F_-^\alpha T f|| \to 0$ for $a \to \infty$.

Theorem 4.

$$\lim_{a\to\infty} T(-a_1 \cdots -a_n)ST(a_1 \cdots a_n) = S^C = \prod_{k=1}^{n} S^k,$$

in the strong sense on R_-^C.

Proof. It is sufficient to prove the theorem for states $f \in R_-^{C,\beta}$, β any channel in C. By Theorem 3, it is only necessary to show that $||T^* \sum_{\alpha\neq\beta} S^\alpha T f|| \to 0$ for $a \to \infty$. But this follows from $||\sum_{\alpha\neq\beta} S^\alpha T f|| = ||\sum_{\alpha\neq\beta} F_-^\alpha T f|| = ||(\sum_{\alpha\neq\beta} F_-^\alpha)(1 - F_-^\beta) T f|| \leq ||(1 - F_-^\beta) T f|| \to 0$ for $a \to \infty$, by Lemma 1.

Analogous theorems hold for $S^{\alpha *}$ and S^* on R_+^C. As an illustration, consider the case of N clusters, each one consisting of one single particle. Then $R_\pm^C = \mathfrak{H}$, and

$$\lim_{a\to\infty} T(-a_1 \cdots -a_N)ST(a_1 \cdots a_N) = 1,$$

$$\lim_{a\to\infty} T(-a_1 \cdots -a_N)S^*T(a_1 \cdots a_N) = 1,$$

in the strong sense on $\mathfrak{H}$. Of course it would be nice to know that the other limits, like Lemma 1, Theorems 3 and 4, also hold on the entire Hilbert space $\mathfrak{H}$. This is in fact asserted by the hypothesis of asymptotic completeness, which states that $R_+ = R_- = \mathfrak{H}$. But so far, this hypothesis has been proved only for $N \leq 3$, under somewhat different conditions on the potentials.[7,8] If these conditions are satisfied,

[7] Two-particle case: T. Ikebe, Arch. Rat. Mech. Anal. 5, 1 (1960).

[8] L. D. Faddeev, Mathematical Problems of the Quantum Theory of Scattering for a Three-Particle System, Publications of the Stoklov Mathematical Institute, No. 69 (1963) (in Russian).

we can conclude, at least, that $R_+^C = R_-^C = \mathfrak{K}$, if none of the clusters consists of more than three particles.

MORE GENERAL POTENTIALS

For a two-body potential which behaves like $|x|^{-r}$ for $x \to \infty$, the requirement of square integrability implies $r > \frac{3}{2}$. This condition can be weakened. In fact, all we have proved so far still holds if we include two-body potentials which are locally square integrable and fall off faster than the Coulomb potential at infinity, in the sense that

$$|V(x)| \leq A\ |x|^{-r}, \quad r > 1, \quad \text{for} \quad |x| > \rho. \tag{16}$$

To demonstrate this, let us treat again the example (5). We set, for $R \geq \rho$,

$$V_R(x) = \begin{cases} V(x) & \text{for} \quad |x| \geq R, \\ 0 & \text{for} \quad |x| < R, \end{cases}$$

and define $N_R(a_1 \cdots a_n, \tau)$ in the same way as $N(a_1 \cdots a_n, \tau)$, but with $V_{23}(x_{23})$ replaced by $V_R(x_{23})$. The only thing we have to show is that there exists an exponent $s > 1$ and, for any $\epsilon > 0$, a sufficiently large R such that

$$\mu^{-s} N_R(a_1 \cdots a_n, \tau) < \epsilon, \tag{17}$$

for all $a_1 \cdots a_n$ and all τ in $-\infty < \tau < +\infty$. Starting from (7) and applying the Hoelder inequality to the x_{23} integral, we get

$$\mu^{-s} N_R(a_1 \cdots a_n, \tau)$$
$$\leq \text{const} \times \mu^{3/2-s}\left(\int_0^\infty d\zeta \zeta^2 e^{-p\mu\zeta^2}\right)^{1/p}\left(\int_R^\infty d\zeta \zeta^{2-2qr}\right)^{1/q}$$
$$= \text{const} \times \mu^{3/2q-s}\left(\int_R^\infty d\zeta \zeta^{2-2qr}\right)^{1/q},$$

with $1 < p, q < \infty$, $p^{-1} + q^{-1} = 1$. Evidently, (17) holds if $3/2q - s \geq 0$ and $2qr > 3$, or, since $s > 1$ but arbitrarily close to 1, if $1 < q < \frac{3}{2}$ and $q > \frac{3}{2}r$. These two conditions for q are compatible if $r > 1$.

The fact that Coulomb potentials are still excluded is of course not surprising. It has been shown by Dollard[9] that the case of charged particles can be fitted into the formalism of Jauch, if the definition of convergence for the time limits (11) is modified. A similar modification might be necessary for the spatial limits in the various cluster properties.

Generalizations analogous to (16) can be expected for many-body potentials.

CONCLUDING REMARKS

So far we have not obtained estimates of the rate of convergence in the various cluster properties. This seems to be possible only if we restrict ourselves to special states of the system and, in addition, impose certain conditions on the behavior of the potentials for large separations of the particles.

ACKNOWLEDGMENT

The author gratefully acknowledges the partial financial support received from the Swiss National Fund.

[9] J. D. Dollard, thesis, Princeton University (1963) and J. Math. Phys. 5, 729 (1964).

Relation between the Onsager and Pfaffian Methods for Solving the Ising Problem. I. The Rectangular Lattice

C. A. HURST
University of Adelaide, Adelaide, South Australia
(Received 28 July 1964)

An algebraic proof is given showing the equivalence of the Pfaffian and Onsager methods of solution of the Ising problem for the two-dimensional rectangular lattice with free edge conditions. With cyclic conditions on the rows and columns and with helical edge conditions it is shown how the two solutions differ. The relation between the appearance of crossed long-range bonds and the appearance of unwanted negative signs in the Pfaffian method is shown explicitly for this particular lattice.

1. INTRODUCTION

IT has been rather puzzling that the two methods at present known for finding exact solutions for the Ising problem, namely the algebraic method of Onsager[1-3] and the combinatorial method employing Pfaffians,[4,5] have exactly the same range of application, although they appear so different in approach. Problems which yield to one method yield to the other, whilst problems which are not tractable by one approach also fail to be exactly solved by the other, although the reasons for this failure appear to have completely different mathematical origins. On the one hand, Ising problems which cannot be solved by the Pfaffian method are characterized by the appearance of crossed bonds which produce unwanted negative signs in the combinatorial generating functions, and such crossed bonds are usually manifestations of the topological structure of the lattice being investigated, e.g., the three-dimensional cubic lattice. On the other hand, the Onsager approach breaks down because the Lie algebra encountered in the process of solution cannot be decomposed into sufficiently simple algebras. It is usually stated that such more complicated algebras occur only when the corresponding lattice has crossed bonds,[6] although an explicit proof of this fact does not appear to be published.

The usual way in which the correspondence between the two methods is established is to interpret the partition function of the Ising model as the generating function for the number of closed polygons drawn on a lattice, using the algebraic transformation due to Oguchi,[7] and then to write down another formula which produces the same generating function. The fact that these two different formulas lead to the same answer can be established geometrically by inspection, or algebraically by an induction argument which is essentially a formal statement of the geometrical relationship.[8] Also the fact that both Onsager's method and the Pfaffian method lead to the same expressions for the partition function and the correlation functions is a confirmation of one's belief in the identity of the two methods.

But at present a detailed correspondence between the two is lacking. It is difficult to see why the two methods have exactly the same limitations, and why it is that although the Pfaffian method always permits *some* answer to be written down it is only in special circumstances that it is the right answer. This is in contrast with the Onsager method which either permits the correct answer to be obtained or cannot be treated at all, except by approximate methods. Apparently the Pfaffian method insists on dealing with simple expressions even when they are no longer relevant for an exact approach.

It is the purpose of this and subsequent papers to discuss this question, and to show how the Pfaffian method can be put in detailed correspondence with the Onsager method. It will be shown that in those cases where the Onsager and Pfaffian methods apply they are both concerned with the solution of a *linear problem* whereas when they can no longer be carried through it is because the Onsager method becomes faced with a *nonlinear* problem for which no exact treatment is available, whilst the Pfaffian

[1] L. Onsager, Phys. Rev. **65**, 177 (1944).
[2] B. Kaufman, Phys. Rev. **76**, 1232 (1949); B. Kaufman and L. Onsager, *ibid.*, p. 1244. Also Y. Nambu, Progr. Theoret. Phys. (Kyoto) **5**, 1 (1950).
[3] T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964). This paper will be referred to as S.
[4] C. A. Hurst and H. S. Green, J. Chem. Phys. **33**, 1059 (1960).
[5] P. W. Kasteleyn, J. Math. Phys. **4**, 287 (1963).
[6] For a simple account of the relation between Onsager's method and the reduction of Lie algebras, see the review article by G. F. Newell and E. W. Montroll, Rev. Mod. Phys. **25**, 353 (1953).
[7] T. Oguchi, J. Phys. Soc. Japan **6**, 27 (1951).
[8] H. S. Green and C. A. Hurst, *Order–Disorder Phenomena* (Interscience Publishers, Inc., New York, 1964), Chap. 8, p. 347.

method, rather perversely, concerns itself with a *linear* problem which is no longer the *correct one*. So this is why the Pfaffian method can always be pushed through to a final closed expression (for asymptotically large lattices) but only in the soluble cases has this expression any useful meaning.

In this paper the special, but fundamental, case of the two-dimensional rectangular lattice will be treated in detail. In Sec. 2 the case of free-edge conditions will be discussed, and in Sec. 3 the various complications associated with toroidal and helical boundary conditions will be resolved.

2. THE RECTANGULAR TWO-DIMENSIONAL LATTICE WITH FREE EDGE CONDITIONS

For a rectangular lattice with n rows and m columns and $N = mn$ sites, the Ising problem is the problem of evaluating the partition function:

$$Z = \operatorname*{tr}_{\sigma_1}\operatorname*{tr}_{\sigma_2}\cdots\operatorname*{tr}_{\sigma_N}\exp\{-E(\sigma_1, \cdots, \sigma_N)/kT\}, \tag{1}$$

where

$$E(\sigma_1, \cdots, \sigma_N) = -J\sum_{i=1}^{N}{}'\,\sigma_{i+1}\sigma_i - J'\sum_{i=1}^{N-m}\sigma_{i+m}\sigma_i. \tag{2}$$

Here J and J' are the energies of interaction between neighboring lattice sites in the horizontal and vertical directions respectively, and σ_i is a 2^N-dimensional matrix which can be written as a direct product

$$\sigma_i = I_2 \times I_2 \times \cdots \times \sigma \times I_2 \times \cdots \times I_2, \tag{3}$$

where

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and σ appears in the jth place. The choice of free edge conditions manifests itself in the occurrence of the upper limit of summation, $N - m$, in the second term on the right-hand side of Eq. (2), and in the notation $\sum'$ for the first term. This latter notation means that when $j = rm$ for $r = 1, \cdots, n$ the term $\sigma_{j+1}\sigma_j$ is omitted. The notation $\operatorname{tr}_{\sigma_j}$ means that we are to take the trace over those indices corresponding to the jth place in the direct product (3).

The matrices σ_i have the algebraic properties

$$\sigma_i\sigma_{i'} - \sigma_{i'}\sigma_i \equiv [\sigma_i, \sigma_{i'}]_- = 0, \qquad j \neq j', \qquad \sigma_i^2 = I_{2^N}. \tag{4}$$

The Oguchi transformation replaces Eq. (1) by

$$Z = (\cosh K)^{N-n}(\cosh K')^{N-m}\operatorname*{tr}_{\sigma_1}\cdots\operatorname*{tr}_{\sigma_N}\prod_{j=1}^{N}{}'(1 + x\sigma_{j+1}\sigma_j)\prod_{j=1}^{N-m}(1 + y\sigma_{j+m}\sigma_j) = (\cosh K)^{N-n}(\cosh K')^{N-m}Z_1, \tag{5}$$

where $x = \tanh K$, $y = \tanh K'$, $K = J/kT$, $K' = J'/kT$, and the notation $\prod'$ has the same significance as the notation $\sum'$ in Eq. (2), i.e., the terms $\sigma_{j+1}\sigma_j$ are omitted when $j = rm$.

The relation between Eq. (5) and a combinatorial problem is well known. An alternative expression which is equal to Eq. (5) is given by

$$Z_1 = 2^{-N}\operatorname*{tr}_{\sigma_1^{(1)}}\cdots\operatorname*{tr}_{\sigma_N^{(1)}}\operatorname*{tr}_{\sigma_1^{(2)}}\cdots\operatorname*{tr}_{\sigma_N^{(2)}} \times \prod_{j=1}^{N}{}''(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)} + x\sigma_j^{(1)}\sigma_{j-1}^{(1)} + y\sigma_j^{(2)}\sigma_{j-1}^{(1)} + x\sigma_j^{(1)}\sigma_{j-m}^{(2)} + y\sigma_j^{(2)}\sigma_{j-m}^{(2)} + xy\sigma_j^{(2)}\sigma_j^{(1)} + xy\sigma_j^{(2)}\sigma_j^{(1)}\sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)}), \tag{6}$$

where the notation $\prod''$ denotes the appropriate incorporation of the edge conditions. These conditions here are the following: all terms containing $\sigma_{rm}^{(1)}$ are omitted for $r = 0, 1, \cdots, n$, as also are all terms containing $\sigma_j^{(2)}$ for $j > N - m$ or $j < 0$.

The matrices $\sigma_j^{(1)}$, $\sigma_j^{(2)}$ are a set of 2^{2N}-dimensional matrices with the properties:

$$[\sigma_i^{(i)}, \sigma_{i'}^{(i')}]_- = 0 \quad \text{for} \quad i \neq i' \quad \text{or} \quad j \neq j',$$

and

$$(\sigma_i^{(i)})^2 = I_{2^{2N}}. \tag{7}$$

The equivalence of Eqs. (5) and (6) can be seen by inspecting the structure of the terms which survive on taking the trace. In both cases one has a sum of terms which arise from casting out all terms which contain an odd power of any σ matrix, and each such term can then be made to correspond to a closed polygon (or a set of closed polygons) drawn on the lattice. However for our purposes we will show the correspondence algebraically. It can be shown, by multiplying out and using Eq. (7) that

$$(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)} + x\sigma_j^{(1)}\sigma_{j-1}^{(1)} + y\sigma_j^{(2)}\sigma_{j-1}^{(1)} + x\sigma_j^{(1)}\sigma_{j-m}^{(2)} + y\sigma_j^{(2)}\sigma_{j-m}^{(2)} + xy\sigma_j^{(2)}\sigma_j^{(1)} + xy\sigma_j^{(2)}\sigma_j^{(1)}\sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)}) = (1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)})(1 + x\sigma_j^{(1)}\sigma_{j-1}^{(1)})(1 + y\sigma_j^{(2)}\sigma_{j-1}^{(1)}). \tag{8}$$

Edge conditions make obvious modifications to Eq. (8).

Because the matrices $\sigma_{rm}^{(1)}$ do not appear in Z_1, the trace operation over the indices corresponding to them is trivial and gives a factor 2^n. The factor $\frac{1}{2}(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)})$ has the property of a δ symbol, for it is zero when $\sigma_{j-m}^{(2)}$ and $\sigma_{j-1}^{(1)}$ have opposite signs

and is one when they have the same sign. So if this factor appears in Z_1 it may simply be dropped and in the remaining expression wherever $\sigma_{j-m}^{(2)}$ appears it can be replaced $\sigma_{j-1}^{(1)}$. A formal proof of this statement will be given, for an analogue will be needed later. Equation (6) may be written in the form

$$Z_1 = \operatorname*{tr}_{\sigma_{j-m}^{(2)}} \operatorname*{tr}_{\sigma_{j-1}^{(1)}} \tfrac{1}{2}(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)}) \times (A + B\sigma_{j-1}^{(1)} + C\sigma_{j-m}^{(2)} + D\sigma_{j-1}^{(1)}\sigma_{j-m}^{(2)}), \quad (9)$$

where A, B, C, D are expressions independent of $\sigma_{j-1}^{(1)}$ and $\sigma_{j-m}^{(2)}$. Hence Z_1 is given by

$$Z_1 = \operatorname*{tr}_{\sigma_{j-1}^{(1)}} (A + B\sigma_{j-1}^{(1)} + C\sigma_{j-1}^{(1)} + D). \quad (9')$$

This is just the expression we get when we replace $\sigma_{j-m}^{(2)}$ by $\sigma_{j-1}^{(1)}$ throughout before taking the trace over $\sigma_{j-m}^{(2)}$. So the factor $\frac{1}{2}(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)})$ may be dropped from Eq. (8) for all j, and then Eq. (6) becomes

$$Z_1 = 2^{-m+1} \operatorname*{tr}_{\sigma_1^{(1)}} \cdots \operatorname*{tr}_{\sigma_N^{(1)}} \prod_{j=1}^{N}{}'' (1 + x\sigma_j^{(1)}\sigma_{j-1}^{(1)}) \times (1 + y\sigma_{j+m-1}^{(1)}\sigma_{j-1}^{(1)}). \quad (6')$$

(A number of minor modifications due to edge conditions will be discussed later.) This is the same as the Oguchi expression (5) if we make the correspondence

$$\sigma_j^{(1)} \rightarrow \sigma_{j+1}. \quad (10)$$

The first step in setting up the Pfaffian method is to regard the matrices $\sigma_j^{(i)}$ appearing in Eq. (6) as anticommuting rather than commuting quantities, so that the multiplication relations (7) are replaced by

$$\sigma_j^{(i)}\sigma_{j'}^{(i')} + \sigma_{j'}^{(i')}\sigma_j^{(i)} \equiv [\sigma_j^{(i)}, \sigma_{j'}^{(i')}]_+ = 2\delta_{ii'}\delta_{jj'}. \quad (11)$$

Once again these matrices may be represented as 2^{2N}-dimensional matrices which are direct products of $2N$ two-dimensional matrices. The order of the factors in (6) is to be taken in the order of increasing j reading from right to left. It has been shown[9] that this change does not alter the value of Z. Furthermore, this change does not affect the factorization described in Eq. (8) if the order of factors is the same as given there. An important feature of the factorization given in Eq. (8) is that no term containing $\sigma_{j-m}^{(2)}$ nor $\sigma_{j-1}^{(1)}$ occurs to the left of the factor $(1 + \sigma_{j-m}^{(2)}\sigma_{j-1}^{(1)})$ appearing in Eq. (6). Hence we can write Eq. (6) once again in the form (9) with the correct order of the factors $\sigma_{j-1}^{(1)}$ and $\sigma_{j-m}^{(2)}$. A, B, C, and D again do not depend on these two matrices. Taking the trace of $\sigma_{j-m}^{(2)}$ we find that

$$Z_1 = \operatorname*{tr}_{\sigma_{j-1}^{(1)}} (A + B\sigma_{j-1}^{(1)} - C\sigma_{j-1}^{(1)} + D). \quad (9'')$$

Equation (9'') differs from Eq. (9) in the sign before the coefficient C, but this difference is inessential because after taking the trace over $\sigma_{j-1}^{(1)}$, the terms in B and C will disappear. So once again we find that Z_1 is given by Eq. (6') even for anticommuting matrices. This expression (6') is very reminiscent of that encountered in the Onsager treatment particularly when the form of proof utilized by Kaufman or Nambu[2] is followed. However, in Onsager's method the matrices are only of dimension 2^m rather than 2^N (the operation of taking traces of all the relevant matrices $\sigma_j^{(2)}$ and redundant indices $\sigma_{rm}^{(1)}$ has lowered the order from 2^{2N} to 2^N). In order to complete the correspondence between the two methods, the procedure of Schultz, Mattis, and Lieb[3] of constructing partial density matrices with an associated *transfer matrix* will be followed. This important paper has done much to simplify and clarify the esoteric algebraic manipulations involved in the original presentation of the Onsager method.

We define a partial density matrix

$$P_M(\sigma_{M+1}, \cdots, \sigma_N) = \operatorname*{tr}_{\sigma_1} \cdots \operatorname*{tr}_{\sigma_M} \prod_{j=1}^{M}{}'' (1 + x\sigma_{j+1}\sigma_j)(1 + y\sigma_{j+m}\sigma_j), \quad (12)$$

$$P_0 = 1,$$

and then

$$P_{M+1}(\sigma_{M+2}, \cdots, \sigma_N) = \operatorname*{tr}_{\sigma_{M+1}} (1 + x\sigma_{M+2}\sigma_{M+1})(1 + y\sigma_{M+m+1}\sigma_{M+1})P_M. \quad (13)$$

Because of the structure of the bond connections, P_M can only depend on the matrices $\sigma_{M+1}, \cdots, \sigma_{M+m}$, and if $M = rm + k$ with $0 \leq r < n$, $1 \leq k \leq m$, we will denote the matrices $\sigma_{M+1}, \cdots, \sigma_{M+m+1}$ by $\sigma_{k+1}', \sigma_{k+2}', \cdots, \sigma_m', \sigma_1, \cdots, \sigma_{k-1}, \sigma_k, \sigma_{k+1}$. So if we write

$$P_M = A_M + B_M\sigma_{k+1}',$$

where A_M, B_M depend on the matrices $\sigma_{k+2}', \cdots, \sigma_k$, and are of even, odd, degree in these matrices respectively, then we have

$$P_{M+1} = \operatorname*{tr}_{\sigma_{k+1}'} [(1 + x\sigma_{k+2}'\sigma_{k+1}' + y\sigma_{k+1}\sigma_{k+1}' - xy\sigma_{k+2}'\sigma_{k+1})(A_M + B_M\sigma_{k+1}')] = 2(1 - xy\sigma_{k+2}'\sigma_{k+1})A_M - 2(x\sigma_{k+2}' + y\sigma_{k+1})B_M. \quad (14)$$

[9] See Ref. 8, especially Chap. 4. In this reference, instead of matrices $\sigma_j^{(i)}$, a pair of annihilation and creation operators $a_j^{(i)}$, $a_j^{(i)*}$ are used, but the replacement

$$\sigma_j^{(i)} \rightarrow a_j^{(i)*}, \quad \sigma_{j-1}^{(1)} \rightarrow a_{j-1}^{(1)}, \quad \sigma_{j-m}^{(2)} \rightarrow a_{j-m}^{(2)}$$

in the jth term of Eq. (6) of this paper will give the appropriate transcription.

The negative sign comes from the anticommutation properties of σ'_{k+1}. Because the structure of these expressions is such that we need never consider more than m σ's at a time, we may restrict ourselves to a representation with 2^m-dimensional matrices. This representation can be written, with some resemblance to Eq. (3), in the form

$$\sigma_j = \tau \times \tau \times \cdots \tau \times \sigma \times \delta \times \cdots \times \delta, \qquad m \text{ terms} \quad (1 \leq k \leq m) \tag{15}$$

where now we put, in contrast to Eq. (3),

$$\sigma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \delta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{16}$$

and the matrix σ appears in the jth place. In following through this process of reduction it is not strictly necessary to relate P_{M+1} to P_M. Instead one can relate P_{M+m} to P_M, i.e., work along a column instead of a row.

Following S, Eq. (14) can be written as

$$P_{M+1} = 2[(1 - xy\sigma_{k+2}\sigma_{k+1})(1 - \mathsf{N}_{k+1}) + (x\sigma_{k+2}\sigma_{k+1} + y\mathsf{N}_{k+1})]P_M. \tag{17}$$

N_{k+1} is an operator which counts the number of times σ_{k+1} appears in P_M. As pointed out in S there is actually no operator which has these desired properties except when applied to a particular state. The particular state can be chosen to be the "vacuum state" $|0\rangle$ representing a 2^m-rowed column matrix with all the spins upwards, i.e.,

$$\tau_j |0\rangle = |0\rangle, \qquad j = 1, \cdots, m, \tag{18}$$

where $\tau_j = \delta \times \delta \times \cdots \times \delta \times \tau \times \delta \times \cdots \times \delta$. If we define a third anticommuting matrix

$$\rho_j = \tau \times \cdots \times \tau \times \rho \times \delta \times \cdots \times \delta,$$

with

$$\rho = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

in the jth place, then the operator

$$\mathsf{N}_j = \tfrac{1}{2}(1 + i\sigma_j\rho_j) = \tfrac{1}{2}(1 - \tau_j)$$

has the properties

$$(1 - \mathsf{N}_j)|0\rangle = |0\rangle,$$
$$\mathsf{N}_j\sigma_j |0\rangle = \sigma_j(1 - \mathsf{N}_j)|0\rangle = \sigma_j |0\rangle,$$
$$\mathsf{N}_j |0\rangle = 0. \tag{19}$$

Under these circumstances the operator N_{k+1} has the required properties and we can replace Eq. (17) by

$$\begin{aligned} P_{M+1}|0\rangle &= [(1 - xy\sigma_{k+2}\sigma_{k+1})(1 - i\sigma_{k+1}\rho_{k+1}) \\ &\quad + (x\sigma_{k+2}\sigma_{k+1} + y)(1 + i\sigma_{k+1}\rho_{k+1})]P_M|0\rangle \\ &= [(1+y) - i(1-y)\sigma_{k+1}\rho_{k+1}](1 + ix\sigma_{k+2}\rho_{k+1})P_M|0\rangle \\ &= \frac{(2\sinh 2K')^{\frac{1}{2}}}{\cosh K \cosh K'} e^{-iK^*\sigma_{k+1}\rho_{k+1}} e^{iK\sigma_{k+2}\rho_{k+1}} P_M|0\rangle, \end{aligned} \tag{20}$$

where $e^{-2K'} = \tanh K^*$.

Edge conditions make the following modifications: (i) For $M + 1 = rm$ there is no horizontal bond to the right, and we put $x = K = 0$; (ii) For $1 \leq M \leq m$, there are no vertical bonds leading downwards, so, in the calculations leading to (6'), factors like $(1 + \sigma^{(2)}_{j-m}\sigma^{(1)}_{j-1})$ are missing from the first m terms. The typical vertex factor in this case is $(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1} + y\sigma^{(2)}_j\sigma^{(1)}_{j-1} + xy\sigma^{(2)}_j\sigma^{(1)}_j)$ which factorizes to $(1+x\sigma^{(1)}_j\sigma^{(1)}_{j-1})(1+y\sigma^{(2)}_j\sigma^{(1)}_{j-1})$, and with the correspondence (10) this is $(1+x\sigma_{j+1}\sigma_j)(1+y\sigma_{j+m}\sigma_j)$ as before; (iii) For $M = rm + 1$, there are no horizontal bonds to the left so the expression (5) becomes $(1 + x\sigma^{(1)}_{rm+1}\sigma^{(2)}_{(r-1)m+1})(1 + y\sigma^{(2)}_{rm+1}\sigma^{(2)}_{(r-1)m+1})$ and instead of taking the trace over $\sigma^{(2)}_{rm+1}$ we make the replacement $\sigma^{(2)}_{rm+1} \to \sigma_{(r+1)m+1}$; (iv) For $N - m < M \leq N$, there are no vertical bonds leading upwards, and we simply put $y = 0$; (v) For $M = 1$, the only factor is $(1 + xy\sigma^{(2)}_1\sigma^{(1)}_1)$ so we make the replacements $\sigma^{(2)}_1 \to \sigma_{m+1}$, $\sigma^{(1)}_1 \to \sigma_2$ and then write $2(1 + xy\sigma_{m+1}\sigma_2)$ as

$$e^{-iK^*\sigma_1\rho_1} e^{iK\sigma_2\rho_1},$$

using the results

$$(1 - i\sigma_1\rho_1)|0\rangle = 2|0\rangle, \qquad (1 + i\sigma_1\rho_1)|0\rangle = 0.$$

From the relation

$$Z_1 = 2^m \langle 0| P_N |0\rangle,$$

we find, when all edge effects are properly incorporated:

$$Z = (2\sinh 2K')^{\frac{1}{2}(N-m)} \langle 0| \prod_{j=1}^{m-1} e^{iK\sigma_{j+1}\rho_j} V^{n-1} |0\rangle, \tag{21}$$

where

$$V = \prod_{j=1}^{m} e^{-iK^*\sigma_j\rho_j} \prod_{j=1}^{m-1} e^{iK\sigma_{j+1}\rho_j}. \tag{22}$$

The expressions (21) and (22) form the starting point for the treatment of the Ising problem by Kaufman using spinor algebraic methods. So we have established a direct algebraic connection between the Pfaffian method which starts from Eq. (6) and the Onsager method which starts from Eq. (1).

The connection between Eq. (21) and the combinatorial approach can be seen by expanding Eq. (22)

as a polynomial in x and y. A term $ix\sigma_{j+1}\rho_j$ corresponds to a horizontal bond joining the jth and $(j+1)$th columns, with the row designated by the particular V in which it occurs. The term $(1 - i\sigma_j\rho_j)$ corresponds to a vertical bond in the jth column and the row again is designated by the particular V in which it occurs.

Because of the identity of Eqs. (1) and (6) only one determinant is required in the Pfaffian approach to give the correct expression for the partition function. This equality is a consequence of the simple free edge conditions. In the next section it will be shown how considerably greater complication arises if either toroidal or helical boundary conditions are employed.

3. TOROIDAL AND HELICAL EDGE CONDITIONS

The most commonly used edge conditions are the toroidal and helical conditions. For toroidal conditions they are define by replacing Eq. (2) by:

$$E(\sigma_1, \cdots, \sigma_N) = -J \sum_{j=1}^{N}{}' \sigma_{j+1}\sigma_j - J' \sum_{j=1}^{N} \sigma_{j+m}\sigma_j, \tag{23}$$

with $\sigma_{j+m} \equiv \sigma_{j+m-N}$ for $N - m < j \leq N$, and $\sum'$ here means that when $j = rm$ we replace $\sigma_{rm}\sigma_{rm+1}$ by $\sigma_{rm}\sigma_{(r-1)m+1}$.

For helical boundary conditions we replace (2) by

$$E(\sigma_1, \cdots, \sigma_N) = -J \sum_{j=1}^{N} \sigma_{j+1}\sigma_j - J' \sum_{j=1}^{N} \sigma_{j+m}\sigma_j, \tag{24}$$

with $\sigma_{N+1} \equiv \sigma_1$.

Toroidal conditions may be modified by allowing the second summation in Eq. (23) to run only up to $N - m$, so that we have a lattice wrapped on a cylinder and not on a torus. For the Onsager method this has been treated by S.

(i) Cyclic End Conditions on Rows Only: Cylindrical Edge Conditions

We will consider first of all the case just mentioned, i.e., cyclic end conditions on the rows but not on the columns. The considerations of Sec. 2 can be carried through up to the derivation of Eq. (9″). The argument subsequent to that equation which permits the dropping of the terms linear in $\sigma_{j-1}^{(1)}$ is no longer correct without qualification. For if $j = rm + 1$, we must replace $\sigma_{rm}^{(1)}$ by $\sigma_{(r+1)m}^{(1)}$ and such a term also occurs associated with the vertex $j = (r+1)m$. Hence Eq. (9) for this value of j will have the structure

$$\begin{aligned} Z_1 &= \operatorname*{tr}_{\sigma_{(r-1)m+1}^{(2)}} \operatorname*{tr}_{\sigma_{(r+1)m}^{(1)}} \tfrac{1}{2}(A' + B'\sigma_{(r+1)m}^{(1)}) \\ &\quad \times (1 + \sigma_{(r-1)m+1}^{(2)}\sigma_{(r+1)m}^{(1)})(A + B\sigma_{(r+1)m}^{(1)} \\ &\quad + C\sigma_{(r-1)m+1}^{(2)} + D\sigma_{(r+1)m}^{(1)}\sigma_{(r-1)m+1}^{(2)}) \\ &= \operatorname*{tr}_{\sigma_{(r+1)m}^{(1)}} (A' + B'\sigma_{(r+1)m}^{(1)}) \\ &\quad \times [(A + D) + (B - C)\sigma_{(r+1)m}^{(1)}], \end{aligned} \tag{25}$$

and the term $B'C$ appears with the wrong sign. This fact can be expressed in another way by saying that we may replace $\sigma_{(r-1)m+1}^{(2)}$ by $\sigma_{(r+1)m}^{(1)}$ everywhere so long as we take care to note that $\sigma(_{(r+1)m}^{(1)})^2 = 1$ except when one factor comes from the lattice point $j = (r+1)m$ and the other from the lattice point $j = (r-1)m + 1$, and in that case $(\sigma_{(r+1)m}^{(1)})^2 = -1$. In graphical language, the exceptional situation arises when a bond joins $(r+1)m$ and $rm + 1$, and another bond joins $(r-1)m + 1$ and $rm + 1$. At all other points in a row there is no difficulty, and we may carry out on them the reductions leading to Eq. (20), to find

$$\begin{aligned} &P_{(r+1)m}(\sigma_1, \cdots \sigma_m)\,|0\rangle \\ &= \operatorname*{tr}_{\sigma_1'} [(1+y) - i(1-y)\sigma_m\rho_m](1 + ix\sigma_1'\rho_m) \\ &\quad \times \prod_{j=2}^{m-1} [(1+y) - i(1-y)\sigma_j\rho_j](1 + ix\sigma_{j+1}\rho_j) \\ &\quad \times (1 + x\sigma_2\sigma_1')(1 + y\sigma_1\sigma_1')P_{rm}(\sigma_1', \sigma_2, \cdots, \sigma_m)\,|0\rangle, \\ &= \operatorname*{tr}_{\sigma_1'} (1 + ix\sigma_1'\rho_m)(1 + x\sigma_2\sigma_1') \\ &\quad \times (1 + y\sigma_1\sigma_1')(A + B\sigma_1')\,|0\rangle, \end{aligned} \tag{26}$$

where A, B are independent of σ_1, σ_1', and the exceptional condition referred to above is interpreted as saying that the term $ix\sigma_1'\rho_m B\sigma_1'$ must be replaced by $-ix\sigma_1'\rho_m B\sigma_1'$. Evaluation of this trace then gives for $P_{(r+1)m}|0\rangle$

$$\begin{aligned} &[(1+y) - i(1-y)\sigma_1\rho_1](1 + ix\sigma_2\rho_1) \\ &\qquad \times (1 - ix\sigma_1\rho_m)(A + B\sigma_1)\,|0\rangle. \end{aligned} \tag{27}$$

Now we define the matrix

$$U = \tau_1\tau_2 \cdots \tau_m, \tag{28}$$

where τ_j is defined by Eq. (18). This matrix has the property that it counts as $+1$ if the total number of σ's in $A + B\sigma_1$ is even, and as -1 if they are odd. So we can make the replacement

$$(1 - ix\sigma_1\rho_m) \to (1 - ix\sigma_1\rho_m U) = (1 + ix\sigma_1 U\rho_m),$$

because both A and $B\sigma_1$ have an even number of σ's. Then

$$P_{(r+1)m}\,|0\rangle = VP_{rm}\,|0\rangle,$$

with

$$V = \prod_{j=1}^{m} e^{-iK^*\sigma_j\rho_j} \prod_{j=1}^{m} e^{iK\sigma_{j+1}\rho_j}, \tag{29}$$

with $\sigma_{m+1} \equiv \sigma_1 U$. This is the result obtained in the standard treatments of the Onsager approach, so once again the equivalence of the two methods has been demonstrated. It is interesting to notice that the Pfaffian method leads to the combination $\sigma_1\rho_m$ which is quadratic in the anticommuting quantities, whereas the Onsager method leads to the combination $\sigma_1\sigma_m \equiv \sigma_1 U\rho_m$ which is of degree m in the anticommuting quantities. Following S we will call $\sigma_1\rho_m$ a *linear term* in analogy with the correspondence between linear equations and quadratic Hamiltonian forms in mechanics. Similarly $\sigma_1 U\rho_m$ will be called a *nonlinear* term. So for this problem it is possible to replace the correct nonlinear term by the incorrect term because of the properties of the U matrix. The U matrix can be regarded as counting the number of crossing points of the long-range bond joining $rm + 1$ and $(r + 1)m$ with the vertical bonds coming from the row $(r - 1)$. As the number of such crossing points is even the Pfaffian method gives the correct sign. If the number of crossing points were odd then the Pfaffian method would give the wrong sign, and it would not be possible to linearize the Onsager approach in this way. In toroidal boundary conditions just this situation arises.

(ii) Cyclic Conditions on Columns Only

The typical term in the first row is now

$$(1 + \sigma^{(2)}_{N-m+j}\sigma^{(1)}_{j-1})(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1})(1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1}),$$

whilst on the last row it is

$$(1 + \sigma^{(2)}_{N-2m+j}\sigma^{(1)}_{N-m+j-1})(1 + x\sigma^{(1)}_{N-m+j}\sigma^{(1)}_{N-m+j-1}) \times (1 + y\sigma^{(2)}_{N-m+j}\sigma^{(1)}_{N-m+j-1}),$$

with $1 < j \leq m$. For $j = 1$ the terms are

$$(1 + x\sigma^{(1)}_1\sigma^{(2)}_{N-m+1})(1 + y\sigma^{(2)}_1\sigma^{(2)}_{N-m+1})$$

and

$$(1 + x\sigma^{(1)}_{N-m+1}\sigma^{(2)}_{N-2m+1})(1 + y\sigma^{(2)}_{N-m+1}\sigma^{(2)}_{N-2m+1}),$$

respectively.

Now Z_1 can be written as

$$Z_1 = \operatorname{tr} A(1 + \sigma^{(2)}_{N-m+j}\sigma^{(1)}_{j-1})(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1}) \times (1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})B,$$

where A does not depend on $\sigma^{(1)}_{j-1}$ and B does not depend on $\sigma^{(2)}_{N-m+j}$ and they contain even-order products of matrices only. Hence,

$$\begin{aligned} Z_1 &= \operatorname{tr}[A(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1})(1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})B] \\ &\quad - \operatorname{tr}[\sigma^{(1)}_{j-1}A(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1}) \\ &\quad \times (1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})B\sigma^{(2)}_{N-m+j}], \\ &= \operatorname{tr}[A(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1})(1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})B] \\ &\quad + \operatorname{tr}[\sigma^{(2)}_{N-m+j}A(1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1}) \\ &\quad \times (1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})B\sigma^{(1)}_{j-1}], \end{aligned} \tag{30}$$

using the properties of the trace and the anticommutativity of $\sigma^{(1)}_{j-1}$ and $\sigma^{(2)}_{N-m+j}$.

When this process is repeated with another index j', $1 < j < j' \leq m$, Eq. (30) becomes

$$\operatorname{tr} A + \operatorname{tr} \sigma^{(2)}_{N-m+j}A\sigma^{(1)}_{j-1} + \operatorname{tr} \sigma^{(2)}_{N-m+j'}A\sigma^{(1)}_{j'-1} + \operatorname{tr} \sigma^{(2)}_{N-m+j}\sigma^{(2)}_{N-m+j'}A\sigma^{(1)}_{j'-1}\sigma^{(1)}_{j-1},$$

where now we put

$$A = A'(1 + x\sigma^{(1)}_{j'}\sigma^{(1)}_{j'-1})(1 + y\sigma^{(2)}_{j'}\sigma^{(1)}_{j'-1})B' \times (1 + x\sigma^{(1)}_j\sigma^{(1)}_{j-1})(1 + y\sigma^{(2)}_j\sigma^{(1)}_{j-1})C'$$

and A', B', C' do not depend on $\sigma^{(1)}_{j-1}$, $\sigma^{(1)}_{j'-1}$, $\sigma^{(2)}_{N-m+j}$, and $\sigma^{(2)}_{N-m+j'}$ and are of even order in the remaining matrices. This process can be repeated for all the lattice points in the first row. For the first point a slightly modified procedure is necessary.

We introduce a new matrix σ_1, $\sigma_1^2 = 1$, anticommuting with all other matrices and we write this first factor as

$$(1 + xy\sigma^{(2)}_1\sigma^{(1)}_1) + (x\sigma^{(1)}_1 + y\sigma^{(2)}_1)\sigma_1\sigma^{(2)}_{N-m+1},$$

and then

$$\begin{aligned} Z_1 &= \operatorname{tr}[A(1 + xy\sigma^{(2)}_1\sigma^{(1)}_1)] \\ &\quad + \operatorname{tr}[\sigma^{(2)}_{N-m+1}A(x\sigma^{(1)}_1\sigma_1 + y\sigma^{(2)}_1\sigma_1)\sigma_1]. \end{aligned} \tag{31}$$

Now when we take the trace over all the matrices $\sigma^{(2)}$ except those in the first column and the last row, we can replace $\sigma^{(2)}_{j-m}$ and $\sigma^{(1)}_{j-1}$ by σ_j in the usual way, and we also put $\sigma_{(r+1)m+1}$ for $\sigma^{(2)}_{rm+1}$ for $1 \leq r < n$. We then carry out the same procedure as described in Sec. 2. After we have taken the trace over the matrices σ_{N-m+j}, the only surviving ones are the set $\sigma^{(2)}_{N-m+j}$ and following the standard procedure these can now be renamed σ_j. So finally we obtain the expression

$$Z = (2\sinh 2K')^{\frac{1}{2}N}\Big[\langle 0|\, V^n\,|0\rangle + \sum_{j=1}^{m}\langle 0|\,\sigma_j V^n\sigma_j\,|0\rangle + \sum_{1\leq j<j'\leq m}\langle 0|\,\sigma_j\sigma_{j'}V^n\sigma_{j'}\sigma_j\,|0\rangle + \cdots\Big], \tag{32}$$

where V is the matrix defined in Eq. (22). A term such as tr $(\sigma_i V^n \sigma_i)$ may be represented graphically by a set of closed polygons drawn on the lattice with the exception of the jth column where there is a pair of bonds, called *external bonds*, leading upwards from the last row and downwards from the first row, and such a pair of bonds can be interpreted as a *long-range bond joining the first and last row*, and so closing the corresponding polygon to which they are connected. Because

$$(2 \sinh 2K')^{\frac{1}{2}m} \langle 0| e^{-iK^* \sigma_i \rho_j} = (2 \cosh K')^m \langle 0|$$

the first term of Eq. (32) describes the Ising problem with free edge conditions. A complete set of states is given by

$$\sigma_{j_q} \sigma_{j_{q-1}} \cdots \sigma_{j_2} \sigma_{j_1} |0\rangle \quad \text{with} \quad j_1 < j_2 < \cdots < j_q, \quad 0 \leq q \leq m,$$

so that Eq. (32) is nothing but

$$Z = (2 \sinh 2K')^{\frac{1}{2}N} \text{ tr } V^n, \tag{32'}$$

the usual Onsager expression. The fact that Eqs. (32) and (32') are the same can be interpreted graphically by noting that, with the long-range vertical bonds introduced here and the free edge conditions on the rows, there can only be an even number of crossing points and so there can be no unwanted negative signs.

(iii) Cyclic Conditions on Rows and Columns: Toroidal Edge Conditions

If there are cyclic conditions on the ends of the rows then the discussion of Subsection (i) shows that the matrix U must be introduced to relate the two methods. It is necessary that only the eigenvalue $U = +1$ occurs and this is equivalent to the requirement that the σ matrices in P_{rm}, for all $r < n$, occur in even-order products only. Now because of the appearance of additional σ matrices in the initial states $\sigma_{j_q} \cdots \sigma_{j_1} |0\rangle$ this matrix U will only have the correct eigenvalue if the initial state has an even number of σ's. Otherwise the sign will be incorrect. But this is just the condition for an even number of intersections of the long-range vertical and horizontal bonds, and so, in order for the Pfaffian method to give the correct result, a prescription such as given by Potts and Ward[10] must be used. This prescription can be verified algebraically by comparing the two expressions.

[10] R. B. Potts and J. C. Ward, Progr. Theoret. Phys. (Kyoto) **13**, 38 (1955).

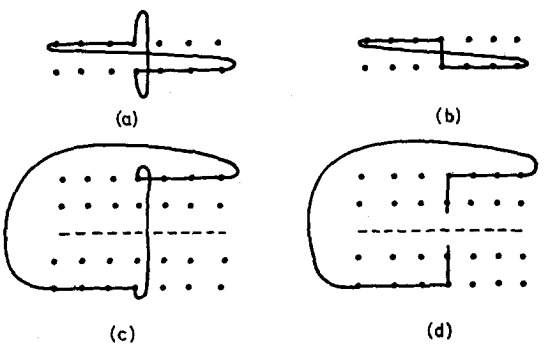

FIG. 1. Relation between long-range bonds and crossing points for helical edge conditions. (a) Long-range vertical bond and helical row connection. (b) Short-range vertical bond and helical row connection. (c) Long-range vertical bond and long-range helical row connection. (d) Short-range vertical bonds and long-range helical row connections.

(iv) Helical Edge Conditions

With helical edge conditions the partition function has the structure

$$Z = \text{tr } (1 + \sigma_{N-m}^{(2)} \sigma_{N-1}^{(1)})(1 + x\sigma_N^{(1)} \sigma_{N-1}^{(1)})(1 + y\sigma_N^{(2)} \sigma_{N-1}^{(1)}) A \times (1 + \sigma_{N-m+1}^{(2)} \sigma_N^{(1)})(1 + x\sigma_1^{(1)} \sigma_N^{(1)})(1 + y\sigma_1^{(2)} \sigma_N^{(1)}),$$

where A does not contain $\sigma_N^{(1)}$ and is even order in all the matrices. Hence we can write as before

$$\begin{aligned} Z = \text{tr } & (1 + \sigma_{N-m}^{(2)} \sigma_{N-1}^{(1)})(1 + x\sigma_N^{(1)} \sigma_{N-1}^{(1)})(1 + y\sigma_N^{(2)} \sigma_{N-1}^{(1)}) A \\ & \times (1 + x\sigma_1^{(1)} \sigma_N^{(1)})(1 + y\sigma_1^{(2)} \sigma_N^{(1)}) \\ & + \text{tr } [\sigma_{N-m+1}^{(2)} (1 + \sigma_{N-m}^{(2)} \sigma_{N-1}^{(1)}) \\ & \times (1 - x\sigma_N^{(1)} \sigma_{N-1}^{(1)})(1 + y\sigma_N^{(2)} \sigma_{N-1}^{(1)}) A \\ & \times (1 + x\sigma_1^{(1)} \sigma_N^{(1)})(1 + y\sigma_1^{(2)} \sigma_N^{(1)}) \sigma_N^{(1)}]. \end{aligned} \tag{33}$$

Now the trace over all $\sigma^{(2)}$ matrices except those in the last row can be taken and then the procedure used in Subsection (iv) followed, after making the correspondences $\sigma_{i-1}^{(1)} \to \sigma_i$, $\sigma_N^{(1)} \to \sigma_1$. The latter correspondence may not appear to be justified because $\sigma_N^{(1)}$ appears at both ends of the product inside the trace. However, by using the properties of the trace, this difficulty can be overcome, although the details are tedious. It is found that instead of Eq. (32') we have

$$Z = (2 \sinh 2K')^{\frac{1}{2}N} \text{ tr } V V'^{n-1}, \tag{34}$$

where V is as defined in Eq. (29) and V' is obtained from V by replacing K by $-K$ for $s = m$. So helical boundary conditions differ from toroidal boundary conditions in that long-range vertical bonds now give the correct sign, indicating no crossing points, whereas short-range vertical bonds give the incorrect sign when there is a long-range horizontal

bond connecting the lattice point rm to $rm + 1$. This is illustrated in Figs. 1(a) and 1(b). A bond connecting $j = 1$ to $j = N$ gives the incorrect sign when a long-range vertical bond is present, and the correct sign when long-range vertical bonds are absent. This is because of the factor $(1 - x\sigma_N^{(1)}\sigma_{N-1}^{(1)})$ in the second term of Eq. (33). Figures 1(c) and 1(d) illustrate these two cases.

4. CONCLUSION

In this paper it has been shown algebraically how the Pfaffian method corresponds with the Onsager method for the rectangular two-dimensional lattice. In all cases in which the two methods do not exactly correspond it has been shown how the failure of the Pfaffian method is due to the appearance of crossed bonds, and that the nonlinearity associated with long-range bonds does not prevent a solution so long as there are no crossing points. In other words the correct nonlinear Onsager formulas can be replaced by the linear Pfaffian formulas without altering the expression for the partition function so long as the matrix U takes only its eigenvalue $+1$ and this will be so if there are no crossed bonds.

In later publications it is hoped to show that this is a general feature of the correspondence between the two methods, and in this way to understand better the limitations of the Pfaffian method. Once these limitations are understood the possibilities of the Pfaffian method as the basis for a solution by successive approximations will be clearer.

ACKNOWLEDGMENTS

This work was commenced while I was a visitor at the Department of Mathematics, University of Toronto, and I would like to thank the members of that department and of the Department of Physics for their hospitality. I would particularly like to thank Professor A. F. Pillow for his stimulating interest.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Improved Self-Consistent Configuration Interaction

SERAFIN FRAGA
Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada
(Received 31 July 1964)

Using general coupling operators the configuration interaction method is reformulated. The new formulation proves that the configuration interaction procedure leads to self-consistent results for the set of configurations chosen and makes it possible to include a self-consistent contribution of the continuum in practical calculations.

INTRODUCTION

THE practical impossibility of solving analytically the electronic Schrödinger equation has imposed the use of approximate methods. The variational procedure, within the self-consistent field (SCF) formalism, and using wavefunctions approximated by antisymmetrized products of one-electron functions, has provided the best way of tackling the problem in actual applications.[1] Unfortunately the approximation used for the functions (with omission of interelectron-distance terms) determines that the lowest energy attainable will be the Hartree–Fock energy for the Hamiltonian used. Configuration interaction (CI) using SCF wavefunctions (or virtual SCF functions) leads to lower energies but the improvement is restricted by the size of the basis set of interacting configurations chosen. To date it has not been possible to overcome this restriction as the continuum contribution has systematically been omitted.

The method developed here permits one to carry out a close-to-exact, close-to-complete,[2] self-consistent, direct configuration interaction treatment, with the basis set of interacting configurations being improved self-consistently towards a close-to-exact

[1] See, e.g., F. W. Birss and S. Fraga, J. Chem. Phys. 38, 2552 (1963): SCF treatment for lowest states. S. Fraga and F. W. Birss, *ibid.* 40, 3207 (1964): SCF treatment for excited states. Pertinent references can be found in both papers.

[2] The term "complete" is used throughout this paper to denote a treatment using a complete basis set of configurations.

bond connecting the lattice point rm to $rm + 1$. This is illustrated in Figs. 1(a) and 1(b). A bond connecting $j = 1$ to $j = N$ gives the incorrect sign when a long-range vertical bond is present, and the correct sign when long-range vertical bonds are absent. This is because of the factor $(1 - x\sigma_N^{(1)}\sigma_{N-1}^{(1)})$ in the second term of Eq. (33). Figures 1(c) and 1(d) illustrate these two cases.

4. CONCLUSION

In this paper it has been shown algebraically how the Pfaffian method corresponds with the Onsager method for the rectangular two-dimensional lattice. In all cases in which the two methods do not exactly correspond it has been shown how the failure of the Pfaffian method is due to the appearance of crossed bonds, and that the nonlinearity associated with long-range bonds does not prevent a solution so long as there are no crossing points. In other words the correct nonlinear Onsager formulas can be replaced by the linear Pfaffian formulas without altering the expression for the partition function so long as the matrix U takes only its eigenvalue $+1$ and this will be so if there are no crossed bonds.

In later publications it is hoped to show that this is a general feature of the correspondence between the two methods, and in this way to understand better the limitations of the Pfaffian method. Once these limitations are understood the possibilities of the Pfaffian method as the basis for a solution by successive approximations will be clearer.

ACKNOWLEDGMENTS

This work was commenced while I was a visitor at the Department of Mathematics, University of Toronto, and I would like to thank the members of that department and of the Department of Physics for their hospitality. I would particularly like to thank Professor A. F. Pillow for his stimulating interest.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Improved Self-Consistent Configuration Interaction

SERAFIN FRAGA
Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada
(Received 31 July 1964)

Using general coupling operators the configuration interaction method is reformulated. The new formulation proves that the configuration interaction procedure leads to self-consistent results for the set of configurations chosen and makes it possible to include a self-consistent contribution of the continuum in practical calculations.

INTRODUCTION

THE practical impossibility of solving analytically the electronic Schrödinger equation has imposed the use of approximate methods. The variational procedure, within the self-consistent field (SCF) formalism, and using wavefunctions approximated by antisymmetrized products of one-electron functions, has provided the best way of tackling the problem in actual applications.[1] Unfortunately the approximation used for the functions (with omission of interelectron-distance terms) determines that the lowest energy attainable will be the Hartree–Fock energy for the Hamiltonian used. Configuration interaction (CI) using SCF wavefunctions (or virtual SCF functions) leads to lower energies but the improvement is restricted by the size of the basis set of interacting configurations chosen. To date it has not been possible to overcome this restriction as the continuum contribution has systematically been omitted.

The method developed here permits one to carry out a close-to-exact, close-to-complete,[2] self-consistent, direct configuration interaction treatment, with the basis set of interacting configurations being improved self-consistently towards a close-to-exact

[1] See, e.g., F. W. Birss and S. Fraga, J. Chem. Phys. 38, 2552 (1963): SCF treatment for lowest states. S. Fraga and F. W. Birss, *ibid.* 40, 3207 (1964): SCF treatment for excited states. Pertinent references can be found in both papers.

[2] The term "complete" is used throughout this paper to denote a treatment using a complete basis set of configurations.

lowest state. A general discussion for exact solutions is also given.

THEORETICAL DEVELOPMENT

The time- and spin-independent, electronic Hamiltonian for an N-electron system is defined by

$$\mathcal{H} = H + R,$$

with

$$H = \sum_{\rho} H_{\rho},$$

$$R = \tfrac{1}{2} \sum_{\rho} \sum_{\tau \neq \rho} (1/r_{\rho\tau}),$$

where the summations are taken over all electrons ρ, τ. H_{ρ} is the one-electron operator formed by kinetic and nuclear potential components; $r_{\rho\tau}$ is the distance between electrons ρ and τ.

The corresponding Schrödinger equation may be written as

$$\mathcal{H}\Psi^N_{a\alpha} \equiv (H + R)\Psi^N_{a\alpha} = E^N_a \Psi^N_{a\alpha}, \tag{1}$$

where the subscript α distinguishes the degenerate eigenfunctions belonging to the same eigenvalue E^N_a; the superscript N stands for the number of electrons.

The CI treatments reduce, essentially, to the selection of a basis set of interacting configurations, the expansion of the wavefunctions $\Psi^N_{k\kappa}$ in terms of these basis functions, the evaluation of the matrix $\mathcal{H}$ in terms of these functions, and the solution of the matrix equation

$$\mathcal{H}\mathbf{c}^N_{k\kappa} = E^N_k \mathbf{S}\mathbf{c}^N_{k\kappa}, \tag{2}$$

where the column vector $\mathbf{c}^N_{k\kappa}$ is formed by the expansion coefficients of $\Psi^N_{k\kappa}$ and $\mathbf{S}$ represents the total overlap matrix defined in terms of the basis functions.

There are two questions which merit some inspection within this formalism: the choice of the basis set of interacting configurations and the self consistency of the results.

Selection of the Basis Set of Configurations

In a limited CI treatment it is customary to expand the functions $\Psi^N_{k\kappa}$ in terms of a restricted basis set of (orthonormal) approximate eigenfunctions of the discrete spectrum of the problem under consideration, i.e.,

$$\Psi^N_{k\kappa} = \sum_{l} \sum_{\lambda} c_{l\lambda,k\kappa} \Phi^N_{l\lambda}.$$

In a complete CI treatment one must expand the functions in terms of a complete basis set of approximate functions. One might write[3,4]

$$\Psi^N_{k\kappa} = \sum_{l} \sum_{\lambda} c_{l\lambda,k\kappa} \Phi^N_{l\lambda} + \sum_{\nu} \int_{\epsilon}^{\infty} c_{\nu,k\kappa}(E) \Phi^N_{\nu}(E)\, dE, \tag{3}$$

where $\Phi^N_{\nu}(E)$ is an approximate eigenfunction of the continuum for the problem under consideration, with the members of a degenerate group labeled by the subscript ν; ϵ is the energy of the ground state of the singly-charged, positive ion. This expansion will only be exact if the functions $\Phi^N_{l\lambda}$, $\Phi^N_{\nu}(E)$ form a complete set of class D functions. One cannot know *a priori* whether such functions form or not a complete set, but it is clear that they are not class D functions with respect to the Hamiltonian under consideration; if they were, the problem would already be solved.

This difficulty can be overcome in the following way. The set of exact functions $\Psi^N_{l\lambda}$, $\Psi^N_{\nu}(E)$ is a complete set of class D eigenfunctions for the N-electron problem. The approximate functions $\Phi^N_{\lambda l}$ (as we have mentioned above) are not class D eigenfunctions, but it is completely legitimate (see the Appendix for a justification) to write

$$\Phi^N_{k\kappa} = \sum_{l} \sum_{\lambda} d_{l\lambda,k\kappa} \Psi^N_{l\lambda} + \sum_{\nu} \int_{\epsilon}^{\infty} d_{\nu,k\kappa}(E) \Psi^N_{\nu}(E)\, dE$$

which can be rewritten as[3]

$$\Phi^N_{k\kappa} = \sum_{l} \sum_{\lambda} d_{l\lambda,k\kappa} \Psi^N_{l\lambda} + \sum_{m} \sum_{\mu} \Psi^{N-1}_{m\mu} \int_{0}^{\infty} d_{m\mu,k\kappa}(E) \Psi'(E)\, dE, \tag{4}$$

where $\Psi'(E)$ is an appropriate free-electron function and $\Psi^{N-1}_{m\mu}$ represents an exact, discrete eigenfunction of the $(N-1)$-electron Hamiltonian. The summations over m, μ extend over the complete discrete spectrum of eigenfunctions of the singly-charged, positive ion; for this reason the energy-integration limits have now been changed to 0 and ∞. The degeneracy indicated by ν is now taken care of by the summations over m, μ; furthermore for $\Psi'(E)$ a linear combination of the two possible independent (degenerate) functions must be taken.

The set of equations represented by Eq. (4) may be solved for the functions $\Psi^N_{l\lambda}$. One can write, in general,

---

[3] See any textbook for further details. In particular, E. C. Kemble, *Fundamental Principles of Quantum Mechanics*, (Dover Publications, Inc., New York, 1958), Chap. VI (especially pp. 211, 212, 215 ff.).

[4] Hereafter, except when reference is made to limited CI treatments, the functions Φ^N are supposed to be approximate functions for all the discrete states of the problem under consideration.

$$\Psi_{k\kappa}^N = \sum_l \sum_\lambda c_{l\lambda,k\kappa}\Phi_{l\lambda}^N + \sum_m \sum_\mu \Psi_{m\mu}^{N-1}\int_0^\infty p_{m\mu,k\kappa}(E)\Psi'(E)\,dE, \quad (5)$$

which is the expansion to be used.

The problem which arises because of the inclusion of the continuum may be solved by the following argument. Let us assume that the expansion for the exact wavefunction $\Psi_{a\alpha}^N$ for the lowest state has been determined. One can then write

$$\langle\Psi_{a\alpha}^N\,|\mathcal{H}|\,\Psi_{a\alpha}^N\rangle = E_a^N$$

(assuming $\Psi_{a\alpha}^N$ to be normalized); i.e., the numerical value E_a^N can be found even if its evaluation, as given by the preceding equation, implies operations with the infinite integrals $\int_0^\infty p_{m\mu,k\kappa}(E)\Psi'(E)dE$. Therefore, these infinite integrals must be convergent, and one can, in principle, replace them by a function $\Omega_{m\mu}$, function only of the positional coordinates of the electron involved. One can then rewrite Eq. (5), in general, as

$$\Psi_{k\kappa}^N = \sum_l \sum_\lambda c_{l\lambda,k\kappa}\Phi_{l\lambda}^N + \sum_m \sum_\mu c_{m\mu,k\kappa}\Psi_{m\mu}^{N-1}\Omega_{m\mu}.$$

A general formula for $\Omega_{m\mu}$ cannot be given at this moment. One can see from Eq. (5) that

$$\left\langle\Psi_{n\nu}^{N-1}\int_{E'}^{E'+\eta}\Psi'(E)\,dE\mid\Psi_{k\kappa}^N\right\rangle = 0$$

$$= \sum_l\sum_\lambda c_{l\lambda,k\kappa}\left\langle\Psi_{n\nu}^{N-1}\int_{E'}^{E'+\eta}\Psi'(E)\,dE\mid\Phi_{l\lambda}^N\right\rangle + \sum_m\sum_\mu\left\langle\Psi_{n\nu}^{N-1}\int_{E'}^{E'+\eta}\Psi'(E)\,dE\mid\Psi_{m\mu}^{N-1}\times\int_0^\infty p_{m\mu,k\kappa}(E)\Psi'(E)\,dE\right\rangle$$

and therefore

$$p_{m\mu,k\kappa}(E') = -\lim_{\eta\to0}\eta^{-1}\sum_l\sum_\lambda c_{l\lambda,k\kappa}\times\left\langle\Psi_{n\nu}^{N-1}\int_{E'}^{E'+\eta}\Psi'(E)\,dE\mid\Phi_{l\lambda}^N\right\rangle$$

which implies that one will be concerned with the infinite integrals

$$\int_0^\infty\left\{-\lim_{\eta\to0}\eta^{-1}\sum_l\sum_\lambda c_{l\lambda,k\kappa}\times\left\langle\Psi_{n\nu}^{N-1}\int_{E'}^{E'+\eta}\Psi'(E)\,dE\mid\Phi_{l\lambda}^N\right\rangle\right\}\Psi'(E')\,dE'$$

to be determined in each case depending on the basis set of functions $\Phi_{l\lambda}^N$.

The functions $\Psi_{m\mu}^{N-1}\Omega_{m\mu}$ will be assumed to satisfy the conditions

$$\langle\Psi_{m\mu}^{N-1}\Omega_{m\mu}\mid\Psi_{l\lambda}^N\rangle = 0,$$
$$\langle\Psi_{m\mu}^{N-1}\Omega_{m\mu}\mid\Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle = \delta_{mn}\delta_{\mu\nu}.$$

Self-Consistency of CI Treatments

Taking into account the results of the preceding section one can write (see the Appendix)

$$R\Psi_{a\alpha}^N = \sum_b\sum_\beta\theta_{b\beta,a\alpha}\Psi_{b\beta}^N + \sum_\delta\int_\epsilon^\infty\xi_{\delta,a\alpha}(E)\Psi_\delta^N(E)\,dE = \sum_b\sum_\beta\theta_{b\beta,a\alpha}\Psi_{b\beta}^N + \sum_c\sum_\gamma\zeta_{c\gamma,a\alpha}\Psi_{c\gamma}^{N-1}\Omega_{c\gamma}.$$

Substituting into Eq. (1) we obtain

$$H\Psi_{a\alpha}^N = E_a^N\Psi_{a\alpha}^N - \sum_b\sum_\beta\theta_{b\beta,a\alpha}\Psi_{b\beta}^N - \sum_c\sum_\gamma\zeta_{c\gamma,a\alpha}\Psi_{c\gamma}^{N-1}\Omega_{c\gamma}. \quad (6)$$

The coefficients $\theta_{b\beta,a\alpha}$ and $\zeta_{c\gamma,a\alpha}$ being given by

$$\theta_{a\alpha,a\alpha} = E_a^N - \langle\Psi_{a\alpha}^N|\,H\,|\Psi_{a\alpha}^N\rangle,$$
$$\theta_{b\beta,a\alpha} = -\langle\Psi_{b\beta}^N|\,H\,|\Psi_{a\alpha}^N\rangle,$$
$$\zeta_{c\gamma,a\alpha} = -\langle\Psi_{c\gamma}^{N-1}\Omega_{c\gamma}|\,H\,|\Psi_{a\alpha}^N\rangle,$$

one can rewrite Eq. (6) as

$$(H\mid\Psi_{a\alpha}^N\rangle - \sum_b\sum_\beta|\Psi_{b\beta}^N\rangle\langle\Psi_{b\beta}^N|\,H\,|\Psi_{a\alpha}^N\rangle - \sum_c\sum_\gamma|\Psi_{c\gamma}^{N-1}\Omega_{c\gamma}\rangle\langle\Psi_{c\gamma}^{N-1}\Omega_{c\gamma}|\,H\,|\Psi_{a\alpha}^N\rangle + |\Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N|\,\mathcal{H}\,|\Psi_{a\alpha}^N\rangle = |\Psi_{a\alpha}^N\rangle E_a^N.$$

Defining a new operator F by

$$F = \sum_k\sum_\kappa|\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N|\,H) - \sum_n\sum_\nu|\Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle\langle\Psi_{n\nu}^{N-1}\Omega_{n\nu}\mid H) - 2\sum_k\sum_l\sum_\kappa\sum_\lambda|\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N|\,H\,|\Psi_{l\lambda}^N\rangle\langle\Psi_{l\lambda}^N| + \sum_k\sum_\kappa|\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N|\,\mathcal{H}\,|\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| - \sum_n\sum_\nu(H\mid\Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle\langle\Psi_{n\nu}^{N-1}\Omega_{n\nu}| + \sum_k\sum_\kappa(H\mid\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N|,$$

it can be seen that

$$(F\mid\Psi_{a\alpha}^N\rangle = |\Psi_{a\alpha}^N\rangle E_a^N. \quad (7)$$

This equation is then equivalent, though not identical, to Eq. (1). The equivalence should be understood in the sense that both equations have the same discrete solutions. This equation can be

best solved for the lowest state by a variational treatment, with the SCF formalism.[5]

The expectation value for the energy is given by

$$E_a^N = \frac{\langle \Psi_{a\alpha}^N | F | \Psi_{a\alpha}^N \rangle}{\langle \Psi_{a\alpha}^N | \Psi_{a\alpha}^N \rangle}$$

where $\Psi_{a\alpha}^N$ is not assumed to be normalized.[6]

Giving an infinitesimal variation to the function $\Psi_{a\alpha}^N$ (but keeping everything else constant) we obtain

$$\langle \Psi_{a\alpha}^N | \Psi_{a\alpha}^N \rangle \delta E_a^N + E_a^N \delta \langle \Psi_{a\alpha}^N | \Psi_{a\alpha}^N \rangle = \delta \langle \Psi_{a\alpha}^N | F | \Psi_{a\alpha}^N \rangle. \quad (8)$$

The orthonormalization conditions for all the wavefunctions must be satisfied at any time. The total constraint to be introduced in the expression for the variation of the energy is

$$\sum_k \sum_\kappa \chi_{k\kappa,a\alpha} \langle \delta\Psi_{a\alpha}^N | \Psi_{k\kappa}^N \rangle + \sum_k \sum_\kappa \chi_{a\alpha,k\kappa} \langle \Psi_{k\kappa}^N | \delta\Psi_{a\alpha}^N \rangle + \sum_n \sum_\nu \omega_{n\nu,a\alpha} \langle \delta\Psi_{a\alpha}^N | \Psi_{n\nu}^{N-1}\Omega_{n\nu} \rangle + \sum_n \sum_\nu \omega_{a\alpha,n\nu} \langle \Psi_{n\nu}^{N-1}\Omega_{n\nu} | \delta\Psi_{a\alpha}^N \rangle = 0. \quad (9)$$

We add Eq. (9) to Eq. (8) and set δE to zero. The resulting equation can be separated, after some manipulation, into two equations, one for $\delta\Psi_{a\alpha}^N$ and the other for $\delta\bar{\Psi}_{a\alpha}^N$, which are complex conjugates of each other. The first one can be written as

$$\langle \delta\Psi_{a\alpha}^N | G^{a\alpha}\Psi_{a\alpha}^N - E_a^N\Psi_{a\alpha}^N - \sum_k \sum_\kappa \chi_{k\kappa,a\alpha}\Psi_{k\kappa}^N - \sum_n \sum_\nu \omega_{n\nu,a\alpha}\Psi_{n\nu}^{N-1}\Omega_{n\nu} \rangle = 0, \quad (10)$$

with

$$G^{a\alpha} = \sum_k \sum_\kappa |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| H) - |\Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N| H) - \sum_n \sum_\nu |\Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle\langle\Psi_{n\nu}^{N-1}\Omega_{n\nu}| H) - \sum_k \sum_l \sum_\kappa \sum_\lambda |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| H |\Psi_{l\lambda}^N\rangle\langle\Psi_{l\lambda}^N| + \sum_k \sum_\kappa |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| \mathcal{K} |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| + |\Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N| \mathcal{K} |\Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N| - \sum_k \sum_l \sum_\kappa \sum_\lambda |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| H |\Psi_{l\lambda}^N\rangle\langle\Psi_{l\lambda}^N| - \sum_n \sum_\nu (H | \Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle\langle\Psi_{n\nu}^{N-1}\Omega_{n\nu}| - (H | \Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N| + \sum_k \sum_\kappa (H | \Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N|.$$

In order that Eq. (10) will be satisfied it is necessary that the coefficients of $\delta\Psi_{a\alpha}^N$ in the integrand vanish. This leads to the equation

$$G^{a\alpha}\Psi_{a\alpha}^N = E_a^N\Psi_{a\alpha}^N + \sum_k \sum_\kappa \chi_{k\kappa,a\alpha}\Psi_{k\kappa}^N + \sum_n \sum_\nu \omega_{n\nu,a\alpha}\Psi_{n\nu}^{N-1}\Omega_{n\nu}, \quad (11)$$

where it can be seen that

$$\chi_{a\alpha,a\alpha} = \langle \Psi_{a\alpha}^N | G^{a\alpha} | \Psi_{a\alpha}^N \rangle - E_a^N,$$
$$\chi_{k\kappa,a\alpha} = \langle \Psi_{k\kappa}^N | G^{a\alpha} | \Psi_{a\alpha}^N \rangle,$$
$$\omega_{n\nu,a\alpha} = \langle \Psi_{n\nu}^{N-1}\Omega_{n\nu} | G^{a\alpha} | \Psi_{a\alpha}^N \rangle.$$

Therefore one can rewrite Eq. (11) as

$$(G^{a\alpha} | \Psi_{a\alpha}^N\rangle - \sum_k \sum_\kappa |\Psi_{k\kappa}^N\rangle\langle\Psi_{k\kappa}^N| G^{a\alpha} |\Psi_{a\alpha}^N\rangle - \sum_n \sum_\nu |\Psi_{n\nu}^{N-1}\Omega_{n\nu}\rangle\langle\Psi_{n\nu}^{N-1}\Omega_{n\nu}| G^{a\alpha} |\Psi_{a\alpha}^N\rangle + |\Psi_{a\alpha}^N\rangle\langle\Psi_{a\alpha}^N| \mathcal{K} |\Psi_{a\alpha}^N\rangle = |\Psi_{a\alpha}^N\rangle E_a^N.$$

Defining a new operator $J^{a\alpha}$ by an expression completely similar to that for F, but replacing H by $G^{a\alpha}$, one obtains

$$(J^{a\alpha} | \Psi_{a\alpha}^N\rangle = |\Psi_{a\alpha}^N\rangle E_a^N. \quad (12)$$

This eigenvalue equation could now be solved for the lowest state, in principle, in the manner common to all SCF procedures. One would assume a set of trial functions, evaluate the operator $J^{a\alpha}$ and solve Eq. (12) for an improved form of the functions. These new functions are used in repetition of the process until self-consistency has been reached.

The solution of Eq. (12), as indicated above, is a close-to-impossible task. Such an obstacle may be overcome in numerical calculations by expansion of the wavefunctions in terms of a basis set of approximate eigenfunctions (LCAE).

In a first approach (limited LCAE approximation) one can use a restricted basis set of discrete, approximate eigenfunctions, i.e.,

$$\Psi_{k\kappa}^N = \sum_l \sum_\lambda c_{l\lambda,k\kappa}\Phi_{l\lambda}^N = \mathbf{\Phi c}_{k\kappa},$$

where $\mathbf{\Phi}$ is a row vector and $\mathbf{c}_{k\kappa}$ is a column vector; the ordering of the elements is done on the basis of the subindices $l\lambda$. In a consistent treatment $\Psi_{n\nu}^{N-1}\Omega_{n\nu}$ should then be expanded in terms of a basis set of approximate eigenfunctions of the continuum, consistent with (and therefore orthogonal to) $\Phi_{l\lambda}^N$. In such a case the continuum contribution within the operator vanishes.

The corresponding matrix equation to be solved is

$$\mathbf{J}^{a\alpha}\mathbf{c}_{a\alpha} = E_a^N\mathbf{S}\mathbf{c}_{a\alpha}, \quad (13)$$

where

$$\mathbf{J}^{a\alpha} = \mathbf{S(D - D}^{a\alpha}\mathbf{)H - S(D - D}^{a\alpha}\mathbf{)HDS + SAS - SDH(D - D}^{a\alpha}\mathbf{)S + H(D - D}^{a\alpha}\mathbf{)S},$$

[5] Only the general description will be given here. For more details, see Ref. 1.

[6] While the variational treatment is discussed, the functions Ψ^N are not the exact functions, but the same notation is maintained, for at self-consistency, the exact eigenfunction for the lowest state will be obtained.

$$\mathbf{D} = \sum_k \sum_\kappa \mathbf{D}^{k\kappa}, \qquad \mathbf{D}^{k\kappa} = \mathbf{c}_{k\kappa}\mathbf{c}_{k\kappa}^*,$$

$$\mathbf{A} = \sum_k \sum_\kappa \mathbf{D}^{k\kappa}\mathfrak{H}\mathbf{D}^{k\kappa}, \quad M_{k\kappa,l\lambda} = \langle \Phi_{k\kappa}^N | M | \Phi_{l\lambda}^N \rangle,$$

with M standing for H, $\mathfrak{H}$, or S.

The solution of Eq. (13) is carried out in the conventional way by trial and error until the desired self-consistency has been reached. It can be seen that at self-consistency the eigenvectors are precisely those which diagonalize the matrix $\mathfrak{H}$. In other words, the normal CI procedure provides a self-consistent treatment within the approximation chosen (use of a limited basis set of approximate functions).

In a complete LCAE treatment we must expand the functions in terms of a complete basis set of approximate functions. One can use (see before) the expansion

$$\Psi_{k\kappa}^N = \sum_l \sum_\lambda c_{l\lambda,k\kappa}\Phi_{l\lambda}^N + \sum_m \sum_\mu c_{m\mu,k\kappa}\Psi_{m\mu}^{N-1}\Omega_{m\mu} = \mathbf{\Phi}\mathbf{c}_{k\kappa}, \qquad (14)$$

where $\mathbf{\Phi}$ is a row vector formed by the functions Φ^N, $\Psi^{N-1}\Omega$, and $\mathbf{c}_{k\kappa}$ is a column vector.

The matrix equation to be solved is

$$\mathbf{J}^{a\alpha}\mathbf{c}_{a\alpha} = E_a^N\mathbf{S}\mathbf{c}_{a\alpha},$$

with

$$\mathbf{J}^{a\alpha} = (\mathbf{X}^{a\alpha} + \mathbf{Y}^{a\alpha}) + (\mathbf{X}^{a\alpha} + \mathbf{Y}^{a\alpha})^\dagger,$$

$$\mathbf{X}^{a\alpha} = \mathbf{S}(\mathbf{D} - \mathbf{D}^{a\alpha})\mathbf{H} - \mathbf{S}(\mathbf{D} - \mathbf{D}^{a\alpha})\mathbf{H}\mathbf{D}\mathbf{S} + \tfrac{1}{2}\mathbf{S}\mathbf{A}\mathbf{S},$$

$$\mathbf{Y}^{a\alpha} = \mathbf{S}^0\mathbf{H}^{0\dagger} - \mathbf{S}(2\mathbf{D} - \mathbf{D}^{a\alpha})\mathbf{H}^0\mathbf{S}^{0\dagger} + \mathbf{S}^0\mathbf{H}^a\mathbf{S}^{0\dagger}.$$

The adjacent diagrams give the structure of the matrices.

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}(\Phi^N, \Phi^N) & \mathbf{M}(\Phi^N, \Psi^{N-1}\Omega) \\ \mathbf{M}(\Psi^{N-1}\Omega, \Phi^N) & \mathbf{M}(\Psi^{N-1}\Omega, \Psi^{N-1}\Omega) \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2^\dagger & \mathbf{M}_3 \end{bmatrix}$$

$$\mathbf{M}^0 = \begin{bmatrix} \mathbf{M}_2 \\ \mathbf{M}_3 \end{bmatrix}, \qquad \mathbf{M}^\omega = \mathbf{M}_3.$$

The solution of this matrix equation would be carried out in the normal way by successive iterations, until the desired self-consistency would have been reached. It can be seen that again in this case, as for the limited LCAE treatment, at self-consistency the solutions are precisely those which diagonalize the matrix $\mathfrak{H}$. That is, a complete CI treatment would also lead to self-consistent results.

SCF EXACT AND CLOSE-TO-EXACT SOLUTION

The SCF exact solution of the electronic Schrödinger equation can therefore be obtained within the formalism of direct configuration interaction, knowing that the solution of Eq. (2) leads to self-consistent results within the basis set chosen. The expansions to be used are given by Eq. (14), which implies that the matrix $\mathfrak{H}$ is finite, its dimension being equal to the number of discrete eigenstates for the N-electron problem plus the number of discrete eigenstates for the $(N-1)$-electron problem.

The practical procedure to be followed can be summarized in this way. The functions Φ^M are determined.[7] The exact functions Ω are determined (see before) or the approximate functions Ω' are arbitrarily chosen. Assuming that the functions Ψ^{M-1} are known, the set of functions $(\Psi^{M-1}\Omega)_0$ or $(\Psi^{M-1}\Omega')_0$ is formed. In this second case the matrix $\mathfrak{H}^\omega$, evaluated with the basis set of functions $(\Psi^{M-1}\Omega')_0$, is first diagonalized; the functions obtained in this way are labeled $(\Psi^{M-1}\Omega)_0$. Though they are just an approximation to the exact functions, the same notation is used in order to simplify the following discussion.

Now the matrix $\mathfrak{H}$ is evaluated using the basis set of functions Φ^M, $(\Psi^{M-1}\Omega)_0$. Its diagonalization will lead to the self-consistent solutions Ψ_1^M, $(\Psi^{M-1}\Omega)_1$. The process is repeated using the functions Ψ_1^M, $(\Psi^{M-1}\Omega)_0$ and the functions Ψ_2^M, $(\Psi^{M-1}\Omega)_2$ are obtained. The procedure is repeated until the desired self-consistency for the basis set has been reached.[8] This self-consistency can be taken to have been reached when $(\Psi^{M-1}\Omega)_n \equiv (\Psi^{M-1}\Omega)_0$, within the limits imposed.

The solution found for the lowest state, when using the exact functions Ω will approximate the corresponding exact solution of the Schrödinger equation within the precision limits determined by the self-consistency conditions.

It must be pointed out that for the solution of the M-electron problem it is necessary to know the functions Ψ^{M-1}. This means that the $(M-1)$-electron problem (for the same skeleton of nuclei as corresponding to the system under consideration) must have been solved previously. But the solution of this problem, to be carried out in a similar way as for the M-electron problem, implies the knowledge of the functions Ψ^{M-2}, etc., etc. Therefore an aufbau process must be followed, with the calculations being carried out successively for $M = 1, 2, 3, \cdots, N-1, N$.

The preceding formulation and discussion applies

[7] The functions Φ^M can be determined by any method, but it would be expected that SCF functions would insure a faster convergency. In this connection see Ref. 1.

[8] Each iteration produces self-consistent solutions for the basis set used; the final solutions correspond furthermore to a self-consistent basis set. The subscripts used in this section label the iterations.

to those cases where the number of discrete eigenvalues is finite; that is, for systems for which the neutral and all the positively-ionized states have only a finite series of discrete negative eigenvalues, the formulation can be applied as presented here and should lead (using the correct functions Ω) to the exact solution.

But there are other cases of more practical interest, namely, when there is an infinite series of discrete, negative eigenvalues.[9] In such cases the infinite series of discrete, negative eigenvalues must be arbitrarily subdivided into two sets, one including a finite number of normal, discrete levels, the other including the vast majority of the terms of the infinite series. This second set can be replaced by a function Υ, which can be defined by an argument similar to the one given for the functions Ω.

PRACTICAL CALCULATIONS

The determination of the functions Ω and Υ is a close-to-impossible task. But in any case the present formulation offers certain definite advantages with respect to the classical CI procedures.

In normal CI treatments there are two deficiencies which cannot be corrected. When proceeding within the framework of the LCAO MO method, the choice of (relatively) small basis sets (because of practical considerations) restricts the number of possible configurations which can be used. But even not all of these configurations (and the situation is worse when larger basis sets are used and higher-lying configurations are taken into account) have a physical meaning. Most of them (and this is common experience in the virtual MO approximation) lie above the groundstate of the theoretically-determined positive ion.

In the present formulation this situation is changed. Using the same basis set one can have a much larger number of physically correct configurations, spanning the complete range of negative energies, from the groundstate of the neutral system to zero. Of all the possible configurations, one uses, e.g., for the neutral system, the ones corresponding to energies lower than the energy of the groundstate of the singly-charged, positive ion; for this system one takes only those configurations with energies below that of the groundstate of the doubly-charged, positive ion, and so on. In this way one has an extended set of interacting configurations.

Instead of the functions $\Phi^{N-1}\Omega$ one can use determinantal functions, properly antisymmetrized, with the ionized electron described by molecular orbitals that, with proper orbital exponents, are only of significance in those regions where interactions are small. A similar formulation is used in the cases when there is more than one ionized electron.[10]

The aufbau process can either be carried in a single-step or in a multiple-step fashion, depending on the practical convenience. In the one-step procedure the wavefunction to be used is given by

$$\Psi^N_{k\kappa} = \sum_n \sum_l \sum_\lambda c^n_{l\lambda,k\kappa}\Phi^n_{l\lambda},$$

(with n running from N to 1) where, as pointed out before, the antisymmetrized functions Φ^n are formed from bound orbitals and as many nonbound orbitals as given by $(N - n)$.

When large basis sets can be used in practice and the approximate wavefunctions for the excited states are determined within the framework of the SCF procedure, the subdivision of the series of discrete eigenvalues (mentioned above) can be carried out in a straightforward manner. The separation may be done at that excited level E_n for which the energy of the groundstate of the next ion lies within the interval $(E_n \pm \Delta^{\frac{1}{2}})$.[11]

APPENDIX

The expansion

$$\Psi = \sum_l \sum_\lambda c_{l\lambda}\Psi_{l\lambda} + \sum_\mu \int_e^\infty c_\mu(E)\Psi_\mu(E)\,dE$$

is valid for a function Ψ which is a quadratically integrable, class-D function.

This type of expansion has been used for the functions $\Phi^N_{k\kappa}$ and $R\Psi^N_{a\alpha}$. These functions are certainly quadratically integrable. In particular, the integral $\langle R\Psi^N_{a\alpha} \mid R\Psi^N_{a\alpha}\rangle$ appears in a modified variational treatment for excited states, and it is expected to have a finite value.[12]

Though for many-electron systems it is not possible to prove the general theorem of the approximation of arbitrary quadratically integrable functions by means of class-D functions, this theorem is generally assumed as a postulate, which we do here.

Therefore, for our purposes, it is enough to know that whether $\Phi^N_{k\kappa}$ and $R\Psi^N_{a\alpha}$ belong to class D or not, it is possible to define quadratically integrable functions, which belong to such class and which approximate them as closely as desired. The expansion, as given above, follows immediately.

[9] See Ref. 3, p. 215. Even in this case it is necessary to introduce the continuum contribution.

[10] Two basis sets of atomic orbitals will be needed: one for the bound electrons, the other for the ionized (nonbound) electrons.

[11] See the second paper of Ref. 1 for complete details on the determination of Δ.

[12] See the second paper of Ref. 1.

Covariant Higher-Spin Equations*

R. A. BERG
Lockheed Research Laboratories, Palo Alto, California
(Received 5 June 1964)

Covariant wave equations are derived for nonzero-mass particles of arbitrary spin, with wavefunctions that involve no redundant components. The Dirac equation is seen to be the first of these, for spin $\frac{1}{2}$; the Proca vector meson equations for spin 1 are the next set. The derivation is based on synthesizing higher-spin particles from particles of lower spin.

INTRODUCTION

THE problem of constructing covariant higher-spin equations has been extensively studied.[1,2] The construction of these equations usually involves redundant components. For example, the Duffin–Kemmer equation for $s = 1$, involves a ten-component wavefunction, and only three of these can be considered as independent, corresponding to spin 1. We give a general formulation for arbitrary spin of a covariant wave equation which does not involve any arbitrary components, i.e., no redundancy in the wavefunction.

Section I considers the Dirac equation in the form desirable for our considerations; Sec. II considers the vector meson and a covariant wave equation for this spin 1 case, which does not involve any redundant components. Section III demonstrates how the spin-1 case can be synthesized out of two spin-$\frac{1}{2}$ cases. Section IV considers the general case and derives the covariant wave equation for arbitrary spin. Section V considers the question of the meaning of the synthesis of higher-spin particles from combinations of lower-spin particles.

We use the word "wavefunction" in the sense of meaning a state vector, depending on the 4-momentum p^μ plus spin indices.

I. THE DIRAC EQUATION

We write the Dirac equation in momentum representation in a form convenient for our purposes:

$$E\psi = \rho_3(\mathfrak{d}\cdot\mathbf{p})\psi + m\rho_1\psi,$$

where

$$\psi = \begin{pmatrix}\varphi\\ \eta\end{pmatrix} \quad (\varphi,\ \eta \text{ transform as spinors}).$$

Thus

$$\begin{aligned} E\varphi &= (\mathfrak{d}\cdot\mathbf{p})\varphi + m\eta \\ E\eta &= -(\mathfrak{d}\cdot\mathbf{p})\eta + m\varphi. \end{aligned} \tag{1.1}$$

The usual transcription is [metric $g^{\mu\nu} = (1, -1, -1, -1)$]

$$\beta = \gamma^0 = \rho_1 \quad \boldsymbol{\alpha} = \beta\boldsymbol{\gamma} = \rho_3\mathfrak{d} \quad \boldsymbol{\gamma} = -i\rho_2\mathfrak{d} \quad \gamma_5 = -i\rho_3$$

and therefore

$$-i\gamma^\mu\,\partial\psi/\partial x^\mu + m\psi = 0.$$

In coordinate space

$$-ip_\mu \leftrightarrow \partial/\partial x^\mu.$$

The more usual representation for the γ matrices[3] is

$$\gamma_D^0 = \rho_3 \quad \boldsymbol{\gamma}_D = -i\rho_2\mathfrak{d}$$

$$\psi_D \equiv \begin{pmatrix}u\\ v\end{pmatrix} \equiv \frac{1}{\sqrt{2}}(1 + i\rho_2)\psi,$$

i.e.,

$$u \equiv (1/\sqrt{2})(\varphi + \eta); \quad v \equiv (1/\sqrt{2})(-\varphi + \eta).$$

The usual representation has the advantage that the nonrelativistic limit of the momentum space solution of the Dirac equation takes an especially simple form. However, the u, v do not transform as two component spinors under the Lorentz group; all four components are mixed.

The representation that we use transforms as two component spinors under the Lorentz group, since Eqs. (1.1) are covariant under the transformations:

Spatial rotation

$$\begin{aligned} &\mathbf{p} \to \mathbf{p} + \boldsymbol{\theta}\times\mathbf{p}, \quad E \to E \\ &\varphi \to \varphi - \tfrac{1}{2}i(\mathfrak{d}\cdot\boldsymbol{\theta})\varphi, \quad \eta \to \eta - \tfrac{1}{2}i(\mathfrak{d}\cdot\boldsymbol{\theta})\eta; \end{aligned} \tag{1.2}$$

Lorentz transformations

$$\begin{aligned} &p \to p - \boldsymbol{\theta}E, \quad E \to E - (\boldsymbol{\theta}\cdot\mathbf{p}), \\ &\varphi \to \varphi - \tfrac{1}{2}(\mathfrak{d}\cdot\boldsymbol{\theta})\varphi, \quad \eta \to \eta + \tfrac{1}{2}(\mathfrak{d}\cdot\boldsymbol{\theta})\eta; \end{aligned} \tag{1.3}$$

* This work was carried out as a part of the Lockheed Independent Research Program.

[1] H. Umezawa, *Quantum Field Theory* (North Holland Publishing Company, Amsterdam, 1956).

[2] H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945); **21**, 451 (1949). The philosophy of the methods used in this paper (synthesis of higher-spin particles from those of lower spin) is also related to the "méthode de fusion" of L. de Broglie, although the methods and explicit formalism are somewhat different. L. de Broglie, *Théorie Générale des Particules a Spin* (Gauthier-Villars, Paris, 1954).

[3] S. S. Schweber, *Introduction to Relativistic Quantum Field Theory* (Row-Peterson, Evanston, Illinois, 1961).

Space and time inversion

$$R_s:\ \varphi \to \lambda_s\eta \qquad R_t:\ \varphi \to -\lambda_t\eta$$
$$\eta \to \lambda_s\varphi \qquad \eta \to \lambda_t\varphi$$
$$(\lambda_s^4 = \lambda_t^4 = +1).$$

We see that φ and η are—separately—spinor representations of the proper Lorentz group.

The Foldy-Wouthuysen transformation is defined in our representation as follows:

$$\psi_{\mathrm{FW}} \equiv U^{-1}\psi \equiv \begin{pmatrix}\chi\\ \zeta\end{pmatrix} \qquad \omega \equiv (m^2 + p^2)^{\frac{1}{2}},$$

where

$$U \equiv (\omega)^{\frac{1}{2}} \exp(i\lambda\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}) \exp(-\tfrac{1}{4}i\pi\rho_2)$$
$$= 2(\omega + m)^{-\frac{1}{2}}[(\omega + m) + (\boldsymbol{\sigma}\cdot\mathbf{p}) - i\rho_2(\omega + m) - (\boldsymbol{\sigma}\cdot\mathbf{p})],$$

i.e.,

$$\varphi(p, E) = 2(\omega + m)^{-\frac{1}{2}}[(\omega + m + \boldsymbol{\sigma}\cdot\mathbf{p})\chi - (\omega + m - \boldsymbol{\sigma}\cdot\mathbf{p})\zeta].$$

Then we have the wave equation in the FW representation

$$E\psi_{\mathrm{FW}} = \rho_3\omega\psi_{\mathrm{FW}} \tag{1.4}$$

which shows that the wavefunctions χ and ζ have positive and negative energy respectively.

The wavefunctions χ and ζ transform according to the canonical representation[4] of the proper Lorentz group. The canonical representation is explicitly unitary, whereas the representation of the proper Lorentz group, Eq. (1.3) is not unitary. This is connected with the fact that the transformation U is not a unitary transformation, due to the presence of the factor $\omega^{\frac{1}{2}}$.

The transformation of the FW wavefunction is given by

Spatial rotation:

$$\mathbf{p} \to \mathbf{p} + \boldsymbol{\theta}\times\mathbf{p}, \quad E \to E,$$
$$\chi \to \chi - \tfrac{1}{2}i(\boldsymbol{\sigma}\cdot\boldsymbol{\theta})\chi, \quad \zeta \to \zeta - \tfrac{1}{2}i(\boldsymbol{\sigma}\cdot\boldsymbol{\theta})\zeta; \tag{1.5}$$

Lorentz transformation:

$$p \to p - \boldsymbol{\theta}E, \qquad E \to E - (\boldsymbol{\theta}\cdot\mathbf{p})$$
$$\chi \to \chi - \frac{i}{2}\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\theta}\times\mathbf{p}}{m + \omega}\right)\chi, \tag{1.6}$$
$$\zeta \to \zeta - \frac{i}{2}\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\theta}\times\mathbf{p}}{-m - \omega}\right)\zeta.$$

[4] L. L. Foldy, Phys. Rev. **102**, 568 (1956).

We see the distinction between the usual Dirac laws of transformation, Eq. (1.3), and the canonical transformation, Eq. (1.6). The Dirac-type transformation has the advantage that in going to the coordinate space representation, the transformations corresponding to Eq. (1.3) are local (since they do not depend on p):

Rotation

$$\varphi'(r', t') = [1 - \tfrac{1}{2}i(\boldsymbol{\sigma}\cdot\boldsymbol{\theta})]\varphi(r, t);$$

Lorentz transformation

$$\chi'(r', t') = \left(1 - \frac{i}{2}\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\theta}\times\mathbf{p}}{(m + i\partial_t)}\right)\chi(\mathbf{r}, t).$$

This is also the reason for the necessity of mixing positive and negative energy states in going from the FW wavefunction ψ_{FW} to the locally transforming Dirac wavefunction ψ. To form a local quantity, i.e., a delta function, one has to mix negative and positive energy states.

As we shall see, this state of affairs will hold for arbitrary spin: simple transformation laws will lead to complicated wave equations, whereas the simple wave equation of the form of Eq. (1.4) will require complicated, nonlocal transformation properties of the wavefunctions.

II. THE SPIN-1 FIELD

The equation for the vector meson field may be written (Proca field) in momentum space as follows: (E = energy operator $\neq \mathbf{E}$)

$$\mathbf{p}\cdot\mathbf{B} = 0, \qquad \mathbf{E} = -i\mathbf{p}\varphi + iE\mathbf{A},$$
$$\mathbf{p}\cdot\mathbf{E} = im^2\varphi, \qquad \mathbf{B} = i\mathbf{p}\times\mathbf{A}, \tag{2.1}$$

and therefore

$$\mathbf{p}\times\mathbf{B} = -E\mathbf{E} + im^2\mathbf{A},$$
$$\mathbf{p}\times\mathbf{E} = E\mathbf{B}.$$

We now consider the 3×3 $s = 1$, representation of the three-dimensional rotation group, R_3, defined by $(S_l)_{mn} \equiv -i\epsilon_{lmn}$.

Therefore

$$\sum_{n=1}^{n=3}(\mathbf{S}\cdot\boldsymbol{\theta})_{mn}X_n = i(\boldsymbol{\theta}\times\mathbf{X})_m.$$
$$\sum_{n=1}^{n=3}[(\mathbf{S}\cdot\boldsymbol{\theta})^2]_{mn}X_n = -[\boldsymbol{\theta}\times(\boldsymbol{\theta}\times\mathbf{X})]_m.$$

Now define the two vectors $\mathbf{Z}_{1,2} \equiv \mathbf{E} \pm i\mathbf{B}$. This pair now has transformation properties that are in exact analogy to our representation of the case $s = \frac{1}{2}$, Eqs. (1.2) and (1.3).

Rotation

$$\mathbf{p} \to \mathbf{p} + \boldsymbol{\theta} \times \mathbf{p}, \qquad E \to E,$$
$$\mathbf{Z}_1 \to \mathbf{Z}_1 - i(\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z}_1, \quad \mathbf{Z}_2 \to \mathbf{Z}_2 - i(\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z}_2; \tag{2.2}$$

Lorentz transformation

$$\mathbf{p} \to \mathbf{p} - \boldsymbol{\theta} E, \qquad E \to E - (\boldsymbol{\theta}\cdot\mathbf{p})$$
$$\mathbf{Z}_1 \to \mathbf{Z}_1 - (\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z}_1, \quad \mathbf{Z}_2 \to \mathbf{Z}_2 + (\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z}_2. \tag{2.3}$$

We now desire a wave equation analogous to the Dirac equation, for the spin-1 case. It must therefore involve only the wavefunctions $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and $\mathbf{p}$, E, with the matrices $\mathbf{S}$ in the combination $(\mathbf{S}\cdot\mathbf{p})$ and be covariant under the transformations, Eqs. (1.2) and (1.3).

Now

$$\mathbf{p}\times(\mathbf{p}\times\mathbf{E}) = E\mathbf{p}\times\mathbf{B},$$
$$\mathbf{p}\times(\mathbf{p}\times\mathbf{B}) = -E\mathbf{p}\times\mathbf{E} + im^2(-i\mathbf{B}), \tag{2.4}$$

and therefore

$$\mathbf{p}\times(\mathbf{p}\times\mathbf{Z}_1) = E[\mathbf{p}\times\mathbf{B} - i\mathbf{p}\times\mathbf{E}] + im^2\mathbf{B},$$
$$\mathbf{p}\times(\mathbf{p}\times\mathbf{Z}_2) = E[\mathbf{p}\times\mathbf{B} + i\mathbf{p}\times\mathbf{E}] - im^2\mathbf{B}, \tag{2.5}$$

which is equivalent to

$$\Sigma_1(p)\mathbf{Z} \equiv [2(\mathbf{S}\cdot\mathbf{p})^2 - 2\rho_3 E(\mathbf{S}\cdot\mathbf{p}) + m^2(1-\rho_1)]\mathbf{Z} = 0, \tag{2.6}$$

where

$$\mathbf{Z} \equiv \begin{pmatrix}\mathbf{Z}_1\\ \mathbf{Z}_2\end{pmatrix}.$$

This equation now contains the vector meson equations. Equation (2.4) yields $\mathbf{p}\cdot\mathbf{B} = 0$.

Now define the auxiliary quantities φ, A in terms of E, B:

$$m^2\varphi \equiv -i\mathbf{p}\cdot\mathbf{E}; \quad m^2\mathbf{A} \equiv -(i\mathbf{p}\times\mathbf{B} + E\mathbf{E}).$$

Then,

$$\mathbf{p}\times\mathbf{A} = -im^{-2}[\mathbf{p}\times(\mathbf{p}\times\mathbf{B}) + E\mathbf{p}\times\mathbf{E}] = -i\mathbf{B},$$
$$E\mathbf{A} = -im^{-2}[E\mathbf{p}\times\mathbf{B} + \omega^2\mathbf{E}] = -i\mathbf{E} + \mathbf{p}\varphi,$$
$$(\omega^2 \equiv m^2 + \mathbf{p}^2).$$

We may, therefore, regard Eqs. (2.5) or (2.6) as the natural counterpart to the Dirac equation for $s = 1$. The adjoint operator is defined as

$$\Sigma_1(p) \equiv 2(\mathbf{S}\cdot\mathbf{p})^2 + 2\rho_3 E(\mathbf{S}\cdot\mathbf{p}) + m^2(1+\rho_1),$$

which acting on Eq. (2.6) yields

$$\hat{\Sigma}_1(p)\Sigma(p)Z = (m^2 - E^2 + p^2)^2\mathbf{Z} = 0.$$

Since our equation, written in vector form, yields the covariant Proca equations, we know that our formalism is covariant. However, it is of interest in a later context, to prove the covariance of the formalism directly.

For the case of rotation, we have

$$\mathbf{p} \to \mathbf{p} + \boldsymbol{\theta} + \mathbf{p},$$
$$\mathbf{Z} \to \mathbf{Z} - i(\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z},$$

so we must have that the over-all change in the wave equation, Eq. (2.6), is zero under this transformation. Thus, we must have

$$2(\mathbf{Sp})(\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p}) + 2(\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p})(\mathbf{S}\cdot\mathbf{p}) - 2i(\mathbf{Sp})^2(\mathbf{S}\boldsymbol{\theta}) - 2\rho_3 E(\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p}) - 2i\rho_3 E(\mathbf{Sp})(\mathbf{S}\boldsymbol{\theta}) + im^2(\rho_1 - 1)(\mathbf{S}\boldsymbol{\theta}) = 0. \tag{2.7}$$

Now from the commutation relations of the representations of the radiation group, R_3, we have

$$[\mathbf{S}\cdot\mathbf{a}, \mathbf{S}\cdot\mathbf{b}] = i\mathbf{S}\cdot\mathbf{a}\times\mathbf{b}$$

and therefore

$$[(\mathbf{Sp})^2, \mathbf{S}\boldsymbol{\theta}] = i(\mathbf{Sp})(\mathbf{S}\cdot\mathbf{p}\times\boldsymbol{\theta}) + i(\mathbf{S}\cdot\mathbf{p}\times\boldsymbol{\theta})(\mathbf{S}\cdot\mathbf{p}).$$

Substituting this in Eq. (2.7), we see that the over-all change in the wave equation is zero, and the equation is covariant under spatial rotations.

For Lorentz transformations, we must prove that under

$$\mathbf{p} \to \mathbf{p} - \boldsymbol{\theta} E$$
$$E \to E - (\boldsymbol{\theta}\cdot\mathbf{p})$$
$$\mathbf{Z} \to \mathbf{Z} - \rho_3(\mathbf{S}\cdot\boldsymbol{\theta})\mathbf{Z},$$

Eq. (2.6) does not alter form. The procedure is analogous to proving covariance under spatial rotations, except that it is necessary to make use of specific properties of the 3×3 representations of R_3. (Note that in proving covariance of the equation for spatial rotations, no use was made of the properties of the S matrices, beyond the commutation relations which are valid for any spin.)

The necessary property, which holds only for the 3×3 representations, is

$$\{(\mathbf{Sp})^2, (\mathbf{S}\cdot\boldsymbol{\theta})\}_+ = \mathbf{p}^2(\mathbf{S}\cdot\boldsymbol{\theta}) + (\boldsymbol{\theta}\cdot\mathbf{p})(\mathbf{S}\cdot\mathbf{p}).$$

This is used in moving $(\mathbf{S}\cdot\boldsymbol{\theta})$ to the left of $(\mathbf{S}\cdot\mathbf{p})^2$ and then making use of the wave equation to eliminate $(\mathbf{S}\cdot\mathbf{p})^2$, one can easily show that the equation is covariant.

This is a characteristic which is later shown to be true of the general wave equation: covariance under spatial rotations does not need to make use of the

specific properties of the representation of spin s being used. Proof of covariance under Lorentz transformations, however, does require making use of the specific properties of the representation being used.

The Taketani–Sakata equation[5] for the spin-1 particle is related to the formulation given here as follows:

Let

$$\Psi \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{A} - im^{-1}\mathbf{E} \\ i\mathbf{A} - m^{-1}\mathbf{E} \end{bmatrix},$$

then

$$\left[i\rho_1 \left\{ -\frac{(\mathbf{S}\cdot\mathbf{p})^2}{m} + \frac{p^2}{2m} \right\} + \rho_3 \left\{ m + \frac{p^2}{2m} \right\} \right] \Psi = E\Psi.$$

The wavefunctions Ψ and $\mathbf{Z}$ are clearly related, and both obey wave equations of somewhat similar form. The transformation properties of Ψ under Lorentz transformations are, however, considerably more complicated than those of $\mathbf{Z}$.

Equation (2.6) when transformed to coordinate space becomes

$$\nabla \times (\nabla \times \mathbf{Z}) + \rho_3(\nabla \times \partial\mathbf{Z}/\partial t) - \tfrac{1}{2}m^2(1 - \rho_1)\mathbf{Z} = 0. \qquad (2.8)$$

Although Eq. (2.8) contains the space and time coordinates in an unsymmetrical manner, it is in fact covariant. This emphasizes the point that treating the space and time variables on a symmetrical footing (as one does in the usual heuristic derivation of the Dirac equation) is sufficient for covariance, but not necessary.

III. SYNTHESIS OF SPIN 1 FROM SPIN 1/2

We now consider the possibility of synthesizing the case $s = 1$ from the $s = \frac{1}{2}$ case. The observation that φ for $s = \frac{1}{2}$ and Z_1 for $s = 1$, both transform according to irreducible representations of the group R_3 indicates that we may try to obtain Z_1 by a Clebsch–Gordon addition of two φ functions, and likewise $\mathbf{Z}_2$ by adding two η functions. Thus, (summation over repeated indices)

$$Z_{1k} \equiv (\tfrac{1}{2}\tfrac{1}{2}1k \mid \tfrac{1}{2}m_1\tfrac{1}{2}m_2)\varphi_{m_1}\varphi_{m_2} \equiv (k \mid m_1m_2)\varphi_{m_1}\varphi_{m_2},$$

$$Z_{2k} \equiv (\tfrac{1}{2}\tfrac{1}{2}1k \mid \tfrac{1}{2}m_1\tfrac{1}{2}m_2)\eta_{m_1}\eta_{m_2} \equiv (k \mid m_1m_2)\eta_{m_1}\eta_{m_2}.$$

Clearly, if φ and η transform according to Eqs. (1.2) and (1.3) then Z_1 and Z_2 will transform according to Eqs. (2.2) and (2.3). The Clebsch–Gordon coefficients for $\frac{1}{2} + \frac{1}{2} = 1$ can be written in terms of the Pauli matrices as follows:

$$(\tfrac{1}{2}\tfrac{1}{2}1k \mid \tfrac{1}{2}m_1\tfrac{1}{2}m_2) = i(\sigma^2\sigma^k)_{m_1m_2}/\sqrt{2}.$$

So

$$\mathbf{Z}_1 = (i/\sqrt{2})\hat{\varphi}\partial\varphi, \qquad \mathbf{Z}_2 = (i/\sqrt{2})\hat{\eta}\partial\eta,$$

where the adjoint wavefunctions are defined as

$$\hat{\varphi} \equiv \tilde{\varphi}\sigma^2; \quad \text{i.e.,} \quad \hat{\varphi}_\alpha \equiv \sum_{\beta=1}^{2} \varphi_\beta(\sigma^2)_{\beta\alpha}.$$

We now try to determine the effect of the various powers of the 3×3 $(\mathbf{S}\cdot\mathbf{p})$ on the $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Since $(\mathbf{Sp})^3 = \mathbf{p}^2(\mathbf{Sp})$, we restrict ourselves to $(\mathbf{Sp})$ and $(\mathbf{Sp})^2$.

Thus,

$$\begin{aligned}
(\mathbf{S}\cdot\mathbf{p})_{kl}Z_{1l} &= \tfrac{1}{2}(k \mid \alpha_1\alpha_2)[(\partial\cdot\mathbf{p})_{\alpha_1\beta_1}1_{\alpha_1\beta_1} \\
&\qquad + 1_{\alpha_1\beta_1}(\partial\cdot\mathbf{p})_{\alpha_2\beta_2}]\varphi_{\beta_1}\varphi_{\beta_2} \\
&= \tfrac{1}{2}(k \mid \alpha_1\alpha_2)[(E\varphi - m\eta)_{\alpha_1}\varphi_{\alpha_2} \\
&\qquad + \varphi_{\alpha_1}(E\varphi - m\eta)_{\alpha_2}], \\
(\mathbf{SP})^2_{kl}Z_{1l} &= \tfrac{1}{2}(k \mid \alpha_1\alpha_2)[\mathbf{p}^2 1_{\alpha_1\beta_1}1_{\alpha_2\beta_2} \\
&\qquad + (\partial\cdot\mathbf{p})_{\alpha_1\beta_1}(\partial\cdot\mathbf{p})_{\alpha_2\beta_2}]\varphi_{\beta_1}\varphi_{\beta_2} \\
&= \tfrac{1}{2}(k \mid \alpha_1\alpha_2)[(\mathbf{p}^2 + E^2)\varphi_{\alpha_1}\varphi_{\alpha_2} + m^2\eta_{\alpha_1}\eta_{\alpha_2} \\
&\qquad - mE(\varphi_{\alpha_1}\eta_{\alpha_2} + \eta_{\alpha_1}\varphi_{\alpha_2})], \\
-E(\mathbf{SP})_{kl}Z_{1l} &= \tfrac{1}{2}(k \mid \alpha_1\alpha_2)[-E^2(\varphi_{\alpha_1}\varphi_{\alpha_2} + \varphi_{\alpha_1}\varphi_{\alpha_2}) \\
&\qquad + mE(\varphi_{\alpha_1}\eta_{\alpha_2} + \eta_{\alpha_1}\varphi_{\alpha_2})].
\end{aligned}$$

We obtain [carrying out the same procedure for Z_2 and eliminating the $(\eta\varphi + \varphi\eta)$ terms] exactly the spin 1 wave equation, Eq. (2.6). We have therefore synthesized $s = 1$ from two $s = \frac{1}{2}$ cases.

In terms of the vector field quantities, we have

$$\mathbf{E} = m\hat{\psi}\partial\psi, \qquad \mathbf{B} = -im\hat{\psi}\rho_3\partial\psi$$
$$\mathbf{A} = -i\hat{\psi}\rho_1\partial\psi, \quad \varphi = -\hat{\psi}\rho_2\psi.$$

This indicates how a synthesis of two spin-$\frac{1}{2}$ particles will yield the spin-1 case. This is a seeming paradox, in that we seem to be combining two mass-m particles, to yield a mass-m particle, whereas we might expect a particle of mass $2m$ to appear. The resolution of this seeming paradox is given in Sec. V.

In the next section, we consider how this synthesis can be carried through in general, and show that one can indeed obtain a covariant wave equation for arbitrary spin.

The FW-type transformation for $s = 1$ can be derived from the FW transformation for $s = \frac{1}{2}$. We define the quantities

$$W_{1k} \equiv (k \mid \alpha_1\alpha_2)\chi_{\alpha_1}\chi_{\alpha_2} \qquad \mathbf{W} \equiv \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix}, \qquad (3.1)$$
$$W_{2k} \equiv (k \mid \alpha_1\alpha_2)\zeta_{\alpha_1}\zeta_{\alpha_2}$$

[5] S. Sakata and M. Takatani, Proc. Phys. Math. Soc. (Japan) 22, 757 (1940).

which will transform according to the canonical representation of the Lorentz group:

Rotation

$$\mathbf{W} \to \mathbf{W} - i(\mathbf{S}\cdot\boldsymbol{\theta})W;$$

Lorentz transformation

$$\mathbf{W}_{1,2} \to \mathbf{W}_{1,2} - i[\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p}/(m+E)]\mathbf{W}_{1,2}.$$

The three-component wavefunctions $\mathbf{W}_1$ and $\mathbf{W}_2$ will obey the canonical wave equation

$$E\mathbf{W} = \rho_3\omega\mathbf{W}, \quad \text{where} \quad \omega = (m^2+\mathbf{p}^2)^{\frac{1}{2}}.$$

Expressing χ, ζ in terms of φ, η by means of the spin-$\frac{1}{2}$ FW transformation, and then expressing φ, η in terms of Z, we obtain the spin-1 FW transformation

$$\mathbf{W} = \frac{1}{2m}\left[\left(1-\frac{(\mathbf{Sp})^2}{\omega(m+\omega)}\right)\times(1+\rho_1)+\frac{(\mathbf{S}\cdot\mathbf{p})}{m}(\rho_1-1)+\frac{E\rho_3}{\omega}\left(1+\frac{(\mathbf{S}\cdot\mathbf{p})^2}{m(\omega+m)}\right)(1+\rho_1)\right]\mathbf{Z},$$

$$\mathbf{Z} = \frac{m}{2}\left[\left(1+\frac{(\mathbf{S}\cdot\mathbf{p})^2}{m(m+\omega)}\right)(1+\rho_1)+\frac{(\mathbf{S}\cdot\mathbf{p})}{m}(1-\rho_1)\right]\mathbf{W}.$$

IV. THE GENERAL SYNTHESIS

We are now in a position to determine the general synthesis of the higher spin covariant wave equation. We proceed in exact analogy to the $1 = \frac{1}{2}+\frac{1}{2}$ case, by combining k $s = \frac{1}{2}$ particles, to a total spin of $\frac{1}{2}k$.

We define (summing over repeated α and β),

$$Z_1^A \equiv (A \mid \alpha_1\cdots\alpha_k)\varphi_{\alpha_1}\cdots\varphi_{\alpha_k}$$

$$Z_2^A \equiv (A \mid \alpha_1\cdots\alpha_k)\eta_{\alpha_1}\cdots\eta_{\alpha_k},$$

where the generalized Clebsch–Gordon coefficients $(A \mid \alpha\cdots)$ express the addition of k $s = \frac{1}{2}$ particles to form a $2s+1$ entity which transforms according to the irreducible $(2s+1)$ by $(2s+1)$ representation of R_3. Note that the order of adding the particles together is important in the sense that two different orderings, define two different Z wavefunctions, which are connected by a unitary transformation. (For $k = 3$, they are connected by the $6j$ symbols.)

The Z will now transform according to the $(2s+1)$ by $(2s+1)$ matrices:

Rotation:

$$\begin{aligned} Z_1 &\to Z_1 - i(\mathbf{S}\cdot\boldsymbol{\theta})Z_1, \\ Z_2 &\to Z_2 - i(\mathbf{S}\cdot\boldsymbol{\theta})Z_2; \end{aligned} \tag{4.1}$$

Lorentz transformations:

$$\begin{aligned} Z_1 &\to Z_1 - (\mathbf{S}\cdot\boldsymbol{\theta})Z_1, \\ Z_2 &\to Z_2 + (\mathbf{S}\cdot\boldsymbol{\theta})Z_2. \end{aligned} \tag{4.2}$$

We must determine the effects of various powers of $(\mathbf{S}\cdot\mathbf{p})$ on the Z, and then eliminate elements of the type $(\varphi\cdots\eta\cdots\eta\cdots\varphi)$. For this we must consider the algebraic properties of the $(\mathbf{S}\cdot\mathbf{p})$:

$$(\mathbf{S}\cdot\mathbf{p})^{AB} \equiv \tfrac{1}{2}(A \mid \alpha_1\cdots\alpha_k)\times[(\boldsymbol{\sigma}\cdot\mathbf{p})_{\alpha_1\beta_1}\delta_{\alpha_2\beta_2}\cdots\delta_{\alpha_k\beta_k}+\cdots+\delta_{\alpha_1\beta_1}\cdots(\boldsymbol{\sigma}\cdot\mathbf{p})_{\alpha_k\beta_k}](\beta_1\cdots\beta_k \mid B)$$

or

$$(\mathbf{S}\cdot\mathbf{p})^{AB} \leftrightarrow \tfrac{1}{2}[(\boldsymbol{\sigma}\cdot\mathbf{p})\times 1\times 1\cdots\times 1+\cdots+1\times 1\times\cdots\times(\boldsymbol{\sigma}\cdot\mathbf{p})].$$

Now define $(\boldsymbol{\sigma}_q\cdot\mathbf{p})$ as the sum of expressions of the form

$$1\times\cdots\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times\cdots\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times\cdots$$

(k factors, of which q are $\boldsymbol{\sigma}\cdot\mathbf{p}$), where we sum over all permutations of position of the q $(\boldsymbol{\sigma}\cdot\mathbf{p})$ factors.

Thus

$$\begin{aligned}
(\boldsymbol{\sigma}_1\cdot\mathbf{p}) &= (\boldsymbol{\sigma}\cdot\mathbf{p})\times 1\times\cdots+\cdots+1\times\cdots\times(\boldsymbol{\sigma}\cdot\mathbf{p}) \\
(\boldsymbol{\sigma}_2\cdot\mathbf{p}) &= (\boldsymbol{\sigma}\cdot\mathbf{p})\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times\cdots+1\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times 1\times(\boldsymbol{\sigma}\cdot\mathbf{p})+\cdots+1\times\cdots\times 1\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times(\boldsymbol{\sigma}\cdot\mathbf{p}) \\
&\ \ \vdots \qquad\qquad \vdots \\
(\boldsymbol{\sigma}_k\cdot\mathbf{p}) &= (\boldsymbol{\sigma}\cdot\mathbf{p})\times(\boldsymbol{\sigma}\cdot\mathbf{p})\times\cdots\times(\boldsymbol{\sigma}\cdot\mathbf{p}) \\
(\boldsymbol{\sigma}_q\cdot\mathbf{p}) &\equiv 0 \qquad (q>k).
\end{aligned}$$

Then we have

$$\begin{aligned}
(\mathbf{S}\cdot\mathbf{p}) &\leftrightarrow \tfrac{1}{2}(\boldsymbol{\sigma}_1\cdot\mathbf{p}) \\
(\mathbf{S}\cdot\mathbf{p})^2 &\leftrightarrow \tfrac{1}{4}k\mathbf{p}^2+\tfrac{1}{2}(\boldsymbol{\sigma}_2\cdot\mathbf{p}) \\
(\mathbf{S}\cdot\mathbf{p})^3 &\leftrightarrow \mathbf{p}^2\tfrac{1}{4}(3k-2)(\mathbf{S}\cdot\mathbf{p})+\tfrac{3}{4}(\boldsymbol{\sigma}_3\cdot\mathbf{p}) \\
&\ \ \vdots \qquad\qquad \vdots \\
(\mathbf{S}\cdot\mathbf{p})^n &\leftrightarrow \sum_{i=1}^{[n/2]}(\mathbf{p}^2)^i f_i^{(n)}(\mathbf{S}\cdot\mathbf{p})^{n-2i}+\frac{n!}{2^n}(\boldsymbol{\sigma}_n\cdot\mathbf{p}) \\
&\ \ \vdots \qquad\qquad \vdots \\
(\mathbf{S}\cdot\mathbf{p})^{k+1} &\leftrightarrow \sum_{i=1}^{l} f_i^{(k+1)}(\mathbf{S}\cdot\mathbf{p})^{k+1-2i}(\mathbf{p}^2)^i, \\
&\qquad (k=2l \ \text{ or } \ 2l+1),
\end{aligned} \tag{4.3}$$

which defines the coefficients $f_i^{(n)}$.

Consideration of recurrence relations [obtained by multiplying $(\mathbf{S}\cdot\mathbf{p})^n$ by $(\mathbf{S}\cdot\mathbf{p})$] yields (with $f_0^{(n)} \equiv -1$)

$$f_j^{(n+1)} = f_j^{(n)} - \tfrac{1}{4}n(k-n+1)f_{j-1}^{(n-1)} \qquad (4.3a)$$

and

$$f_j^{(n)} = 0 \quad \text{for} \quad j > \tfrac{1}{2}n.$$

A discussion of these algebraic relations is given in the Appendix.

We now proceed to the study of the general wave equation. Define X_q as the sum of the generalized Clebsch–Gordon combination of q η-functions and $(k-q)$ φ-functions, summed over all permutations; define Y_q as the same with φ and η interchanged. Thus,

$$X_0^A \equiv Z_1^A \equiv (A \mid \alpha_1 \cdots \alpha_k)\varphi_{\alpha_1}\cdots\varphi_{\alpha_k}$$

$$Y_0^A \equiv Z_2^A \equiv (A \mid \alpha_1 \cdots \alpha_k)\eta_{\alpha_1}\cdots\eta_{\alpha_k}$$

$$X_1^A \equiv (A \mid \alpha_1 \cdots \alpha_k)(\eta_{\alpha_1}\varphi_{\alpha_2}\cdots + \cdots + \varphi_{\alpha_1}\cdots\varphi_{\alpha_{k-1}}\eta_{\alpha_k})$$

$$Y_1^A \equiv (A \mid \alpha_1 \cdots \alpha_k)(\varphi_{\alpha_1}\eta_{\alpha_2}\cdots + \cdots + \eta_{\alpha_1}\cdots\eta_{\alpha_{k-1}}\varphi_{\alpha_k})$$

$$\vdots \qquad\qquad \vdots$$

Define, as for the case $s = 1$, the general wave-function Z

$$Z^A \equiv \begin{pmatrix} Z_1^A \\ Z_2^A \end{pmatrix} \equiv W_0^A$$

and likewise

$$W_r^A \equiv \begin{pmatrix} X_r^A \\ Y_r^A \end{pmatrix} \qquad (r = 1 \cdots k).$$

We see that for k = even $= 2l$ ($s = l$) (since, e.g., $X_{k-1} = Y_1$)

$$W_{l+p} = \rho_1 W_{l-p} \qquad (p = 1 \cdots l) \qquad (4.4)$$

and for k = odd $= 2l+1$ ($s = l + \frac{1}{2}$)

$$W_{l+p} = \rho_1 W_{l-(p-1)} \qquad (r = 1 \cdots l+1). \qquad (4.5)$$

Now consider the action of the operator $(\delta_q\cdot\mathbf{p})$ on Z_1:

$$(A \mid \alpha_1 \cdots \alpha_k)(\delta_q\cdot\mathbf{p})_{\alpha_1\cdots,\beta_1\ldots}(\beta_1 \cdots \mid B)(B \mid \gamma_1 \cdots)\varphi_{\gamma_1}\cdots$$

$$= (A \mid \alpha_1 \cdots \alpha_k)\Big[\sum_{(p)} \{\cdots (\delta\cdot\mathbf{p})_{\alpha_s\beta_s}\varphi_{\beta_s}\cdots(\delta\cdot\mathbf{p})_{\alpha_r\beta_r}\varphi_{\beta_r}\cdots\}\Big]$$

$$= (A \mid \alpha_1 \cdots)\Big[\sum_{(p)} \{\cdots (E\varphi_{\alpha_s} - m\eta_{\alpha_s})\cdots(E\varphi_{\alpha_r} - m\eta_{\alpha_r})\cdots\}\Big]$$

$$= (A \mid \alpha_1 \cdots \alpha_k)\sum_{r=0}^{q}\binom{k-r}{q-r}E^{q-r}(-m)^r X_r,$$

where the sum is over the permutations.

The analogous term for Z_2 results in the combined equation

$$(A \mid \alpha_1 \cdots)(\delta_q\cdot\mathbf{p})_{\alpha\cdots,\beta\ldots}(\beta_1 \cdots \mid B)Z^B = \rho_3^q \sum_{r=0}^{q}\binom{k-r}{q-r}E^{q-r}(-m)^r W_r^A, \qquad (q = 1 \cdots k),$$

This array of equations can now be solved for W_s^A, and we then obtain

$$W_s^A = \frac{1}{m^s}\sum_{i=0}^{s}(-\rho_3)^{s-i}E^i\binom{k-s+j}{j} \times (A \mid \alpha \cdots)(\delta_{s-i}\cdot\mathbf{p})_{\alpha\ldots\beta\ldots}(\beta \cdots \mid B)Z^B.$$

Now we must express the $(\delta_{s-i}\cdot\mathbf{p})$ in terms of the $(\mathbf{S}\cdot\mathbf{p})$ and eliminate the W_s^A using Eqs. (4.4) and (4.5), thus obtaining a wave equation for Z^A involving only powers of $(\mathbf{S}\cdot\mathbf{P})$ and $\mathbf{p}$, E.

Making use of Eq. (4.3), we have

$$(\delta_{s-i}\cdot\mathbf{p}) \leftrightarrow -\frac{2^{s-i}}{(s-j)!}\sum_{r=0}^{[(s-i)/2]} f_r^{(s-i)}(\mathbf{S}\cdot\mathbf{P})^{s-i-2r}(\mathbf{p}^2)^r.$$

So

$$W_s^A = -m^{-s}\sum_{i=0}^{s}\sum_{r=0}^{[(s-i)/2]}\frac{2^{s-i}}{(s-j)!}(\mathbf{p}^2)^r(-\rho_3)^{s-i}E^i \times \binom{k-s+j}{j} f_r^{(s-i)}[(\mathbf{S}\cdot\mathbf{P})^{s-i-2r}]^{AB}Z^B. \qquad (4.6)$$

Now use Eqs. (4.4) and (4.5) to eliminate W_s^A. Let $k = 2l + \delta$ (k = even, $\delta = 0$; k = odd, $\delta = 1$), then

$$W_{l+1}^A = \rho_1 W_{l-1+\delta}^A.$$

Use Eq. (4.6) to express W_{l+1}^A and $\rho_1 W_{l-1+\delta}^A$ and equate the two. Then after algebraic manipulation, we obtain

$$\Sigma_s^{AB}(p)Z^B \equiv \sum_{q=0}^{l+1} T_q\left\{\left[-\frac{\rho_3(\mathbf{S}\cdot\mathbf{P})}{E}\right]^q\right\}^{AB}Z^B = 0, \qquad (4.7)$$

where

$$T_q \equiv E^2F_q^{(1)}(\mathbf{p}^2/E^2) - \rho_1 m^2(E/m)^\delta F_q^{(2)}(\mathbf{p}^2/E^2)$$

$$\equiv E^2D_q^{(1)}(m^2/E^2) - \rho_1 m^2(E/m)^\delta D_q^{(2)}(m^2/E^2),$$

TABLE I. Covariant wave equations for $s \leq 4$.

Spin	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^0$	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^1$	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^2$	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^3$	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^4$	$(-\rho_3 \mathbf{S}\cdot\mathbf{p})^5$
$\frac{1}{2}$	$E-\rho_1 m$	2				
1	$m^2(1-\rho_1)$	$+2E$	2			
$\frac{3}{2}$	$3(m^2+E^2)-6mE\rho_1$	$4(2E-m\rho_1)$	4			
2	$6m^2E(1-\rho_1)$	$(4E^2+2m^2)+3m^2(1-\rho_1)$	$6E$	2		
$\frac{5}{2}$	$15[E(E^2+3m^2)$ $-\rho_1 m(3E^2+m^2)]$	$46E^2+26m^2-48mE\rho_1$	$12(3E-m\rho_1)$	8		
3	$9m^2(4E^2+m^2)(1-\rho_1)$	$6E(2E^2+3m^2)$ $+30m^2E(1-\rho_1)$	$22E^2+8m^2$ $+6m^2(1-\rho_1)$	$12E$	2	
$\frac{7}{2}$	$105[(E^4+6m^2E^2+m^4)$ $-4\rho_1 mE(E+m^2)]$	$8[4E(11E^2+19m^2)$ $-\rho_1 m(71E^2+19m^2)]$	$8[43E^2+17m^2$ $-30mE\rho_1]$	$32(4E-m\rho_1)$	16	
4	$60Em^2(4E^2+3m^2)$ $\times(1-\rho_1)$	$6(8E^4+24m^2E^2+3m^4)$ $+5m^2(52E^2+11m^2)(1-\rho_1)$	$10[(10E^2+11m^2)$ $+9m^2(1-\rho_1)]$	$10[7E^2+2m^2$ $+m^2(1-\rho_1)]$	$20E$	2

and

$$F_q^{(1)}(x) \equiv \sum_{i=0}^{[(l+1-q)/2]} B_{qi}^{(1)}(-x)^i$$

$$F_q^{(2)}(x) \equiv \sum_{i=0}^{[(l+\delta-1-q)/2]} B_{qi}^{(2)}(-x)^i,$$

where

$$B_{qj}^{(1)} \equiv \frac{2^q}{(q+2j)!}\begin{bmatrix} 2l+\delta-q-2j \\ l+1-q-2j \end{bmatrix} C_j^{q+2j}$$

$$B_{qj}^{(2)} \equiv \frac{2^q}{(q+2j)!}\begin{bmatrix} 2l+\delta-q-2j \\ l+\delta-1-q-2j \end{bmatrix} C_j^{q+2j}$$

$$D_q^{(1)}(x) \equiv \sum_{i=0}^{[(l+1-q)/2]} \sum_{s=i}^{[(l+1-q)/2]} (-1)^s \binom{s}{j} B_{qs}^{(1)}(-x)^i$$

$$D_q^{(2)}(x) \equiv \sum_{i=0}^{[(l-1+\delta-q)/2]} \sum_{s=i}^{[(l-1+\delta-q)/2]} (-1)^s \binom{s}{j} B_{qs}^{(2)}(-x)^i .$$

The coefficients C_j^{q+2j} are determined from the recursion relations in the Appendix. We illustrate with two examples, $k = 1, 2$:

(1) $k = 1, \quad l = 0, \quad \delta = 0, \quad s = \frac{1}{2}$

$$F_1^{(1)} = 2 \quad F_0^{(1)} = 1 \quad F_0^{(2)} = 1$$

$$T_1 = 2E^2 \quad T_0 = E^2 - \rho_1(Em)$$

$$(E^2 - \rho_1 Em) + 2E^2(-\rho_3 \mathbf{S}\cdot\mathbf{p}/E)$$

$$= E[(E - m\rho_1) - 2\rho_3(\mathbf{S}\cdot\mathbf{p})] = 0.$$

This is just the Dirac equation, since for $s = \frac{1}{2}$, we have

$$(\mathbf{S}\cdot\mathbf{p}) = \tfrac{1}{2}(\boldsymbol{\sigma}\cdot\mathbf{p});$$

(2) $k = 2, \quad l = 1, \quad \delta = 0, \quad s = 1$

$$F_0^{(1)} = \left(1 - \frac{\mathbf{p}^2}{E^2}\right) \qquad F_0^{(2)} = 1$$

$$T_0 = E^2(1 - \mathbf{p}^2/E^2) - \rho_1 m^2 = m^2(1 - \rho_1)$$

$$F_1^{(1)} = 2 \qquad T_1 = 2E^2$$

$$F_2^{(1)} = 2 \qquad T_2 = 2E^2$$

$$\therefore\ m^2(1 - \rho_1) + 2E^2(-\rho_3 \mathbf{S}\cdot\mathbf{p}/E) + 2E^2(-\rho_3 \mathbf{S}\cdot\mathbf{p}/E)^2 = 0.$$

which is just our spin-1 equation, Eq. (2.6).

The general problem of constructing wave equations of higher spin leads to extensive complexities as the spin increases. For this reason, an IBM 7090 computer was programmed for computation of the higher-spin equations. The computation was carried out up to $k = 20$, corresponding to spin 10. Only results for lower spins are given here since the coefficients start to run up into six integers. Table I indicates the structure of the wave equations for spin ≤ 4.

Thus, we have succeeded in constructing a general higher-spin covariant wave equation, with no redundant components for the wavefunction. That these equations are covariant follows directly from their construction from the spin-$\frac{1}{2}$ Dirac equation. We regard these equations as the natural generalization of the Dirac equation, and the $2s + 1$ component entity Z as the natural generalization of the Dirac wavefunction.

We give an explicit proof of the covariance of the general equation, Eq. (4.7), under spatial rotations. Under rotation,

$$\mathbf{p} \to \mathbf{p} + (\boldsymbol{\theta} \times \mathbf{p}),$$

$$Z^A \to Z^A - i(\mathbf{S}\cdot\boldsymbol{\theta})^{AB} Z^B.$$

The functions T_q will not change, since they depend only on $\mathbf{p}^2$ and E. Thus, we have for covariance

$$\sum_{q=0}^{l+1} T_q \left[-\frac{\rho_3}{E}\right]^q \{(\mathbf{S}\cdot\mathbf{p})^{q-1}(\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p}) + \cdots + (\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p})(\mathbf{S}\cdot\mathbf{p})^{q-1} - i(\mathbf{S}\cdot\mathbf{p})^q(\mathbf{S}\cdot\boldsymbol{\theta})\}Z = 0.$$

Now from the angular-momentum commutation relations, we have

$$+ i[(\mathbf{Sp})^q, (\mathbf{S\theta})] = (\mathbf{S}\cdot\mathbf{p})^{q-1}(\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p}) + \cdots + (\mathbf{S}\cdot\boldsymbol{\theta}\times\mathbf{p})(\mathbf{S}\cdot\mathbf{p})^{q-1}$$

and, therefore, we obtain

$$(\mathbf{S}\cdot\boldsymbol{\theta})\left[\sum_{q=0}^{l+1} T_q\left(\frac{-\rho_3}{E}\right)^q (Sp)^q\right]\mathbf{Z} = 0,$$

and the general equation is indeed covariant under spatial rotations.

The problem of proving covariance under the Lorentz transformations is much more complex. One must make use of explicit form of the $(2s + 1)$ by $(2s + 1)$ S matrices for each spin s; For example,

$$s = \tfrac{1}{2}: \quad \{(\mathbf{S}\cdot\mathbf{a}), (\mathbf{S}\cdot\mathbf{b})\}_+ = \tfrac{1}{2}(\mathbf{a}\cdot\mathbf{b}).$$

In addition, the quantities T_q will also change, and the interrelationships among them, expressed by their dependence on the C_j^n and the recursion relations of the C_j^n, will be involved in the proof.

The adjoint operator for the first three cases is given by

$$\hat{\Sigma}_{\frac{1}{2}}(p) = \rho_2\Sigma_{\frac{1}{2}}(p)\rho_2$$
$$\hat{\Sigma}_1(p) = \rho_2\Sigma_1(p)\rho_2$$
$$\hat{\Sigma}_{\frac{3}{2}}(p) = \rho_2\Sigma_{\frac{3}{2}}(p)\rho_2$$

and we conjecture that this is true in general.

From the adjoint operator we can define the solution of the plane-wave problem; i.e., the solution of $\Sigma_s^{AB}(p)Z^B(p) = 0$ is then $Z^B(p) = \hat{\Sigma}^{BC}(p)\chi^C$, where χ^C is one of a set of $2(2s + 1)$ arbitrary $2(2s + 1)$ component vectors. We illustrate for $s = \frac{1}{2}$:

$$\Sigma_{\frac{1}{2}}^{AB}(p)Z^B = [2\rho_3(\mathbf{S}\cdot\mathbf{p}) - E + m\rho_1]^{AB}Z^B(p) = 0,$$
$$\therefore Z^B(p) = [2\rho_3(\mathbf{S}\cdot\mathbf{p}) + E + m\rho_1]^{BC}\chi^C.$$

There are four solutions to the equation. If we take as our χ^C the set of four constant vectors with 1 in successive rows, then the four solutions will become the four columns of the array

$$\begin{bmatrix} p^3 + E & p^1 - ip^2 & m & 0 \\ p^1 + ip^2 & -p^3 + E & 0 & m \\ m & 0 & -p^3 + E & -p^1 + ip^1 \\ 0 & m & -p^1 - ip^2 & p^3 + E \end{bmatrix}$$

(aside from a normalization factor).

We note that each equation contains the "germ" of the lower-spin equations. We illustrate this idea for $s = \frac{3}{2}$ and $s = 2$.

(a) $s = \frac{3}{2}$:

We may write the $s = \frac{3}{2}$ equation in the form

$$(\mathbf{S}\cdot\mathbf{p})^2 - (\tfrac{1}{4})\mathbf{p}^2 = (\tfrac{1}{2})(2\rho_3 E + im\rho_2) \times [2(\mathbf{S}\cdot\mathbf{p}) - \rho_3 E + im\rho_2]. \tag{4.8}$$

Now the $s = \frac{1}{2}$ equation obeys

$$(\mathbf{S}\cdot\mathbf{p})^2 - \tfrac{1}{4}\mathbf{p}^2 = 0$$
$$2(\mathbf{S}\cdot\mathbf{p}) - \rho_3 E + im\rho_2 = 0,$$

so it *a fortiori* obeys the $s = \frac{3}{2}$ equation. We may also write Eq. (4.8) as

$$[3(E - m\rho_1) - 2\rho_3(\mathbf{S}\cdot\mathbf{p})] \times [(E - m\rho_1) - 2\rho_3(\mathbf{S}\cdot\mathbf{p})] = 0.$$

(b) $s = 2$:

The wave equation can be written

$$(\mathbf{S}\cdot\mathbf{p})^3 - p^2(\mathbf{S}\cdot\mathbf{p}) = \tfrac{3}{2}E\rho_3[2(\mathbf{S}\cdot\mathbf{p})^2 - 2\rho_3 E(\mathbf{S}\cdot\mathbf{p}) + m^2(1 - \rho_1)] - \tfrac{3}{4}m^2[(1 + \rho_1)(\mathbf{S}\cdot\mathbf{p}) - E\rho_3(1 - \rho_1)]$$

or

$$[2E - \rho_3(\mathbf{S}\cdot\mathbf{p})][m^2(1 - \rho_1) - 2E\rho_3(\mathbf{S}\cdot\mathbf{p}) + 2(\mathbf{Sp})^2] + 4m^2(1 - \rho_1)[E - \rho_3(\mathbf{Sp})] = 0.$$

The spin-1 particle obeys

$$(\mathbf{Sp})^3 = \mathbf{p}^2(\mathbf{Sp}),$$
$$2(\mathbf{Sp})^2 - 2\rho_3 E(\mathbf{Sp}) + m^2(1 - \rho_1) = 0, \tag{4.9}$$
$$(1 + \rho_1)(\mathbf{Sp}) - E\rho_3(1 - \rho_1) = 0. \tag{4.10}$$

Equation (4.10) follows from Eq. (4.9) as follows

$$2\mathbf{p}^2(\mathbf{S}\cdot\mathbf{p}) = 2(\mathbf{S}\cdot\mathbf{p})^3 = 2\rho_3 E(\mathbf{S}\cdot\mathbf{p})^2 - m^2(1 - \rho_1)(\mathbf{S}\cdot\mathbf{p})$$
$$= \rho_3 E[2\rho_3 E(\mathbf{S}\cdot\mathbf{p}) - m^2(1 - \rho_1)] - m^2(1 - \rho_1)(\mathbf{S}\cdot\mathbf{p})$$
$$\therefore 2(\mathbf{S}\cdot\mathbf{p})m^2 - m^2E\rho_3(1 - \rho_1) + m^2(1 - \rho_1)(\mathbf{S}\cdot\mathbf{p}) = 0$$
$$\therefore (1 + \rho_1)(\mathbf{S}\cdot\mathbf{p}) - E\rho_3(1 - \rho_1) = 0$$

or

$$(1 - \rho_1)[E - \rho_3(\mathbf{S}\cdot\mathbf{p})] = 0.$$

Thus the spin-1 particle obeys the spin-2 equation. In general the s = integer (half-integer) particle will obey all higher-integer (half-integer) equations.

The integer-spin equations cannot be solved for E since the integer-spin S matrices are singular. The half-integer spin equations can be solved for E and yield for the first two cases:

(a) spin $\frac{1}{2}$

$$E = \rho_1 m + 2\rho_3(\mathbf{S}\cdot\mathbf{p}),$$

(b) spin $\frac{3}{2}$

$$E = \rho_1 m + 2\rho_3(\mathbf{Sp}) - \tfrac{2}{3}[(\mathbf{Sp})^2 - \tfrac{1}{4}\mathbf{p}^2] \times [m^2 + 4\mathbf{p}^2]^{-1}[3m\rho_1 + 4\rho_3(\mathbf{S}\cdot\mathbf{p})].$$

The half-integral spin wave equations can be written in the form:

$$(s = \tfrac{1}{2}) \quad [(E - m\rho_1) + \mathcal{S}] = 0 \qquad \mathcal{S} \equiv -2\rho_3(\mathbf{S}\cdot\mathbf{p})$$

$$(s = \tfrac{3}{2}) \quad [3(E - m\rho_1) + \mathcal{S}][(E - m\rho_1) + \mathcal{S}] = 0$$

$$(s = \tfrac{5}{2}) \quad [5(E - m\rho_1) + \mathcal{S}][3(E - m\rho_1) + \mathcal{S}] \times [(E - m\rho_1) + \mathcal{S}] = 0.$$

We *conjecture* that this is true in general, and have therefore for the general wave equation of spin $r/2$ (r = odd integer)

$$\prod_{t=0}^{r-1} [(2t + 1)(E - m\rho_1) + \mathcal{S}] = 0$$

(In proper order).

(A corresponding expression for integer spin does not seem to exist.)

V. COMMENTS ON THE SYNTHESIS

There is an at first puzzling aspect of this synthesis. We seem to be combining in a sense k mass-m, spin-$\frac{1}{2}$ particles, to get a particle, again of mass m, and of spin $k/2$, whereas we might expect the composite particle to have mass km. We can of course ignore the process whereby we obtained Eq. (4.7), and simply regard it as given. Then proving that it represents a particle of mass m and spin $s = k/2$ would involve (a) showing the existence of an adjoint operator $\hat{\Sigma}_s$ such that we have $m^2 = E^2 - \mathbf{p}^2$; (b) showing the covariance of Eq. (4.7) under the transformations, Eqs. (4.1) and (4.2), of rotation and the Lorentz transformations. Nothing in principle, except algebraic complexities, would prevent this course.

We consider instead the mass paradox. Consider a state vector which represents a particle of mass m and spin s, with momentum p (α represents the spin coordinate: $\alpha = -s, \cdots, +s$)

$$|m, s; \mathbf{p}, \alpha\rangle.$$

Consider the problem of combining two such states to form a state of mass m and spin s:

$$|m, s; \mathbf{p}, \alpha\rangle = \int d^4p_1 \int d^4p_2 (s_1 s_2 \alpha \mid s_1\alpha_1 s_2\alpha_2) \times f(p_1, p_2)\, |m_1 s_1; p_1\alpha_1\rangle\, |m_2 s_2; p_2\alpha_2\rangle.$$

We require that the infinitesimal generators corresponding to displacement yield p as eigenvalue:

$$P_\mu\, |m, s; p, \alpha\rangle = p_\mu\, |m, s; p, \alpha\rangle$$

$$P_\mu = (p_\mu \times 1) + (1 \times p_\mu);$$

so

$$p_\mu = p_{1\mu} + p_{2\mu}.$$

Now

$$P_\mu P^\mu = (1 \times p^\mu + p^\mu \times 1)(1 \times p_\mu + p_\mu \times 1) = (p^\mu p_\mu) \times 1 + 1 \times (p^\mu p_\mu) + 2(p^\mu \times p_\mu),$$

so

$$m_1^2 + m_2^2 + 2(p_1 p_2) = m^2$$

This simplest solution to this problem is to take

$$p_1^\mu = (m_1/m)p^\mu \qquad p_2^\mu = (m_2/m)p^\mu,$$

and our state vectors are therefore taken to be Clebsch–Gordon combinations of state vectors evaluated at

$$\mathbf{p}_1 = (m_1/m)\mathbf{p} \qquad \mathbf{p}_2 = (m_2/m)\mathbf{p},$$

i.e., for $s = 1$:

$$Z^A(\mathbf{p}) = (A \mid \alpha_1\alpha_2)\varphi_{\alpha_1}(m_1, (m_1/m)\mathbf{p})\varphi_{\alpha_2}(m_2, (m_2/m)\mathbf{p}).$$

Now the Dirac equation for $\varphi(m_1, \mathbf{p})$ is

$$E\varphi = (\boldsymbol{\vartheta}\cdot\mathbf{p})\varphi + m_1\eta$$

and, therefore, the equation governing

$$\varphi(m_1, (m_1/m)\mathbf{p})$$

is

$$(m_1/m)E\varphi = (m_1/m)(\boldsymbol{\vartheta}\cdot\mathbf{p})\varphi + m_1\eta$$

or

$$E\varphi = (\boldsymbol{\vartheta}\cdot\mathbf{p})\varphi + m\eta.$$

This is now the reason why our formalism works. We are working with Dirac wavefunctions of mass m_1 but evaluated at $(m_1/m)\mathbf{p}$. Thus, it is true that the wavefunction φ obeys the Dirac equation for mass m_1, but the quantities we actually use are $\varphi((m_i/m)\mathbf{p})$ which obeys the Dirac equation for mass m.

We are therefore combining a particle of mass m_1 and a particle of mass m_2 ($m = m_1 + m_2$), each with spin $\frac{1}{2}$ to obtain a particle of mass m and spin 1. Likewise for the general case, we combine k particles of mass $m_i (i = 1, \cdots, k)$ where $\sum \Re\, m_i = m$.

A clearer view of this process may be had by going

to the coordinate space representation:

$$Z_A(m, \mathbf{p}) = (A \mid \alpha_1\alpha_2)\varphi_{\alpha_1}(m_1, (m_1/m)\mathbf{p})\varphi_{\alpha_2}(m_2, (m_2/m)\mathbf{p}).$$

Define

$$\varphi_{\alpha_i}(x) \equiv \int d^4p e^{-i(px)}\varphi_{\alpha_i}(p);$$

$$Z_A(x) = \int d^4p e^{-i(px)} Z_A(p).$$

Then

$$Z_A(x) = (A \mid \alpha_1\alpha_2)(2\pi)^{-4} \times \left(\frac{m_2}{m}\right)^4 \int d^4y \varphi_{\alpha_1}\left(x + \frac{m_2}{m}y\right)\varphi_{\alpha_2}\left(x - \frac{m_1}{m}y\right).$$

Thus the coordinate state vector at x is a nonlocal combination of $\varphi_{\alpha_1}(x)$ and $\varphi_{\alpha_2}(x)$ with weighting distance dependent on the masses involved.

VI. CONCLUSION

We have shown how to synthesize wave equations for particles of arbitrary spin, with no redundant components in the wave equation. The Dirac equation is the first of these equations and the vector meson equation (in terms of the field strengths, not the potential) constitutes the natural generalization of the Dirac equation to spin 1: higher-spin equations involve higher powers of derivatives, when considered in the context of a local field theory. The space and time dependence enter the structure of the equations in an unsymmetrical way, but they are nevertheless covariant, since space–time symmetrization is sufficient for covariance, but not necessary.

APPENDIX

We define

$$C_i^n \equiv 4^i(-1)^{i+1} f_i^{(n)}$$

and the recursion relation Eq. (4.3a) becomes ($C_0^n \equiv +1$, $C_i^r = 0$, $r \geq 2j$)

$$C_i^{n+1} = C_i^n + n(k - n + 1)C_{i-1}^{n-1}.$$

The solution for $j = 1, 2$ is

$$C_1^n = k\binom{n}{2} - 2\binom{n}{3}$$

$$C_2^n = (3k^2 - 6k)\binom{n}{4} + (24 - 20k)\binom{n}{5} + 40\binom{n}{6}$$

which suggests a general solution of the form

$$C_i^n = \sum_{r=0}^{i}\sum_{s=0}^{r} A_{rs}^{(i)} k^s \binom{n}{3j - r} = \sum_{s=0}^{i}\sum_{r=s}^{i} A_{rs}^{(i)} k^s \binom{n}{3j - r}.$$

In fact, this obeys the recursion relations, provided we have

$$A_{rs}^{(i)} = -(3j - 2 - r)(3j - 1 - r)(A_{rs}^{(i-1)} + A_{r-1,s}^{(i-1)}) + (3j - r - 1)A_{r-1,s-1}^{(i-1)}.$$

A few values:

$$A_{00}^{(0)} = 1 \quad A_{00}^{(1)} = -2 \quad A_{10}^{(1)} = 0 \quad A_{11}^{(1)} = 0$$

$$A_{00}^{(2)} = 40 \quad A_{10}^{(2)} = 24 \quad A_{11}^{(2)} = -20 \quad A_{21}^{(2)} = -6 \quad A_{20}^{(2)} = 0 \quad A_{22}^{(2)} = 3.$$

In practice, it is more convenient to use the C_i^n directly, and compute the C_i^n arrays for each value of k separately. This was done for the computation of the functions $F_q^{(1)}$ and $F_q^{(2)}$ of Sec. IV.

Position and Intrinsic Spin Operators in Quantum Theory*

R. A. BERG
Lockheed Research Laboratories, Palo Alto, California
(Received 21 July 1964)

The problem of defining position and intrinsic spin operators in terms of the generators of the inhomogeneous Lorentz group is considered. These operators are taken to have the properties usually attributed to them in nonrelativistic quantum theory: their commutation relations must have the commonly accepted form. These commutation relations are then shown to define the intrinsic spin uniquely. The position operator is shown to be also essentially uniquely determined. Explicit forms of the spin and position operators for some special representations are exhibited. The relation of these operators to the spin and position operators of the spin-$\frac{1}{2}$ Dirac theory in the Foldy–Wouthuysen representation is considered. An Appendix gives a heuristic derivation of the spin operator.

1. INTRODUCTION

THE concept of position and intrinsic spin operators in quantum theory has been considered in various aspects.[1] In this paper, we show that assuming some requirements considered as desirable for these operators leads to an essentially unique determination of them in terms of the generators of the inhomogeneous Lorentz group (ILG).

For the special case of $s = \frac{1}{2}$ we consider the relation between the Foldy–Wouthuysen (FW) position and spin operators[2] and the operators considered in this paper.

I. THE GENERATORS FOR THE INHOMOGENEOUS LORENTZ GROUP

The generators of the ILG[3] are determined by the unitary transformations corresponding to the infinitesimal Lorentz transformations. We have

$$x'_\mu = \omega_{\mu\nu}x^\nu + \epsilon_\mu + x_\mu$$

corresponding to

$$U = 1 - i\omega^{\mu\nu}M_{\mu\nu} - i\epsilon^\mu P_\mu$$

[metric: $(1, -1, -1, -1)$, $x^0 = t$, $a^\mu b_\mu = a^0b^0 - \mathbf{a}\cdot\mathbf{b}$].

The communication relations obeyed by the generators are

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(g_{\nu\rho}M_{\mu\sigma} - g_{\mu\rho}M_{\nu\sigma} + g_{\mu\sigma}M_{\nu\rho} - g_{\nu\sigma}M_{\mu\rho}),$$

$$[M_{\mu\nu}, P_\rho] = i(g_{\nu\rho}P_\mu - g_{\mu\rho}P_\nu).$$

We now define

$$M^1 = M_{23}, \quad \text{etc.}$$
$$N^i = M_{0i},$$
$$E = P^0.$$

Then

$$\begin{aligned} [M^i, M^j] &= -[N^i, N^j] = i\epsilon^{ijk}M^k, \\ [M^i, N^j] &= i\epsilon^{ijk}N^k, \\ [M^i, E] &= 0, \qquad [N^i, E] = iP^i, \\ [M^i, P^j] &= i\epsilon^{ijk}, \quad [N^i, P^j] = i\delta^{ij}E. \end{aligned} \tag{1.1}$$

We concern ourselves in this paper with the representations for real, nonzero mass: the eigenvalues of $P_\mu P^\mu$ are greater than zero.

II. REQUIREMENTS ON THE INTRINSIC SPIN OPERATOR AND THE POSITION OPERATOR

We want to define an intrinsic spin operator $\mathbf{S}$ and a position operator $\mathbf{Q}$ with the following properties:

$$[Q^i, E] = iP^iE^{-1} \equiv iv^i, \tag{2.1}$$
$$[Q^i, P^j] = i\delta^{ij}, \tag{2.2}$$
$$[Q^i, Q^j] = 0, \tag{2.3}$$
$$[S^i, Q^j] = 0, \tag{2.4}$$
$$[S^i, P^\mu] = 0, \tag{2.5}$$
$$[S^i, S^j] = i\epsilon^{ijk}S^k, \tag{2.6}$$
$$[S^i, M^j] = i\epsilon^{ijk}S^k, \tag{2.7}$$
$$[Q^i, M^j] = i\epsilon^{ijk}Q^k. \tag{2.8}$$

Equation (2.1) expresses the requirement that the commutator of position operator with the energy is the velocity $\mathbf{P}/E$. Equations (2.2) and (2.3) are the usually required commutation relations of position and momentum, and position with itself. Equa-

* This work was carried out as a part of the Lockheed Independent Research Program.
[1] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. 21, 400 (1949); M. H. L. Pryce, Proc. Roy. Soc. (London) A195, 62 (1948).
[2] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. 78, 29 (1950).
[3] E. P. Wigner, Am. Math. 40, 149 (1939); E. P. Wigner and V. Bargmann, Proc. Natl. Acad. Sci. U. S. 34, 211 (1948).

tions (2.4) and (2.5) express the commutation of intrinsic spin with position and 4-momentum. Equation (2.6) is the desired property of an intrinsic spin operator and Eqs. (2.7) and (2.8) express the requirement that S and Q are 3-vectors under spatial rotations.

In addition we require that the operator of total angular momentum be the sum of an orbital and an intrinsic part:

$$\mathbf{M} = \mathbf{Q}\times\mathbf{P} + \mathbf{S}. \tag{2.9}$$

These requirements differ from those of a recent paper[4] on the position operator, where Eqs. (2.1) and (2.9) are not required, but the requirement

$$[Q^i, N^j] = \tfrac{1}{2}\{Q^j[Q^i, E] + [Q^i, E]Q^j\}$$

is imposed. These differing requirements lead to different formulations for the position operators, which however agree in the case of the FW position operator for spin $\frac{1}{2}$ (see Sec. VII).

III. THE INTRINSIC SPIN OPERATOR

We must express S as some combination of the basic operators of the ILG: **P**, E, **M**, **N**. We have from Eq. (2.6) that the highest power of **M** or **N** that occurs in the expression for S is 1. (In the following the order of the operators ($\mathbf{P}^2$, **M**, **N**, etc.) must be considered; for example, $[\mathbf{N}, \mathbf{P}^2] = 2iP^0\mathbf{P} \neq 0$.)

Since we require that S be a 3-vector under spatial rotations, Eq. (2.7), we must therefore write S in the form

$$\mathbf{S} = A\mathbf{M} + B\mathbf{N} + C\mathbf{P} + D\mathbf{P}\times\mathbf{M} + G\mathbf{P}\times\mathbf{N}, \tag{3.1}$$

where A, B, C, D, and G are three-dimensional scalars. Consideration of the powers of **M**, **N** shows that A, B, D, G are operators depending only on $\mathbf{P}^2$ and E, and C must have the form

$$C = C_1 + (\mathbf{P}\cdot\mathbf{M})C_2 + (\mathbf{P}\cdot\mathbf{N})C_3,$$

where $C_{1,2,3}$ are functions of $\mathbf{P}^2$ and E.

We now require that [from Eq. (2.5)]

$$[\mathbf{S}, E] = 0, \qquad [\mathbf{S}, \mathbf{P}] = 0, \tag{3.2}$$

which implies $B = -C_3\mathbf{P}^2$, and

$$0 = [S^i, P^j] = i(A - GE)\epsilon^{ijk}P^k + i(BE + D\mathbf{P}^2)\delta^{ij} + i(C_3E - D)P^iP^j,$$

therefore

$$A = GE,$$

$$BE + D\mathbf{P}^2 = 0,$$

$$C_3E = D = -BE/\mathbf{P}^2,$$

and so

$$\mathbf{S} = A[\mathbf{M} + E^{-1}(\mathbf{P}\times\mathbf{N})] + B[\mathbf{N} - (\mathbf{P}^2)^{-1}(\mathbf{P}\cdot\mathbf{N})\mathbf{P} - E(\mathbf{P}^2)^{-1}(\mathbf{P}\times\mathbf{M})] + [C_1 + (\mathbf{P}\cdot\mathbf{M})C_2]\mathbf{P} = \mathbf{S}^*.$$

We define (* indicates Hermitian adjoint operator).

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}^* = \mathbf{M} + E^{-1}(\mathbf{P}\times\mathbf{N}),$$

$$\boldsymbol{\beta} = \boldsymbol{\beta}^* = \mathbf{N} - (\mathbf{P}^2)^{-1}(\mathbf{P}\cdot\mathbf{N})\mathbf{P} - E(\mathbf{P}^2)^{-1}(\mathbf{P}\times\mathbf{M}) = E(\mathbf{P}^2)^{-1}(\mathbf{P}\times\boldsymbol{\alpha}),$$

$$\boldsymbol{\gamma} = 2(\mathbf{N} - \mathbf{M})[1 + m^2(\mathbf{P}^2)^{-1}] + (\mathbf{P}\times\mathbf{M})(2E/\mathbf{P}^2) + \mathbf{P}(\mathbf{P}\cdot\mathbf{M})(m^2/\mathbf{P}^4),$$

then

$$\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 0, \qquad \mathbf{P}\cdot\boldsymbol{\beta} = 0, \qquad \mathbf{P}\cdot\boldsymbol{\alpha} = \mathbf{P}\cdot\mathbf{M},$$

$$[\boldsymbol{\alpha}, P^\mu] = [\boldsymbol{\beta}, P^\mu] = 0,$$

$$[\alpha^i, \alpha^j] = i\epsilon^{ijk}[\alpha^k - E^{-2}P^k(\mathbf{P}\cdot\mathbf{M})],$$

$$[\alpha^i, \beta^j] = i[E^{-1}P^jM^i + m^2E^{-1}(\mathbf{P}^2)^{-1}\delta^{ij}(\mathbf{P}\cdot\mathbf{M}) - E(\mathbf{P}^2)^{-1}P^i\alpha^j],$$

$$[\beta^i, \beta^j] = i\epsilon^{ijk}\gamma^k,$$

and therefore

$$\mathbf{S} = A\boldsymbol{\alpha} + B\boldsymbol{\beta} + [C_1 + (\mathbf{P}\cdot\mathbf{M})C_2]\mathbf{P}.$$

We now impose the commutation relations, Eq. (2.6), for the S and obtain

$$A^2[\alpha^k - E^{-2}P^k(\mathbf{P}\cdot\mathbf{M})] + AB[-\beta^k - E^{-1}(\mathbf{P}\times\mathbf{M})^k] + BC_2\mathbf{P}^2\beta^k + AC_2[\mathbf{P}^2\alpha^k - (\boldsymbol{\alpha}\cdot\mathbf{P})P^k] + B^2\gamma^k = A\alpha^k + B\beta^k + (C_1 + (\mathbf{P}\cdot\mathbf{M})C_2)P^k. \tag{3.3}$$

Taking the product with P^k and summing, we have

$$(\mathbf{P}\cdot\mathbf{M})m^2A^2/B + B^2[2(\mathbf{P}\cdot\mathbf{N})(1 + m^2/\mathbf{P}^2) - 2(\mathbf{P}\cdot\mathbf{M}) - m^2(\mathbf{P}\cdot\mathbf{M})/\mathbf{P}^2] = A(\mathbf{P}\cdot\mathbf{M}) + (C_1 + (\mathbf{P}\cdot\mathbf{M})C_2)\mathbf{P}^2.$$

Since A, B only involve $\mathbf{P}^2$ and E, we must therefore have

$$C_1 = B = 0 \qquad A^2m^2/E^2 = A + \mathbf{P}^2C_2.$$

Then Eq. (3.3) shows that

$$A^2[\alpha^k - E^{-2}P^k(\mathbf{P}\cdot\mathbf{M})] + AC_2[\mathbf{P}^2\alpha^k - (\mathbf{P}\cdot\mathbf{M})P^k] = A\alpha^k + C_2(\mathbf{P}\cdot\mathbf{M})P^k$$

[4] T. F. Jordan and N. Mukunda, Phys. Rev. 132, 1842 (1963).

or

$$A^2 + AC_2\mathbf{P}^2 = A,$$
$$A^2 + AC_2E^2 = -C_2E^2,$$

and therefore

$$A = \pm(E/m),$$
$$C_2 = -1/[m(m \pm E)].$$

Our expression for $\mathbf{S}$ is therefore determined by its commutation relations and is

$$\begin{aligned}\mathbf{S} &= (E/m)(\mathbf{M} + E^{-1}(\mathbf{P}\times\mathbf{N})) \\ &\quad - \mathbf{P}((\mathbf{P}\cdot\mathbf{M})/[m(m+E)]), \\ \mathbf{S} &= (-E/m)(\mathbf{M} + E^{-1}(\mathbf{P}\times\mathbf{N})) \\ &\quad - \mathbf{P}((\mathbf{P}\cdot\mathbf{M})/[m(m-E)]).\end{aligned} \tag{3.4}$$

The two expressions are related by $E \to -E$, $\mathbf{N} \to -\mathbf{N}$. (The commutation relations of the ILG are unchanged by this transformation: it corresponds to time reversal. Space inversion corresponds to $\mathbf{P} \to -\mathbf{P}$, $\mathbf{N} \to -\mathbf{N}$ and leaves $\mathbf{S}$ unchanged.)

We have therefore derived a unique (up to time reversal) expression for the intrinsic angular momentum from the commutation rules of the ILG. A heuristic derivation is given in the Appendix.

We have from the structure of $\mathbf{S}$:

$$[S^i, N^j] = i(m+E)^{-1}(P^iS^j - \delta^{ij}(\mathbf{P}\cdot\mathbf{M})),$$

which indicates that $\mathbf{S}$ does not transform as the first three components of a 4-vector under Lorentz transformations, since the condition that a 4-component entity, a^μ, transform as a 4-vector is

$$\begin{aligned}[M^i, a^j] &= i\epsilon^{ijk}a^k, \qquad [M^i, a^0] = 0, \\ [N^i, a^j] &= i\delta^{ij}a^0, \qquad [N^i, a^0] = ia^i.\end{aligned} \tag{3.5}$$

IV. THE POSITION OPERATOR

We now want to write [in agreement with our requirement, Eq. (2.9)], the total angular momentum $\mathbf{M}$ as the sum of two parts, an "orbital" and an "intrinsic":

$$\mathbf{M} = \mathbf{Q}\times\mathbf{P} + \mathbf{S}. \tag{4.1}$$

By insertion of the expression for $\mathbf{S}$ we see that this determines $\mathbf{Q}$ up to a term $\mathbf{P}F$

$$m\mathbf{Q} = -(m+E)^{-1}\mathbf{P}\times\mathbf{M} + \mathbf{N} + m\mathbf{P}F, \tag{4.2}$$

where F is some operator, not necessarily commuting with $\mathbf{P}$.

We want $\mathbf{Q}$ to have the commutation relations indicated in Sec. II and we take these in order. We have from Eqs. (2.1) and (2.2)

$$m[F, E] = i(mE^{-1} - 1), \tag{4.3}$$

whereas Eq. (2.4) implies

$$[S^i, F] = 0. \tag{4.4}$$

Equation (2.3) gives an expression of some complexity. We consider first the equations resulting from Eqs. (4.3) and (4.4). From Eq. (4.3) we see that F must be linear in $\mathbf{N}$ (this follows from the commutation relations of E and $\mathbf{N}$). Since $\mathbf{Q}$ must be a 3-vector under spatial rotations, the operator F must be a (three-dimensional) scalar, and must be formed of $\mathbf{P}^2$, E, $\mathbf{P}\cdot\mathbf{N}$. It however cannot have terms of the type $\mathbf{P}\cdot\mathbf{M}$, since this would contradict Eq. (4.4). So the operator F must be of the form

$$F = X_1(\mathbf{P}^2, E)\cdot(\mathbf{P}\cdot\mathbf{N}) + X_2(\mathbf{P}^2, E).$$

Then Eq. (4.3) determines X_1:

$$X_1 = -(mE)^{-1}(m+E)^{-1}$$

and X_2 is still undetermined.

We now require that $\mathbf{Q}$ be a Hermitian operator: $\mathbf{Q}^* = \mathbf{Q}$, and this will determine the imaginary part of X_2:

$$\begin{aligned}\mathbf{Q} &= -m^{-1}(m+E)^{-1}\mathbf{P}\times\mathbf{M} + m^{-1}\mathbf{N} \\ &\quad + \mathbf{P}[-(mE)^{-1}(m+E)^{-1}\mathbf{P}\cdot\mathbf{N}] - i(2E^2)^{-1}\mathbf{P} + \mathbf{P}X_2^R,\end{aligned}$$

where X_2^R is real and undetermined.

We may rewrite Q as

$$\begin{aligned}\mathbf{Q} &= -E^{-1}(m+E)^{-1}(\mathbf{P}\times\mathbf{S}) + E^{-1}\mathbf{N} \\ &\quad - (i/2E^2)\mathbf{P} + \mathbf{P}X_2^R.\end{aligned} \tag{4.5}$$

We have not as yet considered the commutation relations for the components of Q, Eq. (2.3).

We may avoid tedious algebra by writing:

$$\mathbf{N} = E\mathbf{Q} + (m+E)^{-1}(\mathbf{P}\times\mathbf{S}) + i(2E)^{-1}\mathbf{P} - E\mathbf{P}X_2^R$$

and substituting this in the commutation relations for $\mathbf{N}$. We see then that Eq. (1.1) can only be true if $[Q^i, Q^j] = 0$ and X_2^R is arbitrary.

Thus we have determined a position operator with all the desired properties, and it is unique up to an additive term: $\mathbf{P}X_2^R$. It is simplest to set this term equal to zero: $X_2^R = 0$.

We have

$$\begin{aligned}[Q^i, N^j] &= iE^{-1}Q^iP^j + im(m+E)^{-1}E^{-1}S^k \\ &\quad + iE^{-1}(m+E)^{-2}(\mathbf{P}\cdot\mathbf{S})P^k \\ &\quad - i(m+E)^{-2}E(\mathbf{P}\times\mathbf{S})^iP^j \\ &\quad + (2E^3)^{-1}(P^iP^j + E^2\delta^{ij})\end{aligned} \tag{4.6}$$

showing that Q^i does not transform as the first three components of a 4-vector.

V. SPIN AND POSITION IN A GENERAL REPRESENTATION OF ILG

In general, the generators $\mathbf{M}$, $\mathbf{N}$ are expressed as follows (in a representation space for which P_μ is diagonal)

$$\mathbf{M} = -i\mathbf{P}\times\nabla_P + \mathbf{M}_S,$$
$$\mathbf{N}_S = iE\nabla_P + \mathbf{N}_S,$$

where $\mathbf{M}_s$, $\mathbf{N}_s$ represent the "spin" part of the representation.

Substituting in our expressions for $\mathbf{S}$ and $\mathbf{Q}$ we have

$$m\mathbf{S} = -(m+E)^{-1}\mathbf{P}\times(\mathbf{P}\times\mathbf{M}_S) + m\mathbf{M}_S + \mathbf{P}\times\mathbf{N}_S,$$
$$= E\mathbf{M}_S + \mathbf{P}\times\mathbf{N}_S - (m+E)^{-1}\mathbf{P}(\mathbf{P}\cdot\mathbf{M}_S), \quad (5.1)$$

$$\mathbf{Q} = (i\nabla_P - (i/2E^2)\mathbf{P}) + (1/m)[\mathbf{N}_S - (m+E)^{-1}\mathbf{P}\times\mathbf{M}_S - E^{-1}(m+E)^{-1}\mathbf{P}(\mathbf{P}\cdot\mathbf{N}_s)]. \quad (5.2)$$

The action of the generator on a state function $|p\rangle$ is

$$|\mathbf{p}\rangle \to (1 - i\mathbf{M}\cdot\boldsymbol{\theta})\,|\mathbf{p}\rangle,$$
$$= (1 - \boldsymbol{\theta}\times\mathbf{P}\cdot\nabla_P - i\mathbf{M}_S\cdot\boldsymbol{\theta})\,|\mathbf{p}\rangle,$$
$$|\mathbf{p}\rangle \to (1 - i\mathbf{N}\cdot\boldsymbol{\theta})\,|\mathbf{p}\rangle = (1 + E\boldsymbol{\theta}\cdot\nabla_P - i\mathbf{N}_S\cdot\boldsymbol{\theta})\,|\mathbf{p}\rangle,$$

for an infinitesimal spatial rotation and Lorentz transformation, respectively.

If we go to a new representation of ILG:

$$|\mathbf{p}\rangle' \equiv U(\mathbf{P})\,|\mathbf{p}\rangle,$$

then we have [$U(\mathbf{p})$ = unitary]

$$1 - i\mathbf{M}'\cdot\boldsymbol{\theta} = U(P)[1 - i\mathbf{M}\cdot\boldsymbol{\theta}]U^{-1}(P'),$$
$$1 - i\mathbf{N}'\cdot\boldsymbol{\theta} = U(P)[1 - i\mathbf{N}\cdot\boldsymbol{\theta}]U^{-1}(P'),$$

and therefore

$$\boldsymbol{\theta}\cdot\mathbf{M}'_S = i[U(\mathbf{P})U^{-1}(\mathbf{P} - \boldsymbol{\theta}\times\mathbf{P}) - 1] + U(\mathbf{P})(\mathbf{M}_S\cdot\boldsymbol{\theta})U^{-1}(\mathbf{P}),$$
$$\boldsymbol{\theta}\cdot\mathbf{N}'_S = i[U(\mathbf{P})U^{-1}(\mathbf{P} + \boldsymbol{\theta}E) - 1] + U(\mathbf{P})(\mathbf{N}_s\cdot\boldsymbol{\theta})U^{-1}(\mathbf{P}).$$

Also

$$\mathbf{Q}' = UQU^{-1}; \qquad \mathbf{S}' = USU^{-1}.$$

VI. EXPLICIT FORMS FOR INTRINSIC ANGULAR MOMENTUM AND POSITION OPERATORS IN SOME SPECIAL REPRESENTATIONS OF THE ILG

Canonical Representation

In the canonical representation,[5] where we have (for spin s)

$$\mathbf{M}_s = \mathbf{J}, \qquad \mathbf{N}_s = -(m+E)^{-1}(\mathbf{J}\times\mathbf{P}) \quad (6.1)$$

[the J's are the usual $(2s+1)$ by $(2s+1)$ dimensional angular momentum matrices], the intrinsic angular momentum and position operators are

$$\mathbf{S} = \mathbf{J}; \qquad \mathbf{Q} = i\nabla_p - i(2E^2)^{-1}\mathbf{P}.$$

Helicity Representation

In the helicity representation [$\mathbf{e}_3$ is the unit vector along the 3-axis, $\mathbf{n}$ is a unit vector in the direction of $\mathbf{P}$, and $\mathbf{e}_2 \equiv (\mathbf{e}_3\times\mathbf{n})/(1-n_3^2)^{\frac{1}{2}}$].

$$\mathbf{M}_s = (\mathbf{e}_3 + \mathbf{n})J_3/(1+n_3)$$
$$\mathbf{N}_s = [(1-n_3^2)^{-\frac{1}{2}}][-\mathbf{e}_3 mJ_2/P + \mathbf{n}n_3 mJ_2/P + \mathbf{e}_2\{-m(\mathbf{J}\cdot\mathbf{n}) + n_3 mJ_2 - E(1-n_3)J_3\}P^{-1}].$$

The helicity representation is useful, because the helicity is diagonal in the usual representation for the J_3 matrix:

$$(\mathbf{M}\cdot\mathbf{P}) = J_3\,|\mathbf{P}| \equiv J_3 P.$$

It is convenient to use the set of axes $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ instead of $(\mathbf{n}, \mathbf{e}_2, \mathbf{e}_3)$, where $\mathbf{e}_1 = \mathbf{e}_2\times\mathbf{e}_3$. Then

$$\mathbf{N}_s = -\frac{E}{P}\left(\frac{1-n_3}{1+n_3}\right)^{\frac{1}{2}}J_3\mathbf{e}_2 + \frac{m}{P}\left(\mathbf{e}_1 J_2 - \mathbf{e}_2 J_1 - \left(\frac{1-n_3}{1+n_3}\right)^{\frac{1}{2}}\mathbf{e}_3 J_2\right),$$
$$\mathbf{M}_s = \left(\mathbf{e}_3 + \mathbf{e}_1\left(\frac{1-n_3}{1+n_3}\right)^{\frac{1}{2}}\right)J_3,$$
$$\mathbf{S} = \mathbf{e}_1[(\mathbf{e}_3\cdot\mathbf{n})J_1 + (1-n_3^2)^{\frac{1}{2}}J_3] + \mathbf{e}_2 J_2 + \mathbf{e}_3[(\mathbf{e}_3\cdot n)J_3 - J_1(1-n_3^2)^{\frac{1}{2}}],$$
$$\mathbf{S}^2 = J_1^2 + J_2^2 + J_3^2,$$

and the position operator is

$$\mathbf{Q} = i\nabla_p + \left[\mathbf{e}_1 J_2\left[n_3 - \frac{m}{E}(1+n_3)\right] - \mathbf{e}_2\left(J_1 + J_3\left(\frac{1-n_3}{1+n_3}\right)^{\frac{1}{2}}\right) - \mathbf{e}_3(1-n_3)^{\frac{1}{2}}J_3\left(1 - \frac{mn_3}{E(1+n_3)}\right)\right]\frac{1}{P} - \frac{i}{2E^2}\mathbf{P}.$$

[5] L. L. Foldy, Phys. Rev. 102, 568 (1956).

VII. RELATION TO THE FW EXPRESSIONS FOR SPIN AND POSITION

We define a four-component entity for spin-$\frac{1}{2}$ by defining

$$\chi \equiv \begin{cases} |\mathbf{p}, \omega; |m|\rangle \\ |\mathbf{p}, -\omega; -|m|\rangle \end{cases} \tag{7.1}$$

in terms of the two-component state vectors: $|\mathbf{p}, p^0; m\rangle$, where $\omega = +(|m|^2 + \mathbf{p}^2)^{\frac{1}{2}}$

Then we have for the canonical representation of the ILG [see Eq. (6.1)]:

$$\begin{aligned} \mathbf{M}_s &= \mathbf{J}, \\ \mathbf{N}_s &= -[\mathbf{J} \times \mathbf{P}/(|m| + \omega)]\rho_3. \end{aligned} \tag{7.1a}$$

The "wave equation" obeyed by the four-component entity, Eq. (7.1), is then

$$E\chi = \omega\rho_3\chi. \tag{7.2}$$

The FW transformation[2] is given in this approach by

$$\begin{aligned} U(p) &\equiv 2^{-\frac{1}{2}}[(a + b\boldsymbol{\sigma}\cdot\mathbf{p}) - i\rho_2(a - b\boldsymbol{\sigma}\cdot\mathbf{p})] \\ &= \exp(i\lambda\rho_2\boldsymbol{\sigma}\cdot\mathbf{p}) \exp(-i\rho_2\pi/4), \end{aligned}$$

where

$$\begin{aligned} a \equiv [(\omega + |m|)/2\omega]^{\frac{1}{2}}, \quad b \equiv 1/[2\omega(\omega + |m|)]^{\frac{1}{2}}, \\ \cos\lambda p = a, \qquad \cos 2\lambda p = |m|/\omega, \end{aligned} \tag{7.3}$$

and we define

$$\psi(\mathbf{p}) = U(\mathbf{p})\cdot\chi(\mathbf{p}).$$

The wave equation satisfied by ψ is found by transforming Eq. (7.2) and is

$$E\psi = (|m|\ \rho_1 + \rho_3\boldsymbol{\sigma}\cdot\mathbf{p})\psi, \tag{7.3a}$$

which is the Dirac equation when considered in the context of a local field theory.

To determine the operators $\mathbf{M}_s'$, $\mathbf{N}_s'$ which generate the ILG when acting on ψ, we must use

$$\begin{aligned} &U(\mathbf{p})U^{-1}(\mathbf{p} - \boldsymbol{\theta}\times\mathbf{p}) \\ &U(\mathbf{p})^{-1}(\mathbf{p} + \boldsymbol{\theta}E). \end{aligned} \tag{7.4}$$

For spatial rotations we obtain

$$\mathbf{M}_s' = \tfrac{1}{2}\boldsymbol{\sigma} \tag{7.5a}$$

and for Lorentz transformations we have

$$\begin{aligned} \mathbf{N}_s' &= \frac{-\rho_1(\boldsymbol{\sigma}\times\mathbf{p})}{2(|m| + \omega)} - \frac{(\boldsymbol{\sigma}\times\mathbf{p})E}{2\omega(|m| + \omega)} + \frac{\rho_2 E\boldsymbol{\sigma}}{2\omega} \\ &\quad - \frac{\rho_2(\boldsymbol{\sigma}\cdot\mathbf{p})E}{2\omega^2(|m| + \omega)}\mathbf{P}. \end{aligned}$$

$$= -\left(\frac{1}{2} i\rho_3\boldsymbol{\sigma} - i\frac{1}{2E}\mathbf{P}\right) + \left(\frac{1}{2E} i\rho_3\boldsymbol{\sigma} - \frac{\rho_1\boldsymbol{\sigma}\times\mathbf{P}}{2E(|m| + \omega)}\right)(E - |m|\ \rho_1 - \rho_3\boldsymbol{\sigma}\cdot\mathbf{P}). \tag{7.5b}$$

Operating on the wavefunction ψ we see that

$$\mathbf{N}_s'\psi = [-\tfrac{1}{2}i\rho_3\boldsymbol{\sigma} + (i/2E)\mathbf{P}]\psi \tag{7.6}$$

and the operator N_s' is Hermitian: $\mathbf{N}_s'^* = \mathbf{N}_s'$.

By defining a new ψ_1 by

$$\psi_1 \equiv \omega^{\frac{1}{2}}\psi, \tag{7.7}$$

we can eliminate the second term in Eq. (7.6) and have

$$\mathbf{N}_s''\psi_1 = (-\tfrac{1}{2}i\rho_3\boldsymbol{\sigma})\psi_1, \tag{7.8}$$

which is the usual Dirac Lorentz transformation.

This N_s'' is however not Hermitian, and the corresponding $(1 - i\mathbf{N}_s''\cdot\boldsymbol{\theta})$ is therefore not a unitary transformation. This is so since Eq. (7.7) is not unitary.

From the canonical representation, Eq. (6.1), we have

$$\mathbf{Q}' = U(\mathbf{P})[i\nabla_p - (i/2E^2)\mathbf{P}]U^{-1}(\mathbf{P}).$$

We may write this as (for some infinitesimal quantity θ)

$$\begin{aligned} \boldsymbol{\theta}\cdot\mathbf{Q}' &= i[U(p)\cdot E^{-1}\cdot U^{-1}(\mathbf{P} + \boldsymbol{\theta}E) - E^{-1}] \\ &\quad - (i/2E^2)\mathbf{P}\cdot\boldsymbol{\theta}. \end{aligned}$$

Comparing this with Eq. (7.5), we have

$$\mathbf{Q}' = \rho_2\frac{1}{2\omega}\boldsymbol{\sigma} - \frac{\omega\boldsymbol{\sigma}\times\mathbf{P} + \rho_2\mathbf{P}(\boldsymbol{\sigma}\cdot\mathbf{P})}{2\omega^2(|m| + \omega)} + i\nabla_p - \frac{i}{2E^2}\mathbf{P}.$$

Likewise from Eq. (7.1a)

$$\begin{aligned} \mathbf{S}' &= U(\mathbf{P})\tfrac{1}{2}\boldsymbol{\sigma}U^{-1}(\mathbf{P}) \\ &= -\frac{\rho_2\boldsymbol{\sigma}\times\mathbf{P}}{2\omega} + P\frac{\boldsymbol{\sigma}\cdot\mathbf{P}}{2\omega(|m| + \omega)} + \frac{|m|}{2\omega}\boldsymbol{\sigma}. \end{aligned}$$

These are the usual spin intrinsic angular momentum and position generators of the FW transformation.

They cannot be obtained directly from Eqs. (5.1) and (5.2) by substituting Eq. (7.8) since it is not Hermitian. They can be obtained by substituting Eqs. (7.5a) and (7.5b) directly in Eqs. (5.1) and (5.2), and making use of the fact that the mass operator m is such that $m\chi = |m|\ \rho_3\chi$ and $m\psi = |m|\ E/\omega\psi$ plus the wave equation, Eq. (7.3a).

VIII. CONCLUSIONS

We have demonstrated that from the structure of the operators which define the transformations of

the inhomogeneous Lorentz group, there follows a unique intrinsic spin and position operator, with the properties usually accepted as desirable.

In terms of the generators of the ILG these operators are

$$m\mathbf{S} = E(\mathbf{M} + E^{-1}\mathbf{P}\times\mathbf{N}) - \mathbf{P}(\mathbf{P}\cdot\mathbf{M})(m+E)^{-1},$$

$$m\mathbf{Q} = -(m+E)^{-1}\mathbf{P}\times\mathbf{M} + \mathbf{N} - \mathbf{P}E^{-1} \times (m+E)^{-1}(\mathbf{P}\cdot\mathbf{N}) - im(2E^2)^{-1}\mathbf{P} + im\nabla_P.$$

APPENDIX

The 4-vector y^μ is defined as follows ($\epsilon^{0123} = -\epsilon_{0123} = 1$)

$$y^\mu \equiv \tfrac{1}{2}\epsilon^{\mu\alpha\beta\sigma}M_{\alpha\beta}P_\sigma$$

and therefore

$$y^0 = -(\mathbf{P}\cdot\mathbf{M}),$$

$$\mathbf{y} = -E\mathbf{M} - \mathbf{P}\times\mathbf{N}.$$

The eigenvalues of the operators $p^\mu p_\mu$ and $y^\mu y_\mu$ define the representation space in which the state vectors are defined.

In the rest system, $\mathbf{p}' = 0$, (denoted by primes), $\mathbf{y}' = -m\mathbf{M}$.

We now heuristically define the intrinsic angular momentum operator as $\mathbf{M}$ in the rest system, since it is whatever angular momentum is left over after the motion has been subtracted.

Thus we define $\mathbf{S} \equiv -(1/m)\mathbf{y}'$.

Now an arbitrary 3-vector, $\mathbf{a}$, in a system with momentum $\mathbf{p}$, has the value $\mathbf{a}'$ in the rest system ($\mathbf{p}' = 0$), where

$$\mathbf{a}' = \mathbf{a} - \mathbf{p}[(ap) + ma^0]/m(m+E).$$

Therefore

$$\mathbf{y}' = -E\mathbf{M} - \mathbf{P}\times\mathbf{N} - \mathbf{P}[-(\mathbf{P}\cdot\mathbf{M})]/(m+E) \equiv -m\mathbf{S}$$

and we can therefore define the intrinsic angular momentum as

$$\mathbf{S} = m^{-1}E\mathbf{M} - m^{-1}(m+E)^{-1}\mathbf{P}(\mathbf{P}\cdot\mathbf{M}) + m^{-1}(\mathbf{P}\times\mathbf{N})$$

for any system.

This expression is that derived by a more formal argument in Sec. III.

We have for the value of the "spin invariant"

$$y^\mu y_\mu = -(p^\mu p_\mu)\mathbf{S}^2 = -m^2\mathbf{S}^2.$$

On the Number of Bound States of a Given Interaction

GIAN CARLO GHIRARDI AND ALBERTO RIMINI
Istituto di Fisica dell'Università, Trieste, Italy
(Received 21 July 1964)

A method for deriving bounds to the number of bound states of a given interaction is built up. The class of interactions for which the method works includes the nonlocal interactions besides the local ones. Problems with many channels and spin-dependent interactions can also be treated. Some bounds are explicitly given.

1. INTRODUCTION

IT is known that the system consisting of a particle in a local field of force admits only a finite number of bound states provided that an appropriate integral involving the potential is convergent. In fact Bargmann[1] has derived in the case of a rotationally invariant potential $V(r)$, the inequality

$$n_l < \frac{1}{2l+1} \int_0^\infty dr\, r\, |V(r)|, \tag{1}$$

where n_l is the number of bound states in the lth partial wave (for fixed magnetic quantum number). Schwinger,[2] together with a new derivation of the result of Bargmann, has given for the total number N of bound states, in the general case of a local potential $V(\mathbf{r})$, the upper bound

$$N < \frac{1}{(4\pi)^2} \iint d^3r\, d^3r' \frac{V^{(-)}(\mathbf{r}) V^{(-)}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^2}. \tag{2}$$

Here $V^{(-)}(\mathbf{r})$ is the negative part of $V(\mathbf{r})$ defined by

$$V^{(-)}(\mathbf{r}) = V(\mathbf{r}) \quad \text{for } \mathbf{r} \text{ such that } V(\mathbf{r}) < 0,$$
$$0 \quad \text{for } \mathbf{r} \text{ such that } V(\mathbf{r}) \geq 0.$$

Moreover, Schwinger has shown how to obtain bounds in the case of the presence of the tensor interaction.

The above mentioned methods are based on the introduction, beside the given interaction V, of a comparison interaction W which is negative definite and satisfies the inequality

$$(\varphi, W\varphi) \leq (\varphi, V\varphi)$$

for any state vector φ. The comparison interaction W has certainly at least as many bound states as V, and is introduced in order to deal with an interaction such that any increase of its strength causes a lowering of the energies of the bound states and does not lessen their number. The determination of the comparison interaction, requires in general, the diagonalization of the given interaction V. This makes such methods not suitable for the case of interactions which are not given in diagonal form. This is the case for nonlocal forces or for many-channel problems. Spin-dependent interactions are also difficult to deal with if they give rise to matrices of high rank.

In this paper a method is builded up for obtaining bounds on the number of bound states which avoids the introduction of the negative definite comparison interaction W and works directly on V. The class of interactions for which the method works is large enough to include all the above-mentioned difficult cases besides the local one. The method is based on the fact that, under fairly general conditions, the energies of the bound states decrease when the strength of the interaction is increased, also if the interaction itself is not a negative definite operator.

2. DERIVATION OF THE BOUNDS ON THE NUMBER OF BOUND STATES

We first consider the case of a spinless particle. Let the representative of the interaction be $V(\mathbf{p}, \mathbf{p}')$ [which becomes $V(\mathbf{p} - \mathbf{p}')$ in the local case]. Then, in suitable units, the Schrödinger equation for negative energy states reads

$$p^2\psi(\mathbf{p}) + \int d^3p' V(\mathbf{p}, \mathbf{p}')\psi(\mathbf{p}') = -k^2\psi(\mathbf{p}). \tag{3}$$

We introduce the equation

$$p^2\psi(\mathbf{p}) + \lambda \int d^3p' V(\mathbf{p}, \mathbf{p}')\psi(\mathbf{p}') = -k^2\psi(\mathbf{p})$$

which, defining $X(\mathbf{p}) = (p^2 + k^2)^{\frac{1}{2}}\psi(\mathbf{p})$, becomes

$$\lambda^{-1}(k^2)X(k^2, \mathbf{p}) = -\int d^3p \frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}} X(k^2, \mathbf{p}'). \tag{4}$$

The solutions of this equation for $\lambda = 1$ correspond to the solutions of the original equation (3). We

[1] V. Bargmann, Proc. Natl. Acad. Sci. U. S. 38, 961 (1952).
[2] J. Schwinger, Proc. Natl. Acad. Sci. U. S. 47, 122 (1961).

assume that the integral

$$A^2(k^2) = \iint d^3p\, d^3p' \frac{|V(\mathbf{p}, \mathbf{p}')|^2}{(p^2 + k^2)(p'^2 + k^2)} \tag{5}$$

converges. Clearly it is a decreasing function of k^2.

Equation (4) is an integral equation with a square integrable Hermitian kernel which depends on the parameter k^2. Its eigenvalues lie in the interval $(-A(k^2), A(k^2))$ and form a denumerable sequence $\lambda_i^{-1}(k^2)$ whose only limit point is zero. It is proved in the Appendix that the eigenvalues and the eigenfunctions of Eq. (4), are continuous functions of k^2. In order to investigate the movement of the eigenvalues $\lambda_i^{-1}(k^2)$ when k^2 is varied, let us take into account two values, k^2 and k'^2, of the parameter k^2. From Eq. (4) one has

$$\lambda_i^{-1}(k^2) \int d^3p X_i^*(k^2, \mathbf{p}) \left(\frac{p^2 + k'^2}{p^2 + k^2}\right)^{\frac{1}{2}} X_i(k'^2, \mathbf{p})$$

$$= -\iint d^3p\, d^3p' X_i^*(k^2, \mathbf{p}) \times \frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}} \left(\frac{p'^2 + k'^2}{p'^2 + k^2}\right)^{\frac{1}{2}} X_i(k'^2, \mathbf{p}')$$

$$\lambda_i^{-1}(k'^2) \int d^3p X_i^*(k^2, \mathbf{p}) \left(\frac{p^2 + k'^2}{p^2 + k^2}\right)^{\frac{1}{2}} X_i(k'^2, \mathbf{p})$$

$$= -\iint d^3p\, d^3p' X_i^*(k^2, \mathbf{p}) \left(\frac{p^2 + k'^2}{p^2 + k^2}\right)^{\frac{1}{2}} \times \frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k'^2)(p'^2 + k'^2)]^{\frac{1}{2}}} X_i(k'^2, \mathbf{p}').$$

Subtraction of the second of these equations from the first one gives

$$[\lambda_i^{-1}(k^2) - \lambda_i^{-1}(k'^2)] \times \int d^3p X_i^*(k^2, \mathbf{p}) \left(\frac{p^2 + k'^2}{p^2 + k^2}\right)^{\frac{1}{2}} X_i(k'^2, \mathbf{p})$$

$$= -\iint d^3p\, d^3p' X_i^*(k^2, \mathbf{p}) \frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}} \times \frac{k'^2 - k^2}{[(p'^2 + k^2)(p'^2 + k'^2)]^{\frac{1}{2}}} X_i(k'^2, \mathbf{p}')$$

$$= (k'^2 - k^2)\lambda_i^{-1}(k^2) \int d^3p \frac{X_i^*(k^2, \mathbf{p}) X_i(k'^2, \mathbf{p})}{[(p^2 + k^2)(p^2 + k'^2)]^{\frac{1}{2}}}.$$

Since $\lambda_i^{-1}(k^2)$ and $X_i(k^2, \mathbf{p})$ are continuous functions of k^2, one has, in the limit as $k'^2 \to k^2$

$$-\frac{d\lambda_i^{-1}(k^2)}{d(k^2)} = \lambda_i^{-1}(k^2) \int d^3p \frac{|X_i(k^2, \mathbf{p})|^2}{p^2 + k^2}. \tag{6}$$

The eigenfunctions $X_i(k^2, \mathbf{p})$ have been assumed to be normalized. Eq. (6) shows that the positive eigenvalues are increasing functions of $-k^2$. Note that this is equivalent to say that the increase of the strength of the interaction lowers the energies of the bound state. This is proved, irrespectively of their negative definite character, for the local or nonlocal interactions for which the integral (5) exists.

If $A(\bar{k}^2) = 1$, there is no bound state of energy less than $-\bar{k}^2$ since $\lambda_i^{-1}(k^2) < 1$ for $-k^2 < -\bar{k}^2$. When $-k^2$ is increased from $-\bar{k}^2$ up to zero, the positive $\lambda_i^{-1}(k^2)$ increase (continuously), as it has just been shown. Therefore it follows that the number of bound states of Eq. (3) whose energy is less than an arbitrary fixed $-\bar{k}^2$ is equal to the number of eigenvalues $\lambda_i^{-1}(\bar{k}^2)$ which lie above 1. This number is less than $\sum_i \lambda_i^{-2}(\bar{k}^2)$, i.e., the trace of the iterated kernel of Eq. (4). Therefore, for the total number N of bound states of an interaction whose momentum representative is $V(\mathbf{p}, \mathbf{p}')$ one has the bound

$$N < \iint d^3p\, d^3p' \frac{|V(\mathbf{p}, \mathbf{p}')|^2}{p^2 p'^2}. \tag{7}$$

Defining the Green operator $G(-k^2)$ as the operator whose momentum representation is

$$\delta^3(\mathbf{p} - \mathbf{p}')/(p^2 + k^2)$$

the inequality (7) can also be written

$$N < \mathrm{Tr}\,\{[G(0)V]^2\}.$$

It is easy to construct interactions which approach the above limit as close as one wishes. In fact, taking into consideration a nonlocal attractive separable potential $V(\mathbf{p}, \mathbf{p}') = -v(\mathbf{p})v(\mathbf{p}')$, the bound (7) becomes

$$N < \left[\int d^3p \frac{v^2(\mathbf{p})}{p^2}\right]^2$$

and, on the other hand, the condition

$$\int d^3p \frac{v^2(\mathbf{p})}{p^2} > 1$$

is sufficient for the existence of the unique possible bound state of such an interaction.[3]

For a local potential $V(\mathbf{r})$ the bound (7) becomes, in the coordinate representation

$$N < \frac{1}{(4\pi)^2} \int d^3r\, d^3r' \frac{V(\mathbf{r})V(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^2}. \tag{8}$$

This bound differs from the bound (2) given by Schwinger in that it involves the potential itself instead of its negative part $V^{(-)}(\mathbf{r})$. Note that the bound (8) may be better or worse than the bound (2)

[3] G. C. Ghirardi and A. Rimini, J. Math. Phys. 5, 722 (1961).

depending on the specific form of the potential. This is easily seen by considering a potential

$$V(\mathbf{r}) = \alpha V^{(+)}(\mathbf{r}) + V^{(-)}(\mathbf{r}),$$

where $V^{(+)}(\mathbf{r}) \geq 0$ and $V^{(-)}(\mathbf{r}) \leq 0$ and moreover for any value of $\mathbf{r}$ at least one of them is equal to zero. Then the bound (8) becomes

$$N < \frac{1}{(4\pi)^2}\alpha^2 \iint d^3r\, d^3r' \frac{V^{(+)}(\mathbf{r})V^{(+)}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^2} + 2\alpha \frac{1}{(4\pi)^2} \iint d^3r\, d^3r' \frac{V^{(+)}(\mathbf{r})V^{(-)}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^2} + \frac{1}{(4\pi)^2} \iint d^3r\, d^3r' \frac{V^{(-)}(\mathbf{r})V^{(-)}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^2}. \quad (9)$$

Since the coefficients of α^2 and α have opposite signs the contribution from the first two terms in Eq. (9) can be made positive or negative.

3. BOUNDS IN SOME FURTHER SPECIAL CASES

In this section we shall give explicitly the bounds obtained, following essentially the method of Sec. 2, in some further specific cases. These are the case of the lth partial wave for a rotationally invariant spin-independent potential and the case of a many two-body channels problem. Finally, as an example of the application of the method of Sec. 2 to a spin-dependent case, we shall take into account a central plus tensor interaction.

If one assumes a rotationally invariant spin-independent interaction $V(p, p', \cos \widehat{\mathbf{pp}'})$ it is possible to reduce the problem in the eigenmanifolds of the orbital angular momentum operator. The above procedure, applied to the eigenvalue equation for the lth partial wave

$$p^2\psi_l(p) + \int_0^\infty p'^2\, dp' V_l(p, p')\psi_l(p') = -k^2\psi_l(p),$$

gives for the number n_l of the l-wave bound states, a part from the $(2l+1)$-fold degeneracy

$$n_l < \int_0^\infty \int_0^\infty dp\, dp'\, |V_l(p, p')|^2. \quad (10)$$

In the preceding formulas we have put

$$V_l(p, p') = 2\pi \int_{-1}^{+1} d\cos\theta\, P_l(\cos\theta) V(p, p', \cos\theta).$$

In the local case, the above bound becomes

$$n_l < \int_0^\infty \int_0^\infty dr\, dr' V(r)V(r')g_l^2(r, r'), \quad (11)$$

where

$$g_l(r, r') = \frac{1}{2l+1} r_<^{l+1} r_>^{-l}. \quad (12)$$

A rather crude estimate shows that the integral appearing in (11) is majorized by $[2/(2l+1)^2]\,[\int_0^\infty dr\, r\,|V(r)|]^2$.

If one is concerned with a problem with many two-body channels, the Schrödinger equation reads

$$\left(\frac{p^2}{2\mu_\alpha} + E_\alpha\right)\psi_\alpha(\mathbf{p}) + \sum_\beta \int d^3p' V_{\alpha\beta}(\mathbf{p}, \mathbf{p}')\psi_\beta(\mathbf{p}') = E\psi_\alpha(\mathbf{p}).$$

In this equation $\mathbf{p}$ is the relative momentum of the two particles, E_α is the energy of the αth threshold, μ_α is the reduced mass in channel α. The evident generalization of the above arguments gives

$$N < \sum_{\alpha\beta} \iint d^3p\, d^3p' \times \frac{|V_{\alpha\beta}(\mathbf{p}, \mathbf{p}')|^2}{(p^2/2\mu_\alpha + E_\alpha - E_1)(p^2/2\mu_\beta + E_\beta - E_1)} \quad (13)$$

for the total number N of bound states of energy less than E_1.

We come now to the case of a central plus tensor interaction

$$V_c(r) + V_T(r)S_{12}$$

$$S_{12} = 3\frac{(\boldsymbol{\sigma}_1\cdot\mathbf{r})\cdot(\boldsymbol{\sigma}_2\cdot\mathbf{r})}{r^2} - (\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2).$$

The operator S_{12} is zero when acting on a singlet ($S = 0$) state. In the triplet space S_{12} has the matrix representative

	$L = J-1$	$L = J$	$L = J+1$
$L = J-1$	$\frac{-(J-1)}{2J+1}$	0	$\frac{6[J(J+1)]^{\frac{1}{2}}}{2J+1}$
$L = J$	0	2	0
$L = J+1$	$\frac{6[J(J+1)]^{\frac{1}{2}}}{2J+1}$	0	$\frac{-(J+2)}{2J+1}$

(14)

and does not couple the state $S = 1$, $L = J$ with the remaining triplet states $S = 1$, $L = J \pm 1$. One obtains immediately, for the numbers of bound states with J, J_z fixed and, respectively, $S = 0$, $S = 1$ and parity $(-1)^J$, and $S = 1$ and parity $(-1)^{J+1}$

$$n^0_{JJ_z} < \int_0^\infty \int_0^\infty dr\, dr' V_c(r)V_c(r')g_J^2(r, r'),$$

$$n^L_{JJ_z} < \int_0^\infty \int_0^\infty dr\, dr'[V_c(r) + 2V_T(r)] \times [V_c(r') + 2V_T(r')]g_J^2(r, r'), \quad (15)$$

$$n^{\pm}_{JJ_z} < \int_0^\infty \int_0^\infty dr\, dr'$$
$$\times \sum_{LL'} [V_c(r)\delta_{LL'} + V_T(r)S_{12}(L, L')]$$
$$\times [V_c(r')\delta_{L'L} + V_T(r')S_{12}(L', L)]g_L(r, r')g_{L'}(r', r).$$

The functions $g_L(r, r')$ are defined by Eq. (12). The matrix $S_{12}(L, L')$ is the 2×2 matrix obtained from Eq. (14) by taking out the middle row and column. The third inequality reads, in displayed form,

$$n^{\pm}_{JJ_z} < \iint dr\, dr' \Big\{ g^2_{J-1}(r, r') \Big[V_c(r)V_c(r') - \frac{2(J-1)}{2J+1} [V_c(r)V_T(r') + V_c(r')V_T(r)] + \frac{4(J-1)^2}{(2J+1)^2} V_T(r)V_T(r') \Big] + g^2_{J+1}(r, r') \times \Big[V_c(r)V_c(r') - \frac{2(J+2)}{2J+1} [V_c(r)V_T(r') + V_c(r')V_T(r)] + \frac{4(J+2)^2}{(2J+1)^2} V_T(r)V_T(r') \Big] + \frac{72J(J+1)}{(2J-1)(2J+3)} g^2_J(r, r')V_T(r)V_T(r') \Big\}.$$

The consideration following Eq. (12) shows that no bound state can be present for sufficiently high values of J.

ACKNOWLEDGMENT

We thank Professor L. Fonda for his interest in our work.

APPENDIX

We show here the continuity with respect to the parameter k^2 of the eigenvalues and eigenfunctions of Eq. (4). Writing

$$N(k^2, \mathbf{p}, \mathbf{p}') = -\frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}},$$

one has for any square integrable function φ,

$$|(\varphi, [N(k'^2) - N(k^2)]\varphi)| \leq ||\varphi||^2 \cdot ||N(k'^2) - N(k^2)|| \leq ||\varphi||^2\, ||N(k^2)||\, \frac{|k'^2 - k^2|}{k'^2},$$

with obvious meaning of the symbols. This means that

$$\lim_{k'^2 \to k^2} (\varphi, [N(k'^2) - N(k^2)]\varphi) = 0, \qquad \text{(A1)}$$

even if φ depends on k'^2, provided it is of bounded norm.

Let $\lambda^{-1}_{1+}(k^2)$, $X_{1+}(k^2)$ be the greatest of the positive eigenvalues and the corresponding normalized eigenfunction of the kernel $N(k^2)$. Then one has

$$\lambda^{-1}_{1+}(k^2) = (X_{1+}(k^2), N(k^2)X_{1+}(k^2)) = \sup (\varphi, N(k^2)\varphi).$$

In this equation "sup" denotes the least upper bound and φ runs over the whole set of the normalized square integrable functions. Introducing the step function $\theta(x)$ which is 0 for $x < 0$ and 1 for $x \geq 0$, it results

$$\begin{aligned} |\lambda^{-1}_{1+}(k'^2) - \lambda^{-1}_{1+}(k^2)| &= \theta(\lambda^{-1}_{1+}(k^2) - \lambda^{-1}_{1+}(k'^2)) \\ &\quad \times [(X_{1+}(k^2), N(k^2)X_{1+}(k^2)) \\ &\quad - (X_{1+}(k'^2), N(k'^2)X_{1+}(k'^2))] \\ &\quad + \theta(\lambda^{-1}_{1+}(k'^2) - \lambda^{-1}_{1+}(k^2)) \\ &\quad \times [(X_{1+}(k'^2), N(k'^2)X_{1+}(k'^2)) \\ &\quad - (X_{1+}(k^2), N(k^2)X_{1+}(k^2))] \\ &\leq \theta(\lambda^{-1}_{1+}(k^2) - \lambda^{-1}_{1+}(k'^2))(X_{1+}(k^2), \\ &\quad \times [N(k^2) - N(k'^2)]X_{1+}(k^2)) \\ &\quad + \theta(\lambda^{-1}_{1+}(k'^2) - \lambda^{-1}_{1+}(k^2))(X_{1+}(k'^2), \\ &\quad \times [N(k'^2) - N(k^2)]X_{1+}(k'^2)). \end{aligned}$$

Owing to Eq. (A1) this implies

$$\lim_{k'^2 \to k^2} \lambda^{-1}_{1+}(k'^2) = \lambda^{-1}_{1+}(k^2), \qquad \text{(A2)}$$

i.e., the continuity of the first eigenvalue.

From Eqs. (A1) and (A2) it follows

$$\lim_{k'^2 \to k^2} (X_{1+}(k'^2), N(k^2)X_{1+}(k'^2)) = \lambda^{-1}_{1+}(k^2). \qquad \text{(A3)}$$

Introducing the expansion

$$X_{1+}(k'^2) = \sum_i c_i(k'^2)X_i(k^2), \qquad \sum_i |c_i(k'^2)|^2 = 1$$

and denoting by δ the difference between $\lambda^{-1}_{1+}(k^2)$ and the eigenvalue nearest to it we have

$$\sum_{i \neq 1+} |c_i(k'^2)|^2\, \delta \leq \sum_i |c_i(k'^2)|^2 (\lambda^{-1}_{1+}(k^2) - \lambda^{-1}_i(k^2)) = \lambda^{-1}_{1+}(k^2) - (X_{1+}(k'^2), N(k^2)X_{1+}(k'^2)).$$

Equation (A3) then implies that

$$\lim_{k'^2 \to k^2} \sum_{i \neq 1+} |c_i(k'^2)|^2 = 0.$$

Therefore, if the arbitrary phase of $X_{1+}(k'^2)$ is chosen in such a way that $c_{1+}(k'^2)$ is real and positive, one has

$$\lim_{k'^2 \to k^2} ||X_{1+}(k'^2) - X_{1+}(k^2)||^2 = \lim_{k'^2 \to k^2} |c_{1+}(k'^2) - 1|^2 = \lim_{k'^2 \to k^2} |(1 - \sum_{i \neq 1+} |c_i(k'^2)|^2)^{\frac{1}{2}} - 1|^2 = 0.$$

This means that the first eigenfunction is continuous.

Taking into account the kernel

$$N_1(k^2, \mathbf{p}, \mathbf{p}') = -\frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}} - \lambda_{1+}^{-1}(k^2)X_{1+}(k^2, \mathbf{p})X_{1+}^*(k^2, \mathbf{p}'),$$

whose greatest positive eigenvalue is the second positive eigenvalue of $N(k^2)$ one evidently has

$$\lim_{k'^2 \to k^2} (\varphi, [N_1(k'^2) - N_1(k^2)]\varphi) = 0.$$

The continuity of the second eigenvalue and eigenfunction can then be established by means of the same procedure used for the first ones. The repetition of the above argument proves the continuous dependence on k^2 of all the eigenvalues and eigenfunctions.

Note added in proof: While this work was in course of publication, we noticed that its main ideas were already present in the paper on scattering problems by S. Weinberg in Phys. Rev. **131**, 440 (1963). Our work, however, contains rigorous self-contained proofs which are not given in Weinberg's paper. Some questions concerning our specific problem and various particular formulas are also considered here.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Perturbation Theory of Microwave Interaction with Gyroelectric Plasmas*

ELLIOTT R. NAGELBERG
Bell Telephone Laboratories, Incorporated, Whippany, New Jersey
(Received 30 June 1964; final manuscript received 18 August 1964)

As a means of determining the effects of a uniform but arbitrarily directed magnetic field on cylindrical and spherical wave propagation in a cold, homogeneous plasma, we regard the magnetic field as a small perturbation. Assuming an expansion for the electric and magnetic fields in powers of the parameter $i\omega_g/\omega$, where ω_g is the static gyrofrequency of the electron, we solve for the linear terms. This solution is carried out under the assumption that the fields are known for the limit of vanishing static magnetic field.

The first-order theory is then applied to cylindrical and spherical systems. When the approximate solution for the axially magnetized column is compared with the exact result, agreement is obtained provided that the static magnetic field is weak, as expected.

I. INTRODUCTION

IN the following article, the results are presented of a theoretical investigation into certain aspects of microwave interaction with a bounded, homogeneous, gyroelectric plasma, i.e., a homogeneous plasma in which a uniform static magnetic field is maintained. It has been shown[1,2] that such a medium may be characterized, within the framework of a phenomenological theory, by a relative permeability equal to unity and a permittivity given by a second-rank tensor.

Among boundary-value problems involving plasmas with cylindrical or spherical shape which are of interest here, only the axially magnetized cylinder is amenable to rigorous analysis.[3,4] It is therefore of interest to search for approximate methods which permit the study of a broader class of problems. We develop one such method, applicable when the static field is "small," in a sense to be defined precisely later on.

Since emphasis is placed on the solution of boundary-value problems rather than on the detailed physics of the plasma–field interaction, discussion is confined to the collisionless case. In addition, it is assumed that the frequency of the electromagnetic field is high enough that the induced motion of positive ions may be neglected. Thus, the only effect

* Excerpted from the dissertation, "Microwave Interaction with Bounded Gyroelectric Plasmas," submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the California Institute of Technology, Pasadena, California.

[1] H. W. Nichols and J. C. Schelling, Bell System Tech. J. 4, 215 (1925).

[2] C. H. Papas, "A Note Concerning a Gyroelectric Medium," Tech. Report No. 4, Antenna Laboratory, California Institute of Technology (1954). Also see C. H. Papas, *Theory of Electromagnetic Wave Propagation* (McGraw-Hill Book Company, Inc., to be published), Chap. 6.

[3] P. M. Platzman and H. T. Ozaki, J. Appl. Phys. **31**, 1957 (1960).

[4] K. H. B. Wilhelmsson, J. Res. Natl. Bur. Std. (U. S.), **66D**, 439 (1962).

This means that the first eigenfunction is continuous.

Taking into account the kernel

$$N_1(k^2, \mathbf{p}, \mathbf{p}') = -\frac{V(\mathbf{p}, \mathbf{p}')}{[(p^2 + k^2)(p'^2 + k^2)]^{\frac{1}{2}}} - \lambda_{1+}^{-1}(k^2)X_{1+}(k^2, \mathbf{p})X_{1+}^*(k^2, \mathbf{p}'),$$

whose greatest positive eigenvalue is the second positive eigenvalue of $N(k^2)$ one evidently has

$$\lim_{k'^2 \to k^2} (\varphi, [N_1(k'^2) - N_1(k^2)]\varphi) = 0.$$

The continuity of the second eigenvalue and eigenfunction can then be established by means of the same procedure used for the first ones. The repetition of the above argument proves the continuous dependence on k^2 of all the eigenvalues and eigenfunctions.

Note added in proof: While this work was in course of publication, we noticed that its main ideas were already present in the paper on scattering problems by S. Weinberg in Phys. Rev. **131**, 440 (1963). Our work, however, contains rigorous self-contained proofs which are not given in Weinberg's paper. Some questions concerning our specific problem and various particular formulas are also considered here.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Perturbation Theory of Microwave Interaction with Gyroelectric Plasmas*

ELLIOTT R. NAGELBERG
Bell Telephone Laboratories, Incorporated, Whippany, New Jersey
(Received 30 June 1964; final manuscript received 18 August 1964)

As a means of determining the effects of a uniform but arbitrarily directed magnetic field on cylindrical and spherical wave propagation in a cold, homogeneous plasma, we regard the magnetic field as a small perturbation. Assuming an expansion for the electric and magnetic fields in powers of the parameter $i\omega_g/\omega$, where ω_g is the static gyrofrequency of the electron, we solve for the linear terms. This solution is carried out under the assumption that the fields are known for the limit of vanishing static magnetic field.

The first-order theory is then applied to cylindrical and spherical systems. When the approximate solution for the axially magnetized column is compared with the exact result, agreement is obtained provided that the static magnetic field is weak, as expected.

I. INTRODUCTION

IN the following article, the results are presented of a theoretical investigation into certain aspects of microwave interaction with a bounded, homogeneous, gyroelectric plasma, i.e., a homogeneous plasma in which a uniform static magnetic field is maintained. It has been shown[1,2] that such a medium may be characterized, within the framework of a phenomenological theory, by a relative permeability equal to unity and a permittivity given by a second-rank tensor.

Among boundary-value problems involving plasmas with cylindrical or spherical shape which are of interest here, only the axially magnetized cylinder is amenable to rigorous analysis.[3,4] It is therefore of interest to search for approximate methods which permit the study of a broader class of problems. We develop one such method, applicable when the static field is "small," in a sense to be defined precisely later on.

Since emphasis is placed on the solution of boundary-value problems rather than on the detailed physics of the plasma–field interaction, discussion is confined to the collisionless case. In addition, it is assumed that the frequency of the electromagnetic field is high enough that the induced motion of positive ions may be neglected. Thus, the only effect

* Excerpted from the dissertation, "Microwave Interaction with Bounded Gyroelectric Plasmas," submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the California Institute of Technology, Pasadena, California.

[1] H. W. Nichols and J. C. Schelling, Bell System Tech. J. 4, 215 (1925).

[2] C. H. Papas, "A Note Concerning a Gyroelectric Medium," Tech. Report No. 4, Antenna Laboratory, California Institute of Technology (1954). Also see C. H. Papas, *Theory of Electromagnetic Wave Propagation* (McGraw-Hill Book Company, Inc., to be published), Chap. 6.

[3] P. M. Platzman and H. T. Ozaki, J. Appl. Phys. **31**, 1957 (1960).

[4] K. H. B. Wilhelmsson, J. Res. Natl. Bur. Std. (U. S.), **66D**, 439 (1962).

of these heavier particles is to make the total average charge density equal to zero.

Rationalized mks units and the harmonic time dependence $e^{-i\omega t}$ will be used.

II. POWER-SERIES EXPANSION FOR THE DIELECTRIC TENSOR

In this section, we derive an expansion for the dielectric tensor in powers of the parameter $i\omega_g/\omega$, where ω_g is the electron gyrofrequency corresponding to the static magnetic field. The series is somewhat unconventional in that the terms are matrices rather than scalars. However, this presents no essentially mathematical difficulty, since the theory of functions of a matrix is analogous in many respects to the corresponding theory for scalar variables.

Maxwell's equations for a plasma are

$$\begin{aligned} \nabla \times \mathbf{E} &= i\omega\mu_0 \mathbf{H}, \\ \nabla \times \mathbf{H} &= -i\omega\epsilon_0 \mathbf{E} - Ne\mathbf{v}. \end{aligned} \tag{1}$$

In Eq. (1), μ_0 and ϵ_0 are the permeability and permittivity of vacuum and $\mathbf{v}$ is the average induced velocity of the electrons, of which there are assumed to be N per cubic meter. Since $\mathbf{v}$ will be found to be a linear vector function of $\mathbf{E}$, it is convenient to write the second of Maxwell's equations as

$$\nabla \times \mathbf{H} = -i\omega\boldsymbol{\varepsilon}\mathbf{E}, \tag{2}$$

thereby defining an effective dielectric tensor $\boldsymbol{\varepsilon}$.

In order to derive the expansion for $\boldsymbol{\varepsilon}$, we examine the motion of electrons acted upon simultaneously by a harmonic electric field $\mathbf{E}e^{-i\omega t}$ and an arbitrarily directed static magnetic field $B_0\mathbf{e}_B$. The symbol $\mathbf{e}$ denotes a unit vector, $\mathbf{e}_B$ representing a unit vector in the direction of the static magnetic field. From Newton's law of motion and the Lorentz force equation we obtain the dyadic expression

$$\mathbf{v} = (e/i\omega m)[\mathbf{U} - i(\omega_g/\omega)\mathbf{e}_B \times]^{-1}\mathbf{E}, \tag{3}$$

where $\mathbf{U}$ is the unit dyad. We now regard the operator $[\mathbf{U} - i(\omega_g/\omega)\mathbf{e}_B \times]^{-1}$ as a function of the matrix $i(\omega_g/\omega)\mathbf{e}_B \times$, in analogy with the function $(1 - s)^{-1}$ of the scalar complex variable $s = \alpha + i\beta$.

There exists a matrix–scalar correspondence principle[5] which states that, if $f(s)$ has the Taylor series expansion

$$f(s) = \sum_{n=0}^{\infty} a_n s^n, \tag{4}$$

then the same function, but with a matrix $\mathbf{S}$ in place of the scalar s, will have the expansion,

$$\mathbf{f}(\mathbf{S}) = \sum_{n=0}^{\infty} a_n \mathbf{S}^n, \tag{5}$$

which converges, provided that the eigenvalues of $\mathbf{S}$ all lie inside the circle of convergence of Eq. (4). Since the circle of convergence of the series, which is

$$(1 - s)^{-1} = \sum_{n=0}^{\infty} s^n, \tag{6}$$

is the unit circle centered about the origin, it follows that the representation

$$\left(\mathbf{U} - i\frac{\omega_g}{\omega}\mathbf{e}_B \times\right)^{-1} = \sum_{n=0}^{\infty}\left(i\frac{\omega_g}{\omega}\mathbf{e}_B \times\right)^n \tag{7}$$

is valid provided that the eigenvalues of $i(\omega_g/\omega)\mathbf{e}_B \times$ satisfy the inequality $|\lambda_i| < 1$ $(i = 1, 2, 3)$. It is straightforward to show that these eigenvalues are $\lambda_1 = 0$, $\lambda_2 = (\omega_g/\omega)$, $\lambda_3 = -(\omega_g/\omega)$. The criterion for validity of Eq. (7) is thus that $\omega_g/\omega < 1$, and it is assumed in what follows that this condition is satisfied.

By substituting Eq. (7) into Eq. (3) and using the definition of $\boldsymbol{\varepsilon}$, we can show that the required expansion is

$$\boldsymbol{\varepsilon} = \epsilon\left[\mathbf{U} - \zeta\sum_{n=1}^{\infty} g^n(\mathbf{e}_B \times)^n\right], \tag{8}$$

where we have used the notation

$$\begin{aligned} \epsilon &= \epsilon_0(1 - \omega_p^2/\omega^2), \quad \zeta = \omega_p^2/(\omega^2 - \omega_p^2), \\ g &= i\omega_g/\omega, \quad \omega_p^2 = Ne^2/m\epsilon_0. \end{aligned} \tag{9}$$

In the limit of vanishing static magnetic field Eq. (8) reduces to the well-known result for a collisionless isotropic plasma.

It is straightforward to show that, in a coordinate system aligned with the z axis along $\mathbf{e}_B$, the representation of Eq. (8) gives the same result as an element-by-element expansion of the conventional formula for $\boldsymbol{\varepsilon}$.[2] However, Eq. (8) has the advantage of being a vector representation and is therefore valid in any coordinate system.

III. PARTIAL FIELD EXPANSION

Using the result of the previous section, we write Maxwell's equations inside the plasma as

$$\begin{aligned} \nabla \times \mathbf{E} &= i\omega\mu_0 \mathbf{H}, \\ \nabla \times \mathbf{H} &= -i\omega\epsilon\mathbf{E} + i\omega\epsilon\zeta\left[\sum_{n=1}^{\infty} g^n(\mathbf{e}_B \times)^n\right]\mathbf{E}. \end{aligned} \tag{10}$$

We next assume expansions for the electric and

[5] F. R. Gantmacher, *Theory of Matrices* (Chelsea Publishing Company, New York, 1959), Vol. 1, p. 113.

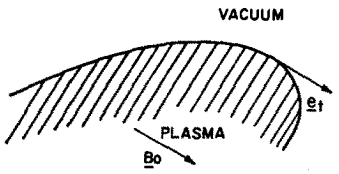

FIG. 1. Boundary between plasma and vacuum.

magnetic fields of the form

$$\mathbf{E} = \sum_{m=0}^{\infty} g^m \mathbf{E}^{(m)}, \qquad \mathbf{H} = \sum_{m=0}^{\infty} g^m \mathbf{H}^{(m)}, \tag{11}$$

where $\mathbf{E}^{(m)}$ and $\mathbf{H}^{(m)}$ are referred to as the mth-order partial fields. Substituting Eq. (11) into Eq. (10) and equating powers of the expansion parameter g, we conclude that the partial fields obey the following equations:

$$\begin{aligned} \nabla \times \mathbf{E}^{(0)} &= i\omega\mu_0 \mathbf{H}^{(0)}, \\ \nabla \times \mathbf{H}^{(0)} &= -i\omega\epsilon \mathbf{E}^{(0)}, \\ \nabla \times \mathbf{E}^{(n)} &= i\omega\mu_0 \mathbf{H}^{(n)}, \\ \nabla \times \mathbf{H}^{(n)} &= -i\omega\epsilon \mathbf{E}^{(n)} \\ &\quad + i\omega\epsilon\zeta \sum_{m=0}^{n-1} (\mathbf{e}_B \times)^{n-m} \mathbf{E}^{(m)}, \qquad n > 0. \end{aligned} \tag{12}$$

The formulation in terms of partial fields differs from the conventional statement of Maxwell's equations for a gyroelectric plasma, both mathematically and physically. From a mathematical point of view, it represents a change from the problem of solving a pair of homogeneous partial differential equations which involve a tensor operator to the problem of solving an infinite sequence of inhomogeneous equations in which the source terms depend on solutions of lower order. Physically, we have introduced a description of the electromagnetic field inside the plasma as being composed of a sum of fields. These are arranged in a hierarchy of complexity in which the fields of lower order "interact" with the static magnetic field to produce those of higher order.

The advantage of present formulation is that, if the static field is weak, the more complex fields may be ignored since they are of higher order in g. In this work we consider in detail only the zero-order field which, as can be seen from Eq. (12), would exist if there were no static magnetic field, and the first order component which is linear in g. Thus, it is assumed that B_0 is small enough so that terms of order $(\omega_g/\omega)^2$ are negligible.

IV. A NOTE ON BOUNDARY CONDITIONS

In the following paragraphs, we consider problems which involve a boundary between a plasma and a vacuum, as in Fig. 1. According to Eq. (11), the fields inside the plasma will be given by

$$\begin{aligned} \mathbf{E}_i &= \mathbf{E}_i^{(0)} + g\mathbf{E}_i^{(1)} + g^2\mathbf{E}_i^{(2)} + \cdots, \\ \mathbf{H}_i &= \mathbf{H}_i^{(0)} + g\mathbf{H}_i^{(1)} + g^2\mathbf{H}_i^{(2)} + \cdots. \end{aligned} \tag{13}$$

Because of the boundary conditions, there must exist a corresponding field on the vacuum side of the interface,

$$\begin{aligned} \mathbf{E}_e &= \mathbf{E}_e^{(0)} + g\mathbf{E}_e^{(1)} + g^2\mathbf{E}_e^{(2)} + \cdots, \\ \mathbf{H}_e &= \mathbf{H}_e^{(0)} + g\mathbf{H}_e^{(1)} + g^2\mathbf{H}_e^{(2)} + \cdots, \end{aligned} \tag{14}$$

and by equating tangential components we obtain the result that

$${}_tE_e^{(n)} = {}_tE_i^{(n)}, \qquad {}_tH_e^{(n)} = {}_tH_i^{(n)} \tag{15}$$

for all n. Hence, the boundary conditions must be satisfied at each step of the perturbation procedure. The physical interpretation is that the internal interactions between the fields of a given order and $\mathbf{B}_0$ produce waves which are partially transmitted and partially reflected at the plasma–vacuum boundary.

V. ON THE CRITERIA FOR VALIDITY OF THE PERTURBATION EXPANSION

In general, the solution to a physical or mathematical problem depends on several parameters, with the result that an asymptotic form is rarely uniformly valid. For example, the usual asymptotic formulas for Bessel functions with large argument do not apply when the order is of comparable size.

It has been implied that the perturbation theory for gyroelectric plasmas is applicable, provided that the magnetic field is weak in the sense that $\omega_g/\omega \ll 1$. We now explore the conditions for validity more throughly in order to determine how the other parameters, namely the plasma frequency and physical dimensions of the interaction zone, affect the convergence of the expansion of Eq. (11).

Dependence on Plasma Frequency

In order to determine the effect of plasma frequency, we eliminate $\mathbf{H}^{(n)}$ in Eqs. (12) and obtain as the equation for $\mathbf{E}^{(n)}$, $n > 0$,

$$\nabla \times \nabla \times \mathbf{E}^{(n)} - k^2 \mathbf{E}^{(n)} = -k^2\zeta \sum_{m=0}^{n-1} (\mathbf{e}_B \times)^{n-m} \mathbf{E}^{(m)}, \tag{16}$$

k being the wavenumber in the isotropic plasma

$\omega(\mu_0\epsilon)^{\frac{1}{2}}$. Using mathematical induction, we can show that any particular solution for the nth-order field ($n > 0$) will be of the form

$$\mathbf{E}_p^{(n)} = \sum_{m=1}^{n} \zeta^m \mathbf{f}_m^{(n)}, \tag{17}$$

where $\zeta = \omega_p^2/(\omega^2 - \omega_p^2)$, so that in the expansion for $\mathbf{E}^{(n)}$ there will be a term proportional to ζ^n. This parameter becomes large at frequencies close to the plasma frequency and we reason that here the expansion will converge poorly, if at all. The physical explanation is in the fact that the plasma frequency is a "resonant" frequency at which the electrons undergo large-amplitude oscillations. Hence, even if the static magnetic field is weak, the electron path length over which it operates may be large enough that its effect is important.

Physical Dimensions of the Interaction Zone

Because of the difficulty of attacking a completely general problem, we demonstrate the effect of physical dimensions by considering the specific example of a TEM wave propagating in the direction of a static magnetic field. The conclusions which are extrapolated nevertheless seem plausible on physical grounds.

The coordinate system and the unperturbed wave $\mathbf{E}^{(0)}$, $\mathbf{H}^{(0)}$ are shown in Fig. 2. Using Eq. (16), we obtain the inhomogeneous equation for $\mathbf{E}^{(1)}$

$$\nabla^2\mathbf{E}^{(1)} + k^2\mathbf{E}^{(1)} = \zeta k^2 E_0 e^{ikz}\mathbf{e}_y, \tag{18}$$

which gives the first-order change in the electric field

$$(\omega_g \zeta E_0 kz/2\omega)e^{ikz}\mathbf{e}_y. \tag{19}$$

This term, proportional to kz, produces a rotation of the electric vector in the transverse plane, an effect known as Faraday rotation.

In a similar manner, it may be shown that $\mathbf{E}^{(2)}$ will have a term proportional to $(kz)^2$, $\mathbf{E}^{(3)}$ to $(kz)^3$, and so on. Thus in solving the problem of propagation through a longitudinally magnetized slab of width L, one obtains a first-order field proportional to kL, and for a cylinder or sphere of radius a, a corresponding result proportional to ka. The size of the plasma, therefore, cannot be allowed to become too large for then the assumption of small changes will be violated.

On the basis of the preceding results, we reason that the condition $\omega_g/\omega \ll 1$ is not sufficient; the static magnetic field, operating frequency, plasma frequency, and characteristic dimension L must be such that

$$|\zeta(\omega_g/\omega)kL| \ll 1. \tag{20}$$

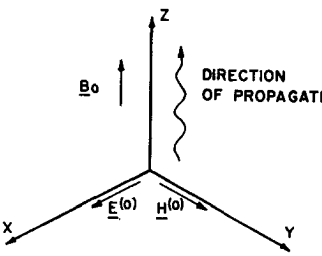

FIG. 2. Unperturbed TEM wave propagating in the direction of the static magnetic field.

If this inequality is satisfied, the perturbation method should yield good results.

VI. SOLUTION FOR THE FIRST-ORDER FIELDS

Having established what appear to be sufficient conditions for validity of the perturbation theory, we proceed to determine the general solutions for the first-order fields, which satisfy

$$\begin{aligned} \nabla\times\mathbf{E}^{(1)} &= i\omega\mu_0\mathbf{H}^{(1)}, \\ \nabla\times\mathbf{H}^{(1)} &= -i\omega\epsilon\mathbf{E}^{(1)} + i\omega\epsilon\zeta\mathbf{e}_B\times\mathbf{E}^{(0)}. \end{aligned} \tag{21}$$

The general solution to Eqs. (21) consist of two parts, a particular integral and a complementary integral. In problems involving bounded plasmas, which are of interest here, both parts have physical significance. The particular solution is the field which arises directly from the interaction between the zero-order wave and the static magnetic field. On the other hand, complementary solutions, satisfying the homogeneous equations,

$$\begin{aligned} \nabla\times\mathbf{E}_c^{(1)} &= i\omega\mu_0\mathbf{H}_c^{(1)}, \\ \nabla\times\mathbf{H}_c^{(1)} &= -i\omega\epsilon\mathbf{E}_c^{(1)}, \end{aligned} \tag{22}$$

are superimposed in order to satisfy boundary conditions and thus may be interpreted as fields which are reflected from the boundary back into the plasma. Although the complementary solutions required for problems involving cylindrical or spherical plasmas may be obtained by known methods, particular integrals, denoted by $\mathbf{E}_P$ and $\mathbf{H}_P$, present a more formidable problem.

We partition the field $\mathbf{E}_P^{(1)}$ into two parts, one accounting for the divergence and a remainder which is solenoidal. The first is found by using the divergence operator on the second equation of Eq. (21), from which we find that

$$\nabla\cdot\mathbf{E}_P^{(1)} = \zeta\nabla\cdot(\mathbf{e}_B\times\mathbf{E}^{(0)}). \tag{23}$$

We are thus led to write that

$$\mathbf{E}_P^{(1)} = \zeta\mathbf{e}_B\times\mathbf{E}^{(0)} + \mathbf{E}_a^{(1)}, \tag{24}$$

where $\mathbf{E}_a^{(1)}$ is solenoidal. Substituting Eq. (24) into

Eqs. (21) and eliminating $\mathbf{H}_P^{(1)}$, we determine that $\mathbf{E}_a^{(1)}$ satisfies

$$\nabla^2\mathbf{E}_a^{(1)} + k^2\mathbf{E}_a^{(1)} = -iZ_0\zeta k_0^2(\mathbf{e}_B\cdot\nabla/k_0)\mathbf{H}^{(0)}, \quad (25)$$

where Z_0 is the impedance of vacuum. The use of the normalized gradient operator, ∇/k_0, is desirable for maintaining dimensional consistency.

Assume now that the fields for the isotropic problem are known and, in accordance with one method for generating solutions to Maxwell's equations,[6] $\mathbf{H}^{(0)}$ is given by

$$\mathbf{H}^{(0)} = \frac{\nabla}{k_0}\times(\mathbf{u}W_1^{(0)}) + \frac{\nabla}{k_0}\times\left[\mathbf{u}\times\left(\frac{\nabla}{k_0}W_2^{(0)}\right)\right]. \quad (26)$$

In Eq. (26), $\mathbf{u}$ denotes one of the rectangular unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$, or the (normalized) radius vector $k_0\mathbf{r}$, the particular choice depending on the geometry. $W_1^{(0)}$ and $W_2^{(0)}$ must, of course, satisfy the scalar wave equation

$$\nabla^2 W + k^2 W = 0. \quad (27)$$

Let $\mathbf{E}_a^{(1)}$ be represented as

$$\mathbf{E}_a^{(1)} = iZ_0\zeta\left(\mathbf{e}_B\cdot\frac{\nabla}{k_0}\right)\left\{\frac{\nabla}{k_0}\times(\mathbf{u}V_1^{(1)}) + \frac{\nabla}{k_0}\times\left[\mathbf{u}\times\left(\frac{\nabla}{k_0}V_2^{(1)}\right)\right]\right\}, \quad (28)$$

where $V_1^{(1)}$ and $V_2^{(1)}$ are scalar functions. It then follows, from the fact that $\mathbf{e}_B$ is a constant vector, and from the commutative properties of the Laplacian,[7] that Eq. (28) will be a particular solution to Eq. (25) provided that the scalar functions $V_1^{(1)}$ and $V_2^{(1)}$ are related to the functions $W_1^{(0)}$ and $W_2^{(0)}$ by the differential equations

$$\begin{aligned}\nabla^2V_1^{(1)} + k^2V_1^{(1)} &= -k_0^2W_1^{(0)},\\ \nabla^2V_2^{(1)} + k^2V_2^{(1)} &= -k_0^2W_2^{(0)}.\end{aligned} \quad (29)$$

Combining these results, we determine that a particular solution to Eq. (21) is

$$\mathbf{E}_P^{(1)} = \zeta\mathbf{e}_B\times\mathbf{E}^{(0)} + iZ_0\zeta\left(\mathbf{e}_B\cdot\frac{\nabla}{k_0}\right)\left\{\frac{\nabla}{k_0}\times(\mathbf{u}V_1^{(1)}) + \frac{\nabla}{k_0}\times\left[\mathbf{u}\times\left(\frac{\nabla}{k_0}V_2^{(1)}\right)\right]\right\}, \quad (30)$$

where $\mathbf{E}^{(0)}$ is the zero-order electric field and $V_1^{(1)}$ and $V_2^{(1)}$ are derived from the generating functions for the zero-order magnetic field as prescribed by Eq. (29).

[6] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), 2nd ed., Chaps. 14 and 15.

[7] Reference 6, p. 265.

It is important to note that the function $V_1^{(1)}$ generates a transverse electric wave (with respect to the vector $\mathbf{u}$) while $V_2^{(1)}$ generates a transverse magnetic wave, the opposite from their respective zero-order counterparts, $W_1^{(0)}$ and $W_2^{(0)}$. This shows very clearly why TE and TM modes are not, in general, solutions to Maxwell's equations for a gyroelectric plasma; a TE mode will "interact" with $\mathbf{B}_0$ to produce a TM mode, and vice versa. The physical origin of this interaction is, of course, the effect of the static magnetic field on the motion of the electrons, but to examine the field structure on the basis of the individual orbits would be prohibitively complicated. The macroscopic, or phenomenological, approach is a convenient alternative.

An example worth mentioning at this point is the case of a plane TE wave normally incident on a longitudinally magnetized cylinder. Here $\mathbf{e}_B\cdot\nabla \equiv 0$, and $\mathbf{e}_B\times\mathbf{E}^{(0)}$ is in the transverse plane, so that the TE character of the incident wave is retained throughout. The case in which the cylinder has a circular cross section was discussed by Platzman and Ozaki.[3]

VI. APPLICATIONS TO CYLINDRICAL AND SPHERICAL SYSTEMS

To determine the fields generated by the interaction between the zero-order wave and static magnetic field, it is necessary to know the functions $V_1^{(1)}$ and $V_2^{(1)}$. These shall now be determined for cylindrical and spherical systems.

Cylindrical Coordinates

For boundary-value problems involving cylindrical geometry, $W_1^{(0)}$ and $W_2^{(0)}$ are conveniently expanded in a series of the form

$$W = \sum_{n=-\infty}^{+\infty} a_n J_n(\beta\rho)e^{in\varphi}e^{i\gamma z}, \quad (31)$$

where $\beta = (k^2 - \gamma^2)^{\frac{1}{2}}$. Then $V_1^{(1)}$ and $V_2^{(1)}$ will have the expansion

$$V = \sum_{n=-\infty}^{+\infty} a_n u_n(\rho, \varphi, z), \quad (32)$$

where the u_n satisfy

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial u_n}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 u_n}{\partial^2\varphi} + \frac{\partial^2 u_n}{\partial z^2} + k^2u_n = -k_0^2J_n(\beta\rho)e^{in\varphi}e^{i\gamma z}. \quad (33)$$

We assume that the φ and z dependences of u_n are the same as on the right, and represent the

solution as

$$u_n(\rho, \varphi, z) = R_n(\beta\rho)e^{in\varphi}e^{i\gamma z}. \tag{34}$$

Substituting Eq. (34) into Eq. (33), we determine that $R_n(\beta\rho)$ must satisfy the ordinary differential equation

$$v^2 \frac{d^2R_n(v)}{dv^2} + v \frac{dR_n(v)}{dv} + (v^2 - n^2)R_n(v) = -\frac{v^2}{\nu^2} J_n(v), \tag{35}$$

where $v = \beta\rho$ and $\nu = \beta/k_0$. This is an inhomogeneous form of Bessel's equation and may be solved using the method of variation of parameters.[8] The solution is

$$R_n(\beta\rho) = (\beta\rho/2\nu^2)J_n'(\beta\rho). \tag{36}$$

It follows that, if $W_1^{(0)}$ is given by

$$W_1^{(0)} = \sum_{n=-\infty}^{+\infty} a_n^{(0)} J_n(\beta\rho)e^{in\varphi}e^{i\gamma z}, \tag{37}$$

then, for $V_1^{(1)}$, we have

$$V_1^{(1)} = \frac{\beta\rho}{2\nu^2} \sum_{n=-\infty}^{+\infty} a_n^{(0)} J_n'(\beta\rho)e^{in\varphi}e^{i\gamma z}, \tag{38}$$

with an analogous result for $V_2^{(1)}$ corresponding to $W_2^{(0)}$.

Spherical Coordinates

Here $W_1^{(0)}$ and $W_2^{(0)}$ will have the form

$$W = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} a_{mn}^{(0)} \mathfrak{z}_n(kr)P_n^m(\cos\theta)e^{im\varphi}, \tag{39}$$

in terms of the radial coordinate r, the azimuthal angle φ, and coaltitude angle θ. $\mathfrak{z}_n(kr)$ denotes any spherical Bessel function and P_n^m an associated Legendre polynomial.

By analogy with the cylindrical case, we are led to solve

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial u_n^m}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial u_n^m}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u_n^m}{\partial\varphi^2} + k^2 u_n^m = -k_0^2 \mathfrak{z}_n(kr)P_n^m(\cos\theta)e^{im\varphi}. \tag{40}$$

Assume now that the φ and θ dependences are the same as on the right so that

$$u_n^m(r, \theta, \varphi) = R_n(kr)S_n^m(\theta, \varphi), \tag{41}$$

where

$$S_n^m(\theta, \varphi) = P_n^m(\cos\theta)e^{im\varphi}. \tag{42}$$

[8] E. R. Nagelberg, "Microwave Interaction with Bounded Gyroelectric Plasmas," Technical Report No. 31, Antenna Laboratory, California Institute of Technology, Appendix D (1964).

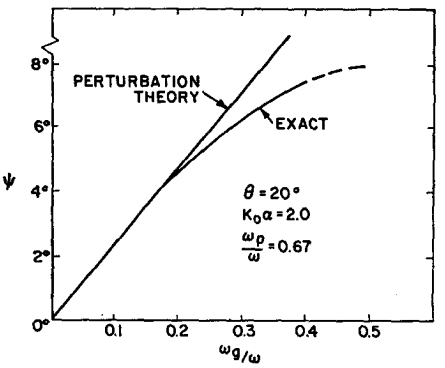

FIG. 3. Orientation angle ψ of the polarization ellipse; comparison between exact and first-order results.

Using the eigenvalue identity

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial S_n^m}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 S_n^m}{\partial\varphi^2} = -n(n+1)S_n^m, \tag{43}$$

we determine that R_n must satisfy the ordinary differential equation

$$v^2 \frac{d^2R_n(v)}{dv^2} + 2v\frac{dR_n}{dv} + [v^2 - n(n+1)]R_n(v) = -\frac{v^2}{\nu^2}\mathfrak{z}_n(v), \tag{44}$$

with $v = kr$ and $\nu = k/k_0$. If we make the substitutions

$$\begin{aligned} R_n(v) &= (\pi/2v)^{\frac{1}{2}}G_n(v), \\ \mathfrak{z}_n(v) &= (\pi/2v)^{\frac{1}{2}}Z_{n+\frac{1}{2}}(v), \end{aligned} \tag{45}$$

we obtain the same equation for $G_n(v)$ as in the previous section, except that the index is now half-integral. We may still write, however, that

$$G_n(v) = (kr/2\nu^2)Z_{n+\frac{1}{2}}'(kr), \tag{46}$$

and hence that

$$u_n^m(kr, \theta, \varphi) = (kr/2\nu^2)(\pi/2kr)^{\frac{1}{2}}Z_{n+\frac{1}{2}}'(kr)S_n^m(\theta, \varphi). \tag{47}$$

Using Eq. (46) and the series for $W_1^{(0)}$ and $W_2^{(0)}$, we may obtain the corresponding series for $V_1^{(1)}$ and $V_2^{(1)}$.

As mentioned earlier, the final step is to superimpose the usual vector wavefunctions inside and outside the plasma in such a way that the components of electric and magnetic fields tangent to the interface are continuous. For the cylindrical case this will be essentially straightforward, owing to the simple dependence on φ and z. It will be found that, when the static magnetic field is not aligned with the z axis, the operators $\mathbf{e}_B \times$ and $\mathbf{e}_B \cdot \nabla$ which occur in Eq. (30) will cause a coupling between

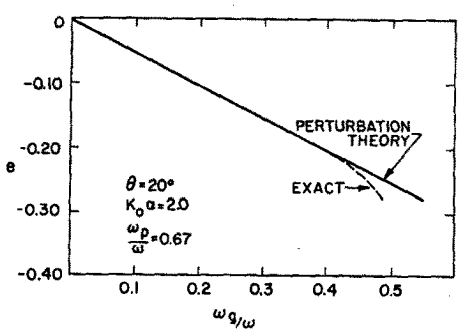

FIG. 4. Eccentricity e of the polarization ellipse; comparison between exact and first-order results.

functions of the form $e^{in\varphi}$ and $e^{i(n+1)\varphi}$ and $e^{i(n-1)\varphi}$. This is analogous to problems in quantum mechanics, where a perturbation which does not preserve cylindrical symmetry causes a coupling between states with different z components of angular momentum.

The difficulty is compounded in the case of spherical coordinates because of the complicated dependence on the angle θ. However if, without loss of generality, we align the coordinate system so that the z axis is in the direction of the static magnetic field, then there is coupling only between different n values. The case where all quantities are independent of φ has been examined and it is found that wavefunctions corresponding to a given n value couple into functions of order $n-1$ and $n+1$.[9]

VII. A VERIFICATION OF THE THEORY

In order to verify the previous results, the perturbation theory was applied to the problem of scattering of an obliquely incident plane wave by an axially magnetized cylinder with circular cross section.[10] The incident wave was polarized with the electric field perpendicular to the plane of incidence, i.e., the plane determined by the incident propagation vector and the axis of the cylinder.

The far-zone scattered field consists of TEM waves propagating in the direction of specular reflection. However, whereas in the isotropic case $\omega_g/\omega = 0$ the scattered field in the plane of incidence is polarized in the same direction as the incident wave, as the static magnetic field increases the scattered wave becomes elliptically polarized. Both the orientation and eccentricity of the ellipse depend on B_0. Figures 3 and 4 illustrate the angle of orientation ψ with respect to the incident electric vector, and the magnitude of the eccentricity e as a function of ω_g/ω for $\omega_p/\omega = 0.67$, $k_0a = 2.0$, and $\theta = 20°$; a is the radius of the cylinder; and θ is the angle between the incident propagation vector and the normal to the axis. As anticipated, the perturbation theory correctly predicts the behavior for small values of magnetic field.

VIII. SUMMARY AND CONCLUSIONS

The purpose of this article is to present a theory of microwave interaction with gyroelectric plasmas in which the biasing magnetic field is regarded as a perturbation. Accordingly, we represent the electromagnetic field as being made up of a sum of partial fields arranged in order of increasing complexity. The fields of lower order are presumed to interact with the static magnetic field to produce those of higher order. Such a formulation has the advantage that, under suitable conditions, the more complex fields may be neglected. These conditions are that the ratio ω_g/ω be less than unity and that the operating frequency, plasma frequency, and characteristic dimension L be such that the inequality,

$$|\omega_g\omega_p^2 kL/\omega(\omega^2 - \omega_p^2)| \ll 1,$$

be satisfied. This is equivalent to the physical requirement that the static magnetic field have a relatively small effect, i.e., that the additional fields which result from the gyroelectric character of the plasma be small compared to those which would exist if the medium were isotropic.

By solving formally for the first-order fields we find that TE and TM modes are not, in general, solutions to Maxwell's equations for a gyroelectric medium. The physical reason for this is that the effect of the static magnetic field on the electron motion induced by a wave of one type is such as to produce a wave of the other type.

The first-order theory is then applied to cylindrical and spherical systems. The fields induced by the first interaction are essentially determined in terms of appropriate coordinates, thus permitting the first-order solution of boundary-value problems. A comparison is made between the exact and perturbation solutions for the problem of scattering by an axially magnetized column. As anticipated, the approximate solution is valid for small values of magnetic field.

ACKNOWLEDGMENT

The author is indebted to his research advisor, Professor Charles H. Papas of the California Institute of Technology, who contributed encouragement as well as guidance to this project.

[9] Reference 8, Chap. VI.
[10] Reference 8, Chaps. III and V.

Summation of Partial Waves*

HENRY BRYSK†
Physics Department, Adelphi University, Garden City, New York
(Received 14 July 1964)

For high-energy potential scattering, when the partial wave expansion converges too slowly to be directly useful, a technique is proposed for extracting the differential cross section more effectively from the phase shifts by means of a weight function and orthogonal polynomial approach.

I. INTRODUCTION AND CONCEPT

THE exact solution of a potential scattering problem normally involves separation of variables followed by numerical solution of the equations for the partial waves (which may be coupled for noncentral potentials). At high energies (i.e., when the ratio of range to wavelength is large), many partial waves contribute significantly. This gives rise to the twin problems that a prohibitively large number of partial waves (or at least their asymptotic representation in terms of phase shifts) must be computed, and that (especially for large angles of scattering) heroic measures are required to achieve sufficient precision to cope with near cancellations.

A mathematically similar difficulty was encountered in the problem of diffusion of atomic radiation in matter, and considerable success was achieved by a technique involving a spatial weight function and expansion of the distribution function in terms of polynomials orthogonal relative to the weight function.[1] A related approach is here proposed for the scattering amplitude f. From the partial wave expansion, f emerges as a Legendre polynomial series. The Legendre polynomials are polynomials in $x = \cos\theta$ which are orthogonal relative to weight function unity. It seems reasonable to rewrite f in terms of a weight function $W(x)$ times an expansion in polynomials orthogonal relative to W. With a judicious choice of W, this would yield a more rapidly convergent series for the angular distribution. Clearly, if W happened to coincide with f there would be maximum convergence (a single term). The more nearly W is a good guess to the exact f, the better the rate of convergence. The choice of W then involves correlating the preliminary available information (solutions of cognate problems, approximations to present one) for a best guess at f. A bootstrap improvement can be achieved by putting adjustable parameters in W and then varying them to improve the tailing-off of the series. A blow-by-blow run-through of the technique will be given for a particular case, followed by a systematic exposition of the procedure.

II. ILLUSTRATION: n–p POTENTIAL

Inasmuch as the technique involves the use of intuition and exploration, the gist of it can best be conveyed in a blow-by-blow solution of an illustrative example. Of necessity, the illustration has to be set up somewhat artificially, since an assessment of success requires the exact solution to be known, which in turn undercuts the need for the analysis. An example simple enough for clarity (and hand computation) yet nontrivial, can be constructed for n–p scattering at about 100 MeV. Swan[2] quotes several such central potentials (with parameters adjusted to yield the experimental scattering length and effective range) and quotes machine computations of the phase shifts at six energies. The case chosen is the Gaussian potential

$$U(r) = -U_0 \exp[-(r/b)^2] \qquad (1)$$

for the ${}^3S$ state ($U_0 = 1.7399 \times 10^{26}$ cm^{-2}, $b = 1.4837 \times 10^{-13}$ cm) at the highest energy given ($k = 1.5 \times 10^{13}$ cm^{-1}). Swan quotes the phase shifts for $l = 0 - 6$, the last one being 0.1% of the first; the differential cross section can be reconstructed from these, yielding the curve labeled "exact" in Fig. 1. The problem is now posed as follows: Suppose that only the first three exact phase shifts were known, i.e.,

$$\delta_0 = 0.6830, \quad \delta_1 = 0.4941, \quad \delta_2 = 0.2290. \qquad (2)$$

From these (neglecting the higher ones) there results the curve labeled "truncated" in Fig. 1, which obviously completely fails at large angles of scatter-

* Work supported by the National Science Foundation under Grant NSF-G 23169.
† Present address: United Nuclear Corporation, White Plains, New York.
[1] L. V. Spencer and U. Fano, J. Res. Natl. Bur. Std. (U. S.) 46, 446 (1951); and many following papers.
[2] P. Swan, Nucl. Phys. 18, 245 (1960).

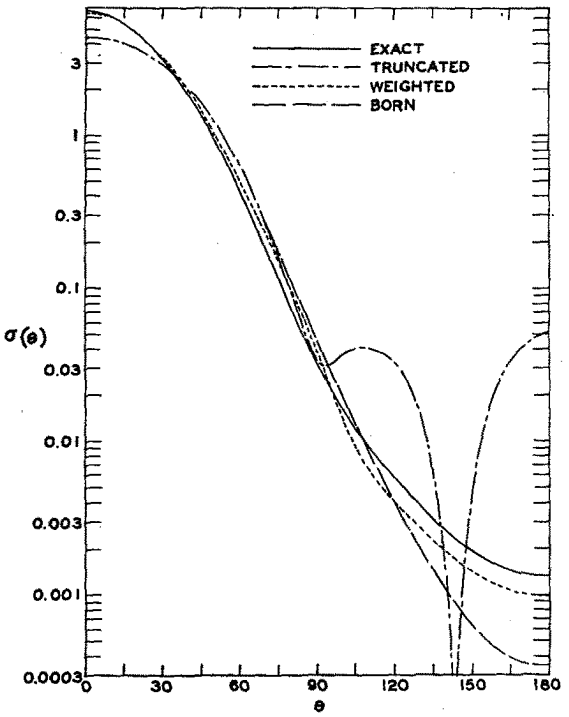

FIG. 1. Differential cross section for an n–p scattering potential.

ing. Can this failure be rectified with the orthogonal polynomial technique?

The Born approximation cross section for the Gaussian is well known,[3] namely

$$\sigma(\theta) = (\pi/16)U_0^2 b^6 \exp[-k^2b^2(1 - \cos\theta)]. \quad (3)$$

It is also plotted in Fig. 1; characteristically, it is pretty good at small angles, too small at large angles. Incidentally, summing the Born-approximation phase shifts (also given by Swan) yields a quite different cross-section curve, comparably in error in the opposite directions for large angles. The prescription of using the Born approximation with the first few phase shifts corrected to their exact values[4] does not work here.

The scattering amplitude f, is an expansion in Legendre polynomials $P_l(x)$ (where $x = \cos\theta$), i.e., in polynomials orthogonal relative to a weight function unity. The object is now to switch to a different weight function. An obvious first choice is the Born approximation, i.e., (since a multiplicative constant does not affect the result) to use

$$W(x) = \exp(\tfrac{1}{2}k^2b^2x), \quad (4)$$

with f (or more precisely with its real and imaginary parts). When this is tried out, the expansion coefficients are found to increase rather than decrease (the ϕ_n's are of order unity), namely for the real part of f

$$D_1/D_0 = -1.453, \qquad D_2/D_0 = -6.409; \quad (5)$$

and for the imaginary part of f

$$D_1/D_0 = -.3483, \qquad D_2/D_0 = -8.368. \quad (6)$$

This behavior immediately signals that the weight function is not adequate, that a more sophisticated one will be required.

It is known that in general the Born approximation will be better at small angles, and will give too small a cross section at large angles.[4] (To play the present game fairly, such background information can be used, but a peek at the exact curve is not allowed.) Thus, the weight function should be modified in that direction, i.e., the new weight function should differ little from the Born approximation at small angles and give a larger result at large angles. This can be achieved with a minimum of new computational effort by choosing

$$W(x) = \exp(\tfrac{1}{2}k^2b^2x) + A\exp(-\tfrac{1}{2}k^2b^2|x|). \quad (7)$$

The constant A is best used as an adjustable parameter to be selected to make the sequence of D_n's drop off as fast as possible. The simplest prescription is to choose A such that $D_2 = 0$; this means using a different value of A for the real part of f than for the imaginary part of f. (Formally, better convergence criteria can be set up, but they are more laborious.) In the present case, this constraint on A leads to a quadratic equation, and the average of the results obtained with the two roots has been used. Carrying through this calculation, the curve labeled "weighted" in Fig. 1 is obtained. The agreement with the "exact" curve is respectable.

In the last round, the D_1 values obtained are still of comparable magnitude to the D_0 values. This serves as internal evidence that there is yet a discrepancy. The next systematic step would be to further refine the weight function, aiming to achieve $D_1 \ll D_0$ as well as $D_2 = 0$; success in this effort would provide confidence that results close to the exact have been obtained.

III. SYSTEMATIC PROCEDURE

The illustrative example has illuminated the approach—the sequence of steps and the relevant considerations. The outline of the procedure can now be systematized, with an eye to making the variation of parameters amenable to machine programming.

[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Pergamon Press Ltd., London, 1958), p. 413.

[4] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1949), pp. 191–193.

The starting point is a given potential for which there are available a finite number of phase shifts which do not suffice to determine accurately the differential cross section. At least a rough idea is always attainable as to what the cross section should look like, either from known solutions of cognate problems or from an approximate solution of the present one. The aim is to exploit this auxiliary information plus the internal resources of a variation of parameters technique in order to achieve optimum use of the phase shifts.

The phase shifts yield the real and imaginary parts of the scattering amplitude f as finite Legendre polynomial expansions with real coefficients; these two are handled as separate and distinct, though parallel, problems. A weight function is written down, consisting of the best preliminary guess as to the form of f(real or imaginary part, respectively). The weight function should incorporate one or more adjustable parameters, which need not occur linearly. The expansion in polynomials orthogonal to the weight function is next calculated; on a computer, this means a sequence of calculations with each of a succession of values of the parameters. What is desired is a particular choice of values of the parameters for which the expansion coefficients fall off successively at a reasonably fast rate. This goal can be expressed as a sequence of inequalities of increasing stringency, the computer being programmed to keep varying the parameters until the conditions are met (with obvious auxiliary instructions to narrow the range of the scanning as the conditions start being satisfied). If the convergence requirements fail to be met, the weight function should be changed and the process repeated.

Work is in progress, in collaboration with C. R. Fischer, on applying this technique to the scattering of high-energy electrons by nuclei.

APPENDIX: ORTHOGONAL POLYNOMIALS

In this section, expansion of a function in terms of polynomials orthogonal relative to a weight function is considered, and the transformation from one weight function to another outlined. To simplify the exposition while retaining ample generality for the present purpose, it will be assumed that all functions are real functions of a real variable, the weight functions are nonnegative, and all integrals exist and are finite. These conditions can be considerably relaxed; exhaustive discussions can be found in the standard mathematical literature.[5]

[5] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, New York, 1959).

The scalar product of two functions $f(x)$ and $g(x)$ relative to a weight function $W(x)$ in the interval $a \leq x \leq b$ is defined as

$$(f, g) = \int_a^b W(x)f(x)g(x)\,dx. \tag{8}$$

A set of polynomials $\{\phi_n(x)\}$ is orthogonal in this system if

$$(\phi_n, \phi_m) = 0, \qquad n \neq m. \tag{9}$$

As a particular case, $\{P_n(x)\}$ will denote the set of orthogonal polynomials with weight function unity in the same interval, and the corresponding scalar product will be denoted by square brackets:

$$[P_n, P_m] = 0, \qquad n \neq m. \tag{10}$$

Since a polynomial of order n is expressible in terms of any complete set of polynomials of order up to n,

$$P_n(x) = \sum_{k=0}^{n} a_{nk}\phi_k(x). \tag{11}$$

Taking the scalar product with ϕ_m (for $m \leq n$), Eq. (9) yields for the coefficient

$$a_{nm} = (P_n, \phi_m)/(\phi_m, \phi_m). \tag{12}$$

If the $P_n(x)$ are known, the above relations suffice to construct the $\phi_n(x)$, up to an arbitrary constant factor. For computational ease, the latter will be specified by the choice

$$a_{nn} = 1 \tag{13}$$

rather than by normalizing the polynomials. For $m = n$, Eq. (12) then yields

$$(\phi_n, \phi_n) = (P_n, \phi_n). \tag{14}$$

The $\phi_n(x)$ can now be obtained recursively from Eq. (11) upon substituting Eqs. (12)–(14) into it:

$$\phi_0(x) = P_0(x), \tag{15}$$

$$\phi_n(x) = P_n(x) - \sum_{k=0}^{n-1} \phi_k(x)\frac{(P_n, \phi_k)}{(P_k, \phi_k)}. \tag{16}$$

Given an expansion of a function in terms of the set $\{P_n(x)\}$, its expansion in terms of the set $\{\phi_n(x)\}$ relative to the weight function $W(x)$,

$$\sum_{k=0}^{\infty} C_k P_k(x) = W(x) \sum_{m=0}^{\infty} D_m\phi_m(x), \tag{17}$$

can be directly deduced on applying the orthogonality relations. Upon multiplying Eq. (17) by $\phi_n(x)$ and integrating from a to b,

$$\sum_{k=0}^{\infty} C_k[P_k, \phi_n] = \sum_{m=0}^{\infty} D_m(\phi_m, \phi_n) = D_n(\phi_n, \phi_n). \tag{18}$$

The k sum cuts off at $k = n$ because the expansion of ϕ_n in terms of the P_k has no contribution for $k > n$. Substituting Eq. (14),

$$D_n = \left\{\sum_{k=0}^{n} C_k[P_k, \phi_n]\right\}/(P_n, \phi_n). \tag{19}$$

A crucial point here is that D_n depends only on the first n of the C_k, so a given number of terms in the left-hand series of Eq. (17) completely determines an equal number of terms in the right-hand series.

For the sake of illustration, the set $\{P_n(x)\}$ were described as having weight function unity. Actually, no essential use was made of this specialization. The results above permit a direct transformation from an expansion in terms of any weight function to an expansion in terms of any other weight function (subject to the stated analytical limitations).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Hilbert Spaces of Type S*

ALEXANDER GROSSMANN
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 30 June 1964)

Certain Hilbert spaces of generalized functions are examined. They contain the space of tempered distributions, are invariant with respect to the Fourier transformation and contain functions that increase rapidly at infinity.

1. INTRODUCTION

THIS paper is concerned with a family of Hilbert spaces which are closely related to the topological vector spaces "of type S" of Gel'fand and Silov. (See Refs. 1 and 2). These Hilbert spaces share with $L^{(2)}$, and with the space of tempered distributions, the property of being invariant with respect to the Fourier transformation. Their elements are not restricted to polynomial growth at infinity, a fact which makes them sometimes more suitable than tempered distributions.

The definitions and results of Secs. 2, 3, and 4 can be illustrated on a simple special case:

(A) The first—and decisive—stage of the game is the construction of Hilbert spaces of testing functions, denoted here by H_α.

Let S be the vector space of infinitely differentiable functions $u(x)(-\infty < x < \infty)$ such that

$$\sup |x^k D^m u(x)| < \infty \quad (k, m = 1, 2, \cdots).$$

In S, consider the sequence of operators

$$\begin{aligned} M_0 &= 1 \\ M_1 &= x^2 + p^2 \\ &\vdots \quad\quad \vdots \\ M_n &= xM_{n-1}x + pM_{n-1}p \\ &\vdots \quad\quad \vdots \end{aligned} \tag{1.1}$$

Here $p = i^{-1}D$, and D is the operator of differentiation. The expression (u, v) means $\int u^*(x)v(x)dx$.

For every $u \in S$, $u \neq 0$, the numbers $(u, M_n u)$ are positive.

Now let α be a positive number. Define H_α as the set of functions $u \in S$ for which

$$\sum_n [\Gamma(\alpha n)]^{-2}(u, M_n u) < \infty. \tag{1.2}$$

Here Γ is Euler's gamma function.

With the obvious scalar product, H_α becomes a Hilbert space (Sec. 3B). If $\alpha < \frac{1}{2}$, then H_α consists of the element $u = 0$ only; if $\alpha > \frac{1}{2}$, then H_α is infinite dimensional (Sec. 4A).

The Fourier transformation is a unitary operator in H_α.

The Hermite functions h_k are eigenfunctions of the operator of Fourier transformation. The Fourier invariance of H_α suggests that the condition $u \in H_\alpha$ can be restated as a restriction on the absolute values of the expansion coefficients (u, h_k). This is indeed

* This paper was sponsored by the Ford Foundation.
[1] I. M. Gel'fand and G. E. Silov, *Generalized Functions* (Russian) (State Publisher of Physical & Math. Literature, Moscow, 1958), Vol. II.
[2] S. Mandelbrojt, Ann. Sci. Ec. Norm. Sup. (3e serie) 77, 41 (1960).

The k sum cuts off at $k = n$ because the expansion of ϕ_n in terms of the P_k has no contribution for $k > n$. Substituting Eq. (14),

$$D_n = \left\{\sum_{k=0}^{n} C_k[P_k, \phi_n]\right\}/(P_n, \phi_n). \tag{19}$$

A crucial point here is that D_n depends only on the first n of the C_k, so a given number of terms in the left-hand series of Eq. (17) completely determines an equal number of terms in the right-hand series.

For the sake of illustration, the set $\{P_n(x)\}$ were described as having weight function unity. Actually, no essential use was made of this specialization. The results above permit a direct transformation from an expansion in terms of any weight function to an expansion in terms of any other weight function (subject to the stated analytical limitations).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Hilbert Spaces of Type S*

ALEXANDER GROSSMANN
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received 30 June 1964)

Certain Hilbert spaces of generalized functions are examined. They contain the space of tempered distributions, are invariant with respect to the Fourier transformation and contain functions that increase rapidly at infinity.

1. INTRODUCTION

THIS paper is concerned with a family of Hilbert spaces which are closely related to the topological vector spaces "of type S" of Gel'fand and Silov. (See Refs. 1 and 2). These Hilbert spaces share with $L^{(2)}$, and with the space of tempered distributions, the property of being invariant with respect to the Fourier transformation. Their elements are not restricted to polynomial growth at infinity, a fact which makes them sometimes more suitable than tempered distributions.

The definitions and results of Secs. 2, 3, and 4 can be illustrated on a simple special case:

(A) The first—and decisive—stage of the game is the construction of Hilbert spaces of testing functions, denoted here by H_α.

Let S be the vector space of infinitely differentiable functions $u(x)(-\infty < x < \infty)$ such that

$$\sup |x^k D^m u(x)| < \infty \quad (k, m = 1, 2, \cdots).$$

In S, consider the sequence of operators

$$\begin{aligned} M_0 &= 1 \\ M_1 &= x^2 + p^2 \\ &\vdots \quad\quad \vdots \\ M_n &= xM_{n-1}x + pM_{n-1}p \\ &\vdots \quad\quad \vdots \end{aligned} \tag{1.1}$$

Here $p = i^{-1}D$, and D is the operator of differentiation. The expression (u, v) means $\int u^*(x)v(x)dx$.

For every $u \in S$, $u \neq 0$, the numbers $(u, M_n u)$ are positive.

Now let α be a positive number. Define H_α as the set of functions $u \in S$ for which

$$\sum_n [\Gamma(\alpha n)]^{-2}(u, M_n u) < \infty. \tag{1.2}$$

Here Γ is Euler's gamma function.

With the obvious scalar product, H_α becomes a Hilbert space (Sec. 3B). If $\alpha < \frac{1}{2}$, then H_α consists of the element $u = 0$ only; if $\alpha > \frac{1}{2}$, then H_α is infinite dimensional (Sec. 4A).

The Fourier transformation is a unitary operator in H_α.

The Hermite functions h_k are eigenfunctions of the operator of Fourier transformation. The Fourier invariance of H_α suggests that the condition $u \in H_\alpha$ can be restated as a restriction on the absolute values of the expansion coefficients (u, h_k). This is indeed

* This paper was sponsored by the Ford Foundation.
[1] I. M. Gel'fand and G. E. Silov, *Generalized Functions* (Russian) (State Publisher of Physical & Math. Literature, Moscow, 1958), Vol. II.
[2] S. Mandelbrojt, Ann. Sci. Ec. Norm. Sup. (3e serie) 77, 41 (1960).

the case (Sec. 3A); consequently the space H_α has a simple characterization in the representation due to Bargmann.[3]

If $u \in H_\alpha$ then $u(x) = O(|x|^2 \exp(-|x|^{1/\alpha}))$ for large $|x|$ (Sec. 4B).

If $\alpha < 1$, then every $u \in H_\alpha$ can be analytically continued to a function $u(x + iy) = u(z)$ which is entire analytic of order $\rho \leq (1 - \alpha)^{-1}$ (Sec. 4E).

If $\alpha = 1$ then every $u \in H_\alpha$ is analytic in a strip along the real axis (Sec. 4F).

If $\alpha > 1$, then H_α contains functions of compact support.

(B) To every Hilbert space H_α of testing functions there corresponds a Hilbert space $H_{\dot\alpha}$ of generalized functions (Sec. 3F). Since every Hilbert space is self-dual, it is possible to identify H_α and $H_{\dot\alpha}$ by the canonical correspondence between the element $u \in H_\alpha$ and the functional $v \to (u, v)_\alpha$. It is convenient not to do this, but to identify a locally integrable function $f(x)$ to the functional $v \to \int f^*(x)v(x)dx$. Hilbert spaces of generalized functions are discussed by Lax,[4] Gel'fand and Vilenkin,[5] and Berezanskii[6] where further references can be found.

The properties of the elements of $H_{\dot\alpha}$ follow easily from properties of H_α. The space S' of tempered distributions is contained in every $H_{\dot\alpha}(\alpha > \frac{1}{2})$. The Fourier transformation is a unitary operator in $H_{\dot\alpha}$. The space $H_{\dot\alpha}$ contains all locally integrable functions $f(x)$ such that $f(x) = O(\exp(|x|^{1/\alpha - \epsilon}))$ for some $\epsilon > 0$. In particular, if $\frac{1}{2} < \alpha < 1$ then $H_{\dot\alpha}$ contains all exponentials. By Fourier invariance, it also contains all delta functions δ_a with arbitrary complex a; their definition is immediate since the testing functions $u \in H_\alpha(\frac{1}{2} < \alpha < 1)$ are entire analytic.

(C) The family of Hilbert spaces H_α, $L^{(2)}$, $H_{\dot\alpha}$ $(\frac{1}{2} < \alpha < \infty)$ forms a convenient framework in various questions of mathematical physics (e.g., in the study of nonnormalizable states and of the propagators of field theory). Some of the applications have been sketched in a previous paper[7]; others will follow.

In actual calculations one mostly deals with bounded operators between Hilbert spaces of the family, and not with bounded operators in individual spaces. Unbounded operators do not occur. Information about norms of certain classes of operators can be found in Secs. 3E, 4C, and 4D.

The families of Hilbert spaces considered in Sec. 4 form what Gel'fand[5] calls an "osnashchennoe" Hilbert space. A slight modification of this concept is described in Sec. 5.

This paper is self-contained in the sense that the proofs are based on explicit calculations and on the simplest properties of Hilbert space. There are two exceptions: we use the concept of order of an entire analytic function (see Ref. 8) and the elementary fact that the space S is complete.

2. THE NUMBERS $(u, M_n u)$

To every function $u \in S$ one can associate an increasing sequence of numbers $(u, M_n u)$. The Hilbert spaces $S(\beta) \subset S$—to be introduced in Sec. 3A—are defined by restrictions on this sequence. The study of the spaces $S(\beta)$ is based on elementary properties of the numbers $(u, M_n u)$, derived in this section. The numbers $(u, M_n u)$ are Fourier invariant (Sec. 2B); they are related to the numbers

$$\sup |x^m D^k u|$$

(Sec. 2C) and to the expansion coefficient of u in a Hermite series (Sec. 2D).

A. Notations and Definitions

Let $x = \{x_1, \cdots, x_d\}$ denote a point in R^d. Write $D_i = \partial/\partial x_i$ and $p_i = i^{-1}D_i$. As usual, x^m denotes a monomial $x_1^{m_1} \cdots x_d^{m_d}$ and $|m|$ stands for $m_1 + \cdots + m_d$. The symbol (u, v) means $\int u^*(x)v(x)dx$. The space of infinitely differentiable functions $u(x)$ such that $\sup_x |x^m D^r u(x)| < \infty$ for all m, r, is denoted by S. The Fourier transformation operator is defined, in S, by

$$(Fu)(y) = (2\pi)^{-\frac{1}{2}d} \int e^{-iy\cdot x} u(x)\, dx.$$

For $n = 0, 1, 2, \cdots$ define, in S, an operator M_n by

$$\begin{aligned} M_0 &= 1 \\ M_1 &= \sum_{i=1}^{d} (x_i^2 + p_i^2) \\ &\vdots \qquad\qquad \vdots \\ M_n &= \sum_{i=1}^{d} (x_i M_{n-1} x_i + p_i M_{n-1} p_i) \\ &\vdots \qquad\qquad \vdots \end{aligned} \tag{2.1}$$

Notice that, for $u \in S$, $u \neq 0$,

$$(u, M_n u) > 0. \tag{2.2}$$

[3] V. Bargmann: Commun. Pure Appl. Math. 14, 187 (1961).
[4] P. D. Lax: Commun. Pure Appl. Math. 8, 615 (1955).
[5] I. M. Gel'fand and N. Ya. Vilenkin, Ref. 1, Vol. IV.
[6] Yu. M. Berezanskii: Usp. Mat. Nauk 18, No. 1, 63 (1963).
[7] A. Grossmann: J. Math. Phys. 5, 1025 (1964).
[8] R. P. Boas, *Entire Functions* (Academic Press Inc., New York, 1954), p. 8.

If $n = 0$, this is obvious, If $n > 0$ it follows from

$$(u, M_n u) = \sum_{i=1}^{d} [(x_i u, M_{n-1} x_i u) + (p_i u, M_{n-1} p_i u)] \tag{2.3}$$

by complete induction. Actually this argument shows that M_n is the sum of $(2d)^n$ terms each of which is positive definite.

B. Fourier Invariance

Theorem 2.1: For every $u \in S$ and every n

$$(u, M_n u) = (Fu, M_n Fu).$$

Proof: For $n = 0$, this is the Parseval equality. For $n > 0$, use complete induction and the operator equalities $p_i F = -F x_i$, $F p_i = x_i F$:

$$(Fu, M_n Fu)$$

$$= \sum_{i=1}^{d} [(x_i Fu, M_{n-1} x_i Fu) + (p_i Fu, M_{n-1} p_i Fu)]$$

$$= \sum_{i=1}^{d} [(F p_i u, M_{n-1} F p_i u) + (F x_i u, M_{n-1} F x_i u)]$$

$$= \sum_{i=1}^{d} [(p_i u, M_{n-1} p_i u) + (x_i u, M_{n-1} x_i u)] = (u, M_n u),$$

which proves the assertion.

C. Relationship between $(u, M_n u)$ and $\sup |x^m D^k u|$

Notice that here n denotes one integer while m and k are d-tuples of integers.

The theorems below show that, for any given $u \in S$, the sequence of numbers $(u, M_n u)^{\frac{1}{2}}$ is not drastically different from the (multiple) sequence $\sup_x |x^m D^k u|$.

Theorem 2.2: Denote by s the integer $s = [\frac{1}{2} d] + 1$ and by C_d the number

$$C_d = \left\{ \int \left[1 + \left(\sum_{i=1}^{d} x_i^2 \right)^s \right]^{-1} dx \right\}^{\frac{1}{2}} \tag{2.4}$$

(s has been chosen so as to make C_d finite). Then

$$\sup |x^m D^k u(x)| \leq (2\pi)^{-\frac{1}{2}d} C_d [(u, M_{|m|+|k|} u) + (u, M_{|m|+|k|+s} u)]^{\frac{1}{2}} \tag{2.5}$$

for every $u \in S$.

Proof: For every $u \in S$ and for every x we have

$$|u(x)| \leq (2\pi)^{-\frac{1}{2}d} \int |(Fu)(y)| \, dy.$$

Furthermore, for every $v \in S$ and with $r^2 = \sum_{i=1}^{d} x_i^2$

$$\int |v(x)| \, dx = \int |v(x)| (1 + r^{2s})^{\frac{1}{2}} (1 + r^{2s})^{-\frac{1}{2}} \, dx$$

$$\leq C_d \left\{ \int |v(x)|^2 (1 + r^{2s}) \, dx \right\}^{\frac{1}{2}}$$

$$\leq C_d [(v, (M_0 + M_s) v)]^{\frac{1}{2}} = C_d [(Fv, (M_0 + M_s) Fv)]^{\frac{1}{2}}.$$

So

$$\sup |u(x)| \leq (2\pi)^{-\frac{1}{2}d} C_d [(u, (M_0 + M_s) u)]^{\frac{1}{2}}.$$

From (2.3) and the positivity of $(u, M_n u)$ it follows that $(x_i u, M_n x_i u) \leq (u, M_{n+1} u)$ and $(p_i u, M_n p_i u) \leq (u, M_{n+1} u)$. So

$$\sup |x_i u(x)| \leq (2\pi)^{-\frac{1}{2}d} C_d [(x_i u, (M_0 + M_s) x_i u)]^{\frac{1}{2}} \leq (2\pi)^{-\frac{1}{2}d} C_d [(u, (M_1 + M_{s+1}) u)]^{\frac{1}{2}}.$$

Repeated application of this inequality and of the corresponding inequality for $D_i u$ gives (2.5).

Remark: It follows from (2.5) that

$$\sup |x^m D^k u(x)| \leq 2^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}d} C_d (u, M_{|m|+|k|+s} u)^{\frac{1}{2}} \tag{2.6}$$

since the numbers $(u, M_n u)$ increase with n (See Sec. 2E).

An immediate consequence of Theorem 2.2 is

Theorem 2.3: Let $u^{(\nu)} (\nu = 1, 2, \cdots)$ be a sequence of elements of S such that, for every (fixed) n

$$\lim (u^{(\nu)}, M_n u^{(\nu)}) = 0.$$

Then the sequence $u^{(\nu)}$ converges to zero in the topology of S.

Estimates in the other direction are provided by

Theorem 2.4: Denote by $|u|_{2n}$ the number

$$|u|_{2n} = \sum_{m,k} \sup |x^m D^k u|$$

the sum being taken over all m, k such that $|m| + |k| \leq 2n$. Let $|u|_L$ be the number

$$|u|_L = \int |u(x)| \, dx.$$

Here u is any element of S. Then: Given any $\epsilon > 0$, there exists a $n_0 = n_0(\epsilon)$ such that

$$(u, M_n u) \leq n! \, (2 + \epsilon)^n \, |u|_{2n} \, |u|_L \tag{2.7}$$

for all $n > n_0$.

Proof: One can commute in M_n the operators x_i to the left and write

$$M_n = \sum_{m,k} c_n(m; k) x^m D^k. \tag{2.8}$$

The coefficients $c_n(m; k)$ are different from zero only if $0 \leq |m| + |k| \leq 2n$ and if $|m| + |k|$ is even. Because of (2.1), they satisfy the recursion relation

$$c_{n+1}(m; k) = \sum_{i=1}^{d} \{c_n(m - 2_i; k) - c_n(m; k - 2_i) + (k_i + 1)c_n(m - 1_i; k + 1_i) - (m_i + 1)c_n(m + 1_i; k - 1_i)\} \quad (2.9)$$

which start with $c_0(0; 0) = 1$. Here 1_i is the d-tuple which has 1 at the ith place and 0 elsewhere.

Denote by c_n the largest of the numbers $|c_n(m; k)|$ for given n. Then (2.9) gives

$$|c_{n+1}(m; k)| \leq \sum_{i=1}^{d} \{2c_n + (k_i + 1)c_n + (m_i + 1)c_n\} = (4d + |k| + |m|)c_n \leq (4d + 2n)c_n = 2n(1 + 4d(2n)^{-1})c_n$$

so that, for sufficiently large n, $c_n < (2 + \epsilon)^n n!$. Then

$$(u, M_n u) = \sum_{m,k} c_n(m; k)(u, x^m D^k u) \leq \{\sum_{m,k} |c_n(m; k)| \sup |x^m D^k u|\} \int |u(x)|\, dx$$

proves the assertion.

Corollary: If $u^{(\gamma)} \in S$ for $\gamma = 1, 2, \cdots$ and if $u^{(\gamma)} \to 0$ in S, then, for every n, $\lim (u^{(\gamma)}, M_n u^{(\gamma)}) = 0$.

D. Expressing $(u, M_n u)$ in terms of (u, h_k)

It will now be shown that the numbers $(u, M_n u)$ can be readily expressed in terms of the Fourier coefficients of u with respect to Hermite functions.

Let

$$\xi_j = 2^{-\frac{1}{2}}(x_j + ip_j) \quad (j = 1, 2, \cdots, d)$$
$$\eta_j = 2^{-\frac{1}{2}}(x_j - ip_j).$$

The definition (2.1) becomes

$$M_n = \sum_{i=1}^{d} (\xi_i M_{n-1} \eta_i + \eta_i M_{n-1} \xi_i). \quad (2.10)$$

Denote by h_k the Hermite functions

$$h_k = (k!)^{-\frac{1}{2}} \eta^k h_0 = \pi^{-\frac{1}{4}d} \prod_{i=1}^{d} (2^{k_i} k_i!)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x_i^2) H_{k_i}(x_i)$$

$$h_0 = \pi^{-\frac{1}{4}d} \exp\left(-\frac{1}{2} \sum_{i=1}^{d} x_i^2\right),$$

where the $H_{k_i}(x_i)$ are Hermite polynomials.

Notice that

$$(h_k, M_n h_m) = 0 \quad \text{if} \quad k \neq m. \quad (2.11)$$

For $n = 0$, this is the orthogonality property of Hermite functions. For arbitrary n, (2.11) follows from (2.10) by the definition of h_k and by induction.

Denote by $a(n; k)$ the number

$$a(n; k) = (h_k, M_n h_k). \quad (2.12)$$

Theorem 2.5: For every $u \in S$ and for every n,

$$(u, M_n u) = \sum_k a(n; k)\, |(h_k, u)|^2. \quad (2.13)$$

Proof: Notice that $u = \sum_k h_k(h_k, u)$ and use Eqs. (2.11) and (2.12).

E. Estimates of $a(n; k)$

It follows from (2.12) and (2.10) that

$$a(0; k) = 1 \quad \text{for every} \quad k, \quad (2.14)$$

that $a(n; k)$ depends only on $|k|$ and that

$$a(n; |k|) = (d + |k|)a(n - 1; |k| + 1) + |k|\, a(n - 1; |k| - 1), \quad (n \geq 1). \quad (2.15)$$

The recursion relations (2.15) give the estimates

$$(d + |k|)_n \leq a(n; k) \leq 2^n(\tfrac{1}{2}d + |k|)_n. \quad (2.16)$$

Here

$$(c)_n = c(c + 1) \cdots (c + n - 1) = \Gamma(c + n)/\Gamma(c).$$

3. THE HILBERT SPACES $S(\beta)$ AND $S(\bar{\beta})$

To every sequence $\{\beta\} = \{\beta_1, \beta_2, \cdots\}$ of positive numbers one can associate a Hilbert space $S(\beta) \subset S$ defined by (3.1). The elements of this space play the role of testing functions. If the numbers β_n do not decrease sufficiently fast, then the space $S(\beta)$ consists of the element $u = 0$ only. If the numbers β_n satisfy the condition (3.4), then $S(\beta)$ is infinite-dimensional. The natural embedding of $S(\beta)$ into into $L^{(2)}$ is an operator of trace class. The polar decomposition of this operator allows a simple construction of the space $S(\bar{\beta}) \supset S'$ of generalized functions (Sec. 3F); $S(\bar{\beta})$ is also a Hilbert space. The properties of elements of $S(\beta)$ and $S(\bar{\beta})$ are studied in Secs. 3C, 3D, and 3F. The Fourier transformation is a unitary operator in $S(\beta)$ and in $S(\bar{\beta})$. The multiplication by certain entire analytic functions defines bounded linear transformations between spaces $S(\beta)$ (Sec. 3E).

The spaces $S(\beta)$ and $S(\bar{\beta})$ may be called Hilbert spaces by type S.

A. Definitions

The space $S(\beta)$ can be defined either by the condition (3.1) on $(u, M_n u)$ or by the condition (3.3) on (h_k, u).

Definition: Let $\{\beta\} = \{\beta_0, \beta_1, \cdots, \beta_n, \cdots\}$ be a sequence of positive numbers. Denote by $S(\beta)$ the set of all $u \in S$ which are such that

$$\sum_{n=0}^{\infty} \beta_n (u, M_n u) < \infty . \tag{3.1}$$

Since $(u, M_n u) > 0$ for $u \in S$, $u \neq 0$ (See Sec. 2A), it is clear that $S(\beta)$ is a pre-Hilbert space (i.e., a space with scalar product but not necessarily complete).

Another characterization of $S(\beta)$ is given by

Theorem 3.1: Define a (multiple) sequence of numbers λ_k $(0 \leq \lambda_k < \infty)$ by

$$\lambda_k^{-2} = \sum_{n=0}^{\infty} \beta_n a(n; k), \tag{3.2}$$

where the $a(n; k)$ are given by (2.12) [or, equivalently, by (2.14) and (2.15)]. Then $S(\beta)$ consists exactly of the $u \in S$ for which

$$\sum_k |(u, h_k)|^2 \lambda_k^{-2} < \infty . \tag{3.3}$$

Proof: The assertion is an immediate consequence of Theorem 2.5.

B. Nontriviality and Completeness

If the numbers β_n decrease sufficiently rapidly, then $S(\beta)$ is an infinite-dimensional Hilbert space (Theorem 3.2). If the numbers β_n decrease too slowly, then $S(\beta)$ consists of the element $u = 0$ only (Theorem 3.3). $S(\beta)$ is a Hilbert space.

Theorem 3.2: Assume that the sequence $\{\beta\}$ is such that, for every k,

$$\lambda_k^{-2} \equiv \sum_n \beta_n a(n; k) < \infty , \tag{3.4}$$

i.e., that $\lambda_k \neq 0$ for all k. Then $S(\beta)$ is a separable infinite-dimensional Hilbert space with the scalar product

$$(u, v)_\beta = \sum_n \beta_n (u, M_n v) = \sum_k (u, h_k) \lambda_k^{-2} (h_k, v). \tag{3.5}$$

The elements $\lambda_k h_k$ form an orthonormal basis of $S(\beta)$.

Proof: The condition (3.4) insures that the scalar product (h_k, h_i) is defined and equal to $\lambda_k^{-2} \delta_{ki}$. So $S(\beta)$ is infinite dimensional and the functions $\lambda_k h_k$ are orthonormal in $S(\beta)$. It remains to be shown that $S(\beta)$ is complete with respect to the norm $|| \; ||_\beta$ defined by (3.5). If $u \in S(\beta)$, then the sequence of numbers $\varphi_k = \lambda_k^{-1}(h_k, u)$ belongs to $l^{(2)}$ (i.e., satisfies $\sum_k |\varphi_k|^2 < \infty$), and $||u||_\beta = ||\varphi||^{l^{(2)}}$. The completeness of $S(\beta)$ will be proved if we show that, for every $\varphi \in l^{(2)}$, there exists a $u \in S(\beta)$ such that $u = \sum_k \lambda_k \varphi_k h_k$. This means: we have to prove that the series $\sum_k \lambda_k \varphi_k h_k$ converges in $S(\beta)$.

Consider the partial sums

$$s_\varphi^{(N)} = \sum_{|k| \leq N} \lambda_k \varphi_k h_k$$

They belong to $S(\beta)$ and form a Cauchy sequence in $S(\beta)$, since

$$||s^{(N)} - s^{(N+K)}||_\beta = \sum_{k=N}^{N+K} |\varphi_k|^2 . \tag{3.6}$$

Notice that

$$(u, u)_\beta \geq \beta_n (u, M_n u). \tag{3.7}$$

By Theorem 2.3, then, the elements $s^{(N)}$ form a Cauchy sequence in S. Since S is complete, there exists an element

$$u_\varphi = \lim_{N \to \infty} s_\varphi^{(N)} \in S,$$

the limit being taken in the topology of S. From the construction of u_φ it follows that $(h_k, u_\varphi) = \lambda_k \varphi_k$ for every k. Consequently

$$(u_\varphi, u_\varphi)_\beta = \sum_k (u_\varphi, h_k) \lambda_k^{-2} (h_k, u_\varphi) = \sum_k |\varphi_k|^2$$

which shows that $u_\varphi \in S(\beta)$ and completes the proof of Theorem 3.2.

Theorem 3.3: Let β_n be a sequence of positive numbers such that, for every k,

$$\sum_n \beta_n a(n; k) = +\infty \tag{3.8}$$

Then $S(\beta)$ consists of the single element $u = 0$.

Proof: If $u \in S$ and $u \neq 0$, then there exists at least one k such that $(h_k, u) \neq 0$. It follows then from (2.13) that $(u, M_n u) \geq a(n; k) \; |(h_k, u)|^2$ for every n. Consequently

$$(u, u)_\beta = \sum_n \beta_n (u, M_n u) \geq |(h_k, u)|^2 \sum_n \beta_n a(n; k) = \infty \tag{3.9}$$

which shows that u does not belong to $S(\beta)$.

Theorem 3.3a: If $S(\beta)$ is infinite dimensional, then (3.4) holds for all k.

Proof: The numbers $a(n; k)$ increase with $|k|$. Consequently, if (3.4) fails to hold for some k, it

also fails to hold for all k' such that $|k'| > |k|$, and $S(\beta)$ is finite dimensional.

From now on it will be assumed—unless the contrary is stated—that $S(\beta)$ is infinite dimensional, so that (3.4) holds for all k.

C. Behavior of $\sup_x |x^m D^k u|$

By the results of Sec. 2C, restrictions on the sequence $(u, M_n u)$ entail restrictions on the sequence $\sup |x^m D^k u|$.

Theorem 3.4: If $u \in S(\beta)$, then, for every m, k

$$\sup |x^m D^k u| \leq 2^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}d} C_d \, ||u||_\beta \, \beta_{m+|k|+s}^{-\frac{1}{2}}, \quad (3.10)$$

where $s = [\frac{1}{2}d] + 1$ and C_d is given by (2.4).

Proof: The assertion follows from (2.6) and (3.7).

Theorem 3.5: Let $u \in S$. Define $|u|_{2n}$ as

$$|u|_{2n} = \sum_{|m|+|k| \leq 2n} \sup |x^m D^k u|.$$

If the series

$$\sum_n n! \, (2 + \epsilon)^n \, |u|_{2n} \, \beta_n \quad (3.11)$$

converges for some $\epsilon > 0$, then $u \in S(\beta)$.

Proof: This theorem is an immediate consequence of (2.7).

D. Condition for Analyticity of $u \in S(\beta)$

Sufficiently stringent conditions on the sequence $\sup |D^k u|$ give analyticity of u.

Theorem 3.6: Assume that the sequence $\{\beta\}$ defining $S(\beta)$ is such that, for sufficiently large n, and some $\epsilon > 0$,

$$\beta_n > n^{-(2-\epsilon)n}. \quad (3.12)$$

This means that

$$\limsup (2n \log n)^{-1} \log (\beta_n^{-1}) \equiv \alpha < 1. \quad (3.13)$$

Then every $u(x) \in S(\beta)$ can be analytically continued to a function $u(x + iy) = u(z)$ which is entire analytic of order $\rho \leq (1 - \alpha)^{-1}$ in every one of its arguments.

Proof: Notice that, by (3.10)

$$|D^k u(x)| \leq C_u \beta_{|k|+s}^{-\frac{1}{2}} \quad (3.14)$$

for every x and every k. In order to show that u is entire analytic, use (3.14) and the Taylor formula with remainder. In order to estimate the order of u, use the formula

$$\rho = \limsup \frac{k \log k}{\log (1/|a_k|)} \quad (3.15)$$

for the order of the entire function $u(z) = \sum_k a_k z^k$.

Remark: It will be seen that (3.12) is compatible with (3.4) so that there exist infinite-dimensional spaces $S(\beta)$ consisting of entire analytic functions.

Remark: If $u \in S(\beta)$ then the complex conjugate $u^*(x)$ also belongs to $S(\beta)$, since $(u^*, M_n u^*) = (u, M_n u)$. If u has the analytic continuation $u(z)$, then u^* has the analytic continuation $u^*(z^*)$.

E. Entire Functions as Multipliers

In applications one often considers simultaneously several Hilbert spaces of type S. It is then important to know whether an operator (e.g., a differentiation operator or the multiplication operator by a function) is a bounded transformation from one Hilbert space into the other. A criterion for this will now be given.

Let $\{\beta\}$ and $\{\beta'\}$ be two sequences that both satisfy (3.4) so that the Hilbert spaces $S(\beta)$ and $S(\beta')$ are both infinite dimensional.

Let $B(z) = B(x + iy) = \sum_k B_k z^k$ be an entire analytic function of $z = \{z_1, \cdots, z_d\}$. Assume that the numbers B_k satisfy the condition

$$\sum_{n=0}^{\infty} \left[\sum_k |B_k| \left(\frac{\beta_n'}{\beta_{n+|k|}} \right)^{\frac{1}{2}} \right]^2 \equiv N^2 < \infty. \quad (3.16)$$

Theorem 3.7: If the above conditions are satisfied then

(a) For every $u \in S(\beta)$ the product $B(x)u(x)$ belongs to $S(\beta')$.

(b) The bound norm of the mapping $u(x) \to B(x)u(x)$ of $S(\beta)$ into $S(\beta')$ does not exceed the number N defined by (3.16).

(c) The order of the function $B(z)$ (in any one of its d complex arguments) does not exceed 2.

Remark: In the statement of Theorem 3.7, the multiplication operator $B(x)$ can be replaced by a differentiation operator $B(D)$. This follows from the fact that $D = iF^{-1}xF$ and that the Fourier transformation F is a unitary operator in $S(\beta)$. (See Sec. 3G).

Proof: Lemma 1: The numbers B_k satisfy

$$\sum_k |B_k| \, (|k!|)^{\frac{1}{2}} < \infty. \quad (3.17)$$

Proof of Lemma 1: The estimates (2.16) show that $n! \leq a(n; k)$ for all n and k. Consequently $\sum_n \beta_n n! \leq \sum_n \beta_n a(n; k) < \infty$. This means that

$$\beta_n^{-\frac{1}{2}} = \gamma_n^{-1}(n!)^{\frac{1}{2}}, \quad \text{with} \quad \{\gamma_n\} \in \rho^{(2)}.$$

So

$$\beta_n^{-\frac{1}{2}} \geq (\gamma_{\max})^{-1}(n!)^{\frac{1}{2}},$$

where $\gamma_{\max}$ is the largest of the numbers γ_n. Consequently

$$\sum_k |B_k|\, \beta_{n+|k|}^{-\frac{1}{2}} \geq (\gamma_{\max})^{-1} \sum_k |B_k|\, [(n+|k|)!]^{\frac{1}{2}}. \quad (3.18)$$

By the assumption (3.16), the series on the lhs of (3.18) is convergent. So

$$\sum_k B_k\, (|k!|)^{\frac{1}{2}} \leq \sum_k B_k [(n+|k|)!]^{\frac{1}{2}} < \infty \qquad \text{q.e.d.}$$

The assertion (c) follows from the lemma by (3.15).

In order to prove (b) and (a), denote by $s_r (r = 1, 2, \cdots)$ the partial sums

$$s_r = \sum_{|k|=0}^{r} B_k x^k u(x).$$

Introduce the notation

$$I(n; v) = (v, M_n v)^{\frac{1}{2}} \qquad (v \in S).$$

By the results of Sec. 2A, we have

$$I(n; \lambda v) = |\lambda|\, I(n; v) \qquad (\lambda \text{ arbitrary})$$
$$I(n; v_1 + v_2) \leq I(n; v_1) + I(n; v_2)$$
$$I(n; x^k v) \leq I(n + |k|; v).$$

Lemma 2: For every fixed n and for every fixed $u \in S(\beta)$, the functions $I(n; s_r)$ form a Cauchy sequence with respect to the norm $I(n; s_r)$.

Proof of Lemma 2: Notice that $u \in S(\beta)$ means $I(n; u) = \beta_n^{-\frac{1}{2}} \eta_n$ with $\{\eta_n\} \in l^{(2)}$. So

$$I(n; s_{r+q} - s_r) = I\Big(n; \sum_{|k|=r}^{r+q} B_k x^k u\Big)$$
$$\leq \sum_{|k|=r}^{r+q} |B_k|\, I(n + |k|; u) = \sum_{|k|=r}^{r+q} |B_k|\, \beta_{n+|k|}^{-\frac{1}{2}} \eta_{n+|k|}$$
$$\leq |\eta_{\max}| \sum_{|k|=r}^{r+q} |B_k|\, \beta_{n+|k|}^{-\frac{1}{2}}. \quad (3.19)$$

By the assumption (3.16), the series $\sum_k |B_k| \beta_{n+k}^{-\frac{1}{2}}$ is convergent. So the rhs of (3.19) can be made arbitrarily small by the choice of sufficiently large r. This proves the lemma.

By Theorem 2.3, it follows that the partial sums s_r converge, in S, to an element of S. This element is $B(x)u(x)$ since $\lim s_r(x) = B(x)u(x)$ for every x.

We have now to show that $Bu \in S(\beta')$ and find a bound on its norm. For every n,

$$I(n; Bu) = \lim_{r\to\infty} I(n; s_r)$$
$$\leq \sum_k |B_k|\, I(n + |k|; u) \leq \sum_k |B_k|\, \beta_{n+|k|}^{-\frac{1}{2}} \eta_{n+|k|}$$
$$\leq |\eta_{\max}| \sum_k |B_k|\, \beta_{n+|k|}^{-\frac{1}{2}} \quad [\eta_n = \beta_n^{\frac{1}{2}} I(n; u); \{\eta_n\} \in l^{(2)}].$$

So

$$(\beta_n')^{\frac{1}{2}} I(n; Bu) \leq |\eta_{\max}| \sum_k |B_k|\, (\beta_n')^{\frac{1}{2}} (\beta_{n+|k|})^{-\frac{1}{2}}$$

and

$$||Bu||_{\beta'}^2 = \sum_n \beta_n' I^2(n; Bu) \leq |\eta_{\max}|^2 \sum_n \Big(\sum_k |B_k|\, (\beta_n')^{\frac{1}{2}} \times (\beta_{n+|k|})^{-\frac{1}{2}}\Big)^2 = N^2\, |\eta_{\max}|^2 \leq N^2\, ||u||_\beta^2$$

since

$$|\eta_{\max}|^2 \leq \sum_n |\eta_n|^2 = \sum_n \beta_n I^2(n; u) = ||u||_\beta^2.$$

This completes the proof of the theorem.

A straightforward consequence of Theorem 3.7 concerns the inclusion relations between the various spaces $S(\beta)$ and is obtained by setting $B(x) = 1$ (i.e., $B_k = \delta_{k0}$). It is

Theorem 3.8: Let $\{\beta\}$ and $\{\beta'\}$ be such that

$$\sum_{n=0}^{\infty} \frac{\beta_n'}{\beta_n} \equiv N^2 < \infty. \quad (3.20)$$

Then $S(\beta) \subseteq S(\beta')$. The bound norm of the natural embedding of $S(\beta)$ into $S(\beta')$ does not exceed N.

F. The Space $S(\bar{\beta})$

The Hilbert space $S(\bar{\beta})$ is, roughly speaking, the space of distributions corresponding to the space $S(\beta)$ of testing functions.

Since $S(\beta)$ is a Hilbert space, it is dual to itself. This means that there exists a canonical one-to-one correspondence between the elements $u \in S(\beta)$ and the continuous linear functionals

$$v \to (u, v)_\beta; \qquad v \in S(\beta). \quad (3.21)$$

The element u can be identified with the functional (3.21).

It is often more convenient to identify u with the functional

$$v \to \int u^*(x) v(x)\, dx, \quad (3.22)$$

which involves the scalar product in the Hilbert space $L^{(2)} \supset S(\beta)$ If this is done, then arbitrary continuous linear functionals on $S(\beta)$ form a Hilbert space $S(\bar{\beta}) \supset L^{(2)}$ which will be studied in the present section.

Denote by $E_{o\beta}$ the natural embedding operator of $S(\beta)$ into $L^{(2)}$, i.e., the operator which to every $u \in S(\beta)$ associates the same u, considered as an element of $L^{(2)}$.

Let $(E_{o\beta})_{\beta o}^*$ be the adjoint of $E_{o\beta}$, i.e., the mapping from $L^{(2)}$ into $S(\beta)$ defined by

$$(u, (E_{o\beta})_{\beta o}^* f)_\beta = (E_{o\beta} u, f) \quad (3.23)$$

for every $u \in S(\beta)$ and every $f \in L^{(2)}$.

Denote by $L_{oo}^{(\beta)}$ the positive definite operator

$$L_{oo}^{(\beta)} = [E_{o\beta}(E_{o\beta})_{\beta o}^{*}]^{\frac{1}{2}}. \tag{3.24}$$

Definition: Let $S(\bar{\beta})$ denote the completion of $L^{(2)}$ with respect to the norm defined by the scalar product

$$(f, g)_{\bar{\beta}} = (L_{oo}^{(\beta)} f, L_{oo}^{(\beta)} g). \tag{3.25}$$

Denote by $E_{\bar{\beta}o}$ the natural embedding of $L^{(2)}$ into $S(\bar{\beta})$.

Theorem 3.9: (a) The operator $L_{oo}^{(\beta)}$ (acting in $L^{(2)}$) is given by

$$L_{oo}^{(\beta)} g = \sum_k h_k \lambda_k (h_k, g), \tag{3.26}$$

where the numbers λ_k are defined by (3.2) and where the h_k form the orthonormal basis of $L^{(2)}$ defined in Sec. 2D. (b) The polar decomposition (i.e., the decomposition into an isometric factor and a positive semidefinite factor) of $E_{o\beta}$ is

$$E_{o\beta} = L_{oo}^{(\beta)} U_{o\beta}, \tag{3.27}$$

where the unitary operator $U_{o\beta}$ transforms the orthonormal basis $\{\lambda_k h_k\}$ of $S(\beta)$ into the orthonormal basis $\{h_k\}$ of $L^{(2)}$. (c) The polar decomposition of $E_{\bar{\beta}o}$ is

$$E_{\bar{\beta}o} = U_{\bar{\beta}o} L_{oo}^{(\beta)}, \tag{3.28}$$

where the unitary operator $U_{\bar{\beta}o}$ transforms the orthonormal basis $\{h_k\}$ of $L^{(2)}$ into the orthonormal basis $\lambda_k^{-1} h_k$ of $S(\bar{\beta})$.

The assertions of Theorem 3.9 are special cases of known elementary results and need not be proved here.

Theorem 3.10: The natural embeddings $E_{o\beta}$ and $E_{\bar{\beta}o}$ are of trace class.

Proof: It is sufficient to prove that $\sum_k \lambda_k < \infty$. By (2.16),

$$\lambda_k^{-2} \geq \sum_n \beta_n (d + |k|)_n > \beta_{n_0} (d + |k|)_{n_0} > \beta_{n_0} |k|^{n_0}$$

for every n_0. Consequently

$$\lambda_k < \beta_{n_0}^{-\frac{1}{2}} |k|^{-\frac{1}{2} n_0}. \tag{3.29}$$

The choice of a sufficiently large n_0 gives the assertion.

Remark: The above proof shows that $\sum_k \lambda_k^{\gamma} < \infty$ for every $\gamma > 0$.

Remark: The fact that $E_{o\beta}$ is of trace class allows the application of known results[5,6] on generalized eigenvalue expansions.

The bilinear functional to be defined now corresponds to the value that a distribution takes on a testing function.

Definition: Let $f \in S(\bar{\beta})$ and $u \in S(\beta)$. Define the number $\langle f \mid u \rangle = \langle u \mid f \rangle^*$ by

$$\langle f \mid u \rangle = (U_{o\bar{\beta}} f, U_{o\beta} u), \tag{3.30}$$

where $U_{o\bar{\beta}} = (U_{\bar{\beta}o})_{o\bar{\beta}}^{*} = (U_{\bar{\beta}o})_{o\bar{\beta}}^{-1}$ and the unitary operators $U_{o\beta}$ and $U_{\bar{\beta}o}$ are described in Theorem 3.9.

Notice that the scalar product on the rhs of (3.30) is in $L^{(2)}$.

From the definition (3.30) it follows immediately that $|\langle f \mid u \rangle| \leq ||f||_{\bar{\beta}} \, ||u||_{\beta}$, that $\langle f \mid u \rangle = 0$ for all $u \in S(\beta)$ entails $f = 0$, and that $\langle f \mid u \rangle = 0$ for all $f \in S(\bar{\beta})$ entails $u = 0$.

The bilinear form (3.30) is in a sense an extension of the scalar product in $L^{(2)}$. This is shown by

Theorem 3.11: If $f \in L^{(2)} \subset S(\bar{\beta})$ then for every $u \in S(\beta)$,

$$\langle f \mid u \rangle = \int f^*(x) u(x) \, dx. \tag{3.31}$$

Proof: The assumption means that $f \in S(\bar{\beta})$ is in the range of the natural embedding operator $E_{\bar{\beta}o}$. Write $f = E_{\bar{\beta}o} f_0$. Then

$$\langle f \mid u \rangle = (U_{o\bar{\beta}} E_{\bar{\beta}o} f_0, U_{o\beta} u) = (U_{o\bar{\beta}} U_{\bar{\beta}o} L_{oo}^{(\beta)} f_0, U_{o\beta} u)$$

$$= (L_{oo}^{(\beta)} f_0, (L_{oo}^{(\beta)})^{-1} E_{o\beta} u) = (f_0, E_{o\beta} u) = \int f^*(x) u(x) \, dx$$

which proves the theorem.

The element $f \in S(\bar{\beta})$ will be identified with the linear functional $u \to \langle f \mid u \rangle$ on $S(\beta)$. As usual, the identification correspondence is antilinear.

Theorem 3.12: The space S' of tempered distributions is contained in $S(\bar{\beta})$.

Proof: Let $u \to l(u)$ be any element of S'. Consider the restriction of l to $S(\beta) \subset S$. Since $S(\beta)$ is dense in S (it contains all finite linear combinations of h_k), the restriction of l to $S(\beta)$ determines l completely. Let $\{u_\nu\}$ be a sequence of elements of $S(\beta)$ such that $||u_\nu||_\beta \to 0$. Then, by (3.10), $u_\nu \to 0$ in S and $l(u_\nu) \to 0$. This means that $l(u)$ is a continuous linear functional over $S(\beta)$. Consequently there exists a $v_l \in S(\beta)$ such that $l(u) = (v_l, u)_\beta$. Then the element $f = U_{\bar{\beta}o} U_{o\beta} v_l$ where $f \in S(\bar{\beta})$ satisfies $\langle f \mid u \rangle = l(u)$ q.e.d.

Remark: It can be shown that the natural embedding of S' into $S(\bar{\beta})$ is continuous.

Some locally integrable functions belong to $S(\bar{\beta})$. This is shown by

Theorem 3.13: Let $f(x)$ be locally integrable with respect to the Lebesgue measure. Assume that there exists a constant C_f such that, for every $u \in S(\beta)$,

$$\left| \int f^*(x)u(x)\,dx \right| \leq C_f\, ||u||_\beta. \qquad (3.32)$$

Then there exists one and only one $f \in S(\bar{\beta})$ such that

$$\langle f \mid u \rangle = \int f^*(x)u(x)\,dx. \qquad (3.33)$$

Proof: By the assumption (3.32), the correspondence $u \to \int f^* u dx$ is a continuous linear functional on $S(\beta)$. The proof proceeds as that of Theorem 3.12.

G. Fourier Invariance

By Theorem 2.1 and the definition of $S(\beta)$, the Fourier transformation F is a unitary operator in $S(\beta)$.

Definition: The Fourier transformation F is defined in $S(\bar{\beta})$ by

$$\langle Ff \mid Fu \rangle = \langle f \mid u \rangle; \qquad f \in S(\bar{\beta}) \qquad (3.34)$$

for every $u \in S(\beta)$.

Theorem 3.14: The Fourier transformation is a unitary operator in $S(\bar{\beta})$.

Proof: It was seen above that $U_{\beta_0}h_k \equiv (U_{0\beta})^{-1}_{\beta_0}h_k = \lambda_k h_k$ and that $U_{\bar{\beta}_0}h_k = \lambda_k^{-1}h_k$. Because of $Fh_k = i^{-k}h_k$, the unitary operators $U_{\bar{\beta}_0}$ and $U_{0\beta}$ commute with F. So

$$\begin{aligned}(Ff, Fg)_{\bar{\beta}} &= (U_{\beta_0}U_{0\bar{\beta}}Ff,\ U_{\beta_0}U_{0\bar{\beta}}Fg)_\beta \\ &= (FU_{\beta_0}U_{0\bar{\beta}}f,\ FU_{\beta_0}U_{0\bar{\beta}}g)_\beta \\ &= (U_{\beta_0}U_{0\bar{\beta}}f,\ U_{\bar{\beta}_0}U_{0\bar{\beta}}g)_\beta = (f, g)_{\bar{\beta}} \quad \text{q.e.d.}\end{aligned}$$

H. Estimates of the Numbers λ_k

In the study of relationships between various spaces of type S it is often useful to know the asymptotic behavior of the numbers λ_k as $|k| \to \infty$.

Notice that $a(n; k)$ is a polynomial in $\kappa = |k|$. [See (2.15).] Consequently λ_k^{-2} is an entire function of κ.

Theorem 3.15: Assume that the sequence β_n is such that the function $\sum_n \zeta^n \beta_n$ is entire analytic (in ζ) of order $\rho < 1$.

Then the function

$$g(\kappa) = \lambda_k^{-2} = \sum_n \beta_n a(n; k); \qquad \kappa = |k| \qquad (3.35)$$

is entire analytic of the same order ρ.

Proof: By (2.16),

$$\sum_{n=0}^{\infty} (d + \kappa)_n \beta_n \leq g(\kappa) \leq \sum_{n=0}^{\infty} 2^n (\tfrac{1}{2}d + \kappa)_n \beta_n = \sum_{n=0}^{\infty} \beta'_n (\tfrac{1}{2}d + \kappa)_n$$

with $\beta'_n = 2^n \beta_n$. The assertion of the theorem follows from the

Lemma: Let γ_n be a sequence of positive numbers such that $\sum_n \gamma_n \zeta^n = f(\zeta)$ is entire analytic of order $\rho < 1$. Then $f_1(\zeta) = \sum_n \gamma_n (\zeta)_n$ is also entire analytic of order ρ.

Proof of the lemma: Denote by ρ_1 the order of f_1. Clearly $\rho_1 \geq \rho$ (since $f_1 = \sum_n \gamma'_n \zeta^n$ and $\gamma'_n \geq \gamma_n$). In order to prove that $\rho_1 \leq \rho$ we shall show that f_1 is majorized, on the positive real axis (which is the direction of fastest growth of all the functions that appear in this lemma) by a function of order ρ. Namely

$$\begin{aligned}f_1(\zeta) &= \sum_n \gamma_n (\zeta)_n \leq \sum_n \gamma_n (\zeta + n)^n \\ &= \sum_{n=0}^{\infty} \sum_{s=0}^{n} \gamma_n \zeta^s n^{n-s} \binom{n}{s} = \sum_{n=0}^{\infty} \sum_{s=0}^{n} 2^n \gamma_n n^n \zeta^s n^{-s}.\end{aligned}$$

By the assumption on f, the numbers $2^n\gamma_n$ satisfy $2^n\gamma_n < n^{-\eta n}$ for every η such that $\eta < \rho^{-1}$ and for sufficiently large n. Chose $\eta > 1$. Then

$$\begin{aligned}f_1(\zeta) &\leq \text{const} + \sum_{s=0}^{\infty} \zeta^s \sum_{n=s}^{\infty} n^{-(\eta-1)n} n^{-s} \\ &= \text{const} + \sum_{s=0}^{\infty} a_s \zeta^s, \qquad (3.36)\end{aligned}$$

where

$$a_s = \sum_{n=s}^{\infty} n^{-(\eta-1)n} n^{-s} = s^{-\eta s}(1 + \cdots) \leq 2s^{-\eta s}$$

for sufficiently large s. This shows that the series on the rhs of (3.36) defines a function of order $\leq \rho$ and completes the proof of the theorem.

4. FAMILIES OF HILBERT SPACES $H(\alpha; A)$

A Hilbert space $H(\alpha; A)$ is a space $S(\beta)$ corresponding to the special choice (4.1) of the sequence $\{\beta_n\}$. Results about $H(\alpha; A)$ are obtained by specialization or by sharpening of results about arbitrary spaces $S(\beta)$ (Secs. 4A–4D). The family of all spaces $H(\alpha; A)$ and $H(\bar{\alpha}; \bar{A})$ is totally ordered with respect to inclusion. It decomposes naturally into three subfamilies, denoted by $\mathfrak{H}^{(\frac{1}{2})}$, $\mathfrak{H}^{(1)}$, and $\mathfrak{H}^{(1^*)}$, respectively. The properties and applications of each of these are shortly discussed in Secs. 4E and 4F.

A. Definition and Dimension of $H(\alpha; A)$

Let α and A be positive numbers. Denote by $H(\alpha; A)$ the Hilbert space $S(\beta)$ with

$$\beta_n = [A^n\Gamma(\alpha n)]^{-2} . \tag{4.1}$$

That is: $H(\alpha; A)$ consists of the functions $u \in S$ such that

$$(u, u)_\alpha = \sum_n (u, M_n u)A^{-2n}\Gamma^{-2}(\alpha n) = \sum_k (u, h_k)\lambda_k^{-2}(h_k, u) < \infty \tag{4.2}$$

The numbers λ_k^{-2} are given by (3.2) which becomes

$$\lambda_k^{-2} = \sum_n A^{-2n}\Gamma^{-2}(\alpha n)a(n; k) \tag{4.3}$$

where the numbers $a(n; k)$ are defined by (2.12).

Theorem 3.8 shows that $H(\alpha'; A') \subseteq H(\alpha''; A'')$ if $\alpha' \leq \alpha''$ or if $\alpha' = \alpha''$, $A' \leq A''$. This makes it convenient to order lexicographically the set of pairs $\{\alpha; A\}$. So $\{\alpha'; A'\} < \{\alpha''; A''\}$ means that either $\alpha' < \alpha''$ or $\alpha' = \alpha''$ and $A' < A''$.

Theorems 3.2 and 3.3 may be used to study the dimension of the spaces $H(\alpha; A)$.

Theorem 4.1: If $\{\alpha; A\} \leq \{\frac{1}{2}; 2^{-\frac{1}{2}}\}$ then $H(\alpha; A)$ contains only the element $u = 0$. If $\{\alpha; A\} > \{\frac{1}{2}; (8e)^{\frac{1}{2}}\}$ then $H(\alpha; A)$ is an infinite-dimensional Hilbert space in which the functions $\lambda_k h_k(x)$ form an orthonormal basis.

Proof: By Theorems 3.2 and 3.3, one needs only to investigate the convergence of the series (4.3). The first inequality (2.16) shows that $a(n; k) \geq n!$ for all n and all k. The second inequality (2.16) gives $a(n; k) \leq (4n)^n$ for $n > \frac{1}{2}d + |k| + 1$. These estimates and the Stirling formula allow a straightforward study of (4.3) which yields the assertion.

Remark: The above theorem does not say anything about the case $\alpha = \frac{1}{2}$, $2^{-\frac{1}{2}} < A \leq (8e)^{\frac{1}{2}}$, which can be studied with the help of sharper estimates.

B. Properties of $u \in H(\alpha; A)$

Theorem[1] *4.2:* There exists a constant $C(\alpha; A)$ such that, for every x and for every $u \in H(\alpha; A)$,

$$|u(x)| \leq C(\alpha; A)\,||u||_\alpha\,(x_1 \cdots x_d)^{[\frac{1}{2}d]+2} \times \exp\left(-A^{-1/\alpha}d^{-1}\sum_{i=1}^{d}|x_i|^{1/\alpha}\right). \tag{4.4}$$

Proof: Consider the inequality (3.10) with $k = 0$ and $m = \{m_1, 0 \cdots 0\}$. It is

$$|x^{m_1}u| \leq c\Gamma(\alpha m_1 + \alpha s)A^{m_1}\,||u||_\alpha \qquad (s = [\tfrac{1}{2}d] + 1). \tag{4.5}$$

The same letter c will denote various constants. It is convenient to replace m_1 by $m_1 + \Delta m_1$ where $|\Delta m_1| \leq \frac{1}{2}$ and where m_1 is no longer required to be an integer. Write $\xi_i = |x_i|/A$ $(i = 1, \cdots, d)$ and $\mu = \alpha(m_1 + s + \frac{1}{2})$. Then (4.5) becomes $|x_1^{m_1+\Delta m_1}u| \leq cA^{m_1}\Gamma(\alpha m_1 + \alpha s + \frac{1}{2}\alpha)\,||u||_\alpha$ or

$$|u| \leq cx_1^{|\Delta m|}\xi^{s+\frac{1}{2}}\xi^{-\mu/\alpha}\Gamma(\mu)\,||u||_\alpha. \tag{4.6}$$

By Stirling's formula,

$$\log(\xi^{-\mu/\alpha}\Gamma(\mu)) \leq \text{const} + \chi(\mu),$$

where $\chi(\mu) = \mu(\log\mu - \alpha^{-1}\log\xi - 1)$. Now μ will be chosen so as to give $\chi(\mu)$ its minimum value. The condition $\chi'(\mu) = 0$ gives $\mu_0 = \xi^{1/\alpha}$ and $\chi(\mu_0) = -\xi^{1/\alpha}$. Substitution into (4.6) gives

$$|u(x)| \leq c\,|x_1|^{s+1}\exp(-A^{-1/\alpha}\,|x_1|^{1/\alpha})\,||u||_\alpha. \tag{4.7}$$

The assertion (4.4) follows; replace x_1 by x_i $(i = 1, \cdots, d)$, take the product of the d inequalities so obtained and extract the dth root.

The results of Sec. 3H can be used to display functions that belong to certain of the spaces $H(\alpha; A)$ and not to others. Notice that the function $\sum_n A^{-2n}\Gamma^{-2}(\alpha n)\zeta^n$ is of order $(2\alpha)^{-1}$. By Theorem 3.15 we have

$$\lambda_k^{-2} = O(\exp(|k|^{1/2\alpha+\epsilon})) \tag{4.8}$$

for every positive ϵ and no negative ϵ. For any μ such that $0 < \mu < 1$ let $u_\mu(x)$ be the function

$$u_\mu(x) = \sum_k \exp(-|k|^\mu)h_k(x). \tag{4.9}$$

Then $(u_\mu, u_\mu)_\alpha = \sum_k \exp(-2\,|k|^\mu)\lambda_k^{-2}$ and so (4.8) gives

Theorem 4.3: If $\alpha > (2\mu)^{-1}$ then $u_\mu \in H(\alpha; A)$. If $\alpha < (2\mu)^{-1}$ then $u_\mu \notin H(\alpha; A)$. Here A is arbitrary.

Another consequence of (4.8) is

Theorem 4.4: If $\alpha' > \alpha$, then the natural embedding of $H(\alpha; A)$ into $H(\alpha'; A')$ is of trace class. Here A and A' are arbitrary.

Actually a stronger result holds: The eigenvalues of (the positive definite part in the polar decomposition of) the natural embedding operator not only have a finite sum but they behave, roughly speaking, as

$$\exp(-|k|^{1/2\alpha} + |k|^{1/2\alpha'}) \sim \exp(-|k|^{\frac{1}{2}\alpha}).$$

C. The Spaces $H(\bar{\alpha}; \bar{A})$

Denote by $H(\bar{\alpha}; \bar{A})$ the space $S(\bar{\beta})$ corresponding to $S(\beta) = H(\alpha; A)$. It is a Hilbert space. An orthonormal basis of $H(\bar{\alpha}; \bar{A})$ consists of the functions

$\lambda_k^{-1}h_k$ where $\lambda_k = \{\sum_n A^{-2n}\Gamma^{-2}(\alpha n)a(n;\, k)\}^{-\frac{1}{2}}$. The Fourier transformation is a unitary operator in $H(\bar{\alpha};\, \bar{A})$. If $\{\alpha';\, A'\} \geq \{\alpha;\, A\}$ (See Sec. 4A), then $H(\bar{\alpha}';\, \bar{A}') \subseteq H(\alpha;\, A)$. It is convenient to write then, by convention, $\{\bar{\alpha}';\, \bar{A}'\} \leq \{\bar{\alpha};\, \bar{A}\}$.

The results of Secs. 4B and 3F show that functions (respectively kernels) which are locally integrable and which do not increase too fast at infinity, belong to $(H\bar{\alpha};\, \bar{A})$ [respectively define a bounded operator from $H(\alpha;\, A)$ into $H(\bar{\alpha};\, \bar{A})$].

For typographical convenience, write

$$E(\alpha;\, A;\, x) = (x_1 \cdots x_d)^{[\frac{1}{2}d]+2} \times \exp\left(-A^{-1/\alpha}d^{-1}\sum_{i=1}^{d}|x_i|^{1/\alpha}\right). \tag{4.10}$$

Then (4.4) becomes

$$|u(x)| \leq C(\alpha;\, A)E(\alpha;\, A;\, x)\,||u||_\alpha \quad \text{for every } u \in H(\alpha;\, A).$$

Theorem 4.5a: Let $f(x)$ be measurable (with respect to the Lebesgue measure) and such that

$$I_f = \int |f(x)|\, E(\alpha;\, A;\, x)\, dx < \infty. \tag{4.11}$$

Then $f \in H(\bar{\alpha};\, \bar{A})$ and $||f||_{\bar{\alpha}} \leq C(\alpha;\, A)I_f$.

Theorem 4.5b: Let $B(x;\, x')$ be measurable and such that

$$I_B = \iint E(\alpha;\, A;\, x)B(x;\, x') \times E(\alpha';\, A';\, x')\, dx\, dx' < \infty. \tag{4.12}$$

Then $u(x) \to \int B(x;\, x')u(x')dx'$ is a bounded operator from $H(\alpha';\, A')$ into $H(\bar{\alpha};\, \bar{A})$. Its bound norm does not exceed $C(\alpha;\, A)C(\alpha';\, A')I_B$.

Theorem 4.5c: Let $G(x)$ be measurable and such that $G(x)E(\alpha;\, A;\, x)$ belongs to $L^{(2)}$. Then $u(x) \to G(x)u(x)$ is a bounded operator from $H(\alpha;\, A)$ into $L^{(2)}$. Its bound norm does not exceed $C(\alpha;\, A)$ times the $L^{(2)}$ norm of $G(x)E(\alpha;\, A;\, x)$.

The proofs are straightforward and will be left to the reader.

D. Entire Functions as Multipliers

The results of Secs. 3E and 3F can easily be specialized to the case of the spaces $H(\alpha;\, A)$. An additional result is given by Theorem 4.8.

Theorem 3.7 gives

Theorem 4.7: Let $B(z) = \sum_k B_k z^k$ be entire analytic of order $\rho < 2$. (Here $z = x + iy$). Let $\{\alpha;\, A\}$ and $\{\alpha';\, A'\}$ be such that

$$\sum_n \left(\frac{A}{A'}\right)^{2n}\Gamma^{-2}(\alpha' n) \times [\sum_k |B_k|A^{|k|}\Gamma(\alpha n + \alpha\,|k|)]^2 \equiv N^2 < \infty. \tag{4.9}$$

Then $u(x) \to B(x)u(x)$ is a bounded operator from $H(\alpha;\, A)$ into $H(\alpha';\, A')$. Its bound norm does not exceed the number N defined by (4.9).

The remark before Theorem 3.7 shows that $B(x)$ can be replaced by a differentiation operator $B(D)$.

In particular:

If $\{\alpha';\, A'\} > \{\alpha;\, A\}$ then the bound norm of the natural embedding of $H(\alpha;\, A)$ into $H(\alpha';\, A')$ does not exceed the number

$$N = \{\sum_n (A/A')^n[\Gamma(\alpha n)/\Gamma(\alpha' n)]^2\}^{\frac{1}{2}}. \tag{4.10}$$

If $B(x)$ is a polynomial then $u \to Bu$ is a bounded operator from $H(\alpha;\, A)$ into any $H(\alpha';\, A')$ such that $\{\alpha';\, A'\} > \{\alpha;\, A\}$.

Let $B(x)$ be such that $u \to Bu$ is a bounded operator from $H(\alpha;\, A)$ into $H(\alpha';\, A')$. For every $f \in H(\bar{\alpha}';\, \bar{A}')$ define $Bf \in H(\bar{\alpha};\, \bar{A})$ by

$$\langle Bf \mid u\rangle = \langle f \mid B^*u\rangle \tag{4.11}$$

for every $u \in H(\alpha;\, A)$. Here $B^*(x)$ is the complex conjugate of $B(x)$. Then $f \to Bf$ is a bounded operator from $H(\bar{\alpha}';\, \bar{A}')$ into $H(\bar{\alpha};\, \bar{A})$. The bound norm of this operator is the same as that of $u \to Bu$. Statements about (4.11) can be easily deduced from the corresponding statements about $u \to Bu$ and will be mostly omitted.

It will now be shown that every analytic function of order $\rho < 2$ defines a bounded operator between some of the spaces $H(\alpha;\, A)$.

Theorem 4.8: If $B(z)$ is an entire function of order $\rho < 2$ (in every one of its arguments) and if $\alpha < \rho^{-1}$, then $u \to Bu$ is a bounded operator from $H(\alpha;\, A)$ into any $H(\alpha';\, A')$ such that $\alpha' > \alpha$.

Proof: For all k except a finite number of them and for some $\epsilon > 0$ we have

$$|B_k| \leq \prod_{i=1}^{d} k_i^{-(\alpha+\epsilon)k_i} \leq |k|^{-(\alpha+\epsilon)\,|k|}.$$

Notice that, by the definition of Euler's beta function

$$\Gamma(\alpha n + \alpha\,|k|) = \Gamma(\alpha n)\Gamma(\alpha\,|k|)/B(\alpha n;\, \alpha\,|k|) \leq \Gamma(\alpha n)\Gamma(\alpha\,|k|)\alpha^2 n\,|k|\, 2^{n+|k|-1}(n + |k|)^{-1}$$

so that

$$\sum_k |B_k| A^{|k|}\Gamma(\alpha n + \alpha |k|) \leq \text{const} + \alpha^2 n\Gamma(\alpha n)2^{n-1} \otimes \sum_k \Gamma(\alpha |k|)A^{|k|}k^{-(\alpha+\epsilon)|k|}2^{|k|}$$

which shows that (4.9) converges. This proves the theorem.

E. The Family $\mathfrak{K}^{(\frac{1}{2}1)}$

The spaces $H(\alpha; A)$ fall naturally into three families corresponding to $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$.

Definition: Denote by $\mathfrak{K}^{(\frac{1}{2}1)}$ the family consisting of the Hilbert spaces $H(\alpha; 1)(\frac{1}{2} < \alpha < 1)$, the space $L^{(2)}$, and the spaces $H(\bar{\alpha}; \bar{1})(\frac{1}{2} < \alpha < 1)$.

In this section, the space $H(\alpha; 1)(\frac{1}{2} < \alpha < 1)$ will be denoted by H_α, and $H(\bar{\alpha}; \bar{1})$ by $H_{\bar{\alpha}}$. It will be shown that elements of H_α have entire analytic continuations, that they decrease fast on parallels to the real "axis" and that $\mathfrak{K}^{(\frac{1}{2}1)}$ admits complex translations.

Theorem 4.9: If $u \in H_\alpha$ then $u(x)$ can be analytically continued to a function $u(z) = u(x + iy)$ which is entire analytic (in every one of its arguments) of order $\rho \leq (1 - \alpha)^{-1}$

Proof: This theorem follows from Theorem 3.6.

Because of analyticity and Fourier invariance, complex translations of $u \in H_\alpha$ are easily studied. Let $a = \{a_1 \cdots a_d\}$ be arbitrary complex. Define $T_a u$ by

$$(T_a u)(x) = u(x - a) = (F^{-1}e^{-iax}Fu)(x),$$

where F is the operator of Fourier transformation and e^{-iax} is the operator of multiplication by e^{-iax}. Since the exponential function is of order 1, Theorem 4.8 shows that $T_a u \in H_{\alpha'}$ for every $\alpha' > \alpha$. In particular, a can be pure imaginary. This gives

Theorem 4.10: If $u \in H_\alpha$, then $u(x + iy)$, considered in its dependence on x belongs to $H_{\alpha'}$ for every $\alpha' > \alpha$ and for every y.

By Theorem 4.2 this means, in particular, that the analytic continuation of $u \in H$ decreases as $\exp(-\sum_{i=1}^d |x_i|^{1/\alpha})$ on parallels to the real "axis."

These properties of $u \in H_\alpha$ are useful in the study of integrals of the form

$$\int f(\zeta; x)u(x)\,dx \qquad (u \in H_\alpha), \tag{4.11}$$

where f depends on a complex parameter ζ. They allow deformations of the path of integration in (4.11) and so facilitate the study of holomorphic families of elements of $H_{\bar{\alpha}}$. Simple examples may be found in Ref. 7. A fuller discussion will be published elsewhere.

F. The Family $\mathfrak{K}^{(1)}$

If $\alpha = 1$, then the elements of $H(1; A)$ are analytic in Cartesian product of strips.

Definition: Denote by $\mathfrak{K}^{(1)}$ the family consisting of the Hilbert spaces $H(1; A)(0 < A < \infty)$, the space $L^{(2)}$ and the spaces $H(\bar{1}; \bar{A})$.

Theorem 4.10: If $u \in H(1; A)$, then $u(x)$ can be analytically continued to a function $u(z) = u(x + iy)$ which is holomorphic in the domain

$$|y_i| < A^{-1} \qquad (i = 1, \cdots, d). \tag{4.12}$$

Proof: It follows from (3.10) that

$$|D^k u(x)| < \text{const}\ ||u||_A\ \Gamma(|k|)A^{|k|} \tag{4.13}$$

for every x, every k and every $u \in H(1; A)$. The assertion is then obtained by examination of the Taylor series of u.

Entire functions of order 1 and finite type are multipliers between spaces of $\mathfrak{K}^{(1)}$. Many of the expressions previously introduced can be evaluated in closed form for $\alpha = 1$. The resulting expressions are not given here.

Another topic which will be treated in a later paper is the family $\mathfrak{K}^{(1\infty)}$ corresponding to $1 < \alpha < \infty$. The main feature there is that a support can be defined for every $f \in H_{\bar{\alpha}}$.

5. NESTED HILBERT SPACES

Many of the above constructions can be set into a general context with the help of the notion of nested Hilbert space (see Introduction). This is, technically speaking, a special kind of inductive limit of a family of Hilbert spaces (see Sec. 5c). It keeps many Hilbert space properties. For example, the adjoint of an operator A acts on the same space as A (Sec. 5F).

The results of this section are not used in the preceding sections. For this reason the proofs are omitted. They are simple and will be given elsewhere.

A. Notation

$H_i, \cdots, H_j \cdots$ are Hilbert spaces.

$f_i, \cdots, g_i \cdots$ are elements of H_i.

A_{ji} is a linear operator from H_i into H_j.

$(A_{ij})^*_{ij}$ is the adjoint of A_{ji}, defined by

$$(g_i, (A_{ji})^*_{ij}f_j) = (A_{ji}g_i, f_j). \tag{5.1}$$

On the lhs of (5.1) the scalar product is in H_i;

on the rhs it is in H_j. There is no need to indicate this by $(\ ,\ )_i$ or $(\ ,\ )_j$ since the subscripts are carried by the elements and the operators.

B. Nesting Operators

An operator E_{ji} from H_i into H_j will be called a nesting operator if

(a) E_{ji} is continuous and defined for all elements of H_i.

(b) E_{ji} is one-to-one.

(c) The range of E_{ji} is dense in H_j.

If the range of E_{ji} is not the whole of H_j then E_{ji} will be called a proper nesting operator.

Theorem 5.1: A continuous linear operator from H_i into H_j is a nesting operator if and only if, in the polar decomposition

$$E_{ji} = L_{jj}^{(j)} U_{ji} \qquad (L_{jj}^{(j)} = [E_{ji}(E_{ji})_{ji}^*]^{\frac{1}{2}}$$

the operator $L_{jj}^{(j)}$ is injective (i.e., 1-to-1 into) and the operator U_{ji} is unitary (rather than only isometric).

Theorem 5.2: The product of two nesting operators is a nesting operator. The adjoint of a nesting operator is a nesting operator.

C. Nested Hilbert Space

A nested Hilbert space consists, roughly speaking, of a totally ordered family of Hilbert spaces together with nesting operators that identify elements of these spaces. In Secs. 3 and 4, these nesting operators were the natural embeddings. One of the spaces, denoted by H_0, has a distinguished role; in Secs. 3 and 4 this was $L^{(2)}$. The other spaces come in pairs that are dual to each other in the sense of Sec. 3F.

Let I be a totally ordered set, with an order-reversing involution $i \leftrightarrow \bar{\imath}$.

Assume that there exists an element $o \in I$ such that $\bar{o} = o$.

For every $i \in I$, let H_i be a Hilbert space. For all pairs $i, j \in I$ such that $i > j$ (in the sense of the order in I), let E_{ij} be a nesting operator from H_j into H_i. Assume

$$E_{ij}E_{jk} = E_{ik} \qquad (i > j > k). \tag{5.2}$$

For every $i \in I$, define E_{ii} as the identity in H_i.

Assume that, for every $i \leq o$, the operators E_{oi} and E_{io} are related by

$$E_{oi}(E_{oi})_{io}^* = (E_{io})_{oi}^* E_{io}. \tag{5.3}$$

In the union $\bigcup_{i \in I} H_i$ define an equivalence relation $\equiv$ by:

$f_i \equiv f_j$ means: There exists a k such that $k \geq i$, $k \geq j$ and that $E_{ki}f_i = E_{kj}f_j$.

Denote by H_I the set of classes so obtained and by $f, g, \cdots$ elements of H_I. If $f_i \in H_i$ belongs to the class f, then f_i is called a representative of f.

Definition: H_I is called a nested Hilbert space.

For every $i \in I$ there exists the canonical embedding E_{Ii} of H_i into H_I. If $i \geq j$ then $E_{Ii}H_i \supseteq E_{Ij}H_j$.

If $f \in H_I$ denote by $J(f) \subseteq I$ the set of all $i \in I$ such that f has a representative f_i in H_i. That is:

$$J(f) = \{j : f \in E_{Ij}H_j\}.$$

D. The Numbers $\langle f \mid g \rangle$

For every $i \leq o$, define U_{oi} as the unitary operator appearing in the polar decomposition of E_{oi}:

$$E_{oi} = L_{oo}^{(i)} U_{oi}$$
$$(L_{oo}^{(i)} = [E_{oi}(E_{oi})_{io}^*]^{\frac{1}{2}}). \tag{5.4}$$

For every $i \geq o$, define U_{oi} as the unitary operator appearing in the polar decomposition of $(E_{io})_{oi}^*$:

$$(E_{io})_{oi}^* = [(E_{io})_{oi}^* E_{io}]^{\frac{1}{2}} U_{oi} = L_{oo}^{(i)} U_{oi}. \tag{5.5}$$

The last equality follows from the assumption (5.3).

Given any $i \in I$, define U_{io} as $U_{io} = (U_{oi})_{io}^*$ and write $U_{ii} = U_{io}U_{oi}$.

Definition: Let f and g be two elements of H_I such that $J(f) \cap \bar{J}(g)$ is nonempty. [Here $\bar{J}(g)$ is the set of all $\bar{\jmath}$ with $j \in J(g)$.] Chose an arbitrary $i \in J(f) \cap \bar{J}(g)$ and define

$$\langle f \mid g \rangle = (U_{oi}f_i, U_{oi}g_i).$$

Theorem 5.3: The number $\langle f \mid g \rangle$ does not depend on the choice of $i \in J(f) \cap \bar{J}(g)$. It is linear in g, antilinear in f. It satisfies $\langle f \mid g \rangle = \langle g \mid f \rangle^*$.

E. Operators in H_I

An operator in H_I is defined so that its domain is not arbitrary but is a union of the Hilbert spaces H_i that determine H_I. Furthermore, the image of every H_i must be contained in some H_j. These conditions are always satisfied in the applications.

Preliminary remark: Since the subspaces $E_{Ii}H_i$ increase with i, an arbitrary union $\bigcup E_{Ii}H_i (i \in I_1 \subseteq I)$ is equal to a union $\bigcup E_{Ii}H_i$ $(i \in D, D$ an initial subset of I, i.e, a subset such that $i \in D$, $j < i$ entails $j \in D)$.

Definition: An operator A in H_I is a linear transformation such that

(a) Its domain $\mathfrak{D}(A) \subseteq H_I$ is of the form

$$\mathfrak{D}(A) = \bigcup_i E_{Ii}H_i\,, \tag{5.6}$$

where $i \in D(A)$ and $D(A)$ is a nonempty initial subset of I.

(b) For every $i \in D(A)$ there exists a $j = j(i)$ such that

$$AE_{Ii}H_i \subseteq E_{Ij}H_j. \tag{5.7}$$

The set $J(i; A)$ of $j \in I$ such that (5.7) holds is a final subset of I. If $i \in D(A)$ and if $j \in J(i; A)$ then A defines the linear transformation A_{ji} from H_i into H_j by $A_{ji}f_i = (Af)_j$. The operator A_{ji} is called a representative of A. Notice that A_{ji} is defined on the whole Hilbert space H_i. In order to avoid pathology, assume

(c) A_{ji} is closed for every $i \in D(A)$ and for every $j \in J(i; A)$.

It follows then that A_{ji} is bounded.

The study of A is equivalent to the study of the collection $A_{ji}(i \in D(A),\ j \in J(i; A))$ of bounded operators between Hilbert spaces.

Denote by $R(A) \subseteq I$ the final subset

$$R(A) = \bigcup_{i \in D(A)} J(i; A).$$

F. The Adjoint

It is easy to verify that the operators in H_I form a vector space (i.e., that the sum of two operators and the scalar multiples of an operator are, again, operators in the sense of Sec. 5E). It will be shown elsewhere that every operator in H_I has an adjoint which is also an operator in H_I. This is to be contrasted with the case of continuous operators in more general spaces (where the adjoint acts on the dual space) and with the case of arbitrary operators in Hilbert space (where the adjoint need not be defined).

G. The Algebra $\mathfrak{a}$

The set of all operators in H_I is not an algebra since the product of two operators need not be an operator in the sense of Sec. 5E.

Definition: An operator A in H_I is said to belong to $\mathfrak{a}$ if $D(A) = R(A) = I$.

Theorem: $\mathfrak{a}$ is an algebra with involution. If $A \in \mathfrak{a}$ and if C is any operator in H_I, then AC and CA are operators in H_I; they need not belong to $\mathfrak{a}$.

Remark: Another algebra of operators consists of all the operators in H_I such that both $D(A)$ and $R(A)$ contain the element o. This algebra is isomorhpic to the algebra of all bounded operators in H_o. It neither contains nor is contained in the algebra $\mathfrak{a}$.

The reader may find it worth while to examine the operators of Sec. 4 and to determine the sets $D(A)$ and $R(A)$.

ACKNOWLEDGMENTS

This work was begun while the author was supported by funds from the National Science Foundation at the Institute for Advanced Study in Princeton. He has received much help and encouragement both there and at the Courant Institute. A conversation with Professor V. Bargmann was particularly valuable.

Continuous-Representation Theory. V. Construction of a Class of Scalar Boson Field Continuous Representations

JOHN R. KLAUDER AND JAMES MCKENNA
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 2 July 1964)

A large class of continuous representations of separable Hilbert spaces is constructed with the aid of representations of the canonical commutation relations (CCR) for a scalar boson field $\varphi(f)$ and its canonical conjugate $\pi(g)$. A representation of the CCR for a scalar boson field consists of two operator-valued functions $V[g]$ and $W[f]$, defined for all f, g in Schwartz's space $\mathcal{S}$ of real test functions, where $V[g]$ and $W[f]$ are unitary operators defined on some separable Hilbert space $\mathfrak{H}$, and which satisfy the commutation relations $V[g]W[f] = e^{-i(f,g)}W[f]V[g]$. These unitary operators are related to the field and its momenta by $V[g] = e^{-i\pi(g)}$, $W[f] = e^{i\varphi(f)}$. We explicitly construct a family of such representations with the help of von Neumann's theory of infinite direct products of Hilbert spaces, the pertinent parts of which are reviewed. A continuous representation $\mathfrak{C}$ of the Hilbert space $\mathfrak{H}$ is composed of a linear vector space of complex, bounded, continuous functionals defined on $\mathcal{S} \times \mathcal{S}$. These functionals are defined for all $\Psi \in \mathfrak{H}$ by $\psi(f, g) = (V[g]W[f]\Phi_0, \Psi)$. In this definition, Φ_0 is a fixed unit vector in $\mathfrak{H}$. The properties of the functions in $\mathfrak{C}$ depend on the choice of the representation of the CCR and on the choice of Φ_0. When $\mathfrak{C}$ is constructed with the aid of an irreducible representation of the CCR, an inner product can be defined for all pairs of functionals in $\mathfrak{C}$ by an intuitively meaningful, rigorously defined integral in the sense of Friedrichs and Shapiro. With this inner product, $\mathfrak{C}$ is a complete Hilbert space congruent with $\mathfrak{H}$. As in all continuous representations, a reproducing kernel exists and determines the functions in the continuous representation. One such space is closely related to a space of analytic functionals introduced by Segal and Bargmann. The representation of various operators as kernels and as functional derivatives is discussed. Finally, the construction of a vast number of unitary invariants for a representation of the CCR is used to establish the unitary inequivalence of uncountably many of the representations that we construct.

1. INTRODUCTION

IN the present paper we initiate a study of continuous representations of Hilbert space[1-3] for scalar boson fields,[4] extending our previous analysis of phase-space continuous representations for finitely many degrees of freedom.[5] It is well known that a field theory is in part more complicated than a many-particle problem for two reasons, namely, that a field has an infinite number of degrees of freedom, and that in many actual applications the representation of the operator algebra must be chosen reducible.[6-10] The first complication has as a consequence the existence of infinitely many, inequivalent, irreducible representations of the canonical commutation relations (CCR),[6,7,9-14] which is in striking contrast to the existence of only a single irreducible representation for finitely many degrees of freedom.[15] The second complication of fields makes it mandatory that we analyze and be prepared to employ representations other than just the standard Fock representations. To this end this paper is largely devoted to a rigorous construction of myriads of inequivalent, irreducible representations of the CCR together with their intimately associated Hilbert spaces of bounded, continuous functions—the continuous representations—on which the particular representation of the CCR is the natural, regular representation. Vast numbers of reducible representations of the CCR may, of course, be obtained by direct sums or direct integrals of various irreducible representations, but we do not primarily concern ourselves with these questions here.

Operator properties for a scalar field are summarized in the CCR, a proper statement of which is

[1] J. R. Klauder, J. Math. Phys. 4, 1055 (1963), referred to as I.
[2] J. R. Klauder, J. Math. Phys. 4, 1058 (1963).
[3] J. R. Klauder, J. Math. Phys. 5, 177 (1964).
[4] A preliminary account of this work appeared in J. McKenna and J. R. Klauder, Bull. Am. Phys. Soc. 9, 85 (1964).
[5] J. McKenna and J. R. Klauder, J. Math. Phys. 5, 878 (1964) referred to as IV. References to this paper carry the prefix IV.
[6] K. O. Friedrichs, *Mathematical Aspects of the Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1953).
[7] R. Haag, Kgl. Danske Videnskab. Selskab, Mat. Fys. Medd. 29, 12 (1955).
[8] Analogous results hold for fermions as well: R. Haag, Nuovo Cimento 25, 287 (1962); H. Ezawa, J. Math. Phys. 5, 1078 (1964).
[9] H. Araki, J. Math. Phys. 1, 492 (1960).
[10] H. Araki and E. J. Woods, J. Math. Phys. 4, 637 (1963).
[11] L. Gärding and A. S. Wightman, Proc. Nat. Acad. Sci. 40, 617 (1954); A. S. Wightman and S. S. Schweber Phys. Rev. 98, 812 (1955).
[12] I. E. Segal, Trans. Am. Math. Soc. 88, 12 (1958).
[13] J. M. Cook, J. Math. Phys. 2, 33 (1961).
[14] J. S. Lew, thesis, Princeton University, 1960 (unpublished).
[15] J. von Neumann, Math. Ann. 104, 570 (1931).

given in "Weyl form" in terms of unitary operators, and which is defined as follows:[12,14]

Definition 1.1. Let $\mathfrak{U}$ be a countably infinite-dimensional real linear vector space, called the "test function space," and let $f \times g \to (f, g), f, g \in \mathfrak{U}$ be a nondegenerate, bilinear map of $\mathfrak{U} \times \mathfrak{U}$ into R, the field of real numbers. Let $\mathfrak{H}$ be a complex Hilbert space, and $\mathfrak{G}$ the group of all unitary transformations of $\mathfrak{H}$ onto itself, where the unit operator is denoted by I. Then two maps, $\mathfrak{U} \xrightarrow{W} \mathfrak{G}$, and $\mathfrak{U} \xrightarrow{V} \mathfrak{G}$ are said to be a representation of the canonical commutation relations if they satisfy:

$$(1)\quad W[f]W[f'] = W[f + f'], \quad W[0] = I; \tag{1.1}$$

$$(2)\quad V[g]V[g'] = V[g + g'], \quad V[0] = I; \tag{1.2}$$

$$(3)\quad V[g]W[f] = e^{-i(f,g)}W[f]V[g]; \tag{1.3}$$

(4) For each fixed $f, g \in \mathfrak{U}$, the operators $W[tf]$ and $V[tg]$ are weakly continuous functions of t, $-\infty < t < \infty$.

We later indicate that Condition (4) is sufficient to guarantee the existence of self-adjoint, smeared field operators $\varphi(f)$ and their canonical momenta $\pi(g)$ such that $V[g] = e^{-i\pi(g)}$, $W[f] = e^{i\varphi(f)}$. In a sense to be made precise later, $\varphi(f)$ and $\pi(g)$ satisfy the familiar commutation relations $[\varphi(f), \pi(g)] = i(f, g)I$. It should be noted that $U[f, g] \equiv V[g]W[f]$ gives a representation up to a factor of the additive group of $\mathfrak{U} \times \mathfrak{U}$.

While the preceding definition is not the most general, it is more than adequate for our purposes. Indeed, we limit ourselves at the outset only to representations defined on separable Hilbert spaces. Furthermore, since it is conventionally assumed that a scalar boson field describes a dynamical system with countably many degrees of freedom, we shall employ a countably-infinite direct product space as a natural setting for rigorously defining representations of such a field. Superficially these two goals appear to be incompatible for if $\{\mathfrak{H}_n : n \in \Delta\}$, where Δ is some index set, is a set of separable, infinite-dimensional, complex Hilbert spaces, the complete direct product space (CDPS) $\prod_{n\in\Delta} \otimes \mathfrak{H}_n$ as defined by von Neumann,[16] is *not* a separable space when Δ has a countable infinity of elements. However, as von Neumann has shown, the CDPS $\prod_{n\in\Delta} \otimes \mathfrak{H}_n$ can be decomposed into an uncountable number of separable, mutually orthogonal, closed subspaces. These spaces are called incomplete direct product spaces (IDPS), and the representations of the CCR we explicitly and rigorously construct in Sec. 2 are defined on incomplete direct product spaces. Our construction of the CCR on an IDPS is aided considerably by the extensive investigation of such Hilbert spaces by von Neumann. However, we are under no illusion that such constructions provide an adequate means to discuss all representations of the CCR.

As a further specialization of the conditions of Definition 1.1, we take our test function space $\mathfrak{U}$ to be the space of all real-valued functions of three real variables $f(x_1, x_2, x_3) = f(\mathbf{x})$, which are infinitely differentiable and which decrease at infinity faster than any inverse power of $|\mathbf{x}|$. The mathematical properties of this space, usually denoted by $\mathcal{S}$, have been studied by Schwartz,[17] and the physical motives involved in the choice of $\mathcal{S}$ as a test function space for field theory have been discussed recently by Wightman.[18]

The simultaneous study of large classes of representations of the CCR's and the Hilbert spaces in which they operate is particularly convenient in the continuous-representation formalism. It is our purpose in Sec. 3 to rigorously define such spaces, and to establish the intuitive and natural representation of the CCR that scalar-field continuous representations provide, for the representations defined in Sec. 2. A continuous representation $\mathfrak{C}$ of the IDPS $\mathfrak{H}$ is composed of a linear vector space of complex-valued, bounded functionals defined on $\mathcal{S} \times \mathcal{S}$, which are defined for all $\Psi \in \mathfrak{H}$ by

$$\psi(f, g) = (\Phi[f, g], \Psi), \tag{1.4}$$

wherein the vectors of the overcomplete family of states (OFS) $\mathfrak{S}$ which "generates" the representation[1] are defined by

$$\Phi[f, g] \equiv V[g]W[f]\Phi_0 \equiv U[f, g]\Phi_0. \tag{1.5}$$

In (1.5), the "fiducial vector" Φ_0 may be chosen as one of a generally large class of unit vectors in $\mathfrak{H}$, and the particular choice will have a certain influence on the properties of $\mathfrak{C}$. For each such choice of Φ_0, every $\psi(f, g) \in \mathfrak{C}$ will be a continuous function on $\mathcal{S} \times \mathcal{S}$, if $\mathcal{S}$ is supplied with the usual locally convex topology and $\mathcal{S} \times \mathcal{S}$ is supplied with the product topology.

The particular representation of the CCR enters, essentially, in two ways. Firstly, it is necessary that the family of operators $U[f, g]$ be cyclic, i.e., that a vector Φ_0 exists such that the closed space spanned

[16] J. von Neumann, Composito Math. **6**, 1 (1938).

[17] L. Schwartz, *Theorie des distributions* (Hermann & Cie., Paris, 1957), Vol. II.

[18] A. S. Wightman, *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963), pp. 11–58.

by $U[f, g]\Phi_0$, for all $f, g \in \mathcal{S}$, is $\mathfrak{H}$ itself. This is a relatively weak restriction, and is in fact realized in all physical applications analyzed to date.[8-10] Secondly, the representation of $V[g]$ and $W[f]$ influences the form of the inner product in $\mathfrak{C}$. Since $\mathcal{S} \times \mathcal{S}$, considered as an additive group is *not* locally compact, no Haar measure is known to exist. Thus one cannot talk of the square integrability of the functionals $\psi(f, g)$ over the group in the conventional sense. However, when an irreducible representation is used, it is true that an integral and measure in the sense of Friedrichs and Shapiro[19] exist, i.e., an integral defined as the limit of a sequence of integrals. It is in this extended sense that we employ the terminology integral and measure. However, the relevant limits by which they are defined will always be carefully stated.

In particular, we shall establish that $\mathfrak{C}$ is a separable, complete Hilbert space for which the inner product $(\ ,\ )_0$ is explicitly given by

$$(\psi, \lambda)_0 = \lim_{N\to\infty} \int \psi^*(f_{(N)}, g_{(N)})\lambda(f_{(N)}, g_{(N)}) \times d\mu(f_{(N)}, g_{(N)}) = (\Psi, \Lambda), \tag{1.6}$$

if $\mathfrak{C}$ is defined with the aid of an irreducible representation and if $\psi(f, g) = (\Phi[f, g], \Psi)$ and $\lambda(f, g) = (\Phi[f, g], \Lambda)$ are arbitrary elements in $\mathfrak{C}$. In this expression

$$d\mu(f_{(N)}, g_{(N)}) = \prod_{|n|\leq N} (dp_n\, dq_n/2\pi), \tag{1.7}$$

$$f_{(N)}(\mathbf{x}) = \sum_{|n|\leq N} p_n h_n(\mathbf{x}); \quad p_n = (h_n, f), \tag{1.8}$$

$$g_{(N)}(\mathbf{x}) = \sum_{|n|\leq N} q_n h_n(\mathbf{x}); \quad q_n = (h_n, g), \tag{1.9}$$

$$n = (n_1, n_2, n_3), \quad n_i \geq 0, \quad i = 1, 2, 3; \quad |n| = n_1 + n_2 + n_3, \tag{1.10}$$

and

$$h_n(\mathbf{x}) = h_{n_1}(x_1)h_{n_2}(x_2)h_{n_3}(x_3) \tag{1.11}$$

is a basis for $\mathcal{S}$ composed of Hermite functions. The plausibility and intuitive meaning of (1.6) is straightforward. Since we are dealing only with continuous functions $\psi(f, g)$, it is clear that each function is completely determined by its values on a dense set. The integral (1.6) is defined simply as the limit of a sequence of integrals whose integrands are evaluated on a dense set as N becomes arbitrarily large, since test functions of the form (1.8) and (1.9), for all finite N, are dense in $\mathcal{S}$. Consequently, $\mathfrak{C}$ is a complete, separable Hilbert space composed of bounded, continuous functionals of test functions $\psi(f, g)$ (and *not* of equivalence classes of functionals) with an inner product given by (1.6). We call $\mathfrak{C}$ a continuous representation since it is also congruent,[20] i.e., isomorphic and isometric, with the original Hilbert space.

A single space $\mathfrak{C}$ by no means contains all bounded, continuous functionals that are square integrable in the sense of (1.6); in fact, uncountably many mutually orthogonal continuous representations $\mathfrak{C}$ exist. As an example of one such space, we briefly discuss in Sec. 3 the special continuous representation that is closely connected with the space of analytic functionals studied by Segal[21] and Bargmann.[22]

In every space $\mathfrak{C}$ there exists a natural, canonical representation of the operators $V[g]$ and $W[f]$ satisfying Definition 1.1. In particular, the transformations defined by

$$(V[g]\psi)(f', g') = \psi(f', g' - g), \tag{1.12a}$$

$$(W[f]\psi)(f', g') = e^{i(f, g')}\psi(f' - f, g') \tag{1.12b}$$

are unitary in the inner product (1.6) and fulfill the Weyl relations (1.1)–(1.3). In fact $U[f, g]$ constructed from (1.12) is just the *regular* representation up to a factor of the additive group of $\mathcal{S} \times \mathcal{S}$. From this point of view, different representations are not distinguished by *how* they act on a common set of functions, but rather on *which* set of functions they act, which in turn depends on the choice of the fiducial vector Φ_0. Of course, a congruent map of one space $\mathfrak{C}'$ onto a different space $\mathfrak{C}$ always exists; but under this map the image of the regular representation would no longer be given by (1.12). Thus the two views of "how" or "which" are completely complementary, although there is some advantage to be gained by a consistent use of the simple, canonical form (1.12), particularly with a view toward constructing reducible representations of the CCR by direct sums or direct integrals.

While this paper is devoted to continuous representations for a single scalar field, brief mention is made of analogous representations pertaining to K independent scalar fields. Not only do such multifield representations have their own intrinsic interest, but they provide a rapid and simple method to generate reducible representations of the single field CCR by a process akin to that of Lie group contraction.[23]

[19] K. O. Friedrichs and H. N. Shapiro, *Integration of Functionals,*" Lecture Notes, New York University Institute of Mathematical Sciences, 1957, Chap. 1.

[20] A. E. Taylor, *Introduction to Functional Analysis* (John Wiley & Sons, Inc., New York, 1958).

[21] I. E. Segal, Illinois J. Math. **6**, 500 (1962); *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, Rhode Island, 1963), Chap. VI.

[22] V. Bargmann, Proc. Nat. Acad. Sci. **48**, 199 (1962).

[23] E. Inönü and E. P. Wigner, Proc. Nat. Acad. Sci. **39**, 510 (1953); E. J. Saletan, J. Math. Phys. **2**, 1 (1961).

The smeared field operators $\varphi(f)$ and $\pi(g)$, whose existence is ensured by Condition 4 of Definition 1.1, are a common starting point for physical applications. Their representations, which Eq. (1.12) correctly implies may be expressed with the help of smeared functional derivatives, are taken up in Sec. 4.

Finally we discuss an important set of unitary invariant "tags" for a representation of the CCR that frequently enables the inequivalence of two representations of the CCR's to be ascertained. Furthermore, these tags may be completely derived from the functions in the continuous representation. In particular, we prove that the class of spaces and their canonical transformations of the form (1.12) which we rigorously define corresponds to uncountably-many inequivalent representations of the CCR. It is ultimately to be hoped that the simple functional methods we advocate can be used to study the inequivalence of representations of the CCR by a generalization of the classical group orthogonality relations.[24] Support for this hope is suggested by the proof by Bargmann[25] of the von Neumann uniqueness theorem for representations of the CCR for finitely many degrees of freedom based simply on the square integrability of the functions in $\mathfrak{C}$ for any fiducial vector, which we proved in Part IV.

Applications of a special continuous representation have been made by Glauber[26] and by Sudarshan[27] in the analysis of the coherence properties of light. Segal[21] has used his closely associated space of analytic functionals to study properties of free, relativistic scalar boson fields. A broader application of continuous representations to model field theories with nontrivial interactions is currently under study[28] extending the first named author's earlier dynamical studies for finitely many degrees of freedom.[2,3]

2. DEFINITION OF IRREDUCIBLE SCALAR FIELD REPRESENTATIONS

A. Discussion of Incomplete Direct Product Spaces

Since we have to make use of some of the detailed properties of the IDPS, we outline its construction and list a few of its properties here. For further information we refer the reader to the cited paper of von Neumann.[16]

Let Δ be an arbitrary index set, and to each $n \in \Delta$ assign the separable Hilbert space $\mathfrak{H}_n$.

Definition 2.1. A sequence of vectors $\{\chi_n\}$, $\chi_n \in \mathfrak{H}_n$, $n \in \Delta$, is called a c_0 sequence if and only if

$$\sum_{n\in\Delta} | \, ||\chi_n|| - 1| < \infty .$$

To each c_0 sequence is assigned an element $\prod_{n\in\Delta} \otimes \chi_n$, called a c_0 vector, with the norm $||\prod_{n\in\Delta} \otimes \chi_n|| = \prod_{n\in\Delta} ||\chi_n||$.

Definition 2.2. Two c_0 vectors are said to be equivalent $\prod_{n\in\Delta} \otimes \chi_n \approx \prod_{n\in\Delta} \otimes \lambda_n$, if and only if

$$\sum_{n\in\Delta} |(\chi_n, \lambda_n) - 1| < \infty . \tag{2.1}$$

The inner product of two equivalent c_0 vectors is

$$(\prod_{n\in\Delta} \otimes \chi_n, \prod_{n\in\Delta} \otimes \lambda_n) = \prod_{n\in\Delta} (\chi_n, \lambda_n), \tag{2.2}$$

while the inner product of two inequivalent c_0 vectors is zero. (We use the convention that inner products are linear in the second variable and conjugate linear in the first variable.)

The set of all finite linear combinations of c_0 vectors with complex coefficients forms a linear vector space. The inner product defined for c_0 vectors is extended by linearity so that this space becomes an inner-product space. The closure of this space yields the complete direct-product space, $\prod_{n\in\Delta} \otimes \mathfrak{H}_n$.

The relation $\approx$ is an equivalence relation which decomposes the set of all c_0 vectors into mutually disjoint equivalence classes. The equivalence class of a given c_0 vector, $\prod_{n\in\Delta} \otimes \chi_n$, is denoted by $E(\prod_{n\in\Delta} \otimes \chi_n)$.

Definition 2.3. The closed linear subspace of $\prod_{n\in\Delta} \otimes \mathfrak{H}_n$ which is spanned by all the c_0 sequences in a given equivalence class E is called an incomplete direct-product space and is denoted by

$$\prod_{n\in\Delta}^{E} \otimes \mathfrak{H}_n .$$

If the power of the index set Δ is countably infinite, $\prod_{n\in\Delta} \otimes \mathfrak{H}_n$ *is not* a separable Hilbert space, but each IDPS *is* a separable, infinite-dimensional Hilbert space.

The following lemma provides a very useful, alternative construction of a given IDPS.

Lemma 2.1. From a given equivalence class E pick a c_0 vector, $\mathrm{X} = \prod_{n\in\Delta} \otimes \chi_n$, $||\chi_n|| = 1$ for all $n \in \Delta$. Then

$$\prod_{n\in\Delta}^{E} \otimes \mathfrak{H}_n$$

[24] L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, 1946), Chap. IV.
[25] V. Bargmann (to be published).
[26] R. J. Glauber, Phys. Rev. Letters **10**, 84 (1963); Phys. Rev. **131**, 2766 (1963).
[27] E. C. G. Sudarshan, Phys. Rev. Letters **10**, 277 (1963).
[28] A preliminary report on part of this work appeared in J. R. Klauder, Bull. Am. Phys. Soc. **9**, 85 (1964).

is the closed linear subspace of $\prod_{n\in\Delta}\otimes \mathfrak{H}_n$ which is the closure of the set $\mathfrak{G}_X$ of all finite linear combinations of vectors, with complex coefficients, of the form $\Lambda = \prod_{n\in\Delta}\otimes \lambda_n$, $\lambda_n \in \mathfrak{H}_n$, $n \in \Delta$, where $\lambda_n = \chi_n$ for all but a finite number of indices n.

Definition 2.4. We call the c_0 vector $X = \prod_{n\in\Delta}\otimes \chi_n$ of Lemma 2.1 the *product reference vector*, and the corresponding $\chi_n \in \mathfrak{H}_n$, $n \in \Delta$, are called the *product components* of X. The dense set of vectors $\mathfrak{G}_X$ of Lemma 2.1 is called the *reference set.*

Lemma 2.2. Let z_n, $n \in \Delta$ be a sequence of complex numbers, and let $\prod_{n\in\Delta}\otimes \chi_n$ be a c_0 sequence. Then $\prod_{n\in\Delta}\otimes z_n\chi_n$ is again a c_0 sequence satisfying

$$\prod_{n\in\Delta}\otimes \chi_n \approx \prod_{n\in\Delta}\otimes z_n\chi_n$$

if and only if $\sum_{n\in\Delta}|z_n - 1| < \infty$. If this condition holds, then $\prod_{n\in\Delta}\otimes z_n\chi_n = (\prod_{n\in\Delta} z_n)(\prod_{n\in\Delta}\otimes \chi_n)$.

B. The Space $\mathcal{S}$ of Test Functions

In our construction of the CCR, an alternate characterization of the test function space $\mathcal{S}$ is needed, and we include a discussion of it here. First we introduce the following standard notation: We set $n = (n_1, n_2, n_3)$, where the n_i, $i = 1, 2, 3$, are nonnegative integers, and $|n| = n_1 + n_2 + n_3$. If $\mathbf{x} = (x_1, x_2, x_3)$ is any real three-vector, then

$$x^n = x_1^{n_1}x_2^{n_2}x_3^{n_3}, \quad d^3x = dx_1\,dx_2\,dx_3,$$

and

$$D^n = \partial^{|n|}/\partial x_1^{n_1}\,\partial x_2^{n_2}\,\partial x_3^{n_3}.$$

Definition 2.5. The space $\mathcal{S}$ is the real, linear topological space whose elements consist of all those real valued, C^∞ functions $f(\mathbf{x})$, which in addition satisfy the condition that $x^nD^mf(\mathbf{x})$ is bounded for all n and m. A locally convex topology is defined in $\mathcal{S}$ by the family of seminorms

$$\rho_{ij}(f) = \sup|x^nD^mf(\mathbf{x})|, \quad |n| \le j, \quad |m| \le k, \tag{2.3}$$

where sup is taken over $\mathbf{x}$, $|n| \le j$, and $|m| \le k$, for $j, k = 0, 1, 2, \cdots$. The pairing of f and $g \in \mathcal{S}$ is given by

$$(f, g) = \int_{R_3} f(\mathbf{x})g(\mathbf{x})\,d^3x. \tag{2.4}$$

The space $\mathcal{S}$ is a *non*locally compact, complete, Hausdorff space with a countable basis.[17] Since $\mathcal{S}$ has a countable basis, it is only necessary to consider sequences (and not directed nets) when proving the convergence of functionals on $\mathcal{S}$.[29]

Let Δ be the set of all possible triples, $n = (n_1, n_2, n_3)$, of nonnegative integers, and let $h_r(x)$, $r = 0, 1, 2, \cdots$ be the set of Hermite functions. Then the set of functions $h_n(\mathbf{x}) = h_{n_1}(x_1)h_{n_2}(x_2)h_{n_3}(x_3)$, $n \in \Delta$, is a basis for $\mathcal{S}$, and every $f(\mathbf{x}) \in \mathcal{S}$ possesses the expansion

$$f(\mathbf{x}) = \sum_{n\in\Delta} p_nh_n(\mathbf{x}), \tag{2.5}$$

where

$$p_n = (h_n, f), \tag{2.6}$$

and the series (2.5) converges uniformly and absolutely.[30]

Lemma 2.3. A sequence of real numbers $\{p_n\}$, $n \in \Delta$, is related to a function $f(\mathbf{x}) \in \mathcal{S}$ by the relations (2.5) and (2.6) if and only if

$$\lim_{|n|\to\infty} n_1^{r_1}n_2^{r_2}n_3^{r_3}p_n = 0 \tag{2.7}$$

for every set of nonnegative integers r_1, r_2, and r_3.[17] Corresponding to each sequence satisfying (2.7) there is exactly one function in $\mathcal{S}$.

If $f(\mathbf{x}) = \sum_{n\in\Delta} p_nh_n(\mathbf{x})$ and $g(\mathbf{x}) = \sum_{n\in\Delta} q_nh_n(\mathbf{x})$ are two elements of $\mathcal{S}$, then it is easily shown that

$$(f, g) = \sum_{n\in\Delta} p_nq_n. \tag{2.8}$$

The following lemma relates convergence criteria in the two characterizations of $\mathcal{S}$.

Lemma 2.4. Let $f_j(\mathbf{x}) = \sum_{n\in\Delta} p_n^{(j)}h_n(\mathbf{x})$, $j = 1, 2, \cdots$ be a sequence of functions in $\mathcal{S}$.

Let

$$\epsilon_{nm}^{(j)} = \sup_{\mathbf{x}}|x^nD^mf_j(\mathbf{x})|,$$

and

$$\delta_r^{(j)} = \sup_n n_1^{r_1}n_2^{r_2}n_3^{r_3}\,|p_n^{(j)}|.$$

Then the statement $\lim_{j\to\infty} f_j(\mathbf{x}) = 0$ in the topology of $\mathcal{S}$ is given equivalently by either

(1) $\lim_{j\to\infty} \epsilon_{nm}^{(j)} = 0$ for every $n, m \in \Delta$,

or

(2) $\lim_{j\to\infty} \delta_r^{(j)} = 0$ for every $r \in \Delta$.

In the remainder of this paper, we occasionally refer to a sequence $\{p_n\}$ as an element of $\mathcal{S}$, meaning thereby the function $f(\mathbf{x}) \in \mathcal{S}$ which has $\{p_n\}$ as its "coordinates."

C. Construction of the Operators $V[g]$ and $W[f]$

Let $L^2(R)$ denote the separable Hilbert space whose elements are (equivalence classes of) complex

[29] J. L. Kelley, *General Topology* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1955), Chap. 2.

[30] G. Sansone, *Orthogonal Functions* (Interscience Publishers, Inc., New York, 1959).

valued, Lebesgue square-integrable functions defined on the real line R.[20] Let $f(x) \in L^2(R)$; then a canonical, irreducible representation of the Weyl form of the single-particle commutation relations is defined by[15]

$$(V_0[q]f)(x) = f(x - q), \quad (W_0[p]f)(x) = e^{ipx}f(x). \tag{2.9}$$

The two continuous, one-parameter groups of unitary operators $V_0[q]$ and $W_0[p]$ satisfy the commutation relation

$$U_0[p, q] \equiv V_0[q]W_0[p] = e^{-ipq}W_0[p]V_0[q]. \tag{2.10}$$

In addition, we can write

$$V_0[q] = e^{-iqP}, \qquad W_0[p] = e^{ipQ},$$

where P and Q are self-adjoint operators whose existence is ensured by Stone's theorem.[31] It is well known that P and Q are defined by

$$(Pf)(x) = -i\, df(x)/dx, \qquad (Qf)(x) = xf(x);$$

the domain of P, $\mathfrak{D}_P$ is the set of all $f(x) \in L^2(R)$ such that $f(x)$ is absolutely continuous on every bounded interval, and $df(x)/dx \in L^2(R)$, and $\mathfrak{D}_Q$, the domain of Q, is the set of all $f(x) \in L^2(R)$ such that $xf(x) \in L^2(R)$. It was proved in Sec. IV.3 of Ref. 5 that every vector in $L^2(R)$ is cyclic with respect to $U_0[p, q]$.

Let Δ be the countable set of all triples of nonnegative integers, and assign to each $n \in \Delta$ the Hilbert space $\mathfrak{H}_n = L^2(R)$. According to Lemma 2.1, an IDPS is characterized by the choice of a product reference vector. Among all such vectors we confine our attention to a distinguished subset defined as follows:

Definition 2.6. Denote by $\mathfrak{R}$ the set of all c_0 vectors $\mathrm{X} = \prod_{n\in\Delta} \otimes \chi_n$ whose product components satisfy the following conditions:

(1) $||\chi_n|| = 1$, $n \in \Delta$;
(2) $\chi_n \in \mathfrak{D}_P \cap \mathfrak{D}_Q$, $n \in \Delta$;
(3) $||P\chi_n|| + ||Q\chi_n|| \leq A(|n|)$, $n \in \Delta$, where $A(|n|)$ is some polynomial depending on X.

We now begin our construction of a representation of the CCR on $\mathfrak{H}_\mathrm{X}$ for $\mathrm{X} \in \mathfrak{R}$ as follows. For each

$$\Lambda = \sum_{j=1}^{M} \alpha_j \Big(\prod_{n\in\Delta} \otimes \lambda_n^{(j)}\Big) \in \mathfrak{G}_\mathrm{X},$$

the reference set, put

$$V[g]\Lambda = \sum_{j=1}^{M} \alpha_j \Big(\prod_{n\in\Delta} \otimes V_0[q_n]\lambda_n^{(j)}\Big), \tag{2.11a}$$

$$W[f]\Lambda = \sum_{j=1}^{M} \alpha_j \Big(\prod_{n\in\Delta} \otimes W_0[p_n]\lambda_n^{(j)}\Big), \tag{2.11b}$$

where $f \in \mathcal{S}$, $g \in \mathcal{S}$, and $\{p_n\}$ and $\{q_n\}$ are the coordinates of f and g, respectively.

We next establish that Conditions 1–3 on $\mathrm{X} \in \mathfrak{R}$ imply that the vectors on the right-hand sides of (2.11a) and (2.11b) are indeed elements of $\mathfrak{H}_\mathrm{X}$. It is clearly sufficient to show that each term in the sum is an element of $\mathfrak{H}_\mathrm{X}$. Consider, for example $\Lambda^{(j)} = \prod_{n\in\Delta} \otimes V_0[q_n]\lambda_n^{(j)}$, a term in the sum (2.11a). This a c_0 vector, for

$$\sum_{n\in\Delta} |\,||V_0[q_n]\lambda_n^{(j)}|| - 1| = \sum_{n\in\Delta} |\,||\lambda_n^{(j)}|| - 1|, \tag{2.12}$$

and the series on the right of (2.12) is a convergent series because $\lambda_n^{(j)} \neq \chi_n$ for at most a finite number of values of n, and $\prod_{n\in\Delta} \otimes \chi_n = \mathrm{X}$ is by definition a c_0 vector. Furthermore, we can show that $\Lambda^{(j)} \approx \mathrm{X}$; by Definition 2.2, this is true if and only if

$$\sum_{n\in\Delta} |(V_0[q_n]\lambda_n^{(j)}, \chi_n) - 1| < \infty, \tag{2.13}$$

but again because $\lambda_n^{(j)} \neq \chi_n$ for at most finitely-many n values, Eq. (2.13) is true if and only if

$$\sum_{n\in\Delta} |(V_0[q_n]\chi_n, \chi_n) - 1| < \infty. \tag{2.14}$$

In Sec. IV 4 we showed that the function $F_n(q_n) = (V_0[q_n]\chi_n, \chi_n)$ is a continuous, bounded function, possessing a continuous bounded derivative, and for which

$$F_n'(q_n) = i(V_0[q_n]P\chi_n, \chi_n). \tag{2.15}$$

The mean-value theorem, coupled with the fact $F_n(0) = 1$, yields

$$F_n(q_n) - 1 = iq_n(V_0[\theta q_n]P\chi_n, \chi_n), \tag{2.16}$$

where $0 < \theta < 1$. Schwartz's inequality and Properties 1 and 3 of Definition 2.6 then lead, in view of the unitarity of $V_0[q]$, to

$$|F_n[q_n] - 1| \leq |q_n|\, ||P\chi_n|| \leq |q_n|\, A(|n|). \tag{2.17}$$

Therefore

$$\sum_{n\in\Delta} |(V_0[q_n]\chi_n, \chi_n) - 1| \leq \sum_{n\in\Delta} |q_n|\, A(|n|) < \infty, \tag{2.18}$$

convergence following because $\{q_n\}$ satisfies (2.7) and $A(|n|)$ is a polynomial. Thus $\Lambda^{(j)} \approx \mathrm{X}$, and hence by Definition 2.3, $\Lambda^{(j)} \in \mathfrak{H}_\mathrm{X}$. The proof that each term in the sum (2.11b) is an element of $\mathfrak{H}_\mathrm{X}$ is the same, and we omit it.

Equations (2.11a) and (2.11b) thus define two obviously linear transformations of the dense set $\mathfrak{G}_\mathrm{X}$

[31] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955).

into $\mathfrak{H}_X$. From the unitarity of $V_0[g]$ and $W_0[p]$ and the form of the inner product in $\mathfrak{H}_X$, it is simply shown that $||V[g]\Lambda|| = ||W[f]\Lambda|| = ||\Lambda||$ for all $\Lambda \in \mathfrak{G}_X$. In standard fashion $V[g]$ and $W[f]$ can be uniquely extended by continuity to isometric mappings defined on all of $\mathfrak{H}_X$.[32] It is obvious that $V[0] = W[0] = I$, where I is the unit operator.

We now show that if $\Phi = \prod_{n\in\Delta} \otimes \varphi_n \in \mathfrak{H}_X$ is an arbitrary c_0 vector (we now allow $\varphi_n \neq \chi_n$ for infinitely many n values), then we can make the definition of $V[g]$ and $W[f]$ more explicit than above; namely that

$$V[g]\Phi = \prod_{n\in\Delta} \otimes V_0[q_n]\varphi_n, \tag{2.19}$$

and

$$W[f]\Phi = \prod_{n\in\Delta} \otimes W_0[p_n]\varphi_n. \tag{2.20}$$

Again we prove only one case, that of (2.19), the proof of (2.20) being just the same. Consider the sequence $\Phi_N = \prod_{n\in\Delta} \otimes \varphi_n^N$, $N = 1, 2, \cdots$, where $\varphi_n^N = \varphi_n$, $|n| \leq N$, $\varphi_n^N = \chi_n$, $|n| > N$. It is clear that $\Phi_N \in \mathfrak{G}_X$, $N = 1, 2, \cdots$, and it is readily shown that $\prod_{n\in\Delta} \otimes V_0[q_n]\varphi_n \approx \prod_{n\in\Delta} \otimes V_0[q_n]\chi_n \in \mathfrak{H}_X$, that $\lim_{N\to\infty} ||\Phi - \Phi_N|| = 0$, and that

$$\lim_{N\to\infty} ||\prod_{n\in\Delta} \otimes V_0[q_n]\varphi_n - V[g]\Phi_N|| = 0.$$

Since $V[g]\Phi = \lim_{N\to\infty} V[g]\Phi_N$, this completes the proof.

We can immediately apply this fact to deduce several important properties of $V[g]$ and $W[f]$. For any vector $\Lambda \in \mathfrak{G}_X$, and $g, g' \in \mathcal{S}$, we can apply $V[g']$ to both sides of (2.11a), and make use of linearity and (2.19) to obtain

$$V[g']V[g]\Lambda = \sum_{j=1}^{M} \alpha_j(\prod_{n\in\Delta} \otimes V_0[q_n']V_0[q_n]\lambda_n^{(j)}). \tag{2.21}$$

However, $V_0[q_n']V_0[q_n] = V_0[q_n' + q_n]$, $n \in \Delta$, and $\{q_n' + q_n\}$ is just the coordinate sequence of $g' + g$. Therefore, for all $\Lambda \in \mathfrak{G}_X$ we have $V[g']V[g]\Lambda = V[g' + g]\Lambda$. Again, since $\mathfrak{G}_X$ is dense in $\mathfrak{H}_X$, this relation must hold by continuity for all $\Lambda \in \mathfrak{H}_X$. Therefore

$$V[g']V[g] = V[g' + g]. \tag{2.22}$$

In just the same fashion we can prove that

$$W[f']W[f] = W[f' + f]. \tag{2.23}$$

In particular, for every $\Phi \in \mathfrak{H}_X$, we have

$$\Phi = V[g](V[-g]\Phi), \quad \text{and} \quad \Phi = W[f](W[-f]\Phi),$$

which proves that the ranges of the isometric operators $V[g]$ and $W[f]$ are both equal to the whole space $\mathfrak{H}_X$. In other words, $V[g]$ and $W[f]$ are unitary operators for every $f, g \in \mathcal{S}$.

The fulfillment of the Weyl form of the CCR is also a consequence of Eqs. (2.19) and (2.20). If $\Lambda \in \mathfrak{G}_X$, then

$$V[g]W[f]\Lambda = \sum_{j=1}^{M} \alpha_j(\prod_{n\in\Delta} \otimes V_0[q_n]W_0[p_n]\lambda_n^{(j)}), \tag{2.24}$$

and

$$W[f]V[g]\Lambda = \sum_{j=1}^{M} \alpha_j(\prod_{n\in\Delta} \otimes W_0[p_n]V_0[q_n]\lambda_n^{(j)}). \tag{2.25}$$

Next notice that the easily proved inequality $|e^{-ip_nq_n} - 1| \leq |p_n|\cdot|q_n|$ implies that

$$\sum_{n\in\Delta} |e^{-ip_nq_n} - 1| \leq \sum_{n\in\Delta} (|p_n|\cdot|q_n|) < \infty, \tag{2.26}$$

where the series on the right of (2.26) converges because $\{p_n\}$ and $\{q_n\}$ satisfy (2.7). Then, according to Eq. (2.10) and Lemma 2.2, each term in the sum on the right-hand side of (2.24) is just

$$\prod_{n\in\Delta} e^{-ip_nq_n} = e^{-i\sum_{n\in\Delta} p_nq_n} = e^{-i(f,g)} \tag{2.27}$$

times the corresponding term in the sum on the right-hand side of (2.25). In other words, we have

$$V[g]W[f]\Lambda = e^{-i(f,g)}W[f]V[g]\Lambda. \tag{2.28}$$

We have proved (2.28) for all $\Lambda \in \mathfrak{G}_X$, but by continuity it holds for all $\Lambda \in \mathfrak{H}_X$. We collect the results so far proved into

Theorem 2.1. Let $X \in \mathfrak{R}$ be a product reference vector satisfying Conditions 1–3 of Definition 2.6. Then X determines an IDPS $\mathfrak{H}_X$ and, by means of Eqs. (2.11a) and (2.11b), two functions, $V[g]$ and $W[f]$, are defined each of which map all of $\mathcal{S}$ into the group $\mathfrak{G}$ of all unitary operators on $\mathfrak{H}_X$. These two operator-valued functions on $\mathcal{S}$ satisfy Conditions 1, 2, and 3 of Definition 1.1.

In what follows we shall refer to the $V[g]$ and $W[f]$ of Theorem 2.1 as the *unitary maps determined by* X.

Continuity of Operators

We next show that $V[g]$ and $W[f]$ are strongly continuous functions on $\mathcal{S}$. Again we give the proof only for $V[g]$; the proof that $W[f]$ is continuous is just the same. Because $V[g]$ is unitary and

$$V[g + g'] = V[g]V[g'],$$

it is sufficient to prove continuity at the origin. Let $\{g^j\}$, $j = 1, 2, \cdots$ be any sequence of functions

[32] S. Bochner and K. Chandrasekharan, *Fourier Transforms* (Princeton University Press, Princeton, New Jersey, 1949), pp. 92–93.

converging to zero in the topology of $\mathcal{S}$. We show first that

$$\lim_{j\to\infty} ||X - V[g^j]X|| = 0, \qquad (2.29)$$

where $X = \prod_{n\in\Delta} \otimes \chi_n$ is the product reference vector. Using the definition of $V[g^j]$, and noting that X is a unit vector we immediately find that

$$||X - V[g^j]X||^2 = 2 \operatorname{Re}\,[1 - \prod_{n\in\Delta} (V_0[q_n^j]\chi_n, \chi_n)], \qquad (2.30)$$

where $\{q_n^j\}$ are the coordinates of g^j. Since $\{q_n^j\}$ are the coordinates of a null sequence, given an arbitrary triple of positive integers, $r = (r_1, r_2, r_3)$, according to Lemma 2.4 there exists a constant c_r, which is independent of j, and which satisfies

$$(n_1^{r_1} + 1)(n_2^{r_2} + 1)(n_3^{r_3} + 1)\ |q_n^j| < c_r, \quad n \in \Delta. \qquad (2.31)$$

The infinite product in (2.30) converges absolutely and uniformly with respect to $\{q_n^j\}$ if and only if the series $\sum_{n\in\Delta} |(V_0[q_n^j]\chi_n, \chi_n) - 1|$ converges uniformly.[33] However, according to (2.18), this series is majorized by $\sum_{n\in\Delta} |q_n^j|\ A(|n|)$, which in turn, according to (2.31), is majorized by

$$c_r \sum_{n\in\Delta} A(|n|)/(n_1^{r_1} + 1)(n_2^{r_2} + 1)(n_3^{r_3} + 1)$$

for a suitable choice of r. This last series is convergent independently of j which proves the uniform convergence.

Because of this uniform convergence, given any $\epsilon > 0$, there exists an integer N_ϵ, which is independent of j, such that

$$\prod_{|n|>N_\epsilon} (V_0[q_n^j]\chi_n, \chi_n) = 1 + \delta_j, \quad |\delta_j| < \epsilon. \qquad (2.32)$$

Each of the functions $(V_0[q_n^j]\chi_n, \chi_n)$ is continuous, and since $q_n^j \to 0$ as $j \to \infty$ for each n, we can pick an integer J_ϵ such that

$$\prod_{|n|\le N_\epsilon} (V_0[q_n^j]\chi_n, \chi_n) = 1 + \rho_j, \quad |\rho_j| < \epsilon \text{ for all } j \ge J_\epsilon. \qquad (2.33)$$

From (2.30), (2.32), and (2.33), it follows that for all $j \ge J_\epsilon$,

$$||X - V[g^j]X||^2 \le 2(|\delta_j| + |\rho_j| + |\delta_j|\,|\rho_j|) < 4\epsilon + 2\epsilon^2. \qquad (2.34)$$

Since ϵ was arbitrary, this proves (2.29).

If $\Lambda \in \mathfrak{G}_X$, then Λ consists of a finite sum of product vectors, each of whose products components differ from the product components of X for only a finite number of n values. Therefore it is readily seen that

$$\lim_{j\to\infty} ||\Lambda - V[g^j]\Lambda|| = 0 \qquad (2.35)$$

holds for all $\Lambda \in \mathfrak{G}_X$. Since $\mathfrak{G}_X$ is dense in $\mathfrak{H}_X$, and $V[g]$ is unitary, it follows that (2.35) holds for all $\Lambda \in \mathfrak{H}_X$. We summarize these results as

Theorem 2.2. If $X \in \mathfrak{N}$ is an arbitrary product reference vector, then the unitary maps $V[g]$ and $W[f]$ determined by X are strongly continuous in the topology of $\mathcal{S}$.

We now state two immediate corollaries:

Corollary 1. For each fixed $g \in \mathcal{S}$, and $f \in \mathcal{S}$, $V[tg]$ and $W[tf]$ are strongly continuous functions of t, $-\infty < t < \infty$, in the usual topology for R.

Proof. Let $t \in R$, and let t_i be any sequence of real numbers such that $\lim_{i\to\infty} t_i = t$. Then for any $f \in \mathcal{S}$, $t_i f \to tf$ in the topology of $\mathcal{S}$, because

$$\lim_{i\to\infty} \sup_x |x^n D^m[t_i f(x) - tf(x)]| = \lim_{i\to\infty} |t_i - t| \sup_x |x^n D^m f(x)| = 0$$

for all n, m. Then by Theorem 2.2 it follows that $V[tg]$ and $W[tf]$ are the strong limits of $V[t_i g]$ and $W[t_i f]$, respectively, which completes the proof.

Corollary 2. The unitary map of $\mathcal{S} \times \mathcal{S} \to \mathfrak{G}$ defined by $\{f, g\} \to V[g]W[f] \equiv U[f, g]$ is strongly continuous on $\mathcal{S} \times \mathcal{S}$ supplied with the product topology.

Proof. The proof is almost identical to the simple proof of Lemma IV 3.2, a similar result for finitely many degrees of freedom, and is omitted.

Irreducibility

We now show that if $V[g]$ and $W[f]$ are defined by (2.11), then the self-adjoint set of operators $\mathfrak{M} = \{V[g], W[f]\}$ (for all f, g in $\mathcal{S}$) is irreducible. We show this by proving the equivalent result that $\mathfrak{M}''$, the bicommutator of $\mathfrak{M}$, is equal to $\mathfrak{B}$, the ring of all bounded operators in $\mathfrak{H}_X$. Let n be a fixed element of Δ, and let $\mathfrak{M}_n = \{V[q_n h_n], W[p_n h_n]\}$, where q_n and p_n take on all real values and $h_n(x)$ is a fixed basis element of $\mathcal{S}$. Then $\mathfrak{M}_n$ is a self-adjoint set of operators, and since $\mathfrak{M}_n \subset \mathfrak{M}$, then $\mathfrak{M}_n'' \subset \mathfrak{M}''$. If we set $\mathfrak{N}_n = \{V_0[q_n], W_0[p_n]\}$ for all real p_n and q_n, then $\mathfrak{N}_n$ is a self-adjoint set of unitary operators in $\mathfrak{H}_n[\equiv L^2(R)]$, and there exists a straightforward isomorphism between the rings $\mathfrak{M}''$ and $\mathfrak{N}_n''$.[16] However, the set $\mathfrak{N}_n$ is irreducible by assumption,

[33] K. Knopp, *Theory and Application of Infinite Series* (Blackie & Son Ltd., London, 1928), p. 381.

so $\mathfrak{M}_n' = \mathfrak{B}_n$, the ring of all bounded operators in $\mathfrak{H}_n$. Thus $\overline{\mathfrak{M}}_n'$ is the image, $\overline{\mathfrak{B}}_n$, of the ring of operators $\mathfrak{B}_n$ under this isomorphism for each $n \in \Delta$. However, von Neumann has shown that the ring generated by all the $\overline{\mathfrak{B}}_n$, $n \in \Delta$, is just the ring $\overline{\mathfrak{B}}$ of all bounded operators in $\mathfrak{H}_X$.[16] Since $\overline{\mathfrak{B}}_n \subset \overline{\mathfrak{M}}''$ for each $n \in \Delta$, this implies that $\overline{\mathfrak{M}}'' = \overline{\mathfrak{B}}$, and the proof is completed.

Since $\overline{\mathfrak{M}}$ is irreducible, it follows that every $\Psi \in \mathfrak{H}_X$ is a cyclic vector for $\overline{\mathfrak{M}}$. Furthermore, if we let $\overline{\mathfrak{N}} = \{U[f, g]\}$ (all $\{f, g\} \in \mathcal{S} \times \mathcal{S}$) then clearly $\overline{\mathfrak{M}} \subset \overline{\mathfrak{N}}$, so $\overline{\mathfrak{N}}$ is also an irreducible set of unitary operators. We sum up these results on irreducibility in

Theorem 2.3. If $X \in \mathfrak{R}$ is an arbitrary product reference vector, and $V[g]$ and $W[f]$ are the unitary maps determined by X, then the self-adjoint set of operators $\overline{\mathfrak{M}} = \{V[g], W[f]\}$ is irreducible. Furthermore, the set of operators $\overline{\mathfrak{N}} = \{U[f, g]\}$ is also irreducible. As a consequence every $\Psi \in \mathfrak{H}_X$ is a cyclic vector for both $\overline{\mathfrak{M}}$ and $\overline{\mathfrak{N}}$.

This completes our construction of the Weyl form of the CCR and we state this result as

Theorem 2.4. The strongly continuous unitary maps $V[g]$ and $W[f]$ determined by any product reference vector $X \in \mathfrak{R}$, $f, g \in \mathcal{S}$, form an irreducible representation of the CCR.

It is clear from our construction, that if $X_1 \in \mathfrak{R}$ and $X_2 \in \mathfrak{R}$ are equivalent (Definition 2.2), then they determine identical representations of the CCR. For in that case $\mathfrak{H}_{X_1} \equiv \mathfrak{H}_{X_2}$, and then Eqs. (2.19) and (2.20) show that both representations are identical when confined to product vectors (whose finite linear sums are dense in $\mathfrak{H}_X$). It would thus be more appropriate to say that each equivalence class in $\mathfrak{R}$ determines a distinct representation of the CCR, and in the future, when we speak of the representation determined by X, we shall mean the representation determined by the equivalence class of which X is a representative.

Discussion

We should like to emphasize the essential simplicity of the representations of the CCR which we have constructed and the close conceptual relationship of them to the Fock representation. Indeed, if the product reference vector is chosen so that $\chi_n \equiv \varphi_0$, where φ_0 is the ground state of an harmonic oscillator, then the resulting representation of the CCR determined by X is just a Fock representation.[10] However, the representations constructed here are *not* all unitarily equivalent to Fock representations; this important fact is proved in Sec. 4.

It might be thought that a more general representation of the CCR could be constructed if instead of requiring that $\mathfrak{H}_n \equiv L^2(R)$ for all $n \in \Delta$, we let $\mathfrak{H}_n$ be an arbitrary, separable Hilbert space, and if furthermore in each $\mathfrak{H}_n$ we pick an *irreducible but otherwise arbitrary* representation of the single-particle commutation relations, V_n and W_n. The construction of $V[g]$ and $W[f]$ could then be carried through just as before. However, a theorem of von Neumann[15] guarantees the existence of a unitary map T_n of $\mathfrak{H}_n$ onto $L^2(R)$, with the property that $T_n V_n[q_n] T_n^{-1} = V_0[q_n]$ and $T_n W_n[p_n] T_n^{-1} = V_0[p_n]$ for each $n \in \Delta$. It is then easy to show that the "more general" representation is in fact unitarily equivalent to one of our canonical representations.

D. Field Operators

We turn now to a discussion of the infinitesimal generators of the operators $W[f]$ and $V[g]$ determined by a product reference vector $X \in \mathfrak{R}$, that is, to the field operators $\varphi(f)$ and $\pi(g)$. Since $W[f]$ and $V[g]$ satisfy Conditions 1, 2, and 4 of Definition 1.1, it follows that for any fixed f and $g \in \mathcal{S}$, $W[tf]$ and $V[tg]$ form two continuous, one-parameter groups of unitary operators.[31] Then a theorem of Stone[31] asserts the existence of two self-adjoint (generally unbounded) operators $\varphi(f)$ and $\pi(g)$ such that

$$W[tf] = e^{it\varphi(f)}, \qquad V[tg] = e^{-it\pi(g)}. \tag{2.36}$$

We denote by $\mathfrak{D}_{\varphi(f)}$ and $\mathfrak{D}_{\pi(g)}$ the domains of the respective operators $\varphi(f)$ and $\pi(g)$. From Stone's theorem

$$\varphi(f) = \lim_{t \to 0} (1/it)(W[tf] - I), \tag{2.37}$$

where the limit is taken in the strong sense, and $\Psi \in \mathfrak{D}_{\varphi(f)}$ if and only if $\lim (1/it)\ (W[tf] - I)\Psi$ exists. The same statements are true of $V[tg]$ and $\pi(g)$, except that it must be replaced by $-it$ in the denominators of the corresponding expressions.

It has been shown by Lew[14] that given *any* representation of the CCR, $V[g]$ and $W[f]$, which satisfies the conditions of Definition 1.1, then the field operators $\varphi(f)$ and $\pi(g)$, defined with the aid of Stone's theorem, are "linear functionals on $\mathcal{S}$" in the following restricted sense: Let $\mathfrak{H}$ be the Hilbert space in which the given representation is defined. Let $(f_1, \cdots, f_n)$ be any finite set of functions in $\mathcal{S}$. Then there exists a dense linear manifold $\mathfrak{K} \subset \mathfrak{H}$, invariant under $V[tf_i]$ and $W[tf_i]$ on which all finite linear combinations and finite products of $\varphi(f_i)$ and

$\pi(f_i)$ are defined. Furthermore, on this linear manifold, φ and π satisfy

$$\begin{aligned} \varphi(\alpha f_i + \beta f_k) &= \alpha\varphi(f_i) + \beta\varphi(f_k); \\ \pi(\alpha f_i + \beta f_k) &= \alpha\pi(f_i) + \beta\pi(f_k), \end{aligned} \tag{2.38}$$

$$\begin{aligned} [\varphi(f_i), \varphi(f_k)] &= [\pi(f_i), \pi(f_k)] = 0; \\ [\varphi(f_i), \pi(f_k)] &= i(f_i, f_k)I, \end{aligned} \tag{2.39}$$

where α and β are real numbers and $j, k = 1, 2, \cdots, n$.

An additional fact, which is of use later in this paper, is the following. Let $\Psi \in \mathfrak{D}_{\varphi(f_0)}$; then for all $f, g \in \mathcal{S}$, $U[f, g]\Psi \in \mathfrak{D}_{\varphi(f_0)}$. For from the commutation relations satisfied by $V[g]$ and $W[f]$, and from the definition of $U[f, g]$, it can easily be shown that

$$\begin{aligned} &(1/it)\{W[tf_0] - I\}U[f, g]\Psi \\ &\quad = U[f, g]e^{it(f_0, g)}(1/it)\{W[tf_0] - I\}\Psi \\ &\qquad + U[f, g](1/it)[e^{it(f_0, g)} - 1]\Psi. \end{aligned} \tag{2.40}$$

Therefore

$$\begin{aligned} &\lim_{t\to 0} (1/it)\{W[tf_0] - I\}U[f, g]\Psi \\ &\quad = U[f, g]\varphi(f_0)\Psi + (f_0, g)U[f, g]\Psi. \end{aligned} \tag{2.41}$$

This implies that $U[f, g]\Psi \in \mathfrak{D}_{\varphi(f_0)}$, and that further

$$\varphi(f_0)U[f, g] = U[f, g][\varphi(f_0) + (f_0, g)I]. \tag{2.42}$$

In the same way, it can be shown that if $\Psi \in \mathfrak{D}_{\pi(g_0)}$, then $U[f, g]\Psi \in \mathfrak{D}_{\pi(g_0)}$, and also

$$\pi(g_0)U[f, g] = U[f, g][\pi(g_0) + (f, g_0)I]. \tag{2.43}$$

We discuss now, without proof, several additional properties of the field operators for a representation determined by a product reference vector X. We omit the proofs, for while not difficult, they are a little lengthy, and the results proved are not essential for the purposes of this paper. They do, however, shed some light on the nature and structure of the representations we have constructed.

Let $\varphi(f)$ and $\pi(g)$ be the field operators in a representation determined by the product reference vector $\mathrm{X} = \prod_{n\in\Delta} \otimes \chi_n$. Let $\mathfrak{A}_\mathrm{X}$ be the set of unit product vectors $\Lambda = \prod_{n\in\Delta} \otimes \lambda_n$, such that $\lambda_n \in \mathfrak{D}_p \cap \mathfrak{D}_Q$, all $n \in \Delta$, $||Q\lambda_n|| + ||P\lambda_n|| < A(|n|)$ for some fixed polynomial $A(|n|)$, and $\lambda_n \neq \chi_n$ for only a finite number of values of n. Then $A(|n|)$ can be chosen so that $\mathrm{X} \in \mathfrak{A}_\mathrm{X}$, so that the set of all finite linear combinations of elements in $\mathfrak{A}_\mathrm{X}$ is dense in $\mathfrak{H}_\mathrm{X}$, and so that

$$\mathfrak{A}_\mathrm{X} \subset \Big(\bigcap_{f\in\mathcal{S}} \mathfrak{D}_{\varphi(f)}\Big) \cap \Big(\bigcap_{g\in\mathcal{S}} \mathfrak{D}_{\pi(g)}\Big).$$

Furthermore, if $\Lambda \in \mathfrak{A}_\mathrm{X}$, then

$$\varphi(f)\Lambda = \sum_{n\in\Delta} p_n(Q\lambda_n \otimes \prod_{m\in\Delta-(n)} \otimes \lambda_m), \tag{2.44}$$

$$\pi(g)\Lambda = \sum_{n\in\Delta} q_n(P\lambda_n \otimes \prod_{m\in\Delta-(n)} \otimes \lambda_m). \tag{2.45}$$

Assume now that the product reference vector X satisfies, in addition to Conditions 1, 2, and 3 of Definition 2.7, the two conditions:

(4) $\chi_n \in \mathfrak{D}_{P^2} \cap \mathfrak{D}_{Q^2} \cap \mathfrak{D}_{QP} \cap \mathfrak{D}_{PQ}$, $\quad n \in \Delta$,

(5) $||P^2\chi|| + ||Q^2\chi|| + ||PQ\chi|| + ||QP\chi|| \leq B(|n|)$, $\quad n \in \Delta$,

where $B(|n|)$ is a fixed polynomial. Let $\mathfrak{B}_\mathrm{X} \subset \mathfrak{A}_\mathrm{X}$ be the set of all $\Lambda \in \mathfrak{A}_\mathrm{X}$ that also satisfy Conditions 4 and 5 above. Then $B(|n|)$ can be chosen so that the set of all finite linear combinations of elements in $\mathfrak{B}_\mathrm{X}$ is also dense in $\mathfrak{H}_\mathrm{X}$, and each $\Lambda \in \mathfrak{B}_\mathrm{X}$ is in the domain of all operators of the form $\varphi(f_1)\varphi(f_2)$, $\pi(g_1)\pi(g_2)$, $\varphi(f_1)\pi(g_1)$, and $\pi(g_1)\varphi(f_1)$. The CCR are naturally satisfied in $\mathfrak{B}_\mathrm{X}$. One can clearly continue in this fashion and construct representations of the CCR determined by a product reference vector X so that a dense set containing X is in the domain of all polynomials in $\varphi(f)$ and $\pi(g)$ of some fixed degree.

E. A Generalization to Multifield Representations

We sketch now in briefest outline a construction parallel to the foregoing development that yields additional representations of the CCR. We are motivated in part by a possible formulation pertinent to a finite number, K, of independent scalar fields for which some special invariance is desired, such as a rotational invariance for an isovector field. Also several examples of von Neumann[16] and of Araki and Woods[10] suggest this generalization as well.

The generalization we wish to consider is the following. Instead of assuming that in each $\mathfrak{H}_n$ there is defined an irreducible representation of the *single*-particle commutation relations, we assume that there is defined in each $\mathfrak{H}_n$ an irreducible representation of the commutation relations for K degrees of freedom, and that the test function space $\mathcal{S}$ will be replaced by $\mathcal{S}_K$, the K-fold direct product of $\mathcal{S}$ with itself supplied with the product topology.

The elements $\mathbf{f}$ of $\mathcal{S}_K$ are now ordered K-tuples of test functions in $\mathcal{S}$, $\mathbf{f} = (f_1, \cdots, f_K)$, and a pairing is introduced into $\mathcal{S}_K$ by

$$(\mathbf{f}; \mathbf{g}) \equiv \sum_{i=1}^{K} (f_i, g_i). \tag{2.46}$$

It is not difficult to show that $\mathcal{S}_K$ is isomorphic to a

space of infinite sequences of K-tuples of real numbers, the isomorphism being given by $\mathbf{f} \approx \{\mathbf{p}_n\}$, $n \in \Delta$, where $\mathbf{p}_n = (p_{1n}, p_{2n}, \cdots, p_{Kn})$, and p_{in} is the nth coordinate of f_i in an expansion in Hermite functions. We shall call $\mathbf{p}_n$ the nth coordinate of $\mathbf{f}$. Let

$$\mathbf{p}_n \cdot \mathbf{q}_n = \sum_{i=1}^{K} p_{in} q_{in}, \qquad |\mathbf{p}_n|^2 = \mathbf{p}_n \cdot \mathbf{p}_n;$$

then $\{\mathbf{p}_n\}$ is the coordinate sequence of some element in $\mathcal{S}_K$ if and only if

$$\lim_{|n| \to \infty} n_1^{r_1} n_2^{r_2} n_3^{r_3} \, |\mathbf{p}_n| = 0$$

for every set of nonnegative integers r_1, r_2, r_3. If $\mathbf{f} \approx \{\mathbf{p}_n\}$, $\mathbf{g} \approx \{\mathbf{q}_n\}$, then $(\mathbf{f}; \mathbf{g}) = \sum_{n \in \Delta} \mathbf{p}_n \cdot \mathbf{q}_n$. Furthermore, if $\mathbf{f}^{(j)} = \{\mathbf{p}_n^{(j)}\}$, $j = 1, 2, \cdots$, then $\lim_{j \to \infty} \mathbf{f}^{(j)} = 0$ in the topology of $\mathcal{S}_K$ if and only if

$$\limsup_{j \to \infty} n_1^{r_1} n_2^{r_2} n_3^{r_3} \, |\mathbf{p}_n^{(j)}| = 0$$

for every set of nonnegative integers r_1, r_2, and r_3.

We now set $\mathfrak{H}_n \equiv L^2(R_k)$ for each $n \in \Delta$, where $L^2(R_K)$ is the Hilbert space of square integrable functions of K real variables. We adopt the canonical, irreducible representation of the K-particle commutation relations given by

$$(V_0[\mathbf{q}]f)(\mathbf{x}) = f(\mathbf{x} - \mathbf{q}); \quad (W_0[\mathbf{p}]f)(\mathbf{x}) = e^{i\mathbf{p}\cdot\mathbf{x}} f(\mathbf{x}),$$

for which Eq. (2.9) is replaced by

$$U_0[\mathbf{p}, \mathbf{q}] \equiv V_0[\mathbf{q}] W_0[\mathbf{p}] = e^{-i\mathbf{p}\cdot\mathbf{q}} W_0[\mathbf{p}] V_0[\mathbf{q}]. \tag{2.47}$$

Recourse to Stone's theorem[31] shows the existence of $2K$ self-adjoint operators P_j and Q_j, $j = 1, 2, \cdots, K$ such that

$$V_0[\mathbf{q}] = \exp\left(-i \sum_{j=1}^{K} q_j P_j\right),$$

$$W_0[\mathbf{p}] = \exp\left(i \sum_{j=1}^{K} p_j Q_j\right).$$

Now replace Definition 2.6 by

Definition 2.7. Denote by $\mathfrak{R}_K$ the set of all c_0 vectors, $\mathbf{X} = \prod_{n \in \Delta} \otimes \chi_n \in \mathfrak{H}$, whose product components satisfy the following conditions:

(1) $||\chi_n|| = 1, \qquad n \in \Delta;$

(2) $\chi_n \in \bigcap_{j=1}^{K} (\mathfrak{D}_{P_j} \cap \mathfrak{D}_{Q_j}), \qquad n \in \Delta;$

(3) $\sum_{j=1}^{K} (||Q_j \chi_n|| + ||P_j \chi_n||) \leq A(|n|), \qquad n \in \Delta,$

where $A(|n|)$ is some polynomial depending on $\mathbf{X}$.

If now everywhere in the construction of $V[g]$ and $W[f]$, p_n and q_n are replaced by $\mathbf{p}_n$ and $\mathbf{q}_n$, $V_0[q_n]$ and $W_0[p_n]$ are replaced by $V_0[\mathbf{q}_n]$ and $W_0[\mathbf{p}_n]$, and it is always assumed that $\mathbf{X} \in \mathfrak{R}_K$, then it will be seen that new representations, $V[\mathbf{g}]$, $W[\mathbf{f}]$, of the CCR are defined. The unitary operators $V[\mathbf{g}]$ and $W[\mathbf{f}]$ are strongly continuous in the topology of $\mathcal{S}_K$, the set $\{V[\mathbf{g}], W[\mathbf{f}]\}$ is irreducible, and

$$V[\mathbf{g}] W[\mathbf{f}] = e^{-i(\mathbf{f};\mathbf{g})} W[\mathbf{f}] V[\mathbf{g}], \tag{2.48}$$

which is the generalization of the CCR to K independent scalar fields.

The representations (2.48) for K scalar fields also can be used to generate representations of the CCR for a *single* field, which in general are reducible. For this purpose, we introduce two linear homomorphisms

$$F : \mathbf{f} \to \mathbf{f}^c, \qquad G : \mathbf{g} \to \mathbf{g}^c \tag{2.49a}$$

of $\mathcal{S}_K$ into itself whose image spaces we denote by $\mathcal{S}_K^f$ and $\mathcal{S}_K^g$, i.e., $F : \mathcal{S}_K \equiv \mathcal{S}_K^f$, $G : \mathcal{S}_K \equiv \mathcal{S}_K^g$, respectively. We next require two linear, invertible isomorphisms

$$L : \mathbf{f}^c \to f, \qquad M : \mathbf{g}^c \to g \tag{2.49b}$$

of $\mathcal{S}_K^f$ and $\mathcal{S}_K^g$ onto $\mathcal{S}$, respectively, such that

$$(\mathbf{f}^c; \mathbf{g}^c) = (f, g) \tag{2.50}$$

for each pair $\mathbf{f}^c$, $\mathbf{g}^c$. The transformations (2.49) are analogues of singular transformations of Lie group parameters,[23] and we call such a mapping, as restricted by (2.50), a *contraction*. If we put $V[g] \equiv V[\mathbf{g}^c]$ and $W[f] \equiv W[\mathbf{f}^c]$, then it is clear from (2.48) and (2.50) that these operators obey the CCR for a *single* field. An example of one of the simplest contractions is obtained if we set $F = G$, $L = M$, and for which, if $\mathbf{f} = (f_1, f_2, \cdots, f_K)$, we put $F(\mathbf{f}) \equiv \mathbf{f}^c = (f_1, 0, \cdots, 0)$ and $L : (f_1(\mathbf{x}), 0, \cdots, 0) = f_1(\mathbf{x}) \in \mathcal{S}$. It is quite possible that contracted single field CCR can be catalogued as to algebraic type in terms of the product components $\chi_n \in L^2(R_K)$, $n \in \Delta$, of the product reference vector, along lines similar to those used by Bures[34] to analyze the classic example of von Neumann.[16]

3. CONSTRUCTION OF IRREDUCIBLE SCALAR FIELD CONTINUOUS REPRESENTATIONS

A. Structure of the Continuous Representation

A continuous representation of Hilbert space as defined in Part I is a realization of a given Hilbert space by a space of bounded, continuous functions defined on a topological space called the label space.

[34] D. J. C. Bures, Composito Math. 15, 169 (1963).

The construction is carried out with the aid of a family of unit vectors, called an overcomplete family of states (OFS), which satisfies the three postulates listed in I. We construct in Sec. 3 continuous representations of $\mathfrak{H}_X$, the IDPS determined by the product reference vector $X \in \mathfrak{R}$. The label space is $\mathfrak{L} = \mathfrak{S} \times \mathfrak{S}$, the direct product of test function space with itself, when supplied with the product topology. We shall adopt for our OFS the collection $\mathfrak{S}$ of unit vectors

$$\Phi[f, g] = U[f, g]\Phi_0 \tag{3.1}$$

for all $\{f, g\} \in \mathfrak{L}$, where $\Phi_0 \in \mathfrak{H}_X$ is an arbitrary (but fixed) unit vector in $\mathfrak{H}_X$, $||\Phi_0|| = 1$, called the fiducial vector, and $U[f, g] = V[g]W[f]$ is defined by the representation of the CCR determined by X.

For each $\Psi \in \mathfrak{H}_X$ we define the functional

$$\psi(f, g) \equiv (\Phi[f, g], \Psi), \tag{3.2}$$

which in view of Schwartz's inequality is bounded,

$$|\psi(f, g)| \leq ||\Psi||, \tag{3.3}$$

and which is continuous on $\mathfrak{L}$ by Corollary 2 of Theorem 2.2. We denote by $\mathfrak{C}$ the set of functionals $\psi(f, g)$ determined by all $\Psi \in \mathfrak{H}_X$. Equation (3.2) may be viewed as a map C of $\mathfrak{C}_X$ onto $\mathfrak{C}$, $C : \Psi \to \psi(f, g)$. This map is clearly linear, so that $\mathfrak{C}$ forms a complex, linear vector space composed of bounded, continuous functionals. It should be remarked that while $\mathfrak{C}$ appears to depend on both X and Φ_0, actually Φ_0 alone determines $\mathfrak{C}$. This is because Φ_0 uniquely determines $\mathfrak{H}_X$, and $\mathfrak{H}_X$ in turn uniquely determines the representation of the CCR. When we wish to emphasize the dependence of $\mathfrak{C}$ on Φ_0, we refer to the set $\mathfrak{C}$ corresponding to the fiducial vector Φ_0, or add to $\mathfrak{C}$ distinguishing decorations.

We now introduce an inner product into $\mathfrak{C}$ which will turn $\mathfrak{C}$ into a complete Hilbert space. When this is done, we shall be able to show that $\mathfrak{C}$ is congruent to $\mathfrak{H}_X$ (i.e., the map C is one-to-one and isometric) and so $\mathfrak{C}$ is indeed a continuous representation of $\mathfrak{H}_X$ as defined in Part I.

B. Introduction of Inner Product

A Sequence of Projections

We introduce into $\mathfrak{H}_X$ a sequence of projection operators A_M, $M = 1, 2, \cdots$. These projections are defined for each c_0 vector $\Lambda = \prod_{n\in\Delta} \otimes \lambda_n \in \mathfrak{H}_X$ by

$$A_M\Lambda = \prod_{|n|\leq M} \otimes \lambda_n \otimes \prod_{|n|>M} \otimes (\chi_n, \lambda_n)\chi_n. \tag{3.3}$$

The definition is then extended by linearity and continuity to all of $\mathfrak{H}_X$. It is readily proved that A_M is a projection operator, that $A_1 < A_2 < \cdots$, and that $\lim_{M\to\infty} A_M = I$ (strong convergence). Furthermore, the equation

$$B_M A_M \Lambda = \prod_{|n|>M} (\chi_n, \lambda_n) \cdot \prod_{|n|\leq M} \otimes \lambda_n \tag{3.4}$$

defines a one-to-one, isometric mapping, B_M, of the closed subspace $A_M\mathfrak{H}_X$ onto $\prod_{|n|\leq M} \otimes \mathfrak{H}_n$.

Projection Functionals

Our steps now roughly resemble the methods described by Friedrichs and Shapiro.[19] For every $\Psi \in \mathfrak{H}_X$, we consider the sequence of vectors $\Psi_M = A_M\Psi \in \mathfrak{H}_X$, $M = 1, 2, \cdots$, and the corresponding functions

$$\psi_M(f, g) = (\Phi[f, g], A_M\Psi) \in \mathfrak{C},$$

for all M. In addition, we introduce for each $\{f, g\} \in \mathfrak{L}$ the *truncated* test functions

$$f_{(N)} \equiv \sum_{|n|\leq N} p_n h_n, \quad g_{(N)} \equiv \sum_{|n|\leq N} q_n h_n, \quad N = 1, 2, \cdots, \tag{3.5}$$

where $p_n = (f, h_n)$, $q_n = (g, h_n)$, and $f_{(\infty)} \equiv f$, $g_{(\infty)} \equiv g$. Let us then consider the set of functions

$$\psi_M(f_{(N)}, g_{(N)}) = (U[f_{(N)}, g_{(N)}]\Phi_0, A_M\Psi) \tag{3.6}$$

for all M and N.

Assume first that $N \geq M$. Then since $A_N A_M = A_M$, (3.6) becomes

$$\begin{aligned}\psi_M(f_{(N)}, g_{(N)}) &= (A_N U[f_{(N)}, g_{(N)}]\Phi_0, A_M\Psi) \\ &= (U[f_{(N)}, g_{(N)}]A_N\Phi_0, A_M\Psi).\end{aligned} \tag{3.7}$$

The last result holds in view of the commutation of the operators $A_{N'}$ and $U[f_{(N)}, g_{(N)}]$ whenever $N' \geq N$, as is clear from (3.3), (2.19), and (2.20). By assumption, the set of operators $V_0[q_n]W_0[p_n]$ is irreducible; thus in the closed subspace $A_N\mathfrak{H}_X \subset \mathfrak{H}_X$, the family of operators induced by $U[f_{(N)}, g_{(N)}]$ is irreducible. Consequently, each function $\psi_M(f_{(N)}, g_{(N)})$ in (3.7), $N \geq M$, is (apart from a uniform scale factor $||A_N\Phi_0||$) of the form of those functions in a continuous representation based on an *irreducible* representation of the canonical commutation relations for *finitely many*, $\pi(N)$, degrees of freedom, where $\pi(N) = (N + 1)(N + 2)(N + 3)/6$, the number of triples n such that $|n| \leq N$. Such continuous representations have been studied exhaustively in Part IV. Theorem IV 3.1 then states that such functions are square integrable, and that moreover

$$\begin{aligned}\int \psi_M^*(f_{(N)}, g_{(N)})\lambda_M(f_{(N)}, g_{(N)}) \prod_{|n|\leq N} \left(\frac{1}{2\pi}\right) dp_n\, dq_n \\ = ||A_N\Phi_0||^2\,(A_M\Psi, A_M\Lambda),\end{aligned} \tag{3.8}$$

whenever $N \geq M$, $\Psi, \Lambda \in \mathfrak{H}_X$. If we adopt the convention,

$$d\mu\,(f_{(N)}, g_{(N)}) = \prod_{|n| \leq N} \left(\frac{1}{2\pi}\right) dp_n\, dq_n, \tag{3.9}$$

then from the properties of the projection operators A_M, we find that

$$(\Psi, \Lambda) = \lim_{M\to\infty} \lim_{N\to\infty} \int \psi^*_M(f_{(N)}, g_{(N)}) \times \lambda_M(f_{(N)}, g_{(N)})\, d\mu\,(f_{(N)}, g_{(N)}). \tag{3.10}$$

This formula is characteristic of Friedrichs–Shapiro-type integrals. However, due to our restricted class of "cylinderlike" functionals—for instance, $\psi_M(f_{(N)}, g_{(N)})$ does not even run over all continuous $L^2(R_{\pi(N)} \times R_{\pi(N)})$ functions—we can say considerably more.

Interchangeability of Limits

We next show that the order of taking limits in (3.10) can be interchanged. We recall the standard criterion[35] for interchanging the order of limits in a double sequence: If a_{MN} and $\lim_{M\to\infty} \lim_{N\to\infty} a_{MN} = l$ exist, then $\lim_{N\to\infty} \lim_{M\to\infty} a_{MN}$ exists and has the value l if $\lim_{M\to\infty} a_{MN} = b_N$ exists uniformly in N and $\lim_{N\to\infty} b_N$ exists.

In the application of this criterion to (3.10) we must first examine

$$a_{MN}[\Psi, \Lambda] \equiv \int \psi^*_M(f_{(N)}, g_{(N)}) \times \lambda_M(f_{(N)}, g_{(N)})\, d\mu\,(f_{(N)}, g_{(N)}) \tag{3.11}$$

when $M > N$. From (3.6) it follows that

$$\psi_M(f_{(N)}, g_{(N)}) = (U[f_{(N)}, g_{(N)}]A_M\Phi_0, A_M\Psi), \tag{3.12}$$

and a similar expression holds for $\lambda_M(f_{(N)}, g_{(N)})$. In this case the family of operators induced by $U[f_{(N)}, g_{(N)}]$ in $A_M\mathfrak{H}_X$ is reducible since $M > N$. Theorem IV 3.1 cannot be used to evaluate the integral (3.11), but we can appeal to the theorem of von Neumann [15] on the representation of the commutation relations for finitely many degrees of freedom to get a bound on (3.11). This theorem states that *any* representation for $\pi(N)$ degrees of freedom may be decomposed into a direct sum of irreducible representations for $\pi(N)$ degrees of freedom, which may contain a countable infinity of terms. Thus we can introduce the following decomposition:

$$\psi_M(f_{(N)}, g_{(N)}) = \sum_{r=1}^{L} (U_r[f_{(N)}, g_{(N)}]\Phi^M_{0r}, \Psi^M_r), \tag{3.13a}$$

$$\lambda_M(f_{(N)}, g_{(N)}) = \sum_{r=1}^{L} (U_r[f_{(N)}, g_{(N)}]\Phi^M_{0r}, \Lambda^M_r), \tag{3.13b}$$

where, for all $\Phi \in \mathfrak{H}_X$,

$$A_M\Phi \equiv \Phi^M = \sum_{r=1}^{L} \oplus \Phi^M_r, \qquad ||\Phi^M||^2 = \sum_{r=1}^{L} ||\Phi^M_r||^2, \tag{3.14a}$$

and

$$U[f_{(N)}, g_{(N)}] = \sum_{r=1}^{L} \oplus U_r[f_{(N)}, g_{(N)}] \tag{3.14b}$$

wherein L is a positive integer or infinity, and each $U_r[f_{(N)}, g_{(N)}]$ corresponds to an irreducible representation for $\pi(N)$ degrees of freedom. From our earlier work,[5] it follows that

$$\psi_{Mr}(f_{(N)}, g_{(N)}) \equiv (U_r[f_{(N)}, g_{(N)}]\Phi^M_{0r}, \Psi^M_r), \tag{3.15}$$

as well as $\lambda_{Mr}(f_{(N)}, g_{(N)})$, is a continuous and square-integrable function on $R_{\pi(N)} \times R_{\pi(N)}$ for each r. Then

$$\begin{aligned}\int |\psi^*_{Mr}(f_{(N)}, g_{(N)})\lambda_{Ms}(f_{(N)}, g_{(N)})|\, d\mu\,(f_{(N)}, g_{(N)}) &\leq \left\{\int |\psi_{Mr}(f_{(N)}, g_{(N)})|^2\, d\mu\,(f_{(N)}, g_{(N)}) \times \int |\lambda_{Ms}(f_{(N)}, g_{(N)})|^2\, d\mu\,(f_{(N)}, g_{(N)})\right\}^{\frac{1}{2}} \\ &= ||\Phi^M_{0r}||\, ||\Psi^M_r||\, ||\Phi^M_{0s}||\, ||\Lambda^M_s||. \end{aligned} \tag{3.16}$$

Whence

$$\begin{aligned}\sum_{r,s=1}^{L} \int |\psi^*_{Mr}(f_{(N)}, g_{(N)})\lambda_{Ms}(f_{(N)}, g_{(N)})|\, d\mu\,(f_{(N)}, g_{(N)}) &\leq \left\{\sum_{r=1}^{L} ||\Phi^M_{0r}||\, ||\Psi^M_r||\right\}\left\{\sum_{s=1}^{L} ||\Phi^M_{0s}||\, ||\Lambda^M_s||\right\} \\ &\leq \left\{\sum_{r=1}^{L} ||\Phi^M_{0r}||^2\right\}^{\frac{1}{2}}\left\{\sum_{r=1}^{L} ||\Psi^M_r||^2\right\}^{\frac{1}{2}} \\ &\quad \times \left\{\sum_{s=1}^{L} ||\Phi^M_{0s}||^2\right\}^{\frac{1}{2}}\left\{\sum_{s=1}^{L} ||\Lambda^M_s||^2\right\}^{\frac{1}{2}} \\ &= ||\Phi^M_0||^2\, ||\Psi^M||\, ||\Lambda^M||, \end{aligned} \tag{3.17}$$

which shows that

$$\psi^*_M(f_{(N)}, g_{(N)})\lambda_M(f_{(N)}, g_{(N)}) = \sum_{r=1}^{L}\sum_{s=1}^{L} \psi^*_{Mr}(f_{(N)}, g_{(N)})\lambda_{Ms}(f_{(N)}, g_{(N)})$$

is an integrable function, and that the series may be integrated term by term.[36] In particular, we find

[35] R. Courant, *Differential and Integral Calculus* (Blackie & Son Ltd., London, 1949), p. 105.

[36] P. R. Halmos, *Measure Theory* (Van Nostrand, Princeton, New Jersey, 1950), p. 114.

the interesting result when $M > N$ that

$$\int |\psi_M^*(f_{(N)}, g_{(N)})\lambda_M(f_{(N)}, g_{(N)})|\, d\mu\,(f_{(N)}, g_{(N)}) \leq ||A_M\Phi_0||^2\, ||A_M\Psi||\, ||A_M\Lambda|| \leq ||A_M\Psi||\, ||A_M\Lambda|| \leq ||\Psi||\, ||\Lambda||. \quad (3.18)$$

As a special case of this formula, consider its application to the norm of $\Lambda \equiv \Psi - A_P\Psi$, $P < M$. Then since $A_M\Lambda = A_M\Psi - A_P\Psi$, (3.12) and (3.18) lead to

$$\int |\psi_M(f_{(N)}, g_{(N)}) - \psi_P(f_{(N)}, g_{(N)})|^2\, d\mu\,(f_{(N)}, g_{(N)}) \leq ||A_M\Psi - A_P\Psi||^2. \quad (3.19)$$

Equation (3.8) shows that the inequalities (3.18) and (3.19) also hold if $M \leq N$. Now, making use of (3.18) and (3.19) we easily get the inequality

$$|a_{MN}[\Psi, \Lambda] - a_{PN}[\Psi, \Lambda]| \leq ||\Psi||\, ||A_M\Lambda - A_P\Lambda|| + ||\Lambda||\, ||A_M\Psi - A_P\Psi||. \quad (3.20)$$

This shows that a_{MN} is a Cauchy sequence in M and that the convergence is uniform with respect to N. Thus the existence of a unique limit $b_N = \lim_{M\to\infty} a_{MN}$ is established, and the convergence is uniform in N. Furthermore, $|b_N| \leq ||\Psi||\, ||\Lambda||$ since a_{MN} satisfies the same inequality for all M and N [apply Schwartz's inequality to (3.11) and employ (3.18)]. Since every bounded sequence of complex numbers has at least one convergent subsequence, there always exists a set $\{N_i\}$ such that $\lim_{i\to\infty} b_{N_i}$ exists. Since

$$\lim_{M\to\infty} \lim_{i\to\infty} a_{MN_i}[\Psi, \Lambda] = (\Psi, \Lambda)$$

for any subsequence such that $N_i \to \infty$ as $j \to \infty$, an application of the criterion for interchanging the order of taking limits shows that for any subsequence such that $\lim_{i\to\infty} b_{N_i}$ exists, the limit is the same. However, any bounded sequence, all of whose convergent subsequences have the same limit, converges.[37] Consequently, we have proved that

$$\lim_{N\to\infty} \lim_{M\to\infty} \int \psi_M^*(f_{(N)}, g_{(N)})\lambda_M(f_{(N)}, g_{(N)}) \times d\mu(f_{(N)}, g_{(N)}) = (\Psi, \Lambda). \quad (3.21)$$

We now notice from (3.19) that any projection functional $\psi_M(f_{(N)}, g_{(N)})$ converges in the mean in $L^2(R_{\pi(N)} \times R_{\pi(N)})$ to a function

$$\bar{\psi}(f_{(N)}, g_{(N)}) \in L^2(R_{\pi(N)} \times R_{\pi(N)}),$$

and in addition, a subsequence $\{M_i\}$ exists such that

$$\lim_{i\to\infty} \psi_{M_i}(f_{(N)}, g_{(N)}) = \bar{\psi}(f_{(N)}, g_{(N)})$$

almost everywhere. However, Schwartz's inequality applied to (3.6) gives $|\psi_M(f_{(N)}, g_{(N)}) - \psi(f_{(N)}, g_{(N)})| \leq ||A_M\Psi - \Psi||$, which shows that

$$\lim_{M\to\infty} \psi_M(f_{(N)}, g_{(N)}) = \psi(f_{(N)}, g_{(N)})$$

everywhere in $R_{\pi(N)} \times R_{\pi(N)}$. Therefore, we can identify $\bar{\psi}(f_{(N)}, g_{(N)})$ and $\psi(f_{(N)}, g_{(N)})$, and in (3.21) we can then take the limit with respect to M under the integral sign. Hence for all ψ, $\lambda \in \mathfrak{C}$, and not merely for projection functionals,

$$\lim_{N\to\infty} \int \psi^*(f_{(N)}, g_{(N)})\lambda(f_{(N)}, g_{(N)}) \times d\mu(f_{(N)}, g_{(N)}) = (\Psi, \Lambda). \quad (3.22)$$

If $\psi(f, g)$, $\lambda(f, g) \in \mathfrak{C}$, we define their inner product to be

$$(\psi, \lambda)_c \equiv \lim_{N\to\infty} \int \psi^*(f_{(N)}, g_{(N)})\lambda(f_{(N)}, g_{(N)})\, d\mu\,(f_{(N)}, g_{(N)}). \quad (3.23)$$

(We have appended a "c" to the inner product in $\mathfrak{C}$ to differentiate it from the inner product in $\mathfrak{H}_X$.) Now by definition $\psi(f, g)$ and $\lambda(f, g)$ are the images respectively of Ψ, $\Lambda \in \mathfrak{H}_X$ under the mapping C defined by (3.2), and from (3.22) we have $(\psi, \lambda)_c = (\Psi, \Lambda)$. With this equality it is a simple matter to show that $(\ ,\ )_c$ has all the properties required of an inner product. Furthermore, the map C of all of $\mathfrak{H}_X$ onto $\mathfrak{C}$ is an isometry, so C is one-one, and $\mathfrak{C}$ being the isometric image of a complete inner product space is itself a complete inner product space, i.e., a Hilbert space. We summarize these results as

Theorem 3.1. Every $\psi(f, g)$ defined by (3.2) is a bounded, continuous function defined on the label space $\mathfrak{L} = \mathfrak{S} \times \mathfrak{S}$. The set $\mathfrak{C}$ of all such functions forms a Hilbert space whose inner product is defined by (3.23). The space $\mathfrak{C}$ is congruent to $\mathfrak{H}_X$ and is called the continuous representation of $\mathfrak{H}_X$.

For the sake of brevity in what follows, we introduce

Definition 3.1. Let

$$\int \psi^*(f, g)\lambda(f, g)\, d\mu\,(f, g) = \lim_{N\to\infty} \int \psi^*(f_{(N)}, g_{(N)})\lambda(f_{(N)}, g_{(N)})\, d\mu\,(f_{(N)}, g_{(N)}). \quad (3.24)$$

[37] K. Knopp, Ref. 33, p. 394.

C. Inverse Map

Let $\psi(f, g) = (\Phi[f, g], \Psi)$ for some fixed $\Psi \in \mathfrak{H}_X$. Consider the sequence of linear functionals defined for all $\Lambda \in \mathfrak{H}_X$ and each positive integer N by

$$\Psi_N(\Lambda) = \int \psi^*(f_{(N)}, g_{(N)}) \times (\Phi[f_{(N)}, g_{(N)}], \Lambda)\, d\mu\,(f_{(N)}, g_{(N)}). \quad (3.25)$$

If in inequality (3.18) we let $M \to \infty$, the resulting inequality implies that $|\Psi_N(\Lambda)| \leq ||\Lambda||\,||\Psi||$. Then by the Riesz representation theorem[31] there exists a unique vector Ψ_N such that $\Psi_N(\Lambda) = (\Psi_N, \Lambda)$. Therefore, we can write

$$\Psi_N = \int \psi^*(f_{(N)}, g_{(N)})\Phi[f_{(N)}, g_{(N)}]\, d\mu\,(f_{(N)}, g_{(N)}),$$

where this integral is a Pettis integral.[38] We have already proved that $\lim_{N\to\infty}(\Psi_N, \Lambda) = (\Psi, \Lambda)$ for all Λ [see (3.22)]. Therefore, Ψ is the weak limit of Ψ_N. We list this result as

Theorem 3.2. Let $\Psi \in \mathfrak{H}_X$, then

$$\Psi = \lim_{N\to\infty} \int (\Phi[f_{(N)}, g_{(N)}], \Psi) \times \Phi[f_{(N)}, g_{(N)}]\, d\mu\,(f_{(N)}, g_{(N)}), \quad (3.26)$$

where each integral is defined in the sense of Pettis, and the limit is taken in the weak sense.

This result enables us to give the inverse of the map C of $\mathfrak{H}_X$ onto $\mathfrak{C}$:

$$C : \Psi \to \psi(f, g) = (\Phi[f, g], \Psi),$$
$$C^{-1} : \psi(f, g) \to \Psi = \lim_{N\to\infty} \Psi_N.$$

Just as in Part IV, we can define for each positive integer N the operator on $\mathfrak{H}_X$

$$I_N = \int \Phi[f_{(N)}, g_{(N)}]\Phi[f_{(N)}, g_{(N)}]^\dagger\, d\mu\,(f_{(N)}, g_{(N)}) \quad (3.27)$$

which maps each $\Psi \in \mathfrak{H}_X$ onto

$$I_N\Psi = \int (\Phi[f_{(N)}, g_{(N)}], \Psi)\Phi[f_{(N)}, g_{(N)}]\, d\mu\,(f_{(N)}, g_{(N)}). \quad (3.28)$$

We have proved in Theorem 3.2 that

$$\lim (I_N\Psi, \Lambda) = (\Psi, \Lambda)$$

for all $\Psi, \Lambda \in \mathfrak{H}_X$. We list this result as

Lemma 3.1. The unit operator I in $\mathfrak{H}_X$ is the weak limit of the sequence of operators I_N defined by (3.27) and (3.28).

We can now assert that the set $\mathfrak{S}$ of unit vectors $\Phi[f, g] \equiv U[f, g]\Phi_0$ satisfies the three postulates (given in Part I) which define an overcomplete family of states. Postulates 1 and 2 simply assert that $\mathfrak{S}$ is arcwise connected (obvious) and that $\Phi[f, g]$ is a weakly continuous function on $\mathfrak{L}$ (Theorem 2.2, Corollary 2). The third postulate states that the set $\mathfrak{S}$ must span $\mathfrak{H}_X$ and that a resolution of the identity into an integral over projection operators exists. The set $\mathfrak{S}$ does span $\mathfrak{H}_X$, for if $(\Phi[f, g], \Psi) \equiv 0$, $\{f, g\} \in \mathfrak{L}$, then by (3.22), $\Psi = 0$. The resolution into projection operators is given in Lemma 3.1. We list this result as

Lemma 3.1. The family $\mathfrak{S}$ of unit vectors, $\Phi[f, g] \equiv U[f, g]\Phi_0$, where $\{f, g\} \in \mathfrak{L} = \mathfrak{s} \times \mathfrak{s}$, and $U[f, g] = V[g]\,W[f]$ is constructed from the irreducible representation of the CCR determined by the product reference vector $X \in \mathfrak{R}$, forms an OFS for any fiducial vector $\Phi_0 \in \mathfrak{H}_X$.

D. Reproducing Kernels and Aronszajn Spaces

Reproducing Kernel

Let $\mathfrak{C}$ be a continuous representation of $\mathfrak{H}_X$ corresponding to the fiducial vector Φ_0. We associate with $\mathfrak{C}$ a functional $\mathcal{K}(f', g'; f, g)$, called the reproducing kernel, and defined by

$$\mathcal{K}(f', g'; f, g) = (U[f', g']\Phi_0, U[f, g]\Phi_0). \quad (3.29)$$

For fixed $\{f, g\} \in \mathfrak{L}$, $\mathcal{K}(f', g'; f, g)$ is an element of $\mathfrak{C}$ when considered as a function of $\{f', g'\}$, as follows directly from (3.29). If

$$\psi(f', g') = (U[f', g']\Phi_0, \Psi) \in \mathfrak{C},$$

then for every $\{f, g\} \in \mathfrak{L}$, it follows from Theorem 3.1 and Eq. (3.22) that

$$(\mathcal{K}(f', g'; f, g), \psi(f', g')) = \int \mathcal{K}(f, g; f', g')\psi(f', g')\, d\mu\,(f', g') = (U[f, g]\Phi_0, \Psi) = \psi(f, g), \quad (3.30)$$

a relation which gives to $\mathcal{K}$ its name. The same relation yields the idempotent property

$$\int \mathcal{K}(f', g'; f'', g'')\mathcal{K}(f'', g''; f, g)\, d\mu\,(f'', g'') = \mathcal{K}(f', g'; f, g). \quad (3.31)$$

In addition, the reproducing kernel has a number

[38] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups* (American Mathematical Society, Providence, 1957), pp. 76–78.

of special properties and we list several of the more important.

(1) If α_k, $k = 1, 2, \cdots, M$, is an arbitrary, finite set of M complex numbers and $\{f_k, g_k\}$ are M arbitrary points in $\mathcal{S} \times \mathcal{S}$, then

$$\sum_{j=1}^{M} \sum_{k=1}^{M} \alpha_j^* \alpha_k \mathcal{K}(f_k, g_j; f_k, g_k) \geq 0. \tag{3.32}$$

The sum in (3.32) is just

$$\left\| \sum_{j=1}^{M} \alpha_j U[f_j, g_j] \Phi_0 \right\|^2,$$

which is obviously nonnegative.

(2) From the definition of $\mathcal{K}$ we obtain

$$\mathcal{K}(f, g; f, g) \equiv 1. \tag{3.33}$$

(3) The kernel $\mathcal{K}(f', g'; f, g)$ is a continuous function of its variables. It is continuous in $\{f, g\}$ uniformly with respect to $\{f', g'\}$, and vice versa. This is a consequence of the strong continuity of the vector-valued function $\Phi[f, g]$, proved in Theorem 2.2, Corollary 2.

(4) As a consequence of the definition (3.29) and the CCR, $\mathcal{K}$ satisfies the relation

$$\mathcal{K}(f', g'; f, g) = e^{i(f, g' - g)} \mathcal{K}(f' - f, g' - g; 0, 0). \tag{3.34}$$

Continuous Representations and Aronszajn Spaces

The reproducing kernel is important because it very possibly completely determines the corresponding continuous representation. A general theory of Hilbert spaces of functions which possess a reproducing kernel has been developed by Aronszajn,[39] and this theory can be applied to our spaces $\mathfrak{S}$ of functionals on $\mathcal{S} \times \mathcal{S}$.

Let us assume that we are given a function $\mathcal{K}(f', g'; f, g)(f, g, f', g' \in \mathcal{S})$ which satisfies Eq. (3.31) and Conditions 1 to 4. We proceed to show how far we can go in reconstructing $\mathfrak{S}$ from $\mathcal{K}$. Aronszajn has shown that starting with a function $\mathcal{K}$ that satisfies only Condition 1 [Eq. (3.32)], it is possible to construct a Hilbert space of functions on $\mathcal{L} = \mathcal{S} \times \mathcal{S}$ for which $\mathcal{K}$ is the reproducing kernel. We denote this space by $\mathfrak{A}$. This construction, which is independent of the nature of the topological space on which the functions are defined, has been outlined in Sec. IV 5, and is not repeated here. Aronszajn's construction, starting with a kernel function $\mathcal{K}$ which is known only to the extent of satisfying Condition 1, does yield a Hilbert space of functions, but specific properties of the functions comprising this space as well as a more explicit form for its inner product cannot be determined unless further properties of the kernel function are known.

In our case, the fact that $\mathcal{K}$ satisfies Conditions 2 and 3 implies that all the functions of $\mathfrak{S}$ are continuous and bounded. Because $\mathcal{K}$ satisfies Condition 4 in addition, we can very readily construct a representation of the CCR in $\mathfrak{A}$, for which the operators $V[g]$ and $W[f]$ are defined by

$$(V[g]\psi)(f', g') = \psi(f', g' - g), \tag{3.35}$$

$$(W[f]\psi)(f', g') = e^{i(f, g')} \psi(f' - f, g'). \tag{3.36}$$

Furthermore, if we set $\varphi_0(f, g) = \mathcal{K}(f, g; 0, 0)$ then we have

$$\mathcal{K}(f', g'; f, g) = (U[f', g']\varphi_0(f'', g''), U[f, g]\varphi_0(f'', g''))_a, \tag{3.37}$$

where, as usual, $U[f, g] \equiv V[g]W[f]$, and $(\ ,\ )_a$ denotes the inner product in $\mathfrak{A}$. The irreducibility of this representation has not been established.

The proofs of the foregoing statements are almost identical with the proofs of the corresponding statements made about Aronszajn spaces in Sec. IV 5, and are not given again here. It can further be shown, since $\mathcal{K}$ satisfies Eq. (3.31), that the inner product in $\mathfrak{A}$ is also given by

$$(\psi, \lambda)_a = \lim_{N \to \infty} \int \psi^*(f_{(N)}, g_{(N)}) \times \lambda(f_{(N)}, g_{(N)})\, d\mu\,(f_{(N)}, g_{(N)}),$$

where $\psi(f, g), \lambda(f, g) \in \mathfrak{A}$. The proof of this fact is relatively straightforward, but it will not be given here because it depends on the details of Aronszajn's construction.

The final step in reconstructing $\mathfrak{S}$ from a kernel function $\mathcal{K}$, satisfying Eq. (3.31) and Condition 1 to 4 is to show that the representation of the CCR defined in (3.35), (3.36) is irreducible. In analogy with the results of Part IV, we conjecture that because $\mathcal{K}$ satisfies Eq. (3.31) the representation (3.35), (3.36) is irreducible, but we have not yet proved this.

E. A Space of "Analytic" Functionals

In order to show the connection between our work and that of Bargmann[22] and Segal[21] we consider the following example of a continuous representation. Set $\mathfrak{H}_n = L^2(R)$ for each $n \in \Delta$, and for each $n \in \Delta$, choose the basis set $\Psi_{n, N_n} \equiv h_{N_n}(x)$, $N_n = 0, 1, \cdots$, where $h_{N_n}(x)$ is the N_nth Hermite function. Let the

[39] N. Aronszajn, Proc. Cambridge Phil. Soc. **39**, 133 (1943); Trans. Am. Math. Soc. **68**, 337 (1950).

product reference vector X which determines the representation of the CCR be $X = \prod_{n\in\Delta} \otimes \Psi_{n,0}$, and set $\Phi_0 = X$, where Φ_0 is the fiducial vector determining the continuous representation of $\mathfrak{H}_X$. Furthermore, let Γ be the countable collection of all sequences, $N = \{N_n : n \in \Delta\}$, of nonnegative integers such that $N_n \neq 0$ for at most a finite number of values of n. Then the countable set of vectors $\Psi_N = \prod_{n\in\Delta} \otimes \Psi_{n,N_n}$, $N \in \Gamma$ forms an orthonormal basis for $\mathfrak{H}_X$,[16] and hence their images $\psi_N(f, g) = (U[f, g]\Phi_0, \Psi_N)$ form an orthonormal basis for the continuous representation $\mathfrak{C}_0$ of $\mathfrak{H}_X$. The explicit expression for $\psi_N(f, g)$ can be easily computed:

$$\psi_N(f, g) = \left\{\prod_{n\in\Delta} \frac{1}{(N_n!)^{\frac{1}{2}}} \left(\frac{q_n - ip_n}{\sqrt{2}}\right)^{N_n}\right\} \times \exp\left[-\sum_{n\in\Delta} (\tfrac{1}{4}q_n^2 + \tfrac{1}{4}p_n^2 - \tfrac{1}{2}ip_nq_n)\right]. \quad (3.38)$$

It should be noted that the product in (3.38) is a finite product, since $N_n = 0$ for all but a finite number of values of n, and the exponential factor is the same for each $\psi_N(f, g)$. It is also clear from (3.38) that the domain of definition of each $\psi_N(f, g)$ can be extended from $\mathcal{S} \times \mathcal{S}$ to $L^2(R_3) \times L^2(R_3)$. If we write $z_n = (1/\sqrt{2})(q_n - ip_n)$ and

$$u_{[N]}(z) = \prod_{n\in\Delta} (N_n!)^{-\frac{1}{2}} z_n^{N_n},$$

then an arbitrary $\psi(f, g) \in \mathfrak{C}_0$ has the expansion

$$\psi(f, g) = \{\sum_{N\in\Gamma} a_N u_{[N]}(z)\} \times \exp\left[-\sum_{n\in\Delta} (\tfrac{1}{4}q_n^2 + \tfrac{1}{4}p_n^2 - \tfrac{1}{2}ip_nq_n)\right], \quad (3.39)$$

where $\sum_{N\in\Gamma} |a_N|^2 < \infty$. However, the functions $\sum_{N\in\Gamma} a_N u_{[N]}(z) = f(z)$ are precisely the "analytic" functions comprising the space which Bargmann denotes $\mathfrak{F}_\infty$.[22] Thus each $\psi(f, g) \in \mathfrak{C}_0$ is equal to a function $f(z) \in \mathfrak{F}_\infty$ multiplied by the common factor $\exp\{-\sum_{n\in\Delta} (\frac{1}{4}q_n^2 + \frac{1}{4}p_n^2 - \frac{1}{2}ip_nq_n)\}$, and the absolute square of this latter function can be identified with the Gaussian weighting factor used by Bargmann and Segal in forming the inner product. The relationship (3.39) defines a one-to-one, isometric mapping of $\mathfrak{C}_0$ onto $\mathfrak{F}_\infty$.

F. Generalizations to Several Fields

The generalization of the CCR discussed in Sec. 2E can also be used to construct continuous representations. Let $X \in \mathfrak{R}_K$ be a product reference vector, let $\mathfrak{H}_X$ be the IDPS determined by X, and let $V[\mathbf{g}]$ and $W[\mathbf{f}]$ be the representation of the CCR determined by X, where $\{\mathbf{f}, \mathbf{g}\} \in \mathcal{S}_K \times \mathcal{S}_K$. If $\Phi_0 \in \mathfrak{H}_X$, $||\Phi_0|| = 1$, is the fiducial vector, and $\Phi[\mathbf{f}, \mathbf{g}] \equiv V[\mathbf{g}]W[\mathbf{f}]\Phi_0$, then the elements of $\mathfrak{C}_K$ are the bounded, continuous functionals

$$\psi(\mathbf{f}, \mathbf{g}) = (\Phi[\mathbf{f}, \mathbf{g}], \Psi), \quad (3.40)$$

for all $\Psi \in \mathfrak{H}_X$. An inner product can be introduced into $\mathfrak{C}_K$ by

$$(\psi, \lambda)_c = \lim_{N\to\infty} \int \psi^*(\mathbf{f}_{(N)}, \mathbf{g}_{(N)}) \times \lambda(\mathbf{f}_{(N)}, \mathbf{g}_{(N)})\, d\mu\,(\mathbf{f}_{(N)}, \mathbf{g}_{(N)}), \quad (3.41a)$$

where

$$d\mu\,(\mathbf{f}_{(N)}, \mathbf{g}_{(N)}) = \prod_{k=1}^{K} \prod_{|n|\leq N} \left(\frac{1}{2\pi}\right) dp_{kn}\, dq_{kn}. \quad (3.41b)$$

With respect to this inner product, $\mathfrak{C}_K$ is a complete Hilbert space. The family of unitary operators

$$(V[\mathbf{g}]\psi)(\mathbf{f}', \mathbf{g}') = \psi(\mathbf{f}', \mathbf{g}' - \mathbf{g}), \quad (3.42a)$$

$$(W[\mathbf{f}]\psi)(\mathbf{f}', \mathbf{g}') = e^{i(\mathbf{f};\mathbf{g}')}\psi(\mathbf{f}' - \mathbf{f}, \mathbf{g}) \quad (3.42b)$$

form an irreducible set of transformations on $\mathfrak{C}_K$ which fulfill the CCR Eq. (2.48).

4. KERNEL AND FUNCTIONAL DERIVATIVE REPRESENTATIONS FOR OPERATORS; INEQUIVALENCE OF CCR REPRESENTATIONS

A. Kernel Representation of Bounded Linear Operators on $\mathfrak{C}$

Just as in the case of a finite number of degrees of freedom (cf. Sec. IV 6), each bounded linear operator on $\mathfrak{C}$ can be represented by a kernel. Since $\mathfrak{C}$ is the isometric image of $\mathfrak{H}_X$, each operator $\mathcal{B}_c$ on $\mathfrak{C}$ is the image of an operator $\mathcal{B}$ on $\mathfrak{H}_X$. Given the bounded, linear operator $\mathcal{B}$ on $\mathfrak{H}_X$, define

$$\mathcal{B}(f', g'; f, g) = (U[f', g']\Phi_0, \mathcal{B}U[f, g]\Phi_0). \quad (4.1)$$

For each fixed $\{f, g\} \in \mathcal{S} \times \mathcal{S}$, $\mathcal{B}(f', g'; f, g) \in \mathfrak{C}$. Then if $\psi(f, g) = (U[f, g]\Phi_0, \Psi)$ is the image in $\mathfrak{C}$ of $\Psi \in \mathfrak{H}_X$, we can use Eq. (3.22) to show that

$$(\mathcal{B}\psi)(f, g) = \lim_{N\to\infty} \int \mathcal{B}(f, g; f'_{(N)}, g'_{(N)}) \times \psi(f'_{(N)}, g'_{(N)})\, d\mu\,(f'_{(N)}, g'_{(N)}) = (U[f, g]\Phi_0, \mathcal{B}\Psi). \quad (4.2)$$

Thus $(\mathcal{B}\psi)(f, g)$ is the image in $\mathfrak{C}$ of the vector $\mathcal{B}\Psi$ in $\mathfrak{H}_X$, so Eq. (4.2) correctly describes the action of $\mathcal{B}_c$ in $\mathfrak{C}$. We omit any further discussion of kernel functions describing bounded operators on $\mathfrak{C}$, since the corresponding discussion given in Part IV for the case of a finite number of degrees of freedom applies here with few changes.

B. Representation of Field Operators by Functional Derivatives

We now discuss functional derivatives and their relationship to representation of the field operators $\varphi(f)$ and $\pi(g)$. We assume for the rest of the discussion in Sec. 4B that the fiducial vector Φ_0 which determines $\mathfrak{C}$ is in the domain of all $\varphi(f)$ and $\pi(g)$, i.e.,

$$\Phi_0 \in (\bigcap_{f\in\mathcal{S}} \mathfrak{D}_{\varphi(f)}) \cap (\bigcap_{g\in\mathcal{S}} \mathfrak{D}_{\pi(g)}). \tag{4.3}$$

We indicated in Sec. 2D that in fact such vectors Φ_0 exist. We now define two directional derivatives for each $\psi(f, g) \in \mathfrak{C}$. Let $e \in \mathcal{S}$, then

$$(e, \delta_f)\psi(f, g) = \lim_{t\to 0} (1/t)[\psi(f + te, g) - \psi(f, g)], \tag{4.4}$$

$$(e, \delta_g)\psi(f, g) = \lim_{t\to 0} (1/t)[\psi(f, g + te) - \psi(f, g)]. \tag{4.5}$$

The limits on the right of (4.3) and (4.4) exist for all $\psi \in \mathfrak{C}$ if the fiducial vector Φ_0 defining $\mathfrak{C}$ satisfies (4.3). Using the definition of $\psi(f, g)$, we can write

$$(1/t)\{\psi(f + te, g) - \psi(f, g)\} = (V[g]W[f](1/t)(W[te] - I)\Phi_0, \Psi).$$

In Sec. 2D we showed that $i\varphi(e)$ is the strong limit of $(1/t)(W[te] - I)$, and by hypothesis $\Phi_0 \in \mathfrak{D}_{\varphi(e)}$. Therefore, we can rewrite (4.4) as

$$(e, \delta_f)\psi(f, g) = -i(U[f, g]\varphi(e)\Phi_0, \Psi). \tag{4.6}$$

The limit on the right of (4.5) can be evaluated in the same fashion, using the fact that $\Phi_0 \in \mathfrak{D}_{\varphi(e)} \Rightarrow U[f, g]\Phi_0 \in \mathfrak{D}_{\varphi(e)}$ for all $\{f, g\}$, as shown in Sec. 2D. The result is

$$(e, \delta_g)\psi(f, g) = i(\pi(e)U[f, g]\Phi_0, \Psi). \tag{4.7}$$

Equations (4.6) and (4.7) can both be written in a slightly different form by commuting $\varphi(e)$ or $\pi(g)$ with $U[f, g]$ with the aid of (2.42) and (2.43). In particular, we note the two formulas

$$\varphi_e(f, g; f', g') \equiv (U[f, g]\Phi_0, \varphi(e)U[f', g']\Phi_0) = [i(e, \delta_f) + (e, g)]\mathcal{K}(f, g; f', g'), \tag{4.8}$$

$$\pi_e(f, g; f', g') \equiv (U[f, g]\Phi_0, \pi(e)U[f', g']\Phi_0) = -i(e, \delta_g)\mathcal{K}(f, g; f', g'). \tag{4.9}$$

Since $U[f, g]\Phi_0 \in \mathfrak{D}_{\varphi(e)} \cap \mathfrak{D}_{\pi(e)}$ for all $\{f, g\}$, it is not difficult to show that for all $\Psi \in \mathfrak{D}_{\varphi(e)}$,

$$(\varphi(e)\psi)(f, g) = \lim_{N\to\infty} \int \varphi_e(f, g; f'_{(N)}, g'_{(N)}) \times \psi(f'_{(N)}, g'_{(N)})\, d\mu\, (f'_{(N)}, g'_{(N)}), \tag{4.10}$$

and for all $\Psi \in \mathfrak{D}_{\pi(e)}$,

$$(\pi(e)\psi)(f, g) = \lim_{N\to\infty} \int \pi_e(f, g; f'_{(N)}, g'_{(N)}) \times \psi(f'_{(N)}, g'_{(N)})\, d\mu\, (f'_{(N)}, g'_{(N)}). \tag{4.11}$$

The proof of (4.10) and (4.11) is a straightforward adaption of the discussion given in Sec. IV 6 of the equivalent equations for a finite number of degrees of freedom, and is omitted here.

An alternative characterization of (4.10) and (4.11) may also be won out of (4.6) and (4.7). In particular, for $\Psi \in \mathfrak{D}_{\varphi(e)}$ we have

$$(\varphi(e)\psi)(f, g) = [i(e, \delta_f) + (e, g)]\psi(f, g) \tag{4.12}$$

and for $\Psi \in \mathfrak{D}_{\pi(e)}$ we have

$$(\pi(e)\psi)(f, g) = -i(e, \delta_g)\psi(f, g). \tag{4.13}$$

Thus the kernels for $\varphi(e)$ and $\pi(e)$, as well as their representation on any vector in their domain can be determined by functional differentiation. It should be noted that Eqs. (4.12) and (4.13) have the same functional form independent of $\mathfrak{C}$.

C. Some Criteria for the Unitary Equivalence and Inequivalence of Irreducible Representations of the CCR

As a final application of continuous-representation theory, we employ it as a tool in a discussion of the unitary equivalence and inequivalence of our representations.

We have already seen that all the product reference vectors in $\mathfrak{R}$ can be grouped into equivalence classes, (Definition 2.2) and that the representations determined by two equivalent vectors are identical. It is useful to introduce another equivalence relation into $\mathfrak{H}$ (and hence into $\mathfrak{R}$) as follows.[16]

Definition 4.1. Two c_0 vectors are said to be weakly equivalent, $\prod_{n\in\Delta} \otimes \chi_n \approx \prod_{n\in\Delta} \otimes \lambda_n$, if and only if there is a set of real numbers $\{\theta_n : n \in \Delta\}$ so that $\prod_{n\in\Delta} \otimes \chi_n \approx \prod_{n\in\Delta} \otimes e^{i\theta_n}\lambda_n$.

A necessary and sufficient condition for weak equivalence is that $\sum_{n\in\Delta} |\,|(\chi_n, \lambda_n)| - 1| < \infty$. A weak equivalence class is the union of a collection of complete equivalence classes.

We now show that two representations determined by weakly equivalent product reference vectors are unitarily equivalent. Let $X = \prod_{n\in\Delta} \otimes \chi_n$, $\Lambda = \prod_{n\in\Delta} \otimes \lambda_n$, and assume $X, \Lambda \in \mathfrak{R}$ and $X \approx_w \Lambda$. Then there exists a set of real numbers $\{\theta_n : n \in \Delta\}$ such that if $\Lambda' = \prod_{n\in\Delta} \otimes e^{i\theta_n}\lambda_n$, then $X \approx \Lambda'$. It is easy to see that $\Lambda' \in \mathfrak{R}$, and therefore, the repre-

sentations determined by X and Λ' are identical. Thus we need only demonstrate that the representations determined by Λ and Λ' are unitarily equivalent. However, von Neumann has shown that there exists a unitary map, $T(\theta)$, of $\mathfrak{H}_\Lambda$ onto $\mathfrak{H}_{\Lambda'}$ which maps each product vector $\Phi = \prod_{n\in\Delta} \otimes \varphi_n \in \mathfrak{H}_\Lambda$ onto $T(\theta)\Phi = \prod_{n\in\Delta} \otimes e^{i\theta_n}\varphi_n$.[16] If $V_\Lambda[g]$ and $W_\Lambda[f]$, and $V_{\Lambda'}[g]$ and $W_{\Lambda'}[f]$ are the representations determined by Λ and Λ' respectively, then with the aid of the defining equations (2.11), it is easy to see that

$$V_{\Lambda'}[g] = T(\theta)V_\Lambda[g]T(\theta)^{-1},$$
$$W_{\Lambda'}[f] = T(\theta)W_\Lambda[f]T(\theta)^{-1}.$$

Unitary Invariant Tags

Although we have not yet been able to determine all those among our representations which are unitarily equivalent, we now discuss a technique which is frequently useful in showing the unitary *in*equivalence of two representations of the CCR.

Note added in proof: In collaboration with E. J. Woods, we have succeeded in proving, with the aid of the tag philosophy, that any two representations of this CCR defined in the canonical fashion of this paper on two IDPS's with weakly inequivalent fiducial vectors are unitarily inequivalent representations. A report on this work is in preparation.

Let a representation determined by $X \in \mathfrak{R}$ be given and form the family of unitary operators $U[f, g] = V[g]W[f]$. Then using the commutation relations, we can write

$$U[f_k, g_k]U[f, g] = e^{i[(f_k,g)-(g_k,f)]}U[f, g]U[f_k, g_k]. \tag{4.14}$$

Now suppose it is possible to pick a sequence $\{f_k, g_k\} \in \mathcal{S} \times \mathcal{S}$, $k = 1, 2, \cdots$, which satisfies the two conditions:

(1) $\lim_{k\to\infty} (f_k, g) = \lim_{k\to\infty} (g_k, f) = 0,$ for all $f, g \in \mathcal{S}$,

(2) $\lim_{k\to\infty} (\Psi, U[f_k, g_k]\Phi) = (\Psi, A\Phi),$ for all $\Psi, \Phi \in \mathfrak{H}_X$,

where A is a bounded, linear operator on $\mathfrak{H}_X$. Then taking the limit of both sides of (4.14), we see that

$$AU[f, g] = U[f, g]A. \tag{4.15}$$

Since (4.15) holds for all $\{f, g\} \in \mathcal{S} \times \mathcal{S}$, and $U[f, g]$ is an irreducible set of operators, we can conclude from Schur's theorem[40] that $A = aI$, where a is a complex number depending on the sequence $\{f_k, g_k\}$ and I is the unit operator. It is clear for the representation $U'[f, g] \equiv VU[f, g]V^{-1}$, which is unitarily equivalent to $U[f, g]$, that the weak limit of $U'[f_k, g_k]$ is $VaIV^{-1} = aI$. In other words, a is a *unitary invariant of the representation* of the CCR.

We can now draw the following conclusion: Let $U^{(1)}[f, g]$ and $U^{(2)}[f, g]$ be determined by $X^{(1)} \in \mathfrak{R}$ and $X^{(2)} \in \mathfrak{R}$ respectively. Let $\{f_k, g_k\} \in \mathcal{S} \times \mathcal{S}$, $k = 1, 2, \cdots$ be any sequence satisfying condition 1 above, and suppose further that $A^{(1)} = a^{(1)}I$ is the weak limit of $U^{(1)}[f_k, g_k]$. Then either if the weak limit of $U^{(2)}[f_k, g_k]$ does not exist, or if it does exist and $A^{(2)} = a^{(2)}I$ where $a^{(1)} \neq a^{(2)}$, then the two representations must be *inequivalent.*

As an elementary application of this procedure, consider the special case where $X^{(i)} = \prod_{n\in\Delta} \otimes \chi_n^{(i)} \in \mathfrak{R}$, $i = 1, 2$, have the property that there exist unit vectors $\chi_\infty^{(i)} \in L^2(R)$, such that

$$\lim_{|n|\to\infty} ||\chi_n^{(i)} - \chi_\infty^{(i)}|| = 0, \qquad i = 1, 2.$$

Let us adopt $f_k = ph_{n_k}$, $g_k = qh_{n_k}$, $k = 1, 2, \cdots$, where p and q are arbitrary real numbers, n_k, $k = 1, 2, \cdots$ is any subsequence of Δ such that $\lim_{k\to\infty} |n_k| = \infty$, and h_n, $n \in \Delta$ are the usual Hermite function basis elements of $\mathcal{S}$. It is obvious that the sequence $\{f_k, g_k\}$ satisfies condition 1 above, and it is a straightforward matter to show that $U^{(i)}[f_k, g_k]$ converges weakly and that $a^{(i)}(p, q) = (\chi_\infty^{(i)}, V_0[q]W_0[p]\chi_\infty^{(i)})$, $i = 1, 2$. Thus in order that the two respresentations be unitarily equivalent, it is necessary that $a^{(1)}(p, q) = a^{(2)}(p, q)$ *for all* p, q. With the aid of Eq. (2.9) we can write this last equality as

$$\int_{-\infty}^{\infty} \chi_\infty^{(1)}(y)^* e^{ip(y-q)}\chi_\infty^{(1)}(y - q)\, dy = \int_{-\infty}^{\infty} \chi_\infty^{(2)}(y)^* e^{ip(y-q)}\chi_\infty^{(2)}(y - q)\, dy. \tag{4.16}$$

In Sec. IV 3 we showed that both sides of (4.16) possess Fourier transforms, which yields the equality

$$\chi_\infty^{(1)}(y)^*\chi_\infty^{(1)}(y - q) = \chi_\infty^{(2)}(y)^*\chi_\infty^{(2)}(y - q).$$

This equality can hold for almost all y and q if and only if $\chi_\infty^{(1)}(y) = e^{i\theta}\chi_\infty^{(2)}(y)$ for some real θ. Therefore we can conclude that if $\chi_\infty^{(1)} \neq e^{i\theta}\chi_\infty^{(2)}$ for any real θ, then the representations of the CCR determined by $X^{(1)}$ and $X^{(2)}$ are unitarily inequivalent.

In particular, consider the very special class of product reference vectors of the form $X = \prod_{n\in\Delta} \otimes \chi_n$, $\chi_n \equiv \chi_\infty$, $n \in \Delta$. In this case we can speak of the representation of the CCR determined by χ_∞, and

[40] M. A. Naimark, *Normed Rings* (P. Noordhoff Ltd., Groningen, The Netherlands, 1959), p. 255.

we have just shown that as long as χ_∞ and χ'_∞ are not colinear ($\chi_\infty \neq e^{i\theta}\chi'_\infty$) then the representations they determine are unitarily inequivalent. Clearly, there are an uncountable infinity of unit vectors $\chi_\infty \in L^2$, no two of which are colinear, from which product reference vectors can be formed. Thus the product reference vectors in this restricted class alone determine an uncountable infinity of inequivalent, irreducible representations of the CCR.

It is clear that all the unitary invariant "tags" discussed here can be obtained within the continuous representation $\mathfrak{C}$ by taking appropriate limits of the reproducing kernel. If f_k, g_k are two weakly convergent test function sequences, then it may be shown that condition 2 above for the existence of A is satisfied if and only if

$$\lim_{k\to\infty} \mathcal{K}(f, g; f_k, g_k)$$

exists for all $\{f, g\} \in \mathcal{S} \times \mathcal{S}$. It then follows that the unitary invariant

$$a = \lim_{k\to\infty} (\Phi_0, U[f_k, g_k]\Phi_0) = \lim_{k\to\infty} \mathcal{K}(0, 0; f_k, g_k).$$

ACKNOWLEDGMENTS

The authors have benefited from several conversations with Dr. E. J. Woods regarding the canonical commutation relations and their representations. The generous encouragement of this work by Professor V. Bargmann and Professor A. S. Wightman is warmly appreciated.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

The Structure of Space and the Formalism of Relativistic Quantum Theory. I

YRJÖ AHMAVAARA
Institute for Mathematical Physics, University of Turku, Turku, Finland
(Received 17 March 1964; final manuscript received 21 July 1964)

The structure of a Euclidean space can be approached, with an unlimited accuracy, by a part of a maximally ordered finite linear space. Accordingly, all the physical theories based on the space–time continuum can also be considered in such a finite space–time. The finiteness of the underlying space makes also some new kinds of theories possible. Among them is a purely group theoretical formalism of relativistic quantum theory, including a free-particle theory as well as a group formalism of interaction of particles. The free-particle formalism of a finite space–time is considered here (Part I). An essential difference in comparison with the formalism of continuous space–time is that there is, as a consequence of the relations of Euclidicity to be imposed on observable 4-vectors, a nontrivial spectrum of momentum, mass, and energy in a finite geometry.

I. INTRODUCTORY REMARKS

THERE are reasons[1] to believe that distances larger than about 4×10^{27} cm from the earth are inobservable in principle. One can refer, for instance, to the fact that the red shift in the spectra of the galaxies at that distance already equals the Doppler effect of the velocity of light. On the other hand, one has considered in atomic physics the possibility of the existence of an elementary length d, the smallest observable length. Different orders of magnitude of d have been suggested, all of them obeying $d \leq 10^{-18}$ cm. If a smallest observable length exists, the observable points of space (and time) form a finite set.

Irrespective of the fact whether there are any largest or smallest observable lengths, one can consider the problem whether the physical space–time can be mathematically described by a finite space. The solution of the problem depends on the solution of another problem, viz. the following: Is it possible to approach the structure of a Euclidean space, with an unlimited accuracy, by a finite space? The answer to the latter question is in the affirmative[2] even though the results in question are little known outside the circle of the mathematicians studying finite geometries. These results are reported in Sec. II below.

The significance of the results concerning the approximation of a Euclidean geometry by a finite geometry is in the fact that it makes of finite geom-

[1] For instance, G. Järnefelt, Ann. Acad. Sci. Fennicae Ser. A I, No. 96 (1951).

[2] P. Kustaanheimo, Soc. Sci. Fennica 15, No. 19 (1950).

we have just shown that as long as χ_∞ and χ'_∞ are not colinear ($\chi_\infty \neq e^{i\theta}\chi'_\infty$) then the representations they determine are unitarily inequivalent. Clearly, there are an uncountable infinity of unit vectors $\chi_\infty \in L^2$, no two of which are colinear, from which product reference vectors can be formed. Thus the product reference vectors in this restricted class alone determine an uncountable infinity of inequivalent, irreducible representations of the CCR.

It is clear that all the unitary invariant "tags" discussed here can be obtained within the continuous representation $\mathfrak{C}$ by taking appropriate limits of the reproducing kernel. If f_k, g_k are two weakly convergent test function sequences, then it may be shown that condition 2 above for the existence of A is satisfied if and only if

$$\lim_{k\to\infty} \mathcal{K}(f, g; f_k, g_k)$$

exists for all $\{f, g\} \in \mathcal{S} \times \mathcal{S}$. It then follows that the unitary invariant

$$a = \lim_{k\to\infty} (\Phi_0, U[f_k, g_k]\Phi_0) = \lim_{k\to\infty} \mathcal{K}(0, 0; f_k, g_k).$$

ACKNOWLEDGMENTS

The authors have benefited from several conversations with Dr. E. J. Woods regarding the canonical commutation relations and their representations. The generous encouragement of this work by Professor V. Bargmann and Professor A. S. Wightman is warmly appreciated.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

The Structure of Space and the Formalism of Relativistic Quantum Theory. I

YRJÖ AHMAVAARA
Institute for Mathematical Physics, University of Turku, Turku, Finland
(Received 17 March 1964; final manuscript received 21 July 1964)

The structure of a Euclidean space can be approached, with an unlimited accuracy, by a part of a maximally ordered finite linear space. Accordingly, all the physical theories based on the space–time continuum can also be considered in such a finite space–time. The finiteness of the underlying space makes also some new kinds of theories possible. Among them is a purely group theoretical formalism of relativistic quantum theory, including a free-particle theory as well as a group formalism of interaction of particles. The free-particle formalism of a finite space–time is considered here (Part I). An essential difference in comparison with the formalism of continuous space–time is that there is, as a consequence of the relations of Euclidicity to be imposed on observable 4-vectors, a nontrivial spectrum of momentum, mass, and energy in a finite geometry.

I. INTRODUCTORY REMARKS

THERE are reasons[1] to believe that distances larger than about 4×10^{27} cm from the earth are inobservable in principle. One can refer, for instance, to the fact that the red shift in the spectra of the galaxies at that distance already equals the Doppler effect of the velocity of light. On the other hand, one has considered in atomic physics the possibility of the existence of an elementary length d, the smallest observable length. Different orders of magnitude of d have been suggested, all of them obeying $d \leq 10^{-18}$ cm. If a smallest observable length exists, the observable points of space (and time) form a finite set.

Irrespective of the fact whether there are any largest or smallest observable lengths, one can consider the problem whether the physical space–time can be mathematically described by a finite space. The solution of the problem depends on the solution of another problem, viz. the following: Is it possible to approach the structure of a Euclidean space, with an unlimited accuracy, by a finite space? The answer to the latter question is in the affirmative[2] even though the results in question are little known outside the circle of the mathematicians studying finite geometries. These results are reported in Sec. II below.

The significance of the results concerning the approximation of a Euclidean geometry by a finite geometry is in the fact that it makes of finite geom-

[1] For instance, G. Järnefelt, Ann. Acad. Sci. Fennicae Ser. A I, No. 96 (1951).

[2] P. Kustaanheimo, Soc. Sci. Fennica 15, No. 19 (1950).

etry a "realistic model" in physics. All the existing physical theory can be now approximated, with an unlimited accuracy, by theory in a finite space. More interesting than just a reformulation of old theories is, of course, to study whether the use of finite space can open new possibilities for physical theory.

A purely group theoretical formalism of relativistic quantum theory, possible in a finite space–time, will be studied in this and two subsequent papers (to be called Part I, Part II, and Part III, respectively). "Purely group theoretical" means that also the interactions of particles are described in this formalism by means of an invariance and transformation theory, based on geometry and covariance only. A well-known earlier formalism of this type is the formalism of general relativity where the gravitational interactions are described by equations based on geometry and covariance only. As in general relativity, the laws of interaction (there the equations for the gravitational field, here the S operators predicting the interactions and composed of the creation and the annihilation operators of the particles in question) are constructed by means of tensorial contractions. The covariance of the laws of interaction with respect to the relativity group is, together with the geometry of space–time, the only assumption used for the construction of the S matrices in this formalism (see Part II for details of construction).

Such a formalism may or may not have use in physical theory. Here the formalism is investigated as a mathematical model, as a logical possibility connected with a finite space–time. The existence of this formalism was predicted in an earlier paper.[8]

The group formalism of interaction will be considered in Part II. An essential point of the interaction formalism is based, however, on the existence of nontrivial spectrum of observable momentum, mass, and energy in finite geometry; the existence of this spectrum is obtained as a result of the consideration of the free-particle formalism of finite space–time which is given in the present paper (Part I).

II. THE FINE STRUCTURE OF SPACE-TIME

Irrespective of the fact whether any largest or smallest observable lengths exist, one can approximate the observable points of space–time, with an unlimited accuracy, by the points of a cubic lattice E contained in a four-dimensional Euclidean space R^4. The side length of an elementary cube must *at least* be smaller than 10^{-13} cm in the spatial dimensions, and smaller than $(10^{-13}\text{ cm})/c = 10^{-24}$ sec in the dimension of time, but it can of course be made also much smaller than these upper limits. The total number of successive points in both spatial and time dimensions must thus exceed $8\,10^{27}/10^{-13} = 10^{41}$ at least. The finite lattice so characterized may be called the *Euclidean lattice* of the observable points.

The finiteness of E arises the question whether a finite four-dimensional linear space $EG(q, 4)$ over a finite Galois field $GF(q)$ exists, such that $EG(q, 4)$ contains a subset isomorphic with the lattice E. Obviously, the problem is equivalent to the problem whether a number ($>10^{41}$) of successive integers of $GF(q)$ can be transitively ordered, and thus put into an isomorphic correspondence with a sequence of equal-distanced real numbers.

Consider a Galois field $GF(q)$. It contains q elements (q is a prime), viz. the q rest classes of the ordinary integers modulo q. Let us denote these rest classes by $0_q, 1_q, 2_q, \cdots, (q-1)_q$, where $a_q = \{n;\ n = a \pmod q;\ n = \text{integer}\}$. There are always in a field $GF(q)$ primitive elements ρ for which the integer $\nu = q - 1$ is the lowest exponent for which $\rho^\nu = 1_q$ holds true. Every nonzero element a_q of $GF(q)$ is either an even or an odd potence of a given primitive element: $a_q = \rho^{2h}$ or $a_q = \rho^{2h+1}$, with an integer h. The even potences will be called "squares", the odd potences "not-squares." There are exactly $(q-1)/2$ squares and $(q-1)/2$ not-squares in a field $GF(q)$. Since the product of two squares, and of two not-squares is always a square, and the product of a square and a not-square is always a not-square, one can consider the squares as the positive numbers, and the not-squares as the negative numbers of the field $GF(q)$: $a_q > 0_q$ if $a_q = \rho^{2h}$, and $a_q < 0_q$ if $a_q = \rho^{2h+1}$. The relation $\rho^{q-1} = 1_q$ guarantees the positiveness of the unit element.

The relations "greater than" and "smaller than" can be defined for any two elements $a_q \in GF(q)$ and $b_q \in GF(q)$, $a_q \neq b_q$, by

$$\begin{aligned} a_q > b_q \quad &\text{if} \quad a_q - b_q > 0_q, \\ a_q < b_q \quad &\text{if} \quad a_q - b_q < 0_q, \end{aligned} \tag{2}$$

provided that the condition $-1_q < 0_q$ holds true. In terms of a symbol of Legendre this condition reads:

$$\left(\frac{-1}{q}\right) = -1. \tag{3}$$

It guarantees that the cases $a_q > b_q$ and $a_q < b_q$ exclude one another.

[8] Y. Ahmavaara, Ann. Acad. Sci. Fennicae Ser. A VI Nos. 85, 95, and 106 (1961)–(1962).

The sum of two squares is obviously not necessarily a square and, accordingly, the relation of order defined by (2) is not transitive in general. Therefore the elements of $GF(q)$ do not in general form ordered sequences of more than two elements. The sequence $0_q < 1_q < 2_q$, for instance, is transitively ordered only if, in addition to the trivial relations $2_q - 1_q = 1_q - 0_q = 1_q > 0_q$ (which are always true), also the condition $2_q - 0_q = 2_q > 0_q$ holds true. Obviously, any sequence $a_q < (a+1)_q < (a+2)_q < \cdots < (a+q_{k+1}-1)_q$ of q_{k+1} successive integers of $GF(q)$ is transitively ordered by (2) if, and only if all the nontrivial relations $2_q > 0_q$, $3_q > 0_q$, $\cdots$, $(q_{k+1}-1)_q > 0_q$ hold true simultaneously. This implies that all the elements $(q_i)_q$ of $GF(q)$, corresponding to the first k primes q_i, are squares. Using the symbol of Legendre this condition can be written as

$$\left(\frac{q_i}{q}\right) = 1, \qquad q_i = 2, 3, 5, 7, \cdots, q_k \text{ (successive primes).} \tag{4}$$

It was shown by Kustaanheimo[2] that the conditions (3) and (4) are simultaneously satisfied if one chooses the fundamental prime q of the field $GF(q)$ to be of the general form

$$q = 8q_1q_2\cdots q_k - 1 \text{ (the } q_i \text{ the } k \text{ first primes).} \tag{5}$$

The existence of an infinite set of solutions q of (5) is guaranteed by the theorem of Dirichlet (cf. Kustaanheimo, Ref. 2). If the element $(q_{k+1})_q$ is not-square, then the greatest possible length of "Euclidean chains"[4] of $GF(q)$ satisfying (5) is exactly q_{k+1} elements. For instance, $q = 47$ satisfies the condition (5), there being $q_k = 3$, $q_{k+1} = 5$, and $5_{47} < 0_{47}$. Accordingly, any sequence of five successive integers of $GF(47)$ is transitively ordered, and no sequence of six successive integers can be ordered in this field.

A four-dimensional linear space $EG(q, 4)$ over $GF(q)$, where q is of the form (5), and where q_{k+1} is of the magnitude of 10^{41} at least, thus contains a subset isomorphic with the lattice E. The *fine structure of space–time* is thus expressed by a space $EG(q, 4) \supset E$.

III. THE RELATIVITY GROUP OF A FINITE SPACE-TIME

Let us now consider a linear model W of space–time, that is, a four-dimensional linear space W over a field K of numbers, such that the Euclidean lattice E is contained in W. It is evident from the Sec. II that one can choose for K either the field R of the real numbers, or a Galois field $GF(q)$ satisfying the condition of Kustaanheimo [Eq. (5)], or a field of rational numbers, for instance. Let us introduce the following notations:

Λ : a 4×4 matrix with elements in K,
x : a point of W indicated by a column vector containing the time coordinate x_0 and the spatial coordinates x_1, x_2, and x_3.
a : another 4-vector of W,
$\tilde{\Lambda}$, $\tilde{x}$, etc.: the transposes of Λ, x, etc., (6)

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} : \text{the metric matrix of } W, \text{ with elements 0, 1, and } -1 \text{ in the field } K,$$

$$x \cdot y = \tilde{x}gy = x_1y_1 + x_2y_2 + x_3y_3 - x_0y_0; \quad x, y \in W,$$
$$x^2 = x \cdot x, \quad x \in W.$$

The transformations (Λ, a) of W onto itself, defined by

$$(\Lambda, a) : x \to \Lambda x + a, \tag{7}$$

form a group where the unit is $1 = (I, 0)$, $I =$ the 4×4 unit matrix. The group multiplication and the inverse are given, respectively, by

$$(\Lambda_1, a)\cdot(\Lambda_2, b) = (\Lambda_1\Lambda_2, \Lambda_1 b + a),$$
$$(\Lambda, a)^{-1} = (\Lambda^{-1}, -\Lambda^{-1}a). \tag{8}$$

In particular, we are interested in the translation group $\mathfrak{Z}$, the Lorentz group $\mathfrak{L}$, the Coish group $\mathfrak{C}$, the Poincaré group (or the inhomogeneous Lorentz group) $\mathfrak{P}$, and the Dieudonné group $\mathfrak{D}$ defined by

$$\begin{aligned} \mathfrak{Z} &= \{(I, a); a \in W\}, \\ \mathfrak{L} &= \{(\Lambda, 0); \tilde{\Lambda}g\Lambda = g\}, \\ \mathfrak{C} &= \{(\Lambda, 0); \tilde{\Lambda}g\Lambda = \pm g\}, \\ \mathfrak{P} &= \{(\Lambda, a); \tilde{\Lambda}g\Lambda = g, a \in W\}, \\ \mathfrak{D} &= \{(\Lambda, a); \tilde{\Lambda}g\Lambda = \pm g, a \in W\}. \end{aligned} \tag{9}$$

By neglecting the elements containing P or T, the inversions of space or time, one obtains the corresponding *restricted groups* $\mathfrak{L}_0$, $\mathfrak{C}_0$, $\mathfrak{P}_0$, and $\mathfrak{D}_0$. The relations between the full and the restricted groups are indicated by semidirect products:

$$\mathfrak{L} = \mathfrak{L}_0 \times \mathfrak{s}, \qquad \mathfrak{C} = \mathfrak{C}_0 \times \mathfrak{s},$$
$$\mathfrak{P} = \mathfrak{P}_0 \times \mathfrak{s}, \qquad \mathfrak{D} = \mathfrak{D}_0 \times \mathfrak{s}. \tag{10}$$

Here $\mathfrak{s}$ is the inversion group composed of the elements 1, P, T, and PT.

[4] This term for the transitively ordered sequences of $GF(q)$ was introduced by F. Levi, Zentr. Math. 39, 156 (1951) [A review of Ref. 2].

For $K = R$ (the field of the real numbers) one has, of course, $\mathfrak{L} = \mathfrak{C}$, and $\mathfrak{P} = \mathfrak{D}$. For $K = GF(q)$ it follows from a theorem of Dieudonné[5] that $\mathfrak{L} \neq \mathfrak{C}$, and $\mathfrak{P} \neq \mathfrak{D}$. The Λ matrices obeying $\tilde{\Lambda} g \Lambda = -g$ can all be generated by the Lorentz matrices together with a particular class of the non-Lorentz matrices, for instance, with the following class given by Coish[6]:

$$\Lambda_0 = \begin{pmatrix} 0_q & 0_q & 0_q & -1_q \\ 0_q & -\alpha_q & \beta_q & 0_q \\ 0_q & \beta_q & \alpha_q & 0_q \\ 1_q & 0_q & 0_q & 0_q \end{pmatrix}, \quad \alpha_q^2 + \beta_q^2 = -1_q. \tag{11}$$

Since the $(\Lambda_0, 0)$ transform $x^2 \to -x^2$ they may be called *reflections in the light cone*. Since the light cone itself is invariant in these reflections, it seems consequent to count the elements $(\Lambda_0, 0)$ to the relativity group of the linear space–time W over $GF(q)$, this group being thus defined by the Dieudonné group $\mathfrak{D}$ rather than by the Poincaré group.

IV. THE UNITARY REPRESENTATIONS OF THE DIEUDONNÉ GROUP

To consider the consequences of the fine structure of space–time in relativistic quantum theory one must first construct the unitary representations of the Dieudonné group $\mathfrak{D}$ over $GF(q)$ in a linear space over the ordinary complex numbers. The general method of Wigner and Mackey[7] can be applied.

By (8) one has

$$U(\Lambda, a) = U[(I, a)\cdot(\Lambda, 0)] = U(I, a)U(\Lambda, 0). \tag{12}$$

Since $\mathfrak{Z}$ is a commutative finite group one has by well-known theorems

$$\begin{aligned} U(I, a) &\sim \bigoplus_p U_p(a), \\ U_p(a) &= e^{(2\pi i/q)p\cdot a}, \qquad p \in EG(q, 4). \end{aligned} \tag{13}$$

Introducing the little groups $\mathfrak{C}(p)$ and their left cosets $K_{p'}(p)$ by

$$\begin{aligned} \mathfrak{C}(p) &= \{(\Lambda, 0);\ \Lambda p = p\}, \\ K_{p'}(p) &= \{(\Lambda, 0);\ \Lambda p = p'\}, \end{aligned} \tag{14}$$

one obtains from the elements $(\Lambda, 0)$ of the Coish group $\mathfrak{C}$ in a unique way elements $(\Lambda(p), 0)$ of the little Coish group $\mathfrak{C}(p)$ by the construction

$$(\Lambda(p), 0) = (\bar{\Lambda}_p(\Lambda p'), 0)\cdot(\Lambda, 0)\cdot(\bar{\Lambda}_{p'}(p), 0). \tag{15}$$

Here $(\bar{\Lambda}_p(\Lambda p'), 0)$ and $(\bar{\Lambda}_{p'}(p), 0)$ are freely chosen but once for all fixed elements of the respective cosets $K_p(\Lambda p')$ and $K_{p'}(p)$. Solving (15) for $(\Lambda, 0)$, using the evident rule $(\Lambda_{p'}(p), 0)^{-1} = (\Lambda_p(p'), 0)$, and exchanging the indices p and p' for convenience, one has the following composition of the unitary representation $U(\Lambda, 0)$:

$$U(\Lambda, 0) = U(\bar{\Lambda}_{\Lambda p}(p'), 0)U(\Lambda(p'), 0)U(\bar{\Lambda}_{p'}(p), 0). \tag{16}$$

Consider a linear space H over the ordinary complex numbers,

$$H = \bigoplus_p H_p, \qquad p \in EG(q, 4), \tag{17}$$

spanned by a basis system

$$\{\Phi_{p,\sigma};\ p \in EG(q, 4);\ \sigma = 1, 2, \cdots, \sigma(p)\},$$

so that $\Phi_{p,\sigma} \in H_p$. One can always choose this system so that one has

$$\begin{aligned} U(I, a)\Phi_{p,\sigma} &= U_p(a)\Phi_{p,\sigma}, \\ U(\bar{\Lambda}_{p'}(p), 0)\Phi_{p,\sigma} &= \Phi_{p',\sigma}, \\ U(\Lambda(p), 0)\Phi_{p,\sigma} &= \sum_\rho \mathfrak{D}_{\sigma\rho}(\Lambda(p), 0)\Phi_{p,\rho}. \end{aligned} \tag{18}$$

Here $(\Lambda(p), 0) \to \mathfrak{D}(\Lambda(p), 0)$ is an ordinary complex-number matrix representation of the little group $\mathfrak{C}(p)$. Combining (12), (13), (16), and (18) one has the *standard form* of the operator $U(\Lambda, a)$ defined by

$$U(\Lambda, a)\Phi_{p,\sigma} = e^{(2\pi i/q)\Lambda p\cdot a} \sum_\rho \mathfrak{D}_{\sigma\rho}(\Lambda(\Lambda p), 0)\Phi_{\Lambda p,\rho}. \tag{19}$$

In a Galois field $GF(q)$ satisfying the condition (5) of Kustaanheimo, the element -1_q is not-square and, accordingly, for every element $M > 0_q$ one has $-M < 0_q$. Let us now decompose the general space (17) as follows:

$$\begin{aligned} H &= H_0 \bigoplus_M H(M), \qquad H(M) = \bigoplus_p H_p, \\ p^2 &= \pm M, \qquad M \geq 0_q, \qquad p \neq 0. \end{aligned} \tag{20}$$

From $U(\Lambda, a)\Phi_{p,\sigma} \in H_{\Lambda p}$ and from $\Lambda p\cdot\Lambda p = \pm p^2$ one concludes that each of the spaces $H(M)$ carries a separate representation of the Dieudonné group $\mathfrak{D}$. Of course, the same holds for the space H_0 corresponding to the case $p = 0$.

From (14) it follows that $\mathfrak{C}(0) = \mathfrak{C}$ so that H_0 carries a representation of the Coish group $\mathfrak{C}$. This space is now less interesting in comparison with the spaces $H(M)$. Each subspace H_p of $H(M)$ evidently carries a representation $(\Lambda(p), 0) \to \mathfrak{D}(\Lambda(p), 0)$ of the little group $\mathfrak{C}(p)$. An element $(\Lambda_p(p'), 0)$ of the coset $K_p(p')$ conducts, by

[5] Dieudonné, Mem. Amer. Math. Soc. No. 2, p. 51 (1949).
[6] Coish, Phys. Rev. 114, 383 (1959).
[7] E. Wigner, Ann. Math. 40, 149 (1939); G. Mackey, Acta Math. 99, 265 (1958); for a similar application see H. Joos, J. Math. Phys. 5, 155 (1964).

$(\Lambda(p), 0) \to (\Lambda_p(p'), 0)^{-1} \cdot (\Lambda(p), 0) \cdot (\Lambda_p(p'), 0) \in \mathfrak{C}(p')$,

an isomorphism from $\mathfrak{C}(p)$ to $\mathfrak{C}(p')$ for every pair (p, p') such that $p'^2 = \pm p^2$ (notice that in the case $p'^2 = -p^2$ this result has no counterpart in the theory of continuous Lorentz group). This implies that all the representations of the little groups $\mathfrak{C}(p)$ in the respective spaces H_p are equivalent with one another and, accordingly, the little groups themselves are isomorphic to one another. One can thus introduce an abstract little group $\mathfrak{C}(M)$, depending only on the constant M, and isomorphic to all the little groups $\mathfrak{C}(p)$ with $p^2 = \pm M$.

If an index S is introduced to label the irreducible unitary representations of the abstract little group $\mathfrak{C}(M)$, one obtains the decomposition

$$\mathfrak{D}(\Lambda(p), 0) \sim \bigoplus_S \mathfrak{D}^{(S)}(\Lambda(p), 0) \tag{21}$$

of the $\mathfrak{D}$ matrices of Eq. (19), and the corresponding decomposition

$$H(M) = \bigoplus_S H(M, S) \tag{22}$$

of the spaces $H(M)$.

Each of the spaces $H(M, S)$ now obviously carries an irreducible unitary representation of the Dieudonné group $\mathfrak{D}$. In particular, this applies to the spaces $H(0_q, S)$ under the condition that the vectors $\Phi_{p,\sigma}$ corresponding to $p = 0$ are not included in $H(0_q, S)$ (but in H_0).

If one decomposes an irreducible manifold $H(M, S)$, for $M > 0_q$, of the Dieudonné group so that

$$H(M, S) = H(+M, S) \oplus H(-M, S), \tag{23}$$

where $H(+M, S)$ is spanned by those vectors $\Phi_{p,\sigma} \in H(M, S)$ for which $p^2 = +M$, and $H(-M, S)$ by those $\Phi_{p,\sigma} \in H(M, S)$ for which $p^2 = -M$, then each of the spaces $H(+M, S)$ and $H(-M, S)$ carries an irreducible unitary representation of the Poincaré group (the inhomogeneous Lorentz group) $\mathfrak{P}$. In particular, the inversions P and T are represented by the unitary operators

$$U(P)\Phi_{p,\sigma} = \Phi_{Pp,\sigma}, \qquad U(T)\Phi_{p,\sigma} = \Phi_{Tp,\sigma}. \tag{24}$$

Here Pp is the space-inverted vector p, and Tp the time-inverted p. In a reflection in the light cone, (11), the two spaces $H(+M, S)$ and $H(-M, S)$ are exchanged with one another.

An open problem so far is the construction of the ordinary complex-number matrices $\mathfrak{D}(\Lambda(p), 0)$. Evidently, this task presupposes the construction of the irreducible representations of the finite Coish group $\mathfrak{C}$ in terms of ordinary complex-number matrices. The present situation in this problem is the following. The non-Lorentz Coish transformations $(\Lambda_0, 0)$ of (11) are outer automorphisms of the Lorentz group $\mathfrak{L}$. The representations of the whole Coish group $\mathfrak{C}$ could be therefore determined if just the ordinary complex-number matrix representations of the finite Lorentz group $\mathfrak{L}$ were known. The latter group belongs to a class of "orthogonal groups" investigated indeed since the work of Dickson,[8] but the interest has been so far restricted to modular representations only. Fortunately, the situation will perhaps change soon.[9]

So far only some properties of the $\mathfrak{D}$ matrices are known. There is a result of Coish[6] which can be formulated as follows.

Theorem: The irreducible modular representations of the restricted Coish group $\mathfrak{C}_0$ over $GF(q)$ are multivalued, every element $(\Lambda, 0)$ of $\mathfrak{C}_0$ being represented by several matrices in the following way:

$$(\Lambda, 0) \to \omega^{Qk} M(\Lambda, 0);$$
$$k = 0, 1, 2, \cdots, q; \quad \omega \in GF(q); \tag{25}$$
$$\omega^{q+1} = 1_q; \quad (\Lambda, 0) \to M(\Lambda, 0) \text{ is a function;}$$

The integer Q determining the multivaluedness is different for different nonequivalent representations, and it has the values $Q = 0, \pm 1, \pm 2, \cdots$.

The theorem implies that one can define a "covering group" $\mathfrak{C}_0'$ of $\mathfrak{C}_0$ in order to make the representations (25) univalued, and that it is the covering Coish group $\mathfrak{C}'$ rather than $\mathfrak{C}$ which is relevant in quantum theory. The unitary irreducible representations of the extended little groups $\mathfrak{C}'(p)$ in terms of ordinary complex-number matrices will obviously be of the form

$$(\Lambda(p), 0) \to e^{2\pi i Qk/(q+1)} \mathfrak{D}^{(S)}(\Lambda(p), 0);$$
$$k = 0, 1, 2, \cdots, q. \tag{26}$$

Here the matrices $\mathfrak{D}^{(S)}(\Lambda(p, 0))$ give a univalued irreducible unitary representation of the little Coish group $\mathfrak{C}(p)$ for every value of S.

When the right member of (26), with the obvious change $p \to \Lambda p$, is substituted for the $\mathfrak{D}$ matrix appearing in (19), the unitary operator $U(\Lambda, a)$ defined by (19) gives an irreducible unitary representation of the covering Dieudonné group $\mathfrak{D}'$. Such a representation is characterized by the three labels M, S, and Q.

It is shown in the next section that the label M is connected with the notion of *rest mass*. On the

[8] L. Dickson, *Linear Groups* (Dover Publications, Inc., New York, 1958) (first printing in 1901).

[9] A private communication from Professor R. Brauer.

other hand, since the abstract little groups $\mathfrak{C}(M)$ are rotation groups, one must expect the label S to be connected with the notion of *spin*. What is then the physical interpretation of the label Q?

To consider this question, consider the Q transformations of the Hilbert space which are given by

$$\Phi_{p,\sigma,Q} \to e^{2\pi i Qk/(q+1)} \Phi_{p,\sigma,Q}, \quad k = 0, 1, 2, \cdots, q. \tag{27}$$

Since these are finite analogies to the continuous gauge transformations the label Q could be interpreted to represent a charge quantum number like, for instance, the *electric charge* number (this is the interpretation of Coish[6]).

V. THE SPECTRAL FUNCTIONS OF REST MASS, ENERGY, AND MOMENTUM

Following the conventional physical interpretation of the formalism of relativistic quantum theory one must consider an irreducible manifold $H(M, S, Q)$ of the covering Dieudonné group $\mathfrak{D}'$ as the Hilbert space composed of the single-particle states of the particle of the species $\lambda = (M, S, Q)$.

The label M has the range of values composed of the $\frac{1}{2}(q+1)$ nonnegative elements of a Galois field $GF(q)$. Such an element is either zero or a square and, accordingly, can always be written as

$$M = \mu^2, \qquad \mu \in GF(q). \tag{28}$$

A vector $\Phi_{p,\sigma} \in H(M, S, Q)$ must be considered as a state of an exact 4-momentum p (and of the variable σ) of the particle (M, S, Q). In particular, the component p_0 of p must be considered as representing the energy of the particle (M, S, Q) in the state $\Phi_{p,\sigma}$. The components p_1, p_2, and p_3 together form the 3-vector momentum of this particle in this state. When doing these interpretations one must, however, remind that one is giving the names of "energy" and "momentum" to Galois numbers: all the components of p are now elements of $GF(q)$.

In particular, it follows from (28) that the rest mass of the particle (M, S, Q) can be defined as the Galois number μ but only for the states $\Phi_{p,\sigma} \in H(-M, S, Q)$ [cf. (23)]: only these states have the correct relation

$$p^2 = p_1^2 + p_2^2 + p_3^2 - p_0^2 = -\mu^2. \tag{29}$$

Let us now decompose the space $H(-M, S, Q)$ so that

$$H(-M, S, Q) = H_1(-M, S, Q) \oplus H_2(-M, S, Q), \tag{30}$$

where $H_1(-M, S, Q)$ is the space spanned by those vectors $\Phi_{p,\sigma} \in H(-M, S, Q)$ for which the element $p_1^2 + p_2^2 + p_3^2$ is either zero or a square of $GF(q)$. To every state $\Phi_{p,\sigma} \in H_1(-M, S, Q)$ one can thus associate a scalar momentum κ defined by

$$p_1^2 + p_2^2 + p_3^2 = \kappa^2, \qquad \kappa \in GF(q). \tag{31}$$

Let the particular field $GF(q)$ now under consideration be so chosen that the condition (5) of Kustaanheimo is satisfied. There are then Euclidean chains in $GF(q)$, that is, there are sequences

$$a_q < (a+1)_q < \cdots < (a + q_{k+1} - 1)_q$$

of q_{k+1} successive integers of $GF(q)$ which can be mapped isomorphically to the q_{k+1} ordinary integers $a < a+1 < \cdots < a + q_{k+1} - 1$. Of course, the "physical domain" of each of the variables p_1, p_2, p_3, p_0, κ, and μ must be represented by the respective Euclidean chains E_1, E_2, E_3, E_0, E_κ, and E_μ of $GF(q)$.

From the nonnegativeness of the physical domains of p_0, κ, and μ, and from the symmetry of the physical domains of p_1, p_2, and p_3 with respect to the origin (of the observers reference frame) it follows that the Euclidean chains representing the physical domains must be chosen as follows:

$$\begin{aligned}
E_i &= \{p_i; p_i = (k_i)_q, k_i = 0, \pm 1, \cdots, \pm \tfrac{1}{2}(q_{k+1} - 1)\}, \\
E_0 &= \{p_0; p_0 = (k_0)_q, k_0 = 0, 1, 2, \cdots, q_{k+1} - 1\}, \\
E_\kappa &= \{\kappa; \kappa = k_q, k = 0, 1, 2, \cdots, q_{k+1} - 1\}, \\
E_\mu &= \{\mu; \mu = m_q, m = 0, 1, 2, \cdots, q_{k+1} - 1\}.
\end{aligned} \tag{32}$$

The states $\Phi_{p,\sigma} \in H_1(-M, S, Q)$, for which each of the variables p_1, p_2, p_3, p_0, κ, and μ belongs to the respective physical domain, may be called the *observable states of momentum.*

The relations (29) and (31) are valid for any states $\Phi_{p,\sigma} \in H_1(-M, S, Q)$. In particular, when applied to the observable states of momentum, they give the fundamental conditions

$$m^2 = k_0^2 - k^2, \qquad k^2 = k_1^2 + k_2^2 + k_3^2. \tag{33}$$

For every solution $(k_1, k_2, k_3, k_0, k, m)$ of (33) by the integers allowed by (32) there is an observable 4-momentum p.

If the results of measurement are indicated by using a natural unit of measurement, viz. the smallest difference ϵ of energy, then the observable values of the variables p_1, p_2, p_3, p_0, κ, and μ corresponding to a given solution $(k_1, k_2, k_3, k_0, k, m)$ of (33) are given by

$$\begin{aligned}
(p_1)_{\text{obs}} &= k_1\epsilon, \quad (p_2)_{\text{obs}} = k_2\epsilon, \quad (p_3)_{\text{obs}} = k_3\epsilon, \\
(p_0)_{\text{obs}} &= k_0\epsilon, \quad \kappa_{\text{obs}} = k\epsilon, \quad \mu_{\text{obs}} = m\epsilon.
\end{aligned} \tag{34}$$

The conditions (33) determine the distribution of

the observable states of momentum. This distribution can be expressed by the spectral function F from $E_1 \times E_2 \times E_3 \times E_0 \times E_\kappa \times E_\mu$ to the numbers 1 and 0, defined to have the value 1 for every solution $(k_1, k_2, k_3, k_0, k, m)$ of (33) by the numbers (32), and otherwise the value 0. The spectral functions F_m and $\sum$ of energy and rest mass, respectively, are then given by

$$F_m(k_0) = \sum_{E_1} \sum_{E_2} \sum_{E_3} F(k_1, k_2, k_3, k_0, k, m), \qquad \Sigma(m) = \sum_{E_0} F_m(k_0). \tag{35}$$

The number $F_m(k_0)$ gives the number of observable 4-momenta for the fixed values $m\epsilon$ and $k_0\epsilon$ of the observable rest mass and energy. The number $\Sigma(m)$ gives the total number of observable 4-momenta for the fixed value $m\epsilon$ of observed rest mass. The existence such nontrivial spectral functions of geometrical origin is obviously a fundamental consequence of the fine structure of space–time.

Examples: For $m = 0$ one has $F_0(k_0) > 0$ for every value $k_0 = 1, 2, 3, \cdots, q_{k+1} - 1$. Since the number q_{k+1} is larger than 10^{41}, the value of $\Sigma(0)$ is very large, indicating the existence of particle(s) having the rest mass zero. Evidently, the peak of the Σ function at $m = 0$ is of the maximal possible magnitude. The energy spectrum of the particle $m = 0$ begins with the values $F_0(1) = F_0(2) = F_0(3) = F_0(4) = 6$, $F_0(5) = 30$. The mass spectrum $\Sigma(m)$ has, in the immediate vicinity of $m = 0$, the following course: $\Sigma(1) = \Sigma(2) = 1$, $\Sigma(3) = \Sigma(4) = 7$, $\Sigma(5) = 31$. The values of $\Sigma(m)$ for $m \neq 0$ are thus in this domain vanishing in comparison with the value $\Sigma(0) \geq 10^{41}$, which shows that there is practically no observable rest mass in the nearest vicinity of $m = 0$.

Taking into account the smallness of the mass unit here employed, it is the behavior of the function Σ for very large m which is interesting in the connection with the problem of the mass spectrum of elementary particles. About this problem a preliminary remark can be made here.

Since (33) can be rewritten as $(k_0+k)(k_0-k)=m^2$, $k^2 = k_1^2 + k_2^2 + k_3^2$, there is a solution (k_0, k) for each integer divisor of m. Indeed, if $m^2 = ab$, where a and b are positive integers such that $a \geq b$, then there is one and only one solution (k_0, k) of (33) for each pair (a, b), viz., that one given by $k_0 = \frac{1}{2}(a + b)$, $k = \frac{1}{2}(a - b)$. Accordingly, the number of the different observable energies of a particle having the rest mass $m\epsilon$ is given by the number of the major divisors of m. Thus the problem of the mass spectrum is closely connected with the well-known number-theoretical problem concerning the distribution $d(m)$ of the divisibility of integer m. One can easily see that the functions $d(m)$ and $\Sigma(m)$ have, for large values of m, discrete peaks having irregular mutual intervals.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Invariant Approach to the Geometry of Spaces in General Relativity*

CARL H. BRANS
Loyola University, New Orleans, Louisiana
(Received 18 September 1964)

A procedure is described for obtaining a complete, invariant classification of the local, analytic geometries and matter fields in general relativity by a finite number of algebraic steps. The approach is based on an extension of the classification scheme to include differential invariants of all orders and to provide maximally determined standard frames of vectors at each point. It is further shown that the resultant invariant functions can be replaced, in a finite number of algebraic steps, by special invariant functions which, while still uniquely representative of the geometry, can be assigned arbitrarily to produce all possible local, analytic solutions to the Einstein equations, in this representation. It is suggested that this type and special function scheme, obtainable from ideal geometric measurements in a finite number of steps, could be useful in general relativity. Unfortunately, due to the extensive algebra involved, this scheme has not yet been explicitly calculated, even for empty spaces.

I. INTRODUCTION

THE usual approach to the study of the differential geometric properties of a manifold is through the metric, which is effectively described by the functions representing the components of the metric tensor in each coordinate system. Another procedure is to use the structure of the infinitesimal parallel displacement as described by the connection. In standard general relativity, where the connection is required to be torsion free and metric, these descriptions are equivalent. Of the two it is perhaps the metric approach which seems to be more natural and intuitive and more closely related to ideal physical measurements. Further, the approach through the connection form has the disadvantage that it is not effectively productive of a geometry in the sense that the functions that are to represent the components of the connection in a particular coordinate system cannot be assigned arbitrarily but must satisfy differential equations expressing the condition that the connection be metric. The direct metric approach, on the other hand, does not give rise to this difficulty and geometries can be produced at will merely by assigning the functions to represent the components of the metric tensor in a particular coordinate system, subject only to qualitative conditions such as differentiability, signature, and the nonvanishing of the determinant.

However, both the metric and connection form approaches do have some drawbacks, particularly in applications in general relativity. In the first place, the description of the geometry in terms of the functions representing the components of the metric tensor or connection form is clearly coordinate dependent. This fact leads to many theoretical and practical difficulties, for example, the practical problem of an effective test for whether or not two such sets of functions represent geometries that are equivalent, at least locally. Secondly, the structure of the Einstein equations, even in empty space, is of such complexity that the study of their properties has been very difficult and slow.

The purpose of this paper is to demonstrate explicitly the existence of an alternate approach to the geometry of spaces satisfying the Einstein equations. In this approach the local geometry of such a space is described by first placing it in one of a finite number of discrete types and then assigning to it a finite set of scalar functions, the special invariant functions, which can be chosen arbitrarily, but which are uniquely representative of the local geometry. It should be emphasized that these are not merely the usual second-order invariant functions such as the Petrov scalars. These latter are *not* effectively productive of a geometry since they can only be chosen subject to differential conditions, and thus are not entirely suitable to represent the geometry. Further, they are complete only in certain cases.

This approach has several advantages. First, it provides an invariant, one-to-one representation of the local geometry. That is, two spaces satisfying the Einstein equations are locally equivalent geometrically if, and only if, they are of the same type and their special invariant functions agree in form, not just in value, providing an effective test of the local equivalence of such spaces. Further, for a

* Supported by a grant from the National Science Foundation.

given matter tensor, all geometries solving the Einstein equations can be locally represented by such a type and set of special invariant functions. Finally, within each type the functions can be prescribed arbitrarily, subject only to qualitative conditions, as the components of the metric tensor as discussed above. Thus all local solutions to the Einstein equations can be produced at will within this representation, which of course, leads to many advantages, e.g., a combination law for solutions, such as the addition of special invariant functions, and other possibilities discussed in the conclusion below.

Of course, this type and special function scheme is admittedly further removed from ordinary geometric measurements than either the metric or connection approach. However, the type and special invariant functions can be obtained from local metric or parallel displacement measurements by a finite number of algebraic steps and a finite number of differentiations. Thus, in principle, following the prescription outlined below, a device could be constructed which would directly read off the type and special invariant functions from local measurements. Conversely, given a type and set of special invariant functions, the analytic expression for the components of the metric tensor can be obtained from the analytic expression for the special invariant functions by algebraic steps (possibly infinite in number).

Thus, this type and special function scheme is as legitimate a representation for local geometry as the assignment of functions to represent the components of the metric tensor. Further, it has the added advantages of providing this representation invariantly, with a "built-in" test for the local equivalence of geometries, and of giving a complete classification of the local solutions to Einstein equations.

Unfortunately, because of the extensive amount of algebraic manipulation required, this scheme has not been explicitly carried through in closed form, even for empty spaces. Thus, this paper is essentially in the form of a proof of an existence theorem stating that such a scheme can be obtained by a finite number of algebraic steps and providing a definite prescription for doing so. It should be noted that since these are essentially algebraic steps, a computer might conceivably be useful in completing the task.

The central theorem used in this paper, that the geometry is characterized by the Riemann tensor and its covariant derivatives, is well known. For a statement and proof of it see Cartan.[1] A canonical form for the Riemann tensor in Einstein spaces was first obtained by Petrov,[2] using the special algebraic properties of such a tensor. Komar[3] showed how the Petrov functions could be used to establish canonical coordinates when they are independent. For a further discussion of these problems and a guide to the associated literature see Bergmann,[4] or Petrov.[5]

The main features of this paper are as follows. First, (Sec. II), it presents an explicit procedure for obtaining a canonical or standard form for *any* Riemann tensor and its sequence of covariant derivatives. This procedure does not depend on special algebraic properties, such as Petrov's for the Weyl tensor, giving it more generality for use with higher-order invariants, but causing it to be more complex in practical applications. This makes possible the effective, general application of the equivalence theorem mentioned above. Secondly, (Sec. III), from an analysis of the resulting algebraic differential equations, it gives a definite prescription for replacing the resulting invariant functions, which *cannot* be chosen arbitrarily by the nonredundant and independent special invariant functions which are arbitrary, but which still provide an invariant one-to-one representation of the local geometry, within each type. Section IV mentions briefly some further problems and possible applications. The Appendix contains a proof of a theorem on product representations needed in Sec. II.

Before proceeding, we should explicitly state the more important restrictions which we are assuming throughout this paper. First, we are only concerned with *local* and *analytic* solutions. Second, the matter theories and tensors must be *algebraic* in the matter fields and their derivatives. Finally, we are only concerned with points in the manifold having the property that the rank and other discrete algebraic structures of analytic matrices are constant over some neighborhood of the point. Since such discrete properties can change only discontinuously, and thus across boundaries of dimensions less than 4, it is felt that these jumps can be dealt with later as limiting cases of situations studied here. They will naturally play a central role in possible global

---

[1] E. Cartan, *Leçons sur la geometrie des espaces de Riemann* (Gauthier-Villars, Paris, 1951), 2nd ed.

[2] A. Petrov, Sci. Not. Kazan State Univ. **114**, 55 (1954).

[3] A. Komar, Phys. Rev. **111**, 1182 (1958).

[4] P. G. Bergmann, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1962), Vol. IV, pp. 203–272.

[5] A. S. Petrow (Petrov), *Einstein-Raume*, (Akademie-Verlag, Berlin, 1964).

extensions involving the matching of metric and matter fields of different type.

II. CANONICAL FORM FOR RIEMANN TENSOR AND ITS DERIVATIVES

We now sketch an effective procedure for determining whether two metric and matter fields are coordinate equivalent by giving an invariant description of them. The approach is based on the well-known fact that the Riemann tensor and its covariant derivatives provide a complete set of invariants for the metric field. The problem, however, is that the components of these tensors must be expressed in some standard frame and the resulting functions in some standard coordinate system before this test is effective. In this section, we solve this problem by studying the direct sum of product (tensor) representations of the Lorentz group and by showing how to obtain a canonical form for a vector (i.e., a sequence of tensors of increasing rank) under such a representation.

The geometric formalism used here is that due to Cartan.[1] The metric is expressed in terms of differential forms, $\omega^a = \omega^a_i dx^i$; $a, i = 0, 1, 2, 3$,

$$ds^2 = \eta_{ab}\omega^a\omega^b; \quad \eta_{ab} = \text{diag}\,(-1, +1, +1, +1) \tag{1}$$

satisfying the structure equations

$$d\omega^a = \omega^b \wedge \omega^a_b; \qquad \omega_{ab} + \omega_{ba} = 0 \tag{2}$$

$$d\omega^b_a - \omega^c_a \wedge \omega^b_c = \tfrac{1}{2}R_a{}^b{}_{cd}\omega^c \wedge \omega^d. \tag{3}$$

In the following we will refer to such a set of forms as a *frame*, although, strictly speaking, it is the isomorphic, dual set of tangent vectors which should be so denoted. Other terms which have been used for such frames are *tetrad* and *vierbein*.

In addition, we consider matter variables, ρ_A, which may include scalars, vectors, tensors, satisfying covariant equations

$$F(\rho_A, \rho_{A;a} \cdots) = 0 \tag{4}$$

and yielding a matter tensor $T_{ab}(\rho_A)$. We assume further that the ρ_A themselves are directly observable (once the frame is determined) and do not include gauge-dependent quantities.

Consider now the sequence consisting of the components of the Riemann tensor and its first N covariant derivatives relative to the frame ω^a.

$$R_{abcd}; \cdots; R_{abcd;a_1\cdots a_N}. \tag{5}$$

A convenient method for ordering quantities with such multiple sets of indices is to order the indices lexicographically. Thus if α represents the set $(a_1 \cdots a_n)$ and β represents the set $(b_1 \cdots b_n)$ then we will say that α precedes β if the first nonzero term in the sequence $a_1 - b_1 \cdots a_n - b_n$ is less than zero. This is of course just the procedure used for ordering words in a dictionary. Hence, using Greek indices to denote such sequences so ordered, the expressions in (5) could be denoted by $v^{(N)}_\alpha$ where the superscript N denotes the number of covariant derivatives represented and α runs from 1 to $\nu \equiv \sum_{i=4}^{N+4} 4^i$.[6] Similarly, let ω^α represent the corresponding base forms associated with the tensor products of the appropriate number of ω^a. Lorentz transformations of the ω^a then result in transformations of the ω^α, yielding a (reducible) representation of the Lorentz group, $R^N(L)$, over the vector space $V^{(N)}$, spanned by the ω^α. Note that the dimension of $V^{(N)}$ is not N but is ν.

The next step is to choose frames obtainable from the ω^α by transformations of $R^N(L)$ in such a way that $v^{(N)}_\alpha$ takes some standard form. Of course, if $R^N(L)$ were the full group of linear transformations over $V^{(N)}$, we could choose a transformation g with $\bar{\omega}^\alpha = g\omega^\alpha$ so that $v^{(N)} \equiv \bar{v}^{(N)}_\alpha \bar{\omega}^\alpha$ is along the first axis, that is, so that $\bar{v}^{(N)}_1 \neq 0$ while the other components are zero. This would then determine a natural standard form for $v^{(N)}_\alpha$ and the corresponding frames would be standard. Since $R^N(L)$ is not so large, however, we cannot expect this to be possible for a general $v^{(N)}_\alpha$ and must thus obtain a weakened substitute.

In order to obtain such a generally applicable procedure for standardizing $v^{(N)}_\alpha$ we will proceed as follows. First, for each admissible frame ω^α (i.e., one obtained from the appropriate products of members of a Lorentz 4-frame) consider the sequence

$$\Sigma \equiv \{V_1 \cdots V_\nu, V_{12} \cdots V_{\alpha\cdots\alpha_j}, \cdots V_{1\cdots}\} \tag{6}$$

consisting of the distinct vector subspaces of $V^{(N)}$ generated by all possible combinations of subsets of the basic set $\{\omega^\alpha\}$. The ordering of the subspaces is important and for each Σ is defined as follows: If $V_{\alpha_1\cdots\alpha_j} \in \Sigma$ and $V_{\beta_1\cdots\beta_k} \in \Sigma$ then $V_{\alpha_1\cdots\alpha_j}$ will precede $V_{\beta_1\cdots\beta_k}$ if $j < k$ or if $j = k$ and $\alpha_1 \cdots \alpha_j$ precedes $\beta_1 \cdots \beta_k$ lexicographically.

Clearly, for each Σ, there is a first $V_{\alpha_1\cdots\alpha_j}$ with $v^{(N)} \in V_{\alpha_1\cdots\alpha_j}$. After a transformation of $R^N(L)$ to another base ω^α with corresponding Σ we will obtain another set of indices $\beta_1 \cdots \beta_k$ with $V_{\beta_1\cdots\beta_k}$ the first element of Σ for which $v^{(N)} \in V_{\beta_1\cdots\beta_k}$. Further, we can compare $\alpha_1 \cdots \alpha_j$ to $\beta_1 \cdots \beta_k$ by the same ordering as was used for the sequence

[6] These components clearly will not all be independent for actual Riemann tensors because of the Bianchi identities, etc., However, these restrictions will be automatically imposed by the differential equations (2) and (3) themselves.

in Σ. Using this ordering we will now let $\alpha_1 \cdots \alpha_n$ represent the first such set of indices for all possible Σ. Thus we have, for some ω^α,

$$v^{(N)} = \lambda_1\omega^{\alpha_1} + \cdots + \lambda_n\omega^{\alpha_n}; \qquad \lambda_i \neq 0 \tag{7}$$

while, for any $g \in R^N(L)$, if $\bar{\omega}^\beta = g\omega^\alpha$ and

$$v^{(N)} = \bar{\lambda}_1\bar{\omega}^{\beta_1} + \cdots + \bar{\lambda}_k\bar{\omega}^{\beta_k}; \qquad \bar{\lambda}_i \neq 0, \tag{8}$$

then we must have either $k > n$ or else $k = n$ in which case we must have that the sequence $\beta_1 \cdots \beta_k$ does not precede $\alpha_1 \cdots \alpha_n$ lexicographically. Any such set ω^α satisfying this condition will be called a *frame adapted to* $v^{(N)}$.

Note that this set of indices $\alpha_1 \cdots \alpha_n$ and thus the set of frames adapted to $v^{(N)}$, is well defined and independent of the original choice of the Lorentz frame ω^α. It might perhaps be thought that this is not true and that the ordering of the ω^α might be relevant, because of the dependence of the definitions on ordering. That this is not the case is easily seen from the fact that a change of ordering of the ω^α can be brought about by a transformation of $R^N(L)$ since we have not restricted ourselves to the proper Lorentz group. Thus, since $\alpha_1 \cdots \alpha_n$ is obtained by a minimization procedure over a set including such changes, the result is independent of the original choice of ω^α, or ω^a.

We next define H_0^N to be the set of transformations $g \in R^N(L)$ which take the adapted frame ω^α into another such. Thus, $g \in H_0^N$ if and only if

$$v^{(N)} = \bar{\lambda}_1\bar{\omega}^{\alpha_1} + \cdots + \bar{\lambda}_n\bar{\omega}^{\alpha_n}; \qquad \bar{\lambda}_i \neq 0, \tag{9}$$

where $\bar{\omega}^\alpha = g\omega^\alpha$. Of course, H_0^N so defined could conceivably depend on the choice of original adapted frame ω^α. However, if H_0^N is the resulting set defined by a different choice, $\bar{\omega}^\alpha = g_0\omega^\alpha$, of adapted frame, then clearly

$$\bar{H}_0^N = H_0^N g_0^{-1}, \quad \text{with} \quad g_0 \in H_0^N. \tag{10}$$

Hence $\bar{H}_0^N$ is homeomorphic to H_0^N and in the following H_0^N can be considered to be defined only up to such a homeomorphism. What is of direct physical significance, of course, is the set of adapted frames (which will be later narrowed down to the set of standard frames) which is completely and uniquely defined. H_0^N is merely a certain subset of the transformation between such frames. It is clear that the sequence $\alpha_1 \cdots \alpha_n$ and H_0^N can be obtained by a finite number of algebraic operations. In fact, the expression of condition (7) for a given set $\alpha_1 \cdots \alpha_n$ together with the condition that $g \in R^N(L)$ consititute a finite set of simultaneous algebraic equations. Since, for fixed N, there are only a finite set of sequences of indices, the first one for which these equations are consistent can be found algebraically as can the resulting set H_0^N, which is thus an algebraic variety. Of course, we may not yet have arrived at a definite standard form for the $v_\alpha^{(N)}$ since for some $g \in H_0^N$ the $\bar{\lambda}_i$ as defined in (9) may not be numerically equal to the λ_i. Hence let us now define the functions $\lambda_i(g)$ for $g \in H_0^N$ from the following equation

$$v^{(N)} = \lambda_1(g)g\omega^{\alpha_1} + \cdots + \lambda_n(g)g\omega^{\alpha_n}; \quad g \in H_0^N. \tag{11}$$

We must thus deal with the possibility that some of the $\lambda_i(g)$ may vary with $g \in H_0^N$, and we must further refine H_0^N.

From the Appendix, it is seen that the only possibilities are that $\lambda_1(g)$ either is constant, achieves the value 1 or -1, or reaches a maximum as g ranges over H_0^N. These alternatives, in that order, will be called a standard form for λ_1. Sets H_i^N, $i = 1 \cdots n$ will then be defined inductively as those elements of H_{i-1}^N for which λ_i achieves its standard form. For convenience, denote by s_N the sequence $\{\alpha_1 \cdots \alpha_n, j_1 \cdots j_p, k_1 \cdots k_q\}$ where the j's and k's are those indices for which $\lambda_j = +1$, $\lambda_k = -1$. Further, let f_p^N denote the remaining set of determined values of the λ's, and let H_n^N be denoted by G^N. The f_p^N may be nonconstant functions as may the sequence s_N. However, as mentioned in the introduction, we are only considering regions over which s_N is constant.

Again we must show that these results are independent of our original choice of adapted frames. Suppose that we had started with a different choice, $\bar{\omega}^\alpha = g_0\omega^\alpha$. Then the functions $\bar{\lambda}_i(g)$ for $g \in \bar{H}_0^N$ would have been defined by an equation analogous to (11) and it is easy to see that

$$\bar{\lambda}_i(\bar{g}) = \lambda_i(\bar{g}g_0) \quad \text{for} \quad \bar{g} \in \bar{H}_0^N \tag{12}$$

so that the functions $\bar{\lambda}_i$ and λ_i have homeomorphic domains and take the same value at corresponding points under this homeomorphism. As a consequence, the standard form and values obtained for the $\bar{\lambda}_i$ using the procedure described above would be precisely the same as those obtained from the λ_i. Thus we have that the sequence s_N and the values f_p^N are uniquely defined scalars, independent of the original choice of adapted frame.

Finally, and most important for our applications, we have the result that *two sets of matrices of the form in expression (5) can be transformed into each other by a Lorentz transformation if, and only if, they have the same* s_N *and* f_p^N.

The basic equivalence theorem[1] states that two

geometries are locally equivalent if, and only if, their corresponding sequences of Riemann tensor and its covariant derivatives, relative to Lorentz frames, can be transformed into each other by a Lorentz transformation. Actually, in the statement given by Cartan, a stronger result is obtained, namely, that there is an integer k such that only those covariant derivatives of order less than or equal to k need be compared in order to determine local geometric equivalence. This integer k is the smallest integer for which the covariant derivatives of order k can all be expressed as a function of the lower-order covariant derivatives and the Riemann tensor itself. The concise and beautiful proof of this theorem by Cartan is an excellent example of the power of the method of differential forms in differential geometry.

In order to apply this to our results above, let x^i be any admissible set of coordinates over the region considered and let $r(N)$ be the rank of the matrix

$$\{\partial f^N_p/\partial x^i\}. \tag{13}$$

(Note that in our notation f^N_p contains information on all derivatives up to order N.) Clearly the integer k is the first integer for which

$$r(k-1) = r(k) \quad \text{and} \quad G^{k-1} = G^k. \tag{14}$$

For the special case of Einstein spaces, Kerr[7] has related k and G^k to the set of motions in the manifold. The set of frames $\tilde{\omega}^a = g^{-1}\omega^a$ for $g \in G^k$ we will call the *set of standard frames.*

In these terms, then, the test in the equivalence theorem can be stated in terms of the equality of k, s_k, f^k_p for two geometries. This is still not an effective test, however, since the f^k_p may be nonconstants and may be expressed in different coordinates. To obviate this, standard coordinate systems[3,4] can be introduced as follows. Let $p_1 \cdots p_{r(k)}$ be the first set of indices for which the matrix

$$\{\partial f^k_{p_1}/\partial x^i \cdots \partial f^k_{p_{r(k)}}/\partial x^i\} \tag{15}$$

has rank $r(k)$. The *standard coordinate systems* are then defined to be those for which

$$f^k_{p_s} = x^s; \qquad s = 1 \cdots r(k). \tag{16}$$

Let the sequence s_k with p_s adjoined be called s'_k and the remaining set of f^k_p, depending only on the determined part of the standard coordinates x^s be denoted by f'^k_p. Hence, *two geometries are locally equivalent if, and only if, they have the same k, s'_k, and their functions f'^k_p have the same form. This is thus an effective test, locally applicable to any analytic geometry.*

Clearly, this procedure can be easily extended to include the matter variables merely by adding the components of the tensors representing the matter variables to the end of the sequence of covariant derivatives of the Riemann tensor, and making possible further reductions of these to standard forms. This will yield a possible reduction in G^k and a possible extension of the determined part of the standard coordinates to x^t, $t = 1 \cdots \bar{r} \geq r(k)$.

In summary, then, we have given an explicit procedure for locally determining in a finite number of algebraic steps, invariant integers k, sequences S, and functions F_Q for each set of metric and matter fields in such a way that two sets of metric and matter fields can be locally transformed into each other if, and only if, they have the same k, S and their functions F_Q agree *in form.*[8] Again, this is clearly an effective test.

The integer k and the sequence S, determine what we call the *type* of the metric and matter fields. Thus the type and the form of the invariant functions, F_Q, provide an invariant and unique local representation of the metric and matter fields. However, this is not satisfactory yet since we cannot choose the functions F_Q arbitrarily and then be sure that they will correspond to a possible geometry and set of matter fields, since they must clearly satisfy differential conditions, even without the field equations themselves. In other words, a geometry and set of matter fields cannot be constructed from an arbitrary choice for these invariant functions, as for example, the geometry can from an arbitrary choice (modulo qualitative conditions) of components for the metric tensor. Further, we do not yet know which types do actually correspond to possible geometries and matter fields.

The purpose of the next section is to show how to obviate these difficulties by replacing the F_Q with functions which can be chosen arbitrarily and which can be obtained from the F_Q by finite number of algebraic steps and differentiations.

III. THE SPECIAL INVARIANT FUNCTIONS

We now study the differential equations relating the function F_Q to the standard frame, ω^a, and the connection forms, $\omega^b_a = \omega^b_{ai}dx^i$, together with the Einstein and matter equations themselves, for each possible type k, S. These will clearly consist of (2), (3), and (4) together with

[7] R. P. Kerr, J. Math. Mech. 12, 33 (1963).

[8] Note that the F_Q depend only on the determined part of the standard coordinate system.

$$R_{ab} - \eta_{ab}R/2 = \kappa T_{ab} \tag{17}$$

$$R_{abcd;e_1\cdots;e_{i+1}} = R_{abcd;e_1\cdots e_i,e_{i+1}} - \omega^f_{ae_{i+1}}R_{fbcd;e_1\cdots e_i} - \cdots;\ i = 0\cdots k-1 \tag{18}$$

$$\partial F_Q/\partial x^j = 0 \quad \text{for} \quad j > j_Q, \tag{19}$$

in which the R_{abcd} and derivatives are expressed in terms of the functions F_Q and x^i (i.e., only the determined part of the standard coordinate systems) in a definite algebraic manner determined by the type k, S. The same is true for the matter variables, ρ_A. These equations are then to be solved for the ω^a_i, ω^b_{ai} and the F_Q, which are to be regarded as the unknowns. For brevity let us group these unknowns under the symbol y^z. The differential equations (2), (3), (4), (17), (18), and (19) can then be written

$$G^M(y^z, y^z{}_{,t}, x^t) = 0, \tag{20}$$

where the G^M are linear in the first derivatives of the y^z, linear in the x^t (the determined part of the standard coordinates) and at most quadratic in the y^z themselves.

This is precisely the form of differential equation for which a thorough analysis is available.[9] The result of this analysis will be to show that by an algebraic operation on (20), the functions y^z can be replaced by new functions, which we call the special invariant functions, which can be assigned arbitrarily. These special invariant functions can be obtained from the y^z by a finite number of differentiations and, conversely, the analytic expressions for the y^z are uniquely determined from those of the special invariant functions.

The first step is to assign an order to the set of derivatives of the y^z. Assume that an order has been placed on y^z themselves. We say that $\partial^{n_0+\cdots+n_3}y^z/(\partial x^0)^{n_0}\cdots(\partial x^3)^{n_3}$ *precedes* $\partial^{m_0+\cdots+m_3}y^w/(\partial x^0)^{m_0}\cdots(\partial x^3)^{m_3}$ if $m_0+\cdots+m_3 > n_0+\cdots+n_3$, or if $m_0+\cdots+m_3 = n_0+\cdots+n_3$ and $z < w$, or if $z = w$ and $n_0+\cdots+n_3 = m_0+\cdots+m_3$ and the sequence $n_0\cdots n_3$ precedes $m_0\cdots m_3$ lexicographically. Next we assume that each of the equations (20) are solved for their highest derivatives in terms of all others. The resulting equations are then differentiated and the results written in the same form, i.e., solved for their highest derivatives. If more than one expression is obtained for the same derivative, the equality of the two expressions is added as another condition. The process is then iterated, the equations being solved for highest derivatives, and duplications eliminated. From a basic theorem of Riquier[9] this procedure must terminate after a finite number of steps, either in inconsistent equations or in equations for which further differentiations do not impose essentially new conditions. Let us now assume that this has been completed and that the resultant equations are

$$\partial^{n_0+\cdots+n_3}y^z/(\partial x^0)^{n_0}\cdots(\partial x^3)^{n_3} = H^{n_0\cdots n_3,z} \tag{21}$$

in which the H's on the right side depend only on the x^t and on the derivatives of the y's of rank less than that of the derivative appearing on the left side, and in which no derivative of a quantity on the left side of one equations occurs in the right side of another.

If the equations are consistent, the y^z can then be replaced by an equivalent set of arbitrary functions as follows: for fixed z, let $\{n^1_i\}\cdots\{n^v_i\}$ be the set of sequences of derivatives of y^z occurring on the left side of the equations (21). We now decompose y^z into a finite number of terms

$$y^z = f^z_0 + x^0 f^z_1 + \cdots + (x^0)^a f^z_a, \tag{22}$$

where a is the largest of the integers $n^1_0\cdots n^v_0$, and where $f^z_0\cdots f^z_{a-1}$ do not depend on x^0 (f^z_a may depend on x^0). The operation is then repeated on each of the f^z_i in terms of x^1, so that

$$f^z_i = f^z_{i0} + x^1 f^z_{i1} + \cdots + (x^1)^b f^z_{ib}, \tag{23}$$

where now b is the largest number of those n^j_1 for which $n^j_0 > i$. The process is then iterated and the final result can be written

$$y^z = \sum (x^0)^r\cdots(x^3)^s g^z_{r\ldots s} + \sum (x^0)^u\cdots(x^3)^v h^z_{u\ldots v}, \tag{24}$$

where each of the sums is over a finite number of terms and the functions g and h may not depend on all the variables. The division into two separate sums is made on the following basis. The second sum, involving the h's, contains those monomials which can be written as multiples of a monomial in the x's in which the exponents of x^i are one of the sequences $\{n^1_i\}\cdots\{n^v_i\}$. The first sum consists of the remainder of terms. Notice that the g's and h's can be obtained from a given y^z by a finite number of differentiations and algebraic steps. Finally, if the sequence $\{n^1_i\}\cdots\{n^v_i\}$ is vacuous, that is, if y^z itself and none of its derivatives are solved for in the equations, then $r = \cdots = s = 0$ and $h^z = 0$.

A straightforward consideration[9] of the differential equations then shows that *the g's can be arbitrarily chosen (apart from analyticity considerations) and the*

[9] J. Ritt, *Algebraic Differential Equations* (American Mathematical Society Colloquium Publications, New York, 1932), Vol. XIV, especially Chap. IX and X.

h's, and thus y's, are then uniquely determined, their analytic expression being obtainable from that of the g's. Thus the g's provide a complete and unique representation of the solutions to (21). By referring the y^s back to the ω_i^a, ω_{ai}^b, and F_Q it is then possible to determine which of the g's have the property of changing some of the F_Q when they are changed and which do not, i.e., those which only change the ω_i^a, ω_{ai}^b. Without loss of generality, then, we could prescribe some definite choice (e.g., zero) for these latter g's, yielding a further, and final, determination of the canonical frames and coordinate systems. The remaining g's are then called the *special invariant functions.*

IV. CONCLUSION

In summary, then, we have given an explicit prescription for obtaining a complete, one-to-one, and invariant classification and representation of the local, analytic metric and matter fields by a finite number of algebraic steps. The special invariant functions themselves can be chosen arbitrarily and can be obtained from the usual metric and matter fields by a finite number of algebraic steps and differentiations.

Clearly the amount of algebra required to make this scheme explicit, even for the empty-space case, is very formidable. However, the procedure would essentially be a trial and error one in which all possible types are tried, leading to all possible combinations of functions F_Q and variables x^t substituted for the Riemann components and derivatives in (2), (3), (17), (18), (19). The resulting equations would then be differentiated and tested for consistency and the special invariant functions picked out, essentially a repetitive-type task. Since the equations are at most quadratic in dependent and linear in the independent variables, a digital computer might be taught how to substitute all possible types to give all possiblitities for (15), then differentiate them and test for consistency, by arranging them as matrices. The feasibility of such a program is currently being studied.

In addition, there are many other possibilities. Can the classification procedure be better handled in the holonomy group approach?[10] Can the entire problem be given more succinct and satisfactory treatment in the bundle theory of connections? What about global extensions and matching of types across boundaries? Can this procedure be applied to give an effective test for whether or not a matter theory can be "geometrized" in the sense of Misner and Wheeler?[11] What is the significance of the "superposition" process of adding special invariant functions? What are the representations and significance of geodesics in this approach?

[10] J. Schell, J. Math. Phys. **2**, 202 (1961).

ACKNOWLEDGMENTS

Finally, the author wishes to express his gratitude to the National Science Foundation for its support of this work.

APPENDIX

In this Appendix we consider the problems arising in Sec. II concerned with the establishment of a standard form for each of the components of a "vector" in a direct sum of product spaces under the induced representation of the Lorentz group. It will not be convenient here to discuss the effects of transforming the basis as leaving the vector fixed but changing its components. Rather, we use the fully equivalent, "active," picture of the transformation as changing the vector but leaving the base fixed. Thus, for $g \in R^N(L)$, rather than considering $v_\alpha^{(N)}$ as the components of the same vector relative to a new frame, we consider them as components of a new vector relative to the same frame, which, with an unimportant abuse of notation, we denote by $gv^{(N)}$ so that $gv^{(N)} \equiv \bar{v}_\alpha^{(N)}\omega^\alpha$. Further, for simplicity, we consider only one of the terms in this direct sum, that is, only the direct product of a fixed number of terms, m. The extension to the general case with a sum over m is immediate.

Thus, let the $\omega^{\alpha_1} \cdots \omega^{\alpha_n}$ of (6) each be product of m of the ω^a. For convenience, let us relabel these to simply be $\omega^1 \cdots \omega^n$. Letting A run from 1 to n, A' from $n+1$ to 4^m, and x from 1 to 4^m, we then have that $\omega^{A'}$ constitute the complementary set of base vectors in this tensor space of rank n. Further let V be the space spanned by ω^A and W the space spanned by the $\omega^{A'}$. Hence (dropping the superscript N) we have that $v \in V$ with no v_A equal to zero. The result we must prove in this Appendix can then be stated

Theorem: If $v \in V$, if $H \equiv \{g: gv \in V; g \in R(L)\}$, and if no component of gv is zero for $g \in H$, then as g ranges over H, any particular component of gv is either constant, reaches the value 1 or -1, or achieves a maximum.

First, note that since H is algebraically defined it has at most a finite number of connected components. If we can prove this result for one component the extension to the full H itself is immediate. Thus,

[11] C. Misner and J. Wheeler, Ann. Phys. (N. Y.) **2**, 525 (1957).

for the remainder, we will consider only the component of the identity of H, which we will denote by H_0. Since H_0 is arcwise connected, the variation of the components of gv for $g \in H_0$ can be studied by considering the variation of gv for g along all possible paths in H_0 passing through the origin. The differential equations for the components of gv along such a path would then be

$$dv_x(t)/dt = X^y_x(t)v_y(t), \tag{A1}$$

where, since along the path $v(t) \in V$, $v_{A'} = 0$. Thus as far as the variation in the components of v_x are concerned we need only consider those paths for which

$$X^y_x(t) = \begin{bmatrix} M(t) & 0 \\ 0 & 0 \end{bmatrix}; \quad M(t) = n \times n \quad \text{matrix.} \tag{A2}$$

In other words, as far as variations of v_x are concerned we need only consider those paths in H_0 for which the tangent vector is in the subspace spanned by the Lie algebra of the connected group leaving V invariant. We may assume that H_0 has been restricted to this group.

Hence, we need only consider those $X^y_x(t)$ of the form (A2) where

$$M^B_A(t) = L^{b_1}_{a_1}\delta^{b_2}_{a_2} \cdots \delta^{b_m}_{a_m} + \cdots + \delta^{b_1}_{a_1} \cdots \delta^{b_{m-1}}_{a_{m-1}} L^{b_m}_{a_m}, \tag{A3}$$

where the sequences $a_1 \cdots a_m$ and $b_1 \cdots b_m$ correspond to the indices A and B respectively, and where the L^b_a represent constant elements of the basic Lorentz Lie algebra. Thus our problem is reduced to finding which L^b_a will yield an element of the form (A2). To this end let $I_1(a_2 \cdots a_m)$ be the set of indices, a, for which

$$\omega^a \otimes \omega^{a_2} \otimes \cdots \otimes \omega^{a_m} \in V \tag{A4}$$

and define similarly $I_2(a_1, a_3 \cdots a_m)$ etc. Thus the only admissible L^b_a for (A3) will be those for which

$$L^b_a = 0 \tag{A5}$$

unless

$$\begin{aligned} &b \in I_1(a_2 \cdots a_m) \text{ implies } a \in I_1(a_2 \cdots a_m) \\ &b \in I_2(a_1, a_3 \cdots a_m) \\ &\qquad \text{implies } a \in I_2(a_1, a_3 \cdots a_m), \text{ etc.,} \end{aligned} \tag{A6}$$

for all $a_1 \cdots a_m$. In other words, the only L^b_a in (A3) we need consider are those representing Lorentz or spatial rotations in planes described by pairs of indices (a, b), satisfying (A6).

If the set of such admissible L's constitutes a subalgebra of the compact group of pure spatial rotations, then the desired result is obtained since in this case each component of gv will each achieve a maximum value, if it does not pass through 1 or -1.

However, if infinitesimal Lorentz velocity transformations are among the admissible L's, the group is not compact and it is not immediately obvious that the list of alternatives for gv in the theorem is exhaustive. Actually, the proof following will show that in this case each component of gv is either constant or reaches 1 or -1.

For this purpose, let us assume that the indices are so ordered that the pair (0, 1) is among those satisfying (A6). Thus we have available the one-dimensional subgroup of velocity transformations along the first space axis. We will now show that under the assumptions of the theorem, as g ranges over this group, each component of gv is either constant or reaches 1 or -1.

First assume that all irrelevant indices (i.e., those other than 0, 1) have been deleted, and let $i = 0, 1$. Next, writing the components of v in expanded, tensor form, $v_{i\ldots j}$, let us relate these components to the null components defined as follows:

$$v_{i\ldots j} = C^{\alpha\cdots\beta}_{i\cdots j} w_{\alpha\ldots\beta}; \qquad \alpha = p, q, \tag{A7}$$

where the C's are defined inductively,

$$C^\alpha_i = (-1)^{\sigma_\alpha(i+1)}; \qquad \sigma_\alpha \equiv \begin{cases} 0 & \text{if } \alpha = p \\ 1 & \text{if } \alpha = q \end{cases}, \tag{A8}$$

and

$$C^{\alpha\beta\cdots\mu}_{ij\cdots k} = (-1)^{\sigma_\alpha(i+1)} C^{\beta\cdots\mu}_{j\cdots k}. \tag{A9}$$

A simple analysis of (A1) in this case shows that the variations of the w's along the path will be of the form

$$w_{\alpha\ldots\beta}(t) = \exp(r_{\alpha\ldots\beta}t) w_{\alpha\ldots\beta}(0); \quad \text{(no sum)}. \tag{A10}$$

$r_{\alpha\ldots\beta} \equiv$ number of p's among $\alpha \cdots \beta$ minus the number of q's among $\alpha \cdots \beta$. In terms of the v's this becomes

$$v_{i\ldots j}(t) = C^{\alpha\cdots\beta}_{i\cdots j} w_{\alpha\ldots\beta}(0) \exp(r_{\alpha\ldots\beta}t). \tag{A11}$$

Thus, as was expected, the $v(t)$ are polynomials in exponentials of t. Now, using the hypothesis of the theorem that none of the $v_{i\ldots j}(t)$ can be zero, we show that actually only one power of exp (t) will contribute to the right side of (A11).

Let r_1 be the minimum and r_2 the maximum value of $r_{\alpha\ldots\beta}$ for which the coefficient of exp $(r_{\alpha\ldots\beta}t)$ in the right side of (A11) is not zero. Thus as $t \to \pm\infty$ the sign of $v_{i\ldots j}(t)$ approaches that of the coefficient of exp $(r_2 t)$ or that of the coefficient of exp $(r_1 t)$ respectively. First, assume that $r_1 \neq r_2$. We show

that this implies that the signs of $v_{i\ldots j}(t)$ as $t \to +\infty$ and $-\infty$ are opposite, and that hence $v_{i\ldots j}$ is zero for some finite value of t, contradicting the hypothesis. What we consider is the product of the asymptotic parts of $v_{i\ldots j}(t)$ for $t \to \pm\infty$. Such products will clearly be of the form

$$X_{i\ldots j} = C_{i}^{\alpha\cdots\beta} C_{i\ldots j}^{\mu\cdots\nu} Y_{\alpha\cdots\beta,\mu\cdots\nu} \tag{A12}$$

in which for each nonzero term in the sum on the right the number of indices $\alpha \cdots \beta$ equal to q is *not* equal to the number of indices $\mu \cdots \nu$ equal to q, since we are assuming $r_1 \neq r_2$.

We now show by induction on m (the number of indices $i \cdots j$) that if some $X_{i\ldots j}$ is not equal to zero then there is another set of indices, $i' \cdots j'$, with

$$X_{i\ldots j} X_{i'\ldots j'} < 0, \tag{A13}$$

so that at least one of them is negative. This is clearly true for $m = 1$. If $m > 1$, use (A9) to write (A12) as

$$X_{ik\ldots j} = C_{k}^{\gamma\cdots\beta} C_{k\ldots i}^{\rho\cdots\nu} (Y_{\rho\gamma\cdots\beta,\rho,\rho\cdots\nu} + Y_{\alpha\gamma\cdots\beta,\alpha,\rho\cdots\nu}) + (-1)^{i+1} Z_{k\ldots j} \tag{A14}$$

or, grouping the first terms under the symbol $X'_{k\ldots j}$,

$$X_{ik\ldots j} = X'_{k\ldots j} + (-1)^{i+1} Z_{k\ldots j}. \tag{A15}$$

From the induction hypothesis there is another set of indices, $k' \cdots j'$, with $X'_{k\ldots j} X'_{k'\ldots j'} < 0$. Simple algebra then shows that not all possible products of two of the following, $X_{0k\ldots j}$, $X_{1k\ldots j}$, $X_{0k'\ldots j'}$, $X_{1k'\ldots j'}$ can be positive, so at least one must be negative, and the induction process is established for all m.

In terms of the v's this means that if $r_1 \neq r_2$, the product of the signs of the asymptotic parts of $v_{i\ldots j}$ for plus and minus infinity is negative for at least one set of indices. Since this would imply that $v_{i\ldots j}(t) = 0$ for some finite t, contradicting the hypothesis of the minimal nature of V, we thus must have $r_1 = r_2$.

Hence, the behavior of each component under this one-dimensional group is that of (A10) in which *only one power of the exp* (t) *appears.* If this is zero, this component is constant under these transformations. If it is not zero, this component will achieve the value 1 if it is positive or -1 if it is negative, for one value of t.

This thus shows that the possibilities listed in the theorem are exhaustive and completes the proof.

Solution of Einstein's Field Equations for a Rotating, Stationary, and Dust-Filled Universe*

JAMES P. WRIGHT
Mathematics Research Center, University of Wisconsin, Madison, Wisconsin
(Received 25 February 1964)

The solution of Einstein's field equations, $R_{ij} - \frac{1}{2}g_{ij}R = 8\pi\kappa\rho u_i u_j + \lambda g_{ij}$, for a line element of the form $ds^2 = (dx^0)^2 - \delta^2(x^1)(dx^1)^2 + 2\beta(x^1)dx^0dx^2 - \gamma^2(x^1)(dx^2)^2 - \alpha^2(x^1)(dx^3)^2$ is found. The density, ρ, may be a function of position, and the cosmological constant λ is not necessary in order to have a finite density. The solution reduces to that of Gödel if the variable α is constant. If the requirement for an empty universe is made ($R_{ij} = 0$), the solution is conformally flat. The characteristics of the conformal curvature tensor are also obtained.

ONE would like to have a general metric satisfying Einstein's field equations

$$R_{ij} - \tfrac{1}{2}g_{ij}R = 8\pi\kappa\rho u_i u_j + \lambda g_{ij}, \qquad (1)$$

and which allows a geodesic congruence which has expansion, rotation, shear, and an arbitrary choice (including zero) for the cosmological constant λ.[1]

In the process of constructing general models, it is useful to construct more restricted models to gain insight into the problem. A relatively simple metric which satisfies all of Eqs. (1) has been constructed. The model has zero expansion and shear, but has nonvanishing rotation and density whether or not the cosmological constant vanishes. A brief description of the solution is as follows.

The line element which has these properties is first expressed in the form

$$ds^2 = (dx^0)^2 - (dx^1)^2 + 2\beta\, dx^0\, dx^2 - \gamma^2(dx^2)^2 - \alpha^2(dx^3)^2, \qquad (2)$$

where α, β, and γ are functions which depend only on the variable x^1. This choice gives a possible nonzero "rotation" vector

$$\omega^i = (-g)^{-1/2}\epsilon^{ijkl}u_j\frac{\partial u_k}{\partial x^l}, \qquad \epsilon^{0123} = +1, \qquad (3)$$

with zero expansion and shear. The quantities R_{00}, R_{11}, R_{22}, and R_{33} do not necessarily vanish identically. The matrices of the covariant g_{ij} and the contravariant g^{ij} are

$$\begin{Vmatrix} 1 & 0 & \beta & 0 \\ 0 & -1 & 0 & 0 \\ \beta & 0 & -\gamma^2 & 0 \\ 0 & 0 & 0 & -\alpha^2 \end{Vmatrix}$$

and

$$\begin{Vmatrix} \gamma^2/(\beta^2+\gamma^2) & 0 & \beta/(\beta^2+\gamma^2) & 0 \\ 0 & -1 & 0 & 0 \\ \beta/(\beta^2+\gamma^2) & 0 & -1/(\beta^2+\gamma^2) & 0 \\ 0 & 0 & 0 & -\alpha^{-2} \end{Vmatrix}, \qquad (4)$$

respectively. The determinant, $|g_{ij}|$, of the metric is $g = -\alpha^2(\beta^2 + \gamma^2)$.

The nonvanishing Christoffel symbols[2] of the first and second kind may be computed from the derivatives of the metric tensor and are

$$\Gamma_{0,12} = \tfrac{1}{2}\beta_1, \qquad \Gamma_{1,02} = -\tfrac{1}{2}\beta_1,$$
$$\Gamma_{1,22} = \gamma\gamma_1, \qquad \Gamma_{2,12} = -\gamma\gamma_1,$$
$$\Gamma_{2,01} = \tfrac{1}{2}\beta_1, \qquad \Gamma_{1,33} = \alpha\alpha_1,$$

and

$$\Gamma_{3,13} = -\alpha\alpha_1; \qquad (5)$$
$$\Gamma^0_{01} = \tfrac{1}{2}\beta\beta_1/(\beta^2+\gamma^2),$$
$$\Gamma^0_{12} = \tfrac{1}{2}\gamma(\gamma\beta_1 - 2\beta\gamma_1)/(\beta^2+\gamma^2),$$
$$\Gamma^1_{02} = \tfrac{1}{2}\beta_1,$$
$$\Gamma^1_{22} = -\gamma\gamma_1,$$
$$\Gamma^1_{33} = -\alpha\alpha_1,$$
$$\Gamma^2_{01} = -\tfrac{1}{2}\beta_1/(\beta^2+\gamma^2),$$
$$\Gamma^2_{12} = \tfrac{1}{2}(\beta\beta_1 + 2\gamma\gamma_1)/(\beta^2+\gamma^2),$$

and

$$\Gamma^3_{13} = \alpha_1/\alpha.$$

The subscripts denote the derivative with respect to the variable x^1.

The components of the contracted Riemann-Christoffel tensor (summation convention used),

$$R_{ij} = \frac{\partial\Gamma^\kappa_{ij}}{\partial x^\kappa} - \frac{1}{2}\frac{\partial^2(\log g)}{\partial x^i\,\partial x^j} + \tfrac{1}{2}\Gamma^\kappa_{ij}\frac{\partial(\log g)}{\partial x^\kappa} - \Gamma^\kappa_{mi}\Gamma^m_{\kappa j}, \qquad (6)$$

* Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No. DA-11-022-ORD-2059.

[1] O. Heckman and E. Schucking, in *Gravitation—An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962), pp. 438–469.

[2] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1926), notes that $\Gamma_{i,jk} = [jk, i]$ and $\Gamma^i_{jk} = \left\{ {i \atop jk} \right\}$.

may now be calculated. The ten independent components are

$$R_{00} = \tfrac{1}{2}\beta_1^2/(\beta^2+\gamma^2),$$
$$R_{02} = \tfrac{1}{2}[\alpha_1\beta_1\alpha^{-1} + \beta_{11} - \beta_1\gamma\gamma_1/(\beta^2+\gamma^2)],$$
$$R_{01} = R_{03} = R_{12} = R_{13} = R_{23} = 0,$$
$$R_{11} = -\alpha_{11}\alpha^{-1} - \tfrac{1}{2}(\beta_1^2 + 2\gamma_1^2 + 2\beta\beta_{11} + 2\gamma\gamma_{11})/(\beta^2+\gamma^2) + (\beta\beta_1+\gamma\gamma_1)^2/(\beta^2+\gamma^2)^2, \quad (7)$$
$$R_{22} = -\gamma\gamma_{11} - \gamma\gamma_1\alpha_1\alpha^{-1} - \tfrac{1}{2}(2\beta^2\gamma_1^2 + \gamma^2\beta_1^2 - 2\beta\gamma\beta_1\gamma_1)/(\beta^2+\gamma^2),$$

and

$$R_{33} = -\alpha\alpha_{11} - \alpha\alpha_1(\beta\beta_1+\gamma\gamma_1)/(\beta^2+\gamma^2).$$

We may solve for the functions α, β, and γ satisfying Eqs. (1) in the co-moving coordinate system $u^i = \delta_0^i (u_i = g_{ik}u^k = g_{0i})$. The equations which must be satisfied are (omitting the equations which vanish identically):

$$R_{00} - \tfrac{1}{2}R = \rho' + \lambda, \quad R_{02} - \tfrac{1}{2}\beta R = \beta\rho' + \beta\lambda,$$
$$R_{11} + \tfrac{1}{2}R = -\lambda, \quad R_{22} + \tfrac{1}{2}\gamma^2 R = \beta^2\rho' - \gamma^2\lambda, \quad (8)$$

and

$$R_{33} + \tfrac{1}{2}\alpha^2 R = -\alpha^2\lambda, \quad \text{where} \quad \rho' = 8\pi\kappa\rho.$$

These equations give rise to the following relations

$$\beta R_{00} - R_{02} = 0, \quad (9)$$
$$\beta R_{02} - R_{22} + (\beta^2+\gamma^2)R_{11} = 0, \quad (10)$$
$$R_{33} - \alpha^2 R_{11} = 0, \quad (11)$$
$$R = g^{ij}R_{ij} = R_{00} - 3\alpha^{-2}R_{33}, \quad (12)$$
$$\rho' = R_{00} + \alpha^{-2}R_{33}, \quad (13)$$

and

$$-\lambda = \tfrac{1}{2}(R_{00} - \alpha^{-2}R_{33}).$$

The functional form of the R_{ij}'s may be substituted into the relations (9), (10), and (11) to obtain the relations

$$\beta_1(\beta\beta_1+\gamma\gamma_1)/(\beta^2+\gamma^2) - \beta_{11} - \beta_1\alpha_1\alpha^{-1} = 0, \quad (14)$$

and

$$\alpha_1(\beta\beta_1+\gamma\gamma_1)/(\beta^2+\gamma^2) - \alpha_{11} = 0. \quad (15)$$

The two relations (14) and (15) may then be integrated to yield

$$\beta_1^2\alpha^2 = A^2(\beta^2+\gamma^2) \quad (\text{if } \beta_1 \neq 0), \quad (16)$$

and

$$\alpha_1^2 = B^2(\beta^2+\gamma^2) \quad (\text{if } \alpha_1 \neq 0), \quad (17)$$

where A and B are constants of integration.

The use of Eqs. (7), (13), (16), and (17) enables us to write down the equations

$$\beta_1^2 = (A^2/B^2)\alpha_1^2/\alpha^2 \quad (18)$$

and

$$2\alpha_{11}/\alpha + \tfrac{1}{2}A^2/\alpha^2 = -2\lambda, \quad (19)$$

which are integrated to give

$$\beta = \pm(A/B)\ln\alpha + C \quad (20)$$

and

$$\alpha_1^2 = D - \lambda\alpha^2 - \tfrac{1}{2}A^2\log\alpha \quad (\lambda = \text{const}), \quad (21)$$

where C and D are integration constants.

The quantities of most interest, ρ' and ω^i, are [from Eqs. (3), (7), (13), and (16)]

$$\rho' = A^2/\alpha^2 + 2\lambda = \tfrac{1}{2}A^2/\alpha^2 - 2\alpha_{11}/\alpha. \quad (22)$$
$$\omega^3 = \alpha^{-1}\beta_1/(\beta^2+\gamma^2)^{\frac{1}{2}} = \pm A/\alpha^2,$$
$$\text{and} \quad |\omega|^2 = A^2/\alpha^2. \quad (23)$$

We see from Eq. (22) that the density, ρ', need not depend on the cosmological constant λ, unless α is constant, which would require $\lambda = -\frac{1}{4}A^2/\alpha^2 = -\frac{1}{2}\rho'$.

In order to put the results in a more convenient form, a change in variables may be taken such that $\alpha = \exp[-y^2]$; the important quantities are then

$$ds^2 = (dx^0)^2 + 2\beta\,dx^0\,dx^2 - \delta^2\,dy^2 - \gamma^2(dx^2)^2 - \alpha^2(dx^3)^2, \quad (24)$$
$$\alpha = \exp[-y^2], \quad \beta = \pm(A/B)y^2 + C, \quad (25)$$
$$\gamma^2 = \tfrac{1}{2}(A/B)^2y^2 - (\lambda/B^2)\exp[-2y^2] + D/B^2 - [(A/B)y^2 + C]^2, \quad (26)$$
$$\delta^2 = 4y^2 \times \exp[-2y^2]/\{D + \tfrac{1}{2}A^2y^2 - \lambda\exp[-2y^2]\}, \quad (27)$$
$$\rho' = A^2\exp[2y^2] + 2\lambda, \quad (28)$$

and

$$\omega^3 = \pm A\exp[2y^2]. \quad (29)$$

This solution [expressed in Eqs. (24)–(29)], which satisfies all of Eqs. (1), does not require λ to have a specific value if α is not constant.

For α constant (let $\alpha = 1$), we may solve the general equations by using Eq. (16) and the fact that $R_{11} = R_{33} = 0$. The result is that β must satisfy the equation

$$\beta_{111} - \tfrac{1}{2}A^2\beta_1 = 0,$$

or

$$\beta_{11} - \tfrac{1}{2}A^2(\beta - d) = 0, \quad (30)$$

or

$$(\beta - d)_{11} - \tfrac{1}{2}A^2(\beta - d) = 0.$$

Equation (30) is readily integrated, and the results are

$$\alpha = 1, \quad \beta = b[\exp(ax^1) + c\exp(-ax^1)] + d,$$
$$\gamma^2 = \tfrac{1}{2}b^2[\exp(ax_1) - c\exp(-ax_1)]^2 - \beta^2. \quad (31)$$
$$\rho' = R = -2\lambda = a^2,$$

and $\omega^3 = \sqrt{2}a$, where $a^2(=\frac{1}{2}A^2)$, b, c, and d are integration constants. This solution is the general form of Gödel's solution.[3] The first form of Gödel's solution is equivalent to letting $a = b = 1$ and $c = d = 0$. The properties of Gödel's solution are easier to understand if we study this general solution.

A solution of Eqs. (1) for the line element (2) with the conditions $R_{ij} = 0$ also requires that the Riemann-Christoffel tensor $R^i_{jkl} = 0$, and the space is conformally flat.

The Weyl curvature tensor[2]

$$C^{ij}_{kl} = R^{ij}_{kl} + \tfrac{1}{2}(\delta^i_k R^j_l - \delta^i_l R^j_k) + \tfrac{1}{2}(\delta^j_l R^i_k - \delta^j_k R^i_l) + \tfrac{1}{6}R(\delta^i_l \delta^j_k - \delta^i_k \delta^j_l) \quad (32)$$

is useful in the classification[4] of the different solutions of Einstein's field equations. The classification corresponds to the Petrov[5,6] classification for vacuum field solutions since in that case $C^{ij}{}_{kl} = R^{ij}{}_{kl}$. The Weyl curvature tensor has the properties

$$C^{ij}_{kl} = 0, \qquad C^i_{jkl} + C^i_{klj} + C^i_{ljk} = 0, \quad (33)$$

and

$$C^{ij}_{kl} = C^{ij}_{lk} = -C^{ij}_{lk} = -C^{ij}_{kl} .$$

The symmetry properties allow us to write the tensor as a six-by-six matrix C^α_β, where the labels α and β take the values

01–1, 02–2, 03–3, 23–4, 31–5, and 12–6.

The eigenvalues of this matrix are found by solving the equation

$$|C^\alpha_\beta - \lambda I| = 0. \quad (34)$$

The sum of the eigenvalues will be zero (from $C^{ij}_{kj} = 0$).

The nonvanishing components of the Weyl curvature tensor for the line element (24) are

$$\begin{aligned} C^{01}_{01} &= C^{23}_{23} = C^1_1 = C^4_4 \\ &= -\tfrac{1}{12}A^2/\alpha^2 + \tfrac{1}{2}\alpha_1\beta\beta_1/[\alpha(\beta^2+\gamma^2)], \\ C^{01}_{12} &= C^{03}_{23} = C^1_6 = C^3_4 \\ &= \tfrac{1}{4}\beta A^2/\alpha^2 - \tfrac{1}{2}\alpha_1\beta_1(2\beta^2+\gamma^2)/[\alpha(\beta^2+\gamma^2)], \\ C^{12}_{01} &= C^{23}_{03} = C^6_1 = C^4_3 = \tfrac{1}{2}\alpha_1\beta_1/[\alpha(\beta^2+\gamma^2)], \\ C^{02}_{02} &= C^{31}_{31} = C^2_2 = C^5_5 = -\tfrac{1}{12}A^2/\alpha^2, \end{aligned} \quad (35)$$

and

$$\begin{aligned} C^{03}_{03} &= C^{12}_{12} = C^3_3 = C^6_6 \\ &= \tfrac{1}{6}A^2/\alpha^2 - \tfrac{1}{2}\alpha_1\beta\beta_1/[\alpha(\beta^2+\gamma^2)]. \end{aligned}$$

The eigenvalues of (34) are therefore

$$\lambda_1 = \lambda_4 = \tfrac{1}{2}(A^2/\alpha^2)[\tfrac{1}{12} + (\tfrac{1}{16} - \alpha_1^2/A^2)^{\frac{1}{2}}],$$
$$\lambda_2 = \lambda_5 = -\tfrac{1}{12}A^2/\alpha^2, \quad (36)$$

and

$$\lambda_3 = \lambda_6 = \tfrac{1}{2}(A^2/\alpha^2)[\tfrac{1}{12} - (\tfrac{1}{16} - \alpha_1^2/A^2)^{\frac{1}{2}}],$$

where the quantity α_1^2 is determined by Eq. (21).

The eigenvalues λ_1 and λ_3 may be taken as the independent characteristics of the Weyl curvature tensor since $\lambda_1 + \lambda_3 = -\lambda_2$. The characteristics corresponding to Gödel's solution are obtained by letting α be constant ($\alpha = 1$). The result is that $\lambda_1 = \frac{1}{6}A^2$ and $\lambda_3 = -\frac{1}{12}A^2$. The classification of the solution of Einstein's equations for the line element (24) would correspond to the Petrov type 1.

If we interpret y as a radial coordinate, x^2 as an angular coordinate, x^3 as a z coordinate, and x^0 as a time coordinate, the line element (24) may be written in the form

$$ds^2 = dt^2 + 2\beta' r\, dt\, d\varphi - (\delta')^2\, dr^2 - (\gamma')^2 r^2\, d\varphi^2 - (\alpha')^2\, dz^2,$$

with

$$\rho' = A^2 \exp[2r^2] + 2\lambda$$

and

$$\omega^3 = \pm A \exp[2r^2]. \quad (37)$$

The line element in this form shows the cylindrical symmetry about the z axis. The rotation vector and the density are functions of the distance, r, from the z axis. The Gödel universe also has cylindrical symmetry, but the rotation vector and the density are constant.

The line element (24) has at least three Killing vectors, one of which must be timelike. The Killing vectors are $\xi_{(1)}{}^i = \delta^i_0$, $\xi_{(2)}{}^i = \delta^i_2$, and $\xi^i_{(3)} = \delta^i_3$. The space admits the following transformations into itself:

$$\text{(I)} \quad x^0 = x'^0 + a, \quad x^i = x'^i, \ i \neq 0; \qquad \text{(II)} \quad x^2 = x'^2 + a, \quad x^i = x'^i, \ i \neq 2;$$

and

$$\text{(III)} \quad x^3 = x'^3 + a, \quad x^i = x'^i, \ i \neq 3,$$

where a is an arbitrary real number.

ACKNOWLEDGMENTS

The author wishes to thank S. Chandrasekhar and A. H. Taub for stimulating discussions; Robert Jastrow for the hospitality of the Goddard Institute for Space Studies, where part of this work was completed, and the hospitality of the Mathematics Research Center, where the work was completed.

[3] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).

[4] J. Ehlers and W. Kundt, *Gravitation: An Introduction To Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962) p. 49ff.

[5] A. Z. Petrov, *Recent Developments in General Relativity* (Pergamon Press, Inc., New York, 1962) p. 371ff.

[6] M. Landau and S. Lifshitz, *The Classical Theory of Fields*, (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962), 2nd. ed., pp. 305–307.

Method for Treating Fluctuation of Dynamical Variables*

GERALD L. JONES
University of Notre Dame, Notre Dame, Indiana
(Received 16 July 1964)

An upper bound, on the probability that a fluctuation of a given size will occur in given time interval, has been derived. The bound is useful because it is small for some cases of interest. The fluctuations of the kinetic energy in a canonical ensemble have been considered as an example.

I. INTRODUCTION

ONE of the problems in the Gibbsian or ensemble form of statistical mechanics is in the connection between the numbers the theory allows one to compute, which are the ensemble mean values of dynamical variables, and the numbers one actually measures in an experiment. In particular, one can ask what the time dependence of the average value of a variable implies about the time behavior of a single system in the ensemble. For the case of equilibrium ensembles, where the average value of most variables of interest is independent of time, the most famous attack on this question has been the ergodic theory, which attempts to prove that the ensemble average of a variable is the same as the time average of that variable for any single member of the ensemble, the physical rational being that it is the time average of a single system that is actually measured. For classical systems, which are the only kind we shall treat here, Khinchin's[1] version of the ergodic theory is probably the most satisfactory. Using Birkoff's theorem and Tschebysheff's inequality, Khinchin proves that, for large systems, the time average of certain types of variables (those with small dispersion) is very near the ensemble average, for the great majority of the systems in the ensemble.[2] This approach to the problem leaves open the question of fluctuations, that is, the question of how long one should expect to watch a given system from the ensemble before the variable of interest deviates from the ensemble mean value by a certain amount. Attempts to investigate this point usually involve the estimation of mean recurrence times of certain nonequilibrium states for various models.[3-5] These calculations, while illuminating, are extremely specialized and do not seem to provide any general method of approach.

Even if one regards the above considerations and related ones as being satisfactory for the equilibrium case, this approach is not suitable for nonequilibrium ensembles. In particular one would like to know, in the nonequilibrium case, that the behavior of an individual member of the ensemble is in some sense well represented over time intervals long compared to the characterestic times (relaxation times) of the physical situation. We shall present an approach here which is sort of in the spirit of Khinchin's treatment of the ergodic problem but which is equally suitable to equilibrium and nonequilibrium ensembles. In particular, we have found an upper bound for the probability of selecting a system, from the ensemble, which will undergo a fluctuation of a given size in a given time interval. The bound is a useful one because, for large systems and appropriate dynamical variables, it is small.

In Sec. II we give a precise statement of the problem and find the bound. In Sec. III we show why the bound is a useful one.

II. DERIVATION OF THE BOUND

We let x stand for a point in the phase space of the system (Γ space). Then $x = \{x_1, x_2, \cdots\}$, where the x_i are the individual coordinates and momenta of the system. The dynamics of the system are contained in the transformation T_t which takes the phase point x (at time 0) into the point $T_t x$ (at time t). By Liouville's theorem we know the Jacobian of this transformation is 1. A general dynamical variable is a function $A(x)$. We can indicate the time dependence of $A(x)$, due to the natural motion of the system, by the notation $A(x, t) = A(T_t x)$. If H is the Hamiltonian of the system then the equations of motion in Poisson bracket notation are

$$\partial A(x, t)/\partial t = \{A(x, t), H(x)\}. \quad (1)$$

The state of the system at $t = 0$ is given by a normalized probability density function $f(x)$ defined on Γ. If an event is represented by a set G then the

* Supported in part by the U. S. Atomic Energy Commission.

[1] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, Inc., New York, 1949).
[2] C. Truesdale, *Ergodic Theories* (Academic Press Inc., New York, 1961).
[3] M. Kac, Bull. Am. Math. Soc. 53, 1002 (1947).
[4] P. Hemmer, L. C. Maximon, and H. Wergeland, Phys. Rev. 111, 689 (1958).
[5] E. Montroll, *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1961), Vol. III.

probability of the event is the measure or volume of the set, $mG = \int_G f(x)dV$. The average value of a dynamical variable will be denoted by

$$\bar{A}(t) = \langle A(x, t)\rangle = \int_\Gamma A(x, t)f(x)\, dV. \qquad (2)$$

We shall use the standard notation $\{x: \cdots\}$ for the set of points x such that whatever condition stands to the right of the colon is satisfied.

We define a set $R(\alpha, \tau)$ where α and τ are fixed positive real numbers by

$$R(\alpha, \tau) = \{x : |A(x, t) - \bar{A}(t)| \geq \alpha \quad \text{for some} \quad 0 \leq t \leq \tau\}. \qquad (3)$$

This is the set of points for which the dynamical variable A deviates from its ensemble average by more than α at least once during the time interval 0 to τ. If the probability of this event, $P(\alpha, \tau) = mR(\alpha, \tau)$ is a small number for suitable choices of A, α and τ, then the ensemble mean behavior of A is a good picture of the actual behavior of A for most of the systems in the ensemble, at least over the time interval τ. We shall find a useful upper bound for $P(\alpha, \tau)$.

In the argument that follows we shall not maintain high standards of rigor but will keep things as simple and as geometrical as possible. Therefore without going into it any further, we shall simply assume that the distribution function $f(x)$ and the dynamical variable $A(x, t)$ are sufficiently well behaved in x and t to justify the required operations. This seems to be verifiable for most applications. We shall assume also that $A(x, t)$, $H(x)$, and $f(x)$ are such that $m\{x : |A(x, t)| = \infty\} = 0$ for all t.

We have directly from the definition of the set $R(\alpha, \tau)$ the following properties

$$dP(\alpha. \tau)/d\alpha \leq 0, \qquad (4a)$$

$$dP(\alpha, \tau)/d\tau \geq 0, \qquad (4b)$$

$$P(\alpha, \tau) \geq 0. \qquad (4c)$$

In order to bound $P(\alpha, \tau)$ we first look for a bound of $dP(\alpha, \tau)/d\tau$. To do this we introduce a new set

$$S(\alpha, t) = \{x: |A(x, t) - \bar{A}(t)| \geq \alpha\}. \qquad (5)$$

It is clear that $R(\alpha, \tau)$ is the union of the family of $S(\alpha, t)$ for all $t \geq 0$ and $t \leq \tau$. $P(\alpha, \tau)$, which is the volume of $R(\alpha, \tau)$ (using $f(x)\, dV$ as the volume element) is then the volume generated by the set $S(\alpha, t)$ as t varies from 0 to τ. We assume the sets $R(\alpha, \tau)$ and $S(\alpha, t)$ both have well-defined boundary surfaces which divide Γ into parts exterior to and interior to the sets, and such that to get from the interior to the exterior one must cross the boundary

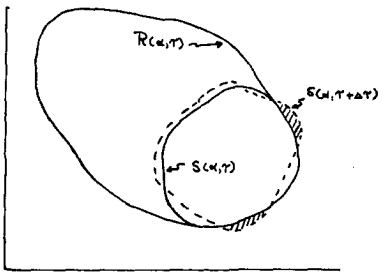

FIG. 1. Γ space.

surface. Now since $R(\alpha, \tau)$ is generated by the moving set $S(\alpha, t)$, it is clear that any increase of the size of $R(\alpha, \tau)$ in the time interval τ to $\tau + \Delta\tau$ must come from the region where the boundaries of $R(\alpha, \tau)$ and $S(\alpha, \tau)$ coincide. In Fig. 1 we indicate geometrically the situation, where we have shown the boundaries of the three sets $R(\alpha, \tau)$, $S(\alpha, \tau)$, and $S(\alpha, \tau + \Delta\tau)$. The increase in volume of $R(\alpha, \tau)$ in the time interval $\Delta\tau$ is

$$P(\alpha, \tau + \Delta\tau) - P(\alpha, \tau) = \Delta P(\alpha, \tau)$$

and is shown in Fig. 1 as the cross-hatched areas. We have somewhat oversimplified the geometry in Fig. 1 since the set $S(\alpha, \tau)$ is actually bounded by two distinct surfaces, $A(x, \tau) = \bar{A}(\tau) \pm \alpha$. Let us consider the contribution of the surface $A(x, \tau) = \bar{A}(\tau) + \alpha$ to $\Delta P(\alpha, \tau)$ first. If $\Delta\tau$ is very small we can write the volume element in the cross hatched region as an element dS of the surface of $S(\alpha, \tau)$ times the normal distance to the surface of $S(\alpha, \tau + \Delta\tau)$. If in time $\Delta\tau$ a point x on the first surface goes into a point $x + \Delta x$ on the second then we must have $\nabla A(x, \tau)\cdot\Delta x + [\partial A(x, \tau)/\partial\tau]\Delta\tau = [\partial\bar{A}(\tau)/\partial\tau]\Delta\tau$. The outward pointing normal to the surface is given by $-\nabla A(x, \tau)/|\nabla A(x, \tau)|$ and therefore the normal distance dn between the surfaces at the point x is given by

$$dn = -\frac{\nabla A}{|\nabla A|}\cdot\Delta x = \frac{1}{|\nabla A|}\frac{\partial(A - \bar{A})}{\partial\tau}\Delta\tau.$$

The increase in the volume from the neighborhood of x is then

$$f(x)\, dS\, dn = [f(x)\, |\nabla A|^{-1}\, \partial(A - \bar{A})/\partial\tau]\, \Delta\tau\, dS.$$

We must integrate this over that portion of the surface of $S(\alpha, \tau)$ which coincides with the boundary of $R(\alpha, \tau)$ and for which $\partial(A - \bar{A})/\partial\tau > 0$ [only these portions give an increase in the volume of $R(\alpha, \tau)$]. This region of integration, which is part of the surface $A(x, \tau) = \bar{A}(\tau) + \alpha$ we shall call G_1. The contribution from this surface to $\Delta P(\alpha, \tau)$ is then

$$\int_{G_1} f(x)[\partial(A - \bar{A})/\partial\tau]\, |\nabla A|^{-1}\, dS\, \Delta\tau.$$

By exactly the same kind of argument we get the contribution to $\Delta P(\alpha, \tau)$ from the surface $A(x, \tau) = \bar{A}(\tau) - \alpha$ to be

$$\int_{G_2} f(x)[-\partial(A - \bar{A})/\partial\tau]\ |\nabla A|^{-1}\ dS\ \Delta\tau$$

where G_2 is that part of the surface $A(x, \tau) = \bar{A}(\tau) - \alpha$ which coincides with the boundary of $R(\alpha, \tau)$ and for which $-\partial(A - \bar{A})/\partial\tau > 0$. The sum of these two contributions is the total change $\Delta P(\alpha, \tau)$. Letting $\Delta\tau \to 0$ we obtain

$$dP(\alpha, \tau)/d\tau = \int_{G_1} f(x)[\partial(A - \bar{A})/\partial\tau]\ |\nabla A|^{-1}\ dS + \int_{G_2} f(x)[-\partial(A - \bar{A})/\partial\tau]\ |\nabla A|^{-1}\ dS. \quad (5)$$

Since the first integral has $\partial(A - \bar{A})/\partial\tau > 0$ in G_1 and in the second $-\partial(A - \bar{A})/\partial\tau > 0$ in G_2, we can replace $\pm\partial(A - \bar{A})/\partial\tau$ by $|\partial(A - \bar{A})/\partial\tau|$ without changing the integrals. But the integrands are now positive and therefore we can only increase the integrals if we enlarge the regions of integration. Hence the inequality

$$dP(\alpha, \tau)/d\tau \leq \int_{A=\bar{A}+\alpha} f(x)\ |\partial(A - \bar{A})/\partial\tau|\ |\nabla A|^{-1}\ dS + \int_{A=\bar{A}-\alpha} f(x)\ |\partial(A - \bar{A})/\partial\tau|\ |\nabla A|^{-1}\ dS, \quad (6)$$

where the integration now goes over the entire surfaces $A(x, \tau) = \bar{A}(\tau) \pm \alpha$. If one integrates both sides of this expression a bound for $P(\alpha, \tau)$ is obtained, however it is not very useful since it involves surface integrals which are not convenient to evaluate. In order to modify this we shall need the following elementary theorem which relates an $(n - 1)$-dimensional surface integral to the derivative of an n-dimensional volume integral.

$$\int_{B(x)=a} g(x)\ |\nabla B|^{-1}\ dS = \frac{d}{dy}\int_{B(x)<y} g(x)\ dV|_{y=a}. \quad (7)$$

The integral on the left is over the surface defined by $B(x) = a$ and the integral on the right is over the volume where $B(x) < y$ and the derivative of the volume integral is evaluated at a. A brief discussion of (7) which is quite simple to prove is given in Ref. 1. Applying (7) to (6) we obtain

$$dP(\alpha, \tau)/d\tau \leq \frac{d}{dy}\int_{A(x,\tau)<y} f(x)\ |\partial(A - \bar{A})/\partial\tau|\ dV|_{y=\bar{A}+\alpha} + \frac{d}{dy}\int_{A(x,\tau)<y} f(x)\ |\partial(A - \bar{A})/\partial\tau|\ dV|_{y=\bar{A}-\alpha}. \quad (8)$$

Now let us integrate both sides of (8) over α, from α_1, to ∞. We obtain

$$\int_{\alpha_1}^{\infty} [dP(\alpha, \tau)/d\tau]\ d\alpha \leq \int_{A(x,\tau)<\infty} K(x, \tau)\ dV - \int_{A(x,\tau)<\bar{A}(\tau)+\alpha_1} K(x, \tau)\ dV + \int_{A(x,\tau)<\bar{A}(\tau)-\alpha_1} K(x, \tau)\ dV - \int_{A(x,\tau)<-\infty} K(x, \tau)\ dV,$$

where

$$K(x, \tau) = f(x)\ |\partial\{A(x, \tau) - \bar{A}(\tau)\}/\partial\tau|.$$

We now assume that the set where $A(x, \tau) < \infty$ differs from Γ by at most a set of zero volume and that the set where $A(x, \tau) < -\infty$ has zero volume. The previous formula becomes

$$\int_{\alpha_1}^{\infty} [dP(\alpha, \tau)/d\tau]\ d\alpha \leq \int_{\Gamma} K(x, \tau)\ dV - \int_{A<\bar{A}+\alpha_1} K(x, \tau)\ dV + \int_{A<\bar{A}-\alpha_1} K(x, \tau)\ dV.$$

If we combine the first two terms on the right we obtain

$$\int_{\alpha_1}^{\infty} [dP(\alpha, \tau)/d\tau]\ d\alpha \leq \int_{A>\bar{A}+\alpha_1} K(x, \tau)\ dV + \int_{A<\bar{A}-\alpha_1} K(x, \tau)\ dV, \quad (9)$$

or, combining the two regions of integration and putting in the explicit form of $K(x, \tau)$

$$\int_{\alpha_1}^{\infty} [dP(\alpha, \tau)/d\tau]\ d\alpha \leq \int_{|A(x,\tau)-\bar{A}(\tau)|>\alpha_1} |\partial\{A(x, \tau) - \bar{A}(\tau)\}/\partial\tau|\ f(x)\ dV. \quad (10)$$

The right-hand side can be put in more convenient form if we replace the time derivative terms by Poisson brackets (1). We can also transfer the time dependence into the function f, by the usual trick of changing the variable of integration from x to $T_\tau x$ and using Liouville's theorem. Equation (10) becomes

$$\int_{\alpha_1}^{\infty} [dP(\alpha, \tau)/d\tau]\ d\alpha \leq \int_{A|(x)-\bar{A}(\tau)|>\alpha_1} |\{A(x), H(x)\} - dA(\tau)/d\tau|\ f(x, \tau)\ dV. \quad (11)$$

To get the right-hand side in its final form, we multiply the integrand by $[A(x) - \bar{A}(\tau)]^2\alpha_1^{-2} > 1$ and then allow the integration over all of Γ.

$$\int_{\alpha_1}^{\infty} \frac{dP(\alpha, \tau)}{d\tau}\ d\alpha < \alpha_1^{-2}\int_{\Gamma} \left|\frac{\{A(x), H(x)\} - d\bar{A}(\tau)}{d\tau}\right| \times [A(x) - \bar{A}(\tau)]^2 f(x, \tau)\ dV. \quad (12)$$

We shall define

$$D(\tau) = \int_\Gamma |\{A(x), H(x)\} - d\bar{A}(\tau)/d\tau| \times [A(x) - \bar{A}(\tau)]^2 f(x, \tau)\, dV, \quad (13)$$
$$= \langle |\{A(x), H(x)\} - \langle \{A(x), H(x)\}\rangle| \times [A(x) - \langle A(x)\rangle]^2\rangle.$$

We shall see in the next section that it is possible to investigate $D(\tau)$ for some cases of interest. Now let us integrate (12) from 0 to τ to obtain (assuming that we can interchange the order of the integrations)

$$\int_{\alpha_1}^\infty P(\alpha, \tau)\, d\alpha \le \int_{\alpha_1}^\infty P(\alpha, 0)\, d\alpha + \alpha_1^{-2} \int_0^\tau D(\tau')\, d\tau'. \quad (14)$$

Now from the definition of P, R, and S we have

$$P(\alpha, 0) = mR(\alpha, 0) = mS(\alpha, 0) = \int_{|A(x)-\bar{A}(0)|>\alpha} f(x)\, dV,$$
$$\le \alpha^{-2} \int_\Gamma [A(x) - \bar{A}(0)]^2 f(x)\, dV.$$

If we set

$$C = \int_\Gamma [A(x) - \bar{A}(0)]^2 f(x)\, dV = \langle [A(x) - \langle A(x)\rangle]^2\rangle \quad (15)$$

then

$$P(\alpha, 0) \le \alpha^{-2} C. \quad (16)$$

One should note that the averages in the definition of $D(\tau)$ use $f(x, \tau)$ while those in C use $f(x, 0) = f(x)$. Now using (16) in (14) and doing the integration on the right we have

$$\int_{\alpha_1}^\infty P(\alpha, \tau)\, d\alpha < \alpha_1^{-1} C + \alpha_1^{-2} \int_0^\tau D(\tau')\, d\tau'. \quad (17)$$

But P is a positive decreasing function of α therefore from the bound of Eq. (17) one can get a bound for the function $P(\alpha, \tau)$ using the simple theorem proved in the Appendix. From Eq. (A3) we obtain

$$P(\alpha, \tau) < 4C\alpha^{-2} + 8\alpha^{-3} \int_0^\tau D(\tau')\, d\tau'. \quad (18)$$

One could decrease the numerical coefficients somewhat and still maintain the inequality. This is the final form of the bound. If we have an equilibrium situation then $f(x, \tau) = f(x)$ and from (13) one can see that $D(\tau)$ is independent of τ. Hence (18) becomes for equilibrium distribution

$$P(\alpha, \tau) < 4C\alpha^{-2} + 8\alpha^{-3}\tau D. \quad (19)$$

Since $d\bar{A}(\tau)/d\tau = 0$ when the system is in equilibrium, we have

$$D = \int_\Gamma |\{A(x), H(x)\}|\, [A(x) - \bar{A}]^2 f(x)\, dV. \quad (20)$$

III. APPLICATIONS

In this section let us consider a simple application of (19) for illustrative purposes.

We shall consider an interacting gas of point molecules described by a canonical distribution function. We shall look at the fluctuations in the kinetic energy of the system. For this purpose we have

$$x = \{\mathbf{r}_1, \mathbf{p}_1 \cdots \mathbf{r}_N, \mathbf{p}_N\}, \quad (21a)$$

$$H(x) = K + V = \sum_i p_i^2/2m + \tfrac{1}{2} \sum_{i,j}{}' V(|\mathbf{r}_i - \mathbf{r}_j|), \quad (21b)$$

$$f(x) = Z^{-1} e^{-\beta H(x)} = (\beta/2\pi m)^{3N/2} e^{-\beta H(x)} \Big/ \int e^{-\beta V}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \quad (21c)$$

$$A(x) = K(x) = \sum_i p_i^2/2m, \quad (21d)$$

$$\{A(x), H(x)\} = -\sum_{i,j}{}' \mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|)\cdot \mathbf{p}_i/m, \quad (21e)$$

where

$$\mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|) = \partial V(|\mathbf{r}_i - \mathbf{r}_j|)/\partial \mathbf{r}_i.$$

The spatial volume the system occupies will be Ω and we have assumed spherically symmetric forces. We have the usual results for $\langle K\rangle$ and $\langle K^2\rangle$.

$$\langle K\rangle = \sum_i \langle p_i^2\rangle/2m = (\tfrac{3}{2})NkT, \quad (22)$$

and

$$\langle K^2\rangle = \sum_{i,j} \langle p_i^2 p_j^2\rangle/4m^2 = \langle K\rangle^2 + (\tfrac{3}{2})N(kT)^2, \quad (23)$$

$$C = (\langle K^2\rangle - \langle K\rangle^2) = (\tfrac{3}{2})N(kT)^2. \quad (24)$$

The calculation of D is more tedious. From Eqs. (20) and (21) we have

$$D = (8m^3 Z)^{-1} \sum_{k,l} \int \Big| \sum_{i,j}{}' \mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|)\cdot \mathbf{p}_i \Big| \times (p_k^2 p_l^2 - \langle p_k^2\rangle\langle p_l^2\rangle) e^{-\beta H(x)}\, d\mathbf{r}_1 \cdots d\mathbf{p}_N$$
$$\le (8m^3 Z)^{-1} \sum_{k,l} \sum_{i,j}{}' \int \Big| \mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|)\cdot \mathbf{p}_i \Big| \times (p_k^2 p_l^2 - \langle p_k^2\rangle\langle p_l^2\rangle) e^{-\beta H(x)}\, d\mathbf{r}_1 \cdots d\mathbf{p}_N. \quad (25)$$

If one defines the position pair distribution function, as usual, by

$$f_2(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \times \int e^{-\beta V}\, d\mathbf{r}_3 \cdots d\mathbf{r}_N \Big/ \int e^{-\beta V}\, d\mathbf{r}_1 \cdots d\mathbf{r}_N, \quad (26)$$

then the spatial part of the integrals in (25) can be written as

$$(N^2 - N)^{-1} \int \Big| \mathbf{F}(|\mathbf{r}_i - \mathbf{r}_j|)\cdot \mathbf{p}_i \Big| f_2(\mathbf{r}_i, \mathbf{r}_j)\, d\mathbf{r}_i\, d\mathbf{r}_j. \quad (27)$$

If $\mathbf{F}$ is of finite range and the volume is large f_2 will be a function of $|\mathbf{r}_i - \mathbf{r}_j|$ alone and to order Ω^{-1}, (27) can be replaced by

$$\Omega(N^2 - N)^{-1} \int_\Omega \left|\mathbf{F}(|\mathbf{r}|)\cdot\mathbf{p}_i\right| f_2(|\mathbf{r}|)\, d\mathbf{r}$$
$$= 2\pi\Omega(N^2 - N)^{-1} p_i \int_0^\infty |F(r)|\, f_2(r) r^2\, dr. \quad (28)$$

We set

$$Q = \int_0^\infty |F(r)|\, f_2(r) r^2\, dr. \quad (29)$$

Using (28) in (25) we can write

$$D \leq \pi Q\Omega(4m^3N)^{-1} \times \sum_{i,k,l} [\langle p_i p_k^2 p_l^2\rangle - \langle p_i\rangle\langle p_k^2\rangle\langle p_l^2\rangle], \quad (30)$$

where

$$\langle p_i p_j p_k \cdots\rangle = (\beta/2\pi m)^{3N/2} \int p_i p_j p_k \cdots e^{-\beta K}\, d\mathbf{p}_1 \cdots d\mathbf{p}_N.$$

In Eq. (30) those terms for which i, j, and k are all different vanish, because of the independence of the integrations. One considers then the possibilities $i=k$ and $i=l$, both of which give the same result, and $k = l$. There are also terms for which $i = k = l$ but these are negligible (of order N^{-1}) compared to the others. Then

$$D \leq \pi Q\Omega(4m^3N)^{-1} \sum_{i,j}{}' \{2[\langle p_i^3 p_j^2\rangle - \langle p_i\rangle\langle p_i^2\rangle\langle p_j^2\rangle] + [\langle p_i\rangle\langle p_i^4\rangle - \langle p_i\rangle\langle p_i^2\rangle^2]\}. \quad (31)$$

Evaluating the averages, one obtains

$$D \leq 3\pi^{\frac{1}{2}}(N - 1)\Omega Q(2m/\beta)^{5/2}(2m^3)^{-1}. \quad (32)$$

Now let us consider the size of α in (19). Suppose we look for fluctuations in the kinetic energy which are some fixed fraction of the total average kinetic energy. Then we will have

$$\alpha = \gamma\langle K\rangle = \gamma(\tfrac{3}{2})NkT, \quad (33)$$

where γ is some number, presumably $\ll 1$ but indedent of N. If we put (33), (32), and (24) into (20) we obtain

$$P(\gamma\langle K\rangle, \tau) \leq \tfrac{8}{3}(\gamma^2 N)^{-1} + (128\sqrt{2}/9)(Q\Omega/N)(mkT)^{-\frac{1}{2}}(\gamma^3 N)^{-1}\tau. \quad (34)$$

As one expects, the right side of (34) vanishes in the thermodynamic limit, that is it can be made arbitrarily small if we choose a large enough system. This result depends essentially on two properties; (1) that Q is independent of Ω (or N) as Ω gets large and (2) that we are looking for macroscopic fluctuations (33). We have made a rough calculation for argon gas, using the Lennard-Jones potential.[6] Taking T to be around 300°K and using the low-density approximation.

$$f_2(r) \simeq (N/\Omega)^2 e^{-\beta V(r)}.$$

We find that (34) becomes approximately

$$P(\gamma\langle K\rangle, \tau) \leq (\tfrac{8}{3})(N\gamma^2)^{-1} + 10^{-16}\tau(\Omega\gamma^3)^{-1} \quad (35)$$

in mks units. The only restriction on N and Ω is that of low density; i.e., $Na^3 \ll \Omega$ where a is the range of the potential.

In conclusion we should like to mention the following two points.

(1) The thing that makes the bound (18) useful, is that for macroscopic fluctuation, the bound behaves as N^{-1} (or Ω^{-1}) and therefore can be made small. This property does not depend on the choice of the canonical ensemble or the fact that we considered fluctuations of the kinetic energy. It is true for a certain class of distribution functions and dynamical variables.

(2) One should not regard the bound (18) as giving the actual asymptotic behavior for $P(\alpha, \tau)$ as a function of N. In particular, for certain kinds of variables one can find a much stronger bound. Instead of making the step from (11) to (12) we can define a number $D'(\alpha, \tau)$ by

$$D'(\alpha, \tau) = \int_{|A(x)-\bar{A}(\tau)|>\alpha} |\{A(x), H(x)\} - d\bar{A}(\tau)/d\tau|\, f(x, \tau)\, dV. \quad (36)$$

We can also define $C'(\alpha)$ by

$$C'(\alpha) = P(\alpha, 0) = \int_{|A(x)-\bar{A}(0)|>\alpha} f(x)\, dV. \quad (37)$$

Then by integrating (11) we obtain

$$\int_{\alpha_1}^\infty P(\alpha, \tau)\, d\alpha \leq \int_{\alpha_1}^\infty C'(\alpha)\, d\alpha + \int_0^\tau D'(\alpha_1, \tau')\, d\tau'. \quad (38)$$

By the theorem in the Appendix

$$P(\alpha, \tau) \leq (\alpha - \alpha_0)^{-1} \times \left[\int_{\alpha_0}^\infty C'(\alpha')\, d\alpha' + \int_0^\tau D'(\alpha_0, \tau')\, d\tau'\right] \quad (39)$$

for any $\alpha_0 < \alpha$. For certain special cases one can find much stronger bounds for C' and D' than the ones (C and D) used in this paper.

These two points will be treated in more detail in a following paper.

ACKNOWLEDGMENTS

I would like to thank Dr. Joseph Kahane for several helpful discussions on the subject of this paper.

APPENDIX

In this Appendix we shall consider briefly a simple theorem used in the body of the paper. It is

[6] T. L. Hill, *Introduction to Statistical Thermodynamics* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1960).

the following: If $K(x)$ is a positive nonincreasing function for all $x \geq 0$ and if $\int_{x_0}^{\infty} K(x)dx \leq g(x_0)$ for all $x_0 > 0$ then $K(x) \leq (x - x_0)^{-1}g(x_0)$ for any $x_0 < x$.

Proof. For a given x, let $x_0 < x$. Then

$$g(x_0) \geq \int_{x_0}^{\infty} K(x')\,dx'$$

$$= \int_{x_0}^{x} K(x')\,dx' + \int_{x}^{\infty} K(x')\,dx'$$

$$\geq \int_{x_0}^{x} K(x')\,dx' \geq K(x)(x - x_0). \quad \text{(A1)}$$

In particular, a possible choice of x_0 (not necessarily the best one) is $x_0 = x/2$. Then

$$K(x) \leq 2g(\tfrac{1}{2}x)/x. \quad \text{(A2)}$$

For the particular application in this paper $g(x) = a/x + b/x^2$ so

$$K(x) \leq 4a/x^2 + 8b/x^3. \quad \text{(A3)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Scalar Invariants of a Rotational System in a Lie Algebra

M. C. PEASE
Stanford Research Institute, Menlo Park, California
(Received 14 July 1964)

We consider matrix equations of the form $d\mathbf{W}/dz = [\mathbf{S}, \mathbf{W}]$, where $\mathbf{S}$ is a matrix function of z that is embedded in a given Lie algebra L—i.e., it is a curve in L. If the initial condition on $\mathbf{W}$ is in L, then $\mathbf{W}$ describes a curve in L. If the Killing form is used as a metric on L, then the behavior of the system is a pure rotation about an axis that is a function of z. A set of scalar invariants of such a system is obtained. These invariants form a set of conservation laws that the system obeys regardless of the detailed behavior of $\mathbf{S}(z)$.

If $\mathbf{S}$ describes a curve in some L_1, which is a semisimple subalgebra of the algebra L in which the whole system is embedded, then we can split the initial condition into two parts, one of which is in L_1 and generates a solution in L_1, the other in a subspace that is orthogonal to L_1 and that generates a curve that remains in the subspace. We can, then, obtain conservation laws that apply separately to the two parts.

The results have application to quantum mechanics since the density matrix obeys this type of equation. They also have application to coupled mode theory if we use, instead of the vector the corresponding power density spectrum matrix or the like.

INTRODUCTION

THE purpose of this work is to study the matrix equation

$$d\mathbf{W}/dz = [\mathbf{S}, \mathbf{W}] = \mathbf{SW} - \mathbf{WS}, \quad (1)$$

where $\mathbf{S}$ and $\mathbf{W}$ are matrix-valued functions of z (which may be, instead, the time, t) that describe curves in L, a semisimple Lie algebra[1-4] over a field of characteristic zero. We determine, in particular, the scalar-valued functions of one or more solutions of Eq. (1) that are independent of z. For example, if $\mathbf{U}$ and $\mathbf{V}$ are solutions of Eq. (1), the Killing form, $\langle \mathbf{U}, \mathbf{V} \rangle$, is one such invariant. We also find higher-order ones.

Our purpose in doing this is to find conservation laws that the system obeys and which are a consequence of the algebra involved and not of the detailed behavior of the system. The results, then, are useful in the study of systems for which an exact and explicit solution is unobtainable or impractical.

Equations of the form of Eq. (1) occur in many contexts. In quantum mechanics, for example, the time derivative of the density matrix,[5] as usually defined, is $(1/ih)$ times the commutator of the Hamiltonian and the density matrix.

Another general area of application is in coupled mode theory.[6,7] We may be given the vector dif-

[1] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts 1962).
[2] P. M. Cohn, *Lie Groups*, in *Cambridge Tracts in Mathematics and Mathematical Physics* (Cambridge University Press, Cambridge, England, 1961), No. 46.
[3] E. B. Dynkin, "The Structures of Semi-Simple Algebras" Usp. Math. Nauk (N. S.) 2, No. 4 (20) 1947, Am. Math. Soc. Transl. No. 17, 1950.
[4] N. Jacobson, *Lie Algebras* (Interscience Publishers, John Wiley & Sons, Inc., New York, 1962).
[5] A. Messiah, *Quantum Mechanics* (North-Holland Publishing Company, Amsterdam, 1961), Vol. I, p. 331 ff.
[6] M. C. Pease, J. Appl. Phys. **31**, 1988 (1960).
[7] M. C. Pease, J. Appl. Phys. **32**, 1736 (1961).

the following: If $K(x)$ is a positive nonincreasing function for all $x \geq 0$ and if $\int_{x_0}^{\infty} K(x)dx \leq g(x_0)$ for all $x_0 > 0$ then $K(x) \leq (x - x_0)^{-1}g(x_0)$ for any $x_0 < x$.

Proof. For a given x, let $x_0 < x$. Then

$$g(x_0) \geq \int_{x_0}^{\infty} K(x')\,dx'$$

$$= \int_{x_0}^{x} K(x')\,dx' + \int_{x}^{\infty} K(x')\,dx'$$

$$\geq \int_{x_0}^{x} K(x')\,dx' \geq K(x)(x - x_0). \quad \text{(A1)}$$

In particular, a possible choice of x_0 (not necessarily the best one) is $x_0 = x/2$. Then

$$K(x) \leq 2g(\tfrac{1}{2}x)/x. \quad \text{(A2)}$$

For the particular application in this paper $g(x) = a/x + b/x^2$ so

$$K(x) \leq 4a/x^2 + 8b/x^3. \quad \text{(A3)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 6, NUMBER 1 JANUARY 1965

Scalar Invariants of a Rotational System in a Lie Algebra

M. C. PEASE
Stanford Research Institute, Menlo Park, California
(Received 14 July 1964)

We consider matrix equations of the form $d\mathbf{W}/dz = [\mathbf{S}, \mathbf{W}]$, where $\mathbf{S}$ is a matrix function of z that is embedded in a given Lie algebra L—i.e., it is a curve in L. If the initial condition on $\mathbf{W}$ is in L, then $\mathbf{W}$ describes a curve in L. If the Killing form is used as a metric on L, then the behavior of the system is a pure rotation about an axis that is a function of z. A set of scalar invariants of such a system is obtained. These invariants form a set of conservation laws that the system obeys regardless of the detailed behavior of $\mathbf{S}(z)$.

If $\mathbf{S}$ describes a curve in some L_1, which is a semisimple subalgebra of the algebra L in which the whole system is embedded, then we can split the initial condition into two parts, one of which is in L_1 and generates a solution in L_1, the other in a subspace that is orthogonal to L_1 and that generates a curve that remains in the subspace. We can, then, obtain conservation laws that apply separately to the two parts.

The results have application to quantum mechanics since the density matrix obeys this type of equation. They also have application to coupled mode theory if we use, instead of the vector the corresponding power density spectrum matrix or the like.

INTRODUCTION

THE purpose of this work is to study the matrix equation

$$d\mathbf{W}/dz = [\mathbf{S}, \mathbf{W}] = \mathbf{SW} - \mathbf{WS}, \quad (1)$$

where $\mathbf{S}$ and $\mathbf{W}$ are matrix-valued functions of z (which may be, instead, the time, t) that describe curves in L, a semisimple Lie algebra[1-4] over a field of characteristic zero. We determine, in particular, the scalar-valued functions of one or more solutions of Eq. (1) that are independent of z. For example, if $\mathbf{U}$ and $\mathbf{V}$ are solutions of Eq. (1), the Killing form, $\langle \mathbf{U}, \mathbf{V} \rangle$, is one such invariant. We also find higher-order ones.

Our purpose in doing this is to find conservation laws that the system obeys and which are a consequence of the algebra involved and not of the detailed behavior of the system. The results, then, are useful in the study of systems for which an exact and explicit solution is unobtainable or impractical.

Equations of the form of Eq. (1) occur in many contexts. In quantum mechanics, for example, the time derivative of the density matrix,[5] as usually defined, is $(1/ih)$ times the commutator of the Hamiltonian and the density matrix.

Another general area of application is in coupled mode theory.[6,7] We may be given the vector dif-

[1] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts 1962).
[2] P. M. Cohn, *Lie Groups*, in *Cambridge Tracts in Mathematics and Mathematical Physics* (Cambridge University Press, Cambridge, England, 1961), No. 46.
[3] E. B. Dynkin, "The Structures of Semi-Simple Algebras" Usp. Math. Nauk (N. S.) 2, No. 4 (20) 1947, Am. Math. Soc. Transl. No. 17, 1950.
[4] N. Jacobson, *Lie Algebras* (Interscience Publishers, John Wiley & Sons, Inc., New York, 1962).
[5] A. Messiah, *Quantum Mechanics* (North-Holland Publishing Company, Amsterdam, 1961), Vol. I, p. 331 ff.
[6] M. C. Pease, J. Appl. Phys. **31**, 1988 (1960).
[7] M. C. Pease, J. Appl. Phys. **32**, 1736 (1961).

ferential equation

$$d\mathbf{x}/dz = \mathbf{S}\mathbf{x}, \tag{2}$$

where $\mathbf{x}$ is an n-dimensional vector written as a column matrix. The general solution to Eq. (2) can be obtained from the so-called matricant[8] which is the matrix function of z and z_0 that is the solution of

$$\partial\mathbf{M}(z, z_0)/\partial z = \mathbf{S}(z)\mathbf{M}(z, z_0); \quad \mathbf{M}(z_0, z_0) = \mathbf{I}. \tag{3}$$

If, then, the initial condition of Eq. (2) at z_0 is $\mathbf{x}(z_0)$, the resultant solution is given by

$$\mathbf{x}(z) = \mathbf{M}(z, z_0)\mathbf{x}(z_0). \tag{4}$$

Equation (3) is not of the same form as Eq. (1). If, however, instead of $\mathbf{M}(z, z_0)$, we consider

$$\mathbf{W}(z) = \mathbf{M}\mathbf{W}_0\mathbf{M}^{-1} = \mathbf{M}(z, z_0)\mathbf{W}_0\mathbf{M}(z_0, z), \tag{5}$$

then $\mathbf{W}(z)$ obeys Eq. (1). $\mathbf{W}_0$ is the initial condition on $\mathbf{W}$, i.e., $\mathbf{W}(z_0)$.

Alternatively, we may find that $\mathbf{S}(z)$ is everywhere what we call K-skew-Hermitian.[7] That is, we may discover that there exists a nonsingular constant Hermitian matrix, $\mathbf{K}$, such that, at any z

$$\mathbf{K}\mathbf{S} + \mathbf{S}^\dagger\mathbf{K} = \mathbf{0}, \tag{6}$$

where the dagger indicates the Hermitian conjugate, which is the complex conjugate transpose of $\mathbf{S}$.

If so, we can define $\mathbf{W}$ as a linear combination of dyad solutions:

$$\mathbf{W}(z) = \sum a_{ij}\mathbf{x}_i\mathbf{x}_j^\dagger\mathbf{K}, \tag{7}$$

where the set $\mathbf{x}_i$ are a set of linearly independent solutions of Eq. (2), and the a_{ij} are constants. We find, then, that $\mathbf{W}(z)$, defined by Eq. (7) satisfies Eq. (1).

The rationale behind Eqs. (6) and (7) depends on regarding $\mathbf{K}$ as providing a metric for the space, even though we have not required $\mathbf{K}$ to be positive definite, (i.e., it is not a Hilbert space). We define what we call an *improper* inner product by $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^\dagger\mathbf{K}\mathbf{y}$.

If $\mathbf{A}$ is any operator, we define its adjoint, $\mathbf{A}^*$, by the requirement that, for any $\mathbf{x}$ and $\mathbf{y}$ in the space.

$$\langle \mathbf{A}^*\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{A}\mathbf{y}\rangle.$$

Equation (6) is, then, the requirement that $\mathbf{S}$ be the negative of $\mathbf{S}^*$, so that $\mathbf{S}$ is skew-self-adjoint. The set of all such matrices form the Lie algebra that is associated with the Lie group of what we call the K-unitary matrices—i.e., the matrices, $\mathbf{M}$, such that $\mathbf{M}^* = \mathbf{M}^{-1}$, which requires that

$$\mathbf{M}^\dagger\mathbf{K}\mathbf{M} = \mathbf{K}.$$

[8] F. R. Gantmacher, *The Theory of Matrices* (Chelsea Publishing Company, New York, 1959), Chap. XIV.

If $\mathbf{S}$ describes a curve in this algebra, we can show that $\mathbf{M}(z, z_0)$, defined by Eq. (3), describes a curve in the Lie group. The $\mathbf{W}(z)$ of Eq. (1), then describes a curve in the algebra, providing $\mathbf{W}_0$ is in the algebra.

The $\mathbf{W}(z)$ of Eq. (5) or (7) does not contain all the information contained in $\mathbf{M}(z, z_0)$ or in $\mathbf{x}(z)$. It does, however, contain what is, for many purposes, the essential information. For example, the density matrix of quantum mechanics suppresses the phase information. However, it is usually only the magnitude that is desired anyway, so that the elimination of the phase is a useful simplification of the problem.

Much the same is true of Eq. (5) or (7). Equation (7) is a generalization of the "power density matrix" that has been found useful in the study of noise.[9]

We shall, then consider Eq. (1) and seek scalar functions of its solutions that are invariant and so give us conservation laws of the system.

THE LIE ALGEBRA

We consider an algebra, L, of $n \times n$ matrices over a suitable field F, although the results apply equally to the corresponding abstract algebra.[1-3] We assume throughout that L is finitely dimensioned and that F has characteristic zero.

We assume that a basis has been chosen, consisting of the matrices, $\mathbf{X}_i$. The structure constants are obtained as

$$[\mathbf{X}_i, \mathbf{X}_j] = c_{ij}^k\mathbf{X}_k, \tag{8}$$

where the summation convention applies and is used throughout. The structure constants are skew-symmetric in the lower indices

$$c_{ij}^k = -c_{ji}^k \tag{9}$$

and obey the Jacobi identity:

$$c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n = 0. \tag{10}$$

We expand an arbitrary element, $\mathbf{A}$, in the algebra, on the basis:

$$\mathbf{A} = a^i\mathbf{X}_i. \tag{11}$$

The "Killing form," or the "scalar product of Cartan," of two elements, $\mathbf{A}$ and $\mathbf{B}$, of L will be written as

$$\langle \mathbf{A}, \mathbf{B}\rangle = a^i b^j c_{im}^n c_{jn}^m \tag{12}$$

and the resultant metric as g_{ij}

$$g_{ij} = c_{im}^n c_{jn}^m. \tag{13}$$

This, is, we note, a symmetric operator.

[9] L. D. Smullin and H. A. Haus, *Noise in Electron Devices* (Technology Press, Cambridge, Massachusetts, and John Wiley & Sons, Inc., New York, 1959), Chap. 3.

We also remind the reader of Cartan's criterion that, if, and only if, L is semisimple, g_{ij} is nonsingular so that there exists an operator g^{ij} such that

$$g_{ij}g^{jk} = \delta_i^k. \tag{14}$$

For the main part of the development that follows, we do not need to assume semisimplicity. However, when we come to consider the situation when $\mathbf{S}(z)$ is embedded in a subalgebra, it will be necessary to assume that this subalgebra is semisimple, so that a reciprocal metric exists and Eq. (12) is a nonsingular form.

Using the expansion of Eq. (11), and the fact that the $\mathbf{X}_i$ are linearly independent, being a basis for the algebra, Eq. (1) becomes

$$dw^k/dz = s^i w^j c_{ij}^k. \tag{15}$$

It is this equation that we study.

n-INDEX e FORMS

We define what we call the n-index e form as the contraction of n structure constants:

$$e_{ij\cdots h} = c_{iv}^u c_{ju}^v \cdots c_{hx}^y. \tag{16}$$

The indices on the e form lists the first subscripts of the structure constants in the order of appearance. The superscript of each structure constant is contracted with the second subscript of the following one. The superscript on the last structure constant is contracted with the second subscript of the first.

The metric of Eq. (13) is the two-index e form:

$$g_{ij} = e_{ij}. \tag{17}$$

The n-index e form is evidently unchanged by a cyclic permutation of its indices. In general, distinct n-index e forms are obtained by any other permutation. There are, then, $(n-1)!$ distinct n-index e forms obtained by all possible noncyclic permutations of the indices.

The principal property of the e forms that concerns us is contained in the following:

Lemma 1: With the e forms defined by Eq. (16),

$$c_{ij}^t e_{tkh\cdots} = e_{jikh\cdots} - e_{ijkh\cdots}. \tag{18}$$

This follows immediately from the Jacobi identity, Eq. (10), with Eq. (9):

$$\begin{aligned} c_{ij}^t e_{tkh\cdots} &= c_{ij}^t c_{tv}^u c_{ku}^v \cdots \\ &= -(c_{jv}^t c_{ti}^u + c_{vi}^t c_{tj}^u) c_{ku}^v \cdots \\ &= c_{jv}^t c_{it}^u c_{ku}^v \cdots - c_{iv}^t c_{jt}^u c_{ku}^v \cdots \\ &= e_{jikh\cdots} - e_{ijkh\cdots}. \end{aligned}$$

Thus, the contraction of a structure constant with an n-index e form may be written as the difference of two $(n+1)$-index e forms which differ only in the interchange of the first two indices.

CONSERVATION LAWS

The n-index e forms give us the elementary scalar invariants of Eq. (1). The theorem that applies is the following:

Theorem 1: If $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, $\cdots$ are n solutions of Eq. (1), which may be distinct or not, and if these solutions are expanded on the basis as in Eq. (11), then the scalar quantity

$$p^{(n)} = (u^i v^j w^k \cdots) e_{ijk\cdots} \tag{19}$$

is invariant.

From Eq. (15) we find that

$$\begin{aligned} dp^{(n)}/dz = {} & (s^m u^n c_{mn}^i v^j w^k \cdots + u^i s^m v^n c_{mn}^j w^k \cdots \\ & + u^i v^j s^m w^n c_{mn}^k \cdots + \cdots) e_{ijk} \\ = {} & (s^m u^i v^j w^k \cdots)\{c_{mi}^t e_{tjk\cdots} + c_{mj}^t e_{itk\cdots} \\ & + c_{mk}^t e_{ijt\cdots} + \cdots\} \end{aligned} \tag{20}$$

by relabeling the dummy indices appropriately. Using the invariance of the e form for a cyclic permutation of its indices and Lemma 1, the bracketed terms in Eq. (20) are

$$\begin{aligned} \{\ \} &= e_{imjk\cdots} - e_{mijk\cdots} \\ &\quad + e_{jmk\cdots i} - e_{mjk\cdots i} + e_{km\cdots ij} - e_{mk\cdots ij} + \cdots \\ &= e_{mjk\cdots i} - e_{mijk\cdots} + e_{mk\cdots ij} - e_{mjk\cdots i} \\ &\quad + e_{m\cdots ijk} - e_{mk\cdots ij} + \cdots = 0. \end{aligned}$$

Hence

$$dp^{(n)}/dz = 0 \tag{21}$$

and the theorem is proved.

If we apply the theorem to the two-index e form, which is identical with g_{ij}, we obtain the invariance of

$$p^{(2)} = u^i v^j e_{ij} = u^i v^j g_{ij} = \langle \mathbf{U}, \mathbf{V} \rangle \tag{22}$$

from Eq. (12). Hence, if we regard the Killing form $\langle \mathbf{W}, \mathbf{W} \rangle$ as the "square of the length" of $\mathbf{W}$, and $\langle \mathbf{U}, \mathbf{V} \rangle$ as a generalized dot product of $\mathbf{U}$ and $\mathbf{V}$, we see that Eq. (1) can be interpreted as indicating a *rotation* in the Lie algebra.

This is not, however, the only invariant that is involved. In general any e form yields a scalar invariant.

SYMMETRIZATION OF THE CONSERVATION LAWS

From the theorem, the following is obvious:

Corollary: In Theorem 1, we can replace the n-index

e form by any constant linear combination of the n-index e forms obtained by permuting the subscripts.

We may note that a permutation of the subscripts of the e forms in Eq. (19) is equivalent to permuting the positions of U, V, W, $\cdots$ in Eq. (19).

With this corollary, then, we can set up combinations of n-index e forms that have any desired symmetry. We define the n-index g form as the linear combination of n-index e forms that is fully symmetric. Since the e forms are already cyclicly symmetric, we can simplify the process somewhat, obtaining:

$$\begin{aligned} g_i &= e_i = c^m_{im}, \\ g_{ij} &= e_{ij}, \\ g_{ijk} &= e_{ijk} + e_{ikj}, \\ g_{ijkh} &= e_{ijkh} + e_{ijhk} + e_{ikjh} \\ &\quad + e_{ikhj} + e_{ihjk} + e_{ihkj}, \quad \text{etc.} \end{aligned} \tag{23}$$

(Note that g_{ijkh} as defined here is *not* the same as the g_{ijkh} that is sometimes used in relativity theory,[10] which is defined as $g_{ik}g_{jh} - g_{ih}g_{jk}$ and so is not fully symmetric.)

With the g forms, we obtain a set of conservation laws that are applicable to a single solution of Eq. (1). That is, the scalars

$$\begin{aligned} p^{(1)} &= w^i g_i, \qquad p^{(2)} = w^i w^j g_{ij}, \\ p^{(3)} &= w^i w^j w^k g_{ijk}, \quad \text{etc.} \end{aligned} \tag{24}$$

are conserved and are significant providing the g form involved is not the null operator.

Not all the scalar invariants of Eq. (24) will be nontrivial. Indeed it seems to be true that, for a simple algebra at least, all the odd-index e forms are antisymmetric, so that the odd-index g forms are null.

Neither are all of the invariants independent. The scalars, $p^{(n)}$ of Eq. (24) are related to the coefficients of the characteristic equation of the linear operator $ad_W = [W, -]$. Hence there is an upper limit on the number that can be linearly independent. By example, we do find that the higher invariants may not be trivial. We show such an example later.

We note, finally, the significance of definiteness in the g forms. If any of the forms of Eq. (24) is definite, it then follows that none of the coefficients can grow without limit with z. Neither can all of the coefficients decay simultaneously to zero. Thus the requirement that all of the nontrivial g forms be indefinite is a necessary, although not a sufficient, condition for either amplification or attenuation.

[10] J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

PARTIALLY SYMMETRIC CONSERVATION LAW

Suppose, now, we consider a given symmetric g form of order $(n - 1)$. If we contract it with c^i_{ij}, we obtain a linear combination of n-index e forms. We call such a form a *derived g form*, and symbolize it with a prime, for example:

$$g'_{ijk} = c^i_{ij}g_{ik} = e_{jik} - e_{ijk}. \tag{25}$$

We find that we obtain the same result if we compute $c^i_{ki}g_{ij}$ and $c^i_{jk}g_{ii}$. This is reasonable. Because of the cyclic symmetry of the e forms, there are only two distinct three-index e forms—e.g., e_{ijk} and e_{ikj}. The information contained in them is given fully in the two forms g_{ijk} and g'_{ijk}. Hence there can be no additional linearly independent three-index forms.

We can continue the process and obtain the *doubly derived g* forms

$$\begin{aligned} g''_{ijkh} &= c^i_{ij}g'_{ikh} = c^i_{ij}c^s_{ik}g_{sh} \\ &= e_{ijkh} - e_{ijhk} - e_{jikh} + e_{jihk}. \end{aligned} \tag{26}$$

We observe that g''_{ijkh} is antisymmetric in the interchange of either the first pair or the second pair, and symmetric in the simultaneous interchange of both pairs (and so has the same symmetry as the Riemann curvature tensor[10]).

We can find, by careful analysis, that there are exactly three linearly independent, singly derived, four-index g forms, and two doubly derived ones. Together with the fully symmetric form, these account for the six distinct four-index e forms. Presumably, the same thing can be done with the higher-order forms.

The derived g forms, because of their antisymmetries, do not generate nontrivial conservation laws for a single solution of Eq. (1). They do, however, generate invariant orthogonality relations. If we consider for example, g'_{ijk}, we see that any scalar such as $w^i w^j w^k g'_{ijk}$ or $w^i w^j v^k g'_{ijk}$ must be identically zero. However, $u^i v^j w^k g'_{ijk}$ may be nontrivial, so that its invariance gives us significant information about the mutual behavior of a triad of solutions, U, V, and W.

L NOT SIMPLE

We now consider, in this and the following section, how the problem can be split and conservation laws obtained that apply to separate parts of the

solution. In this section[11] we consider what can be done if L is not a simple Lie algebra. In the following section, we consider how we can make use of the possibility that $\mathbf{S}(z)$ may be embedded in some smaller Lie algebra than L, the range of possible initial conditions.

We consider, first, the case where L is semisimple but not simple. By the structure theorem,[3,4] then, we can write L as

$$L = L_{(1)} \oplus L_{(2)} \oplus \cdots \oplus L_{(k)}, \tag{27}$$

where the $L_{(i)}$ are simple algebras that are ideals of L. We resolve the initial condition, $\mathbf{W}_0$, into

$$\mathbf{W}_0 = \mathbf{W}_{01} + \mathbf{W}_{02} + \cdots \mathbf{W}_{0k}, \tag{28}$$

where $\mathbf{W}_{0i}$ is in $L_{(i)}$.

Also, we can split $\mathbf{S}(z)$ into

$$\mathbf{S} = \mathbf{S}_{(1)} + \mathbf{S}_{(2)} + \cdots + \mathbf{S}_{(k)}, \tag{29}$$

where $\mathbf{S}_i$ is in $L_{(i)}$. If, now, we write

$$\mathbf{W} = \mathbf{W}_{(1)} + \mathbf{W}_{(2)} + \cdots + \mathbf{W}_{(k)}, \tag{30}$$

where $\mathbf{W}_{(i)}$ is in $L_{(i)}$, then $\mathbf{W}_{(i)}$ obeys the equation

$$d\mathbf{W}_{(i)}/dz = [\mathbf{S}_{(i)}, \mathbf{W}_{(i)}]; \qquad \mathbf{W}_{(i)}(0) = \mathbf{W}_{0i} \tag{31}$$

since the commutator of $\mathbf{S}_{(i)}$ and $\mathbf{W}_{(j)}$, $j \neq i$, must vanish, both $L_{(i)}$ and $L_{(j)}$ being ideals of L and disjoint.

Hence, we have split the problem into a set of problems in simple algebras. The conservation laws of $L_{(i)}$ will, then, apply to $\mathbf{W}_{(i)}$.

If L is not semisimple, then we have a more difficult problem. By Levi's theorem,[4] we can write $L = L_{(0)} + L'$, where $L_{(0)}$ is the radical of L—i.e., the maximal soluble ideal of L—and L' is a semisimple algebra. We can, now, decompose L' as in Eq. (27), writing

$$L = L_{(0)} + L_{(1)} + \cdots + L_{(k)}, \tag{32}$$

where $L_{(i)}$ $(i \neq 0)$ is a simple algebra. The algebra $\{L_{(0)} + L_{(i\neq 0)}\}$ is, now, an ideal of L.

We can write the initial condition uniquely, as

$$\mathbf{W}_0 = \mathbf{W}_{00} + \mathbf{W}_{01} + \cdots + \mathbf{W}_{0k}, \tag{33}$$

where $\mathbf{W}_{0i}$ is in $L_{(i)}$. Let $\mathbf{W}_{(0)}$ and $\mathbf{W}_{(i\neq 0)}$ be such that

$$\begin{aligned} d\mathbf{W}_{(0)}/dz &= [\mathbf{S}_1\mathbf{W}_{(0)}] \qquad \mathbf{W}_{(0)}(0) = \mathbf{W}_{00} \\ d\mathbf{W}_{(i)}/dz &= [\mathbf{S}_1\mathbf{W}_{(i)}] \qquad \mathbf{W}_{(i)}(0) = \mathbf{W}_{0i}. \end{aligned} \tag{34}$$

Since $\{L_{(0)} + L_{(i)}\}$ is an ideal, $\mathbf{W}_{(i)}$ remains in this algebra. The solution to the whole problem is, then, given by

$$\mathbf{W} = \mathbf{W}_{(0)} + \mathbf{W}_{(1)} + \cdots + \mathbf{W}_{(k)}. \tag{35}$$

Thus we have split the problem into the sum of one wholly in the radical and a set of components each of which is in a subalgebra that is the sum of the radical and a simple algebra. Again, the conservation laws of these subalgebras apply to the appropriate parts of the solution.

S(z) IN A SEMISIMPLE SUBALGEBRA

Suppose, now, Eq. (1) is embedded in the Lie algebra, L, but $\mathbf{S}(z)$ is embedded in L_1, a subalgebra of L. In particular, suppose we admit the possibility that the initial condition on W of Eq. (1) is not in the smallest Lie algebra covering $\mathbf{S}(z)$. As an example, we have mentioned, in Eq. (6), the possibility that S may be K-skew-Hermitian. Since the set of K-skew-Hermitian matrices is a Lie algebra over the field of real numbers, as may be easily verified, this defines L_1. However, unless we specifically restrict $\mathbf{W}_0$ to be K-skew-Hermitian also, we do not know that W is in L_1, but instead only know that it is in the full algebra of $n \times n$ square matrices.

We can ask, then, how we can use the properties of L_1, instead of those of L to derive conservation laws of the system.

We find that, if L_1 is semisimple, we can specify a linear subspace, M, which is not in general a subalgebra, with the properties that

(A) The intersection of M and L_1 is void, so that they have no element in common.

(B) $[M, L_1] \in M$ so that M is regenerated when it is commuted with L_1. Hence M and L_1 are disjoint, and M is closed under commutation with L_1.

We can anticipate the later results here and observe that we can take M as the orthogonal complement of L_1 under the Killing form. By Cartan's criterion, if L_1 is semisimple, it can contain no element in common with its orthogonal complement. Hence, Property A is obtained. Property B is shown by the theorem that follows.

We do not, however, need always to use the full orthogonal complement of L_1. The problem may permit us to take as the M set a subspace of the orthogonal complement. Hence, we approach the problem slightly differently. We show that we can split any initial condition $\mathbf{W}_0$, into two parts, one of which, $\mathbf{W}_{01}$, is in L_1, the other, $\mathbf{W}_{02}$, is orthogonal to L_1. We then show that by closing $\mathbf{W}_{02}$ with respect to commutation with L_1, we obtain an M set.

[11] We are indebted to E. Norman for calling our attention to the splitting discussed here when L is not simple.

The significance of this splitting of the initial condition is that then the entire problem is split. Provided that $\mathbf{S}$ is sufficiently well behaved to apply Taylor's theorem at least in overlapping segments of the desired interval, a solution of Eq. (1) will be entirely in L_1, or entirely in M, depending on the initial condition. That is, we can now consider the two equations

$$d\mathbf{W}_1/dz = [\mathbf{S}, \mathbf{W}_1], \qquad \mathbf{W}_1(0) = \mathbf{W}_{01}, \tag{36}$$

$$d\mathbf{W}_2/dz = [\mathbf{S}, \mathbf{W}_2], \qquad \mathbf{W}_2(0) = \mathbf{W}_{02}, \tag{37}$$

and know that $\mathbf{W}_1$ is wholly in L_1, and $\mathbf{W}_2$ wholly in M. The solution to Eq. (1) is then

$$\mathbf{W}(z) = \mathbf{W}_1(z) + \mathbf{W}_2(z). \tag{38}$$

For the $\mathbf{W}_1$ part, we can use the conservation laws of L_1. For the $\mathbf{W}_2$ part, we must use the invariants of L, but even they are simplified by the absence of any of the L_1 terms from $\mathbf{W}_2$.

To obtain, now, the proper division of the initial condition, we use what is essentially the Gram–Schmidt process. We want to find $\mathbf{W}_{01}$ and $\mathbf{W}_{02}$ such that

$$\mathbf{W}_0 = \mathbf{W}_{01} + \mathbf{W}_{02}, \qquad \mathbf{W}_{01} \in L_1, \qquad \langle \mathbf{W}_{02}, L_1 \rangle = 0.$$

We find that these equations are satisfied by

$$\mathbf{W}_{01} = \langle \mathbf{W}_0, \mathbf{X}_\beta \rangle g^{\beta\alpha} \mathbf{X}_\alpha, \tag{39}$$

$$\mathbf{W}_{02} = \mathbf{W}_0 - \mathbf{W}_{01}. \tag{40}$$

The Greek letter indices, here, are used to indicate L_1. That is, we assume L_1 to be spanned by the set $\mathbf{X}_\alpha$, whereas L is spanned by the set $\mathbf{X}_i$. Then for any $\mathbf{X}_\gamma$ in L_1,

$$\begin{aligned} \langle \mathbf{W}_{02}, \mathbf{X}_\gamma \rangle &= \langle \mathbf{W}_0, \mathbf{X}_\gamma \rangle - \langle \mathbf{W}_{01}, \mathbf{X}_\gamma \rangle \\ &= \langle \mathbf{W}_0, \mathbf{X}_\gamma \rangle - \langle \mathbf{W}_0, \mathbf{X}_\beta \rangle g^{\beta\alpha} \langle \mathbf{X}_\alpha, \mathbf{X}_\gamma \rangle \\ &= \langle \mathbf{W}_0, \mathbf{X}_\gamma \rangle - \langle \mathbf{W}_0, \mathbf{X}_\beta \rangle g^{\beta\alpha} g_{\alpha\gamma} \end{aligned} \tag{41}$$

by Eq. (12). From Eq. (14), this vanishes.

From $\mathbf{W}_{02}$, as determined by Eq. (40), we can obtain an M set. We do this by closing $\mathbf{W}_{02}$ with respect to commutation with L_1; that is, we form commutators of as high an order as necessary of $\mathbf{W}_{02}$ with the various X_α, and pick out those that are linearly independent over the applicable field. That this process does generate a suitable M set is shown by the following:

Theorem 2: If L_1 is spanned by the set $\mathbf{X}_\alpha$, and if $\mathbf{A}$ is any element of L such that

$$\langle \mathbf{A}, L_1 \rangle = 0 \tag{42}$$

and if we define

$$\mathbf{B}_\alpha = [\mathbf{A}, \mathbf{X}_\alpha], \tag{43}$$

then

$$\langle \mathbf{B}_\alpha, L_1 \rangle = 0. \tag{44}$$

In other words, if $\mathbf{A}$ is orthogonal to L_1 under the Killing form, so is the commutator of $\mathbf{A}$ with any element of L_1. Hence, so is the whole set obtained from $\mathbf{A}$ by closure under commutation with L_1.

Equation (42) requires that, if $\mathbf{A}$ be expanded as $a^i \mathbf{X}_i$ then

$$a^i g_{i\alpha} = 0. \tag{45}$$

Then

$$\mathbf{B}_\alpha = a^i [\mathbf{X}_i, \mathbf{X}_\alpha] = a^i c^j_{i\alpha} \mathbf{X}_j$$

and

$$\begin{aligned} \langle \mathbf{B}_\alpha, \mathbf{X}_\beta \rangle = a^i c^j_{i\alpha} g_{j\beta} &= a^i (e_{\alpha i\beta} - e_{i\alpha\beta}) \\ &= a^i (e_{\beta\alpha i} - e_{\alpha\beta i}) = a^i c^j_{\alpha\beta} g_{ji}. \end{aligned}$$

Now, since L_1 is an algebra, $c^j_{\alpha\beta}$ vanishes unless X_j is in L_1. Hence we can replace j by γ:

$$\langle \mathbf{B}_\alpha, \mathbf{X}_\beta \rangle = c^\gamma_{\alpha\beta} (a^i g_{\gamma i}) = 0 \tag{46}$$

by Eq. (45).

Hence orthogonality to an algebra under the Killing form is preserved by commutation with the algebra.

The element $\mathbf{W}_{02}$ was chosen to be orthogonal to L_1 under the Killing form. The M set was obtained by closure for commutation with L_1. Hence the entire set is orthogonal to L_1.

By Cartan's criterion, the Killing form is nonsingular for a semisimple algebra. Hence, if L_1 is semisimple, no member of the set so obtained can be in L_1, and the M set so obtained is disjoint from L_1. Hence the set so obtained has Properties A and B and is therefore an M set.

This then demonstrates that, if $\mathbf{S}(z)$ is in L_1, a semisimple subalgebra of the algebra L in which Eq. (1) is embedded, then Eq. (1) can be split into two problems, one of which is wholly in L_1, the other in an M set that is orthogonal to L_1. Separate conservation laws will, then, apply to each part.

EXAMPLES

We have mentioned that the property of being K-skew-Hermitian, as defined in Eq. (6), defines an algebra over the real field. If $\mathbf{S}$ is known to be K-skew-Hermitian, then L_1 is this algebra. The appropriate M set is, then the set of K-Hermitian matrices ($\mathbf{KW} - \mathbf{W}^\dagger\mathbf{K} = 0$) which is i times L_1. We can, then split the initial condition, $\mathbf{W}_0$, into its K-skew-Hermitian and K-Hermitian parts and

treat each separately. In this case, we can multiply Eq. (29) by i and obtain a new problem in L_1. Hence we need only study Eq. (1) with the restriction that both $\mathbf{S}$ and $\mathbf{W}$ are everywhere K-skew-Hermitian. The conservation laws of this algebra apply separately to both parts.

As a second example, let us consider the algebra defined by the multiplication table given in Table I. One example of this algebra is given by

$$\mathbf{X}_1 = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad \mathbf{X}_4 = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\mathbf{X}_2 = \tfrac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \mathbf{X}_5 = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

$$\mathbf{X}_3 = \tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \qquad \mathbf{X}_6 = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This is the algebra of all 2×2 matrices with zero trace. (More generally, it is an example of the "full Lorentz algebra.")

We find that

$$g_{ij} = \text{diag}\,(-4, -4, -4, 4, 4, 4). \tag{47}$$

Hence (reverting to a subscript notation) if we set

$$\mathbf{W} = w_1\mathbf{X}_1 + w_2\mathbf{X}_2 + w_3\mathbf{X}_3 + w_4\mathbf{X}_4 + w_5\mathbf{X}_5 + w_6\mathbf{X}_6$$

and $\mathbf{W}$ is a solution of Eq. (1), then

$$p^{(2)} = -4(w_1^2 + w_2^2 + w_3^2 - w_4^2 - w_5^2 - w_6^2) = c_1 \tag{48}$$

is a constant.

We can find, for this algebra, that e_{ijk} is antisymmetric, so that g_{ijk} is identically zero.

The calculation of g_{ijkh} is tedious but straightforward. We find that $p^{(4)}$ is within a scalar constant,

$$p^{(4)} \propto (p^{(2)})^2 - 4(w_1w_4 + w_2w_5 + w_3w_6)^2. \tag{49}$$

Hence it follows that

$$w_1w_4 + w_2w_5 + w_3w_6 = c_2 \tag{50}$$

is a conservation law.

Suppose, now, that $\mathbf{S}(z)$ is embedded, for example, in the linear envelope over the real field of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$. This is a semisimple subalgebra. The corresponding M set is the linear envelope of $\mathbf{X}_4$, $\mathbf{X}_5$, and $\mathbf{X}_6$. Hence, in this case, we have the separate conservation laws that

$$w_1^2 + w_2^2 + w_3^2 = c_1,$$

$$w_4^2 + w_5^2 + w_6^2 = c_2.$$

In addition, Eq. (41) still applies.

TABLE I. Multiplication table.

	$\mathbf{X}_1]$	$\mathbf{X}_2]$	$\mathbf{X}_3]$	$\mathbf{X}_4]$	$\mathbf{X}_5]$	$\mathbf{X}_6]$
$[\mathbf{X}_1$	0	$\mathbf{X}_3$	$-\mathbf{X}_2$	0	$\mathbf{X}_6$	$-\mathbf{X}_5$
$[\mathbf{X}_2$	$-\mathbf{X}_3$	0	$\mathbf{X}_1$	$-\mathbf{X}_6$	0	$\mathbf{X}_4$
$[\mathbf{X}_3$	$\mathbf{X}_2$	$-\mathbf{X}_1$	0	$\mathbf{X}_5$	$-\mathbf{X}_4$	0
$[\mathbf{X}_4$	0	$\mathbf{X}_6$	$-\mathbf{X}_5$	0	$-\mathbf{X}_3$	$\mathbf{X}_2$
$[\mathbf{X}_5$	$-\mathbf{X}_6$	0	$\mathbf{X}_4$	$\mathbf{X}_3$	0	$-\mathbf{X}_1$
$[\mathbf{X}_6$	$\mathbf{X}_5$	$-\mathbf{X}_4$	0	$-\mathbf{X}_2$	$\mathbf{X}_1$	0

In this case, we can express L_1 as the algebra of K-skew-Hermitian matrices with $\mathbf{K} = \mathbf{I}$. We can, however, easily find examples where this is not so.

CONCLUSIONS

We have found that, for a system described by Eq. (1), there exists a set of conservation laws that limit the behavior of the system. These laws are obtained from the structure of the Lie algebra in which the problem is embedded. Hence their application requires no information other than the initial conditions and the algebra that is involved.

A further refinement is possible, also, if the algebra is not simple. If the algebra is semisimple, but not simple, we can split the problem into parts, each of which is wholly within a simple algebra. If it is not semisimple, the solution can also be split into components, each of which is the sum of the radical and a simple algebra. In either case we can obtain invariants for the separated parts.

It is also possible to split the problem if the system operator $\mathbf{S}(z)$ is embedded in a semisimple subalgebra, L_1. In this case, we can split the problem into two problems, one in L_1, the other in what we call an M space, which is a subspace of the orthogonal complement of L_1 with respect to the Killing form. Again we obtain separate invariants for the split parts of the problem.

ACKNOWLEDGMENTS

We are indebted to A. Novikoff, Stanford Research Institute, for stimulating our interest in the application of the theory of Lie algebras. We also wish to acknowledge the support of the Air Force for the research in coupled-mode theory, of which this has been one result.

Formal Solution of a Nonhomogeneous Differential Equation with a Double Transition Point*

P. F. JORDAN AND P. E. SHELLEY
The Martin Company, Baltimore, Maryland
(Received 9 June 1964)

A formal solution of a fairly general nonhomogeneous, linear, second-order differential equation with a large parameter and a double transition point is presented. This equation arises, for instance, in the quasilinear theory of pressurized membrane shells of revolution. The fact that the solution which is of foremost importance in practical applications converges at infinity makes it convenient to use a direct approach, avoiding any transcendental transformation. The solution is described by means of influence functions which arise from a formal inductive process. The more important influence functions are tabulated. The results of two approximate asymptotic procedures are compared with the exact solution.

I. INTRODUCTION

A SPECIFIC form of the differential equation to be discussed in this paper arises in the quasi-linear theory of pressurized membrane shells of revolution.[1,2] The general equation, however, is of interest beyond this specific application.[3]

The subject differential equation is real-valued, linear, and of second order. It has a large parameter, k, and a double transition point. As such, it belongs into the field of interest of the theory of asymptotic expansions[4] (as was pointed out by Sanders and Liepins[5]). Most of the work on this theory has been concerned with homogeneous solutions; however, the quasi-linear membrane shell equation is essentially nonhomogeneous and, in its application, one is mainly interested in one particular nonhomogeneous solution which may be designated as the "convergent" (at infinity) solution. Even in those shell problems where this convergent solution is not by itself sufficient, the role of the homogeneous solution is often confined to that of a boundary layer correction to the convergent solution.

Now, while it is usually considered sufficient to obtain the homogeneous solution (from which the nonhomogeneous solution may then be derived by means of the method of variation of parameters), the particular nonhomogeneous solution that is required in the present problem is much more easily derived directly. The reason for this somewhat unusual situation lies in the fact that the required solution is convergent while the homogeneous solution is exponentially divergent (see Sec. VIII). Accordingly, the present paper is essentially confined to a discussion of the convergent nonhomogeneous solution.

The limit form of the subject differential equation, that is, the differential equation that arises if the limit process $k \to \infty$ is made in the given differential equation, is a Bessel equation of order $\frac{1}{4}$. The convergent nonhomogeneous limit solution is related to (but different from) the Lommel function $S_{\mu,\frac{1}{2}}$ (Appendix A).

The given differential equation, in which k is large but finite, contains additional terms. Usually, in problems of this kind, one deals with such additional terms by means of a transcendental transformation, the aim being to reduce the given equation to its limit form. This procedure, which in general leads to asymptotic expansions, has several aspects which are undesirable from the point of view of practical applications. The transformation has to be evaluated anew numerically for each new case if one requires the solution itself rather than certain isolated aspects of it; the transformation is complicated enough to make it difficult to visualize, in a general manner, the differences between the actual solution and the limit solution. Furthermore, the analytical difficulties of the exact procedure are often formidable enough to require, in practical applications, the introduction of simplifying assumptions; that is, in practice the method of asymptotic expansion is replaced by an approximate procedure, the degree of validity of which is not always readily ascertained (Sec. VII).

The approach used in the present paper is more direct. No transcendental transformation is made. The difference between the required solution and the limit solution is described by influence functions. These influence functions are general functions, that is to say, they are independent of the specific param-

* This work was supported in part by the National Aeronautics and Space Administration.
[1] P. F. Jordan, J. Aerospace Sci. 29, 213 (1962).
[2] P. F. Jordan, AIAA Publication CP-8, 200 (1964), and J. AIAA (to be published).
[3] R. E. Langer, Phys. Rev. 51, 669 (1937).
[4] E.g., A. Erdélyi, J. Math. Phys. 1, 16 (1960).
[5] J. L. Sanders and A. A. Liepins, J. AIAA 1, 2105 (1963).

eters that describe the additional terms in a given equation and they can therefore be tabulated once and for all.

Three points are made: first, the influence functions arise from a straightforward inductive procedure and are largely simple algebraic combinations of already tabulated functions; second, the more significant influence functions show a convenient convergency behavior; third, approximate procedures require caution. Two approximate procedures were checked; though both appear plausible enough, neither of them sufficiently approximates even the first rank of influence functions.

II. FORMULATION

Consider the following generalized form of Liouville's differential equation

$$[\lambda^2 p + q]F + rF_\xi + sF_{\xi\xi} = u, \qquad (1)$$

where all quantities are supposed to be real valued, and λ^2 a large number. The functions p, q, r, s, and u are given functions of the coordinate ξ, and $F \equiv F(\xi)$ is the unknown function ($F_\xi \equiv dF(\xi)/d\xi$, etc.). Equation (1) is said to have a single transition point if $p(\xi)$ changes sign (within the ξ-range to be considered) and is said to have a double transition point if (after a trivial shift of the coordinate, if necessary) the function $p(\xi)$ can be written as

$$p(\xi) = \xi^2 \bar{p}(\xi)$$

with a finite function $\bar{p}$ which has no zero.

We assume here that Eq. (1) has a double transition point,[6] and assume also that $s(0) \neq 0$. Then, dividing by $\lambda^2\bar{p}(\xi)$, we may rewrite Eq. (1) as

$$F\xi^2 - \frac{1}{k^4}\{[1 + \xi d(\xi)]F_{\xi\xi} + e(\xi)F_\xi + g(\xi)F\} = l(\xi), \qquad (2)$$

where the large parameter is now k^4. We assume further that k^4 is positive.

Equation (2) describes specifically the problem[1,2] of a very thin pressurized toroidal shell of uniform wall thickness if

$$k^{-4} = \epsilon_0 a^2,$$

$$d(\xi) = \frac{1}{2a}(3 + \xi/a)(1 - \xi^2) - \xi,$$

$$e(\xi) = \frac{1}{2a}(4 + 3\xi/a)(1 - \xi^2), \qquad (3)$$

$$g(\xi) = \frac{-(1 - \xi^2)}{2a(a + \xi)},$$

$$-1 \leq \xi_1 \leq \xi \leq \xi_2 \leq 1.$$

[6] In this case, Eq. (1) is closely related to the homogeneous problem discussed by R. W. McKelvey, Trans. Am. Math. Soc. 79, 103 (1955).

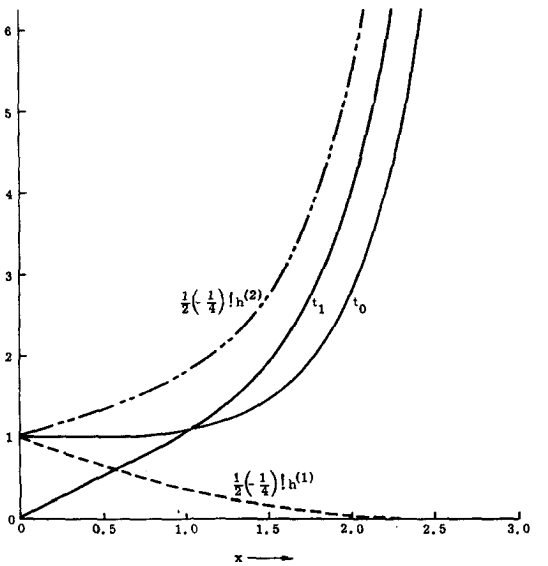

FIG. 1. Homogeneous limit solutions.

Here ϵ_0 is a reference strain and is thus small by definition, while $a > 1$ is the torus opening ratio. Equation (3) becomes modified[2] if the shell has a noncircular meridional cross section, nonuniform wall thickness, etc., and we are not going to assume here that our problem obeys Eq. (3) as given above. We assume, however, that we can conveniently write

$$d(\xi) = \sum_0^\infty d_\nu \xi^\nu; \cdots; l(\xi) = \sum_0^\infty l_\nu \xi^\nu, \qquad (4)$$

and that these series converge rapidly enough, in a sense that will become clear later, in the ξ-range to be considered. Equations (4) mean, in particular, that we will have solved Eq. (2) once we have obtained a set $n = 0, 1, 2 \cdots$ of suitable solutions[7] of the equation

$$F\xi^2 - \frac{1}{k^4}\{[1 + \xi d(\xi)]F_{\xi\xi} + e(\xi)F_\xi + g(\xi)F\} = \xi^n. \qquad (5)$$

Inspection of Eq. (5) indicates two ranges of general behavior of the solution F. In an "outer" range, $|\xi|$ sufficiently large, the "convergency condition" $F(\xi) \approx \xi^{n-2}$ should describe a possible solution. (Below, by making this condition precise, we will define a unique convergent solution.) On the other hand, in an "inner" range, $\xi \approx 0$, the terms in the $\{\ \}$ bracket, and in particular the differential terms, must become important (certainly if $n = 0$ or $n = 1$). In order to inspect more closely this inner range, we make the "stretching" transformation

$$k\xi = x; \qquad k^n F(\xi) = k^2 f(x) \qquad (6)$$

[7] In most applications only one or two solutions will be required. For example, if $l(\xi)$ has no zero, $F = lG$ transforms Eq. (2) into

$$G\xi^2 - \{\cdots lG \cdots\}/k^4 l = 1.$$

which transforms Eq. (5) with Eq. (4) into

$$fx^2 - f'' - \frac{1}{k}\sum_0^\infty \left(\frac{x}{k}\right)^\nu \left\{ x d_\nu f'' + e_\nu f' + \frac{1}{k} g_\nu f \right\} = x^n, \qquad (7)$$

where the prime ($'$) indicates differentiation with respect to x.

In Eq. (7) the limit process $k \to \infty$ can be made without difficulty. Thus arises the "limit form"

$$f_0 x^2 - f_0'' = x^n. \qquad (7a)$$

Applying the inverse of the transformation Eq. (6) to Eq. (7a) we obtain the limit form of Eq. (5)

$$F_0 \xi^2 - \frac{1}{k^4} F_{0,\xi\xi} = \xi^n. \qquad (7b)$$

A solution f_0 of Eq. (7a), respectively a solution F_0 of Eq. (7b), we henceforth refer to as a "limit solution."

Equation (7a) is a special form of the nonhomogeneous Bessel equation. The here relevant properties of f_0 are discussed in the next section; subsequently, the solution f of Eq. (7) is written as a formal expansion in terms of descending powers of k, with f_0 as the starting term.

III. LIMIT SOLUTION

Trying to solve Eq. (7a) by means of a power series in x, one is led to consider the functions $t_m (m = 0, 1, 2 \cdots)$ that are defined by

$$t_m \equiv x^m \left(1 + \frac{x^4}{(m+3)(m+4)} + \cdots\right),$$

TABLE I. Functions $h^{(1)}$, $h^{(2)}$, T_0, T_1, and first derivatives.

x	$(\frac{1}{4})!h^{(1)} \cdot 10^5$	$(\frac{1}{4})!h^{(1)\prime} \cdot 10^5$	$(\frac{1}{4})!h^{(2)}$	$(\frac{1}{4})!h^{(2)\prime}$	T_0	T_0'	T_1	T_1'
0.0	147 934	100 000	1.4793	1.0000	1.31103	0.00000	0.00000	0.59907
0.1	137 935	99 953	1.5794	1.0005	1.30604	-0.099564	0.059741	0.59409
0.2	127 952	99 645	1.6800	1.0043	1.29120	-0.19654	0.11849	0.57931
0.3	118 021	98 871	1.7805	1.0153	1.26690	-0.28844	0.17529	0.55526
0.4	108 198	97 482	1.8830	1.0380	1.23375	-0.37303	0.22926	0.52279
0.5	98 549	95 387	1.9886	1.0774	1.19260	-0.44838	0.27961	0.48301
0.6	89 146	92 550	2.0992	1.1395	1.14444	-0.51296	0.32566	0.43724
0.7	80 063	88 979	2.2175	1.2310	1.09041	-0.56570	0.36691	0.38697
0.8	71 373	84 727	2.3467	1.3596	1.03172	-0.60598	0.40296	0.33372
0.9	63 138	79 877	2.4910	1.5347	0.96963	-0.63367	0.43360	0.27905
1.0	55 413	74 538	2.6555	1.7672	0.90539	-0.64907	0.45877	0.22444
1.1	48 242	68 838	2.8468	2.0708	0.84020	-0.65288	0.47854	0.17123
1.2	41 654	62 909	3.0726	2.4625	0.77517	-0.64613	0.49310	0.12061
1.3	35 663	56 886	3.3429	2.9639	0.71129	-0.63009	0.50278	0.07353
1.4	30 275	50 895	3.6700	3.6030	0.64941	-0.60622	0.50795	0.03075
1.5	25 480	45 054	4.0693	4.4164	0.59025	-0.57604	0.50909	-0.00722
1.6	21 256	39 460	4.5607	5.4527	0.53436	-0.54112	0.50668	-0.04010
1.7	17 577	34 194	5.1694	6.7763	0.48214	-0.50293	0.50124	-0.06781
1.8	14 405	29 317	5.9284	8.4740	0.43384	-0.46288	0.49329	-0.09044
1.9	11 699	24 870	6.8806	10.663	0.38958	-0.42219	0.48331	-0.10826
2.0	9415.4	20 875	8.0827	13.504	0.34939	-0.38192	0.47178	-0.12163
2.1	7508.6	17 337	9.6102	17.214	0.31316	-0.34294	0.45912	-0.13099
2.2	5933.1	14 247	11.564	22.097	0.28073	-0.30591	0.44570	-0.13685
2.3	4644.9	11 585	14.083	28.572	0.25189	-0.27132	0.43185	-0.13972
2.4	3602.7	9322.2	17.351	37.226	0.22638	-0.23948	0.41784	-0.14012
2.5	2768.3	7422.7	21.628	48.886	0.20390	-0.21056	0.40389	-0.13853
2.6	2107.3	5848.7	27.267	64.723	0.18417	-0.18460	0.39019	-0.13540
2.7	1589.1	4560.5	34.767	86.414	0.16688	-0.16153	0.37685	-0.13112
2.8	1187.0	3519.0	44.824	116.37	0.15177	-0.14123	0.36399	-0.12605
2.9	878.27	2687.3	58.431	158.09	0.13855	-0.12349	0.35166	-0.12047
3.0	643.69	2030.9	77.003	216.70	0.12699	-0.10810	0.33990	-0.11463
3.1	467.28	1519.0	102.58	299.72	0.11686	-0.09483	0.32873	-0.10870
3.2	335.98	1124.4	138.13	418.36	0.10796	-0.08342	0.31816	-0.10284
3.3	239.28	823.72	187.94	589.37	0.10012	-0.07364	0.30816	-0.09714
3.4	168.77	597.25	258.58	838.04	0.09319	-0.06528	0.29872	-0.09168
3.5	117.90	428.60	359.44	1202.8	0.08703	-0.05812	0.28982	-0.08651
3.6	81.571	304.41	504.95	1742.7	0.08153	-0.05200	0.28141	-0.08164
3.7	55.894	214.00	716.82	2549.0	0.07660	-0.04675	0.27348	-0.07709
3.8	37.930	148.90	1028.3	3763.9	0.07216	-0.04223	0.26598	-0.07286
3.9	25.491	102.54	1490.5	5610.9	0.06813	-0.03832	0.25890	-0.06893
4.0	16.966	69.897	2183.1	8444.8	0.06448	-0.03493	0.25219	-0.06529
4.1	11.183	47.159	3230.9	12 832	0.06113	-0.03197	0.24583	-0.06192
4.2	7.2996	31.494	4831.3	19 688	0.05807	-0.02938	0.23979	-0.05881
4.3	4.7185	20.818	7299.4	30 500	0.05525	-0.02709	0.23406	-0.05592
4.4	3.0205	13.621	11 143	47 707	0.05264	-0.02506	0.22860	-0.05325
4.5	1.9147	8.8217	17 186	75 347	0.05023	-0.02325	0.22340	-0.05077
4.6	1.2019	5.6553	26 780	120 160	0.04799	-0.02162	0.21844	-0.04847
4.7	0.74717	3.5887	42 160	193 499	0.04590	-0.02016	0.21370	-0.04632
4.8	0.45993	2.2542	67 058	314 644	0.04395	-0.01884	0.20917	-0.04432
4.9	0.28036	1.4016	107 758	516 641	0.04213	-0.01764	0.20483	-0.04246
5.0	0.16923	0.86261	174 939	856 629	0.04042	-0.01654	0.20068	-0.04071

$$\equiv \sum_0^\infty a_\nu x^{m+4\nu}; \quad a_\nu = \frac{a_{\nu-1}}{(m+4\nu-1)(m+4\nu)}, \tag{8}$$

so that

$$t_m x^2 - t_m'' = -m(m-1)x^{m-2}. \tag{8a}$$

Equation (8a) shows in particular that t_0 and t_1 are homogeneous solutions of Eq. (7a). These are shown in Fig. 1. As t_0 is symmetric, t_1 antisymmetric, they form a complete set of solutions. Their Wronskian is $W \equiv 1$.

Often, in particular in boundary layer problems, it is more convenient to work with solutions which converge in the manner of the Hankel functions $H_\nu^{(1)}$ and $H_\nu^{(2)}$. We define corresponding solutions by

$$h^{(1)} \equiv \frac{2t_0}{(-1/4)!} - \frac{t_1}{(1/4)!}$$

$$\left[\equiv x^{\frac{1}{2}} i^{5/4} H_{\frac{1}{4}}^{(1)}\left(\frac{ix^2}{2}\right) \quad \text{if} \quad x > 0 \right], \tag{9}$$

and

$$h^{(2)}(x) = h^{(1)}(-x). \tag{10}$$

Their asymptotic behavior is given by

$$h^{(1)} \sim \frac{2}{(\pi x)^{\frac{1}{2}}}\left(1 - \frac{3}{16x^2} + \cdots\right)e^{-ix^2/2} \quad \text{as} \quad x \to +\infty,$$

$$\sim 2\left(\frac{2}{\pi x}\right)^{\frac{1}{2}}\left(1 + \frac{3}{16x^2} + \cdots\right)e^{ix^2/2} \quad \text{as} \quad x \to -\infty. \tag{11}$$

The two functions $h^{(1)}$ and $h^{(2)}$ are also shown in Fig. 1; they form another complete set of homogeneous solutions of Eq. (7a). A brief tabulation is given in Table I (see Appendix B).

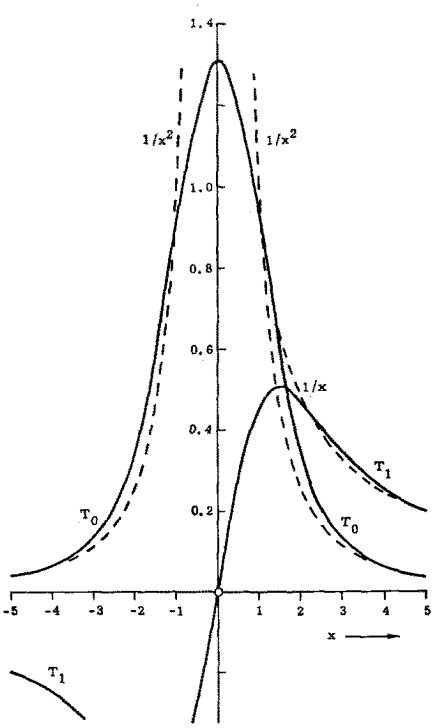

FIG. 2. Nonhomogeneous limit solutions.

TABLE II. T_n functions: formulas.

$$T_0 = T_0$$
$$T_1 = T_1$$
$$T_2 = 1$$
$$T_3 = x$$
$$T_4 = 2T_0 + x^2$$
$$T_5 = 6T_1 + x^3$$
$$T_6 = 12 + x^4$$
$$T_7 = 20x + x^5$$
$$T_8 = 60T_0 + 30x^2 + x^6$$
$$T_9 = 252T_1 + 42x^3 + x^7$$
$$T_{10} = 672 + 56x^4 + x^8$$
$$T_{11} = 1440x + 72x^5 + x^9$$
$$T_{12} = 5400T_0 + 2700x^2 + 90x^6 + x^{10}$$
$$T_{13} = 27\,720T_1 + 4620x^3 + 110x^7 + x^{11}$$
$$T_{14} = 88\,704 + 7392x^4 + 132x^8 + x^{12}$$
$$T_{15} = 224\,640x + 11\,232x^5 + 156x^9 + x^{13}$$

In discussing next the nonhomogeneous solution f_0 of Eq. (7a), we are no longer confined, as we might be in a discussion of Eq. (5), to a given coordinate range $\xi_1 \leq \xi \leq \xi_2$, and can hence state the convergency condition in a precise form. Denoting by T_n a convergent limit solution f_0 for a given number n, we have

$$T_n x^2 - T_n'' = x^n \tag{12}$$

and write the convergency condition as

$$T_n \sim x^{n-2} \quad \text{as} \quad x \to \pm\infty. \tag{12a}$$

As no homogeneous solution obeys a corresponding convergency condition, there can be at most one function T_n for each n.

From Eqs. (12) and (12a) immediately

$$T_2 = 1; \qquad T_3 = x;$$
$$T_{n+4} = x^{n+2} + (n+1)(n+2)T_n. \tag{13}$$

The functions T_n are thus readily written algebraic functions of x if $n = 2 \pmod 4$ or $n = 3 \pmod 4$, and are readily expressed by T_0 if $n = 0 \pmod 4$ and by T_1 if $n = 1 \pmod 4$. It remains to discuss T_0 and T_1.

From Eq. (8a) and from considerations of symmetry it follows that T_0 and T_1, assuming these two functions to exist, can be written as

$$T_0 = c_0 t_0 - \tfrac{1}{2}t_2; \qquad T_1 = c_1 t_1 - \tfrac{1}{6}t_3, \tag{14}$$

with constants c_0 and c_1 still to be determined. To answer the question of existence, and to determine c_0 and c_1, requires knowledge of the asymptotic behavior of the functions t_2 and t_3 [that of t_0 and t_1

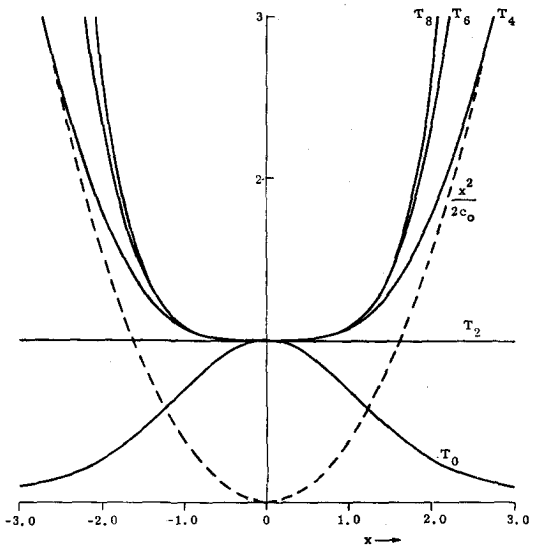

FIG. 3. Normalized functions $T_{2n}(x)/T_{2n}(0)$.

can be read from Eqs. (9) and (11)]. This knowledge is provided by the theory of Lommel functions.[8] The two constants are[9]

$$c_0 = \frac{(1/4)!}{(-1/4)!}\pi^{\frac{1}{2}} = 1.31103\cdots, \qquad c_1 = \frac{\pi}{4c_0} = 0.59907\cdots. \tag{14a}$$

The functions T_0 and T_1 are tabulated in Table I. They are shown graphically in Fig. 2, together with the functions toward which they converge as $|x|\to\infty$. For many practical purposes, convergency is complete if $|x| > 4$.

The first $16T_n$ functions are written out explicitly in Table II. Figure 3 illustrates the behavior of T_n as a function of n. Shown are normalized functions T_{2n} and also the asymptotic curve of T_4.

Formal development in an asymptotic series with descending powers of x leads to

$$T_n = x^{n-2}\left[1 + \frac{(n-2)(n-3)}{x^4} \times \left(1 + \frac{(n-6)(n-7)}{x^4}(1 + \cdots)\right)\right]. \tag{15}$$

Having described the limit solution in terms of the coordinate x, we now return to the coordinate ξ. Figure 4 illustrates the convergent solution F_0 of Eq. (7b) in the case that $n = 0$, i.e., $F_0 = k^2T_0(k\xi)$.

[8] See Appendix A. The functions T_0 and T_1 are not themselves Lommel functions. They were first defined and tabulated by Sanders and Liepins (Ref. 5) who denote them by T_1 and T_2.

[9] Equation (14a) follows directly from Eq. (14) if one requires that, because of Eq. (12a), the νth term in each power series Eq. (14) converges, as $\nu \to \infty$, towards the negative geometric mean of its two neighboring terms.

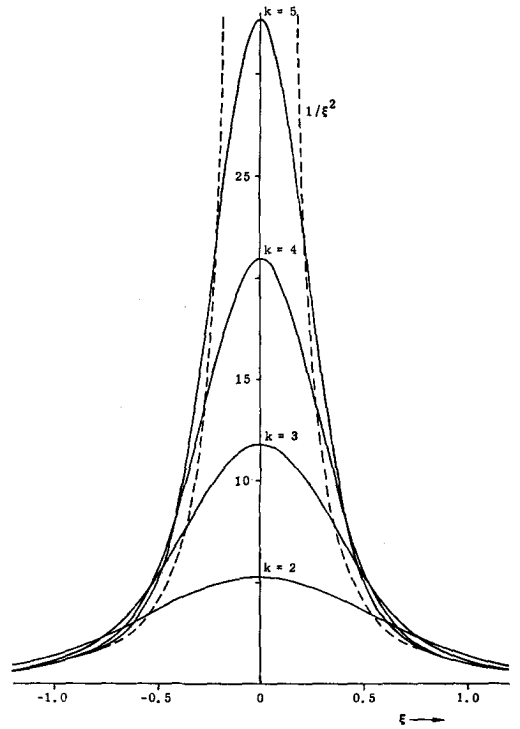

FIG. 4. Limit solution $F_0 = k^2T_0(k\xi)$.

In the outer range this solution converges toward the function $1/\xi^2$ which is independent of k. The inner range contracts as k increases; simultaneously, the central value $F_0(0) = c_0k^2$ increases.

Another illustration, Fig. 5, refers to the equation

$$F_0\xi^2 - \frac{1}{k^4}F_{0,\xi\xi} = (1 - \xi^2)\xi^2. \tag{16}$$

In a problem of this type, one is tempted to divide by ξ^2 and, in the present case, to write $F_0 \approx 1 - \xi^2$. Figure 5 compares this approximation with the exact

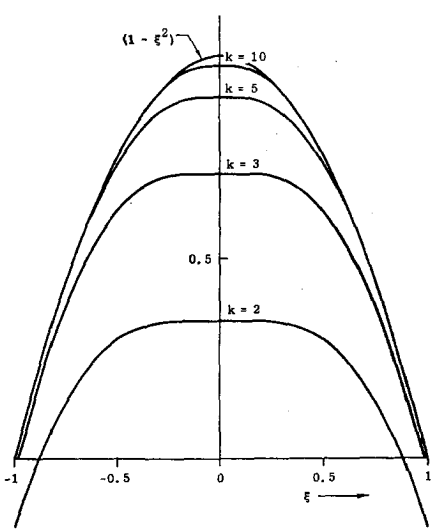

FIG. 5. Solutions of Eq. (16) for several values of k.

convergent solution

$$F_0 = 1 - \xi^2 - \frac{2}{k^2} T_0(k\xi). \tag{16a}$$

Note that, for all values of k, the exact solution F_0 has zero curvature at the transition point $\xi = 0$.

Insofar as Eq. (16) is a representative problem of membrane shell theory, the range of definition might be $-1 \leq \xi \leq 1$, and boundary conditions $F_0(\pm 1) = 0$ might be given. In order to fulfill such (or similar) boundary conditions, a suitable homogeneous solution has to be added to the F_0 curves of Fig. 5. Note that to the ξ-range of Fig. 5 corresponds the range $-k \leq x \leq k$ of Fig. 1, and that the steepness of the homogeneous solutions at the end points $x = \pm k$ increases rapidly as k is increased. Consequently, the practical effect of a homogeneous correction will increasingly be confined to close neighborhoods of the end points.

IV. FORMAL SOLUTION

The limit solution F_0 of Eq. (7b) is the solution of Eq. (5) in the special case that the parameter functions, $d(\xi)$, $e(\xi)$ and $g(\xi)$, are identically zero. We investigate now the modifications to the solution of Eq. (5) which arise from nonzero parameter functions.

We consider first the outer range, writing the solution of Eq. (5) in the form

$$F = F_0 + \frac{1}{k^4} F_1 + \frac{1}{k^8} F_2 + \cdots . \tag{17}$$

We do not, however, imply by this formulation that F_0 be independent of k. Rather, as before, we understand F_0 to be a limit solution, and, furthermore, we confine our attention to the case that F_0 is convergent. From Eqs. (6) and (15) thus

$$F_0 \sim \left[1 + \frac{(n-2)(n-3)}{(k\xi)^4} + \cdots\right]\xi^{n-2}. \tag{17a}$$

Inserting Eqs. (17), (17a) into Eq. (5), multiplying by k^4, and making the transition $k \to \infty$, we obtain

$$\frac{1}{k^4} F_1 = \frac{1}{kx^3}\left[(n-2)((n-3)d(\xi) + e(\xi)) + \xi g(\xi)\right]\xi^{n-2}. \tag{18}$$

By continuing this process we could obtain F_2, then F_3, and so on.

Equation (18) is written such that Eqs. (17a) and (18) have the same factor ξ^{n-2} on the right, and that the two k factors, $1/k^4$ on the left, $1/k$ on the right, explain an apparent discrepancy between Eqs. (5) and (7): according to the former, the large parameter is k^4 while, on transforming to the x coordinate, k rather than k^4 becomes the parameter.

A significant observation can be read from Eqs. (17a) and (18): the corrective term F_1, apart from having the factor $1/k^4$, will itself become small compared with F_0 as $|\xi| \to \infty$ if the parameter functions are such that, for the [ ] bracket in Eq. (18),

$$[\cdots]\xi^{-3} \to 0 \quad \text{for} \quad |\xi| \to \infty . \tag{18a}$$

In order to discuss the inner range, we abbreviate our notation by defining

$$M(f) \equiv fx^2 - f'', \qquad N(f) \equiv \sum_0^\infty \left(\frac{x}{k}\right)^\nu \left\{x d_\nu f'' + e_\nu f' + \frac{1}{k} g_\nu f\right\}, \tag{19}$$

so that Eq. (7) now reads

$$M(f) - \frac{1}{k} N(f) = x^n.$$

Writing, in correspondence with Eq. (17)

$$f = f_0 + \frac{1}{k} f_1 + \frac{1}{k^2} f_2 + \cdots , \tag{20}$$

and again letting f_0 be the convergent limit solution (i.e., $f_0 \equiv T_n$), we find first

$$M(f_1) = N(T_n) \quad \text{or} \quad f_1 = M^{-1}N(T_n) \tag{21}$$

and, generally,

$$f_\mu = M^{-1}N(f_{\mu-1}). \tag{22}$$

The task of determining the sequence of corrective functions f_μ is thus equivalent to the task of successively inverting the operator M. This task is performed in the next section.

Continuing here the discussion of the formal aspect of the solution, we note that, as the parameter k appears explicitly in $N(f)$, Eq. (19), each of the corrections f_μ is again a series in descending powers of k. Combining terms with equal powers of k, we write the solution F of Eq. (5) in the following final form:

$$\begin{aligned} F(\xi) \equiv k^{2-n} f(x) = \{&k^2 T_n + k[\langle d_0\rangle + \langle e_0\rangle] \\ &+ [\langle d_1\rangle + \langle e_1\rangle + \langle g_0\rangle + \langle d_0^2\rangle + \langle d_0 e_0\rangle + \langle e_0^2\rangle] \\ &+ k^{-1}[\langle d_2\rangle + \langle e_2\rangle + \langle g_1\rangle + \langle d_0 d_1\rangle + \langle d_0 e_1\rangle + \langle d_0 g_0\rangle \\ &+ \langle d_1 e_0\rangle + \langle e_0 e_1\rangle + \langle e_0 g_0\rangle + \langle d_0^3\rangle + \langle d_0^2 e_0\rangle \\ &+ \langle d_0 e_0^2\rangle + \langle e_0^3\rangle] + k^{-2}[\cdots] + \cdots\} k^{-n}. \end{aligned} \tag{23}$$

Each term within angular brackets in Eq. (23) represents the influence of the parameter coefficient, or the product of parameter coefficients, that is written within the brackets. In each term, this influence is a function which, for a given exponent n, is a function of x only, and which thus can be determined once and for all. These influence functions we will denote by Greek letters. A few illustrative examples follow:

TABLE III. Survey and convergency of influence functions.

k-Factor in Eq.(23)	Function	Rank μ	Dominant Power of x as $\lvert x\rvert \to \infty$
$k^{2-n}\cdot$	k^1: T_n	0	1
	$\frac{1}{k}$: δ_0, ϵ_0	1	x^{-3}
	$\frac{1}{k^2}$: $\delta_1, \epsilon_1, \gamma_0$	1	x^{-2}
	$\frac{1}{k^2}$: $\langle\epsilon_0^2\rangle, \langle\delta_0\epsilon_0\rangle, \langle\epsilon_0^2\rangle$	2	x^{-6}
	$\frac{1}{k^3}$: $\delta_2, \epsilon_2, \gamma_1$	1	x^{-1}
	$\frac{1}{k^3}$: $\langle\delta_0\delta_1\rangle, \langle\delta_0\epsilon_1\rangle, \langle\delta_0\gamma_0\rangle, \langle\delta_1\epsilon_0\rangle, \langle\epsilon_0\epsilon_1\rangle, \langle\epsilon_0\gamma_0\rangle$	2	x^{-5}
	$\frac{1}{k^3}$: $\langle\epsilon_0^3\rangle, \langle\epsilon_0^2\epsilon_0\rangle, \langle\epsilon_0\epsilon_0^2\rangle, \langle\epsilon_0^3\rangle$	3	x^{-9}
			$\cdot\, x^{n-2}$

$$\langle d_\nu\rangle \equiv d_\nu\cdot\delta_\nu; \qquad \delta_\nu = M^{-1}(x^{\nu+1}T_n'');$$
$$\langle e_\nu\rangle \equiv e_\nu\cdot\epsilon_\nu; \qquad \langle g_\nu\rangle \equiv g_\nu\cdot\gamma_\nu;$$
$$\langle d_\nu^2\rangle \equiv d_\nu^2\cdot\langle\delta_\nu^2\rangle; \qquad \langle d_\nu e_\mu\rangle \equiv d_\nu e_\mu\cdot\langle\delta_\nu\epsilon_\mu\rangle; \tag{23a}$$
$$\langle\delta_\nu\epsilon_\mu\rangle = M^{-1}(x^{\nu+1}e_\mu'' + x^\mu\delta_\nu').$$

As these examples show, angular brackets will be used to distinguish higher-rank influence functions from products. If the number n has to be indicated, this will be done by an additional suffix (e.g., $\delta_{\nu,0} \equiv \delta_\nu$ for $n = 0$).

Table III refers to the discussion above, in connection with Eqs. (18) and (18a), of the convergency behavior of the corrective functions F_μ. Listed in Table III is the dominant power of x for each one of the terms that is written explicitly in Eq. (23). Also indicated is the rank μ, that is, the function f_μ from which each influence function originates.

V. INFLUENCE FUNCTIONS

Equation (23) reduces the task of solving Eq. (5) to the task of determining a set of influence functions. By itself, of course, this achieves little; formal developments are readily written, but to determine analytically the functions that appear in such developments usually remains a rather formidable proposition. The key observation of the present paper is that, in the present case, this task is performed with relative ease.

An illustrative example follows. Owing to

$$M(T_0') \equiv x^2T_0' - (T_0'')' = -2xT_0,$$

we have

$$\gamma_{1,0} = M^{-1}(xT_0) = -\tfrac{1}{2}T_0'.$$

That is, the influence function $\gamma_{1,0}$ is a function that, essentially, has already been tabulated.

This statement can be generalized to a large extent. In order to calculate first f_1 according to Eq. (21), we reduce $N(T_n)$ by means of Table II and Eq. (12) to a sum of members of the following family of functions:

$$x^p;\ x^pT_0;\ x^pT_0';\ x^pT_1;\ x^pT_1' \qquad (p = 0, 1, 2, \cdots).$$

To these the inverse operator M^{-1} has to be applied. As $M^{-1}(x^p) = T_p$ can again be reduced to members of the same family, it remains to try to perform a similar inversion with the other members of the family. To the extent that this can be achieved, we can then proceed to form $N(f_1)$, again in terms of the same family of functions, find f_2 according to Eq. (22), and so on.

In most cases, the inversion is readily performed by starting from the formulas

$$M(x^pT_n) = x^{n+p} - p(p-1)x^{p-2}T_n - 2px^{p-1}T_n', \tag{24}$$
$$M(x^pT_n') = (n+2p)x^{n+p-1} - 2(p+1)x^{p+1}T_n - p(p-1)x^{p-2}T_n',$$

eliminating on the right either T_n or T_n', and applying the inverse operator M^{-1}. Results of this procedure are listed in Table IV. Where the procedure leads to difficulties, this is indicated by the appearance of two new functions, U_0 and U_1, in Table IV. These functions will be discussed later.

Table V lists in detail the formulas that result from this procedure for the more important influence functions, namely, for those which have a (combined) factor k^1, k^0 or k^{-1} in Eq. (23). Table V is confined to $n = 0$ and $n = 1$ because the two influence functions for $n = 2$ which would qualify (that is, $\delta_{0,2}$ and $\epsilon_{0,2}$) are both identically zero. Note that generally

$$\delta_{\nu,2} = \delta_{\nu,3} = \epsilon_{\nu,2} = 0, \qquad \epsilon_{\nu,3} = \gamma_{\nu,2} = \gamma_{\nu-1,3} = T_\nu, \tag{25a}$$

and for $n \geq 4$, from Eq. (13),

$$\delta_{\nu,n} = (n-2)(n-3)[T_{n+\nu-3} + \delta_{\nu,n-4}],$$
$$\epsilon_{\nu,n} = (n-2)[T_{n+\nu-3} + (n-3)\epsilon_{\nu,n-4}], \tag{25b}$$
$$\gamma_{\nu,n} = T_{n+\nu-2} + (n-2)(n-3)\gamma_{\nu,n-4}.$$

Table VI refers to the asymptotic behavior of the influence functions that are listed in Table V but is valid for all n. Given are the first two terms of the respective asymptotic series as read from F_1, Eq. (18), and F_2. Tables VII, VIII, and IX give numerical values; these are plotted in Figs. 6(a), (b).

The analytical presentation, Table V, of the influence functions would become increasingly tedious if it would have to be extended to higher ranks of influence functions; in this case, numerical integration of Eq. (22) would become preferable. In fact, numerical integration had to be used to determine the functions

$$U_n = M^{-1}(T_n) \ (n = 0, 1),$$

TABLE IV. Results of applying the operator M^{-1}.

f	$M^{-1}(f) =$				
(n = 0)	1	T_0	T_0'	T_1	U_0
T_0	0	0	0	0	1
$2x\,T_0$	0	0	-1	0	0
$4x^2\,T_0$	0	2	$-x$	0	0
$6x^3\,T_0$	0	x	$-x^2$	3	0
$8x^4\,T_0$	9/2	$3x^2/2$	$-x^3$	0	3
$10x^5\,T_0$	$6x$	$2x^3$	$-(6+x^4)$	0	0
$12x^6\,T_0$	$15x^2/2$	$30+5x^4/2$	$-(15/2+x^4)\,x$	0	0
$14x^7\,T_0$	$9x^3$	$(10+3x^4)\,x$	$-(10+x^4)\,x^2$	84	0
$2\,T_0'$	0	$-x$	0	1	0
$4x\,T_0'$	1	$-x^2$	0	0	-2
$6x^2\,T_0'$	x	$-x^3$	3	0	0
$8x^3\,T_0'$	x^2	$-(4+x^4)$	$3x$	0	0
$10x^4\,T_0'$	x^3	$-(10/3+x^4)\,x$	$10x^2/3$	-4	0
$12x^5\,T_0'$	$x^4-39/8$	$-(45/8+x^4)\,x^2$	$15x^3/4$	0	-45/4
$14x^6\,T_0'$	$(x^4-26/5)\,x$	$-(42/5+x^4)\,x^3$	$21(6+x^4)/5$	0	0

(n = 1)	1	T_1	T_1'	T_0	U_1
T_1	0	0	0	0	1
$2x\,T_1$	0	0	-1	1	0
$4x^2\,T_1$	0	3	$-x$	0	0
$6x^3\,T_1$	4	x	$-x^2$	0	0
$8x^4\,T_1$	$11x/2$	$3x^2/2$	$-x^3$	0	3
$10x^5\,T_1$	$7x^2$	$2x^3$	$-(6+x^4)$	20	0
$12x^6\,T_1$	$17x^3/2$	$147/2+5x^4/2$	$-(15/2+x^4)\,x$	0	0
$14x^7\,T_1$	$160+10x^4$	$(10+3x^4)\,x$	$-(10+x^4)\,x^2$	0	0
$2\,T_1'$	1	$-x$	0	0	0
$4x\,T_1'$	x	$-x^2$	0	0	-2
$6x^2\,T_1'$	x^2	$-x^3$	3	-1	0
$8x^3\,T_1'$	x^3	$-(3+x^4)$	$3x$	0	0
$10x^4\,T_1'$	$-4/3+x^4$	$-(10/3+x^4)\,x$	$10x^2/3$	0	0
$12x^5\,T_1'$	$(-5/8+x^4)\,x$	$-(45/8+x^4)\,x^2$	$15x^3/4$	0	-45/4
$14x^6\,T_1'$	$(3/5+x^4)\,x^2$	$-(42/5+x^4)\,x^3$	$21(6+x^4)/5$	-24	0

Example: $6\,M^{-1}(x^3\,T_0) = x\,T_0 - x^2\,T_0' + 3\,T_1$.

TABLE V. Influence functions: formulas.

I	A	B	C	D	E	F	G	K	L	M	N	
$6x^2\,\delta_0$	-1	3	-1									
$2x^2\,\epsilon_0$	1	-1										
$16x\,\delta_1$			-2						3			
$4x\,\epsilon_1$									-1			
$2x\,\gamma_0$	1								1			
$192x\langle\delta_0^2\rangle$		16	10	16		-4			-15			
$12x\langle\delta_0\epsilon_0\rangle$	1	-4	1	-1					3			
$8x\langle\epsilon_0^2\rangle$	-2	2							-1			
$10\,\delta_2$	-2		-1		-3							
$12\,\epsilon_2$	2				3							
$4\,\gamma_1$					-1							
$480\langle\delta_0\delta_1\rangle$	79		37	30	96	-15					15	n = 0
$24\langle\delta_0\epsilon_1\rangle$	-3		1		-4						-1	
$12\langle\delta_0\gamma_0\rangle$		-3			1						1	
$96\langle\delta_1\epsilon_0\rangle$	-16		6	-6	-24				-9	9		
$8\langle\epsilon_0\epsilon_1\rangle$					1				1	-1		
$4\langle\epsilon_0\gamma_0\rangle$	-1	1							-1	1		
$3456\langle\delta_0^3\rangle$	-173		-79	-90	-192	65	9	-1			-45	
$384\langle\delta_0^2\epsilon_0\rangle$	48	-16	-26	-6	64	4	-1		15	-15	16	
$48\langle\delta_0\epsilon_0^2\rangle$	-1	8	-1	2	-4				-6	6	-1	
$48\langle\epsilon_0^3\rangle$	4	-6							3	-3		
$6x\,\delta_0$		-1		-1								
$2x\,\epsilon_0$		1										
$16\,\delta_1$				-2						3		
$4\,\epsilon_1$										-1		
$2\,\gamma_0$		1								1		n = 1
$192\langle\delta_0^2\rangle$				10			-1			-15		
$12\langle\delta_0\epsilon_0\rangle$		1		1						3		
$8\langle\epsilon_0^2\rangle$		-2								-1		

Explanations:

Each of the functions, I, listed on the left margin has the form:

$$I = aA + bB + cC + \cdots + nN$$

The coefficients a, b, c, · · · n are the integer numbers that are listed in the body of the table. All coefficients not listed are zero. The functions A, B, C, · · · N are:

$A = (1 - x^2\,T_0)\,x$ $\quad F = (2/3)\left[Ax^4 + 6x\right]$ $\quad N = 2x\,(U_0 - x\,U_0') - 3\,M$

$B = (1 - x\,T_1)\,x$ $\quad G = (8/3)\left[Bx^4 + 2x\right]$

$C = (2 + x^3\,T_0')\,x$ $\quad K = (8/3)\left[Cx^4 + 36x\right]$ $\quad$ Example: $4\gamma_1 = -2\,T_0'$

$D = (1 + x^2\,T_1')\,x$ $\quad L = 2\,U_0 x - A$

$E = 2\,T_0'$ $\quad M = 2\,U_1 - B$

which do not seem to be expressible in terms of T functions. Further functions V_n, W_n $\cdots$ would appear if Table V would be extended to higher-rank influence functions:

$$V_n = M^{-1}(U_n),$$

$$W_n = M^{-1}(V_n) = M^{-2}(U_n) = M^{-3}(T_n).$$

The computation of these functions is described in Appendix B; numerical values are given in Table X.

VI. DISCUSSION

Equation (23), in connection with the graphs of the T_n functions, Figs. 2 and 3, and of the influence functions, Fig. 6, and with Eqs. (25a), (25b), provides a direct insight into the qualitative behavior of the solutions of Eq. (5). To what degree of accuracy this limited number of influence functions determines a numerical solution to a given problem depends upon the parameters involved, that is, the numbers k and n and the parameter coefficients d_ν, e_ν, and g_ν.

In Fig. 6, the influence functions are grouped according to the k factor which they have in Eq. (23). The influence functions are odd if the power of k is odd, and even when this power is even.

To illustrate the application, assume that all parameter coefficients are zero except either d_0, e_0,

TABLE VI. Influence functions: leading terms of asymptotic developments.

		x^{-4}	x^{-8}
$\frac{1}{x^{n-1}}\cdot$	δ_0	$(n-2)(n-3)$	$2(n-2)(n-3)(n-6)^2$
	ϵ_0	$(n-2)$	$2(n-2)(n-4)(n-6)$
$\frac{1}{x^n}\cdot$	δ_1	$(n-2)(n-3)$	$2(n-2)(n-3)\left[n^2-11n+31\right]$
	ϵ_1	$(n-2)$	$2(n-2)\left[n^2-9n+19\right]$
	γ_0	1	$2\left[n^2-7n+13\right]$
	$\langle\delta_0^2\rangle$	0	$(n-2)(n-3)(n-5)(n-6)$
	$\langle\delta_0\epsilon_0\rangle$	0	$(n-2)(n-5)(2n-9)$
	$\langle\epsilon_0^2\rangle$	0	$(n-2)(n-5)$
$\frac{1}{x^{n+1}}\cdot$	δ_2	$(n-2)(n-3)$	$2(n-2)(n-3)\left[n^2-10n+27\right]$
	ϵ_2	$(n-2)$	$2(n-2)(n-3)(n-5)$
	γ_1	1	$2(n-3)^2$
	$\langle\delta_0\delta_1\rangle$	0	$2(n-2)(n-3)(n-5)^2$
	$\langle\delta_0\epsilon_1\rangle$	0	$(n-2)(n-5)(2n-7)$
	$\langle\delta_0\gamma_0\rangle$	0	$2\left[n^2-7n+13\right]$
	$\langle\delta_1\epsilon_0\rangle$	0	$2(n-2)\left[n^2-9n+21\right]$
	$\langle\epsilon_0\epsilon_1\rangle$	0	$(n-2)(2n-9)$
	$\langle\epsilon_0\gamma_0\rangle$	0	$2(n-3)$
	$\langle\delta_0^3\rangle$	0	0
	$\langle\delta_0^2\epsilon_0\rangle$	0	0
	$\langle\delta_0\epsilon_0^2\rangle$	0	0
	$\langle\epsilon_0^3\rangle$	0	0

Example: $\delta_0 \sim (n-2)(n-3)\,x^{n-5}\left(1+2(n-6)^2\,x^{-4}+\cdots\right)$

or g_0. Equation (7) reduces respectively to

$$fx^2 - f''\left(1 + \frac{d_0}{k}x\right) = x^n, \qquad \text{(i)}$$

$$fx^2 - f'\frac{e_0}{k} - f'' = x^n, \qquad \text{(ii)} \qquad (26)$$

$$f\left(x^2 - \frac{g_0}{k^2}\right) - f'' = x^n, \qquad \text{(iii)}$$

and the respective solutions are

$$f = T_n + \begin{cases} \dfrac{d_0}{k}\,\delta_0 + \left(\dfrac{d_0}{k}\right)^2\langle\delta_0^2\rangle + \left(\dfrac{d_0}{k}\right)^3\langle\delta_0^3\rangle + \cdots, \\ \dfrac{e_0}{k}\,\epsilon_0 + \left(\dfrac{e_0}{k}\right)^2\langle\epsilon_0^2\rangle + \left(\dfrac{e_0}{k}\right)^3\langle\epsilon_0^3\rangle + \cdots, \\ \dfrac{g_0}{k^2}\,U_n + \left(\dfrac{g_0}{k^2}\right)^2 V_n + \left(\dfrac{g_0}{k^2}\right)^3 W_n + \cdots. \end{cases} \qquad (26a)$$

For cases (i) and (ii), the first three ranks of influence functions are shown in Fig. 6(a) ($n = 0$);

TABLE VII. Influence functions: numerical values ($n = 0$).

x	δ_0	ϵ_0	δ_1	ϵ_1	γ_0	$\langle\delta_0^2\rangle$	$\langle\delta_0\epsilon_0\rangle$	$\langle\epsilon_0^2\rangle$	δ_2	ϵ_2	γ_1
0.0	0.0000	0.0000	0.0840	-0.4453	1.3906	0.0483	0.2786	-0.2226	0.0000	0.0000	0.0000
0.1	-0.0079	-0.0354	0.0840	-0.4453	1.3840	0.0483	0.2790	-0.2209	0.0200	-0.0333	0.0498
0.2	-0.0149	-0.0699	0.0841	-0.4452	1.3646	0.0482	0.2801	-0.2156	0.0400	-0.0667	0.0983
0.3	-0.0200	-0.1024	0.0847	-0.4449	1.3328	0.0477	0.2815	-0.2071	0.0601	-0.0999	0.1442
0.4	-0.0224	-0.1321	0.0861	-0.4442	1.2896	0.0465	0.2827	-0.1957	0.0806	-0.1330	0.1865
0.5	-0.0217	-0.1583	0.0890	-0.4426	1.2362	0.0441	0.2830	-0.1817	0.1016	-0.1657	0.2242
0.6	-0.0176	-0.1805	0.0938	-0.4400	1.1739	0.0400	0.2817	-0.1658	0.1239	-0.1977	0.2565
0.7	-0.0100	-0.1982	0.1011	-0.4358	1.1044	0.0341	0.2782	-0.1485	0.1478	-0.2285	0.2829
0.8	0.0007	-0.2112	0.1111	-0.4298	1.0294	0.0260	0.2719	-0.1304	0.1741	-0.2577	0.3030
0.9	0.0142	-0.2195	0.1240	-0.4217	0.9507	0.0160	0.2624	-0.1121	0.2031	-0.2846	0.3168
1.0	0.0297	-0.2233	0.1397	-0.4114	0.8700	0.0043	0.2494	-0.0940	0.2354	-0.3088	0.3245
1.1	0.0464	-0.2228	0.1576	-0.3986	0.7890	-0.0086	0.2330	-0.0768	0.2710	-0.3295	0.3264
1.2	0.0636	-0.2186	0.1772	-0.3836	0.7090	-0.0220	0.2136	-0.0607	0.3096	-0.3463	0.3230
1.3	0.0802	-0.2109	0.1977	-0.3663	0.6315	-0.0352	0.1916	-0.0460	0.3506	-0.3588	0.3150
1.4	0.0956	-0.2006	0.2182	-0.3470	0.5576	-0.0473	0.1677	-0.0331	0.3930	-0.3668	0.3031
1.5	0.1090	-0.1881	0.2375	-0.3260	0.4880	-0.0576	0.1426	-0.0219	0.4357	-0.3700	0.2880
1.6	0.1200	-0.1742	0.2549	-0.3038	0.4235	-0.0655	0.1172	-0.0126	0.4770	-0.3687	0.2706
1.7	0.1282	-0.1592	0.2693	-0.2806	0.3645	-0.0706	0.0923	-0.0050	0.5156	-0.3629	0.2515
1.8	0.1335	-0.1438	0.2802	-0.2570	0.3112	-0.0727	0.0686	0.0009	0.5497	-0.3531	0.2314
1.9	0.1357	-0.1284	0.2870	-0.2334	0.2636	-0.0718	0.0469	0.0053	0.5779	-0.3398	0.2111
2.0	0.1352	-0.1135	0.2896	-0.2102	0.2216	-0.0682	0.0275	0.0084	0.5992	-0.3235	0.1910
2.1	0.1321	-0.0993	0.2878	-0.1878	0.1850	-0.0622	0.0109	0.0103	0.6127	-0.3048	0.1715
2.2	0.1268	-0.0860	0.2820	-0.1664	0.1534	-0.0543	-0.0029	0.0113	0.6180	-0.2845	0.1530
2.3	0.1198	-0.0738	0.2724	-0.1464	0.1265	-0.0453	-0.0138	0.0116	0.6150	-0.2631	0.1357
2.4	0.1115	-0.0627	0.2597	-0.1278	0.1037	-0.0356	-0.0219	0.0114	0.6041	-0.2413	0.1197
2.5	0.1023	-0.0529	0.2444	-0.1109	0.0846	-0.0259	-0.0274	0.0107	0.5860	-0.2196	0.1053
2.6	0.0927	-0.0443	0.2273	-0.0956	0.0687	-0.0167	-0.0308	0.0098	0.5617	-0.1985	0.0923
2.7	0.0829	-0.0369	0.2089	-0.0820	0.0557	-0.0083	-0.0322	0.0088	0.5323	-0.1782	0.0808
2.8	0.0734	-0.0305	0.1900	-0.0700	0.0450	-0.0010	-0.0321	0.0077	0.4991	-0.1592	0.0706
2.9	0.0642	-0.0251	0.1711	-0.0595	0.0364	0.0049	-0.0309	0.0066	0.4634	-0.1416	0.0617
3.0	0.0557	-0.0205	0.1527	-0.0504	0.0294	0.0096	-0.0288	0.0056	0.4263	-0.1255	0.0541
3.1	0.0479	-0.0168	0.1351	-0.0427	0.0239	0.0129	-0.0262	0.0046	0.3889	-0.1110	0.0474
3.2	0.0409	-0.0137	0.1187	-0.0361	0.0194	0.0150	-0.0234	0.0038	0.3523	-0.0980	0.0417
3.3	0.0346	-0.0111	0.1037	-0.0305	0.0159	0.0160	-0.0204	0.0031	0.3171	-0.0865	0.0368
3.4	0.0292	-0.0091	0.0901	-0.0258	0.0131	0.0162	-0.0175	0.0025	0.2840	-0.0764	0.0326
3.5	0.0245	-0.0074	0.0780	-0.0219	0.0108	0.0158	-0.0148	0.0020	0.2533	-0.0676	0.0291
3.6	0.0205	-0.0060	0.0673	-0.0187	0.0091	0.0148	-0.0123	0.0015	0.2254	-0.0600	0.0260
3.7	0.0172	-0.0050	0.0580	-0.0160	0.0076	0.0136	-0.0102	0.0012	0.2001	-0.0534	0.0234
3.8	0.0143	-0.0041	0.0499	-0.0137	0.0065	0.0121	-0.0083	0.0009	0.1777	-0.0477	0.0211
3.9	0.0120	-0.0034	0.0431	-0.0119	0.0056	0.0106	-0.0066	0.0007	0.1579	-0.0428	0.0192
4.0	0.0100	-0.0029	0.0372	-0.0103	0.0049	0.0091	-0.0053	0.0005	0.1405	-0.0385	0.0175
4.1	0.0084	-0.0024	0.0322	-0.0090	0.0043	0.0077	-0.0042	0.0004	0.1254	-0.0349	0.0160
4.2	0.0071	-0.0020	0.0281	-0.0080	0.0038	0.0064	-0.0033	0.0003	0.1123	-0.0317	0.0147
4.3	0.0060	-0.0018	0.0245	-0.0071	0.0033	0.0053	-0.0026	0.0002	0.1009	-0.0290	0.0135
4.4	0.0052	-0.0015	0.0215	-0.0063	0.0030	0.0043	-0.0020	0.0002	0.0911	-0.0266	0.0125
4.5	0.0044	-0.0013	0.0190	-0.0056	0.0027	0.0035	-0.0016	0.0001	0.0827	-0.0245	0.0116
4.6	0.0038	-0.0011	0.0169	-0.0051	0.0024	0.0028	-0.0012	0.0001	0.0753	-0.0226	0.0108
4.7	0.0033	-0.0010	0.0151	-0.0046	0.0022	0.0023	-0.0009	0.0001	0.0690	-0.0210	0.0101
4.8	0.0029	-0.0009	0.0135	-0.0042	0.0020	0.0018	-0.0007	0.0001	0.0634	-0.0195	0.0094
4.9	0.0026	-0.0008	0.0122	-0.0038	0.0018	0.0015	-0.0006	0.0001	0.0585	-0.0182	0.0088
5.0	0.0023	-0.0007	0.0110	-0.0035	0.0017	0.0012	-0.0004	0.0000	0.0541	-0.0170	0.0083

TABLE VIII. Influence functions: numerical values ($n = 0$).

x	$\langle\delta_0\delta_1\rangle$	$\langle\delta_0\epsilon_1\rangle$	$\langle\delta_0\gamma_0\rangle$	$\langle\delta_1\epsilon_0\rangle$	$\langle\epsilon_0\epsilon_1\rangle$	$\langle\epsilon_0\gamma_0\rangle$	$\langle\delta_0^3\rangle$	$\langle\delta_0^2\epsilon_0\rangle$	$\langle\delta_0\epsilon_0^2\rangle$	$\langle\epsilon_0^3\rangle$
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	-0.0011	0.0115	-0.0060	0.0183	0.0034	-0.0565	-0.0005	-0.0108	-0.0061	0.0100
0.2	-0.0022	0.0230	-0.0107	0.0366	0.0072	-0.1113	-0.0009	-0.0216	-0.0126	0.0196
0.3	-0.0035	0.0346	-0.0127	0.0547	0.0117	-0.1630	-0.0013	-0.0324	-0.0198	0.0285
0.4	-0.0053	0.0462	-0.0112	0.0726	0.0170	-0.2100	-0.0015	-0.0432	-0.0281	0.0365
0.5	-0.0083	0.0576	-0.0053	0.0898	0.0234	-0.2512	-0.0011	-0.0534	-0.0375	0.0432
0.6	-0.0131	0.0685	0.0051	0.1058	0.0308	-0.2857	0.0000	-0.0627	-0.0480	0.0486
0.7	-0.0206	0.0783	0.0200	0.1200	0.0392	-0.3127	0.0023	-0.0704	-0.0592	0.0525
0.8	-0.0318	0.0866	0.0389	0.1316	0.0484	-0.3321	0.0061	-0.0757	-0.0707	0.0548
0.9	-0.0472	0.0928	0.0610	0.1399	0.0580	-0.3439	0.0115	-0.0780	-0.0822	0.0557
1.0	-0.0671	0.0964	0.0853	0.1442	0.0677	-0.3482	0.0185	-0.0768	-0.0929	0.0552
1.1	-0.0911	0.0970	0.1105	0.1440	0.0771	-0.3457	0.0269	-0.0717	-0.1022	0.0535
1.2	-0.1186	0.0943	0.1352	0.1391	0.0857	-0.3371	0.0361	-0.0628	-0.1097	0.0508
1.3	-0.1481	0.0882	0.1584	0.1294	0.0932	-0.3233	0.0456	-0.0503	-0.1150	0.0472
1.4	-0.1778	0.0790	0.1788	0.1153	0.0994	-0.3052	0.0544	-0.0350	-0.1177	0.0431
1.5	-0.2057	0.0669	0.1956	0.0972	0.1038	-0.2840	0.0618	-0.0176	-0.1178	0.0386
1.6	-0.2297	0.0525	0.2081	0.0760	0.1064	-0.2605	0.0670	0.0008	-0.1152	0.0339
1.7	-0.2477	0.0366	0.2160	0.0528	0.1072	-0.2357	0.0694	0.0191	-0.1103	0.0293
1.8	-0.2582	0.0198	0.2191	0.0286	0.1061	-0.2106	0.0687	0.0361	-0.1032	0.0248
1.9	-0.2601	0.0029	0.2177	0.0045	0.1033	-0.1858	0.0648	0.0510	-0.0945	0.0207
2.0	-0.2531	-0.0132	0.2121	-0.0185	0.0989	-0.1619	0.0580	0.0629	-0.0846	0.0168
2.1	-0.2375	-0.0280	0.2029	-0.0393	0.0934	-0.1394	0.0489	0.0714	-0.0741	0.0135
2.2	-0.2142	-0.0408	0.1907	-0.0573	0.0868	-0.1186	0.0380	0.0763	-0.0633	0.0105
2.3	-0.1846	-0.0514	0.1764	-0.0720	0.0796	-0.0998	0.0263	0.0777	-0.0528	0.0081
2.4	-0.1506	-0.0595	0.1606	-0.0831	0.0720	-0.0831	0.0146	0.0758	-0.0428	0.0060
2.5	-0.1141	-0.0650	0.1441	-0.0906	0.0643	-0.0684	0.0037	0.0713	-0.0337	0.0043
2.6	-0.0772	-0.0680	0.1274	-0.0946	0.0566	-0.0558	-0.0059	0.0647	-0.0256	0.0030
2.7	-0.0416	-0.0688	0.1111	-0.0955	0.0493	-0.0450	-0.0137	0.0565	-0.0187	0.0020
2.8	-0.0090	-0.0677	0.0956	-0.0937	0.0424	-0.0359	-0.0194	0.0477	-0.0129	0.0013
2.9	0.0195	-0.0649	0.0812	-0.0896	0.0360	-0.0284	-0.0230	0.0387	-0.0082	0.0007
3.0	0.0431	-0.0609	0.0681	-0.0838	0.0303	-0.0223	-0.0246	0.0300	-0.0046	0.0003
3.1	0.0614	-0.0560	0.0565	-0.0769	0.0252	-0.0173	-0.0245	0.0220	-0.0019	0.0001
3.2	0.0745	-0.0506	0.0463	-0.0693	0.0208	-0.0133	-0.0230	0.0150	0.0000	-0.0001
3.3	0.0827	-0.0450	0.0376	-0.0614	0.0170	-0.0102	-0.0205	0.0091	0.0013	-0.0002
3.4	0.0865	-0.0394	0.0302	-0.0536	0.0137	-0.0077	-0.0174	0.0043	0.0021	-0.0002
3.5	0.0866	-0.0340	0.0240	-0.0461	0.0110	-0.0058	-0.0140	0.0007	0.0025	-0.0002
3.6	0.0837	-0.0290	0.0189	-0.0392	0.0088	-0.0044	-0.0107	-0.0019	0.0026	-0.0002
3.7	0.0787	-0.0244	0.0148	-0.0329	0.0070	-0.0033	-0.0075	-0.0038	0.0025	-0.0002
3.8	0.0722	-0.0204	0.0115	-0.0273	0.0055	-0.0024	-0.0048	-0.0047	0.0023	-0.0002
3.9	0.0648	-0.0168	0.0089	-0.0224	0.0043	-0.0018	-0.0025	-0.0051	0.0020	-0.0001
4.0	0.0572	-0.0138	0.0068	-0.0183	0.0034	-0.0013	-0.0007	-0.0051	0.0017	-0.0001
4.1	0.0496	-0.0112	0.0052	-0.0148	0.0026	-0.0010	0.0006	-0.0049	0.0015	-0.0001
4.2	0.0425	-0.0090	0.0040	-0.0119	0.0021	-0.0008	0.0015	-0.0044	0.0012	-0.0001
4.3	0.0359	-0.0073	0.0031	-0.0095	0.0016	-0.0006	0.0020	-0.0039	0.0010	-0.0001
4.4	0.0300	-0.0058	0.0023	-0.0076	0.0013	-0.0004	0.0023	-0.0033	0.0008	-0.0000
4.5	0.0249	-0.0047	0.0018	-0.0060	0.0010	-0.0003	0.0023	-0.0028	0.0006	-0.0000
4.6	0.0205	-0.0037	0.0014	-0.0048	0.0008	-0.0003	0.0022	-0.0023	0.0004	-0.0000
4.7	0.0168	-0.0030	0.0011	-0.0038	0.0006	-0.0002	0.0019	-0.0019	0.0003	-0.0000
4.8	0.0138	-0.0024	0.0009	-0.0030	0.0005	-0.0002	0.0015	-0.0015	0.0002	-0.0000
4.9	0.0114	-0.0019	0.0007	-0.0024	0.0004	-0.0001	0.0010	-0.0013	0.0002	-0.0000
5.0	0.0097	-0.0015	0.0006	-0.0019	0.0004	-0.0001	0.0003	-0.0012	0.0001	-0.0000

they are alternatingly odd and even. Their over-all magnitude is considerably smaller than that of T_0, and it does not increase as the rank increases. The inner range contracts as the rank increases, in accordance with Table III. Thus the last term that is written in Eq. (26a) will be negligibly small compared to T_0 if (d_0/k), respectively, (e_0/k), is a reasonably small number. A corresponding statement applies if $n = 1$ [Fig. 6(b)].

Case (iii) belongs into a somewhat different category because k^2 replaces k.[10] Thus the influence functions U_n, V_n, W_n, $\cdots$ are either all even (n even) or all odd (n odd). They are plotted for $n = 0$ and $n = 1$ in Fig. 7. The sequences T_0, U_0, V_0, $\cdots$ and

[10] Case (iii) is not strictly a case of a double transition point. It is a case with two single transition points if $g_0 > 0$, a case with no transition point if $g_0 < 0$ [see, e.g., F. W. J. Oliver, J. Soc. Indust. Appl. Math. 7, 306 (1959)].

$T_1, U_1, V_1, \cdots$ converge; the former to a finite function, the latter to zero. This observation is in agreement with the following: the limits Z_n of the two sequences must (if they exist) obey the differential equation

$$Z_n x^2 - Z_n'' = Z_n,$$

which has the general solution

$$Z_n = \left[C_0 + C_1 \int e^{x^2}\, dx\right] e^{-x^2/2}.$$

In view of the given conditions of symmetry and convergency then

$$Z_0 = C_0 e^{-x^2/2}; \qquad Z_1 \equiv 0.$$

In the scale of Fig. 7, already V_0 is practically indistinguishable from (W_0 and) the function $\sqrt{2}$ exp $(-x^2/2)$, and W_1 already is close to zero.

From the convergency of the sequences $T_n, U_n, \cdots$ (for which no analytical proof is given in this paper) it follows that the series Eq. (26a, iii) converges in

TABLE IX. Influence functions: numerical values ($n = 1$).

x	δ_0	ϵ_0	δ_1	ϵ_1	γ_0	$\langle\delta_0^2\rangle$	$\langle\delta_0\epsilon_0\rangle$	$\langle\epsilon_0^2\rangle$
0.0	-0.3333	0.5000	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	-0.3333	0.4970	-0.0217	0.0121	0.0254	0.0063	0.0045	-0.0188
0.2	-0.3332	0.4882	-0.0434	0.0237	0.0502	0.0125	0.0096	-0.0370
0.3	-0.3329	0.4737	-0.0650	0.0341	0.0739	0.0187	0.0158	-0.0540
0.4	-0.3320	0.4541	-0.0864	0.0429	0.0959	0.0247	0.0235	-0.0694
0.5	-0.3302	0.4301	-0.1073	0.0497	0.1157	0.0304	0.0329	-0.0827
0.6	-0.3270	0.4023	-0.1274	0.0542	0.1330	0.0356	0.0439	-0.0936
0.7	-0.3221	0.3716	-0.1463	0.0562	0.1476	0.0398	0.0565	-0.1019
0.8	-0.3152	0.3388	-0.1633	0.0559	0.1593	0.0427	0.0702	-0.1076
0.9	-0.3060	0.3049	-0.1779	0.0532	0.1679	0.0441	0.0844	-0.1106
1.0	-0.2943	0.2706	-0.1894	0.0485	0.1736	0.0436	0.0986	-0.1110
1.1	-0.2801	0.2368	-0.1975	0.0420	0.1764	0.0412	0.1120	-0.1092
1.2	-0.2637	0.2041	-0.2017	0.0342	0.1766	0.0366	0.1240	-0.1054
1.3	-0.2451	0.1732	-0.2017	0.0254	0.1744	0.0301	0.1339	-0.0999
1.4	-0.2249	0.1444	-0.1976	0.0160	0.1701	0.0219	0.1414	-0.0931
1.5	-0.2034	0.1182	-0.1894	0.0066	0.1641	0.0123	0.1459	-0.0853
1.6	-0.1811	0.0947	-0.1775	-0.0026	0.1567	0.0019	0.1475	-0.0770
1.7	-0.1587	0.0739	-0.1624	-0.0112	0.1482	-0.0087	0.1461	-0.0685
1.8	-0.1365	0.0560	-0.1448	-0.0191	0.1390	-0.0191	0.1419	-0.0600
1.9	-0.1151	0.0409	-0.1253	-0.0259	0.1293	-0.0287	0.1353	-0.0517
2.0	-0.0950	0.0282	-0.1047	-0.0316	0.1195	-0.0370	0.1265	-0.0440
2.1	-0.0764	0.0179	-0.0838	-0.0361	0.1098	-0.0437	0.1163	-0.0369
2.2	-0.0595	0.0097	-0.0633	-0.0394	0.1003	-0.0487	0.1049	-0.0304
2.3	-0.0446	0.0034	-0.0437	-0.0417	0.0912	-0.0517	0.0930	-0.0247
2.4	-0.0317	-0.0014	-0.0256	-0.0430	0.0826	-0.0529	0.0810	-0.0198
2.5	-0.0207	-0.0049	-0.0094	-0.0434	0.0746	-0.0523	0.0693	-0.0156
2.6	-0.0117	-0.0072	0.0047	-0.0430	0.0672	-0.0503	0.0582	-0.0121
2.7	-0.0044	-0.0088	0.0166	-0.0420	0.0604	-0.0470	0.0480	-0.0092
2.8	0.0012	-0.0096	0.0263	-0.0405	0.0542	-0.0429	0.0388	-0.0069
2.9	0.0055	-0.0099	0.0338	-0.0387	0.0387	-0.0382	0.0307	-0.0050
3.0	0.0086	-0.0099	0.0394	-0.0367	0.0438	-0.0332	0.0238	-0.0036
3.1	0.0106	-0.0095	0.0432	-0.0345	0.0394	-0.0282	0.0180	-0.0024
3.2	0.0119	-0.0091	0.0454	-0.0322	0.0355	-0.0234	0.0132	-0.0016
3.3	0.0125	-0.0085	0.0464	-0.0300	0.0370	-0.0189	0.0094	-0.0010
3.4	0.0126	-0.0078	0.0463	-0.0278	0.0289	-0.0149	0.0064	-0.0006
3.5	0.0123	-0.0072	0.0454	-0.0257	0.0262	-0.0114	0.0041	-0.0003
3.6	0.0119	-0.0065	0.0439	-0.0237	0.0239	-0.0084	0.0024	-0.0001
3.7	0.0112	-0.0059	0.0420	-0.0218	0.0217	-0.0061	0.0011	0.0001
3.8	0.0105	-0.0054	0.0398	-0.0201	0.0199	-0.0040	0.0003	0.0001
3.9	0.0097	-0.0048	0.0375	-0.0186	0.0182	-0.0025	-0.0003	0.0002
4.0	0.0089	-0.0044	0.0352	-0.0171	0.0168	-0.0014	-0.0007	0.0002
4.1	0.0081	-0.0039	0.0328	-0.0158	0.0154	-0.0005	-0.0009	0.0002
4.2	0.0074	-0.0036	0.0306	-0.0146	0.0143	0.0001	-0.0010	0.0002
4.3	0.0068	-0.0032	0.0285	-0.0136	0.0132	0.0005	-0.0010	0.0002
4.4	0.0061	-0.0029	0.0265	-0.0126	0.0123	0.0007	-0.0009	0.0001
4.5	0.0056	-0.0027	0.0246	-0.0117	0.0114	0.0008	-0.0009	0.0001
4.6	0.0051	-0.0024	0.0229	-0.0109	0.0107	0.0009	-0.0008	0.0001
4.7	0.0046	-0.0022	0.0213	-0.0102	0.0100	0.0008	-0.0007	0.0001
4.8	0.0042	-0.0020	0.0199	-0.0095	0.0093	0.0008	-0.0006	0.0001
4.9	0.0038	-0.0018	0.0185	-0.0089	0.0087	0.0008	-0.0005	0.0001
5.0	0.0035	-0.0017	0.0173	-0.0083	0.0082	0.0007	-0.0005	0.0001

the classical sense, supposing of course that $|g_0/k^2| < 1$; that is, contrary to what one might have expected, this series is not an asymptotic series. (It is readily shown that, if this is true for $n = 0$ and $n = 1$, then it is true for all n.)

Whether the other two series [Eq. (26a)] are also convergent in the classical sense is not investigated in this paper. From the point of view of practical applications, the answer to this question is of little concern, as both series are useful only if the first few terms suffice. We did confirm, however, in a number of cases that the first few terms of such series do indeed closely approximate the correct solution if the last term considered is sufficiently small. For this purpose, we obtained numerical solutions by integrating Eq. (5) with boundary conditions obtained from outer range relations, e.g., Eq. (18). The numerical solutions agreed to the expected degree of accuracy with the terms written out in Eq. (23).

So far we have discussed three cases where $\nu = 0$. The over-all magnitude of the influence functions in Fig. 6 increases as ν increases. This corresponds to a less rapid convergency in the outer range; we have

$$\begin{aligned} \delta_\nu &\sim (n-2)(n-3)x^{\nu-3}T_n, \\ \epsilon_\nu &\sim (n-2)x^{\nu-3}T_n, \\ \gamma_\nu &\sim x^{\nu-2}T_n, \end{aligned} \tag{27}$$

and the corresponding exponent of the higher rank influence functions is the sum of the respective exponents of the factors involved (compare Table III).

An incorrect conclusion might be drawn from Eq. (27), namely, that no solution of Eq. (5) could converge, as $|x| \to \infty$, to the limit solution T_n if any one of the higher influence coefficients (g_2, d_3, e_3, g_3, $\cdots$) would be different from zero. The proper condition for such convergency is Eq. (18a).

VII. REMARK ON APPROXIMATE PROCEDURES

Differential equations with large parameters and transition points are not infrequently treated by approximate procedures, sometimes referred to as WKB methods. With exact influence functions available, it becomes of interest to compare with these the corresponding results of such approximate procedures.

We report here results of such comparisons for two approximate procedures, described briefly as follows:

Procedure (1)[5]: the standard transcendental transformation

$$F(\xi) = w\tilde{f}(\tilde{x}); \qquad \tilde{x} = kz\xi$$

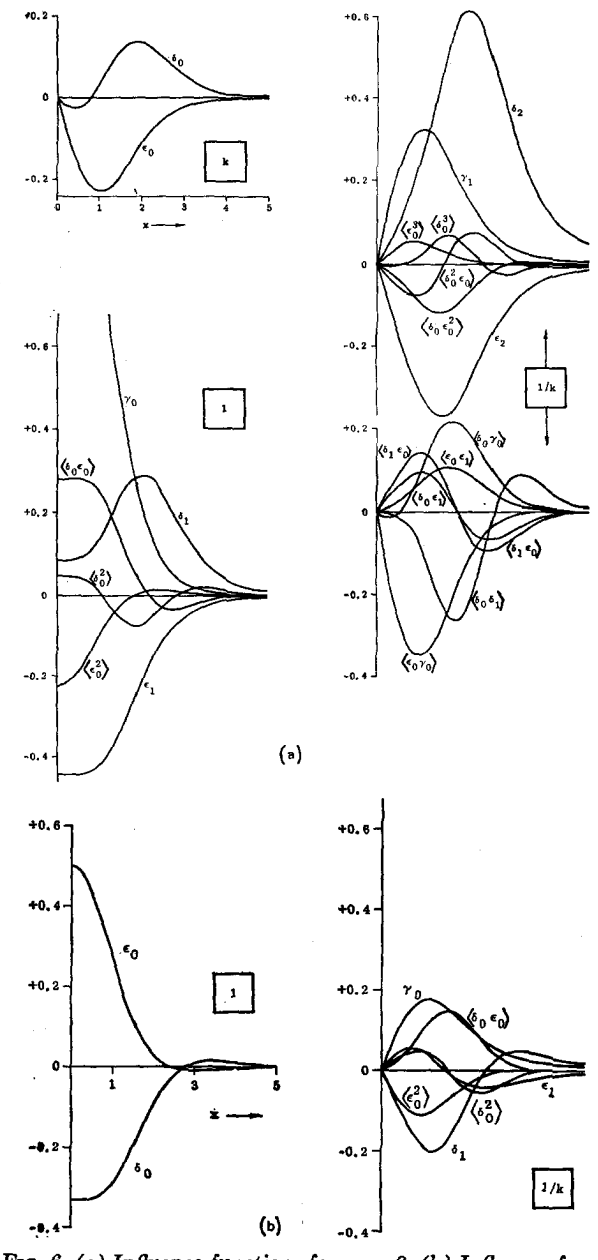

FIG. 6. (a) Influence functions for $n = 0$. (b) Influence functions for $n = 1$.

is applied to Eq. (2), with the functions w and z determined such that Eq. (2) takes approximately the form

$$\tilde{f}\tilde{x}^2 - \tilde{f}_{\tilde{x}\tilde{x}} = Ak^2 + Bk\tilde{x} + \tilde{x}^2 p(\tilde{x}),$$

where A and B are constants, and where a term with f/k^2 is neglected. The solution of this equation is then written as[11]

$$\tilde{f} \approx Ak^2 T_0(\tilde{x}) + BkT_1(\tilde{x}) + p(\tilde{x}).$$

[11] The additional approximative assumption that is here made was discussed above in connection with Eqs. (16), (16a).

TABLE X. Functions U_n, V_n, W_n, and first derivatives.

x	U_0	U_0'	U_1	U_1'	V_0	V_0'	V_1	V_1'	W_0	W_0'	W_1	W_1'
0.0	1.3906	-0.0000	0.0000	0.2551	1.4088	0.0000	0.0000	0.0951	1.4127	0.0000	0.0000	0.0334
0.1	1.3840	-0.1305	0.0254	0.2521	1.4019	-0.1383	0.0095	0.0938	1.4057	-0.1402	0.0033	0.0329
0.2	1.3646	-0.2572	0.0502	0.2433	1.3813	-0.2726	0.0187	0.0901	1.3848	-0.2761	0.0066	0.0315
0.3	1.3328	-0.3766	0.0739	0.2290	1.3477	-0.3989	0.0274	0.0840	1.3507	-0.4040	0.0096	0.0293
0.4	1.2896	-0.4855	0.0959	0.2098	1.3019	-0.5138	0.0345	0.0759	1.3044	-0.5202	0.0124	0.0263
0.5	1.2362	-0.5812	0.1157	0.1865	1.2454	-0.6143	0.0425	0.0661	1.2472	-0.6217	0.0148	0.0226
0.6	1.1739	-0.6616	0.1330	0.1600	1.1796	-0.6981	0.0486	0.0551	1.1806	-0.7062	0.0169	0.0186
0.7	1.1044	-0.7252	0.1476	0.1313	1.1064	-0.7637	0.0535	0.0432	1.1065	-0.7722	0.0185	0.0142
0.8	1.0294	-0.7713	0.1593	0.1014	1.0275	-0.8104	0.0572	0.0310	1.0268	-0.8189	0.0197	0.0097
0.9	0.9507	-0.7999	0.1679	0.0714	0.9449	-0.8382	0.0597	0.0188	0.9434	-0.8464	0.0205	0.0053
1.0	0.8700	-0.8115	0.1736	0.0422	0.8605	-0.8478	0.0610	0.0072	0.8582	-0.8555	0.0208	0.0012
1.1	0.7890	-0.8074	0.1764	0.0147	0.7759	-0.8407	0.0612	-0.0036	0.7729	-0.8474	0.0207	-0.0027
1.2	0.7090	-0.7892	0.1766	-0.0106	0.6928	-0.8185	0.0603	-0.0132	0.6892	-0.8243	0.0203	-0.0060
1.3	0.6315	-0.7590	0.1744	-0.0329	0.6126	-0.7837	0.0586	-0.0214	0.6084	-0.7884	0.0195	-0.0089
1.4	0.5576	-0.7188	0.1701	-0.0521	0.5364	-0.7386	0.0561	-0.0282	0.5318	-0.7421	0.0185	-0.0111
1.5	0.4880	-0.6711	0.1641	-0.0678	0.4652	-0.6858	0.0530	-0.0335	0.4603	-0.6880	0.0173	-0.0128
1.6	0.4235	-0.6180	0.1567	-0.0800	0.3994	-0.6278	0.0495	-0.0372	0.3944	-0.6288	0.0160	-0.0140
1.7	0.3645	-0.5618	0.1482	-0.0889	0.3397	-0.5668	0.0456	-0.0395	0.3346	-0.5668	0.0146	-0.0146
1.8	0.3112	-0.5044	0.1390	-0.0947	0.2861	-0.5050	0.0416	-0.0405	0.2811	-0.5040	0.0131	-0.0147
1.9	0.2636	-0.4475	0.1293	-0.0976	0.2387	-0.4442	0.0376	-0.0404	0.2338	-0.4424	0.0116	-0.0144
2.0	0.2216	-0.3924	0.1195	-0.0981	0.1972	-0.3858	0.0336	-0.0393	0.1925	-0.3835	0.0102	-0.0138
2.1	0.1850	-0.3404	0.1098	-0.0965	0.1614	-0.3311	0.2097	-0.0375	0.1569	-0.3282	0.0089	-0.0130
2.2	0.1534	-0.2921	0.1003	-0.0932	0.1308	-0.2807	0.0261	-0.0351	0.1267	-0.2775	0.0076	-0.0120
2.3	0.1265	-0.2481	0.0912	-0.0887	0.1051	-0.2353	0.0227	-0.0324	0.1013	-0.2319	0.0065	-0.0108
2.4	0.1037	-0.2087	0.0826	-0.0832	0.0836	-0.1949	0.0196	-0.0294	0.0802	-0.1914	0.0055	-0.0097
2.5	0.0846	-0.1740	0.0746	-0.0772	0.0659	-0.1597	0.0168	-0.0263	0.0628	-0.1562	0.0046	-0.0085
2.6	0.0687	-0.1438	0.0672	-0.0709	0.0515	-0.1294	0.0144	-0.0233	0.0487	-0.1260	0.0038	-0.0074
2.7	0.0557	-0.1178	0.0604	-0.0645	0.0399	-0.1037	0.0122	-0.0204	0.0374	-0.1005	0.0031	-0.0063
2.8	0.0450	-0.0958	0.0543	-0.0583	0.0306	-0.0822	0.0103	-0.0176	0.0285	-0.0793	0.0025	-0.0053
2.9	0.0364	-0.0774	0.0487	-0.0523	0.0233	-0.0645	0.0086	-0.0151	0.0215	-0.0618	0.0020	-0.0044
3.0	0.0294	-0.0622	0.0438	-0.0467	0.0176	-0.0501	0.0072	-0.0129	0.0160	-0.0477	0.0016	-0.0037
3.1	0.0239	-0.0497	0.0394	-0.0415	0.0132	-0.0385	0.0061	-0.0109	0.0118	-0.0364	0.0013	-0.0030
3.2	0.0194	-0.0396	0.0355	-0.0368	0.0098	-0.0293	0.0051	-0.0091	0.0087	-0.0275	0.0010	-0.0024
3.3	0.0159	-0.0315	0.0320	-0.0325	0.0073	-0.0221	0.0042	-0.0076	0.0063	-0.0205	0.0008	-0.0019
3.4	0.0131	-0.0250	0.0289	-0.0287	0.0054	-0.0166	0.0035	-0.0063	0.0045	-0.0151	0.0006	-0.0015
3.5	0.0108	-0.0198	0.0262	-0.0254	0.0039	-0.0123	0.0030	-0.0052	0.0032	-0.0111	0.0005	-0.0012
3.6	0.0091	-0.0158	0.0239	-0.0224	0.0029	-0.0090	0.0025	-0.0043	0.0023	-0.0080	0.0004	-0.0009
3.7	0.0076	-0.0126	0.0217	-0.0198	0.0021	-0.0066	0.0021	-0.0035	0.0016	-0.0057	0.0003	-0.0007
3.8	0.0065	-0.0101	0.0199	-0.0175	0.0015	-0.0048	0.0018	-0.0029	0.0011	-0.0041	0.0002	-0.0006
3.9	0.0056	-0.0082	0.0182	-0.0156	0.0011	-0.0035	0.0015	-0.0024	0.0007	-0.0028	0.0002	-0.0004
4.0	0.0049	-0.0067	0.0168	-0.0139	0.0008	-0.0025	0.0013	-0.0020	0.0005	-0.0020	0.0001	-0.0003
4.1	0.0043	-0.0055	0.0154	-0.0124	0.0006	-0.0018	0.0011	-0.0017	0.0003	-0.0014	0.0001	-0.0003
4.2	0.0038	-0.0046	0.0143	-0.0111	0.0005	-0.0013	0.0010	-0.0014	0.0002	-0.0009	0.0001	-0.0002
4.3	0.0033	-0.0038	0.0132	-0.0100	0.0003	-0.0009	0.0008	-0.0011	0.0002	-0.0006	0.0001	-0.0002
4.4	0.0030	-0.0032	0.0123	-0.0090	0.0003	-0.0007	0.0007	-0.0010	0.0001	-0.0004	0.0001	-0.0001
4.5	0.0027	-0.0028	0.0114	-0.0081	0.0002	-0.0005	0.0006	-0.0008	0.0001	-0.0003	0.0000	-0.0001
4.6	0.0024	-0.0024	0.0107	-0.0073	0.0002	-0.0004	0.0006	-0.0007	0.0000	-0.0002	0.0000	-0.0001
4.7	0.0022	-0.0021	0.0100	-0.0067	0.0001	-0.0003	0.0005	-0.0006	0.0000	-0.0001	0.0000	-0.0001
4.8	0.0020	-0.0018	0.0093	-0.0061	0.0001	-0.0002	0.0004	-0.0005	0.0000	-0.0001	0.0000	-0.0000
4.9	0.0018	-0.0016	0.0087	-0.0056	0.0001	-0.0002	0.0004	-0.0004	0.0000	-0.0001	0.0000	-0.0000
5.0	0.0017	-0.0014	0.0082	-0.0051	0.0001	-0.0001	0.0003	-0.0004	0.0000	-0.0000	0.0000	-0.0000

Procedure (2): write

$$F = Ak^2T_0(k\xi) + BkT_1(k\xi) + F^*(\xi),$$

determine the constants A and B such that Eq. (2) reduces to (exactly)

$$F^*\xi^2 - \frac{1}{k^4}\{\cdots F^{*\prime\prime}, F^{*\prime}, F^*, \cdots\} = \xi^2 l^*(\xi),$$

and approximate F^* by

$$F^* \approx l^{(*)}(\xi).$$

Each one of these two procedures was used to determine separately the four influence functions $\delta_{0,0}$, $\delta_{0,1}$, $\epsilon_{0,0}$ and $\epsilon_{0,1}$. The results were

	Procedure (1)	Procedure (2)	Correct
$6\delta_{0,0}$	xT_0-3T_1	$6(xT_0-T_1)$	$xT_0-3T_1-x^2T_0'$
$6\delta_{0,1}$	xT_1-2	$6(xT_1-1)$	$xT_1-2-x^2T_1'$
$2\epsilon_{0,0}$	T_1-xT_0	$2[(T_0'+x)/x^2-T_1]$	T_1-xT_0
$2\epsilon_{0,1}$	$1-xT_1$	$2[(T_1'-c_1)/x^2+c_1T_0]$	$1-xT_1$

A graphical comparison is shown in Fig. 8. Procedure (1) is incorrect in both the inner and the outer ranges in its prediction of $\delta_{0,0}$ and $\delta_{0,1}$, except at the transition point itself.[12] Procedure (1) gives

[12] Presumably this explains the discrepancy around $\phi = -20°$ between accurate solution (Ref. 1) and approximate solution in Fig. 4 of Ref. 5.

the correct result in the two cases, $\epsilon_{0,0}$ and $\epsilon_{0,1}$, where the "transcendental transformation" is not actually transcendental. Procedure (2) predicts all four influence functions incorrectly in the inner range. These discrepancies are worth noting in view of the fact that, on first sight, both procedures would appear plausible enough.

VIII. REMARK ON HOMOGENEOUS SOLUTION

So far we have discussed influence functions which were derived from the convergent limit solution by a formal process and which, together with this convergent limit solution, describe a nonhomogeneous solution of the complete Eq. (5). We remark now that the same formal process can be applied to the homogeneous problem, and can be used to derive influence functions from and for any one homogeneous limit solution h. Indeed, if on the left of Eq. (24) we replace T_n by h, we obtain a similar but slightly simpler relation:

$$M(x^p h) = -p(p-1)x^{p-2}h - 2px^{p-1}h',$$
$$M(x^p h') = -2(p+1)x^{p+1}h - p(p-1)x^{p-2}h'.$$

From this a table similar to Table V can be derived. There is, however, a significant difference between the two types of influence functions. While (the more important ones of) the nonhomogeneous influence functions converge more rapidly than the convergent limit solution as $|x| \to \infty$, the homogeneous influence functions diverge even more rapidly than the homogeneous limit solution.

Two examples: let $h \equiv t_0$; then

$$\gamma_{1,0} = -\tfrac{1}{2}T_0'; \qquad \gamma_{1,h} = -\tfrac{1}{2}t_0'$$

with the suffix h denoting the homogeneous influence function. Formally the two results are identical; however, for the asymptotic behavior as $|x| \to \infty$ we have

$$\gamma_{1,0} \sim \frac{1}{x}T_0; \qquad \gamma_{1,h} \sim -\frac{x}{2}t_0.$$

Similarly: to

$$\delta_{0,0} = \tfrac{1}{6}[xT_0 - x^2T_0' - 3T_1]$$

corresponds

$$\delta_{0,h} = \tfrac{1}{6}[xt_0 - x^2t_0'].$$

In the expression for $\delta_{0,0}$ the terms of order $1/x$ cancel each other, and

$$\delta_{0,0} \sim \frac{6}{x^3}T_0$$

converges with convenient rapidity. On the other hand

$$\delta_{0,h} \sim -\frac{x^3}{6}t_0$$

diverges considerably faster than t_0.

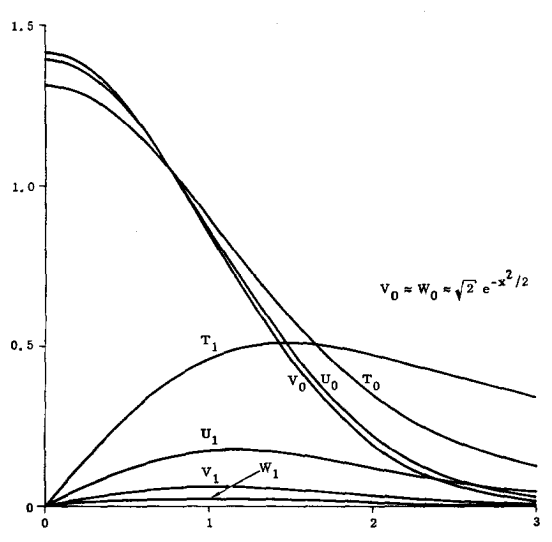

FIG. 7. Sequences $T_0, U_0, V_0, \cdots$ and $T_1, U_1, V_1, \cdots$.

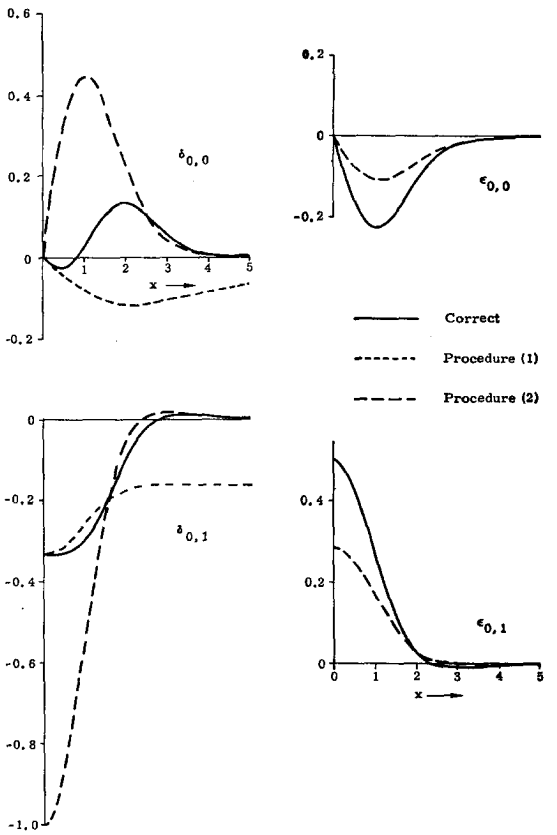

FIG. 8. Results of approximate procedures.

IX. CONCLUSION

A particular solution of a generalized nonhomogeneous Liouville equation has been discussed. The

solution selected is defined by the fact that it converges at infinity. Its asymptotic limit is related to the Lommel function $S_{\mu,\frac{1}{2}}$ but is not itself a Lommel function.

A direct approach was used which avoids a transcendental transformation. The influence which the parameters of the problem exert on the solution has been described by means of a formal development in terms of influence functions. The more important influence functions have been tabulated and have been represented in graphs. They allow a direct insight into the nature of the solution. There is a strong indication that some formal developments are convergent series rather than asymptotic expansions.

In addition to the nonhomogeneous influence functions, both the homogeneous and the convergent nonhomogeneous limit solutions have been tabulated. The homogeneous influence functions have been discussed briefly; their convergency behavior is less convenient than that of the nonhomogeneous influence functions.

The present influence functions are exact in a formal sense. A comparison with results obtained by approximate procedures shows that such approximations require caution.

APPENDIX A. MODIFIED LOMMEL FUNCTIONS

The same considerations that led to Eq. (14) lead to the more general form

$$T_n = c_n t_{\bar{n}} - \frac{t_{n+2}}{(n+1)(n+2)}, \tag{A1}$$

where $\bar{n} = 0$ if n is even, $\bar{n} = 1$ if n is odd; we have $c_2 = c_3 = 1$, and have

$$c_{n+4} = (n+1)(n+2)c_n$$

because of

$$\begin{aligned} t_{m+4} &= (m+3)(m+4)(t_m - x^m), \\ T_{n+4} &= (n+1)(n+2)T_n + x^{n+2}. \end{aligned} \tag{A1a}$$

Watson[13] discusses two types of Lommel functions: $s_{\mu,\nu}(z)$ and $S_{\mu,\nu}(z)$. Setting

$$4u = 2n - 1; \qquad 2z = ix^2; \qquad \nu = \tfrac{1}{4}$$

and, assuming for the time being that x is positive, setting

$$\begin{aligned} x^{\frac{1}{2}} s_{\mu,\nu}(z) &= \left(\frac{i}{2}\right)^{(n-5)/2} s_n^*(x), \\ x^{\frac{1}{2}} S_{\mu,\nu}(z) &= \left(\frac{i}{2}\right)^{(n-5)/2} S_n^*(x), \end{aligned}$$

we find first [Ref. 13, 10.7, Eq. (1)] that the functions s_n^* and S_n^* are solutions of Eq. (7a), and second [Ref. 13, 10.7, Eq. (2)] that

$$s_n^*(x) = -\frac{t_{n+2}}{(n+1)(n+2)}.$$

With this, and taking also into account Eq. (9), we obtain [Ref. 13, 10. 7, Eq. (2)]

$$S_n^*(x) = \left(\frac{n-3}{4}\right)! \left(\frac{n-2}{4}\right)! \, 2^{n-5/2} i^{1-(n/2)} [\cdots] - \frac{t_{n+2}}{(n+1)(n+2)}, \tag{A2}$$

where

$$[\cdots] \equiv \cos\left(\frac{n-1}{4}\pi\right) \frac{2t_0}{(-1/4)!} - i^{\frac{1}{2}} \cos\left(\frac{n}{4}\pi\right) \frac{t_1}{(1/4)!}. \tag{A2a}$$

There is thus a certain correspondence between T_n, Eq. (A1) and S_n^*, Eq. (A2). Indeed, the two functions are identical if we set $n = 2 \pmod 4$ or $n = 3 \pmod 4$; however, S_n^* does not fulfill Eq. (A1) if $n = 0 \pmod 4$ or $n = 1 \pmod 4$.

On the other hand, the function S_n^* converges (for all n) if $x \to +\infty$ [it has the asymptotic development Eq. (15), see Ref. 13, 10. 75 Eq. (1)]. Any solution of Eq. (7a) which converges for both $x \to +\infty$ and $x \to -\infty$ (that is, T_n) must hence have the form

$$T_n = S_n^*(x) + h^{(1)}(x) \cdot \text{const.} \tag{A3}$$

Now Eq. (A2a) can be rewritten as

$$2^{n-5/2} i^{1-(n/2)} [\cdots] = 2^{n-4}[h^{(2)}(x) + (i + (i-1)i^{-n})h^{(1)}(x)]$$

so that, applying Eq. (A3) to Eq. (A2), we can fulfill Eq. (A1) (this proves the existence of T_0 and T_1) and obtain

$$T_n = 2^{n-4}\left(\frac{n-3}{4}\right)! \left(\frac{n-2}{4}\right)! \, [h^{(2)}(x) + (-)^n h^{(1)}(x)] - \frac{t_{n+2}}{(n+1)(n+2)}, \tag{A4}$$

from which the constants c_n can be read directly and can be shown to fulfill Eq. (A1a).

In view of the foregoing, the functions T_n may be designated as "modified Lommel functions." In the present problem, where our attention is confined to real values of the variable x (which twice covers the positive imaginary axis of the variable $z = \frac{1}{2}ix^2$

[13] G. N. Watson, *A Treatise on the Theory of Bessel Functions*. (Macmillan and Company, Inc., New York, 1944), pp. 345–352.

of the Lommel function $S_{\mu,\nu}$), these modified functions are more convenient to use than the Lommel functions proper: they are real valued, convergent for $x \to -\infty$ as well as for $x \to +\infty$, and have desirable symmetry properties.

We note finally the asymptotic development for the functions t_m that is obtained from Eqs. (9), (11), (15), and (A4). Omitting the exponentially convergent term (which is here irrelevant) we find for $x \to +\infty$

$$t_m \sim \frac{2^m}{\sqrt{2\pi x}} \cdot \left(\frac{m-1}{4}\right)! \left(\frac{m}{4}\right)! \times \left(1 + \frac{1}{16x^2} + \cdots\right) e^{x^2/2} - m(m-1)x^{m-4} \times \left[1 + \frac{(m-4)(m-5)}{x^4}(1 + \cdots)\right]. \quad \text{(A5)}$$

This relation is not restricted to integer values m.

APPENDIX B. NUMERICAL METHODS

In this Appendix, the procedure used to prepare Table I is briefly indicated.

The calculations were done on an IBM-7094 computer. Integration of the homogeneous equation

$$M(h) = 0$$

from $x = -7$ to $x = 5$ for two sets of initial conditions: $h(-7) = 1$; $h'(-7) = 0$ and $h(-7) = 0$; $h'(-7) = 1$, led to two functions $h(x)/h'(0)$ which were identical (to eight significant places) in the range $-5.5 \leq x \leq 5$ and which, therefore, were identical to $(\frac{1}{4})!h^{(2)}(x) \equiv (\frac{1}{4})!h^{(1)}(-x)$ in this range.

Integration of the inhomogeneous equation

$$M(\bar{T}_0) = 1$$

from $x = -5$ to $x = 0$, with approximate initial conditions determined by the asymptotic development, Eq. (15), led to a function $\bar{T}_0$ which, in view of symmetry conditions at $x = 0$, could be written as

$$\bar{T}_0 = T_0 + \bar{T}_0'(0)(\tfrac{1}{4})!\, h^{(2)}(x) + \alpha t_0$$

with an unknown constant α. However, as removal of $h^{(2)}$ changed the value of $\bar{T}_0$ at $x = -5$ very little (that is, within the uncertainty range of the starting conditions), $\alpha\, t_0$ had to be very small at the starting point $x = -5$, and in consequence had to be entirely negligible away from this starting point. Consequently

$$T_0 \approx T_0 + \alpha t_0$$

was accepted as the final result.

Continuing the process, $\bar{U}_0$ was determined from

$$M(\bar{U}_0) = T_0,$$

the $h^{(2)}$-component was again eliminated, and so on for V_0 and W_0. The sequence of functions T_1, U_1, $\cdots$ was determined by a corresponding procedure.

On the basis of comparisons of $h(x)$ with NBS tables[14], of $T_0(0)$ and $T_1'(0)$ with their exact values, c_0 and c_1, respectively, of $U_0(0)$ with an independently determined value, and from other indications we concluded that our values for $h(x)$ and T_n were accurate to at least six significant places.[15] The accuracy deteriorated somewhat as the process was continued to determine U_n, V_n, and W_n, but even W_n should be accurate to about the four decimal places given in Table VIII.

Where applicable, our numerical results were compared with the tables of Ref. 5. With the exception of an obvious misprint, all differences could be interpreted as rounding off errors in Ref. 5.

According to Table V, the accuracy of calculating the influence functions decreases as $|x|$ increases. In a few cases, notably $\langle\delta_0^2\rangle$ for $|x| > 3.5$, $\langle\delta_0^3\rangle$ for $|x| > 4$, the last digit of the numerical values given in Table VII might be in error by about one unit.

---

[14] National Bureau of Standards, *Tables of Bessel Functions of Fractional Order* (Columbia University Press, New York, 1949), Vol. II.

[15] An exception is $h^{(2)}$ for $x \geq 4.5$. Here Table I should be accurate to about one unit in the last place given.

Complete High-Energy Behavior for Certain Planar Graphs*

M. MARTINIS†
Imperial College, London, England
(Received 17 July 1964)

The results obtained by Polkinghorne for the set of ladder diagrams is generalized to a certain set of planar graphs. The leading asymptotic term behaves as $s^{-1}(\ln s)^p$, and then the complete set of terms $s^{-1}(\ln s)^m$ is summed over m. The final result allows the writing of an equation for the Regge trajectory function.

I. INTRODUCTION

THE high-energy behavior for planar graphs has been studied by many authors[1-5] giving the sum over leading asymptotic terms only. It is known that the sum over only leading terms need not give, in general, the complete asymptotic behavior of the total Feynman amplitude.

This was extended by keeping certain terms in addition to the leading terms[6] for ladder graphs only, but without the complete sum over all these terms. Later this problem was completely solved for ladder diagrams by Polkinghorne[7] giving the equation for the Regge trajectory.

The aim of this paper is to generalise Polkinghorne's result for a certain number of planar graphs.

In Sec. II a general method and treatment for the high-energy behavior of planar graphs is given. The particular form of the function $f(x)$ which is the coefficient of the energy variable s allows us to write a starting point for all planar graphs with nonintersecting ρ lines.

In Sec. III, the result of Sec. II is applied to the particular case where $\rho = 1$. We give the relation which can be summed over $p \cdot p$ is a number of ρ-lines in a particular graph.

In Sec. IV the summation over p is carried out giving the final formula for the complete high-energy behavior of the $A(s, t)$ in the s variable, and the relation which determined the Regge trajectory.

II. GENERAL METHOD

We consider the parametric representation[1-7] of the scattering amplitude in the variables s and t, ignoring the spin of the particles. In our consideration we include $g\phi^3$ and $G\phi^4$ coupling for three-point and four-point vertices, respectively.

The form of the scattering amplitude $A(s, t)$, which is convenient for investigating the asymptotic behavior in the s channel and the Mellin transform applications, is

$$A(s, t) = K \int dx[\Delta(x)]^{-2} e^{-V(s,t;x)} \qquad (2.1)$$

where the notations are as follows:

K is a constant including the product of certain powers of the coupling g and G;

$$\int dx = \int_0^\infty dx_1 \cdots \int_0^\infty dx_n, \qquad (2.2)$$

where x_i is a Feynman parameter corresponding to an internal line; $\Delta(x)$ is the Feynman numerator usually written as [1,2] $C(x)$ and

$$V(s, t; x) = [f(x)/\Delta(x)]s + \delta(t; x) \qquad (2.3)$$

is the Feynman denominator usually written as[1,2] $D(s, t; x)/C(x)$.

The algebraic structure of the function $f(x)$ will be important for investigating the asymptotic behavior of $A(s, t)$ in the s variable. $\Delta(x)$ is always a positive and nonzero function of x in the region where the integral (2.1) is taken.

For all planar graphs $f(x)$ has the same sign and its structure is given by the relation

$$f(x) = \sum_i \Delta_1(i)\Delta_2(i) \prod_{j \in S_i} x_j. \qquad (2.4)$$

The meaning of each factor in Eq. (2.4) can be seen from Fig. 2.

The sum goes over all possible intermediate states which can be constructed in the s channel for a given graph A. (Fig. 1). By an intermediate state we mean a set of internal lines such that by opening all the lines belonging to it, A (Fig. 2). separates into two parts, where this cannot be done by opening only some of them. The other condition for the existence of an intermediate state is that at least

* The research reported in this document has been sponsored in part by the NSF and the U. S. Air Force Office of Scientific Research, OAR, through the European Office, Aerospace Research, U. S. Air Force.

† On leave of absence from Institute Ruder Bosković, Zagreb, Yugoslavia.

[1] J. C. Polkinghorne, J. Math. Phys. 4, 503 (1963).
[2] P. G. Federbush and M. T. Grisaru, Ann. Phys. (N. Y.) 22, 263 (1963).
[3] I. G. Halliday, Nuovo Cimento 30, 177 (1963).
[4] J. D. Bjorken and T. T. Wu, Phys. Rev. 130, 2566 (1963).
[5] G. Tiktopoulos, Phys. Rev. 131, 480 (1963).
[6] T. L. Trueman and T. Yao, Phys. Rev. 132, 2741 (1963).
[7] J. C. Polkinghorne, J. Math. Phys. 5, 431 (1964).

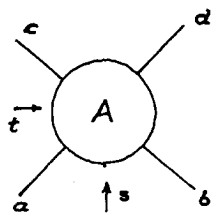

FIG. 1.

one path (Fig. 2) from a to b and one from c to d exists which does not go through the set S_i.

The Mellin transform of (2.1) in the s variable is

$$\mathbf{A}(\alpha, t) = \int_0^\infty ds A(s, t) s^{-\alpha-1}$$

$$= K\Gamma(-\alpha) \int dx (f/\Delta)^\alpha \Delta^{-2} e^{-\delta}. \qquad (2.5)$$

Determination of the asymptotic behavior of (2.1) in the s variable is now equivalent to looking for the singularities of (2.5) in the α variable. One can easily see that they will come from the zero hypersurfaces of the function $f(x)$ in the x variables.

Following Halliday and Tiktopoulos[3,5] we shall call out graph ρ reducible if by putting ρx's equal to zero we make $f = 0$, where this cannot be done for fewer x's. This choice certainly can be made in many ways. In order to get the first leading term, the second leading term and so on, we must choose the shortest ρ line defined by

$$\rho = \min [\rho_1, \rho_2, \cdots] = \min [\rho_i], \qquad (2.6)$$

when one takes the ρ determined by (2.6) one can then make $f = 0$ by putting ρx's equal to zero in p ways. Putting all $p\rho$ lines equal to zero our graph will have a structure as shown in Fig. 3, under the assumption that the ρ lines do not intersect. The λ transformation[5] (Appendix I) of all $p\rho$ lines allows

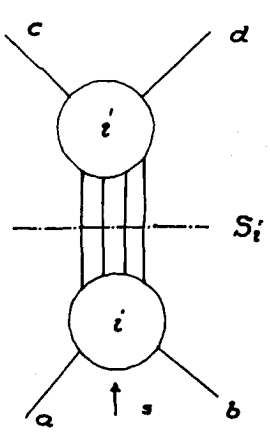

FIG. 2.

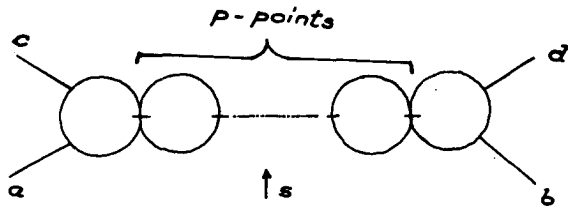

FIG. 3.

the function f/Δ to be written in the form

$$\frac{f}{\Delta} = \left(\prod_{i=1}^{\rho} \lambda_i\right) \frac{\bar{f}}{\Delta}, \qquad (2.7)$$

where $\bar{f}$ is different from zero if we put all λ_i equal to zero. After λ transformation, (2.5) can be rewritten in the form

$$\mathbf{A}_p(\alpha, t)$$

$$= K\Gamma(-\alpha) \int dx\, d\bar{y}^{(j)}\, d\lambda \prod_{i=1}^{p} \lambda_i^{\rho+\alpha-1} J(\alpha), \qquad (2.8)$$

where

$$J(\alpha) = \left(\frac{\bar{f}}{\Delta}\right)^\alpha \Delta^{-2} e^{-\delta}$$

$dx = dx_1 \cdots dx_1$ (no x_i belongs to any of the ρ lines)

$$d\bar{y}^{(j)} = d\bar{y}^{(1)} \cdots d\bar{y}^{(p)} \prod_{i=1}^{p} [\delta(\sum \bar{y}^{(j)} - 1)], \qquad (2.9)$$

where $\bar{y}^{(j)}$ belongs to the (j) ρ line,

$$d\lambda = d\lambda_1 \cdots d\lambda_p.$$

From the structure of (2.8) we see that the integral over λ_i is divergent for $\rho = -\alpha$. We can extend the definition of (2.8) using the partial integration with respect to λ_i, obtaining

$$\mathbf{A}_p(\alpha, t) = K\Gamma(-\alpha) \frac{(-)^p}{(\rho+\alpha)^p} \int dx\, d\bar{y}^{(j)}\, d\lambda$$

$$\times \prod_{i=1}^{p} \lambda_i^{\rho+\alpha} \prod^{p} (\partial/\partial\lambda) J(\alpha) \qquad (2.10)$$

Since (2.10) also contains terms which behave as $(\rho+\alpha)^{-m-1}$ where $m+1 < p$, it is convenient to expand the intergrand in (2.10) in powers of $(\rho+\alpha)$ so that (2.10) is of the form

$$\mathbf{A}_p(\alpha, t) = \sum_{m=0}^{p-1} (\rho+\alpha)^{-m-1} \mathbf{A}_p(m), \qquad (2.11)$$

where

$$\mathbf{A}_p(m) = (-)^p K\Gamma(-\alpha) \sum_{l_1, l_2} \int dx\, d\bar{y}^{(j)}\, d\lambda$$

$$\times \frac{[\ln \prod \lambda_i]^{l_1}}{l_1!} \left(\prod^{p} \partial/\partial\lambda\right) \left\{\frac{1}{l_2!} \frac{[\ln \bar{f}/\Delta]^{l_2}}{[\bar{f}/\Delta]^\rho} e^{-\delta} \Delta^{-2}\right\}, \qquad (2.12)$$

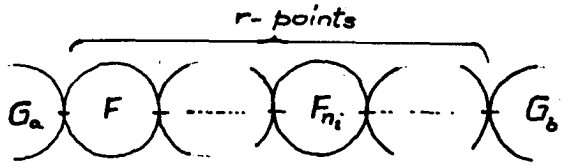

FIG. 4.

where

$$l_1 + l_2 = p - m - 1.$$

Of course, if one considers only leading asymptotic terms then $m = p - 1$. The aim is to find $\mathbf{A}(\alpha, t) = \sum_p \mathbf{A}_p(\alpha, t)$ in the closed form which gives the possibility of writing down the Regge trajectory. The term in (2.12) which contains λ_i in the form of the product can be expanded in the multinominal development

$$[\ln \prod^p \lambda_i]^{l_1} = l_1! \sum_{s_i} \prod^p \frac{[\ln \lambda_i^{s_i}]}{s_i!}, \qquad (2.13a)$$

where

$$\sum s_i = l_1 \qquad (2.13b)$$

and then we may integrate (2.12) over those λ's for which the s_is are equal to zero. This can be performed immediately to give these λ_i zero in the bracket { } of (2.12) and the corresponding differential operators removed.

If there are rs_i's equal to zero, the (2.12) will have the form

$$\mathbf{A}_p(m)_{rs_i'=0} = (-)^p K\Gamma(-\alpha) \sum_{l_1, l_2} \sum_{s_i} \int dx\, d\bar{y}^{(j)}\, d\lambda \times \prod \frac{[\ln \lambda_i]^{s_i}}{s_i!} (\prod \partial/\partial\lambda)[F_{l_1}(Q_i) e^{-\Sigma \delta}(\prod \Delta)^{-2}] \qquad (2.14)$$

with the function $F_{l_1}(Q_i) = F_{l_1}(Q_1, \cdots, Q_\sigma)$ given in the integral form, see Appendix I, as

$$F_{l_1}(Q_i) = \frac{1}{l_2!} \int \prod_{j=1}^{r} \delta(\sum \bar{y}^{(j)} - 1) \times \prod^{\rho} d\bar{y}^{(j)} \frac{\left[\ln \sum_{(i)}^{\sigma} (i_1, \cdots, i_r) Q(i_1, \cdots, i_r)\right]^{l_2}}{\left[\sum_{(i)}^{\sigma} (i_1, \cdots, i_r) Q(i_1, \cdots, i_r)\right]^{\rho}}, \qquad (2.15)$$

where

$$Q(i_1, \cdots i_r) = Q^{(1)}(0, i_1) Q^{(2)}(i_1, i_2) \cdots Q^{(r+1)}(i_r, 0)$$

and $Q^{(j)}(i_{j-1}, i_j)$ depends only on the bubble (j) and the positions i_{j-1} and i_j in the $\bar{y}^{(j-1)}$ and $\bar{y}(j)$ ρ line, respectively, $\sigma = \rho^r$ is the number of terms contained in the sum over (i);

$$(i_1 \cdots i_r) = \bar{y}_{i_1}^{(1)} \cdots \bar{y}_{i_r}^{(r)}.$$

The summation in (2.15) goes over all possible sets $(i_1 \cdots i_r)$ with $i_k = 1, 2, \cdots, \rho$. The integral (2.15) is taken over the $\bar{y}^{(j)}$ corresponding to the r ρ lines put equal to zero.

Unfortunately, we are not able to integrate (2.15) in a satisfactory way for $l_2 \neq 0$ and $\rho > 1$ or to find the most convenient form for F_{l_1} which would allow us to carry out the summation over p and m for general planar graph.

In the case of $l_2 = 0$ we can integrate (2.15) for all ρ, getting the function $F_0(Q_i)$ completely factorized. Because we want $l_2 \neq 0$ we will limit ourselves to the case when $\rho = 1$, which is completely solvable.

III. $\rho = 1$ CASE

When we take $\rho = 1$ the functions $F_{l_1}(Q_i)$ have the very simple form.

$$F_{l_1}(Q) = \frac{1}{l_1!} \frac{[\ln Q]^{l_1}}{Q}, \qquad (3.1)$$

where $Q = Q^{(1)} Q^{(2)} \cdots Q^{(r+1)}$ for rs_i's equal to zero as in (2.14).

The graph which will be considered satisfying the condition $\rho = 1$, is of the form given by Fig. 5.

It is assumed that all shaded blocks are of the same structure and there are no lines of the length one in the shaded blocks joining the $a\,b$ line with $c\,d$. This class of diagram, of course, contains, in particular, ladder diagrams for which Q is just $1/\Delta$.

Since the function (3.1) can be also factorized by putting $r\lambda$'s equal to zero, i.e.,

$$F_{l_1}(Q) = \frac{1}{\prod Q^{(i)}} \sum_{\beta_i} \prod^{r+1} \frac{[\ln Q^{(i)}]^{\beta_i}}{\beta_i!}, \qquad (3.2)$$

where

$$\sum \beta_i = l_2 = \beta, \qquad (3.3)$$

we shall have the completely factorized function $\mathbf{A}_p(m)$ for rs_i's put equal to zero. The conditions

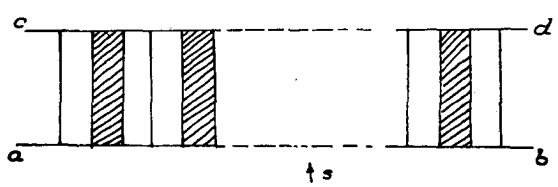

FIG. 5.

(2.13a) and (3.3) taken together satisfy the equation

$$\sum s_i + \beta = p - m - 1. \tag{3.4}$$

Following Polkinghorne[7] we can write $\mathbf{A}_p(m)$ in the form (Appendix II)

$$\mathbf{A}_p(m) = (1+\alpha)^{m+1} \sum_r (1+\alpha)^{-r} \frac{\partial_z^{p-r}}{(p-r)!} \times [(\sum \bar{G})(\sum \bar{F})^{r-1}(\sum \bar{G})]_{z=0}, \tag{3.5}$$

where

$$\bar{F}_n(\alpha) = \sum_{s_i \beta} F_n(s_i, \beta; \alpha)$$

and

$$F_n(s_i, \beta; \alpha) = \left(-\frac{g^2}{16\pi^2}\right)^{n+1} K_n \int dx\, d\lambda \times \prod^n \frac{[\ln \lambda_i]^{s_i}}{s_i!} (1+\alpha)^{s_i - 1} \tag{3.6}$$

$$(\prod^n \partial/\partial\lambda)\left\{\frac{[\ln Q^{(n)}]^\beta}{\beta!} (1+\alpha)^\beta \times [Q^{(n)}]^{-1}\Delta^{-2}(n) e^{-\delta^{(n)}}\right\} \tag{3.7}$$

(and similarly $\bar{G}_n$) where K_n includes only the coupling constants g and G from the shaded blocks (Fig. 5.) coming in a bubble (n). The brackets in (3.5) are polynomials in z for instant

$$\sum \bar{F}) = (\bar{F}_0 + z\bar{F}_1 + z^2\bar{F}_2 + \cdots). \tag{3.8}$$

IV. SUMMARY AND CONCLUSION

Now, it remains to find $\mathbf{A}(\alpha, t)$, defined

$$\mathbf{A}(\alpha, t) = \sum_p \mathbf{A}_p(\alpha, t) = \sum_p \sum_m (1+\alpha)^{-m-1} \mathbf{A}_p(m). \tag{4.1}$$

Putting (3.5) in (4.1), and keeping m and r fixed for a while, the sum over p will give the relation

$$\mathbf{A}(\alpha, t) = \sum_r (1+\alpha)^{-r}[(\sum \bar{G})(\sum \bar{F})^{r-1}(\sum \bar{G})]_{z=1}, \tag{4.2}$$

which is just a geometric series with the sum

$$\mathbf{A}(\alpha, t) = (\sum \bar{G})[1/(\alpha + 1 - \sum \bar{F})](\sum \bar{G}) \tag{4.3}$$

if (4.2), of course, converges. (4.3) has the same structure as was given by Polkinghorne for ladder diagrams. It is easy to see that the Regge poles are determined by the solution of

$$\mathfrak{F}(\alpha) = \alpha + 1 - \sum \bar{F} = 0. \tag{4.4}$$

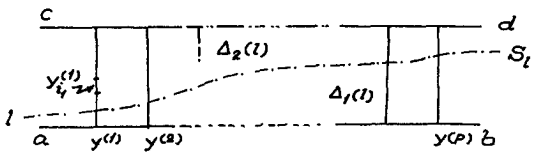

FIG. 6.

Since $\bar{G}$ and $\bar{F}$ are functions of t the convergence of the series (4.2) will depend upon the variable t. This can be investigated only for a particular graph for which the structure of $\delta(x, t)$ is known. We have been, of course, assuming that t is in such a region that the series (4.2) converges.

ACKNOWLEDGMENTS

For very helpful discussions and constant attention I am greatly indebted to Dr. R. F. Streater. I would also like to thank Professor A. Salam and Professor P. T. Matthews for their interest and encouragement, and to I.A.E.A. for the fellowship.

APPENDIX I

Function f/Δ and λ Transformation

The function f appearing as the coefficient of the energy variable s has a structure given by the relation (2.4)

$$f = \sum_i \Delta_1(i)\Delta_2(i) \prod_{i \in S_i} x_i. \tag{AI.1}$$

One can make $f = 0$ by putting a certain number of x's equal to zero, for instance $\rho_i x$'s, where $i = 1, 2, \cdots$ and $\rho_i = 1, 2, \cdots$. Following Halliday and Tiktopoulos[3,5] the graph will be ρ reducible if by putting only ρx's equal to zero we make $f = 0$, where $\rho = \min[\rho_i]$.

If we assume that all ρ lines, of which there are say p, do not intersect we shall have a graph of the structure shown in Fig. 6, where $y^{(j)}$ represents the $(j)\,\rho$ line. Each $y^{(j)}$ has ρ components $y_1^{(j)}, y_2^{(j)}, \cdots, y_\rho^{(j)}$. The remaining internal lines in the graph (Fig. 6) are labeled by x. Each term in (AI.1) will always contain one element from each ρ line in a product $\prod s_i$, i.e., one from the $y^{(1)}\rho$ line, one from $y^{(2)}\rho$ line and so on. We can fix, for a moment the product $y_{i_1}^{(1)} \cdots y_{i_2}^{(2)} \cdots y_{i_p}^{(p)}$ and consider all possible l lines passing through these points, and then sum over all possible products which can be constructed in such a way. This procedure allows (AI.1) to be written in the form

$$f = \sum_{(i)} \bar{y}_{i_1}^{(1)} \cdot \bar{y}_{i_2}^{(2)} \cdots \bar{y}_{i_p}^{(p)} (\sum \Delta_1\Delta_2 \prod x), \tag{AI.2}$$

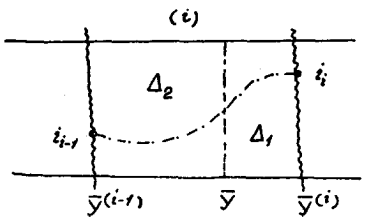

FIG. 7.

where (i) means a sum over all possible combinations of the set $(i_1, \cdots, i_p)$ with $i_k = 1, \cdots, \rho$ and the summation in $(\sum)$ is over all l lines passing through the set $(i_1 \cdots i_p)$ (Fig. 6). If we now replace the $y^{(i)}\rho$ line by $y^{(i)} \cdot \lambda_i$ where λ_i is a constant or a variable then (AI.2) has the form.

$$f = \left(\prod_{i=1}^{p} \lambda_i\right) \sum_{(i)} \bar{y}_{i_1}^{(1)} \cdots \bar{y}_{i_p}^{(p)} \left(\sum \Delta_1 \Delta_2 \prod x\right) = \left(\prod_{i=1}^{p} \lambda_i\right) \bar{f}. \quad \text{(AI.3)}$$

It is easily seen that putting any of the λ_i equal to zero makes $f = 0$. This is the idea of λ-transformation. λ's will certainly come in the function Δ_1 and Δ_2.

Let us suppose that we have put r's equal to zero. The product $\Delta_1 \Delta_2 \prod x$ will factorize as

$$\prod^{r+1} (\Delta_1(i)\Delta_2(i) \prod_i x).$$

Relabeling the contracted ρ lines as $1, 2, \cdots, r$, we can write f from (AI.3) in the form.

$$|_{r'_s=0} = \sum_{(i)} (i_1, \cdots i_r) \times \left[\sum \prod^{r+1} (\Delta_1 \Delta_2 \prod x\bar{y})\right], \quad \text{(AI.4)}$$

where

$(i_1 \cdots i_r) = \bar{y}_{i_1}^{(1)} \cdot \bar{y}_{i_2}^{(2)} \cdots \bar{y}_{i_r}^{(r)}$, and $(\Delta_1 \Delta_2 \prod x\bar{y})$ depends only on a particular bubble. Then it is easy to see that the summation and the product can be exchanged in (AI.4) and we obtain

$$|_{r'_s=0} = \sum_{(i)} (i_1, \cdots i_r) \prod^{r+1} \left(\sum \Delta_1 \Delta_2 \prod x\bar{y}\right). \quad \text{(AI.5)}$$

The summation in (AI.5) graphically means the following (Fig. 7), where the $i_{i-1}i_i$ line is $\prod x\bar{y}$ and the wavy lines are contracted ρ lines. The bracket $(\sum \Delta_1 \Delta_2 \prod x\bar{y})$ depends only upon the element inside one bubble and the way the $\prod x\bar{y}$ is joined with the contracted ρ lines which determined this bubble. Finally, we can write $\bar{f}/\Delta$, for $r\lambda$'s equal to zero, in the form

$$\frac{\bar{f}}{\Delta}\bigg|_{r\lambda'_s=0} = \sum_{(i)} (i_1, \cdots i_r) Q(i_1, \cdots i_r), \quad i_k = 1, 2, \cdots, \rho, \quad \text{(AI.6)}$$

where

$$Q(i_1, \cdots, i_r) = \prod^{r+1} \left(\sum \frac{\Delta_1(i)\Delta_2(i)}{\Delta(i)} \prod_i x\bar{y}\right) = \prod_{i=1}^{r+1} Q^{(i)}. \quad \text{(AI.7)}$$

Since $Q^{(i)}$ depends only on the way in which $Q^{(i)}$ is joined to the two nearest contracted ρ lines, the function $Q(i_1 \cdots i_r)$ has the form

$$Q(i_1, \cdots, i_r) = Q^{(1)}(0, i_1)Q^{(2)}(i_1, i_2) \cdots Q^{(r+1)}(i_r, 0). \quad \text{(AI.8)}$$

APPENDIX II

A derivation of Eq. (3.5) is given. The Eq. (2.14) with (3.2) can be rewritten in the form

$$\mathbf{A}_p(m)_{r s_i' s=0} = K(-)^p \times \sum_{s_i \beta_i} \int dx\, d\lambda \prod \frac{[\ln \lambda_i]^{s_i}}{s_i!} \left(\prod \partial/\partial\lambda\right) \times \left\{\prod^{r+1} \frac{[\ln Q^{(i)}]^{\beta_i}}{\beta_i!} [Q^{(i)}]^{-1} e^{-\sum \delta} \left(\prod \Delta\right)^{-2}\right\}, \quad \text{(AII.1)}$$

where the summation must satisfy the condition

$$\sum s_i + \beta = p - m - 1. \quad \text{(AII.2)}$$

The important point is that (AII.1) is completely factorised and allows the introduction of the functions F and G (Fig. 4)[7] defined as

$$F_{n_i}(s_i, \beta; \alpha) = \left(-\frac{g^2}{16\pi^2}\right)^{n_i} K_{n_i} \int dx\, d\lambda \times \prod^{n_i} \frac{[\ln \lambda_i]^{s_i}}{s_i!} (1+\alpha)^{s_i - 1} \left(\prod^{n_i} \partial/\partial\lambda\right) \times \left\{\frac{[\ln Q^{(n_i)}]^{\beta}}{\beta!} (1+\alpha)^{\beta} [Q^{(n_i)}]^{-1} \Delta^{-2}(n_i) e^{-\delta(n_i)}\right\} \quad \text{(AII.3)}$$

and

$$G_a(s_i, \beta; \alpha) = g\left(-\frac{g^2}{16\pi^2}\right)^{a} K_a \times \int dx\, d\lambda \prod^{a} \frac{[\ln \lambda_i]^{s_i}}{s_i!} (1+\alpha)^{s_i - 1} \left(\prod^{a} \partial/\partial\lambda\right) \times \left\{\frac{[\ln Q^{(a)}]^{\beta}}{\beta!} (1+\alpha)^{\beta} [Q^{(a)}]^{-1} \Delta^{-2}(a) e^{-\delta(a)}\right\}, \quad \text{(AII.4)}$$

where a and, similarly, b are the number of s_i in G_a and G_b, respectively; and n_i is the number of s_i

in F_{n_i}. a, b and n_i always satisfy the relation

$$a + b + \sum_{i=1}^{r-1} n_i + r = p, \qquad \text{(AII.5)}$$

where r is the number of s_i put equal to zero. The conditions (AII.2) and (AII.5) together become

$$m + 1 = -\sum^{a} (s_i - 1) - \beta_a - \sum^{b} (s_i - 1) - \beta_b - \sum_{i=1}^{r-1} [\sum^{n_i} (s_j - 1) + \beta_{n_i}] + r, \qquad \text{(AII.6)}$$

where

$$a, b = 0, 1, \cdots p - 1 \quad ; \quad p > 2$$
$$n_i = 0, 1, \cdots p - 2$$

which was the reason for introducing the functions F and G in the way shown in (AII.3) and (AII.4).

Having the functions (AII.3) and (AII.4), we can rewrite (AII.1) in the form of the product

$$\bar{G}_a \bar{F}_{n_1} \bar{F}_{n_2} \cdots \bar{F}_{n_{r-1}} \bar{G}_b, \qquad \text{(AII.7)}$$

where bars mean the corresponding summation over $s_i > 0$ and β. The sum over all possible ways of choosing rs_i's equal to zero is just

$$\sum_{\delta_i} \frac{2!}{\delta_0! \cdots \delta_k!} \bar{G}_0^{\delta_0} \cdots \bar{G}_k^{\delta_k}$$

$$\times \sum_{\gamma} \frac{(r-1)!}{\gamma_0! \cdots \gamma_{p-r-k}!} \bar{F}_0^{\gamma_0} \cdots \bar{F}_{p-r-k}^{\gamma_{p-r-k}} \qquad \text{(AII.8)}$$

with the conditions

$$\sum_{i=1}^{k} \delta_i = 2; \qquad \sum_{i=1}^{k} i\delta_i = k$$
$$\sum_{i=1}^{p-r-k} \gamma_i = r - 1; \qquad \sum_{i=1}^{p-r-k} i\gamma_i = p - r - k \qquad \text{(AII.9)}$$

so that (AII.8) can be written as

$$\frac{\partial_z^k}{k!} (\sum \bar{G})_{z=0}^2 \frac{\partial_z^{p-r-k}}{(p-r-k)!} (\sum \bar{F})_{z=0}^{r-1}, \qquad \text{(AII.10)}$$

where, for instance

$$(\sum \bar{F}) = \bar{F}_0 + z\bar{F}_1 + z^2\bar{F}_2 + \cdots \qquad \text{(AII.11)}$$

and similarly for $(\sum \bar{G})$.

The sum over k which goes from zero to $p - r$ tells us that (AII.1) summed over all r's which are allowed, is

$$\mathbf{A}_p(m) = (1 + \alpha)^{m+1} \sum_{r} (1 + \alpha)^{-r} \times \frac{\partial_z^{p-r}}{(p-r)!} [(\sum \bar{G})(\sum \bar{F})^{r-1} (\sum \bar{G})]_{z=0}, \qquad \text{(AII.12)}$$

which is just the relation (3.5).

Unitary Symmetry of Oscillators and the Talmi Transformation

BRURIA KAUFMAN
Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel
AND
CORNELIUS NOACK
Institut für Theoretische Physik, Universität Heidelberg, Heidelberg, Germany
(Received 21 July 1964)

The Hamiltonian of an isotropic harmonic oscillator is invariant under unitary transformations in three dimensions. This well-known invariance is exploited in a treatment of the Talmi transformation, viz., the transformation of two-particle oscillator functions to center-of-mass and relative coordinates. A simple and transparent form of this transformation in terms of rotation matrices and Wigner coefficients of SU_3 is given. The calculation of these Wigner coefficients is described and the problem of degeneracies discussed.

A. INTRODUCTION

IN shell-model calculations of nuclear energy levels with an effective two-body interaction one is led to the determination of two-particle matrix elements of the type

$$\langle a_1 a_2 |\ V_{12}\ | a_1' a_2' \rangle$$
$$= \int \phi_{a_1}^*(\mathbf{r}_1)\phi_{a_2}^*(\mathbf{r}_2) V(|\mathbf{r}_1 - \mathbf{r}_2|)\phi_{a_1'}(\mathbf{r}_1)\phi_{a_2'}(\mathbf{r}_2)\, d\tau_1\, d\tau_2. \quad (1)$$

Here ϕ is the wavefunction of a particle in the common average potential, which in practical calculations is often assumed to be that of a harmonic oscillator. The usual way to evaluate these matrix elements is the method of Slater integrals, well known in atomic spectroscopy. For the nuclear case, however, Talmi[1] has developed an alternative method, which is limited to the use of oscillator functions, but in that case is much more practicable. It rests on the fact that the Hamiltonian of two identical oscillators

$$\mathbf{H} = (1/2m)(\mathbf{p}_1^2 + \mathbf{p}_2^2) + \tfrac{1}{2}m\omega^2(\mathbf{r}_1^2 + \mathbf{r}_2^2) \quad (2)$$

is invariant under the transformation to center-of-mass and relative coordinates:

$$\mathbf{R} = (1/\sqrt{2})(\mathbf{r}_1 + \mathbf{r}_2), \quad \mathbf{r} = (1/\sqrt{2})(\mathbf{r}_1 - \mathbf{r}_2), \quad (3)$$

$$\mathbf{H} = (1/2m)(\mathbf{P}^2 + \mathbf{p}^2) + \tfrac{1}{2}m\omega^2(\mathbf{R}^2 + \mathbf{r}^2), \quad (2')$$

so that the solutions of $\mathbf{H}$ in terms of $\mathbf{R}$ and $\mathbf{r}$ are again oscillator functions. The integration over $\mathbf{R}$ is then immediately carried out, and the calculation of $\langle a_1 a_2\ |V_{12}|\ a_1' a_2' \rangle$ is reduced to that of the Talmi integrals[1]

$$I_l = N_l^2 \int_0^\infty V(r)\, e^{-\nu r^2} r^{2l+2}\, dr. \quad (4)$$

In the application of this method, one needs to know the coefficients in the expansion

$$\phi_{n_1 l_1 m_1}(\mathbf{r}_1)\phi_{n_2 l_2 m_2}(\mathbf{r}_2) = \sum_{NLMnlm} t^{NLMnlm}_{n_1 l_1 m_1, n_2 l_2 m_2}\, \phi_{NLM}(\mathbf{R})\, \phi_{nlm}(\mathbf{r}). \quad (5)$$

These coefficients, often called "transformation brackets," have been calculated in various ways by several authors[1-8] and have been extensively tabulated.[9] The methods used in all these calculations do not, however, give a direct insight into the mathematical structure underlying the Talmi transformation. It is the main purpose of the present paper to clarify this structure.

The essential feature of our treatment is the exploitation of the well-known invariance of the Hamiltonian (2) under U_6, the unitary group in six dimensions. (More generally, the Hamiltonian of N identical isotropic oscillators is invariant under U_{3N}.)[10,11] Thus the eigenfunctions of this Hamiltonian,

$$\phi_{n_1 l_1 m_1}(\mathbf{r}_1)\ \phi_{n_2 l_2 m_2}(\mathbf{r}_2) \quad (6)$$

constitute a basis for a representation of U_6. In setting up a classification of wavefunctions (6) according to irreducible representations, it is natural to consider the subgroup $U_2 \times SU_3$ contained in U_6,

[1] I. Talmi, Helv. Phys. Acta 25, 185 (1952).
[2] R. Thieberger, Nucl. Phys. 2, 533 (1956–1957).
[3] K. W. Ford, E. J. Konopinski, Nucl. Phys. 9, 218 (1957–1958).
[4] M. Moshinsky, Nucl. Phys. 13, 104 (1959).
[5] T. A. Brody, Rev. Mex. Fis. 8, 139 (1959).
[6] R. D. Lawson and M. Goeppert-Mayer, Phys. Rev. 117, 174 (1960).
[7] V. V. Balashov and V. A. Eltekov, Nucl. Phys. 16, 423 (1960).
[8] A. Arima and T. Terasawa, Prog. Theoret. Phys. (Kyoto) 23, 115 (1960).
[9] T. A. Brody and M. Moshinsky, *Tables of Transformation Brackets*, (Universidad de Mexico, Mexico City, 1960).
[10] J. M. Jauch and E. L. Hill, Phys. Rev. 57, 641 (1940).
[11] G. A. Baker, Jr., Phys. Rev. 103, 1119 (1956).

since the six coordinates of the functions (6) are physically grouped in 2×3 coordinates. This is done in Sec. C. We shall make use of the subgroup $U_2 \times SU_3$ to define a classification of wavefunctions based upon irreducible representations of SU_3. A reduction matrix (generalized Wigner coefficients) then takes us from the set of functions (6) to the set classified according to SU_3. Section D deals with the definition and actual computation of this reduction matrix.

The scheme defined in Sec. C is particularly adapted to the Talmi transformation (5); this is because the Talmi transformation happens to be a finite element of U_6, which belongs to the subgroup $U_2 \times SU_3$. Thus, once we have set up our scheme in Secs. C and D, the Talmi transformation may be written down immediately, as we see in Sec. E.

The classification of wavefunctions according to irreducible representations of SU_3 is, of course, not new. Some years ago Elliott[12] used this procedure in his model of nuclear rotational spectra, and more recently several papers have dealt extensively with oscillator bases for the irreducible representations of SU_3.[13-15] Because the mathematical problem of reducing the irreducible representations of SU_3 upon restriction to its subgroup O_3^+ cannot be completely defined by purely group-theoretical methods,[16,17] there are many such classifications, and most of them yield nonorthogonal bases. We show in Sec. E that this ambiguity in the definition of our scheme has no influence in the particular case of the Talmi transformation matrix.

B. SYSTEMS OF IDENTICAL OSCILLATORS

We begin by considering a system of N three-dimensional oscillators. While this system is more general than what is required by our problem, the discussion of its algebraic structure is no more complicated than that of the special case $N = 2$.

We define the creation and annihilation operators for oscillator quanta ($\hbar = m = \omega = 1$):

$$\mathbf{a}^*_{\alpha k} \equiv (1/\sqrt{2})(\mathbf{x}_{\alpha k} - i\mathbf{p}_{\alpha k}),$$
$$\mathbf{a}_{\alpha k} \equiv (1/\sqrt{2})(\mathbf{x}_{\alpha k} + i\mathbf{p}_{\alpha k}). \tag{7}$$

The index α refers to the particle number ($\alpha = 1, 2, \cdots, N$), while k denotes the Cartesian components. The commutators of $\mathbf{a}$ and $\mathbf{a}^*$ are

$$[\mathbf{a}_{\alpha k}, \mathbf{a}^*_{\beta l}] = \delta_{\alpha\beta}\delta_{kl}. \tag{8}$$

In terms of these operators, the Hamiltonian of N isotropic oscillators may be written as

$$\mathbf{H} \equiv \sum_{\alpha=1}^{N} \sum_{k=1}^{3} \mathbf{a}^*_{\alpha k}\mathbf{a}_{\alpha k} \tag{9}$$

(the zero-point energy has been dropped).

Now define the set of $(3N)^2$ operators

$$\mathbf{X}^{\alpha\beta}_{kl} \equiv \mathbf{a}^*_{\alpha k}\mathbf{a}_{\beta l}, \tag{10}$$

with the commutators

$$[\mathbf{X}^{\alpha\beta}_{kl}, \mathbf{X}^{\rho\sigma}_{mn}] = \delta_{\beta\rho}\delta_{lm}\mathbf{X}^{\alpha\sigma}_{kn} - \delta_{\alpha\sigma}\delta_{kn}\mathbf{X}^{\rho\beta}_{ml}. \tag{11}$$

The operators $\mathbf{X}^{\alpha\beta}_{kl}$ all commute with $\mathbf{H}$, and indeed are precisely the infinitesimal generators of U_{3N}.[18] In terms of these generators

$$\mathbf{H} = \sum_{\alpha=1}^{N} \sum_{k=1}^{3} \mathbf{X}^{\alpha\alpha}_{kk}. \tag{12}$$

The solutions of the Schrödinger equation $\mathbf{H}\Psi = E\Psi$ for a given integer E may be written

$$\Psi = C \cdot \mathbf{a}^*_{\alpha_1 k_1}\mathbf{a}^*_{\alpha_2 k_2} \cdots \mathbf{a}^*_{\alpha_E k_E} |0\rangle, \tag{13}$$

where the ground state $|0\rangle$ is defined by $a_{\alpha k}|0\rangle = 0$ for all α, k, $\langle 0 \mid 0\rangle = 1$, and C is a normalization coefficient. Since $\mathbf{H}$ is invariant under U_{3N}, the functions Ψ span a representation of this group. They are symmetrical in the permutations of any two creation operators, i.e., their symmetry is that of a completely symmetric tensor of rank E. Hence the representation of U_{3N} spanned by the set of Ψ's is irreducible, and characterized by the Young diagram $\{E, 0, 0, \cdots\}$, consisting of one row of length E.[19]

Within the algebra of U_{3N}, various subsets of the generators form subalgebras. Particularly relevant to our problem (as is shown in Sec. E) are the following two subalgebras: U_N, generated by

$$\mathbf{M}^{\alpha\beta} \equiv \sum_{k=1}^{3} \mathbf{X}^{\alpha\beta}_{kk}, \tag{14}$$

and SU_3, generated by

$$\mathbf{A}_{kl} \equiv \sum_{\alpha=1}^{N} \mathbf{X}^{\alpha\alpha}_{kl} - \tfrac{1}{3}\mathbf{H} \cdot \delta_{kl}. \tag{15}$$

Since these two sets commute:

$$[\mathbf{M}^{\alpha\beta}, \mathbf{A}_{kl}] = 0, \tag{16}$$

it follows that U_{3N} contains as a subgroup, the direct product of its two subgroups U_N and SU_3.

Upon restriction of U_{3N} to the elements of

[12] J. P. Elliott, Proc. Roy. Soc. (London) **245A**, 128, 562 (1958).
[13] V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).
[14] M. Kretzschmar, Z. Phys. **157**, 433 (1960).
[15] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[16] G. Racah, Lectures at the Istanbul Summer School of Theoretical Physics, 1962 (to be published).
[17] R. Sen, "Construction of the Irreducible Representations of SU_3," Ph.D. thesis, Jerusalem, Israel (1963).
[18] G. Racah, Group Theory and Spectroscopy, Lecture Notes, Princeton (1951) (reprint CERN 61-8).
[19] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Pergamon Press, New York, 1962), Chap. 10.

$U_N \times SU_3$, the irreducible representations of U_{3N} decompose into direct sums of irreducible representations of $U_N \times SU_3$. It follows, from theorems proved by Weyl,[20] that for the totally symmetric representations of U_{3N}:

$$[U_N \times SU_3]_{(E,0,0,\cdots)} = \sum_{\{\lambda\}} [U_N]_{\{\lambda\}}[SU_3]_{\{\lambda\}}, \quad (17)$$

where the sum is to be taken over all Young diagrams $\{\lambda\}$ which are partitions of E, with a number of rows $n \leq \min(3, N)$.

Equation (17) imposes a restriction on the Young diagrams $\{\lambda\}$: they must simultaneously characterize irreducible representations of U_N and of SU_3. However, as is well known,[19] all inequivalent irreducible representations of SU_3 may be characterized by Young diagrams of at most two rows. Equation (17) then shows that already in the case of two oscillators *all* of the irreducible representations of SU_3 are involved, as has been pointed out previously by Moshinsky.[21]

C. OPERATORS FOR THE TWO-OSCILLATOR SYSTEM

Specializing now to $N = 2$ we have the reduction $U_6 \supset U_2 \times SU_3$. Of the four generators in U_2, three combinations can be formed which constitute the components of an angular-momentum-type vector (and generate the subgroup SU_2 within U_2). We write this "pseudospin" vector[13] in a spherical basis

$$\mathbf{F}_+ \equiv -(1/\sqrt{2})\mathbf{M}^{12}, \quad \mathbf{F}_0 \equiv \tfrac{1}{2}(\mathbf{M}^{11} - \mathbf{M}^{22}), \quad \mathbf{F}_- \equiv (1/\sqrt{2})\mathbf{M}^{21}, \quad (18)$$

defining, as usual, the square of $\mathbf{F}$ as the scalar operator

$$\mathbf{F}^2 \equiv \sum_{m=0,\pm} (-)^m \mathbf{F}_m \cdot \mathbf{F}_{-m} = (\mathbf{F}_0^2 - \mathbf{F}_0) + \mathbf{M}^{12} \cdot \mathbf{M}^{21}, \quad (19)$$

which commutes with its three components. There remains the fourth combination of generators within U_2,

$$\mathbf{H} = \mathbf{M}^{11} + \mathbf{M}^{22}, \quad (20)$$

which commutes with all other operators of the group. Thus we may select a set of three operators within U_2 ($\mathbf{H}$, $\mathbf{F}^2$, and $\mathbf{F}_0$) which commute with one another; their eigenvalues are denoted by E, $f(f+1)$ and ν, respectively.

These three eigenvalues suffice to characterize the functions which span the representations of U_2 : E and f characterizing the irreducible representation, and ν enumerating the functions within this representation. The characterization by means of E and f is, of course, fully equivalent to that by the Young diagram $\{\lambda\} \equiv \{\lambda_1, \lambda_2\}$; and indeed from (17) we have the relationship

$$E = \lambda_1 + \lambda_2. \quad (21)$$

As for f, we note that the irreducible representations of U_2 must be of dimension $(2f+1)$, because of the angular-momentum character of the $\mathbf{F}$ operators. On the other hand, $\dim [U_2]_{(\lambda_1,\lambda_2)} = (\lambda_1 - \lambda_2 + 1)$ (see Ref. 19). By equating the two, we find

$$f = \tfrac{1}{2}(\lambda_1 - \lambda_2). \quad (22)$$

Turning now to SU_3, we have three combinations of its generators which constitute the angular-momentum operator:

$$\mathbf{J}_k \equiv -i(\mathbf{A}_{lm} - \mathbf{A}_{ml}) \quad (k, l, m \text{ cyclic}), \qquad \mathbf{J}^2 \equiv \sum_{k=1}^{3} \mathbf{J}_k^2. \quad (23)$$

These operators form a subalgebra within SU_3, corresponding to the subgroup $O_3^+ \subset SU_3$. The interest in this subgroup is motivated by physical considerations: the effective interactions in Eq. (1), as well as the Hamiltonian (2) or (9), are spherically symmetric, so that we would like our functions to be eigenfunctions of the angular momentum operators.

A study of the full algebra[17] of SU_3 would have to consider the remaining five combinations of generators which, in the spherical basis, constitute an irreducible tensor set of rank 2.[12] However, for our purposes such a detailed study is not required. We need to know how to characterize irreducible representations of SU_3, and this information is provided by Eq. (17), which in our case ($N = 2$) becomes

$$[U_2 \times SU_3]_{(E,0,\cdots)} = \sum_{(\lambda_1,\lambda_2)} [U_2]_{(\lambda_1,\lambda_2)} \times [SU_3]_{(\lambda_1,\lambda_2)}. \quad (17')$$

Thus (indirectly) we find that the irreducible representations of SU_3 may be characterized by $\{\frac{1}{2}E + f, \frac{1}{2}E - f\}$; and indeed the study of the full algebra of SU_3 would lead to the same result.[22]

[20] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1931), 2nd ed. (revised), Chap. V 12.

[21] M. Moshinsky, Nucl. Phys. **31**, 384 (1962).

[22] The operator $\mathbf{F}^2$ [Eq. (19)] is closely related to the contraction $\mathbf{Q}^2$ of the above-mentioned tensor of rank 2; one has the operator indentity

$$\mathbf{Q}^2 \equiv 2\mathbf{F}^2 + \tfrac{1}{4}\mathbf{H}^2 + \mathbf{H} - \tfrac{1}{2}\mathbf{J}^2.$$

Furthermore, $\mathbf{F}^2$, which is the Casimir operator of U_2, is closely related to the Casimir operator $\mathbf{G}_2$ of SU_3, defined [16] as

$$\mathbf{G}_2 \equiv \tfrac{1}{8}\sum_{i,k} \mathbf{A}_{ik} \cdot \mathbf{A}_{ki}.$$

It may be verified that the following operator identity holds:

$$6\mathbf{G}_2 \equiv \tfrac{1}{2}\mathbf{J}^2 + \mathbf{Q}^2, \text{ so that } 6\mathbf{G}_2 \equiv 2\mathbf{F}^2 + \tfrac{1}{8}\mathbf{H}^2 + \mathbf{H}.$$

Hence the eigenvalue of $\mathbf{G}_2$ in the representation $\{\lambda_1, \lambda_2\}$ is given by

$$g_2 = \tfrac{1}{3}(\lambda_1^2 + \lambda_2^2 - \lambda_1\lambda_2 + 3\lambda_1)$$

in accordance with Eq. (106) of Ref. 18.

D. TWO-PARTICLE FUNCTIONS

The chain of reductions

$$U_6 \supset U_2 \times SU_3 \supset U_2 \times O_3^+$$

has yielded a set of five operators which commute with one another: $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, $\mathbf{J}^2$, and $\mathbf{J}_0$. We now require our two-particle functions to be simultaneous eigenfunctions of these five operators. Such a requirement is not quite sufficient to define the two-particle functions, because, as pointed out in the introduction, the classification of states in an irreducible representation of SU_3 by the quantum numbers of O_3^+ is incomplete. We need three quantum numbers to label the states of SU_3,[18] and O_3^+ yields only two, the total angular momentum $\mathbf{J}^2$ and $\mathbf{J}_0$. In other words: an irreducible representation of SU_3 decomposes upon restriction to O_3^+, and in this decomposition a given representation of O_3^+ may occur several times, so that we need an additional label α to distinguish these states.[23]

This additional quantum number α cannot be defined by the introduction of yet another subgroup G' of SU_3, because G' would in turn have to contain O_3^+, and no such group exists. For the time being we merely consider α to be an unspecified label that distinguishes between representations D_J having the same J.

The multiplicity N_J of a representation D_J of O_3^+ in a representation of SU_3 characterized by the partition $\{\lambda_1, \lambda_2\}$ is given by the formula[25]

$$N_J(\lambda_1, \lambda_2) = P\left(\frac{\lambda_1 + 2 - J}{2}\right) - P\left(\frac{\lambda_2 + 1 - J}{2}\right) - P\left(\frac{\lambda_1 - \lambda_2 + 1 - J}{2}\right), \quad (24)$$

where

$$P(x) = \begin{cases} [x] & \text{for } x \geq 0 \\ 0 & x \leq 0. \end{cases}$$

The two-particle states may now be denoted by

$$|Ef\nu\alpha Jm\rangle, \quad (25)$$

and by the definition of these functions we know not only their eigenvalues under the operations $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, $\mathbf{J}^2$, and $\mathbf{J}_0$, but also their transformation properties under other operations of the group. For, in consequence of the fact that the two-particle functions are eigenfunctions of $\mathbf{F}^2$ and $\mathbf{F}_0$, it follows that under the operations $F_\pm$ they transform according to

$$\mp\sqrt{2}\,\mathbf{F}_\pm\,|Ef\nu\alpha Jm\rangle = [(f \mp \nu)(f \pm \nu + 1)]^{\frac{1}{2}}\,|Ef, \nu \pm 1, \alpha Jm\rangle, \quad (26)$$

and a similar transformation holds for the operators $\mathbf{J}_\pm$ which raise and lower the index m.

In view of Eq. (1) we now attempt to express the two-particle functions (25) in terms of the well-known *single*-particle oscillator functions. For a single oscillator, the total symmetry group of the Hamiltonian is $U_3(N = 1)$. The reduction chain is simply $U_3 \supset SU_3 \supset O_3^+$, and Eq. (24) then gives us the familiar result that the angular momenta l occurring within an oscillator level of energy ϵ (an irreducible representation of U_3) are

$$l = \epsilon, \epsilon - 2, \epsilon - 4, \cdots, 0 \text{ or } 1. \quad (27)$$

The quantum number α is therefore superfluous in the single-particle case. We use the notation

$$|\epsilon lm\rangle \quad (28)$$

for the single-particle oscillator functions, and define these as being simultaneous eigenfunctions of $\mathbf{H}$, $\mathbf{J}^2$, $\mathbf{J}_0$ (defined for the case $N = 1$):

$$\begin{aligned} \mathbf{H}|\epsilon lm\rangle &= \epsilon \cdot |\epsilon lm\rangle, \\ \mathbf{J}^2|\epsilon lm\rangle &= l(l+1) \cdot |\epsilon lm\rangle, \\ \mathbf{J}_0|\epsilon lm\rangle &= m \cdot |\epsilon lm\rangle. \end{aligned} \quad (29)$$

From Eq. (13) it follows that any eigenfunction of $\mathbf{H}$ must be a linear combination of monomials in $\mathbf{a}_k^*$ (of degree ϵ) which operate on the ground state $|0\rangle$; hence $|\epsilon lm\rangle$ must be of the form

$$\Psi_\epsilon = P_\epsilon(\mathbf{a}^*)\,|0\rangle, \quad (30)$$

where P_ϵ is a polynomial of degree ϵ.

Now it has been noted by Bargmann and Moshinsky[13] that if we let $\mathbf{a}_k$ operate on the state $P_\epsilon(\mathbf{a}^*)\,|0\rangle$ the result is the same as if we had applied $\partial/\partial\mathbf{a}_k^*$ to $P_\epsilon(\mathbf{a}^*)$ and then let the result operate on the state $|0\rangle$. Thus the operator $\mathbf{a}_i^*\mathbf{a}_k$ may be replaced by $\mathbf{a}_i^*\partial/\partial\mathbf{a}_k^*$ [acting on $P_\epsilon(\mathbf{a}^*)$], and we see that the angular momentum operators [Eq. (23)] take on the same form as they have in the configurational coordinates, with $\mathbf{a}_k^*$ corresponding to x_k. It then follows that the polynomials $P_\epsilon(\mathbf{a}^*)$ must be solid spherical harmonics in $\mathbf{a}_k^*$; and we may write:

$$|\epsilon lm\rangle = A_{\epsilon lm} r^\epsilon e^{im\phi} P_l^m(\cos\theta)\,|0\rangle, \quad (31)$$

[23] The situation is quite different if we restrict SU_3 to its subgroup U_2, as is done in the applications of SU_3 to strong interaction symmetries.[24] U_2 is labeled by three quantum numbers (hypercharge, isospin, and third component of isospin), and consequently in the decomposition of SU_3 every representation of U_2 occurs at most once, so that the U_2 scheme is completely defined.

[24] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).

[25] G. Racah, Rev. Mod. Phys. **21**, 494 (1949).

where $A_{\epsilon lm}$ is a normalization constant, and r, θ, ϕ are the spherical polar coordinates of the point whose Cartesian coordinates are a_1^*, a_2^*, a_3^*.

The observation of Bargmann and Moshinsky makes available to us all the analytical information about spherical harmonics. In particular, it makes it easy to compute the effect of a_k and a_k^* upon the $|\epsilon lm\rangle$, out of which the matrix representation of these operators in the basis $|\epsilon lm\rangle$ may be obtained.

We construct, out of the operators a_k, an irreducible tensor operator of rank 1, whose spherical components are defined by

$$a_+ \equiv \mp(a_1 \pm ia_2)/\sqrt{2} \qquad a_0 \equiv a_3. \tag{32a}$$

Similarly, from the a_k^* we construct another tensor[26] of rank 1 with spherical components

$$b_+ \equiv \mp(a_1^* \pm a_2^*)/\sqrt{2} \qquad b_0 \equiv a_3^*, \tag{32b}$$

and then we find, for the *reduced* matrix elements (see Appendix B)

$$\langle \epsilon - 1, l + 1 \,||\, a \,||\, \epsilon, l\rangle = +[(l+1)(\epsilon - l)]^{\frac{1}{2}},$$
$$\langle \epsilon - 1, l - 1 \,||\, a \,||\, \epsilon, l\rangle = -[l(\epsilon + l + 1)]^{\frac{1}{2}}, \tag{33}$$

and

$$\langle \epsilon' l' \,||\, b \,||\, \epsilon l\rangle = -\langle \epsilon l \,||\, a \,||\, \epsilon' l'\rangle;$$

all other matrix elements vanish.

The single-particle operator $\mathbf{H} = \sum_k a_k^* \cdot a_k$ is now seen to be the scalar product $(b \cdot a)$. Defining for each particle similar tensors of rank 1, $a^{(\alpha)}$ and $b^{(\alpha)}$, one sees that all the generators of U_2 may be expressed as scalar products in terms of these: $\mathbf{M}^{\alpha\beta} = (b^{(\alpha)} \cdot a^{(\beta)})$. Equations (33) then enable us to compute the effect of the operators $\mathbf{M}^{\alpha\beta}$ upon the product functions $|\epsilon_1 l_1 m_1\rangle \cdot |\epsilon_2 l_2 m_2\rangle$.

Since we wish to obtain combinations of product functions in which the operators $\mathbf{J}^2$ and $\mathbf{J}_0$ are diagonal, it is natural to combine the product functions immediately into

$$|\epsilon_1 l_1 \epsilon_2 l_2 J m\rangle \equiv \sum_{m_1 m_2} (-)^{-l_1+l_2-m}(2J+1)^{\frac{1}{2}} \times \begin{pmatrix} l_1 & l_2 & J \\ m_1 & m_2 & -m \end{pmatrix} \cdot |\epsilon_1 l_1 m_1\rangle\, |\epsilon_2 l_2 m_2\rangle, \tag{34}$$

which by their construction are eigenfunctions of $\mathbf{J}^2$ and $\mathbf{J}_0$. We now ask for the effect of $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, and $\mathbf{F}_\pm$ upon functions of the type (34). We note that these operators are all expressible in terms of $\mathbf{M}^{\alpha\beta}$; in particular, $\mathbf{H} = \mathbf{M}^{11} + \mathbf{M}^{22}$ and $\mathbf{F}_0 = \frac{1}{2}(\mathbf{M}^{11} - \mathbf{M}^{22})$. But from Eq. (29) we see that

$$\mathbf{M}^{\alpha\alpha}\, |\epsilon_\alpha l_\alpha m_\alpha\rangle = \epsilon_\alpha \cdot |\epsilon_\alpha l_\alpha m_\alpha\rangle, \tag{35}$$

so that the functions (34) are eigenfunctions of $\mathbf{H}$ and $\mathbf{F}_0$ with the eigenvalues

$$E = \epsilon_1 + \epsilon_2 \quad \text{and} \quad \nu = \tfrac{1}{2}(\epsilon_1 - \epsilon_2), \tag{36}$$

respectively.

Thus we find that the four operators $\mathbf{H}$, $\mathbf{F}_0$, $\mathbf{J}^2$, $\mathbf{J}_0$ are all diagonal in the functions[27] $|\epsilon_1 l_1 \epsilon_2 l_2 J\rangle$. This is not the case for the operator $\mathbf{F}^2 = (\mathbf{F}_0^2 - \mathbf{F}_0) - \mathbf{M}^{12} \cdot \mathbf{M}^{21}$. The matrix elements of $\mathbf{M}^{12}$ and its adjoint $\mathbf{M}^{21}$ are computable by a straightforward application of standard methods:[28]

$$\langle \epsilon_1 l_1 \epsilon_2 l_2 J |\, \mathbf{M}^{12}\, |\epsilon_1' l_1' \epsilon_2' l_2' J\rangle = (-)^{E+J} \begin{Bmatrix} l_1 l_2 J \\ l_2' l_1' 1 \end{Bmatrix} \times \langle \epsilon_1' l_1' \,||\, a \,||\, \epsilon_1 l_1\rangle \langle \epsilon_2 l_2 \,||\, a \,||\, \epsilon_2' l_2'\rangle. \tag{37}$$

By Eq. (33) we must have $\epsilon_1' = \epsilon_1 - 1$ and $\epsilon_2' = \epsilon_2 + 1$. Because of this fact, the operator $\mathbf{F}_+ = -(1/\sqrt{2})\, \mathbf{M}^{12}$ and its adjoint $-\mathbf{F}_-$ are referred to as "transfer operators,"[13] which remove a unit from the energy index of one particle and add it on to the other. Clearly, the product of the two transfer operators, which appears in the expression for $\mathbf{F}^2$, can have matrix elements only between initial and final states having the *same* energy index, since the unit transferred by $\mathbf{M}^{21}$ is restored by $\mathbf{M}^{12}$. Hence $\mathbf{F}^2$ has as its only nonvanishing matrix elements

$$\langle \epsilon_1 l_1 \epsilon_2 l_2 J |\, \mathbf{F}^2\, |\epsilon_1 l_1' \epsilon_2 l_2' J\rangle,$$

where $l_1' = l_1 + (0 \text{ or } \pm 2)$, and similarly for l_2. Since E and ν are fully determined by ϵ_1, ϵ_2 [Eq. (36)], this means that the term $(\mathbf{F}_0^2 - \mathbf{F}_0)$ which appears in $\mathbf{F}^2$ is merely a multiple of the unit matrix within the subspace with fixed E, ν, J (and, of course, m).

Our task now is to pass from the functions (34), in which four of the operators are diagonal, to linear combinations of these in which also $\mathbf{F}^2$ will be diagonal. This transformation will take place in subspaces of fixed E, ν, J, m, so that the diagonal form of the corresponding four operators is unaffected. From its dependence on $\mathbf{M}^{12}$, $\mathbf{M}^{21}$, one sees that $\mathbf{F}^2$ is a symmetric matrix, hence diagonalizable by an orthogonal transformation.

Now, if $\mathbf{F}^2$ has distinct eigenvalues within the subspace, the diagonalization process is unique, and the resulting functions are the desired two-particle functions. If, however, $\mathbf{F}^2$ has degenerate eigen-

[26] The tensor b is not the Hermitian conjugate of a, but rather $b_{-m} = (-)^m (a_m)^*$.

[27] The matrix elements for all operators under consideration are independent of m; henceforth m will be omitted from the designation of the function.

[28] In establishing the sign of the matrix element, we use the fact that the parity of l_i equals the parity of ϵ_i, and that furthermore, $\epsilon_i' = \epsilon_i \pm 1$ in the matrix elements. This allows us to replace $(-)^{l_1'+l_2}$ by $(-)^{\epsilon_1+\epsilon_2+1} = (-)^{E+1}$.

values,[29] the transforming matrix is no longer unique, and a further requirement must be made to specify the two-particle functions completely. Nevertheless, as has already been mentioned, the computation of the brackets in the Talmi transformation can be carried out without completing the specification.

Assuming that by some process of specification, the transformation from the functions (34) to the two-particle functions has been determined uniquely; we denote that transformation by K, with matrix elements

$$(\epsilon_1 l_1 \epsilon_2 l_2 J \mid E f \nu \alpha J).$$

The indices E, ν are redundant, inasmuch as these are fixed by ϵ_1, ϵ_2. We therefore abbreviate our notation and write

$$K = (\epsilon_1 l_1 \epsilon_2 l_2 \mid f \alpha J). \tag{38}$$

And now, the remarks made above about the lack of uniqueness in the diagonalization of $\mathbf{F}^2$ imply that whatever transformation matrix we might find will be of the form $K \cdot A \equiv R$ where A is *any* matrix which commutes with the diagonal form of $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, $\mathbf{J}^2$, $\mathbf{J}_0$.

In our computations of the matrix R (see Appendix C) we have required R to be an *orthogonal* transformation which brings us from the functions (34) to a basis in which $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, $\mathbf{J}^2$, $\mathbf{J}_0$ are diagonal. We have, in addition, required R to bring the operators $\mathbf{F}_+$ and $\mathbf{F}_-$ to their canonical form (26). These two requirements restrict the freedom of the undetermined factor A, by which R differs from K: A may now be any orthogonal matrix which commutes with the diagonal form of $\mathbf{H}$, $\mathbf{F}^2$, $\mathbf{F}_0$, $\mathbf{J}^2$, $\mathbf{J}_0$, and which also commutes with the canonical forms of $\mathbf{F}_+$ and $\mathbf{F}_-$. The procedure which we have adopted for the computation of R involves $(g-1)$ arbitrary choices of eigenvectors in any g-fold degenerate case (e.g., one choice in the case $E = 6$, $J = 2$).

E. THE TALMI TRANSFORMATION

The transformation τ in configuration space, which was defined by

$$\begin{aligned} \tau : \mathbf{r}_1 \to \mathbf{r} &\equiv (1/\sqrt{2})(\mathbf{r}_1 - \mathbf{r}_2), \\ \mathbf{r}_2 \to \mathbf{R} &= (1/\sqrt{2})(\mathbf{r}_1 + \mathbf{r}_2), \end{aligned} \tag{3'}$$

induces a linear transformation T in the space of the wave functions, enabling us to expand functions of $(\mathbf{r}_1, \mathbf{r}_2)$ in terms of functions of $(\mathbf{r}, \mathbf{R})$ [see Eq. (5)]. In this expansion, the coefficients are simply elements of the matrix which represents τ in a basis spanned by the given wavefunctions.

One sees immediately that τ belongs to U_6, and, moreover, that it is contained in the subgroup $U_2 \subset U_2 \times SU_3 \subset U_6$:

$$\begin{aligned} \tau = \tau_2 \times \mathbf{1} \in U_6 \qquad \mathbf{1} &\equiv \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \in SU_3, \\ \tau_2 \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} &= \exp\left\{ \frac{\pi}{4} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\} \in U_2. \end{aligned} \tag{39}$$

In order to identify the transformation in function space which is induced by τ, we note that the effect of τ on configuration space may be equally well obtained by the *differential* operator

$$\exp\left\{ \frac{\pi}{4} \sum_k \left(-x_{2k} \frac{\partial}{\partial x_{1k}} + x_{1k} \frac{\partial}{\partial x_{2k}} \right) \right\}. \tag{40}$$

By the correspondence between the configurational coordinates and operators $\mathbf{a}_{\alpha k}$, $\mathbf{a}^*_{\alpha k}$ we see that the differential operator in the exponent of (40) corresponds to the operator

$$\begin{aligned} \sum_k (-\mathbf{a}^*_{2k}\mathbf{a}_{1k} + \mathbf{a}^*_{1k}\mathbf{a}_{2k}) &\equiv -\mathbf{M}^{21} + \mathbf{M}^{12} \\ &= -\sqrt{2}(\mathbf{F}_+ + \mathbf{F}_-) \equiv 2i\mathbf{F}_y. \end{aligned} \tag{41}$$

It follows that the operation induced by τ in function space is

$$T = \exp\{\tfrac{1}{2} i \pi \mathbf{F}_y\}. \tag{42}$$

Since we know the explicit representation of $\mathbf{F}_y = (i/\sqrt{2})(\mathbf{F}_+ + \mathbf{F}_-)$ in the basis $|\epsilon_1 l_1 \epsilon_2 l_2 J\rangle$, it is in principle possible to evaluate the matrix representation of T in this basis by exponentiation, as indicated in Eq. (42). However, this method is not practicable. Instead, we make use of the fact that the representation of $\exp\{\frac{1}{2} i \pi \mathbf{F}_y\}$ is well known (and tabulated) in the basis where $\mathbf{F}_+$ and $\mathbf{F}_-$ have their canonical form Eq. (26). The two-particle functions dealt with in the previous section provide precisely such a basis, and we shall follow Edmonds[30] in denoting the irreducible representation matrices of $\exp\{\frac{1}{2} i \pi \mathbf{F}_y\}$ by $\Delta^{(J)}$.

The effect of the transformation τ on the two-particle functions may be thus written down immediately:

$$|E f \nu' \alpha J m\rangle(\mathbf{r}, \mathbf{R}) = \sum_\nu \Delta^{(J)}_{\nu\nu'} \cdot |E f \nu \alpha J m\rangle(\mathbf{r}_1, \mathbf{r}_2). \tag{43}$$

In order to obtain the effect of τ on the original functions $|\epsilon_1 l_1 \epsilon_2 l_2 J m\rangle$, we must transform Eq. (43)

[29] A twofold degeneracy appears for $E \geq 6$, a threefold degeneracy for $E \geq 12$. See Table 2 in Ref. 13.

[30] A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, New Jersey, 1957), Chap. 4.5.

by the Wigner coefficients (38):

$$|\mathcal{E}L\varepsilon lJm\rangle = \sum_{f\alpha}\sum_{\nu}\sum_{l_1l_2}(\mathcal{E}L\varepsilon l \mid f\alpha J)\Delta^{(f)}_{\nu\nu'} \cdot(\epsilon_1 l_1\epsilon_2 l_2 \mid f\alpha J)\cdot|\epsilon_1 l_1\epsilon_2 l_2 J m\rangle, \quad (44)$$
$$2\nu = \epsilon_1 - \epsilon_2, \quad 2\nu' = \mathcal{E} - \varepsilon.$$

The transformation brackets are then

$$t \equiv \langle\epsilon_1 l_1\epsilon_2 l_2 J \mid \mathcal{E}L\varepsilon lJ\rangle = \sum_{f\alpha}(\mathcal{E}L\varepsilon l \mid f\alpha J)\Delta^{(f)}_{\nu\nu'}\cdot(\epsilon_1 l_1\epsilon_2 l_2 \mid f\alpha J), \quad (45)$$

or, in abbreviated notation,

$$t = K\,\Delta\,K^{-1}. \quad (46)$$

If we substitute $R \equiv K\cdot A$ for K in Eq. (46), where A is the undetermined factor discussed in Sec. D, we see that

$$R\,\Delta\,R^{-1} = (KA)\,\Delta\,(KA)^{-1} = K\,\Delta\,K^{-1} = t \quad (47)$$

since A commutes with the canonical form of $(\mathbf{F}_+ + \mathbf{F}_-)$ by the requirements made on R. Hence, the incompleteness of the specification of K has no effect on the matrix t.

Since R was also required to be orthogonal, transformation by R preserves unitarity, so that t is a real unitary matrix.

Since $\Delta^{(f)}_{\nu\nu'} = (-)^{\nu-\nu'}\cdot\Delta^{(f)}_{\nu'\nu}$ (see Ref. 30), it follows (again from the orthogonality of R) that transposing the matrix t can only change the signs of some of its elements. The symmetry rules for the transformation brackets are derived in Ref. 31 by similar group-theoretical considerations.

Finally, we give several examples of the computation of transformation brackets, using Eq. (45) and the tables of R in Appendix D. The results check with those tabulated in Ref. 9. It is to be noted that our functions $|\epsilon_1 l_1\epsilon_2 l_2 J\rangle$ appear as $|n_1 l_1 n_2 l_2 J\rangle$ in the tables of Ref. 9. In order to facilitate comparison we use, in these examples, the quantum numbers n_i in the *brackets*, but replace $(\epsilon_1 l_1\epsilon_2 l_2 \mid f\alpha J)$ by the fully equivalent symbol $(l_1 l_2 \mid EJ;\ \nu \mid f\alpha)$. Here n_i and ϵ_i, l_i are connected through $\epsilon_i \equiv 2n_i + l_i$ while

$$\epsilon_1 + \epsilon_2 = E, \qquad \epsilon_1 - \epsilon_2 = 2\nu.$$

Thus Eq. (45) is rewritten as

$$\langle n_1 l_1 n_2 l_2 J \mid NLnlJ\rangle = \sum_{f,\alpha}(l_1 l_2\,|EJ;\nu|\,f\alpha)(Ll\,|EJ;\nu'|\,f\alpha)\cdot\Delta^{(f)}_{\nu\nu'}. \quad (45')$$

The tabulation of K in Appendix D corresponds to this notation: matrices are labeled by the triplet $(EJ;\ \nu)$, rows by $(l_1 l_2)$, and columns by $(f\alpha)$.

The label α is required only in the case of degeneracy. [The degree of degeneracy is given by $N_J(\frac{1}{2}E + f,\ \frac{1}{2}E - f)$, Eq. (24).] Since $\Delta^{(f)}_{\nu\nu'}$ must vanish if the subscripts are larger than f in their absolute magnitude, only those terms will contribute to the sum where

$$2f \geq \max\{|\epsilon_1 - \epsilon_2|,\ |\mathcal{E} - \varepsilon|\}.$$

EXAMPLES

(a) $E = 4$, $J = 0$. $N_J = 1$ for $f = 2$, and vanishes otherwise.

$$\langle 20,\,00;0 \mid 10,\,10;0\rangle = (00 \mid 40;2 \mid 2)(00 \mid 40;0 \mid 2)\cdot\Delta^{(2)}_{20} = (1)\cdot\left(\frac{5^{\frac{1}{2}}}{3}\right)\cdot\left(\frac{6^{\frac{1}{2}}}{4}\right) = \frac{5^{\frac{1}{2}}}{2(6)^{\frac{1}{2}}}.$$

(b) $E = 6$, $J = 2$. This case is of particular interest, because of the occurrence of degeneracy ($N_J = 2$ for $f = 1$).

$$\langle 04,\,02;2 \mid 22,\,00;2\rangle = \sum_{f=3}(42 \mid 62;1 \mid f\alpha)(20 \mid 62;3 \mid f\alpha)\cdot\Delta^{(f)}_{13} = (1)\cdot\left(\frac{4}{5(7)^{\frac{1}{2}}}\right)\cdot\left(\frac{15^{\frac{1}{2}}}{8}\right) = \left(\frac{3}{140}\right)^{\frac{1}{2}}$$

$$\langle 12,\,02;2 \mid 12,\,10;2\rangle = \sum_{f,\alpha}(22 \mid 62;1 \mid f\alpha)(20 \mid 62;1 \mid f\alpha)\cdot\Delta^{(f)}_{11}$$
$$= \left(\frac{-2(2)^{\frac{1}{2}}}{(5\cdot7)^{\frac{1}{2}}}\right)\cdot\left(\frac{2^{\frac{1}{2}}}{5^{\frac{1}{2}}}\right)\cdot\left(\frac{-1}{8}\right) + \left(\frac{1}{(2\cdot3\cdot7)^{\frac{1}{2}}}\right)\left(\frac{1}{(2\cdot3)^{\frac{1}{2}}}\right)\left(\frac{-1}{2}\right) + \left[\left(\frac{7}{(2\cdot3\cdot11)^{\frac{1}{2}}}\right)\left(\frac{7^{\frac{1}{2}}}{(2\cdot3\cdot11)^{\frac{1}{2}}}\right) + \left(\frac{2^{\frac{1}{2}}}{(5\cdot7\cdot11)^{\frac{1}{2}}}\right)\left(\frac{-3(2)^{\frac{1}{2}}}{(5\cdot11)^{\frac{1}{2}}}\right)\right]\left(\frac{1}{2}\right)$$
$$= 1/3(7)^{\frac{1}{2}}.$$

(For the last two examples it is necessary to know that $\Delta^{(3)}_{11} = -1/8$, $\Delta^{(3)}_{13} = (15)^{\frac{1}{2}}/8$. $\Delta^{(f)}$ is tabulated in Ref. 30 for $f = k/2$, $k \leq 4$.)

ACKNOWLEDGMENTS

The authors take pleasure in thanking Professor V. Bargmann for discussions related to the group theory of harmonic oscillators. One of us (BK) is grateful to Professor G. Racah for an enlightening

[31] M. Moshinsky and T. A. Brody, Rev. Mex. Fis. 9, 181 (1960).

conversation, and is indebted to Y. Bein for his help in the numerical calculations.

APPENDIX A

The matrix t of transformation-brackets decomposes within the space of dimension $\binom{E+5}{5}$ into blocks of fixed E *and* J. The dimension of each such block is

$$d(E, J) = \sum_{k=0}^{[\frac{1}{2}E]} d_{[E-k,k]}[U_2]\cdot N_J(E-k, k), \qquad \text{(A1)}$$

where N_J is the multiplicity of D_J, as given in Eq. (24). Evaluation of the sum is straightforward, if tedious. The result is

$$d(E, J) = \frac{1}{2}\left(J + \frac{1+(-1)^{E-J}}{2}\right) \times \left(\frac{E-J+2}{2}\right)\left(\frac{E-J+4}{2}\right). \qquad \text{(A2)}$$

APPENDIX B: THE REDUCED MATRIX ELEMENTS OF b AND a

It is sufficient to consider any one of the components of the tensor $\mathbf{b}$, say $\mathbf{a}_0^*$.

By the correspondence with configurational coordinates, $\mathbf{a}_0^* \leftrightarrow z$, while $|\epsilon lm\rangle$ corresponds to the solid spherical harmonic

$$\mathcal{Y}_{\epsilon lm} \equiv A_{\epsilon lm} r^\epsilon e^{im\phi} P_l^m(\cos\theta), \qquad \text{(B1)}$$

where

$$A_{\epsilon lm} = (-1)^m\left[\frac{(2l+1)(l-m)!}{(\epsilon+l+1)!!(\epsilon-l)!!(l+m)!}\right]^{\frac{1}{2}}. \qquad \text{(B2)}$$

Now

$$z\cdot\mathcal{Y}_{\epsilon lm} = A_{\epsilon lm} r^{\epsilon+1} e^{im\phi}[\cos\theta\cdot P_l^m(\cos\theta)]$$
$$= \frac{A_{\epsilon lm}}{A_{\epsilon+1,l+1,m}}\cdot\frac{(l-m+1)}{(2l+1)}\cdot\mathcal{Y}_{\epsilon+1,l+1,m} + \frac{A_{\epsilon lm}}{A_{\epsilon+1,l-1,m}}\cdot\frac{(l+m)}{(2l+1)}\cdot\mathcal{Y}_{\epsilon+1,l-1,m}, \qquad \text{(B3)}$$

where we have used the recursion relation for Legendre polynomials:

$$(2l+1)\cos\theta\cdot P_l^m(\cos\theta) = (l-m+1)P_{l+1}^m(\cos\theta) + (l+m)P_{l-1}^m(\cos\theta). \qquad \text{(B4)}$$

Inserting the values of $A_{\epsilon lm}$ into (B3) we find that the coefficient of $\mathcal{Y}_{\epsilon+1,l+1,m}$ is

$$(-)^{l-m-1}\begin{pmatrix} l+1 & l & 1 \\ m & -m & 0\end{pmatrix}[(l+1)(\epsilon+l+3)]^{\frac{1}{2}},$$

while that of $\mathcal{Y}_{\epsilon+1,l-1,m}$ is

$$(-)^{l-m}\begin{pmatrix} l & l-1 & 1 \\ m & -m & 0\end{pmatrix}[l(\epsilon-l+2)]^{\frac{1}{2}}.$$

The correspondence $z \leftrightarrow \mathbf{a}_0^*$ then tells us that

$$\langle\epsilon+1, l+1, m|\,\mathbf{a}_0^*|\epsilon, l, m\rangle = (-)^{l-m-1}\begin{pmatrix} l & l+1 & 1 \\ m & -m & 0\end{pmatrix}[(l+1)(\epsilon+l+3)]^{\frac{1}{2}}, \qquad \text{(B5)}$$

or

$$\langle\epsilon+1, l+1\,||\,\mathbf{b}\,||\,\epsilon, l\rangle = +[(l+1)(\epsilon+l+3)]^{\frac{1}{2}}. \qquad \text{(B6)}$$

Similarly,

$$\langle\epsilon+1, l-1\,||\,\mathbf{b}\,||\,\epsilon, l\rangle = -[l(\epsilon-l+2)]^{\frac{1}{2}}. \qquad \text{(B7)}$$

A similar calculation for the tensor $\mathbf{a}$ shows that

$$\langle\epsilon-1, l+1\,||\,\mathbf{a}\,||\,\epsilon, l\rangle = +[(l+1)(\epsilon-l)]^{\frac{1}{2}}, \qquad \text{(B8)}$$

$$\langle\epsilon-1, l-1\,||\,\mathbf{a}\,||\,\epsilon, l\rangle = -[l(\epsilon+l+1)]^{\frac{1}{2}}. \qquad \text{(B9)}$$

APPENDIX C: ON THE COMPUTATION OF $(\epsilon_1 l_1 \epsilon_2 l_2 \mid f\alpha J)$

By its definition, K (and also $R \equiv K\cdot A$) brings $\mathbf{F}_+$ and $\mathbf{F}_-$ to their canonical form (26). Thus we may write

$$[(f+\nu+1)(f-\nu)]^{\frac{1}{2}}(\epsilon_1 l_1 \epsilon_2 l_2 \mid f\alpha J) = \sum_{l_1' l_2'} (\epsilon_1+1, l_1', \epsilon_2-1, l_2' \mid f\alpha J) \times \langle\epsilon_1 l_1 \epsilon_2 l_2 J|\ \sqrt{2}\mathbf{F}_-\ |\epsilon_1+1, l_1', \epsilon_2-1, l_2' J\rangle$$
$$= (-)^{E+J}\sum_{l_1' l_2'} (\epsilon_1+1, l_1', \epsilon_2-1, l_2' \mid f\alpha J) \times \begin{Bmatrix} l_1 & l_2 & J \\ l_2' & l_1' & 1\end{Bmatrix}\langle\epsilon_1 l_1\,||\,\mathbf{a}\,||\,\epsilon_1+1, l_1'\rangle\langle\epsilon_2-1, l_2'\,||\,\mathbf{a}\,||\,\epsilon_2 l_2\rangle. \qquad \text{(C1)}$$

[Here we have made use of the explicit representation of $\mathbf{M}^{12} = (\sqrt{2}\mathbf{F}_-)^*$ given in Eq. (37); $2\nu \equiv \epsilon_1 - \epsilon_2$, $E = \epsilon_1 + \epsilon_2$.]

Equation (C1) may be reinterpreted as a vector equation, if we introduce the notation $\mathbf{R}(EJ\nu;\, f\alpha)$ for the vector whose components are $(\epsilon_1 l_1 \epsilon_2 l_2 \mid f\alpha J)$. The dimension of such a vector is determined by the number of possibilities $(l_1 l_2)$ which are compatible with the values of E, J, ν. In particular, when $\nu = \frac{1}{2}E$ (and therefore, necessarily, $f = \frac{1}{2}E$) we must have: $\epsilon_2 = 0$, $\epsilon_1 = E$, and therefore $l_2 = 0$, $l_1 = J$ (which means that J must have the parity of E). It follows that there is only one possibility in this case: $(l_1 l_2) = (J, 0)$, i.e., the vector $\mathbf{R}(EJ, \frac{1}{2}E; \frac{1}{2}E)$ is of dimension 1, and accordingly $(EJ00 \mid \frac{1}{2}E, J) = 1$.

Introducing this vector notation into (C1), we

obtain a reinterpretation of the equation as a *recursion* relation:

$$[(f+\nu+1)(f-\nu)]^{\frac{1}{2}}\cdot\mathbf{R}(EJ\nu;f\alpha) = (\sqrt{2}\mathbf{F}_-)\cdot\mathbf{R}(EJ,\nu+1;f\alpha), \quad \text{(C2)}$$

which enables us to compute in succession all vectors $\mathbf{R}(EJ\nu;f\alpha)$ for given E, J, $f\alpha$ provided we have a "starting" vector $\mathbf{R}(EJf;f\alpha)$.

As we have just seen, the vector for $f=\frac{1}{2}E$ is known:

$$\mathbf{R}(EJ,\tfrac{1}{2}E;\tfrac{1}{2}E)=(1), \quad \text{(C3)}$$

hence we can compute in succession

$$\mathbf{R}(EJ,\tfrac{1}{2}E-1;\tfrac{1}{2}E)=\frac{1}{E^{\frac{1}{2}}}(\sqrt{2}\mathbf{F}_-)\cdot\mathbf{R}(EJ,\tfrac{1}{2}E;\tfrac{1}{2}E),$$

$$\mathbf{R}(EJ,\tfrac{1}{2}E-2;\tfrac{1}{2}E)=\frac{1}{[2(E-1)]^{\frac{1}{2}}}(\sqrt{2}\mathbf{F}_-)\cdot\mathbf{R}(EJ,\tfrac{1}{2}E-1;\tfrac{1}{2}E),$$

etc.

Having obtained all vectors for which $f=\frac{1}{2}E$, we go on to the family of vectors for which $f=\frac{1}{2}E-1$. To find the starting vector, we make use of the orthogonality properties of the matrix R reinterpreted in terms of the vector notation:

$$\mathbf{R}(EJ\nu;f\alpha)\cdot\mathbf{R}(EJ\nu;f'\alpha')=\delta_{ff'}\delta_{\alpha\alpha'}. \quad \text{(C4)}$$

Equation (C4) enables us to determine

$$\mathbf{R}(EJ,\tfrac{1}{2}E-1;\tfrac{1}{2}E-1)$$

as the vector orthogonal to $\mathbf{R}(EJ,\frac{1}{2}E-1;\frac{1}{2}E)$ (both are two dimensional, and the determination is unique). From this we then derive, for all $|\nu|\leq\frac{1}{2}E-1$, the family of vectors $\mathbf{R}(EJ\nu;\frac{1}{2}E-1)$.

One could proceed further in this way, if it were not for the occurrence of degeneracies. When $E\geq 6$, a degeneracy occurs for $f=\frac{1}{2}E-2$ at various values of J (see table Ref. 13). In those cases where the degeneracy is twofold, there will be two undetermined vectors $\mathbf{R}(EJ,\frac{1}{2}E-2;\frac{1}{2}E-2,\alpha)$, $\alpha=1,2$. One of these is chosen arbitrarily out of the family of vectors orthogonal to *both* $\mathbf{R}(EJ,\frac{1}{2}E-2;\frac{1}{2}E)$ *and* $\mathbf{R}(EJ,\frac{1}{2}E-2;\frac{1}{2}E-1)$. The other vector is then uniquely determined, and both suffice to determine the full set: $\mathbf{R}(EJ\nu;\frac{1}{2}E-2,\alpha)$. When the degeneracy is threefold (this can occur for $E\geq 12$), two arbitrary choices must be made, etc.

APPENDIX D: TABLES OF $(l_1l_2 \mid EJ;\nu \mid f\alpha)$

Rows are labeled by the pair l_1, l_2. Columns are labeled by the pair f, α; α is omitted in cases where there is no degeneracy.

***E* = 3:**

$J=1$

	$\nu=\frac{1}{2}$	
	$\frac{3}{2}$	$\frac{1}{2}$
21	$\frac{+2}{3}$	$\frac{\sqrt{5}}{3}$
01	$\frac{-\sqrt{5}}{3}$	$\frac{2}{3}$

***E* = 4:**

$J=2$

	$\nu=1$	
	2	1
31	$\frac{\sqrt{3}}{\sqrt{2\cdot 5}}$	$\frac{\sqrt{7}}{\sqrt{2\cdot 5}}$
11	$\frac{-\sqrt{7}}{\sqrt{2\cdot 5}}$	$\frac{\sqrt{3}}{\sqrt{2\cdot 5}}$

	$\nu=0$		
	2	1	0
22	$\frac{\sqrt{2}}{3}$	0	$\frac{\sqrt{7}}{3}$
20	$\frac{-\sqrt{7}}{3\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{3}$
02	$\frac{-\sqrt{7}}{3\sqrt{2}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{3}$

$J = 0$

	$\nu = 0$
	2 0
22	$\frac{2}{3}$ $\frac{-\sqrt{5}}{3}$
00	$\frac{\sqrt{5}}{3}$ $\frac{2}{3}$

$E = 5$:

$J = 3$

	$\nu = \frac{3}{2}$	
	$\frac{5}{2}$	$\frac{3}{2}$
41	$\frac{+3\sqrt{3}}{\sqrt{5\cdot 7}}$	$\frac{2\sqrt{2}}{\sqrt{5\cdot 7}}$
21	$\frac{-2\sqrt{2}}{\sqrt{5\cdot 7}}$	$\frac{3\sqrt{3}}{\sqrt{5\cdot 7}}$

	$\nu = \frac{1}{2}$		
	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$
32	$\frac{-2}{5}$	$\frac{1}{5}$	$\frac{2}{\sqrt{5}}$
30	$\frac{\sqrt{3}}{\sqrt{2\cdot 5}}$	$\frac{-2\sqrt{2}}{\sqrt{3\cdot 5}}$	$\frac{1}{\sqrt{2\cdot 3}}$
12	$\frac{3\sqrt{3}}{5\sqrt{2}}$	$\frac{4\sqrt{2}}{5\sqrt{3}}$	$\frac{1}{\sqrt{2\cdot 3\cdot 5}}$

$J = 2$

	$\nu = \frac{1}{2}$	
	$\frac{3}{2}$	$\frac{1}{2}$
32	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$
12	$\frac{-\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$

$J = 1$

	$\nu = \frac{3}{2}$	
	$\frac{5}{2}$	$\frac{3}{2}$
21	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$
01	$\frac{-2\sqrt{2}}{\sqrt{3\cdot 5}}$	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$

	$\nu = \frac{1}{2}$		
	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{1}{2}$
32	$\frac{2}{5}$	$\frac{2\sqrt{7}}{5\sqrt{3}}$	$\frac{-\sqrt{7}}{\sqrt{3\cdot 5}}$
12	$\frac{-2\sqrt{7}}{5\sqrt{3}}$	$\frac{11}{3\cdot 5}$	$\frac{2}{3\sqrt{5}}$
10	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{2}{3\sqrt{5}}$	$\frac{2}{3}$

E = 6:

$J = 4$

	$\nu = 2$	
	3	2
51	$\frac{\sqrt{5}}{3\sqrt{3}}$	$\frac{\sqrt{2\cdot 11}}{3\sqrt{3}}$
31	$\frac{-\sqrt{2\cdot 11}}{3\sqrt{3}}$	$\frac{5}{3\sqrt{3}}$

	$\nu = 1$		
	3	2	1
42	$\frac{-2\sqrt{2}}{3\sqrt{7}}$	$\frac{\sqrt{11}}{3\sqrt{2\cdot 7}}$	$\frac{\sqrt{11}}{\sqrt{2\cdot 7}}$
40	$\frac{\sqrt{11}}{3\sqrt{5}}$	$\frac{-\sqrt{5}}{3}$	$\frac{1}{\sqrt{5}}$
22	$\frac{\sqrt{2\cdot 11}}{\sqrt{5\cdot 7}}$	$\frac{\sqrt{5}}{\sqrt{2\cdot 7}}$	$\frac{1}{\sqrt{2\cdot 5\cdot 7}}$

	$\nu = 0$		
	3	2	1
33	$\frac{-\sqrt{3}}{5}$	0	$\frac{\sqrt{2\cdot 11}}{5}$
31	$\frac{\sqrt{11}}{5}$	$\frac{-1}{\sqrt{2}}$	$\frac{\sqrt{3}}{5\sqrt{2}}$
13	$\frac{\sqrt{11}}{5}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{5\sqrt{2}}$

$J = 3$

$\nu = 1$

	2	1
42	$\frac{3}{\sqrt{2\cdot 7}}$	$\frac{\sqrt{5}}{\sqrt{2\cdot 7}}$
22	$\frac{-\sqrt{5}}{\sqrt{2\cdot 7}}$	$\frac{3}{\sqrt{2\cdot 7}}$

$\nu = 0$

	2	1	0
33	$\frac{-\sqrt{2}}{\sqrt{5}}$	0	$\frac{\sqrt{3}}{\sqrt{5}}$
31	$\frac{\sqrt{3}}{\sqrt{2\cdot 5}}$	$\frac{-1}{\sqrt{2}}$	$\frac{1}{\sqrt{5}}$
13	$\frac{\sqrt{3}}{\sqrt{2\cdot 5}}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{5}}$

$J = 2$

$\nu = 2$

	3	2
31	$\frac{-\sqrt{2}}{\sqrt{5}}$	$\frac{-\sqrt{3}}{\sqrt{5}}$
11	$\frac{\sqrt{3}}{\sqrt{5}}$	$\frac{-\sqrt{2}}{\sqrt{5}}$

$\nu = 1$

	3	2	1, 1	1, 2
42	$\frac{4}{5\sqrt{7}}$	$\frac{2\sqrt{3}}{\sqrt{5\cdot 7}}$	0	$\frac{3\sqrt{11}}{5\sqrt{7}}$
22	$\frac{-2\sqrt{2}}{\sqrt{5\cdot 7}}$	$\frac{1}{\sqrt{2\cdot 3\cdot 7}}$	$\frac{7}{\sqrt{2\cdot 3\cdot 11}}$	$\frac{\sqrt{2}}{\sqrt{5\cdot 7\cdot 11}}$
20	$\frac{\sqrt{2}}{\sqrt{5}}$	$\frac{1}{\sqrt{2\cdot 3}}$	$\frac{\sqrt{7}}{\sqrt{2\cdot 3\cdot 11}}$	$\frac{-3\sqrt{2}}{\sqrt{5\cdot 11}}$
02	$\frac{\sqrt{7}}{5}$	$\frac{-\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{\sqrt{5}}{\sqrt{3\cdot 11}}$	$\frac{2\sqrt{7}}{5\sqrt{11}}$

$\nu = 0$

	3	2	1, 1	1, 2
33	$\frac{2\sqrt{2}}{5\sqrt{5}}$	0	$\frac{-6\sqrt{7}}{5\sqrt{11}}$	$\frac{3\sqrt{3}}{5\sqrt{5\cdot 11}}$
31	$\frac{-3\sqrt{3}}{5\sqrt{5}}$	$\frac{-1}{\sqrt{2}}$	$\frac{-\sqrt{3\cdot 7}}{5\sqrt{2\cdot 11}}$	$\frac{-13\sqrt{2}}{5\sqrt{5\cdot 11}}$
13	$\frac{-3\sqrt{3}}{5\sqrt{5}}$	$\frac{1}{\sqrt{2}}$	$\frac{-\sqrt{3\cdot 7}}{5\sqrt{2\cdot 11}}$	$\frac{-13\sqrt{2}}{5\sqrt{5\cdot 11}}$
11	$\frac{3\sqrt{7}}{5\sqrt{5}}$	0	$\frac{\sqrt{2}}{5\sqrt{11}}$	$\frac{-4\sqrt{2\cdot 3\cdot 7}}{5\sqrt{5\cdot 11}}$

$J = 1$

$\nu = 0$

	2	0
33	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$
11	$\frac{-\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$

$J = 0$

$\nu = 1$

	3	1
22	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$	$\frac{\sqrt{7}}{\sqrt{3\cdot 5}}$
00	$\frac{-\sqrt{7}}{\sqrt{3\cdot 5}}$	$\frac{2\sqrt{2}}{\sqrt{3\cdot 5}}$

$\nu = 0$

	3	1
33	$\frac{2}{5}$	$\frac{\sqrt{3\cdot 7}}{5}$
11	$\frac{-\sqrt{3\cdot 7}}{5}$	$\frac{2}{5}$

Existence of Proper Modes of Helicon Oscillations*

C. R. LEGÉNDY †
Laboratory of Atomic and Solid State Physics, Department of Physics, Cornell University, Ithaca, New York
(Received 28 April 1964; final manuscript received 28 July 1964)

In this paper it is shown that the class of electromagnetic problems for which the operator $i(\partial/\partial t)$ (where t denotes time) is self-adjoint extends beyond problems involving only insulators and perfect conductors. The class includes problems in which the perfect conductor is generalized to a medium with antisymmetric resistivity tensor. The latter medium approximates media in which *helicon* waves can propagate. Helicon waves are known to propagate in good conductors in a strong magnetic field $\mathbf{B}_0$; it will be found that two necessary conditions for self-adjointness of the operator $i(\partial/\partial t)$ are that the sample carrying helicons must not have a finite portion parallel to $\mathbf{B}_0$, and it must be surrounded by a reflecting surface that prevents energy from escaping.

1. INTRODUCTION

HELICONS are circularly polarized electromagnetic waves, which can propagate almost without attenuation inside a solid of high conductivity, permeated by a strong magnetic field. (The condition for unattenuated propagation is that $\omega_c\tau \gg 1$, where ω_c is the cyclotron frequency of the current carriers and τ is the relaxation time.) In atmospheric physics these waves are well known as *radio whistlers*.[1] The term "helicon," which is presently accepted to denote whistlers in the context of solid state physics, is due to Aigrain,[2] who first proposed achievable experiments to detect them in solids.[3] Observation of helicons was first announced by Bowers, Legéndy, and Rose,[4] at frequencies of order 10 cps. In recent years, helicons have been observed and studied by a number of authors; a short survey of the literature on the subject is given in another article by the present author.[5]

In this article we deal with the abstract boundary-value problem presented by helicons under three idealizing assumptions: (i) The sample carrying helicons has negligible resistivity (but is not a superconductor, so that it does not exclude magnetic flux). (ii) The resistivity tensor is the same as would be for uniform dc fields. (That is, nonlocal effects in space and time are ignored.) (iii) The constitutive equation is linearized. Assumptions (i) to (iii) amount to assuming that (owing to the presence of the external magnetic field mentioned above) the sample carrying helicons is characterized by a fixed, antisymmetric resistivity tensor.

The arrangement we consider consists of a sample (region M_1 in Fig. 1) and a closed reflecting surface (surface S in Fig. 1) surrounding it, to stop any energy from escaping; between the sample and the reflecting surface there is a nonconducting region (region M_2 in Fig. 1). The net charge on the sample, and the charge density in the nonconducting region are assumed to be zero. The sample is required to have smooth boundaries; a further requirement on the boundaries is that they have no finite portion parallel to the external magnetic field.

Under the above assumptions, the operator $-i(\partial/\partial t)$, operating on electromagnetic fields, is shown to be self-adjoint. In the proof it is not necessary to assume that the dielectric constant, magnetic susceptibility, and Hall coefficient are constants throughout the regions of interest; the external magnetic field is not required to be uniform, nor the displacement current negligible. Aside from the restrictions already stated, there is no restriction on the shape of the sample; no use is made of any assumptions to the effect that the sample is connected or simply connected.

The purpose of making the seemingly arbitrary restriction, that the boundary shall have no finite portion tangential to the external magnetic field, is to avoid a certain surface mode of energy absorption[5,6] confined to surfaces tangential to the external field. If the resistivity is assumed to be finite, and is then allowed to tend to zero, the electric currents in this mode increase without bound, and the Ohmic loss does not tend to zero. Therefore, in samples with such surfaces, any free oscillations are bound to be attenuated, the operator $-i(\partial/\partial t)$ cannot have real eigenvalues, and cannot be self-adjoint. (If anoma-

* This work was supported partly by the U. S. Atomic Energy Commission.
† Present address: United Aircraft Corporation Research Laboratories, East Hartford, Connecticut.
[1] L. R. O. Storey, Phil. Trans. Roy. Soc. London **A246**, 113 (1953).
[2] P. Aigrain, Proc. Intern. Conf. Semicond. Phys., Prague, **1960**, 224 (1961).
[3] See, however, O. V. Konstantinov and V. I. Perel', Zh. Eksperim. i Teor. Fiz. **38**, 161 (1960) [English transl.: Soviet Phys.—JETP **11**, 117 (1960)].
[4] R. Bowers, C. R. Legéndy, and F. E. Rose, Phys. Rev. Letters **7**, 339 (1961).
[5] C. R. Legéndy, Phys. Rev. **135**, A 1713 (1964).
[6] J. M. Goodman and C. R. Legéndy (unpublished).

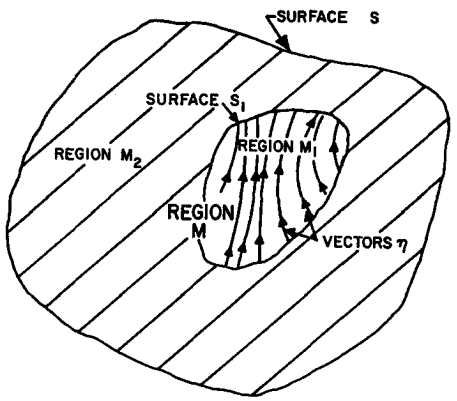

FIG. 1. Notation for Sec. 2. Region M consists of regions M_1 and M_2. M_1 is the sample carrying helicons.

lous skin effect is taken into consideration, the surface loss is found to disappear for low enough resistivity.[5])

Electromagnetic fields will be represented as vectors [see (2.1)] in an abstract vector space with a scalar product [see (2.2)], as was done by Marcuwitz[7] and Wilcox.[8] The present formulation slightly differs from theirs, in that, throughout Sec. 2 we deal only with *instantaneous* electromagnetic fields, and do not even implicitly assume any time dependence [such as exp $(i\omega t)$]. The operator $\hat{T} = -i(\partial/\partial t)$ is rewritten, using two of Maxwell's equations, so that it operates on the spatial coordinates only [see Eq. (2.3)]. None of Maxwell's equations are explicitly used in defining the allowable electromagnetic fields; instead, it is required that the instantaneous field be in the *range* of the operator $\hat{T}$. The two time-independent Maxwell's equations are then automatically satisfied (because of the previously stated assumptions of charge neutrality). After the self-adjointness of $\hat{T}$ is established (and therefore, the existence of a complete set of eigenfunctions is shown), time dependence is introduced by means of the operator exp $(it\hat{T})$. The resulting time-dependent functions automatically satisfy the two time-dependent Maxwell's equations.

In Sec. 2 we give the mathematical definitions and proofs, then make the necessary physical connections in Sec. 3.

2. DEFINITIONS, THEOREM, AND PROOF

Definitions. Let S be a smooth, simply connected, closed surface; M the region composed of all points (x, y, z), inside and on S; S_1 a smooth, closed surface entirely surrounded by S (and not touching S); M_1 the region composed of points inside and on S_1, and let M_2 be the rest of M (see Fig. 1). Define the vector $\mathfrak{y}$ in the region M_1, as an everywhere-bounded, real, and well-behaved function of x, y, z, with the further restriction that on the surface S_1, the scalar product $\mathfrak{y}\cdot\mathbf{n} \neq 0$ (where $\mathbf{n}$ is a normal vector to S_1), except at most on some isolated points or lines. Let $\mathbf{E}(x, y, z)$ and $\mathbf{H}(x, y, z)$ be (possibly complex) vector functions, both defined throughout M such that $\nabla\times\mathbf{E}$ and $\nabla\times\mathbf{H}$ are well defined; let $\epsilon(x, y, z)$ and $\mu(x, y, z)$ be everywhere positive, real and bounded functions, defined throughout M.

Form the six-component vectors

$$\mathsf{F} = \begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix}; \tag{2.1}$$

for any two such vectors, define the scalar product:

$$(\mathsf{F}_1, \mathsf{F}_2) = \frac{1}{2}\int_{\text{over } M} (\epsilon\mathbf{E}_1^*\cdot\mathbf{E}_2 + \mu\mathbf{H}_1^*\cdot\mathbf{H}_2)\, dV, \tag{2.2}$$

and on all such vectors define the operator $\hat{T}$ as follows:

$$\hat{T}\mathsf{F} = \begin{pmatrix}-i\epsilon^{-1}(\nabla\times\mathbf{H} - \mathbf{j})\\ i\mu^{-1}\nabla\times\mathbf{E}\end{pmatrix}, \tag{2.3a}$$

where

$$\mathbf{j} = \begin{cases}\eta^{-2}[\mathfrak{y}(\mathfrak{y}\cdot\nabla\times\mathbf{H}) - \mathbf{E}\times\mathfrak{y}] & \text{in } M_1\\ 0 & \text{in } M_2\end{cases} \tag{2.3b}$$

$$\mathfrak{y} = (\mathfrak{y}_x^2 + \mathfrak{y}_y^2 + \mathfrak{y}_z^2)^{\frac{1}{2}}.$$

Let the domain $\mathfrak{D}$ of $\hat{T}$ be the set of F satisfying the following boundary conditions (almost) everywhere on S and S_1 respectively:

$$(\alpha\mathbf{E} + \beta\mathbf{H})\times\mathbf{n} = 0 \quad \text{on} \quad S \tag{2.4a}$$

$$\left.\begin{matrix}\mathbf{E}\times\mathbf{n}\\ \mathbf{H}\times\mathbf{n}\end{matrix}\right\} \text{ continuous across } S_1 \tag{2.4b}$$

where α and β are fixed, real, scalar functions of the position on S; both differentiable once, and at least one of them differing from zero at each point on S.

Let L^2 be the space of all vectors F for which $(\mathsf{F}, \mathsf{F}) < \infty$. Let $\mathcal{P}$ be the closure of the range of $\hat{T}$. One can show at once that $\mathcal{P}$ is the set of vectors F for which

$$\begin{aligned}&\nabla\cdot\mu\mathbf{H} = 0 \quad \text{throughout } M,\\ &\mathbf{E}\cdot\mathfrak{y} = 0 \quad \text{in } M_1,\\ &\nabla\cdot\epsilon\mathbf{E} = 0 \quad \text{in } M_2,\\ &\oint_{S_2}\epsilon\mathbf{E}\cdot d\mathbf{S} = 0,\end{aligned} \tag{2.5}$$

[7] N. Marcuwitz, *Electromagnetic Waves*, Proceedings of a Symposium Conducted by the Mathematics Research Center, U. S. Army, at the University of Wisconsin, Madison, on 10–12 April 1961; edited by R. E. Langer (The University of Wisconsin Press, Madison, 1962), p. 109.

[8] C. H. Wilcox, Ref. 7, p. 65.

where S_2 is any closed surface in the region M_2, enclosing M_1.

Define the Hilbert space $\mathfrak{F}$ as the set of all L^2 vectors in $\mathcal{P}$.

Lemma. Defined on the domain $\mathfrak{D}$, the operator $\hat{T}$ is symmetric, i.e., if F_1 and F_2 are both in $\mathfrak{D}$, then

$$(\mathsf{F}_1, \hat{T}\mathsf{F}_2) = (\hat{T}\mathsf{F}_1, \mathsf{F}_2).$$

Proof. It is enough to show that for all F in $\mathfrak{D}$,

$$\Delta = \frac{i}{2}[(\mathsf{F}, \hat{T}\mathsf{F}) - (\hat{T}\mathsf{F}, \mathsf{F})] = \mathrm{Re}\,(\mathsf{F}, i\hat{T}\mathsf{F}) = 0. \quad (2.6)$$

For then the substitution $\mathsf{F} = \mathsf{F}_1 + i\mathsf{F}_2$ in (2.6) shows that the real part of $(\mathsf{F}_1, \hat{T}\mathsf{F}_2) - (\hat{T}\mathsf{F}_1, \mathsf{F}_2)$ vanishes, and the substitution $\mathsf{F} = \mathsf{F}_1 + \mathsf{F}_2$ shows that the imaginary part of the same expression vanishes for all F_1 and F_2 in $\mathfrak{D}$.

To prove (2.6), integrate (2.2) by parts, thus splitting up the integral into surface and volume integrals:

$$\Delta = \int_{M_1} A + \int_{M_2} A + \oint_{S_1^{(1)}} B - \oint_{S_1^{(2)}} B + \oint_S B$$

where

$$A = \tfrac{1}{2}\,\mathrm{Re}\,(\mathbf{E}^* \cdot \mathbf{j})\, dV$$

$$B = -\tfrac{1}{2}\,\mathrm{Re}\,(\mathbf{E}^* \times \mathbf{H}) \cdot d\mathbf{S}$$

and $S_1^{(1)}$ and $S_1^{(2)}$ refer to integrals over S_1 as the surface is approached from region M_1 and region M_2 respectively. From the definition of $\mathbf{j}$ in (2.3) one can see at once that A identically vanishes in M_1 as well as M_2, thus, the first two integrals vanish. (It is at this point that the antisymmetry of the resistivity tensor was exploited.) Because of the boundary condition on S_1, (2.4b), the two surface integrals over S_1 cancel. Finally, from the boundary condition over S, (2.4a), the last integral vanishes, which completes the proof of the *lemma*.

Theorem. In the space $\mathfrak{F}$ and on the domain $\mathfrak{D}$, the operator $\hat{T}$ is self-adjoint.

Proof. In view of the *lemma* just proved, it is enough to show that there exist a set of vectors $\{\mathsf{F}_i\}$ that are in $\mathfrak{F}$ as well as $\mathfrak{D}$, and are such that if for some F,

$$(\mathsf{F}, \hat{T}\mathsf{F}_i) = (\hat{T}\mathsf{F}, \mathsf{F}_i) \quad (2.7)$$

for all F_i in the set, then F is necessarily in the domain $\mathfrak{D}$.

To show this, first form two arbitrary, complete sets of everywhere bounded and differentiable vector functions $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$, defined only on the points of S_1, everywhere tangential to S_1 and $\mathbf{b}_i$ identically vanishing wherever $\mathfrak{y}\cdot\mathbf{n} = 0$. Completeness is meant in the sense that, if the vector functions $\mathbf{P}(x, y, z)$ and $\mathbf{Q}(x, y, z)$, defined throughout M are such that

$$\oint_{S_1} (\mathbf{P}\cdot\mathbf{a}_i)\, dS = 0, \quad \oint_{S_1} (\mathbf{Q}\cdot\mathbf{b}_i)\, dS = 0 \quad (2.8)$$

for all i, then (almost) everywhere on S_1,

$$\mathbf{P}\times\mathbf{n} = 0, \qquad \mathbf{Q}\times\mathbf{n} = 0. \quad (2.9)$$

(The latter of these can only be true because, by definition, $\mathfrak{y}\cdot\mathbf{n} \neq 0$ almost everywhere on S_1.) From a_i and b_i construct the six-component vectors

$$\mathsf{G}_i = \begin{pmatrix} 0 \\ \mathbf{p}_i \end{pmatrix}, \qquad \mathsf{K}_i = \begin{pmatrix} \mathbf{q}_i \\ 0 \end{pmatrix}$$

such that $\mathbf{p}_i$ and $\mathbf{q}_i$ identically vanish everywhere on and near S, they are bounded and well-behaved everywhere, and

$$\mathbf{p}_i \times \mathbf{n} = \mathbf{a}_i; \qquad \mathbf{q}_i \times \mathbf{n} = \mathbf{b}_i;$$

on S_1. (Note that, necessarily, G_i and K_i are in the domain $\mathfrak{D}$.) If $\mathbf{p}_i$ and $\mathbf{q}_i$ satisfy the above requirements, G_i and K_i are in the space $\mathfrak{F}$ if, and only if

$$\begin{aligned} \nabla\cdot\mu\mathbf{p}_i &= 0 \quad \text{throught } M, \\ \mathbf{q}_i\cdot\mathfrak{y} &= 0 \quad \text{in } M_1, \\ \nabla\cdot\epsilon\mathbf{q}_i &= 0 \quad \text{in } M_2, \\ \oint_{S_2} \epsilon\mathbf{q}_i\cdot d\mathbf{S} &= 0, \end{aligned} \quad (2.10)$$

where S_2 is any surface enclosing M_1 [Cf. Eq. (2.5)]. Under the assumptions made earlier, the conditions listed [including (2.10)] are not very restrictive, and there is a wide choice of $\mathbf{p}_i$ and $\mathbf{q}_i$ satisfying them. [However, if for any i we had $\mathbf{b}_i \times \mathfrak{y} \neq 0$ at a point where $\mathfrak{y}\cdot\mathbf{n} = 0$, there would exist no $\mathbf{q}_i$ satisfying the second of Eqs. (2.10).]

Substitute for F_i in Eq. (2.7) the vectors G_i and K_i; it follows at once from (2.8) and (2.9) that F in Eq. (2.7) satisfies the boundary condition on S_1.

To carry out the analogous proof for S, form a complete set of vector functions $\mathbf{c}_i$, defined on the points of S, with properties similar to the properties of $\mathbf{a}_i$ previously defined on S_1. Then form

$$\mathsf{L}_i = \begin{pmatrix} \mathbf{e}_i \\ \mathbf{h}_i \end{pmatrix}$$

such that $\mathbf{e}_i$ and $\mathbf{h}_i$ identically vanish in and near M_1, they are bounded and well behaved everywhere, and

$$\mathbf{e}_i \times \mathbf{n} = \beta\mathbf{c}_i; \; \mathbf{h}_i \times \mathbf{n} = -\alpha\mathbf{c}_i,$$

on S [where α and β are as defined in Eq. (2.4a)]. The vector L_i thus defined is necessarily in the domain $\mathfrak{D}$. It is easy to see that if L_i satisfy the above requirements, they are in $\mathfrak{F}$ if, and only if $\nabla\cdot\epsilon\mathbf{e}_i = 0$ and $\nabla\cdot\mu\mathbf{h}_i = 0$ everywhere. Again, the conditions are not very restrictive, and there is a wide choice of L_i satisfying them.

Substituting the vectors L_i for F_i in Eq. (2.7), and using the completeness of the set $\{c_i\}$ through observations such as (2.8) and (2.9), it is easily shown that F in Eq. (2.7) satisfies the boundary condition on S, and the proof is complete.

3. RESULTS AND DISCUSSION

The vectors $\mathbf{E}$ and $\mathbf{H}$ in (2.1) are recognized as the electric and magnetic field; the scalar product of a vector by itself, (F, F), is recognized as the energy in the electromagnetic field F. As was indicated in the Introduction, the operator $\hat{T}$, defined in (2.3) is identified at once as $-i(\partial/\partial t)$ [$\mathbf{j}$ in (2.3) standing for electric current density]. The definition of $\mathbf{j}$ in the region M_1, the region carrying helicons, is so designed as to make $\mathbf{E} = \mathbf{j}\times\mathfrak{y}$, and therefore $\mathfrak{y}\cdot\mathbf{E} = 0$ and $\mathfrak{y}\cdot(\partial\mathbf{E}/\partial t) = 0$ at all times. Physically, $\mathfrak{y} = -R\mathbf{B}_0$, where R is the Hall coefficient and $\mathbf{B}_0$ is the steady, external magnetic field. The boundary condition (2.4a) forces Poynting's vector $\mathbf{E}\times\mathbf{H}$ to be tangential to S, hence the surface S reflects all radiation coming onto it.

Denote the integrand in (2.2) as $\mathsf{F}^*\mathsf{F}$. Then

$$2\,\mathrm{Re}\,(\mathsf{F}, i\hat{T}\mathsf{F}) = \frac{1}{2}\int_M \left(\mathsf{F}^*\frac{\partial\mathsf{F}}{\partial t} + \frac{\partial\mathsf{F}^*}{\partial t}\mathsf{F}\right) = \frac{\partial}{\partial t}(\mathsf{F}, \mathsf{F}). \tag{3.1}$$

Comparing (3.1) with (2.6), it is found that the operator $\hat{T}$ is symmetric if and only if the system conserves energy. We recall from the proof of the *lemma* that symmetry hinged upon three facts: (1) in region M_1 the resistivity tensor is antisymmetric, therefore the current and electric field are perpendicular, and there is no Ohmic loss; (2) by (2.4b), the normal component of Poynting's vector is continuous across S_1; and (3) by (2.4a), the normal component of Poynting's vector is zero over S. The mathematical fact that a symmetric operator has real eigenvalues is translated into the statement that a system conserving energy cannot execute damped or growing oscillations. The mathematical fact that a symmetric operator has orthogonal eigenvectors corresponds to the statement that if the electromagnetic system conserves energy, its total energy is the sum of the energies in the individual modes.

It follows from a remark made below (2.10) that if over a finite portion of the sample's surface $\mathfrak{y}\cdot\mathbf{n} = 0$, it is not possible to establish that F in (2.7) satisfies the boundary condition (2.4b). Indeed, physical considerations[6] show that in that case, under the assumptions on which the present formulation is built [namely, Assumptions (i) to (iii) in the Introduction], the boundary of the sample absorbs energy, and the proof cannot be completed. (However, there is no difficulty in carrying out the proof if the surfaces in question are appropriately tilted or are made slightly "wavy"; for the purposes of the proof it does not matter how slight the distortion is. The artificially introduced roughness of the surface may be thought of as "simulating" anomalous skin effect, in that it eliminates surface loss for low enough resistivities. For this simulation to fit the physical situation best, the depth of roughness must be of the same order of magnitude as the cyclotron radius.)

It is proved in functional analysis[9] that if an operator $\hat{T}$ is self-adjoint in a Hilbert space $\mathfrak{F}$, then the equation

$$\hat{T}\mathsf{F} = \omega\mathsf{F} \tag{3.3}$$

possesses a set of eigenfunctions F which span all of $\mathfrak{F}$. The eigenfunctions are orthogonal and the eigenvalues ω are real. A glance at the definition (2.3) of $\hat{T}$ shows that the two time-dependent Maxwell's equations can be compressed into the form

$$-i\,\partial\mathsf{F}(t)/\partial t = \hat{T}\mathsf{F}(t). \tag{3.4}$$

It follows that once the self-adjointness of $\hat{T}$ is established, it is possible to construct a time-dependent field $\mathsf{F}(t)$ from any instantaneous field F in $\mathfrak{F}$ as follows:

$$\mathsf{F}(t) = e^{it\hat{T}}\mathsf{F}. \tag{3.5}$$

The field (3.5) thus constructed satisfies Maxwell's equations (3.4). The sequence of expressions (3.4), (3.3), (3.5) resembles the sequence of expressions encountered in connection with Schrödinger's equation, with similar causal relations between the successive forms.

The self-adjointness of $\hat{T}$ implies that Eq. (3.5) is meaningful, but it does not imply that the eigenfunctions of (3.3) have finite energy, i.e., that the eigenfrequencies ω form a discrete set. It is hoped that in the near future some author will show that $\hat{T}$ has a unique inverse, and that the inverse is com-

[9] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955).

pletely continuous. This, together with the symmetry property proved in the *lemma,* would imply everything implied by self-adjointness, and would also imply discreteness and square integrability.

We remark that the proofs of Sec. 2 do not make use of the fact that the space $\mathfrak{F}$ is restricted to the range of $\hat{T}$. Both proofs can be repeated without difficulty if the first, third, and fourth of Eqs. (2.5) are dropped, as long as the second is retained and the domain $\mathfrak{D}$ is defined by (2.4). Of course, the results are then not physically meaningful. Also, if $\hat{T}$ has a larger domain than range, it cannot possibly turn out to have an inverse, as was suggested in the previous paragraph.

Eqs. (2.5) can be compressed into the statement that F must be in the closure of the range of $\hat{T}$. The physical interpretation of these equations is clear. The first one is Maxwell's equation; the third is also Maxwell's equation, assuming that region M_2 contains no free charges; the fourth requires that there be no net charge on the sample; the second, combined with (2.3b) means that the resistivity tensor is antisymmetric.

It is a feature of the present formulation that all field equations, more precisely, the four Maxwell's equations and the constitutive equation, are introduced into the problem merely through the definition of a single operator.

In closing, we wish to comment on the reflecting surface S. The operator $\hat{T}^2$ has positive eigenvalues, therefore, if φ is any vector in $\mathfrak{F}$, the quantity

$$(\varphi, \hat{T}^2\varphi)/(\varphi,\varphi) \tag{3.6}$$

is larger than the square of the smallest eigenfrequency. Thus, to make a crude estimate, let

$$\varphi = \begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix}, \qquad \mathbf{E} = (\nabla\times\mathbf{H})\times\mathfrak{y}, \qquad \mathbf{H} = \mu_0^{-1}\nabla\times\mathbf{A},$$

and

$$A_x = A_y = 0,$$

$$A_z = \begin{cases} \left(\cos\frac{\pi x}{l}\right)^4\left(\cos\frac{\pi y}{l}\right)^4\left(\cos\frac{\pi z}{l}\right)^4 \\ \qquad\qquad \text{for} \quad |x| \le \frac{l}{2}, \quad |y| \le \frac{l}{2}, \quad |z| \le \frac{l}{2} \\ 0 \quad \text{otherwise;} \end{cases}$$

choose the origin of the coordinate system in such a way that the cubical region inside which $\varphi \not\equiv 0$ be in the interior of the sample, and choose the edge of the cube, l, to be as large as possible. Substitute the the resulting field φ into (3.6), and assume, for simplicity, that $\epsilon \equiv \epsilon_0$, $\mu \equiv \mu_0$ and $\mathfrak{y}$ is uniform. Then,

$$[(\varphi, \hat{T}^2\varphi)/(\varphi,\varphi)]^{\frac{1}{2}} \sim 14(2\pi/l)^2\mathfrak{y}\mu^{-1}.$$

Thus, the smallest eigenfrequency is necessarily smaller than the latter quantity. Since we chose φ such that $\varphi \equiv 0$ outside the sample, the above estimate only depends on the sample's size, and not on the dimensions of the reflecting surface. The estimate shows that, if the wavelength in free space corresponding to the lowest mode is denoted by λ_0, then, to order of magnitude, $\lambda_0/l \sim (l/2\pi)(\mu o/\epsilon_0)^{\frac{1}{2}}\mathfrak{y}^{-1}$. In the physical situation of Ref. 4 (but not in the situation of Ref. 2), $(l/2\pi)(\mu_0/\epsilon_0)^{\frac{1}{2}}\mathfrak{y}^{-1} \sim 10^8$, i.e., independently of the size of the reflecting shield, the lowest modes can be considered quasistatic (i.e., of essentially infinite vacuum wavelength). To estimate the rate at which energy would leave the region M_1 in the absence of the reflecting surface, consider the fields due to the currents and charges on the sample alone; neglect all but the magnetic dipole radiation, and let the shield be a sphere of radius λ_0. The ratio of the energy crossing the shield in one cycle to the energy inside the shield is then found to be of order $(l/\lambda_0)^3 \sim 10^{-24}$.

For the higher modes the rate of radiation is higher. However, it can be shown that if we formally let the speed of light outside the sample tend to infinity, the set of almost unattenuated modes can be extended to an arbitrarily large part of the spectrum.

ACKNOWLEDGMENTS

The author wishes to express his gratitude toward Professor W. H. J. Fuchs, for his extremely kind and patient guidance through the evolution of this paper. Thanks also go to Mr. J. M. Goodman for a number of helpful suggestions that greatly improved the presentation.

Exact Statistical Thermodynamics of Gravitational Interactions in One and Two Dimensions*

ALLAN M. SALZBERG
School of Chemistry, University of Minnesota, Minneapolis, Minnesota†
(Received 4 August 1964)

The exact equilibrium statistical mechanics of one- and two-dimensional gases, in which the particles interact through gravitational forces, is obtained. It is found that these systems are characterized by nonextensive thermodynamics leading to behavior somewhat reminiscent of the formation of a star from interstellar dust.

1. INTRODUCTION

THE object of this paper is to extend the results of recent studies of the statistical thermodynamics of one- and two-dimensional plasmas[1-6] to analogous systems in which the electrostatic interactions are replaced by gravitational forces. One interesting complication which arises in these gases is the nonextensive nature of the thermodynamic functions.

2. ONE-DIMENSIONAL MASSES

This section deals with a one-dimensional gas of N particles of length d which interact through long-range gravitational forces. The total potential energy U is a sum of two terms U' and U'' where

$$U' = \sum_{i<j} U'_{ij} \tag{2.1}$$

$$U'_{ij} = \infty \quad |X_i - X_j| < d,$$
$$= 0 \quad |X_i - X_j| > d.$$

X_i being the position of the ith particle from the left. The long-range potential U'' is obtained by solving the one-dimensional Poisson equation,[4] obtaining

$$U'' = \sum_{i<j} gm^2 |X_i - X_j|, \tag{2.2}$$

where g is the one-dimensional gravitational constant and m is the mass of each particle. Since all the masses are identical, Eq. (2.2) can be rewritten as

$$U'' = \sum_{i=1}^{N-1} gm^2 n_i |X_{i+1} - X_i|, \tag{2.3}$$
$$n_i = i(N - i).$$

Due to the indistinguishability of the particles, there exists one and only one discernible ordering of the particles; hence, the constant-volume partition function takes on the form

$$Q = \int_0^L \cdots \int_0^{X_3} \exp(-U/kT)\, dX_1 \cdots dX_N, \tag{2.4}$$

where L is the total length of the system. In order to eliminate the constraints $X_i \leq X_{i+1} \leq L$, a Laplace transformation[4] is performed to obtain the constant-pressure partition function $\tilde{Q}$:

$$\tilde{Q} = \int_0^\infty \exp(-PL/kT)Q(L)\, dL$$
$$= \int_0^\infty \cdots \int_0^\infty \exp\left[-\left(U' + \sum_{i=0}^{N} (gm^2 n_i + P)y_i\right)\right]/kT\, dy_0 \cdots dy_N, \tag{2.5}$$
$$y_0 = X_1,$$
$$y_i = X_{i+1} - X_i,$$
$$y_N = L - \sum_{i=0}^{N-1} y_i.$$

As the only effect of U' is to restrict the domain of integration,

$$\tilde{Q} = \left(\frac{kT}{P}\right)^{N+1} \prod_{i=0}^{N} \left(1 + \frac{gm^2 n_i}{P}\right)^{-1} \times \exp\left[-\left(\frac{d(P + gm^2 n_i)}{kT}\right)\right]. \tag{2.6}$$

Since $\tilde{Q}$ is directly related to the Gibbs free energy,[7] the thermodynamics of the ensemble can now be obtained:

$$G_e = -kT \ln \tilde{Q}, \tag{2.7}$$
$$= (N + 1)kT \ln (P/kT) + \sum_{i=0}^{N} \left[d(P + gm^2 n_i) + kT \ln\left(1 + \frac{gm^2 n_i}{P}\right)\right],$$

* From the thesis submitted by A. M. Salzberg to the Department of Chemistry, University of Minnesota, Minneapolis, Minnesota in partial fulfillment of the requirements for the Ph.D. degree.

† Present address: Allegany Ballistics Laboratory, Cumberland, Maryland.

[1] A. Lenard, J. Math. Phys. 2, 682 (1961).
[2] S. F. Edwards and A. Lenard, J. Math. Phys. 3, 778 (1962).
[3] A. Lenard, J. Math. Phys. 4, 533 (1963).
[4] S. Prager, Advan. Chem. Phys. 4, 201 (1961).
[5] A. Salzberg and S. Prager, J. Chem. Phys. 38, 2587 (1963).
[6] M. Kac, Phys. Fluids, 2, 8 (1959).
[7] T. Hill, *Introduction to Statistical Thermodynamics* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1960).

$$H_e = \partial\left(\frac{G_e}{T}\right)/\partial\left(\frac{1}{T}\right)$$

$$= (N+1)kT + N\,dP + dgm^2 \sum_{i=0}^{N} n_i,$$

$$S_e = (H_e - G_e)/T,$$

where G_e, H_e, and S_e are, respectively, the configurational contributions to the Gibbs free energy, the enthalpy and the entropy. Since it is the deviations from ideal-gas behavior which are of interest, the excess functions G', H', and S' are examined:

$$G' = G_e - (N+1)kT \ln (P/kT),$$
$$H' = H_e - (N+1)kT, \qquad (2.8)$$
$$S' = (H' - G')/T.$$

It becomes apparent upon replacing the summations in (2.7) by integrations that the thermodynamic functions are not extensive properties of the system; specifically, for large N,

$$H' \to dgm^2N^3/6,$$
$$S' \to -k[N \ln (gm^2/P) + 2(b + \tfrac{1}{2}N) \ln (b + \tfrac{1}{2}N)$$
$$- 2(b - \tfrac{1}{2}N) \ln (b - \tfrac{1}{2}N)], \qquad (2.9)$$
$$b \equiv [N^2 + p/(gm^2)]^{0.5}.$$

It is of interest to note that if the particles have finite length the degree of deviation from extensive behavior is given primarily by the enthalpy; whereas for point particles, the deviation arises entirely from the entropic terms.

The equation of state can be directly obtained:

$$L = -\left(\frac{\partial G_e}{dP}\right)_T = Nd + kT \sum_{i=0}^{N} \frac{1}{P + gm^2n_i}. \qquad (2.10)$$

It is clear from Eq. (2.10) that, even at constant pressure, the free length of the system is not proportional to the number of particles present. Since the terms of the sum in (2.10) corresponding to $i = 0$ and $i = N$ simply arise from a simultaneous translation of all N particles, it is natural to introduce the variable

$$L^* = L - Nd - 2kT/P, \qquad (2.11)$$

where L^* is the free length between the first and last particle. Although it is clear from (2.10) and (2.11) that

$$\lim_{N\to\infty} (L^*/kT) = 0, \qquad (2.12)$$

it is of interest to examine the asymptotic behavior of the limit. Due to the form of the function to be summed in Eq. (2.10), the following lemma must be established before the standard technique of replacing a sum by an integral is employed.

Lemma 1. If $g, m > 0$, $\lim_{N\to\infty} E_N/I_N = 0$; where

$$I_N = \int_2^{N-2} F(N, i)\,di, \qquad S_N = \sum_{i=2}^{N-2} F(N, i),$$

$F(N, i) = 1/(P + gm^2n_i)$, and $E_N = |I_N - S_N|$.

Proof. (N even). It can be seen from the monotonicity of $F(N, i)$ in the regions $i \in [1, N/2]$, $i \in [N/2, N-1]$ and from the fact that $F(N, N/2 + i) = F(N, N/2 - i)$ that

$$2\int_2^{N/2} diF(N, i) \le \sum_{i=2}^{N-2} F(N, i)$$
$$\le 2\int_2^{N/2} diF(N, i-1), \qquad (2.13)$$

hence

$$E_N \le 2\left|\int_2^{N/2} di[F(N, i) - F(N, i-1)]\right|,$$
$$\le 2\left|\int_1^2 diF(N, i)\right| + 2\left|\int_{N/2-1}^{N/2} diF(N, i)\right|.$$

Since for sufficiently large N, gm^2 and P can be neglected in comparison to gm^2iN

$$E_N \to \ln (2)/(Ngm^2) + o(1/N^2). \qquad (2.14)$$

From the integration of $F(N, i)$, we obtain

$$I_N = \left[\ln\left(\frac{\frac{1}{2}N + b + 2}{-\frac{1}{2}N + b + 2}\right)\right]/(bgm^2), \qquad (2.15)$$

which reduces for $N \gg P/(gm^2)$, $\ln N \gg 1$ to

$$I_N \to 2 \ln (N)/(Ngm^2). \qquad (2.16)$$

Consequently $\lim_{N\to\infty} (E_N/I_N) = 0$, proving the lemma.

From Lemma 1, and Eqs. (2.10), (2.11), and (2.15), the following limiting expression for L^* can be derived:

$$\lim_{N\to\infty} (L^*Ngm^2/2kT \ln N) = 1, \qquad gm^2 > 0, \qquad (2.17)$$
$$\lim_{N\to\infty} (PL^*/NkT) = 1, \qquad gm^2 = 0.$$

One immediate consequence of Eq. (2.16) is that for sufficiently large N, L^* is independent of the external pressure, a result reminiscent of the stability of a star. A similar result is found in the limit $g \to \infty$, and can be interpreted as a gradual aggregation of the N particles into one giant particle.

3. TWO-DIMENSIONAL POINT MASSES

Solution of the two-dimensional Poisson equation[5] leads to a logarithmic potential between two point

masses. In particular,

$$U_{ij} = gm^2 \ln (r_{ij}) \tag{3.1}$$

where r_{ij} is the distance between particles i, j. Thus the configurational partition function Q of point masses confined to an area A is given by

$$Q = N!^{-1} \int \cdots \int_A \prod_{i<j} r_{ij}^{-gm^2/kT} \, da_1 \cdots da_N. \tag{3.2}$$

Although the integration in (3.2) can not be carried out, introduction of the coordinates $x_i' = x_i A^{-.5}$ allows us to write the relation

$$\ln (Q) = (N - \sum_{i<j} m_i m_j / 2kT) \ln (A) + \ln (Q^*) \tag{3.3}$$

in which Q^* is a reduced partition function independent of A. Since the formal development of (3.3) tacitly assumes that Q^* is finite, it is essential to determine its radius of convergence. As the integral represented by Q^* has its domain bounded from above by unity, divergence can occur only at the lower limit corresponding to an arbitrarily close approach of two or more particles. With this in mind we define for every[8] r_{ij}',

$$r_{ij}' = zR_{ij}, \tag{3.4}$$

where R_{ij} is a positive constant and z is a scaling parameter. In this representation

$$Q^* = F(R_{ij}) \int_0^1 z^{-u} \, dz, \tag{3.5}$$

where F is a positive definite function of all the R_{ij} and $u = gm^2N(N - 1)/2 + 1 - 2N$. Clearly the criterion for the convergence of the integral is that

$$Ngm^2/4kT \leq 1 + gm^2/4kT. \tag{3.6}$$

For all finite values of Q^* the equation of state follows at once

$$\begin{aligned} AP &= -AkT[\partial \ln (Q)/\partial A]_T \\ &= NkT(1 - (gm^2N - gm^2)/(4kT)), \end{aligned} \tag{3.7}$$

where the differentiation is to be carried out without changing the shape of the container. Although (3.7) appears to predict negative pressure at sufficiently low T or large N, Q^* is infinite in this region and the treatment leading to the equation of state is not valid. A physical interpetation of this behavior is that as the gravitational forces become dominant, the system coalesces to a single point.

ACKNOWLEDGMENTS

The author is particularly grateful to Dr. Stephen Prager for many helpful discussions on this topic. He would also like to thank both E. I. duPont de Nemours Company and the U. S. Air Force for financial support.

[8] A. Lenard (private communication).

Sufficient Conditions for an Attractive Potential to Possess Bound States

FRANCESCO CALOGERO
Istituto di Fisica dell'Università di Roma,
Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Rome, Italy
(Received 14 June 1964)

Simple conditions on the potential are given, which are sufficient to secure the existence of at least one bound state for each angular momentum $l \leq L$. One such condition is given by the inequality

$$-\int_0^R dr\, rV(r)\left[\frac{r}{R}\right]^{2L+1} - \int_R^\infty dr\, rV(r)\left[\frac{R}{r}\right]^{2L+1} \geq 2L + 1,$$

where R is an arbitrary radius and $V(r)$ an everywhere attractive potential. Upper bounds on the energy of the lower bound state for each angular momentum are also given.

1. INTRODUCTION

LET $V(r)$ be an attractive central potential, such that the integral

$$I = -\int_0^\infty dr\, rV(r) \tag{1.1}$$

is finite. Note that I is positive and dimensionless in the units chosen ($\hbar = 2m = 1$, m being the mass of the particle considered). Jost and Pais have shown that a *necessary* condition for the existence of a bound state is $I \geq 1$.[1] Subsequently Bargmann has proved the more-general inequality

$$n_l < I/(2l + 1), \tag{1.2}$$

where n_l is the number of bound states with angular momentum l.[2] He also showed that this estimate is the best possible, in the sense that potentials may be constructed which saturate it for any given value of n_l. More recently Schwinger has given a new derivation of this inequality and has extended it, deriving *lower* limits for the energies of any bound state.[3] It is the purpose of the present paper to provide *upper* limits for the energies of bound states, and therefore also simple conditions on the potential *sufficient* to guarantee the existence of at least one bound state. We consider only the lower bound state for each angular momentum. These results supplement those of Bargmann and Schwinger and may be useful to obtain quick estimates of the properties of phenomenological potentials.

Upper limits for the energies of bound states, and therefore also conditions on the potential sufficient to guarantee the existence of bound states, may also be obtained from the well-known Rayleigh–Ritz variational principle.[4] We believe that our results may be of some interest nonetheless, in view of their close correspondence with those of Bargmann and Schwinger.

Our treatment is based on the phase approach to scattering theory.[5] The results are stated and proved in Sec. 3, Sec. 2 being used for the proof of a mathematical theorem.

2. ON THE POLES OF THE SOLUTION OF A RICCATI EQUATION

Consider the Riccati equation

$$y'(x) = f(x)[y(x) + g(x)]^2, \tag{2.1}$$

with boundary condition

$$y(0) = 0, \tag{2.2}$$

and with the following limitations on the real functions $f(x)$, $g(x)$:

$$f(x) \geq 0, \tag{2.3a}$$

$$g(x) \geq 0. \tag{2.3b}$$

Let the function $\tilde{g}(x)$ possess the following properties:

$$0 \leq \tilde{g}(x) \leq g(x), \tag{2.4a}$$

$$\lim_{x\to 0} [y(x)/\tilde{g}(x)] = 0, \tag{2.4b}$$

$$\tilde{g}'(x) \geq 0. \tag{2.4c}$$

We now prove the following

Theorem: A sufficient condition for the function $y(x)$ to have at least one pole in the interval of

[1] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[2] V. Bargmann, Proc. Nat. Acad. Sci. U. S. **38**, 961 (1952).
[3] J. Schwinger, Proc. Nat. Acad. Sci. U. S. **47**, 122 (1961).
[4] See, for instance, L. Spruch, "Minimum Principles in Scattering Theory," in *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, Inc., New York, 1962), Vol. IV.
[5] F. Calogero, Nuovo Cimento **27**, 261 (1963).

the real axis between 0 and $X > 0$ is that

$$\int_0^X dx\, \tilde{g}^2(x)f(x) \geq \frac{1}{4}\tilde{g}(X). \tag{2.5}$$

The proof is *per absurdum*. Let us assume that $y(x)$ is finite in the interval 0 to X. We then introduce the function $\tilde{y}(x)$ through

$$\tilde{y}'(x) = f(x)[\tilde{y}(x) + \tilde{g}(x)]^2 \tag{2.6}$$

and

$$\tilde{y}(0) = 0. \tag{2.7}$$

We now note that $\tilde{y}(x)$ is positive and that

$$\tilde{y}(x) \leq y(x) \tag{2.8}$$

as implied by Eqs. (2.1) and (2.6). Therefore $\tilde{y}(x)$ is also finite in the interval 0 to X.

We then introduce

$$z(x) = \tilde{y}(x)/[\tilde{y}(x) + \tilde{g}(x)]. \tag{2.9}$$

Then Eq. (2.4b) implies

$$z(0) = 0, \tag{2.10}$$

while the fact that $\tilde{g}(x)$ and $\tilde{y}(x)$ are positive and finite implies, together with (2.4b) and (2.4c), that

$$0 \leq z(x) < 1 \tag{2.11}$$

in the interval 0 to X. On the other hand $z(x)$ satisfies the differential equation

$$z'(x) = -[\tilde{g}'(x)/\tilde{g}(x)]z(x)[1 - z(x)] + \tilde{g}(x)f(x), \tag{2.12}$$

as follows from Eqs. (2.9) and (2.6). From this equation and Eqs. (2.4a), (2.4c), and (2.11) we see that

$$z'(x) \geq -[\tilde{g}'(x)/\tilde{g}(x)]z(x) + \tilde{g}(x)f(x). \tag{2.13}$$

Finally we introduce $u(x)$ through

$$u'(x) = -[\tilde{g}'(x)/\tilde{g}(x)]u(x) + \tilde{g}(x)f(x) \tag{2.14}$$

and

$$u(0) = 0. \tag{2.15}$$

Then

$$u(x) \leq z(x), \tag{2.16}$$

and from Eq. (2.11),

$$u(x) < 1 \tag{2.17}$$

in the interval 0 to X. But the equation for $u(x)$ is integrable and it yields

$$u(x) = \int_0^x dx' \tilde{g}^2(x')f(x')/\tilde{g}(x). \tag{2.18}$$

Therefore the hypothesis of the theorem, Eq. (2.5), implies

$$u(X) \geq 1. \tag{2.19}$$

This contradicts Eq. (2.17), so that the original assumption that $y(x)$ be finite in the interval 0 to X is disproved, q.e.d.

We also note without proof that a *necessary* condition for the divergence of $y(x)$ in the interval 0 to X is

$$\int_0^X dx g(x)f(x) \geq 1. \tag{2.20}$$

For the validity of this result it is required, besides the conditions Eqs. (2.3a) and (2.3b), that $g(x)$ be a nondecreasing function.

3. APPLICATION TO THE BOUND-STATE PROBLEM

For simplicity we derive the results only for S waves. The generalization to all partial waves is easily done, and we give the relevant results at the end.

The condition for the existence of an S-wave bound state of energy $E \leq -q^2$ is equivalent to the requirement that the solution of the Riccati equation

$$S'(r) = (2q)^{-1}V(r)[S(r)e^{-qr} - e^{qr}]^2 \tag{3.1}$$

with boundary condition

$$S(0) = 1 \tag{3.2}$$

has at least one pole on the positive real axis.[5] We may assume here that the potential $V(r)$ vanishes at infinity faster than exponentially, to dispose of the difficulty due to the divergence in the asymptotic behavior of $S(r)$ which otherwise occurs.[5] Clearly this assumption does not modify any of the physical properties, and in fact it may be forgotten once we have obtained our final results.

It is convenient to make a change in the dependent variable, setting

$$S(r) = 1 - 2qA(r). \tag{3.3}$$

Of course this substitution does not affect the location of the poles, which is the same for $S(r)$ and $A(r)$. On the other hand $A(r)$ satisfies the equation

$$A'(r) = -V(r)e^{-2qr}[A(r) + e^{qr}\sinh qr/q]^2, \tag{3.4}$$

with boundary condition

$$A(0) = 0. \tag{3.5}$$

We now assume the potential to be everywhere attractive

$$V(r) = -|V(r)|. \tag{3.6}$$

We may then apply the theorem of the preceding section, which immediately implies for the energy of the ground state the upper bound

$$E \leq -q^2, \tag{3.7}$$

where q is a solution, if any, of the equation

$$\int_0^\infty dr\, |V(r)|\, e^{-2qr} \tilde{g}^2(r) = \tilde{g}(\infty). \tag{3.8}$$

Here $\tilde{g}(r)$ is an arbitrary function, restricted by the three conditions

$$0 \leq \tilde{g}(r) \leq e^{qr} \sinh qr/q, \tag{3.9a}$$

$$\lim_{r\to 0} [r^{3+\eta}/\tilde{g}(r)] = 0, \tag{3.9b}$$

$$\tilde{g}'(r) \geq 0. \tag{3.9c}$$

The second condition corresponds to Eq. (2.4b) because[5]

$$A(r) \underset{r\to 0}{\longrightarrow} \text{const} \times r^{3+\eta}, \tag{3.10}$$

where η is defined by the behavior of the potential in the origin through

$$V(r) \underset{r\to 0}{\longrightarrow} \text{const} \times r^{\eta}, \qquad \eta > -2. \tag{3.11}$$

We do not elaborate on the possible choices for $\tilde{g}(r)$, which may be suited to the specific problem. We proceed instead to the case $q = 0$, which is particularly interesting, for it yields a sufficient condition on the potential for the existence of at least one bound state. The condition reads

$$\int_0^\infty dr\, |V(r)|\, g_0^2(r) \geq g_0(\infty), \tag{3.12}$$

where now $g_0(r)$ is restricted by the conditions

$$0 \leq g_0(r) \leq r, \tag{3.13a}$$

$$\lim_{r\to 0} [r^{3+\eta}/g_0(r)] = 0, \tag{3.13b}$$

$$g_0'(r) \geq 0. \tag{3.13c}$$

We mention three possible choices for $g_0(r)$. The first is

$$g_0(r) = rR/(r + R). \tag{3.14}$$

This yields, as the condition for the existence of at least one bound state

$$\int_0^\infty dr\, r\, |V(r)|\, rR/(r + R)^2 \geq 1. \tag{3.15}$$

A second choice is

$$g_0(r) = R(1 - e^{-r/R}). \tag{3.16}$$

This yields, as the condition for the existence of at least one bound state

$$R \int_0^\infty dr\, |V(r)|\, (1 - e^{-r/R}) \geq 1. \tag{3.17}$$

For instance, for an exponential potential

$$V(r) = -|V_0| \exp(-r/r_0),$$

setting $R = 1.41 r_0$ we find $|V_0|\, r_0^2 \geq 2.91$. Bargmann's necessary condition yields $|V_0|\, r_0^2 > 1$. The minimum value of $|V_0|\, r_0^2$ for which a bound state becomes actually possible is 1.44.

A third choice, and usually a better one, is

$$\begin{aligned} g_0(r) &= r, \qquad r \leq R, \\ g_0(r) &= R, \qquad r \geq R. \end{aligned} \tag{3.18}$$

This yields, as the condition for the existence of at least one bound state

$$\int_0^R dr\, r\, |V(r)|\, (r/R) + \int_R^\infty dr\, r\, |V(r)|\, (R/r) \geq 1. \tag{3.19}$$

Note that in all cases the distance R is arbitrary; its choice may be delayed to after the integrals have been performed. The more stringent limitation is usually obtained choosing for R a value close to the range of the potential, independently from the strength of the potential.

For a potential which vanishes identically beyond the range r_0 one may chose $R = r_0$ in Eq. (3.18) (although this choice need not be the most convenient one, see below). The sufficient condition for the existence of one bound state becomes then simply

$$\int_0^{r_0} dr\, r\, |V(r)|\, (r/r_0) \geq 1. \tag{3.20}$$

This may be compared with the necessary condition for the existence of one bound state, which in this case is

$$\int_0^{r_0} dr\, r\, |V(r)| > 1. \tag{3.21}$$

For instance, for a square-well potential of depth $|V_0|$ and range r_0, Eq. (3.20) yields $|V_0|\, r_0^2 \geq 3$, while Eq. (3.21) yields $|V_0|\, r_0^2 > 2$. The actual limiting value for the existence of a bound state is $|V_0|\, r_0^2 = 2.46$. A more stringent sufficient condition is obtained using Eq. (3.19) with $R = \frac{3}{4} r_0$ rather than Eq. (3.20), for it yields $|V_0|\, r_0^2 \geq 2.67$.

Another case of some interest is the potential[6]

$$V(r) = -(I/r_0)\delta(r - r_0). \tag{3.22}$$

In fact setting $R = r_0 + \epsilon$ in Eq. (3.19) we find as a sufficient condition for the existence of at least one bound state $I \geq 1$, while the Jost–Pais–Bargmann necessary condition is $I > 1$. This proves that the value of I at which a bound state becomes possible must be 1, a result which may be verified by direct computation. Note that this also implies that the condition of Eq. (3.19) is the best possible, for we have now found a potential which saturates it.

Finally we mention the generalization of these results to higher partial waves. It is achieved substituting everywhere $[\hat{h}_l^{(1)}(iqr)]^{\pm 1}$ in place of $\exp(\mp qr)$ and $-i\hat{j}_l(iqr)$ in place of $\sinh(qr)$. Here $\hat{h}_l^{(1)}(z)$ and $\hat{j}_l(z)$ are the Riccati–Bessel functions, defined as in Ref. 5. With these substitutions, which incidentally keep all the equations real, Eqs. (3.8) and (3.9) yield upper limits for the energy of the lower bound state corresponding to any given value of the angular momentum l.[7] As for the conditions on the potential sufficient to secure the existence of at least one bound state of angular momentum l (and therefore also for all angular momenta up to l), we find, in place of Eq. (3.8), the relation

$$\int_0^\infty dr\, |V(r)|\, r^{-2l} g_l^2(r) \geq g_l(\infty)(2l + 1), \tag{3.23}$$

with $g_l(r)$ restricted by the conditions

$$0 \leq g_l(r) \leq r^{2l+1}, \tag{3.24a}$$

$$\lim_{r\to 0} [r^{2l+3+\eta}/g_l(r)] = 0, \tag{3.24b}$$

$$g_l'(r) \geq 0. \tag{3.24c}$$

Corresponding to Eq. (3.15) we now have

$$\int_0^\infty dr\, r\, |V(r)|\, (rR)^{2l+1}/(r^{2l+1} + R^{2l+1})^2 \geq 2l + 1, \tag{3.25}$$

and corresponding to Eq. (3.19) we have

$$\int_0^R dr\, r\, |V(r)|\, (r/R)^{2l+1} + \int_R^\infty dr r\, |V(r)|\, (R/r)^{2l+1} \geq 2l + 1. \tag{3.26}$$

For instance for a square well of depth $|V_0|$ and range r_0 setting $R = \frac{4}{5}r_0$ in this equation yields $|V_0|\, r_0^2 \geq 11.7$ as the sufficient condition for the occurrence of a P-wave bound state, while Bargmann's necessary condition yields $|V_0|\, r_0^2 > 6$. The exact limiting value is 9.9. It is also easily seen that Eq. (3.26) provides the best possible estimate in the sense that there exists a potential which saturates it. This is the potential equation (3.22) with $I = 2l + 1$, which also saturates the Bargmann condition, Eq. (1.2), for $n_l = 1$.

Finally we observe that it would be easy to prove Bargmann's and Schwinger's results on the basis of Eq. (3.4) (or its generalization to higher partial waves), using the result mentioned at the end of Sec. 2.

[6] This may be considered as the appropriate limit of a finite potential of appropriate shape.

[7] One must also substitute Eq. (3.24b) for Eq. (3.9b).

On the Reduction Formula of Feinberg and Pais

C. G. BOLLINI AND J. J. GIAMBIAGI
Departamento de Física, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina
AND
A. GONZÁLEZ DOMÍNGUEZ
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina
(Received 14 July 1964)

Feinberg and Pais have considered a "reduction" formula for the Fourier transform of functions which depend only on the hyperbolic distance. They have shown that the formula is valid in particular cases and stated that it should be valid in general.

We show that it is valid for any causal distribution and furthermore that it is actually an extension of the well-known Bochner Theorem on the Fourier transform of radial functions.

1. INTRODUCTION

IN their work on weak interactions, Feinberg and Pais[1] have used the following "reduction" formula[2]:

$$I(q^2) = \int \psi(y^2)e^{iq\cdot y}\,d^4y = F(\psi), \tag{1}$$

$$I(q^2) = \frac{4i\pi^2}{q}\left[\int_c H_1^{(1)}(qy)\psi(y^2)y^2\,dy - \int_0^\infty J_1(qy)\psi(y^2)y^2\,dy\right], \tag{2}$$

where

$$y^2 = y_1^2 + y_2^2 + y_3^2 - y_4^2, \qquad q^2 = q_1^2 + q_2^2 + q_3^2 - q_4^2.$$

In all the cases considered by Feinberg and Pais,[1] it turns out that the contour integral vanishes. Furthermore, they state: "We are convinced, that any alternative treatment of the contour integral leads to physical absurdities."[3] This means that, in fact

$$F(\psi(y^2)) = -\frac{4i\pi^2}{q}\int_0^\infty J_1(qy)\psi(y^2)y^2\,dy. \tag{3}$$

We would like to point out that the latter formula is actually the causal version of the following well-known theorem, due to Bochner[4] (which we are writing for a four-dimensional space):

If

$$f(y_1y_2y_3y_4) \in L_1,$$

depends only on $R = (y_1^2 + y_2^2 + y_3^2 + y_4^2)^{\frac{1}{2}}$, then the function

$$\phi(q_1q_2q_3q_4) = \int f(y_1y_2y_3y_4)e^{iq\cdot y}\,d^4y$$

depends only on

$$\rho = (q_1^2 + q_2^2 + q_3^2 + q_4^2)^{\frac{1}{2}}.$$

Furthermore,

$$\phi(\rho) = \frac{4\pi^2}{\rho}\int_0^\infty f(R)J_1(\rho R)R^2\,dR. \tag{4}$$

In the applications to physics of form (3), $\psi(y^2)$ is not, in general, an ordinary function but a distribution. What is actually needed for our purpose is formula (4) for the case in which f is a tempered distribution.[5] This extension has been done by one of us.[6]

2. CAUSAL DISTRIBUTIONS

We shall first define the "causal" distributions.[7] Starting with a radial distribution,[6] which will be simply written as $\psi(y_1^2 + y_2^2 + y_3^2 + y_4^2)$, we introduce a positive parameter a by means of the following definition (φ is a test function $\in S$ [5]):

$$(\psi_a, \varphi) = (\Psi(y_1^2 + y_2^2 + y_3^2 + ay_4^2), \varphi(y_1, y_2, y_3, y_4)) \equiv \frac{1}{a^{\frac{1}{2}}}\left(\Psi(R^2), \varphi\left(y_1, y_2, y_3, \frac{y_4}{a^{\frac{1}{2}}}\right)\right).$$

As usual, Ψ_a will be said to be analytic in a if for any test function φ, the ordinary function (ψ_a, φ) is analytic in a.[8] When Ψ_a can be analytically continued into the whole of the upper half-plane of a, we define

[1] G. Feinberg and A. Pais, Phys. Rev. **131**, 2724 (1963).
[2] Reference 1, p. 2735. See also, Y. Pwu and T. T. Wu, Phys. Rev. **133**, B 778 (1964).
[3] Reference 1, p. 2738.
[4] S. Bochner and K. Chandrasekharan, *Fourier Transform* (Princeton University Press, Princeton, New Jersey, 1949), p. 69.
[5] L. Schwartz, *Theorie des distributions* (Hermann & Cie., Paris, 1951), Vol. 2.
[6] A. González Domínguez (to be published).
[7] We follow the steps discussed by I. M. Guelfand and G. E. Shilov, *Les distributions* (Dunod Cie., Paris, 1962), in particular Chap. III, 2, p. 264.
[8] Reference 7, p. 148.

a causal distribution by the formula

$$\Psi(y^2 + i0) = \lim_{a \to -1+i0} \psi(y_1^2 + y_2^2 + y_3^2 + a y_4^2). \quad (5)$$

An "anticausal" distribution can be defined as the complex conjugate of a causal one.

We now show that it is possible to compute the Fourier transform of a causal distribution by means of Formula (3). In fact,[9]

$$(I, \varphi) \equiv (F(\psi), \varphi) = (\psi, F^{-1}(\varphi)) = \lim_{a \to -1+i0} (\psi_a, F^{-1}(\varphi)) = \lim_{a \to -1+i0} (F(\psi_a), \varphi);$$

i.e.,

$$I = \lim_{a \to -1+i0} F(\psi_a) = \lim_{a \to -1+i0} I_a,$$

where

$$I_a(q_1, q_2, q_3, q_4) = \int \psi(y_1^2 + y_2^2 + y_3^2 + a y_4^2) e^{iq \cdot y} \, d^4y.$$

For positive a we have

$$I\left(q_1, q_2, q_3, \frac{q_4}{a^{\frac{1}{2}}}\right) = \frac{1}{a^{\frac{1}{2}}} \int \psi(y_1'^2 + y_2'^2 + y_3'^2 + y_4'^2) e^{iq' \cdot y'} \, d^4y, \quad (6)$$

where the fourth component of q_a' is $q_4/a^{\frac{1}{2}}$.

Now, using Bochner theorem [Formula (4)], we obtain

$$I(\rho^2) = I\left(q_1^2 + q_2^2 + q_3^2 + \frac{1}{a} q_4^2\right) = \frac{4\pi^2}{a^{\frac{1}{2}} \rho} \int_0^\infty \psi(R^2) J_1(\rho R) R^2 \, dR. \quad (7)$$

Formula (6) is valid for positive a. By making the analytic continuation in the upper half-plane of the parameter a and taking the limit $a \to -1 + i0$, we obtain

$$I(q_1^2 + q_2^2 + q_3^2 + q_4^2) \equiv I(q^2) = \frac{4\pi^2}{iq} \int_0^\infty \psi(R^2) J_1(qR) R^2 \, dR, \quad (8)$$

which coincides with formula (3). [It is understood that $I(q^2)$ means $I(q^2 - i0)$; see Formula (5).]

Formula (8) is valid for causal distributions. For anticausal ones the result is

$$I(q^2) = -\frac{4\pi^2}{iq} \int_0^\infty \psi(R^2) J_1(qR) R^2 \, dR, \quad (9)$$

where it is understood that q^2 means $q^2 + i0$.

3. EXAMPLE

As an example of application we consider now the causal distribution

$$\psi(y^2) = (y^2 + i0)^\lambda, \quad (10)$$

where λ is any complex number. A direct application of Formula (8) gives[10]:

$$I(q^2) = \frac{4\pi^2}{iq} \int_0^\infty J_1(qy) y^{2\lambda+2} \, dy = \frac{4\pi^2}{i} \frac{2^{2\lambda+2} \Gamma(\lambda + 2)}{(q^2)^{\lambda+2} \Gamma(-\lambda)},$$

which is the correct result.[11]

It should be noted from (10) that, near the origin, $\psi(y^2)$ has a singularity worse than y^{-2} when $\lambda < -1$. Nevertheless, Formula (8) is still valid, although the contour integral considered by Feinberg and Pais is not meaningful in this case.

[9] See Ref. 5, p. 106.

[10] W. Groebner and N. Hofreiter, *Integraltafel* (*Bestimmte Integrale*) (Springer-Verlag, Berlin, 1958), p. 196. Also Formula (1), p. 198, analytically extended in κ.

[11] Reference 7, p. 278.