JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Simple Procedure for Determining the Number of Components of an Irreducible Tensor

J. L. PIETENPOL
Department of Physics, Case Western Reserve University, Cleveland, Ohio,
and
*Institut de Physique Théorique, Université de Louvain**
AND
D. SPEISER
*Institut de Physique Théorique, Université de Louvain**

(Received 31 March 1970)

We show that the number of linearly independent components of a tensor in n dimensions with specified symmetry properties is given by a polynomial in n. This polynomial can be determined in a simple way from the Young diagram associated with the tensor.

I. INTRODUCTION

There are two approaches to classifying the irreducible representations of the linear groups GL_n and SU_n. First, one can consider a fixed n, i.e., a space of fixed dimension, and label each representation by a set of $(n-1)$ parameters which run independently from 1 (or 0) to infinity. A standard formula due to Weyl is then available [cf. Eq. (3) below] for computing the dimensionality of the representation. [For SU_3 this is the familiar $pq(p+q)/2$ or $(\lambda_1+1)(\lambda_2+1)\times(\lambda_1+\lambda_2+2)/2$.] This approach is best understood from the point of view of the Cartan–Stiefel diagram,[1] the parameters appearing as coordinates of a lattice point in an $(n-1)$-dimensional space.

Alternatively, the irreducible representations can be built up from the fundamental n-dimensional representation by forming tensors and carrying out the well-known symmetrization and antisymmetrization process on the indices. From this point of view, the representation is described by a Young diagram; this diagram characterizes one irreducible representation of every GL_n (or SU_n), provided n is at least as large as the number of rows in the diagram.

This suggests the following problem: For a given Young diagram, what is the number of linearly independent components of a tensor belonging to that diagram, as a function of n?

For example, a symmetric tensor with two indices has $n(n+1)/2$ components; the fully-covariant Riemann curvature tensor has $n^2(n^2-1)/12$ components; etc. Note that such a formula refers to tensors transforming under *different* groups.

The purpose of this paper is to show that these formulas can be read off from the Young diagram in a simple way. In Sec. II we state the rule; Secs. III and IV contain the proof.

It appears that these results are at least partially known, but we know of no published proof.

II. STATEMENT OF THE RESULT

Consider a Young diagram consisting of N blocks arranged in r rows. Let $P(n)$ be the dimensionality of the associated irreducible representation of GL_n or SU_n. Equivalently, $P(n)$ is the number of linearly independent tensors with N indices (each index running from 1 to n) having the symmetry properties specified by the Young diagram. (When $n < r$, the tensor must vanish identically and we define $P(n) = 0$. In this case, there is no associated representation of the transformation group.)

Our result is as follows: $P(n)$ is a polynomial in n, all of its roots are integers, and the multiplicities of the roots are given by the lengths of the diagonals of the Young diagram.

Specifically, the number of blocks on the main diagonal gives the multiplicity of zero as a root; the number of blocks on the mth diagonal below (above) the main diagonal gives the multiplicity of $+m$ $(-m)$ as a root.

Thus, for the example shown in Fig. 1, we have

$$P(n) = c(n-2)(n-1)^2n^2(n+1)^2(n+2) \times (n+3)(n+4)$$
$$= cn^2(n^2-1)^2(n^2-4)(n+3)(n+4).$$

The leading coefficient c can be determined by comparing this formula with the dimensionality for the lowest GL_n compatible with the diagram, i.e., for $n = r$. In this example, the diagram yields a representation of GL_3 of dimension 15, so we have $c \cdot 1 \cdot 2^2 \cdot 3^2 \cdot 4^2 \cdot 5 \cdot 6 \cdot 7 = 15$, or $c = 1/(2^7 \cdot 3^2 \cdot 7)$.

However, the constant c can also be read off from the Young diagram in a reasonably simple way. We shall show that $c = d/N!$, where N is the total number of blocks in the diagram and d is the dimensionality of the associated representation of the symmetric group S_N. The algorithm for calculating d from the

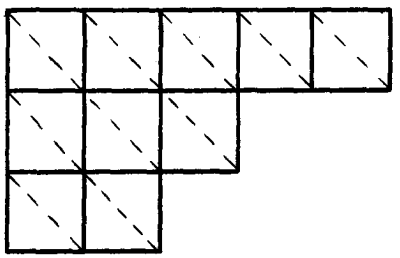


FIG. 1. Young diagram illustrating application of the diagonal rule.

diagram is well known.[2] A less profound but computationally simpler prescription for c will be obtained in the course of the proof [see Eq. (12) below].

III. PROOF OF THE DIAGONAL RULE FOR THE ROOTS

As before, we consider a Young diagram with r rows and N blocks. We also define

$$\begin{aligned} &\mu_i = \text{the length of the } i\text{th row } (\mu_i = 0 \text{ for } i > r), \\ &\lambda_i = \mu_i - \mu_{i+1}, \\ &p_i = \lambda_i + 1. \end{aligned} \tag{1}$$

Note that the definition of a Young diagram requires $\lambda_i \geq 0$, so that the p_i are positive integers. Clearly, we have

$$\begin{aligned} \mu_i &= \sum_{j=i}^{r} \lambda_j, \\ N &= \sum_{i=1}^{r} \mu_i. \end{aligned} \tag{2}$$

Our procedure will consist of beginning with the usual Weyl formula for the dimensionality of a representation of GL_n (or SU_n) and recasting it in such a way as to exhibit its polynomial nature as a function of n. The Weyl formula[3] (for $n \geq r$) is

$$\begin{aligned} P(n) &= p_1 p_2 \cdots p_{n-1} \cdot \frac{(p_1 + p_2)}{2} \cdots \frac{(p_{n-2} + p_{n-1})}{2} \cdots \\ &\qquad \frac{(p_1 + p_2 + \cdots p_{n-1})}{n-1} \\ &\equiv \prod_{i=1}^{n-1} \prod_{k=i}^{n-1} \left(\frac{1}{k - i + 1} \sum_{j=i}^{k} p_j \right). \end{aligned} \tag{3}$$

The n-dependence of this expression is obscured by the presence of n in the upper limits of the product symbols. We therefore proceed by removing these appearances of n. As a first step, we note that the quantity in brackets is equal to 1 whenever $i > r$. Thus, for $n > r$, the upper limit $(n - 1)$ of the first product symbol can be replaced by r, and we have

$$\begin{aligned} P(n) &= \prod_{i=1}^{r} \prod_{k=i}^{n-1} \left(\frac{1}{k - i + 1} \sum_{j=i}^{k} p_j \right) \\ &\equiv \prod_{i=1}^{r} \left(\frac{1}{(n-i)!} \prod_{k=i}^{n-1} \sum_{j=i}^{k} p_j \right). \end{aligned} \tag{4}$$

Although we have justified this step only for $n > r$, it is easily seen that the result is also valid for $n = r$, provided that an "empty product" such as $\prod_{k=r}^{r-1}(\ )$ is interpreted as unity. We shall make this interpretation throughout the remainder of the proof.

To proceed further with the simplification of the n dependence, we use the identity

$$\begin{aligned} \sum_{j=i}^{k} p_j &= \sum_{j=i}^{k} (1 + \lambda_j) = k - i + 1 + \sum_{j=i}^{k} \lambda_j \\ &= k - i + 1 + \mu_i - \mu_{k+1} \end{aligned} \tag{5}$$

and we split the product within the brackets into two parts:

$$\begin{aligned} P(n) = \prod_{i=1}^{r} \Bigg[\frac{1}{(n-i)!} &\left(\prod_{k=i}^{r-1} (k - i + 1 + \mu_i - \mu_{k+1}) \right) \\ &\cdot \left(\prod_{k=r}^{n-1} (k - i + 1 + \mu_i - \mu_{k+1}) \right) \Bigg]. \end{aligned} \tag{6}$$

Now the quantity within the first large parentheses is independent of n, while in the second large parentheses we have $\mu_{k+1} = 0$ (since $k + 1 > r$). Thus,

$$\begin{aligned} &\prod_{k=r}^{n-1} (k - i + 1 + \mu_i - \mu_{k+1}) \\ &\qquad = \prod_{k=r}^{n-1} (k - i + 1 + \mu_i) = \frac{(n - i + \mu_i)!}{(r - i + \mu_i)!}, \end{aligned} \tag{7}$$

and we finally obtain

$$\begin{aligned} P(n) = \left(\prod_{i=1}^{r} \frac{1}{(r - i + \mu_i)!} \prod_{k=i}^{r-1} (k - i + 1 + \mu_i - \mu_{k+1}) \right) \\ \times \left(\prod_{i=1}^{r} \frac{(n - i + \mu_i)!}{(n-i)!} \right). \end{aligned} \tag{8}$$

We have now achieved our goal: The dependence of $P(n)$ on n has been completely removed from the limits of the product symbols. In fact, the expression in the first large parentheses in this result is independent of n, while the expression in the second large parentheses is clearly a polynomial in n with leading coefficient 1. Moreover, the roots of this polynomial agree with the rule stated in Sec. II. In particular, the first row of the Young diagram contributes a factor $n (n + 1) \cdots (n + \mu_1 - 1)$, the second row contributes $(n - 1)n \cdots (n + \mu_2 - 2)$, etc., and this corresponds exactly to the "diagonal rule" for the roots of $P(n)$.

IV. PROOF OF THE FORMULA FOR THE LEADING COEFFICIENT

We now turn our attention to the computation of the leading coefficient of the polynomial, which is given in the first large parentheses of Eq. (8):

$$c = \prod_{i=1}^{r} \left(\frac{1}{(r-i+\mu_i)!} \prod_{k=i}^{r-1} (k - i + 1 + \mu_i - \mu_{k+1}) \right). \tag{9}$$

It is convenient to define

$$a_i = r - i + \mu_i, \quad 1 \leq i \leq r. \tag{10}$$

Then we have

$$c = \prod_{i=1}^{r} \left(\frac{1}{a_i!} \prod_{k=i}^{r-1} (a_i - a_{k+1}) \right) \tag{11}$$

or

$$c = \left(\prod_{i=1}^{r} \frac{1}{a_i!} \right) \left(\prod_{\substack{i,j=1 \\ i<j}}^{r} (a_i - a_j) \right). \tag{12}$$

This equation provides an easy method for computing c. The a's can be read from the Young diagram, since a_i is equal to the number of blocks in the ith row plus the number of blocks in the first column lying below the ith row. The coefficient c is then given by a fraction whose numerator is the product of the positive differences of the a's taken in pairs, and whose denominator is the product of the factorials of the a's.

We shall now prove that $c = d/N!$. Each Young diagram with N blocks is associated with an irreducible representation of the symmetric group S_N; d is the dimensionality of that representation. The following rule is well known[4] and clearly determines d uniquely: The diagram with $N = 1$ has $d = 1$; the value of d for any other diagram equals the sum of the d's associated with all subdiagrams obtained by removing one block.

We begin by summing the c's of all subdiagrams with one block less. An acceptable subdiagram is obtained by removing the last block of (say) the ith row, provided this row is longer than the $(i+1)$th row—i.e., provided $a_i - a_{i+1} > 1$. Then the effect of removing the block is to replace a_i by $(a_i - 1)$, leaving all other a's unchanged.[5] A block cannot be removed from the ith row when it has the same length as the succeeding row (i.e., when $a_i = a_{i+1} = 1$), but in this case the replacement of a_i by $(a_i - 1)$ leads to a set of a's giving $c = 0$ when substituted into Eq. (12); thus no error arises from including these terms in the sum.

If c_i denotes the value of the rhs of Eq. (12) with a_i reduced by 1, then we have

$$\frac{c_i}{c} = a_i \prod_{\substack{j=1 \\ j \neq i}}^{r} \frac{a_i - 1 - a_j}{a_i - a_j}, \tag{13}$$

so that

$$\frac{1}{c} \sum_i c_i = \sum_{i=1}^{r} \left(a_i \prod_{\substack{j=1 \\ j \neq i}}^{r} \left(1 + \frac{1}{a_j - a_i} \right) \right). \tag{14}$$

In an appendix we prove the identity

$$\sum_{i=1}^{r} \left(a_i \prod_{\substack{j=1 \\ j \neq i}}^{r} \left(1 + \frac{1}{a_j - a_i} \right) \right) = -\frac{r(r-1)}{2} + \sum_{i=1}^{r} a_i, \tag{15}$$

which holds for any set of distinct numbers $\{a_i\}$. But we see that the total number of blocks in the diagram is given by

$$N = \sum_{i=1}^{r} \mu_i = \sum_{i=1}^{r} (a_i - r + i) = -\frac{r(r-1)}{2} + \sum_{i=1}^{r} a_i. \tag{16}$$

Combining Eqs. (14)–(16), we find

$$\sum_i c_i = Nc, \tag{17}$$

which can be written

$$\sum_i (N-1)!\, c_i = N!\, c. \tag{18}$$

Moreover, a diagram with just one block has $N!\, c = 1$. Thus, $N!\, c$ satisfies the same recursion as d, so we must have $c = d/N!$, completing the proof.

ACKNOWLEDGMENTS

It is a pleasure to thank L. L. Foldy and C. N. Yang for decisive help. We are grateful to P. B. Kantor and D. Fairlie for discussions. One of us (D. S.) is grateful for the warm hospitality found at Case Western Reserve University. Finally, one of us (J. P.) expresses his gratitude to Prof. M. de Hemptinne and the members of the physics group for a most cordial welcome to the Université de Louvain.

APPENDIX

We now give the proof of Eq. (15).

Theorem: If $a_1, \cdots, a_r$ are any distinct numbers, then

$$\sum_{i=1}^{r} \left(a_i \prod_{\substack{j=1 \\ j \neq i}}^{r} \left(1 + \frac{1}{a_j - a_i} \right) \right) = -\frac{r(r-1)}{2} + \sum_{i=1}^{r} a_i.$$

Proof: The identity may be rewritten as

$$\sum_{i=1}^{r} \left[a_i \left(-1 + \prod_{\substack{j=1 \\ j \neq i}}^{r} \left(1 + \frac{1}{a_j - a_i} \right) \right) \right] = -\frac{r(r-1)}{2}.$$

Consider the lhs of this equation as a function of the complex variable a_r, with $a_1, \cdots a_{r-1}$ as fixed parameters. This is a rational function of a_r, with (at

most) simple poles at $a_1, \cdots a_{r-1}$. By examining the expression, it is easily seen that the residue at each "pole" vanishes. Moreover, as $a_r \to \infty$, the expression approaches the limit

$$-(r-1)+\sum_{i=1}^{r-1}\left[a_i\left(-1+\prod_{\substack{j=1\\ j\neq i}}^{r-1}\left(1+\frac{1}{a_j-a_i}\right)\right)\right].$$

Being a bounded entire function, the expression must be independent of a_r. The result then follows by induction on r.

We mention an interesting related identity:

$$\sum_{i=1}^{r}\prod_{\substack{j=1\\ j\neq i}}^{r}\left(1+\frac{1}{a_j-a_i}\right)=r,$$

which can be proved by a similar method, or else can be obtained as a corollary of the previous identity by evaluating the change when all the a's are incremented by equal amounts.

* Mailing address: Centre de Physique Nucléaire, Parc d'Arenberg, 3030 Heverlee, Belgium.

[1] J.-P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 (1964); **5**, 1560 (1964).

[2] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962), p. 198.

[3] H. Weyl, *Selecta* (Birkhäuser Verlag, Basel-Stuttgart, 1956), p. 262ff.

[4] This rule is related to the Frobenius construction of the representations of S_N; it is clearly equivalent to the combinatorial rule given by Hamermesh in Ref. 2.

[5] A slight complication arises in case the last row has only one block. Its removal then leads to an $(r-1)$-rowed diagram with a's given by $(a_1-1, \cdots a_{r-1}-1)$. The simplest procedure is to treat this as an r-rowed diagram with an empty last row, having a's given by $(a_1, \cdots a_{r-1}, 0)$. Then the discussion in the text applies.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Class of Perturbation Theories of Ordinary Differential Equations

MARTIN KUMMER
Department of Mathematics, The University of Toledo, Toledo, Ohio 43606

(Received 26 June 1967; Revised Manuscript Received 29 April 1970)

A class of perturbation theories of ordinary differential equations is studied in a systematic and rigorous way. This class contains the perturbation theory by Kruskal [J. Math. Phys. **3**, 806 (1962)] and its generalization discussed by Coffey [J. Math. Phys. **10**, 426 (1969)] as well as the formal aspects of the perturbation theory for quasiperiodic solutions by Moser [Math. Ann. **169**, 136 (1967)]. It is shown that the systematic generalization of some algebraic ideas by Sternberg [J. Math. Mech. **10**, 451 (1961)] and Moser provides a framework in which many structures of these perturbation theories become more transparent. Especially, two questions raised by Coffey are answered. Finally, we touch upon the question of convergence of the formal expansions.

1. INTRODUCTION

In his paper, which shows how to generalize Bogoliubov's perturbation theory of ordinary differential equations in the case of systems with all the solutions nearly periodic, Kruskal[1] uses Bogoliubov's[2] averaging method to construct a new system which no longer contains the "angle variable." By solving this new system he is able to construct a solution of the perturbed system which avoids "secular" terms.

This procedure can be described also in terms of formally equivalent systems. Instead of directly aiming at the construction of the perturbed solution, one first constructs a formally equivalent system which if truncated at some finite power of ϵ can be solved directly. This solution then is plugged into the formal transformation of coordinates to give the approximate solution.

Introducing an equivalence relation between formal vectorfields (i.e., formal power series in ϵ of analytic vectorfields) over some open set of R^n, we are able to generalize his idea. Especially, we are concerned with a generalization recently discussed by Coffey.[3] Coffey asks essentially two questions: (1) Can degenerate perturbation theory be made canonical to all orders? (2) How can it be understood that Coffey's approximation of a certain solution of a certain exactly solvable system of differential equations is asymptotic to the exact solution for all times, whereas usually such an approximation is only asymptotic for a time interval of length L/ϵ? We clarify both points by first giving exact conditions under which the first question can be answered affirmatively and secondly proving a theorem [Theorem 2] about the approximation of asymptotically stable solutions which when

most) simple poles at $a_1, \cdots a_{r-1}$. By examining the expression, it is easily seen that the residue at each "pole" vanishes. Moreover, as $a_r \to \infty$, the expression approaches the limit

$$-(r-1)+\sum_{i=1}^{r-1}\left[a_i\left(-1+\prod_{\substack{j=1\\ j\neq i}}^{r-1}\left(1+\frac{1}{a_j-a_i}\right)\right)\right].$$

Being a bounded entire function, the expression must be independent of a_r. The result then follows by induction on r.

We mention an interesting related identity:

$$\sum_{i=1}^{r}\prod_{\substack{j=1\\ j\neq i}}^{r}\left(1+\frac{1}{a_j-a_i}\right)=r,$$

which can be proved by a similar method, or else can be obtained as a corollary of the previous identity by evaluating the change when all the a's are incremented by equal amounts.

* Mailing address: Centre de Physique Nucléaire, Parc d'Arenberg, 3030 Heverlee, Belgium.

[1] J.-P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 (1964); **5**, 1560 (1964).

[2] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962), p. 198.

[3] H. Weyl, *Selecta* (Birkhäuser Verlag, Basel-Stuttgart, 1956), p. 262ff.

[4] This rule is related to the Frobenius construction of the representations of S_N; it is clearly equivalent to the combinatorial rule given by Hamermesh in Ref. 2.

[5] A slight complication arises in case the last row has only one block. Its removal then leads to an $(r-1)$-rowed diagram with a's given by $(a_1-1, \cdots a_{r-1}-1)$. The simplest procedure is to treat this as an r-rowed diagram with an empty last row, having a's given by $(a_1, \cdots a_{r-1}, 0)$. Then the discussion in the text applies.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Class of Perturbation Theories of Ordinary Differential Equations

MARTIN KUMMER
Department of Mathematics, The University of Toledo, Toledo, Ohio 43606

(Received 26 June 1967; Revised Manuscript Received 29 April 1970)

A class of perturbation theories of ordinary differential equations is studied in a systematic and rigorous way. This class contains the perturbation theory by Kruskal [J. Math. Phys. **3**, 806 (1962)] and its generalization discussed by Coffey [J. Math. Phys. **10**, 426 (1969)] as well as the formal aspects of the perturbation theory for quasiperiodic solutions by Moser [Math. Ann. **169**, 136 (1967)]. It is shown that the systematic generalization of some algebraic ideas by Sternberg [J. Math. Mech. **10**, 451 (1961)] and Moser provides a framework in which many structures of these perturbation theories become more transparent. Especially, two questions raised by Coffey are answered. Finally, we touch upon the question of convergence of the formal expansions.

1. INTRODUCTION

In his paper, which shows how to generalize Bogoliubov's perturbation theory of ordinary differential equations in the case of systems with all the solutions nearly periodic, Kruskal[1] uses Bogoliubov's[2] averaging method to construct a new system which no longer contains the "angle variable." By solving this new system he is able to construct a solution of the perturbed system which avoids "secular" terms.

This procedure can be described also in terms of formally equivalent systems. Instead of directly aiming at the construction of the perturbed solution, one first constructs a formally equivalent system which if truncated at some finite power of ϵ can be solved directly. This solution then is plugged into the formal transformation of coordinates to give the approximate solution.

Introducing an equivalence relation between formal vectorfields (i.e., formal power series in ϵ of analytic vectorfields) over some open set of R^n, we are able to generalize his idea. Especially, we are concerned with a generalization recently discussed by Coffey.[3] Coffey asks essentially two questions: (1) Can degenerate perturbation theory be made canonical to all orders? (2) How can it be understood that Coffey's approximation of a certain solution of a certain exactly solvable system of differential equations is asymptotic to the exact solution for all times, whereas usually such an approximation is only asymptotic for a time interval of length L/ϵ? We clarify both points by first giving exact conditions under which the first question can be answered affirmatively and secondly proving a theorem [Theorem 2] about the approximation of asymptotically stable solutions which when

applied to Coffey's system gives an answer to Coffey's second question. By exploiting and generalizing some algebraic ideas of Moser[4] and Sternberg,[5] the algebra of the proofs becomes extremely simple, so that we can concentrate on a rigorous treatment of the analytic aspects of the proofs.

Finally, we touch upon the question of the convergence of the formal expansions and show that this question is strongly related to some research going on in the foundations of classical mechanics initiated by Kolmogorov.[6] We intend to dedicate a future paper to the implications of those works for the question of convergence of Kruskal's perturbation expansions.

2. BASIC DEFINITIONS AND PREPARATIONS

Let R^m be the set of real m-tuples $z = (z_1, \cdots, z_m)$ equipped with the distance

$$|z - z_0| = \max_{\nu=1,2,\cdots,m} |z_\nu - z_{0\nu}|,$$

$\mathfrak{D}$ some connected set of R^m, and the domains

$$\mathfrak{D}_\kappa = \{z \mid z \in C^m; \inf_{z_0 \in \mathfrak{D}} |\mathrm{Re}\,(z - z_0)| < \kappa, |\mathrm{Im}\, z| < \kappa\},$$

$\kappa > 0$, a system of open neighborhoods of $\mathfrak{D}$, considered as a subset of C^m (set of complex m-tuples).

$C_b^\omega(\mathfrak{D})$ (ω stands for analytic, b for bounded) shall denote the ring (with respect to obvious addition and multiplication) of all real-valued functions over $\mathfrak{D}$ with the following two properties.

(i) They are restrictions to $\mathfrak{D}$ of functions which are defined and analytic in some domain $\mathfrak{D}_\delta$. Here δ is a positive number which may depend on f. If we want to show this dependence explicitly, we write $\delta(f)$ instead of simply δ.

(ii)

$$|f|_\kappa \equiv \sup_{z \in \mathfrak{D}_\kappa} |f(z)| \tag{2.1}$$

exists for $0 < \kappa < \delta$.

Example: Let

$$\mathfrak{D}^{(c)} = \{(z_1, z_2) | \, |z_2| < c\} \subset R^2.$$

The function

$$f(z_1, z_2) = (z_2^4 - a^4)^{-1} \cos z_1$$

belongs to $C_b^\omega(\mathfrak{D}^{(c)})$ for each $c < a$, and $\delta(f) = a - c$. It is clear that $f \in C_b^\omega(\mathfrak{D})$ implies $f \in C_b^\omega(\hat{\mathfrak{D}})$, where $\hat{\mathfrak{D}}$ is any connected set contained in $\mathfrak{D}$.

Notice that, in case $\mathfrak{D}$ is bounded, (i) implies (ii). However, this is not the case, in general, as the reader can check by replacing $\cos z_1$ in the example above by z_1.

Let $f_{,\nu_1 \cdots \nu_p}$ be the derivative of f with respect to $z_1, \cdots, z_p$. It also belongs to $C_b^\omega(\mathfrak{D})$ if f does. Indeed, a simple application of Cauchy's integral formula yields

$$|f_{,\nu}|_\kappa \leq 2/(\delta - \kappa)\, |f|_\kappa. \tag{2.2}$$

We will also make use of the following ring of power series in ϵ:

$\mathfrak{F}(\mathfrak{D})$ = ring of formal power series in ϵ over $C_b^\omega(\mathfrak{D})$.

We use the notation

$$f(z, \epsilon) = \sum_{n=0}^{\infty} f^{(n)}(z) \epsilon^n.$$

If $f \in \mathfrak{F}(\mathfrak{D})$, the function $f^{(n)}(z)$ is called the nth coefficient of f, and by definition it belongs to $C_b^\omega(\mathfrak{D})$.

By $\mathfrak{H}(\mathfrak{D})$ we denote the subring of convergent power series or, more exactly, we have $f \in \mathfrak{H}(\mathfrak{D})$ if and only if there exist constants δ and $K(\kappa)$ such that $f(z, \epsilon)$ is real in $\mathfrak{D}$ for real ϵ, holomorphic in

$$\{(z, \epsilon) \mid z \in \mathfrak{D}_\kappa; 0 < \epsilon < [K(\kappa)]^{-1}, 0 < \kappa < \delta\}, \tag{2.3}$$

and uniformly bounded on $\mathfrak{D}_\kappa$ for fixed ϵ. (In order not to overload the text with absolute signs, we do not distinguish between ϵ as a complex number and its modulus. It is always clear from the context what ϵ stands for. If the reader does not like this slightly abusive notation, there is no harm in assuming that ϵ is always a real nonnegative number.)

δ and $K(\kappa)$ may again depend on $f \in \mathfrak{H}(\mathfrak{D})$. To show this dependence, we shall occasionally write $\delta(f)$ and $K(\kappa, f)$ [or shorter $K(f)$], and we shall call δ and K the characteristic constants of f.

To avoid the introduction of a new constant, namely the bound of f on $\mathfrak{D}_\kappa$, we may always choose $K(\kappa)$ in (2.3) in such a way that

$$|f^{(0)}|_\kappa \leq K(\kappa).$$

This choice implies, according to Cauchy's estimate,

$$|f^{(n)}|_\kappa \leq [K(\kappa)]^{n+1} \tag{2.4}$$

for $0 < \kappa < \delta$. Hence, we have

$$|f(z, \epsilon)| \leq \frac{K(\kappa)}{1 - \epsilon K(\kappa)} \tag{2.5}$$

on the set (2.3). Observe that $K(\kappa)$ is a nondecreasing function of κ.

It is easy to see that if f is a formal power series whose coefficients satisfy (2.4) and

$$\delta = \inf_{(n)} \delta(f^{(n)}) > 0,$$

then actually $f \in \mathfrak{H}(\mathfrak{D})$ with characteristic constants δ and $K(\kappa)$. To show that $\mathfrak{H}(\mathfrak{D})$ has really the structure of a ring with respect to the usual addition and multiplication of functions is also trivial and therefore left to the skeptical reader. We also check that if $f_{,\nu_1 \cdots \nu_p}$ stands for the power series obtained from $f \in \mathfrak{H}(\mathfrak{D})$ by differentiating f with respect to $z_{\nu_1}, \cdots, z_{\nu_p}$, then

$$f_{,\nu_1 \cdots \nu_p} \in \mathfrak{H}(\mathfrak{D}).$$

Indeed the corresponding δ is easily seen to be $\delta(f)$. Now let $z \in \mathfrak{D}_\kappa$ $[0 < \kappa < \delta(f)]$. Draw in the z_ν plane a circle of radius $\frac{1}{2}(\delta - \kappa)$ around z_ν and apply Cauchy's integral representation. Then one recognizes that

$$|f^{(n)}_{,\nu_1 \cdots \nu_p}(z)| \leq [K^{(p)}(\kappa)]^{n+1},$$

where

$$\begin{aligned} K^{(p)}(\kappa) &= \left(\frac{2}{\delta - \kappa}\right)^p K(\kappa), \quad \text{for} \quad n + 1 < p, \\ &= \frac{2K(\kappa)}{\delta - \kappa}, \qquad\quad \text{for} \quad n + 1 \geq p. \end{aligned} \tag{2.6}$$

Remark: It is not difficult to see that all polynomials in ϵ with coefficients in $C_b^\omega(\mathfrak{D})$ belong to $\mathfrak{H}(\mathfrak{D})$. If $f \in \mathfrak{F}(\mathfrak{D})$, we define

$$f^{[N]} = \sum_{n=0}^{N} f^{(n)} \epsilon^n,$$

i.e., $f^{[N]}$ stands for the power series f truncated at the $(N + 1)$th coefficient. Clearly, $f^{[N]} \in \mathfrak{H}(\mathfrak{D})$, even if $f \notin \mathfrak{H}(\mathfrak{D})$.

Now we introduce the set of q-tuples over $\mathfrak{F}(\mathfrak{D})$:

$$\mathfrak{F}_q(\mathfrak{D}) = \{f \,|\, f = (f_1, f_2, \cdots, f_q), f_\nu \in \mathfrak{F}(\mathfrak{D})\}.$$

If $f \in \mathfrak{F}_q(\mathfrak{D})$ we refer to f_ν as the νth component of f. $f^{(n)}$ again denotes the nth coefficient of f and is now a q-tuple of members of $C_b^\omega(\mathfrak{D})$. Clearly, $\mathfrak{F}_q(\mathfrak{D})$ can be looked upon as a module over $\mathfrak{F}(\mathfrak{D})$. Similarly, $\mathfrak{H}_q(\mathfrak{D})$, i.e., the set of q-tuples over $\mathfrak{H}(\mathfrak{D})$ can be considered as a module over $\mathfrak{H}(\mathfrak{D})$. By definition, we have $\mathfrak{F}_1(\mathfrak{D}) \equiv \mathfrak{F}(\mathfrak{D})$ and $\mathfrak{H}_1(\mathfrak{D}) \equiv \mathfrak{H}(\mathfrak{D})$. Of special significance is the case $q = m$.

The characteristic constants of $f \in \mathfrak{H}_q(\mathfrak{D})$ are defined by

$$\delta(f) = \min_{\nu=1,\cdots,q} \delta(f_\nu)$$

and

$$K(f) = \max_{\nu=1,\cdots,q} K(f_\nu),$$

respectively.

For later reference we present the following trivial result in form of a lemma.

Lemma 1: Let $f \in \mathfrak{H}_q(\mathfrak{D})$ and $f^{[N]} \doteq 0$ for some $N = -1, 0, 1, 2, \cdots$ (we define $f^{[-1]} = f^{(-1)} \equiv 0$). Let δ and $K(\kappa)$ be the characteristic constants of f; then $|f(z, \epsilon)|$ is majorized on the set (2.3) by

$$K(\kappa) \frac{[K(\kappa)\epsilon]^{N+1}}{1 - \epsilon K(\kappa)} \leq \frac{K(\kappa)}{1 - \epsilon K(\kappa)}.$$

Let $\mathbf{e}$ be the trivial function

$$\mathbf{e}(z, \epsilon) = z. \tag{2.7}$$

Clearly,

$\mathbf{e} \in \mathfrak{H}_m(\mathfrak{D})$ if and only if $\mathfrak{D}$ is bounded.

In any case, we define

$$\mathfrak{F}_m(\mathfrak{D}) = \{\mathbf{f} \,|\, \mathbf{f} = \mathbf{e} + \epsilon f, f \in \mathfrak{F}_m(\mathfrak{D})\}$$

and similarly

$$\mathfrak{H}_m(\mathfrak{D}) = \{\mathbf{f} \,|\, \mathbf{f} = \mathbf{e} + \epsilon f, f \in \mathfrak{H}_m(\mathfrak{D})\}.$$

If $f \in \mathfrak{F}_q(\mathfrak{D})$ and $\mathbf{g} = \mathbf{e} + \epsilon g$, $g \in \mathfrak{F}_m(\mathfrak{D})$, we define a composition $f \circ \mathbf{g}$ which also belongs to $\mathfrak{F}_q(\mathfrak{D})$ by

$$(f \circ \mathbf{g})(z, \epsilon) \equiv \sum_{p=0}^{\infty} \frac{\epsilon^p}{p!} \sum_{\nu_1, \cdots, \nu_p = 1}^{m} f_{,\nu_1 \cdots \nu_p} g_{\nu_1} \cdots g_{\nu_p}, \tag{2.8}$$

where $f_{,\nu_1 \cdots \nu_p}$ was defined above. This composition is the natural extension of the usual composition of analytic functions to formal power series (see Lemma 2).

We recognize from the definition (2.8) that the sum on the right side for each coefficient $(f \circ \mathbf{g})^{(n)}$ is actually finite. We obtain, for $n = 0, 1, 2, \cdots$,

$$\begin{aligned} (f \circ \mathbf{g})^{(0)} &= f^{(0)}, \\ (f \circ \mathbf{g})^{(1)} &= f^{(1)} + \sum_{\nu=1}^{m} f^{(0)}_{,\nu} g^{(0)}_\nu, \\ (f \circ \mathbf{g})^{(2)} &= f^{(2)} + \sum_{\nu=1}^{m} (f^{(0)}_{,\nu} g^{(1)}_\nu + f^{(1)}_{,\nu} g^{(0)}_\nu) \\ &\quad + \tfrac{1}{2} \sum_{\nu,\mu=1}^{m} f^{(0)}_{,\nu\mu} g^{(0)}_\nu g^{(0)}_\mu. \end{aligned} \tag{2.9}$$

Lemma 2: (i) $(f \circ \mathbf{g})^{(n)}$ only contains coefficients $g^{(p)}$ with $p \leq n - 1$. (ii) If $f \in \mathfrak{H}_q(\mathfrak{D})$, $\mathbf{g} \in \mathfrak{H}_m(\mathfrak{D})$

$$[\text{i.e., } \mathbf{g} = \mathbf{e} + \epsilon g, \quad g \in \mathfrak{H}_m(\mathfrak{D})],$$

then $f \circ \mathbf{g}$ also belongs to $\mathfrak{H}_q(\mathfrak{D})$.

Proof: (i) is an immediate consequence of the definition of $f \circ \mathbf{g}$. Under the condition (ii), f and $\mathbf{g}$ stand for at most $2m$ holomorphic functions where (z, ϵ) varies in a domain of the kind (2.3), and therefore (2.8) coincides with the usual composition

$$f(\mathbf{g}(z, \epsilon), \epsilon) = f(z + \epsilon g(z, \epsilon), \epsilon) \tag{2.10}$$

wherever the latter is defined.

To prove (ii), it is therefore sufficient to show that there exist positive constants δ and $K(\kappa)$ such that (2.10) is holomorphic in a domain of the kind (2.3) and is majorized in each $\mathfrak{D}_\kappa$, $0 < \eta < \delta$, by

$$K(\kappa)/[1 - \epsilon K(\kappa)].$$

Let δ_1 be the minimum of all δ's and $K_1(\kappa)$ the maximum of all K's over the set of components of f and g and put

$$\delta = \tfrac{1}{2}\delta_1, \quad K(\kappa) = \max\{[\rho(\kappa)]^{-1}, K_1(2\kappa)\},$$

where $\rho(\kappa)$ is defined to be a real positive number such that for $z \in \mathfrak{D}_\kappa$ and $\epsilon < \rho(\kappa)$ we have

$$|\epsilon g(z, \epsilon)| < \frac{\epsilon K_1(\kappa)}{1 - \epsilon K_1(\kappa)} < \kappa.$$

Then it follows for $z \in \mathfrak{D}_\kappa$, $\epsilon < 1/K(\kappa)$, $0 < \kappa < \delta$, that

$$\mathbf{g}(z, \epsilon) \in \mathfrak{D}_{2\kappa} \subset \mathfrak{D}_\delta.$$

Thus $(f \circ \mathbf{g})(z, \epsilon)$ is holomorphic in each $\mathfrak{D}_\kappa$, $0 < \kappa < \delta$, and majorized by

$$K(\kappa)/[1 - \epsilon K(\kappa)].$$

The lemma is proved.

Remark: Because the composition (2.8) coincides with the usual composition under the assumption (ii) of Lemma 2, it is customary to use the notation (2.10) instead of $(f \circ \mathbf{g})(z, \epsilon)$ even if f and g are only formal power series.

$\mathfrak{H}_m(\mathfrak{D})$ and $\mathfrak{F}_m(\mathfrak{D})$ have an important algebraic structure which is revealed in the following lemma.

Lemma 3: $\mathfrak{H}_m(\mathfrak{D})$ and $\mathfrak{F}_m(\mathfrak{D})$ are groups with unit element $\mathbf{e}$ with respect to the composition

$$\mathbf{f} \circ \mathbf{g} = \mathbf{g} + \epsilon(f \circ \mathbf{g}).$$

Here

$$\mathbf{f} = \mathbf{e} + \epsilon f \quad \text{and} \quad \mathbf{g} = \mathbf{e} + \epsilon g, \quad f, g \in \mathfrak{H}_m(\mathfrak{D}).$$

Proof: The fact that $\mathfrak{H}_m(\mathfrak{D})$ is a semigroup is a corollary of the statement (ii) in Lemma 2. If $\mathbf{f}$, $\mathbf{g} \in \mathfrak{H}_m(\mathfrak{D})$, the composition $\circ$ coincides with the usual composition of holomorphic functions which is known to be associative. Associativity being a purely algebraic property, it extends from $\mathfrak{H}_m(\mathfrak{D})$ to $\mathfrak{F}_m(\mathfrak{D})$. It follows that also $\mathfrak{F}_m(\mathfrak{D})$ is a semigroup. (The fact that $\mathbf{e}$ plays the role of the unit element is trivial.) We shall prove now the existence of an inverse. This is simple for $\mathfrak{F}_m(\mathfrak{D})$. It follows from the fact that the equation

$$g(z, \epsilon) + f(z + \epsilon g, \epsilon) = 0,$$

i.e.,

$$g^{(n)} + (f \circ g)^{(n)} = 0,$$

allows a recursive determination of the coefficients due to the statement (i) of Lemma 2. Because the calculation of $(f \circ g)^{(n)}$ involves only finitely many differentiations and ring operations, the reader easily constructs an inductive proof of the fact that $g^{(n)} \in C_b^\omega(\mathfrak{D})$.

To show that $\mathbf{f} = \mathbf{e} + \epsilon f$ has an inverse in $\mathfrak{H}_m(\mathfrak{D})$, let δ and $K(\kappa)$ be the characteristic constants of f, and define

$$\rho(\kappa) = \frac{\kappa}{1 + \kappa K(2\kappa)} \frac{1}{}.$$

The function spaces

$$\mathfrak{A}_\kappa = \{g \mid g = \text{analytic in the set } \mathfrak{D}_\kappa \times [\epsilon < \rho(\kappa)] \text{ and continuous on its closure}\}$$

equipped with the supnorm, which we denote by $\|g\|_\kappa$, are Banach spaces. The ball $\mathfrak{B}_\kappa$ of radius $\kappa/\rho(\kappa)$ is a compact and convex set of each space $\mathfrak{A}_{\kappa'}$ with

$$0 < \kappa' < \kappa.$$

Therefore, if we can prove that the map T_f defined by

$$T_f g = -f \circ (\mathbf{e} + \epsilon g)$$

maps $\mathfrak{B}_\kappa$ into itself (for some $\kappa > 0$), T_f will have a fixed point in each space $\mathfrak{A}_{\kappa'}$, $0 < \kappa' < \kappa$, by Schauder's fixed point principle,[7] and the function

$$\mathbf{g}_* = \mathbf{e} + \epsilon g_*$$

corresponding to the fixed point g_* will represent the inverse of $\mathbf{f}$ in $\mathfrak{H}_m(\mathfrak{D})$. The coefficients of g_* are the same as those obtained for g in the recursion procedure described above because the latter is unique.

It remains to show that $\mathfrak{B}_\kappa$ is mapped into itself by T_f. Let κ be any real number with the property $0 < \kappa < \frac{1}{2}\delta$; $g \in \mathfrak{B}_\kappa$ implies $\|g\|_\kappa \leq \kappa/\rho(\kappa)$. Hence

$$\|\epsilon g\|_\kappa < \kappa,$$

i.e.,

$$\mathbf{g}(z, \epsilon) \in \mathfrak{D}_{2\kappa} \subset \mathfrak{D}_\delta$$

for

$$z \in \mathfrak{D}_\kappa, \quad \epsilon < \rho(\kappa).$$

Hence, according to (2.5) (or Lemma 1), we have

$$\|T_f g\|_\kappa < \frac{K(2\kappa)}{1 - \rho(\kappa)K(2\kappa)} = \frac{\kappa}{\rho(\kappa)}.$$

The lemma is proved.

Remark: In the sequel we denote the inverse of $\mathbf{f}$ in $\mathfrak{F}_m(\mathfrak{D})$ [or $\mathfrak{H}_m(\mathfrak{D})$] by $\mathbf{f}_{-1}$ and correspondingly its components by $\mathbf{f}_{-1\nu}$, $\nu = 1, 2, \cdots, m$.

We still have not revealed all the algebraic structure of $\mathfrak{F}_m(\mathfrak{D})$. We are led to a structure of a different kind if we interpret the members of $\mathfrak{F}_m(\mathfrak{D})$ geometrically. Obviously they may be considered as formal "vector fields" over some neighborhood of $\mathfrak{D}$, whereas, if $f \in \mathfrak{H}_m(\mathfrak{D})$, f can be interpreted as a 1-parameter family of analytic vector fields.

It is natural to associate with $f \in \mathfrak{F}_m(\mathfrak{D})$ and $g \in \mathfrak{F}_q(\mathfrak{D})$ another member of $\mathfrak{F}_q(\mathfrak{D})$, namely

$$(f \vdash g) \in \mathfrak{F}_q(\mathfrak{D})$$

defined by

$$(f \vdash g)_\mu = \sum_{\nu=1}^{m} f_\nu g_{\mu,\nu}.$$

It can be interpreted as the derivative of g in the direction f. (The symbol $\vdash$ was taken from Ref. 8, where it is used in a similar but more general context.) Finally, we may define

$$[f, g] = f \vdash g - g \vdash f, \quad f, g \in \mathfrak{F}_m(\mathfrak{D}), \quad (2.11)$$

and in this way we imprint on $\mathfrak{F}_m(\mathfrak{D})$ the new structure of a Lie algebra.

Now we are in the position to define a useful partition of $\mathfrak{F}_m(\mathfrak{D})$ into equivalence classes. Let $f, g \in \mathfrak{F}_m(\mathfrak{D})$. We write

$$f \sim g \bmod \mathfrak{F}_m(\mathfrak{D}), \quad (2.12)$$

if there exists

$$\mathbf{W} \in \mathfrak{F}_m(\mathfrak{D})$$

such that

$$f \vdash \mathbf{W} = g \circ \mathbf{W}.$$

To see that this is an equivalence relation and at the same time to get some idea about its usefulness, consider the two formal differential equations

$$\dot{z} = f(z, \epsilon),$$
$$\dot{\zeta} = g(\zeta, \epsilon)$$

and assume that they are transformed into each other by a transformation

$$z = \mathbf{W}(\zeta, \epsilon)$$

with the inverse

$$\zeta = \mathbf{W}_{-1}(z, \epsilon).$$

One easily checks that this implies

$$g \vdash \mathbf{W} = f \circ \mathbf{W},$$
$$f \vdash \mathbf{W}_{-1} = g \circ \mathbf{W}_{-1}, \quad (2.13)$$

i.e., formulas which show the symmetry of the relation (2.12). The transitivity of the relation can be shown in a similar way, and the reflexivity is trivial.

By a q-fold formal integral or asymptotic invariant of the system

$$\dot{z} = f(z, \epsilon), \quad f \in \mathfrak{F}_m(\mathfrak{D}),$$

we mean any member $J \in \mathfrak{F}_q(\mathfrak{D})$ such that

$$\dot{J} = f \vdash J = 0 \quad (2.14)$$

and rank $(J_{\nu,\mu})_{\epsilon=0} = q$.

In any case, we have

$$\begin{aligned} f \vdash J^{[N]} &= (f - f^{[N]}) \vdash J^{[N]} + f^{[N]} \vdash J^{[N]} \\ &= \mathfrak{O}(\epsilon^{N+1}) + (f^{[N]} \vdash J^{[N]})^{[N]} \\ &= [f \vdash J]^{[N]} + \mathfrak{O}(\epsilon^{N+1}). \end{aligned}$$

Hence (2.14) implies

$$f \vdash J^{[N]} = \mathfrak{O}(\epsilon^{N+1}),$$

i.e., a q-fold formal integral defines q (hierarchies of) asymptotic invariants. For the notion of an asymptotic invariant, see Refs. 3, 9, and 10.

Lemma 4: The system of differential equations

$$\dot{z} = f(z, \epsilon),$$

where

$$f \in \mathfrak{F}_m(\mathfrak{D}) \quad \text{and} \quad f_\nu^{(0)} = 0, \quad \nu = 1, 2, \cdots, q,$$

has a q-fold formal integral if and only if

$$f \sim g \bmod \mathfrak{F}_m(\mathfrak{D}),$$

where the first q components of g vanish.

Proof: Let $J \in \mathfrak{F}_q(\mathfrak{D})$ be a q-fold formal integral. Because $f_\nu^{(0)} = 0$, it is no loss of generality to assume $J_\nu^{(0)} = \mathbf{e}_\nu$ for $\nu = 1, \cdots, q$.

Hence there exists $V \in \mathfrak{F}_m(\mathfrak{D})$ such that $\mathbf{V} \in \mathfrak{F}_m(\mathfrak{D})$ defined by

$$\begin{aligned} \mathbf{V}_\nu &= J_\nu, \quad \nu = 1, 2, \cdots, q, \\ &= \mathbf{e}_\nu, \quad \nu = q+1, \cdots, m \end{aligned}$$

has a representation

$$\mathbf{V} = \mathbf{e} + \epsilon V.$$

We define

$$g = (f \vdash \mathbf{V}) \circ \mathbf{V}_{-1}.$$

Hence,

$$g_\nu = (f \vdash \mathbf{V}_\nu) \circ \mathbf{V}_{-1} \text{ vanishes for } \nu = 1, 2, \cdots, q.$$

The inverse statement is an immediate consequence of the second equation of (2.13) which shows

$$g_\nu = 0, \quad \nu = 1, 2, \cdots, q, \quad \text{implies } f \vdash \mathbf{W}_{-1\nu} = 0, \quad \nu = 1, 2, \cdots, q.$$

This proves the lemma.

Closely related to the notion of an asymptotic integral is the notion of an asymptotic surface. Such a surface (or, better, hierarchy of surfaces in analogy to the hierarchy of asymptotic invariants) may be

constructed if the unperturbed system has an invariant surface. To discuss this notion in some detail, assume now that $\mathfrak{D}$ is a closed domain of R^n. Let $a = (a_1, \cdots, a_q)$ be some constant vector of R^q. We define

$$\mathfrak{D}^{(a)} = \{z \mid z \in \mathfrak{D}; z_1 = a_1, \cdots, z_q = a_q\}.$$

Assume that a is chosen in such a way that $\mathfrak{D}^{(a)}$ contains at least one interior point of $\mathfrak{D}$. Then $\mathfrak{D}^{(a)}$ represents an $(m-q)$-dimensional submanifold which is homeomorphic to a closed domain of R^{m-q}.

Now let $C_b^{\omega a}(\mathfrak{D})$ be the subring of functions of $C_b^{\omega}(\mathfrak{D})$ which vanish on $\mathfrak{D}^{(a)}$.

Assume that A is an m-tuple of members of $C_b^{\omega}(\mathfrak{D})$ such that

$$A_\nu \in C_b^{\omega a}(\mathfrak{D}), \quad \nu = 1, 2, \cdots, q.$$

Then the system of differential equations

$$\dot{z} = A(z)$$

has an $(m-q)$-dimensional invariant submanifold $\mathfrak{D}^{(a)}$. [Actually, by definition of $C_b^{\omega a}(\mathfrak{D})$ this manifold has an extension to $\mathfrak{D}_\delta^{(a)}$, $\delta = \delta(A)$.]

This means that any integral curve having a point in common with $\mathfrak{D}^{(a)}$ lies completely in $\mathfrak{D}^{(a)}$ (for all values of t for which it exists).

Let $\mathfrak{F}^{(a)}(\mathfrak{D})$ be the ring of formal power series over $C_b^{\omega a}(\mathfrak{D})$ and consider the perturbed system

$$\dot{z} = f(z, \epsilon), \quad f^{(0)} = A,$$

where $f \in \mathfrak{F}_m(\mathfrak{D})$ [in practical applications we usually even have $f \in \mathfrak{H}_m(\mathfrak{D})$].

We will say that the system has an asymptotic invariant surface if and only if

$$f \sim g \bmod \mathfrak{F}_m(\mathfrak{D}),$$

where

$$g_\nu \in \mathfrak{F}^{(a)}(\mathfrak{D}), \quad \nu = 1, 2, \cdots, q.$$

The importance of this notion can be seen in the following way. Let $\zeta_{(N)}(t, \epsilon)$ be a solution of

$$\dot{\zeta} = g^{[N]}(\zeta, \epsilon)$$

which stays in $\mathfrak{D}^{(a)}$. The corresponding function

$$z_{(N)}(t, \epsilon) = \mathbf{W}^{[N]}(\zeta_{(N)}(t, \epsilon), \epsilon)$$

stays in the analytic submanifold

$$\mathfrak{D}^{(a;N)} = \{z \mid z = \mathbf{W}^{[N]}(\zeta, \epsilon); \zeta \in \mathfrak{D}^{(a)}\}.$$

Now by assumption we have for all $\zeta \in \mathfrak{D}^{(a)}$

$$(f \vdash \mathbf{W}_{-1\nu}) \circ \mathbf{W} = 0, \quad \nu = 1, 2, \cdots, q.$$

Hence we have on $\mathfrak{D}^{(a;N)}$

$$f \vdash \mathbf{W}_{-1\nu}^{[N]}(z, \epsilon) = \mathfrak{O}(\epsilon^{N+1}), \quad \nu = 1, 2, \cdots, q;$$

the functions $\mathbf{W}_{-1\nu}^{[N]}(z, \epsilon)$, if restricted to the submanifold $\mathfrak{D}^{(a;N)}$, are asymptotically invariant under the flow induced by the vector field f in $\mathfrak{D}$. (They are "slowly varying" with time.)

It is obvious from our definitions that a q-fold formal integral defines a q-parametric family of $(m-q)$-dimensional asymptotic surfaces.

3. THE APPROXIMATION OF EXACT SOLUTIONS WITH HELP OF SOLUTIONS OF TRUNCATED FORMALLY EQUIVALENT DIFFERENTIAL EQUATIONS

In this section we define the approximations $z_{(N)}(t, \epsilon)$ and $\tilde{z}_{(N)}(t, \epsilon)$ (which can be looked upon as Nth and $(N - \frac{1}{2})$th approximations) of the exact solution $z(t, \epsilon)$ of a differential equation

$$\dot{z} = f(z, \epsilon), \quad f \in \mathfrak{H}_m(\mathfrak{D}).$$

These approximations are essentially obtained by composing an appropriate solution of a system that is up to order ϵ^N equivalent to the given one with the suitably truncated transformation connecting the two systems. Depending on whether the latter is truncated in such a way as to include powers of ϵ up to order $N-1$ or N, we obtain $\tilde{z}_{(N)}(t, \epsilon)$ or $z_{(N)}(t, \epsilon)$, respectively.

With the general methods developed so far we shall give a proof of a theorem of which several special cases are already known.[1,2,11,12] The essential content of the theorem is that, for solutions $z(t, \epsilon)$ which stay in the domain of analyticity of $f(z, \epsilon)$ for all t large enough, we have

$$|z(t, \epsilon) - \tilde{z}_{(N)}(t, \epsilon)| = \mathfrak{O}(\epsilon^N)$$

for a time interval of length L/ϵ, where L may be arbitrarily large if only ϵ is sufficiently small. This is done in Theorem 1.

Theorem 2 discusses some cases of differential equations and their solutions which allow an approximation of a similar kind as described in Theorem 1, but with the difference that it is valid for a time interval of infinite length.

Because the approximations are only asymptotic, the reader should be warned that the calculation of high-order approximations does not make sense as long as the convergence of the formal expansions is not guaranteed, i.e., as long as we use only formally equivalent and not analytically equivalent systems of differential equations for our construction of the approximation.

Theorem 1: Let

$$f \in \mathfrak{H}_m(\mathfrak{D}) \quad \text{and} \quad f \sim g \bmod \mathfrak{F}_m(\mathfrak{D}),$$

i.e.,

$$g \vdash \mathbf{W} = f \circ \mathbf{W} \quad \text{for some} \quad \mathbf{W} \in \mathfrak{F}_m(\mathfrak{D}),$$

and

$$f^{(0)} = g^{(0)} = A, \quad A \text{ independent of } z.$$

Furthermore let $\zeta_{(N)}(t, \epsilon)$ [with νth component $\zeta_{(N)\nu}(t, \epsilon)$] be a solution of the system

$$\dot{\zeta} = g^{[N]}(\zeta, \epsilon) \tag{3.1}$$

with the property that, for $\epsilon < \rho_1$ and all $t > t_0$, $\zeta_{(N)}(t, \epsilon)$ stays in the domain $\mathfrak{D}_{\kappa_1}$. Here ρ_1, κ_1, and t_0 are real numbers with the properties

$$0 < \rho_1 < 1/K(\kappa_1), \quad 0 < \kappa_1 < \delta,$$

and δ and K are the two characteristic constants of the function:

$$R_{(N)} \equiv g^{[N]} \vdash \mathbf{W}^{[N]} - f \circ \mathbf{W}^{[N]}.$$

Then the function

$$z_{(N)}(t, \epsilon) \equiv \mathbf{W}^{[N]}(\zeta_{(N)}(t, \epsilon), \epsilon)$$

is an $\mathfrak{O}(\epsilon^{N+1})$-approximate solution of

$$\dot{z} = f(z, \epsilon), \tag{3.2}$$

i.e.,

$$\frac{dz_{(N)}(t, \epsilon)}{dt} - f(z_{(N)}(t, \epsilon), \epsilon) = \mathfrak{O}(\epsilon^{N+1}).$$

(Compare Ref. 13, p. 3.) Moreover, there exists an exact solution $z(t, \epsilon)$ of (3.2) such that $z_{(N)}(t, \epsilon)$ and also the function

$$\tilde{z}_{(N)} \equiv \mathbf{W}^{[N-1]}(\zeta_{(N)}(t, \epsilon), \epsilon)$$

deviate from $z(t, \epsilon)$ on each set

$$\{(t, \epsilon) \mid t_0 < t < t_0 + L/\epsilon; \epsilon < \rho(L)\} \tag{3.3}$$

only by a quantity of order ϵ^N, i.e., we have on each set (3.3)

$$|z(t, \epsilon) - z_{(N)}(t, \epsilon)| = \mathfrak{O}(\epsilon^N),$$

$$|z(t, \epsilon) - \tilde{z}_{(N)}(t, \epsilon)| = \mathfrak{O}(\epsilon^N).$$

In (3.3) $\rho(L)$ is a certain nonincreasing function of L.

Remark: Here the symbol $\cdots = \mathfrak{O}(\epsilon^N)$ abbreviates the statement: There exists a real positive constant $C_N(L)$ such that

$$|\cdots| \leq C_N(L)\epsilon^N.$$

Proof: Let $R = \mathrm{g} \vdash \mathbf{W} - f \circ \mathbf{W}$. By assumption $R \equiv 0$ (i.e., all the coefficients of R are zero). Recall the definition $R_{(N)}$ in the statement of the theorem. Lemma 2 implies $R_{(N)}^{[N]} = R^{[N]} = 0$ and $R_{(N)} \in \mathfrak{H}_m(\mathfrak{D})$. According to Lemma 1, $|R_{(N)}|$ is majorized on $\mathfrak{D}_{\kappa_1}$ by

$$\epsilon^{N+1}[K(\kappa_1)]^{N+2}/[1 - \epsilon K(\kappa_1)],$$

i.e., we have for $\epsilon < \rho_1$

$$|R_{(N)}(\zeta_{(N)}(t, \epsilon), \epsilon)| \leq C_N\epsilon^{N+1}$$

with some real positive constant C_N. But obviously

$$\frac{dz_{(N)}(t, \epsilon)}{dt} - f(z_{(N)}(t, \epsilon), \epsilon) = R_{(N)}(\zeta_{(N)}(t, \epsilon), \epsilon).$$

The combination of the last two relations yields the first statement of the theorem.

Using the fact that $f^{(0)}(z) = A$ is independent of z, we conclude from Lemma 1 that $\partial f(z, \epsilon)/\partial z$ on $\{(z, \epsilon): z \in \mathfrak{D}_{\kappa_2}, \epsilon < 1/k\}$ is majorized by ϵk, where κ_2 is some number with the property

$$\kappa_1 < \kappa_2 < \delta \quad \text{and} \quad k = \frac{4}{\delta - \kappa_2} K(\kappa_2).$$

[Observe that, by definition, $\delta < \delta(f)$, $K(\kappa) > K(\kappa, f)$, and also

$$K\left(\kappa, \frac{\partial f}{\partial z}\right) \equiv \max_{\nu,\mu} K(\kappa, f_{\nu,\mu}) < \frac{2}{\delta - \kappa} K(\kappa, f).]$$

Let $z(t, \epsilon)$ be an exact solution of (3.2) which for $t = t_0$ differs from $z_{(N)}(t_0, \epsilon)$ by a quantity of order ϵ^N.

Then by a theorem about ϵ-approximate solutions (see Theorem 2.1 of Ref. 13) there exists a real positive constant b_N such that

$$|z(t, \epsilon) - z_{(N)}(t, \epsilon)| \leq b_N\epsilon^N + \frac{C_N\epsilon^N}{k}(e^{\epsilon k(t-t_0)} - 1).$$

This inequality is true for all $t > t_0$ and all $\epsilon < \rho_2 \equiv \min(1/k, \rho_1)$ for which $z(t, \epsilon)$ does not leave $\mathfrak{D}_{\kappa_2}$.

We now shall show that there exists a function $\rho = \rho(L)$ such that the above inequality is true on the set (3.3). Suppose this would not be the case. Then there would exist a constant L_0 and a function $T(\epsilon)$ such that

$$T(\epsilon) < t_0 + L_0/\epsilon$$

and such that for arbitrarily small ϵ we would have

$$|z(T(\epsilon), \epsilon) - \zeta_{(N)}(T(\epsilon), \epsilon)| = \kappa_3, \tag{3.4}$$

where κ_3 is some fixed number with the property

$$\kappa_1 + \kappa_3 < \kappa_2.$$

In particular,

$$z(T(\epsilon), \epsilon) \in \mathfrak{D}_{\kappa_2}.$$

Hence, if

$$\mathbf{W} = \mathbf{e} + \epsilon W,$$

we have

$$|z(T(\epsilon), \epsilon) - \zeta_{(N)}(T(\epsilon), \epsilon)| \leq [b_N + (C_N/k)(e^{kL_0} - 1)]\epsilon^N + \epsilon\,|W^{[N-1]}[\zeta_{(N)}(T(\epsilon), \epsilon), \epsilon]|.$$

The right side is of order ϵ, thereby contradicting (3.4) and proving the theorem for $z_{(N)}(t, \epsilon)$. That the assertion of the theorem is also true for $\tilde{z}_{(N)}(t, \epsilon)$ follows easily from the fact that $z_{(N)}$ and $\tilde{z}_{(N)}$ differ by a quantity of order ϵ^N. The theorem is proved.

Remark: In most applications of Theorem 1 one would like to approximate a certain solution $z(t, \epsilon)$ of (3.2) characterized by the initial condition $z(t_0, \epsilon) = a$. The question naturally arises: What solutions $\zeta_{(N)}(t, \epsilon)$ of (3.1) have the property that

$$z_{(N)}(t, \epsilon) \equiv \mathbf{W}^{[N]}(\zeta_{(N)}(t, \epsilon), \epsilon)$$

approximates the particular solution $z(t, \epsilon)$ in the sense of the theorem? The only thing we have to observe in choosing $\zeta_{(N)}(t, \epsilon)$ is that

$$|z_{(N)}(t_0, \epsilon) - a|$$

is of order ϵ^N. This, for example, is the case if we choose $\zeta_{(N)}(t, \epsilon)$ such that

$$\zeta_{(N)}(t_0, \epsilon) = \mathbf{W}^{[N]}_{-1}(a, \epsilon).$$

Indeed, we have

$$|z_{(N)}(t_0, \epsilon) - a| = |(\mathbf{W}^{[N]} \circ \mathbf{W}^{[N]}_{-1})(a, \epsilon) - a|.$$

But the right side clearly is of order ϵ^N.

The estimates of Theorem 1 are the best possible for the wide class of differential equations and their solutions that we consider. However, this does not mean that we cannot get better estimates if we restrict the class of differential equations or if we are only concerned with the approximation of a certain class of solutions of these differential equations.

Coffey gives an example where

$$|z(t, \epsilon) - z_{(1)}(t, \epsilon)|$$

is of order ϵ for all t: $-\infty < t < \infty$. Let us have a closer look at his system of differential equations under investigation. His equation [Ref. 3, formula (4.43), p. 435] is trivially equivalent to the system

$$\begin{aligned} \dot{u} &= \epsilon u(1 - u), \\ \dot{z}_1 &= z_2, \\ \dot{z}_2 &= -\omega^2(u) z_1, \end{aligned}$$

where $\omega(u)$ and $\omega^{-1}(u)$ are holomorphic in a certain domain of the u plane containing $u = 0$ and $u = 1$. By the change of variables

$$\begin{aligned} z_1 &= \frac{v}{[\omega(u)]^{\frac{1}{2}} u(1 - u)} \sin x, \\ z_2 &= \frac{v[\omega(u)]^{\frac{1}{2}}}{u(1 - u)} \cos x, \end{aligned}$$

the system is brought into the form

$$\begin{aligned} \dot{u} &= \epsilon u(1 - u), \\ \dot{v} &= \epsilon v(1 - u) - \epsilon v u - \tfrac{1}{2}\epsilon(\ln \omega)' u(1 - u) \cos 2x, \\ \dot{x} &= \omega(u) + \tfrac{1}{2}\epsilon(\ln \omega)' u(1 - u) \sin 2x, \end{aligned} \tag{3.5}$$

where the prime denotes d/du. We see that the system (3.5) has two "closed orbits" with equations $u = 0$, $v = 0$ and $u = 1$, $v = 0$. For $\epsilon > 0$ the first orbit is asymptotically stable for $t \to -\infty$, whereas the second has this property for $t \to \infty$. Observe that

$$u = \mathfrak{O}(e^{\epsilon t}), \quad v = \mathfrak{O}(e^{\epsilon t}), \quad \text{for} \quad t \to -\infty$$

and

$$1 - u = \mathfrak{O}(e^{-\epsilon t}), \quad v = \mathfrak{O}(e^{-\epsilon t}), \quad \text{for} \quad t \to +\infty$$

such that z_1 is well defined for both limits although the transformation $(z_1, z_2) \to (x, v)$ is singular for $u = 0$ and $u = 1$. It follows that at least those solutions of the original system that are under investigation in Ref. 3 can be constructed with the help of solutions of the asymptotically stable differential equation (3.5).

That in the case of asymptotically stable (or unstable) differential equations any perturbation theory which uses solutions of formally equivalent differential equations gives good approximations for an infinite time interval follows from the following considerations (see especially Theorem 2).

Before we state and prove the main result, we mention two lemmas (Lemmas 5 and 6). Lemma 5, whose proof is omitted, can be found, for example, in Ref. 14. The proof of Lemma 6 also follows standard arguments found in Refs. 13 and 14.

Lemma 5: If ψ is differentiable, then

$$\phi(t) \leq \psi(t) + \int_a^t \chi(s)\phi(s)\, ds, \quad \text{for} \quad t \geq a,$$

implies

$$\phi(t) \leq \psi(a) \exp \int_a^t \chi(u)\, du + \int_a^t \exp \int_s^t \chi(u)\, du \psi'(s)\, ds,$$

for $t \geq a$.

Lemma 6: Let $z = (x, y) \in R^r \times R^s$,

$$\mathfrak{D} = R^r \times \{0\}, \quad 0 = \text{origin of } R^s, \quad m = r + s,$$

$$f_\nu(z, \epsilon) = \begin{cases} \omega_\nu + \epsilon F_\nu(x, y, \epsilon), \\ \qquad \nu = 1, 2, \cdots, r \\ (\epsilon\Omega y)_\nu + \epsilon G_\nu(x, y, \epsilon), \\ \qquad \nu = r + 1, \cdots, m \end{cases} \in \mathfrak{H}(\mathfrak{D}), \tag{3.6}$$

where

$$F(x, 0, \epsilon) = 0, \quad G(x, 0, \epsilon) = 0, \quad \frac{\partial G}{\partial y}(x, 0, \epsilon) = 0,$$

and Ω is an $s \times s$ matrix with eigenvalues having negative real parts. If

$$z(t, \epsilon) = [x(t, \epsilon), y(t, \epsilon)]$$

is a real solution of

$$\dot{z} = f(z, \epsilon)$$

and κ is a sufficiently small real positive number, then $z(0, \epsilon) \in \mathfrak{D}_\kappa$ implies $z(t, \epsilon) \in \mathfrak{D}_\kappa$ for all $t > 0$. Moreover,

$$\lim_{t \to \infty} y(t, \epsilon) = 0.$$

Proof: Observe that, for z real, $z \in \mathfrak{D}_\kappa$ is equivalent to $|y| < \kappa$.

Choosing $k(\kappa) = 2K^{(2)}(\kappa, f)$ [see (2.6)], we have

$$|G(x, y, \epsilon)| \leq k(\kappa)\kappa\,|y|$$

for

$$z \in \mathfrak{D}_\kappa, \quad \epsilon < [k(\kappa)]^{-1}.$$

If for the sake of simplicity we make the additional assumption that Ω is diagonal, there exists a real positive number α such that for all $y \in R^s$

$$|e^{\epsilon\Omega t} y| \leq e^{-\epsilon\alpha t}|y|.$$

Thus by standard arguments[13,14] we have

$$e^{\epsilon\alpha t}\,|y(t, \epsilon)| \leq |y_0| + \epsilon k(\kappa)\kappa \int_0^t e^{\epsilon\alpha\tau}\,|y(\tau, \epsilon)|\,d\tau.$$

Applying Lemma 5 to this equation, we find

$$|y(t, \epsilon)| \leq |y_0|\, e^{-\epsilon\beta(\kappa)t},$$

where

$$\beta(\kappa) = \alpha - \kappa k(\kappa).$$

This is a relation from which all the statements of the lemma immediately follow.

Let f be defined as in Lemma 6. Assume $f \sim g$ mod $\mathfrak{F}_m(\mathfrak{D})$, where g is of the same kind as f (i.e., more exactly, g has also the form (3.6), but with F, G replaced by some formal power series).

Let $\zeta_{(N)}(t, \epsilon)$ and $z_{(N)}(t, \epsilon)$ be defined as in Theorem 1. Then $z_{(N)}(t, \epsilon)$ is clearly an exact solution of the equation

$$\dot{z} = f_{(N)}(z, \epsilon),$$

where

$$f_{(N)} = (g^{[N]} \vdash \mathbf{W}^{[N]}) \circ (\mathbf{W}^{[N]})_{-1}.$$

The components of $f_{(N)}$ have the form

$$f_{(N)\nu} = \begin{cases} \omega_\nu + \epsilon F_{(N)\nu}(x, y, \epsilon) \\ (\epsilon\Omega y)_\nu + \epsilon G_{(N)\nu}(x, y, \epsilon). \end{cases} \tag{3.7}$$

Here $F_{(N)}$, $G_{(N)}$ are functions of the same kind as F and G. In addition, we have $f - f_{(N)} = \mathfrak{O}(\epsilon^{N+1})$. More specifically it is not difficult to see that there exists a constant $k(\kappa)$ such that for $z, \zeta \in \mathfrak{D}_\kappa$

$$|F(x, y, \epsilon) - F_{(N)}(\xi, \eta, \epsilon)| \leq k(\kappa)[|y - \eta| + |y|\,|x - \xi| + \epsilon^N |y|] \tag{3.8}$$

and

$$|G(x, y, \epsilon) - G_{(N)}(\xi, \eta, \epsilon)| \leq k(\kappa)[|y - \eta| + \kappa\,|x - \xi| + \epsilon^N\kappa]\,|y|. \tag{3.9}$$

Here we can adapt $k(\kappa)$ in such a manner that it coincides with the constant of Lemma 6 denoted by the same symbol. (Observe that strictly speaking $k(\kappa)$ also depends on N. However, in our discussion we keep N fixed.) In the following we shall keep κ so small that $\beta - 2\kappa k(\kappa) = \alpha - 3\kappa k(\kappa) > 0$, and we assume that $\epsilon < 1/k(\kappa)$.

Theorem 2: Assume that ω is a constant r vector. For each solution $z(t, \epsilon)$ of

$$\dot{z} = f(z, \epsilon) \quad [f \text{ defined in (3.6)}],$$

with $z(0, \epsilon) \in \mathfrak{D}_\kappa$ and κ small enough, there exists a solution $z_{(N)}(t, \epsilon)$ of

$$\dot{z} = f_{(N)}(z, \epsilon) \quad [f_{(N)} \text{ defined in (3.7)}]$$

such that

$$\Delta \equiv |z(t, \epsilon) - z_{(N)}(t, \epsilon)| = \mathfrak{O}(\kappa\epsilon^N) \tag{3.10}$$

for all times $t \geq 0$.

Proof: Assume for the sake of simplicity that Δ vanishes for $t = 0$, i.e., $z_{(N)}(0, \epsilon) = z(0, \epsilon)$. Writing down the differential equations for $z(t, \epsilon)$ and $z_{(N)}(t, \epsilon)$ and using the inequalities (3.8) and (3.9), we easily deduce

$$\Delta_1 \leq \epsilon\kappa\left(\int_0^t \Delta_2\,d\tau + \kappa\int_0^t e^{-\epsilon\beta\tau}\Delta_1\,d\tau + \epsilon^N\kappa\int_0^t e^{-\epsilon\beta\tau}\,d\tau\right),$$

$$\Delta_2 e^{\epsilon\alpha t} \leq \epsilon\kappa k\int_0^t e^{(\alpha-\beta)\tau}\{\Delta_2 + \kappa\Delta_1 + \kappa\epsilon^N\}\,d\tau.$$

Here

$$\Delta_1 \equiv |x(t, \epsilon) - x_{(N)}(t, \epsilon)|,$$
$$\Delta_2 \equiv |y(t, \epsilon) - y_{(N)}(t, \epsilon)|.$$

The right sides of these inequalities are differentiable functions which we denote by ρ_1 and $\rho_2 e^{\epsilon\alpha t}$, respectively.

Then we have

$$\Delta_1 \leq \rho_1, \quad \Delta_2 \leq \rho_2$$

and therefore

$$\dot{\rho}_1 \leq \epsilon k\rho_2 + \epsilon k\kappa e^{-\epsilon\beta t}\rho_1 + \epsilon^{N+1}k\kappa e^{-\epsilon\beta t},$$

$$\dot{\rho}_2 \leq \epsilon k\kappa\rho_2 e^{-\epsilon\beta t} - \epsilon\alpha\rho_2 + \epsilon\kappa^2 k e^{-\epsilon\beta t}\rho_1 + \epsilon^{N+1}\kappa^2 k e^{-\epsilon\beta t}.$$

To simplify these inequalities, we use the abbreviation

$$\gamma = \epsilon k \kappa,$$

and we obtain

$$\dot{\rho}_1 \leq \gamma e^{-\epsilon\beta t}\rho_1 + \epsilon k \rho_2 + \epsilon^N \gamma e^{-\epsilon\beta t},$$

$$\frac{d}{dt}(\rho_2 e^{\epsilon\alpha t}) \leq \gamma\kappa\rho_1 + \gamma\rho_2 + \epsilon^N\gamma\kappa.$$

From the second relation we deduce

$$\rho_2 e^{\epsilon\beta t} \leq \rho_2 e^{\epsilon\alpha t} \leq \epsilon^N \gamma\kappa t + \int_0^t \chi(s) e^{\epsilon\beta s}\rho_2\, ds + \kappa\gamma\int_0^t \rho_1(s)\, ds,$$

where

$$\chi(s) = \gamma e^{-\epsilon\beta s},$$

and therefore

$$\int_s^t ds\chi(s) \leq \int_0^\infty dt\chi(t) = \frac{\kappa k}{\beta} \equiv \ln M.$$

Hence, according to Lemma 5, we have

$$\rho_2 e^{\epsilon\beta t} \leq M\gamma\kappa\int_0^t [\epsilon^N + \rho_1(s)]\, ds.$$

Using this result in the above inequality for $\dot{\rho}_1$, we find

$$\dot{\rho}_1 \leq \left(M\gamma^2\int_0^t [\epsilon^N + \rho_1(s)]\, ds + \gamma[\epsilon^N + \rho_1(t)]\right) e^{-\epsilon\beta t}.$$

A second integration yields

$$\rho_1(t) \leq \int_0^t d\tau e^{-\epsilon\beta\tau}\left(M\gamma^2\int_0^\tau [\epsilon^N + \rho_1(s)]\, ds + \gamma[\epsilon^N + \rho_1(\tau)]\right).$$

Define

$$\lambda(t) = \max_{0\leq s\leq t} \rho_1(s).$$

Then

$$\int_0^\tau [\epsilon^N + \rho_1(s)]\, ds \leq \tau[\epsilon^N + \lambda(\tau)] \leq \tau[\epsilon^N + \lambda(t)]$$

for $\tau \leq t$.

Hence

$$\rho_1(t) \leq \gamma\int_0^t d\tau e^{-\epsilon\beta\tau}(M\gamma\tau + 1)[\epsilon^N + \lambda(t)] < \left(\frac{\kappa k}{\beta}\right)\left(1 + \frac{\kappa k}{\beta}\right)[\lambda(t) + \epsilon^N](1 - e^{-\epsilon\beta t}).$$

It follows that

$$\lambda(t) \leq C[\lambda(t) + \epsilon^N],$$

where

$$C = \left(\frac{\kappa k}{\beta}\right)\left(1 + \frac{\kappa k}{\beta}\right)$$

or

$$\Delta_1 \leq \rho_1 \leq \lambda(t) \leq \frac{\epsilon^N C}{1 - C}$$

for all $t \geq 0$, a result which shows that

$$\Delta_1 = \mathfrak{O}(\epsilon^N\kappa)$$

as long as we keep κ so small that $C < 1$. But the last inequality is implied by the assumption about κ made above Theorem 2.

It is now an easy matter to prove that also Δ_2 is of order $\epsilon^N\kappa$, and, because

$$\Delta = \Delta_1 + \Delta_2,$$

the result (3.10) follows.

Remark 1: Actually our inequalities allow us to prove somewhat more, namely the fact that

$$\rho_1(t) = \mathfrak{O}(\kappa\epsilon^N(1 - e^{-\epsilon\beta t})),$$
$$\rho_2(t) = \mathfrak{O}(\kappa^2\epsilon^{N+1}te^{-\epsilon\beta t}).$$

These relations show that $x_{(N)}$ approximates x the worse the larger t becomes. But this is not so for the y approximation, which is the poorest for

$$t = 1/\epsilon\beta.$$

Of course, this result is not surprising in view of the fact that we approximated a solution which is asymptotically stable for $t \to \infty$.

Remark 2: Notice that we proved the result (3.10) under the assumption that to zeroth order the system has the form

$$\dot{x} = \omega, \quad \dot{y} = 0,$$

where ω is a constant r-vector.

The observation that in Coffey's system (3.5) ω is not constant but a function of u shows that our theorem must be slightly generalized in order to be applicable to the system (3.5).

Indeed, the following generalized version of Theorem 2 makes it understandable that Coffey's construction is asymptotic for all times $-\infty < t < +\infty$.

Theorem 2': Let all the assumptions of Theorem 2 be satisfied except for the condition of constancy of the r-vector ω being replaced by the following conditions: (i) y splits into (u, v) and correspondingly G into (H, K) and $G_{(N)}$ into $(H_{(N)}, K_{(N)})$ in such a way that $H_{(N)} \equiv H$ is a function of the variables u only (i.e., the variables u are not changed in the process of the construction of the formally equivalent system); (ii) ω is a function of the u's only.

Proof: If ω is a function of y, then the inequality (3.8) has to be supplemented by an estimate of the term

$$\omega(y) - \omega(\eta). \tag{3.11}$$

In general, such a term destroys the result of Theorem 2 by producing a term $\exp(C/\epsilon)$ on the right side of (3.10). However, under the condition (i) of Theorem 2′ we have

$$u_{(N)}(t, \epsilon) = u(t, \epsilon).$$

This conclusion together with the condition (ii) shows that the term (3.11) can again be dropped. Hence all the arguments which led us to the result (3.10) of Theorem 2 are also valid under the more general assumptions of Theorem 2′.

Remark 3: A slight modification of the proof shows that the result (3.10) is also true if the system to the zeroth order has the form

$$\dot{x} = \omega(y), \quad \dot{y} = \Omega y,$$

i.e., if we drop the ϵ in front of the matrix Ω in the vector field (3.6).

4. HOW TO CONSTRUCT FORMALLY EQUIVALENT VECTOR FIELDS

We mentioned already that by introducing the bracket by definition (2.11) we imprint on $\mathfrak{F}_m(\mathfrak{D})$ the structure of a Lie algebra. As usual in this connection we associate with each $f \in \mathfrak{F}_m(\mathfrak{D})$ a linear operator

$$\operatorname{ad} f : \mathfrak{F}_m(\mathfrak{D}) \to \mathfrak{F}_m(\mathfrak{D})$$

by

$$\operatorname{ad} fg = [f, g].$$

We have

$$\operatorname{ad} fg = \left[\sum_{n=0}^{\infty} f^{(n)}\epsilon^n, g\right] = \sum_{n=0}^{\infty} \epsilon^n [f^{(n)}, g]$$
$$= \sum_{n=0}^{\infty} \epsilon^n \operatorname{ad} f^{(n)} g.$$

In this sense each operator $\operatorname{ad} f$ is representable as an infinite formal power series. The Jacobi identity implies that $\operatorname{ad} f$ is a derivation of the Lie algebra $\mathfrak{F}_m(\mathfrak{D})$, i.e.,

$$\operatorname{ad} f[g, h] = [\operatorname{ad} fg, h] + [g, \operatorname{ad} fh].$$

We have now

$$\exp(\epsilon \operatorname{ad} f) \equiv \sum_{p=0}^{\infty} \frac{\epsilon^p}{p!} (\operatorname{ad} f)^p.$$

This definition makes sense because, as is easy to check, each coefficient of the power series

$$h = \exp(\epsilon \operatorname{ad} f) g$$

is expressible by finitely many ring operations and differentiations on the coefficients of f and g. We find for $n = 0, 1, 2$

$$h^{(0)} = g^{(0)},$$
$$h^{(1)} = g^{(1)} + [f^{(0)}, g^{(0)}],$$
$$h^{(2)} = g^{(2)} + [f^{(0)}, g^{(1)}] + [f^{(1)}, g^{(0)}] + \tfrac{1}{2}[f^{(0)}, [f^{(0)}, g^{(0)}]].$$

One easily verifies (see, e.g., Ref. 15)

$$[\exp(\epsilon \operatorname{ad} f) g, \exp(\epsilon \operatorname{ad} f) h] = \exp(\epsilon \operatorname{ad} f)[g, h].$$

Every $f \in \mathfrak{F}_m(\mathfrak{D})$ determines an automorphism of $\mathfrak{F}_m(\mathfrak{D})$ with respect to its structure as a Lie algebra. This is not only true for $\mathfrak{F}_m(\mathfrak{D})$ itself but for any subalgebra $\mathfrak{L}$ of $\mathfrak{F}_m(\mathfrak{D})$. The group of automorphisms generated by all the elements of the form

$$\exp(\epsilon \operatorname{ad} f), \quad f \in \mathfrak{L},$$

we denote by $\operatorname{Int} \mathfrak{L}$ and call the group of inner automorphisms of $\mathfrak{L}$.

All these notions are taken from the theory of finite-dimensional Lie algebras (see, e.g., Ref. 15). Infinite-dimensional formal Lie algebras have first been investigated by Sternberg.[5] Nonformal aspects of the theory are investigated and applied in papers by Moser.[4] All generalized perturbation theories of Kruskal's type are based on the following two theorems.

Theorem 3: Let $A = A^{(0)} \in \mathfrak{F}_m(\mathfrak{D})$ and $f \in \mathfrak{F}_m(\mathfrak{D})$. Assume

(i) $\mathfrak{F}_m(\mathfrak{D}) = \mathfrak{P} \oplus \hat{\mathfrak{L}}$ as a vector space,

(ii) $\operatorname{ad} A \mathfrak{F}_m(\mathfrak{D}) \supset \mathfrak{P}$.

Then

$$A + \epsilon f \sim A + \epsilon g \bmod \mathfrak{F}_m(\mathfrak{D}), \tag{4.1}$$

where

$$g \in \hat{\mathfrak{L}}.$$

Remark: The transformation establishing the equivalence (4.1) is only determined up to an element of the null space $\mathfrak{N}_A$ of $\operatorname{ad} A$. In most applications we shall discuss $\mathfrak{N}_A = \hat{\mathfrak{L}}$ and $\operatorname{ad} A \mathfrak{F}_m(\mathfrak{D}) = \mathfrak{P}$, but, as the theorem indicates, the desired conclusion (4.1) is true under less restrictive assumptions.

Proof: Let us assume (4.1) is true. Then with $\mathbf{W} = \mathbf{e} + \epsilon W$ we have

$$(A + \epsilon g) \vdash \mathbf{W} = (A + \epsilon f) \circ \mathbf{W},$$

i.e.,

$$A + \epsilon g + \epsilon(A \vdash W) + \epsilon^2 (g \vdash W) = A \circ \mathbf{W} + \epsilon(f \circ \mathbf{W})$$

or

$$g + [A, W] = F + \epsilon X \equiv Z, \tag{4.2}$$

where

$$X = \epsilon^{-1}[f \circ \mathbf{W} - f] + \epsilon^{-2}[A \circ \mathbf{W} - A - \epsilon W \vdash A] - g \vdash W.$$

Let $\hat{}$ denote the projection of an arbitrary element $f \in \mathfrak{F}_m(\mathfrak{D})$ onto $\hat{\mathfrak{L}}$ corresponding to the decomposition (i). Assuming $W^{[N-1]}$ and $g^{[N-1]}$ are known, then $X^{[N-1]}$ is known [see Lemma 2(i)], and hence $Z^{[N]}$ is known. Define

$$g^{(N)} = \hat{Z}^{(N)}.$$

Because $A \equiv A^{(0)}$, the following equation for $W^{(N)}$ results:

$$\text{ad } A W^{(N)} = Z^{(N)} - \hat{Z}^{(N)}. \tag{4.3}$$

According to assumption (ii), ad A is invertible on $\mathfrak{P}$, and therefore (4.3) can be solved for $W^{(N)}$. This construction shows that W is only determined up to an element of the null space of ad A.

Remark: There is an alternative procedure to construct $\mathbf{W}$ by which $\mathbf{W}$ is represented as an infinite product of successive transformations. In order to sketch this procedure, assume that

$$f - \hat{f} = \epsilon^{N-1} h + \mathfrak{O}(\epsilon^N).$$

By definition $h \in \mathfrak{P}$, and by (ii) there exists $W_{(N)}$ such that

$$[A, W_{(N)}] = \epsilon^{N-1} h.$$

Define

$$g = f - \epsilon^{N-1} h + \epsilon \hat{X}_{(N)},$$

where $X_{(N)}$ is defined as X above except that W is replaced by $W_{(N)}$. Then obviously $g - \hat{g} = \mathfrak{O}(\epsilon^N)$. In this way we recognize that $\mathbf{W}$ is obtained as an infinite product

$$\mathbf{W} = \prod_{N=1}^{\infty} (\circ \mathbf{W}_{(N)}).$$

As an application of Theorem 3 consider

$$\hat{\mathfrak{L}} = \{f \mid f \in \mathfrak{F}_m(\mathfrak{D}), f_\nu = 0; \nu = 1, 2, \cdots, q\},$$

$$\mathfrak{P} = \{f \mid f \in \mathfrak{F}_m(\mathfrak{D}), f_\nu = 0; \nu = q + 1, \cdots, m\}.$$

Assume $A = A^{(0)} \in \hat{\mathfrak{L}}$ and ad A is invertible on $\mathfrak{P}$. Then

$$A + \epsilon f \sim A + \epsilon g \bmod \mathfrak{F}_m(\mathfrak{D}),$$

where

$$g \in \hat{\mathfrak{L}}$$

and according to Lemma 4 the inverse $\mathbf{W}_{-1}$ of the transformation constructed according to the scheme of Theorem 3 defines a q-fold formal integral of the system

$$\dot{z} = A(z) + \epsilon f(z, \epsilon).$$

The next theorem deals with the question of equivalence of vector fields belonging to a subalgebra $\mathfrak{L}$ of $\mathfrak{F}_m(\mathfrak{D})$ with respect to inner automorphisms.

Theorem 4: Let $\mathfrak{L} \subset \mathfrak{F}_m(\mathfrak{D})$ be a Lie subalgebra of $\mathfrak{F}_m(\mathfrak{D})$. Let $A \equiv A^{(0)} \in \mathfrak{L}$ and $f \in \mathfrak{L}$.

Assume

(i) $\mathfrak{L} = \mathfrak{P} \oplus \hat{\mathfrak{L}}$ as a vector space,

(ii) ad $A\mathfrak{L} \supset \mathfrak{P}$.

Then

$$A + \epsilon f \sim A + \epsilon g \bmod \text{Int } \mathfrak{L}, \tag{4.4}$$

where

$$g \in \hat{\mathfrak{L}}$$

(i.e., there exists an element of Int $\mathfrak{L}$ which maps $A + \epsilon f$ onto $A + \epsilon g$).

Proof: We look for an element $U \in \mathfrak{L}$ such that

$$e^{\epsilon \text{ ad } U}(A + \epsilon f) = A + \epsilon g, \quad g \in \hat{\mathfrak{L}}.$$

This equation can easily be seen to be equivalent to

$$g + [A, U] = f + \epsilon X \equiv Z,$$

where

$$X = \epsilon^{-1}(e^{\epsilon \text{ ad } U} - 1)f + \epsilon^{-2}(e^{\epsilon \text{ ad } U} - \epsilon \text{ ad } U - 1)A.$$

Now we repeat the arguments of Theorem 3.

Again U is only determined up to an element of the null space of ad A considered as an operator over $\mathfrak{L}$. It is clear that also in the present case we could construct the transformation establishing the equivalence (4.4) as an infinite product of successive transformations in a similar way to how we sketched it in a remark after Theorem 3. Theorem 4 shows how to construct the field $U \in \mathfrak{L}$ such that

$$A + \epsilon f \quad \text{and} \quad A + \epsilon g$$

are connected by an inner automorphism of $\mathfrak{L}$. What is the corresponding transformation of coordinates $z = \mathbf{W}(\zeta, \epsilon)$? The answer is contained in the following lemma.

Lemma 7: If $\mathbf{W}_t$ is constructed iteratively according to the recipe

$$\mathbf{W}_t = \underset{/}{\mathbf{e}} + \epsilon \int_0^t (U \circ \mathbf{W}_\tau)\, d\tau, \tag{4.5}$$

then the corresponding transformation of coordinate is given by

$$z = \mathbf{W}_t(\zeta, \epsilon)\big|_{t=1}. \tag{4.6}$$

Proof: We show that in general

$$g_t \equiv e^{\epsilon t \operatorname{ad} U} f = (f \vdash \mathbf{W}_{-t}) \circ \mathbf{W}_t .$$

This is certainly true for $t = 0$. It is therefore sufficient to show that both expressions for g_t satisfy the same differential equation.

Now we have

$$\dot{g}_t = \epsilon[U, g_t]$$

from the left side expression for g_t. Differentiating the right side expression gives us the same differential equation. Indeed, using

$$\dot{\mathbf{W}}_t = \epsilon(U \circ \mathbf{W}_t)$$

and Einstein's summation convention, we find

$$\frac{d}{dt}(f \vdash \mathbf{W}_{-t}) \circ \mathbf{W}_t = -\epsilon[f_\nu (U \circ \mathbf{W}_{-t})_{,\nu}] \circ \mathbf{W}_t + \epsilon[(f \vdash \mathbf{W}_{-t})_{,\nu} \circ \mathbf{W}_t](U_\nu \circ \mathbf{W}_t).$$

Taking into account the relation

$$0 = \frac{d}{dt} \mathbf{W}_t \circ \mathbf{W}_{-t} = -\epsilon(\mathbf{W}_{t,\rho} \circ \mathbf{W}_{-t})(U_\rho \circ \mathbf{W}_{-t}) + \epsilon U,$$

i.e.,

$$\mathbf{W}_{t,\rho} U_\rho = \epsilon(U \circ \mathbf{W}_t),$$

we find

$$\begin{aligned} \frac{d}{dt}(f \vdash \mathbf{W}_{-t}) \circ \mathbf{W}_t &= -[(f_\nu \mathbf{W}_{-t\rho,\nu}) \circ \mathbf{W}_t] U_{,\rho} \\ &\quad + [(f \vdash \mathbf{W}_{-t}) \circ \mathbf{W}_t]_{,\rho} U_\rho \\ &= \epsilon[-g_t \vdash U + U \vdash g_t] \\ &= \epsilon[U, g_t]. \end{aligned}$$

Hence the lemma is proved.

It is useful to define a mapping

$$\mathfrak{C}: \operatorname{Int} \mathfrak{F}_m(\mathfrak{D}) \to \mathfrak{F}_m(\mathfrak{D})$$

by

$$\mathfrak{C}(e^{\epsilon \operatorname{ad} U}) = \mathbf{W}_t|_{t=1},$$

where $\mathbf{W}_t|_{t=1}$ was defined in (4.5) and (4.6). We easily check that this mapping has the properties

$$\mathfrak{C}(e^{\epsilon \operatorname{ad} U} e^{\epsilon \operatorname{ad} V}) = \mathfrak{C}(e^{\epsilon \operatorname{ad} V}) \circ \mathfrak{C}(e^{\epsilon \operatorname{ad} U}) \quad (4.7)$$

and

$$W^{(0)} = U, \quad (4.8)$$

where W is defined by the formula

$$\mathfrak{C}(e^{\epsilon \operatorname{ad} U}) = \mathbf{e} + \epsilon W.$$

Now let $\mathfrak{N}_A$ be again the null space of A. $\mathfrak{N}_A$ has the structure of a Lie algebra, and the set defined by

$$\mathfrak{J}(A) = \mathfrak{C}(\operatorname{Int} \mathfrak{N}_A)$$

has the structure of a group. We call it the isotropy group of A. Clearly, we have

$$A \vdash \mathbf{W} = A \circ \mathbf{W}$$

for

$$\mathbf{W} \in \mathfrak{J}(A).$$

We are now in the position to investigate the structure of the set of all transformations of coordinates which establish an equivalence of the type (4.1).

Theorem 5: Assume that (i) and (ii) of Theorem 3 are satisfied and, in addition, assume

(iii) $$\operatorname{Int} \mathfrak{N}_A \hat{\mathfrak{L}} \subset \hat{\mathfrak{L}}.$$

If $\mathbf{W}$ and $\mathbf{W}'$ are two transformations of coordinates establishing an equivalence of $A + \epsilon f$ to an element of $\hat{\mathfrak{L}}$, then there exists a transformation $\mathbf{S} \in \mathfrak{J}(A)$ such that

$$\mathbf{W}' = \mathbf{W} \circ \mathbf{S}.$$

Proof: (The proof was stimulated by a similar proof found in Ref. 4.) Write $\mathbf{W} = \mathbf{e} + \epsilon W$ and $\mathbf{W}' = \mathbf{e} + \epsilon W$. To zeroth order the two transformations coincide. By induction assumption, there exists

$$\mathbf{S}_{(N-1)} \in \mathfrak{J}(A)$$

such that, if we write

$$\mathbf{W} \circ \mathbf{S}_{(N-1)} = \mathbf{e} + \epsilon V_{(N-1)}, \quad (4.9)$$

we have

$$V_{(N-1)}^{[N-1]} = W'^{[N-1]}. \quad (4.10)$$

By (iii) the transformation $\mathbf{W} \circ \mathbf{S}_{(N-1)}$ maps $A + \epsilon f$ into $\hat{\mathfrak{L}}$. Moreover, it leads to the same expression for $Z^{(N)}$ [see (4.2)] as $\mathbf{W}'$. Equation (4.3) implies therefore that the field T defined by

$$T = W'^{(N)} - V_{(N-1)}^{(N)}$$

belongs to $\mathfrak{N}_A$. If we define

$$\mathbf{S}_{(N)} = \mathbf{S}_{(N-1)} \circ \mathfrak{C}(e^{\epsilon^{N+1} \operatorname{ad} T})$$

and $V_{(N)}$ in accordance with (4.9), we obtain, using (4.8),

$$V_{(N)} = V_{(N-1)} + \epsilon^N T + \mathfrak{O}(\epsilon^{N+1})$$

and therefore

$$V_{(N)}^{[N]} = V_{(N-1)}^{[N-1]} + W'^{(N)} = W'^{[N]}$$

by (4.10). This relation completes our induction proof.

As an immediate consequence of Theorem 5 and (iii), we have:

Corollary: Under the conditions (i), (ii), and (iii), the set of all transformations establishing the equivalence (4.1) constitutes a left coset of the isotropy

group of A in the group $\mathfrak{C}(\text{Int } \mathfrak{F}_m(\mathfrak{D}))$. Similarly, the set of transformations establishing an equivalence of the type (4.4) coincides with the intersection of such a left coset with $\mathfrak{C}(\text{Int } \mathfrak{L})$. Consequently, the set of all vector fields in $\hat{\mathfrak{L}}$ that are equivalent to $A + \epsilon f$ can be described as an orbit of $\mathfrak{J}(A)$ in $\hat{\mathfrak{L}}$. (For this notion see Ref. 15.)

5. APPLICATION OF THE THEORY TO SPECIFIC SITUATIONS, PARTICULARLY IN MECHANICS

Let $\mathfrak{D}$ be some connected set of R^{2m} and I the matrix

$$I = \begin{pmatrix} 0 & -E \\ E & 0 \end{pmatrix},$$

where E is the m-dimensional unit matrix. With each element $f \in \mathfrak{F}(\mathfrak{D})$ we may associate an element of $\mathfrak{F}_{2m}(\mathfrak{D})$ by

$$(I \operatorname{grad} f)_\mu = \sum_{\nu=1}^{2m} f_{,\nu} I_{\nu\mu}. \tag{5.1}$$

These vector fields are called globally Hamiltonian, and it may be easily seen that they form a Lie subalgebra of $\mathfrak{F}_{2m}(\mathfrak{D})$.

In fact, we have

$$[I \operatorname{grad} f, I \operatorname{grad} g] = I \operatorname{grad} \{f, g\},$$

where

$$\{f, g\} = \sum_{\nu,\rho=1}^{2m} f_{,\nu} I_{\nu\rho} g_{,\rho} \tag{5.2}$$

is the Poisson bracket.

Let $\mathfrak{F}_0$ be the set of elements in $\mathfrak{F}(\mathfrak{D})$ whose coefficients do not depend on z. By $\mathfrak{F}'(\mathfrak{D}) = \mathfrak{F}(\mathfrak{D})/\mathfrak{F}_0$ we denote the quotient space of $\mathfrak{F}(\mathfrak{D})$ with respect to $\mathfrak{F}_0$, i.e., the elements of $\mathfrak{F}'(\mathfrak{D})$ are classes of the type $f + \mathfrak{F}_0$. If we endow $\mathfrak{F}'(\mathfrak{D})$ with the Poisson bracket $\{\,,\}$, it becomes a Lie algebra which is isomorphic to the Lie algebra of globally Hamiltonian fields.

Let us imagine that we apply Theorem 4 to the case of the Lie algebra of globally Hamiltonian vector fields, i.e., we construct iteratively a field

$$U = I \operatorname{grad} V, \quad V \in \mathfrak{F}'(\mathfrak{D}),$$

such that the operator ad U maps a given Hamiltonian field into some specified subspace. Lemma 7 tells us how to find the corresponding transformation of coordinates: $z = \mathbf{W}(\zeta, \epsilon)$. This procedure is, in general, fairly cumbersome.

We can simplify the procedure in the present case if we define the linear operator Ad V: $\mathfrak{F}'(\mathfrak{D}) \to \mathfrak{F}'(\mathfrak{D})$ by

$$\text{Ad } Vf = \{V, f\}. \tag{5.3}$$

Then the formula

$$\mathbf{W}_\nu(\zeta, \epsilon) = e^{\epsilon \text{ Ad } V} \zeta_\nu \tag{5.4}$$

gives a much simpler construction of $\mathbf{W}_\nu(\zeta, \epsilon)$. (5.4) actually implies

$$e^{\epsilon \text{ Ad } V} f = f \circ \mathbf{W}$$

for all $f \in \mathfrak{F}(\mathfrak{D})$. Indeed, if we set

$$W_{t\nu}(\zeta, \epsilon) = e^{t\epsilon \text{ Ad } V} \zeta_\nu$$

and $F_t = f \circ \mathbf{W}_t$, then

$$\begin{aligned} \frac{dF_t}{dt} &= \epsilon (f_{,\nu} \circ \mathbf{W}_t) e^{t\epsilon \text{ Ad } V} \{V, \zeta_\nu\} \\ &= \epsilon (f_{,\nu} \circ \mathbf{W}_t) \{V, \mathbf{W}_{t\nu}\} \\ &= \epsilon \{V, f \circ \mathbf{W}_t\} \\ &= \epsilon \text{ Ad } V F_t. \end{aligned}$$

Hence

$$F_t = e^{t\epsilon \text{ Ad } V} f,$$

and our statement follows for $t = 1$. Moreover, we find

$$\begin{aligned} \frac{dW_{t\nu}}{dt} &= \epsilon e^{\epsilon t \text{ Ad } V} \{V, \zeta_\nu\} \\ &= \epsilon e^{\epsilon t \text{ Ad } V} U_\nu = \epsilon (U_\nu \circ \mathbf{W}_t), \end{aligned}$$

i.e., a relation which shows that the present construction of $\mathbf{W}$ in the case of Hamiltonian vector fields coincides with the one given in the proof of Lemma 7. However, the transformation of coordinates $\mathbf{W}$ is much more simply related to the power series V than it is to the corresponding Hamiltonian vector field U. Therefore, it would be of great convenience if we could modify Theorem 4 in such a way that it would give a recipe for the direct construction of V. Indeed, using the isomorphism described above, we can do this simply by leaving Theorem 4 exactly as it stands except for the replacement

$$\text{ad } U \to \text{Ad } V,$$

where Ad V is defined within the Lie algebra $\mathfrak{F}'(\mathfrak{D})$ by (5.3).

To see how the method works, let

$$\mathfrak{D} = R^{(2)}, \quad z = (x, y), \quad \zeta = (\xi, \eta),$$
$$H = y(\omega + \epsilon \sin x) \in \mathfrak{H}^{\text{per}}(\mathfrak{D}),$$

where the superscript "per" indicates that we confine ourselves to the case of functions periodic in x with period 2π. A in our case is given by

$$A = y\omega.$$

Furthermore, let

$$\mathfrak{P} = \{f \mid f \in \mathfrak{F}^{\text{per}}(\mathfrak{D}), \text{ where the zeroth Fourier coefficient of } f \text{ with respect to } x \text{ vanishes}\},$$

$$\hat{\mathfrak{L}} = \{f \mid f \in \mathfrak{F}^{\text{per}}(\mathfrak{D}); f \text{ does not depend on } x\}.$$

Correspondingly the projection $\hat{}$ is the operation of averaging over x. Because

$$\{y\omega, f\} = \omega \frac{\partial f}{\partial x}$$

and $\mathfrak{P}$ is spanned by the functions e^{ipx}, $p = \pm 1, \pm 2 \cdots$, Ad A is clearly invertible on $\mathfrak{P}$. Hence the conditions of Theorem 4 are satisfied. It gives a recipe to construct a formal Hamiltonian which no longer depends on an angle variable. We will now construct this Hamiltonian and the corresponding transformation up to second order in ϵ. Let the new Hamiltonian be

$$K = \eta\omega + \epsilon g(\eta, \epsilon), \quad g \text{ linear in } \eta.$$

The condition

$$e^{\epsilon \operatorname{Ad} V} H = K,$$

written up to second order, is

$$g(\eta, \epsilon) + \omega \frac{\partial V}{\partial \xi} = \eta \sin \xi + \epsilon X,$$

where

$$X = \{V, \eta \sin \xi\} + \tfrac{1}{2}\epsilon\{V, \{V, \eta \sin \xi\}\} + \tfrac{1}{2}\omega\{V, \{V, \eta\}\} + \tfrac{1}{6}\epsilon\omega\{V, \{V, \{V, \eta\}\}\} + \cdots.$$

Clearly,

$$g^{(0)} = 0, \quad V^{(0)} = -(\eta/\omega) \cos \xi.$$

Hence we have

$$g^{(1)} + \omega \frac{\partial V^{(1)}}{\partial \xi} = -\frac{1}{\omega}\{\eta \cos \xi, \eta \sin \xi\} + \frac{1}{2\omega}\{\eta \cos \xi, \{\eta \cos \xi, \eta\}\} = -\frac{\eta}{2\omega}.$$

Thus, we find

$$g^{(1)} = -(\eta/2\omega), \quad V^{(1)} = 0,$$

and

$$x = \xi + \epsilon\{V^{(0)}, \xi\} + \tfrac{1}{2}\epsilon^2\{V^{(0)}, \{V^{(0)}, \xi\}\} + \mathfrak{O}(\epsilon^3),$$

and similarly for y. Observe that the coefficients of V: $V^{(0)}, V^{(1)}, \cdots$ actually are only determined up to an arbitrary function of η. We normalized them in such a way that $\hat{V} = 0$. We find

$$K = \eta\omega - \frac{\eta}{2\omega}\epsilon^2 + \mathfrak{O}(\epsilon^3),$$

$$x = \xi - \frac{\epsilon}{\omega}\cos \xi - \frac{\epsilon^2}{4\omega^2}\sin 2\xi + \mathfrak{O}(\epsilon^3),$$

$$y = \eta - \frac{\epsilon}{\omega}\eta \sin \xi + \tfrac{1}{2}\epsilon^2 \frac{\eta}{\omega^2} + \mathfrak{O}(\epsilon^3).$$

The inverse transformation is

$$\eta = y + \epsilon\{V^{(0)}, y\} + \tfrac{1}{2}\epsilon^2\{V^{(0)}, \{V^{(0)}, y\}\} + \mathfrak{O}(\epsilon^3) = y + \frac{\epsilon}{\omega} y \sin x + \tfrac{1}{2}\epsilon^2 \frac{y}{\omega^2} + \mathfrak{O}(\epsilon^3). \tag{5.5}$$

Here V is exactly the same function as above, the only difference being that the variables (ξ, η) have been replaced by (x, y). This follows from the observation that the relation

$$z_\nu = \mathbf{W}_\nu(\zeta, \epsilon) = e^{\epsilon \operatorname{Ad} V}\zeta_\nu$$

implies

$$\zeta_\nu = \mathbf{W}_{-1\nu}(z, \epsilon) = e^{-\epsilon \operatorname{Ad} V} z_\nu.$$

(Our treatment of variables may confuse some readers. We do not interpret each transformation passively, i.e., strictly as a transformation of variables, but find the following point of view more convenient: Treat all transformations as mappings, i.e., do all your calculations in one fixed coordinate system, and only after having accomplished all calculations up to the desired order reinterpret the transformations as transformations of coordinates.)

Because K does not depend on the x-like variable (i.e., ξ) any longer, the expression (5.5) is an asymptotic invariant (formal integral). Recall that V is determined only up to an arbitrary function S of η. But because $\{y, S(y)\} = 0$, the asymptotic invariant (5.5) is uniquely determined. This result is confirmed by an application of Theorem 5 to the present situation. It shows that $\mathfrak{J}(A)$ consists of all transformations of type (5.3), where V is a function of η only. These transformations are phase-shifts of the kind

$$\xi \to \xi + \epsilon v(\eta),$$

where v is an arbitrary function of η, and they affect neither K nor η.

Kruskal's generalized perturbation theory [1,3] results from an application of Theorems 3 and 4 to vector fields over an s-parametric family of r-dimensional tori. In the following paragraphs we will work out this point in some detail.

Let $z = (x, y)$, $x \in R^r$, $y \in R^s$, and $\mathfrak{D} = R^r \times \mathfrak{B}$, where $\mathfrak{B}$ is some compact connected set of R^s. Let I^r be the lattice of r-tuples of integers and I^r_+ the sublattice of r-tuples of nonnegative integers. Consider the following subring of $C^\omega_b(\mathfrak{D})$:

$$C^\omega_{\text{per}}(\mathfrak{D}) = \{f \mid f \in C^\omega_b(\mathfrak{D}); f \text{ periodic in } x_\nu, \nu = 1, 2, \cdots, r, \text{ with period } 1\}.$$

Clearly, each element of $C^\omega_{\text{per}}(\mathfrak{D})$ has a Fourier representation

$$(x, y) = \sum_{p \in I^r} f_p(y) e^{2\pi i(p|x)},$$

where

$$(p \mid x) \equiv \sum_{\nu=1}^{r} p_\nu x_\nu,$$

and the right side is uniformly convergent on each set $\mathfrak{D}_\kappa$, $0 < \kappa < \delta(f)$. Moreover, the Fourier, coefficients f_p belong to $C_b^\omega(\mathfrak{B})$. Indeed, we will show by adapting a proof of Moser to our situation that

$$|f_p|_\kappa = \sup_{y \in \mathfrak{B}_\kappa} |f_p(y)| \leq |f|_\kappa \, e^{-2\pi|p|\kappa}, \qquad (5.6)$$

where $|f|_\kappa$ was defined in (2.1) and $|p|$ stands for $\sum_{\nu=1}^{r} |p_\nu|$. For this purpose we shift the path of integration over each variable ξ_ν in the integral

$$f_p(y) = \int_0^1 \cdots \int_0^1 d^r\xi f(\xi, y) e^{-2\pi i(p|\xi)}$$

from the real axis to $\operatorname{Im} \xi_\nu = -\rho \operatorname{sgn} p_\nu$, where $0 < \rho < \kappa$, and find

$$|f_p(y)| \leq |f|_\kappa \, e^{2\pi(p|\operatorname{Im}\xi)} = |f|_\kappa \, e^{-2\pi|p|\rho}.$$

This is true for all ρ smaller than κ; hence the inequality (5.6) follows.

We are now in the position to show that the generalized Kruskal perturbation theory discussed in Ref. 3 drops out as an application of Theorem 3 to vector fields over $\mathfrak{D}$ (i.e., over an analytic family of tori).

Theorem 6: Let ω be a constant $(r + s)$ vector with the last s components vanishing and assume that for each $p \in I^r$ we either have

$$|(p \mid \omega)| > \gamma(|p|^\tau + 1)^{-1}, \qquad \gamma, \tau \text{ real positive constants}, \quad (5.7)$$

or else $(p \mid \omega) = 0$.

The set of all $p \in I^r$ such that $(p \mid \omega) = 0$ is a module over the integers. We denote it by σ and its complement in I^r by σ'.

Let $f \in \mathfrak{F}_{r+s}^{\mathrm{per}}(\mathfrak{D})$ [i.e., the module of $(r + s)$-tuples of formal power series over $C_{\mathrm{per}}^\omega(\mathfrak{D})$] and

$$\hat{\mathfrak{Q}} \equiv \left\{ f \mid f \in \mathfrak{F}_{r+s}^{\mathrm{per}}; f = \sum_{p \in \sigma} f_p e^{2\pi i(p|x)} \text{ with } f_p \in \mathfrak{B}_{r+s}(\mathfrak{D}) \right\};$$

then

$$\omega + \epsilon f \sim \omega + \epsilon g \bmod \mathfrak{F}_{r+s}(\mathfrak{D})$$

and

$$g \in \hat{\mathfrak{Q}}.$$

Proof: We have to show that the conditions of Theorem 3 are satisfied. If we define

$$\mathfrak{P} = \left\{ f \mid f \in \mathfrak{F}_{r+s}^{\mathrm{per}}(\mathfrak{D}); f = \sum_{p \in \sigma'} f_p e^{2\pi i(p|x)} \text{ with } f_p \in \mathfrak{F}_{r+s}(\mathfrak{B}) \right\},$$

obviously the condition (i) of Theorem 3 is satisfied, i.e., we have

$$\mathfrak{F}_{r+s}^{\mathrm{per}} = \hat{\mathfrak{Q}} \oplus \mathfrak{P}.$$

The only thing left to show is that ad ω is invertible on $\mathfrak{P}$. It is easy to see that this condition boils down to the following statement: If

$$f = \sum_{p \in \sigma'} f_p e^{2\pi i(p|x)} \quad \text{with} \quad f_p \in C_b^\omega(\mathfrak{B})$$

belongs to $C_{\mathrm{per}}^\omega(\mathfrak{D})$, then also

$$g = \sum_{p \in \sigma'} \frac{f_p}{(\omega \mid p)} e^{2\pi i(p|x)}$$

does. Again we adapt a proof of Moser to the present situation. Choose numbers κ, $\delta(g)$, and ρ such that

$$0 < \kappa < \delta(g) < \rho < \delta(f).$$

We have on $\mathfrak{D}_\kappa$

$$\left| \frac{f_p}{(\omega \mid p)} e^{2\pi i(p|x)} \right| < \frac{1}{\gamma}(|p|^\tau + 1) |f|_\rho \, e^{-|p|(\rho-\kappa)}.$$

Hence

$$|g|_\kappa = \sup_{(x,y) \in \mathfrak{D}_\kappa} g(x, y)$$

exists if the sum

$$\sum_{p \in \sigma'} (|p|^\tau + 1) e^{-|p|(\rho-\kappa)}$$

exists. This sum now is bounded by the integral

$$\frac{1}{(\rho - \kappa)^{\tau+r}} \int \cdots \int (|v|^\tau + 1) e^{-|v|} \, dv.$$

Remark: We saw that a condition of the kind (5.7) is sufficient to insure that all the formal power series constructed with the help of Theorem 3 belong to $\mathfrak{F}_{r+s}(\mathfrak{D})$. Such a condition has first been considered by Siegel[16] in connection with the question of stability of an analytic mapping with a fixed point.

From a purely computational point of view, one would probably better redefine the set σ in such a way as to include not only the p's for which $(p \mid \omega) = 0$ but also those for which $(p \mid \omega) < \mathfrak{D}(\epsilon)$. Then for the remaining p's we would have $|(p \mid \omega)| > \epsilon$, a condition which is obviously stronger than (5.7).

Coffey in his paper asks if degenerate perturbation theory can be made canonical to all orders. Under certain conditions this is indeed the case, as will be stated and proven in the next theorem.

The next theorem is an adaption of Theorem 6 to Hamiltonian systems with m degrees of freedom.

Theorem 7: Let ω be a constant $(r+s)$ vector subjected to the same conditions as in Theorem 6. Let $\mathfrak{D}$ be defined as above, with the only difference that now $r+s=2m$, and define

$$\hat{\mathfrak{L}} \equiv \{f \mid f \in \mathfrak{F}^{\mathrm{per}}(\mathfrak{D}); f = \sum_{p \in \sigma} f_p e^{2\pi i(p \mid x)}\}$$

with $f_p \in \mathfrak{F}(\mathfrak{B})$. Then for any $f \in \mathfrak{F}^{\mathrm{per}}(\mathfrak{D})/\mathfrak{F}_0$ we have

$$(\omega \mid \tilde{y}) + \epsilon f \sim (\omega \mid \tilde{y}) + \epsilon g \bmod \mathrm{Int}\, \mathfrak{F}'^{\mathrm{per}}$$

and $g \in \hat{\mathfrak{L}}$, where $\tilde{y} = (y_1, \cdots, y_r)$ and y_ν is canonically conjugate to x_ν, $\nu = 1, 2, \cdots, r$.

Proof: As before, the essential part according to Theorem 4 is to show that Ad $(\omega\, \tilde{y}|\,)$ as a linear operator over the Lie algebra $\mathfrak{F}'^{\mathrm{per}}(\mathfrak{D})$ can be inverted on

$$\mathfrak{P} \equiv \left\{ f \mid f \in \mathfrak{F}(\mathfrak{D}); f = \sum_{p \in \sigma'} f_p e^{2\pi i(p \mid x)}; f_p \in C_b^\omega(\mathfrak{B}) \right\}.$$

(Observe that $p = 0$ belongs to the set σ. Hence the projection of f onto $\mathfrak{P}$ is uniquely determined although its projection onto $\hat{\mathfrak{L}}$ is determined only up to a constant.) This is done exactly the same way as in the proof of Theorem 6.

Theorem 7 is interesting because it shows that for each Hamiltonian $(\omega \mid \tilde{y}) + \epsilon f$ there exists a formal canonical transformation such that the transformed Hamiltonian depends only on the angle combinations

$$(p \mid x), \quad p \in \sigma.$$

Especially, it becomes independent of x whenever σ contains only the origin of the lattice I^r (compare Ref. 3). In this latter case the perturbation theory corresponding to the equivalence established in Theorem 7 is known as nondegenerate perturbation theory. Because the transformed Hamiltonian does not depend on ξ, there exists an r-fold formal integral given by

$$\eta_\nu = \mathbf{W}_{-1\nu}(z, \epsilon), \quad \nu = 1, 2, \cdots, r. \tag{5.8}$$

Theorem 5 implies again that the new Hamiltonian as well as the r formal integrals are uniquely determined. [The isotropy group of the zeroth-order vector field is generated by canonical transformations affecting only the variables $\eta_{r+1}, \cdots, \eta_s$ and by "phase-shifts"

$$\xi_\nu \to \xi_\nu + \epsilon v_\nu(\eta, \epsilon)].$$

Given an analytic system

$$\dot{z} = f(z, \epsilon), \quad f \in \mathfrak{H}_m(\mathfrak{D}),$$

it is, in general, not difficult to construct formally equivalent systems by applying Theorem 3 in a convenient way. To each formally equivalent system there corresponds a perturbation theory of Kruskal's type, and, because the construction of the approximate solution requires that the truncated equivalent system

$$\dot{\zeta} = g^{[N]}(\zeta, \epsilon)$$

can be solved for small values of N, the question arises: Given an analytic vector field, what is the simplest formally equivalent vector field? If one is interested only in relatively rough approximations as described in Theorems 1 and 2, formal equivalence is enough. However, even from a purely computational point of view, the question of convergence of the formal expansions is not unimportant because, if convergence is proved, one is assured that the effort and the time spent in computing higher-order approximations are rewarded by a better knowledge of the exact solution.

The problem now becomes much more difficult and can be expressed as follows: Given $f \in \mathfrak{H}_m(\mathfrak{D})$, find the "simplest" vector field $g \in \mathfrak{H}_m(\mathfrak{D})$ such that

$$f \sim g \bmod \mathfrak{H}_m(\mathfrak{D}).$$

In this generality the question cannot be answered. But in special important cases many results have been established in recent years by Moser and coworkers in New York and Arnold[17] and coworkers in Moscow based on earlier works by Poincaré, Siegel,[16] Kolmogorov,[6] and others. We intend to discuss the implications of those works for the question of convergence of Kruskal's generalized perturbation theory in a future work. Here we only want to point out that the formal aspects of Moser's work about the permanence of quasiperiodic motions under perturbation can also be based on (a somewhat modified version of) Theorem 4. This is a slightly different point of view than that taken by Moser himself.

Theorem 4': Let $\mathfrak{L} \subset \mathfrak{F}_m(\mathfrak{D})$ be a Lie subalgebra of $\mathfrak{F}_m(\mathfrak{D})$. Let $A = A^{(0)} \in \mathfrak{L}$ and $f \in \mathfrak{L}$. Assume

(i) $$\mathfrak{L} = \mathfrak{N}_A \oplus \hat{\mathfrak{L}} \oplus \mathfrak{P}$$

as a vector space, where $\mathfrak{N}_A$ is the null space of ad A,

(ii) $$\mathrm{ad}\, A\mathfrak{L} \supset \mathfrak{P}.$$

Then there exists $N \in \mathfrak{N}_A$ such that

$$A + N + \epsilon f \sim A + \epsilon g \bmod \mathrm{Int}\, \mathfrak{L},$$

where $g \in \hat{\mathfrak{L}}$.

The proof of this theorem is exactly the same as of Theorem 4 except that Z (see 4.2) in each step now splits into three parts.

Notice that by the formal procedure described in Theorem 4′ the original vector field $A + \epsilon f$ is "renormalized," i.e., modified by a member of $\mathfrak{N}_A$, before it becomes equivalent to $A + \epsilon g$, $g \in \hat{\mathfrak{L}}$. As long as we are only concerned with the formal aspects of the equivalence, such a renormalization is not necessary. One can instead make the replacement

$$\mathfrak{N}_A \oplus \hat{\mathfrak{L}} \to \hat{\mathfrak{L}}$$

and proceed as described in Theorem 4. However, such a "renormalization" may become essential if we are concerned with the convergence of the formal expansions. It is exactly this procedure, proposed first by Kolmogorov, together with a systematic generalization of Newton's approximation method, that leads to much deeper insight into many old questions of classical mechanics.

We have already mentioned an example where renormalization in classical mechanics becomes necessary to ensure convergence, namely the perturbation of a quasiperiodic motion, which, for example, can occur in a system of coupled oscillators. The simplest Hamiltonian in R^{2m} describing a quasiperiodic motion with characteristic numbers

$$\begin{aligned}(\omega_1, \cdots, \omega_r, \Omega_1, \cdots, \Omega_r, \Omega_{r+1}, \cdots, \Omega_m, & \\ \Omega_{m+1}, \cdots, \Omega_{2m-r}) & \\ \Omega_\nu = 0, \qquad \nu = 1, 2, \cdots, r, & \\ \Omega_{r+\nu} = -\Omega_{m+\nu}, \quad \nu = 1, \cdots, m-r & \end{aligned}$$

(for this notion, see Ref. 4) is

$$A = \sum_{\nu=1}^{r} \omega_\nu y_\nu + \sum_{\nu=r+1}^{m} \Omega_\nu y_\nu y_{\nu+m-r}.$$

Here we assume that all y_ν, $\nu = 1, \cdots, 2m-r$, vary in a neighborhood of the origin of R^{2m-r}, the variable $y_{m+\nu}$ is canonical conjugate to $y_{r+\nu}$, $\nu = 1, \cdots, m-r$, and the variables canonical conjugate to the y_ν, $\nu = 1, \cdots, r$, are denoted by $x_1, \cdots, x_r$ and assumed to vary over all of R^r. In other words, we put $\mathfrak{D} = R^r \times \{0\}$, where 0 denotes the origin of R^{2m-r} and again we use the notation

$$z = (x, y) x \in R^r, y \in R^{2m-r}.$$

In the Lie algebra of the Poisson brackets over $\mathfrak{D}$ we have

$$\operatorname{Ad} Af = \sum_{\nu=1}^{r} \omega_\nu \frac{\partial f}{\partial x_\nu} + \sum_{\nu=r+1}^{2m-r} \Omega_\nu y_\nu \frac{\partial f}{\partial y_\nu}.$$

Let $\mathfrak{F}^{\mathrm{per}}(\mathfrak{D})$ be the subring of $\mathfrak{F}(\mathfrak{D})$ whose elements have coefficients that are periodic with period 1 in the x-variables.

Define the operator Q by

$$\begin{aligned} Q(f)(x, y, \epsilon) \equiv & f(x, 0, \epsilon) \\ & + \sum_{\nu=1}^{2m-r} \frac{\partial f}{\partial y_\nu}(x, 0, \epsilon) y_\nu \\ & + \frac{1}{2} \sum_{\nu,\mu=r+1}^{2m-r} \frac{\partial^2 f}{\partial y_\nu \partial y_\mu}(x, 0, \epsilon) y_\nu y_\mu. \end{aligned}$$

Observe that, if

$$g \in \hat{\mathfrak{L}} \equiv \{g \mid g \in \mathfrak{F}^{\mathrm{per}}(\mathfrak{D}); Q(g) = 0\},$$

then the canonical equations belonging to the Hamiltonian $A + \epsilon g$ have the form

$$\begin{aligned} \dot{\xi} &= \omega + \mathfrak{O}(\eta), \\ \dot{\eta} &= \Omega\eta + \mathfrak{O}(\eta^2). \end{aligned}$$

Hence, if we assume that for some

$$\begin{aligned} f &\in \mathfrak{H}^{\mathrm{per}}(\mathfrak{D}), \\ A + \epsilon f &\sim A + \epsilon g, \end{aligned}$$

modulo a convergent transformation, the canonical equations belonging to the perturbed Hamiltonian $A + \epsilon f$ have a quasiperiodic solution with the same characteristic numbers as those belonging to the Hamiltonian A.

The eigenvectors of Ad A belonging to the subspace

$$\{f \mid Q(f) = f\}$$

of $\mathfrak{F}^{\mathrm{per}}(\mathfrak{D})$ are

$$\begin{aligned} & e^{2\pi i(p|x)}, \quad e^{2\pi i(p|x)} y_{\nu|\nu=1,\cdots,r}, \\ & e^{2\pi i(p|x)} y_\nu y_{\mu|\nu,\mu=r+1,\cdots,2m-r} p \in I^r \end{aligned}$$

with eigenvalues $2\pi i(p \mid \omega) + \Omega_\nu - \rho\Omega_\mu$, where

$$p \in I^r, \nu, \mu = r+1, \cdots, 2m-r, \rho = 0, 1.$$

If we assume that there exist positive constants γ and τ such that for all $p \in I^r$ different from zero the absolute values of these eigenvalues are larger than

$$\gamma(|p|^\tau + 1)^{-1}$$

and all Ω_ν ($\nu = r+1, \cdots, 2m-r$) are different from zero and from each other, then the only eigenvectors with the eigenvalue zero are

$$1, \quad y_{\nu|\nu=1,2,\cdots,r}, \quad y_{r+\nu} y_{m+\nu|\nu=1,2,\cdots,m-r},$$

where we can disregard the first one because two Hamiltonians differing by a constant are identified. The null space $\mathfrak{N}_A$ of Ad A is the span of these functions and has dimension m. If we define

$$\mathfrak{P} = \{f \mid f \in \mathfrak{F}^{\mathrm{per}}(\mathfrak{D}), Q(f) = f, \quad f \text{ has zero component in } \mathfrak{N}_A\},$$

then a similar argument as applied in the proof of Theorem 6 shows that the assumptions of Theorem 4′ are satisfied. Hence we can conclude that any perturbed system if suitably modified is formally equivalent (via a canonical transformation) to a system which has a quasiperiodic solution with the same characteristic numbers as the unperturbed system.

The essential point of Moser's work is that this equivalence does define a transformation of variables which is not only formal but analytic, thereby proving that a quasiperiodic notion is conserved under perturbation if the Hamiltonian of the original system is modified by a suitable function of a maximal but by assumption finite set of functionally independent integrals (i.e., members of $\mathfrak{N}_A$). This result is of great physical value if the parameters represented by those integrals can be compensated for by the same number of free parameters in the physical problem. For details see Moser's original work[4] and also Ref. 6. The reason why we briefly discussed Moser's work is first to show that the formal aspects of it are based essentially on the same algebraic considerations as Kruskal's perturbation theory and secondly to demonstrate that ensuring any kind of convergence of the formal expansions of the type proposed by Kruskal may only be possible after modifying the original system slightly. There is at least one case in which the perturbation expansion, as originally conceived by Kruskal,[1] converges, namely the case of a system

$$\dot{x} = 1,$$
$$\dot{y} = \epsilon A(x, \epsilon)y, \tag{5.9}$$

where $\mathfrak{D} = R \times R^s$ and A is an $s \times s$ matrix belonging to $\mathfrak{H}_{m^2}(R)$ and periodic in x with period 1. It is easily seen that in this case Kruskal's perturbation theory constructs an analytically equivalent system

$$\dot{\xi} = 1,$$
$$\dot{\eta} = \epsilon B(\epsilon)\eta, \text{ where } B(\epsilon) \text{ is independent of } \xi. \tag{5.10}$$

This follows from a simple application of Floquet's theory to the present situation. Indeed, let $\Phi(x, \epsilon)$ be the fundamental matrix solution of (5.9) with the properties

$$\Phi(x, 0) = \Phi(0, \epsilon) = E \ (= s \times s \text{ unit matrix}).$$

Define

$$\Phi(\epsilon) = [\Phi(1, \epsilon) - E]/\epsilon$$

and

$$\epsilon B(\epsilon) = \ln \Phi(1, \epsilon) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \Phi^k(\epsilon)\epsilon^k,$$

where the right side is convergent for $\|\Phi(\epsilon)\| < 1/\epsilon$ in some norm $\| \ \|$. One checks that (5.9) is analytically transformed into (5.10) by the transformation

$$\xi = x,$$
$$\eta = P(x, \epsilon)y,$$

where $P(x, \epsilon) = \Phi(x, \epsilon)e^{-\epsilon B(\epsilon)x}$, proving the convergence of Kruskal's expansions in this special case.

The convergence question becomes much harder to tackle in the nonlinear case. We intend to come back to this question in a later paper.

6. OUTLOOK

We have demonstrated that any perturbation theory which uses formally equivalent systems for the construction of the approximate solution according to the recipe given in Theorem 1 avoids "secular terms." In general such an approximate solution is asymptotic to the exact solution for a time of the order L/ϵ. However, this time interval actually is of infinite length in the case of stable differential equations (Theorem 2).

It is an open question whether Theorem 2 is also true in the case of arbitrary differential equations provided one approximates a solution whose integral curve lies completely on a stable manifold associated with a singular solution.

We also showed how to construct formally equivalent systems in general by using some algebraic ideas of Moser (Theorems 3 and 4). Applying these ideas to Hamiltonian systems, we demonstrated that degenerate perturbation theory (in the sense of Coffey[3]) can be made canonical to all orders provided that (i) to zeroth order the phase space decomposes into a family of invariant tori, each of them carrying a quasiperiodic motion with frequencies $\omega_1, \cdots, \omega_r$, (ii) for all $q \in I^r$ for which $(q \mid \omega) \neq 0$ the ω's satisfy the conditions (5.7).

Finally we showed that in the case of linear periodic systems the perturbation theory, as originally presented by Kruskal,[1] converges.

ACKNOWLEDGMENT

The author would like to thank Professor R. Jost (Federal Institute of Technology Zurich) for some helpful discussions, Dr. G. W. Ford and Dr. K. M. Case (University of Michigan) for introducing the author to the subject of these fascinating perturbation theories, and last but not least his wife for her patient typing of the manuscript.

[1] M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[2] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods of Nonlinear Oscillations* (Gordon and Breach, New York, 1961).

[3] T. P. Coffey, J. Math. Phys. **10**, 426 (1969).
[4] J. Moser, Math. Ann. **169**, 136 (1967) [See also SIAM Rev. **8**, 145 (1966)].
[5] S. Sternberg, J. Math. Mech. **10**, 451 (1961).
[6] A. N. Kolmogorov, in the book by R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967), Appendix D.
[7] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964).
[8] J. Milnor, *Morse Theory* (Princeton U.P., Princeton, N.J., 1963).
[9] C. S. Gardner, Phys. Rev. **115**, 791 (1959).
[10] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).
[11] J. Berkowitz and C. S. Gardner, Commun. Pure Appl. Math. **12**, 501 (1959).
[12] T. P. Coffey, thesis, University of Michigan, 1966.
[13] E. A. Coddington and N. Levison, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[14] A. Halanay, *Differential Equations* (Academic, New York, 1966).
[15] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[16] C. L. Siegel, *Vorlesungen über Himmelsmechanik* (Springer, Berlin, 1956).
[17] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Extremum Principles for the Equation $\nabla^2\phi = \phi - \phi^3$ *

PETER D. ROBINSON†
Mathematics Research Center and Theoretical Chemistry Institute
University of Wisconsin, Madison, Wisconsin 53706

(Received 11 May 1970)

It has been conjectured that the Lagrangian functional

$$G(\Phi) = \int [\tfrac{1}{2}\Phi(-\nabla^2 + 1)\Phi - \tfrac{1}{4}\Phi^4]\, d\mathbf{r}$$

provides an upper bound for $G(\varphi_0)$, where φ_0 is the ground-state eigenfunction of the nonlinear field equation $-\nabla^2\varphi + \varphi - \varphi^3 = 0$, provided that $G(\Phi)$ is constrained to be stationary with respect to variations in the amplitude of Φ. In this paper we demonstrate that this conjecture is true. The effect of a stationary-scale constraint on $G(\Phi)$ is also shown to guarantee an upper bound. Complementary functionals arising from an Euler–Hamilton approach to the problem are investigated. Unfortunately, these do not (as in more favorable circumstances) provide lower bounds for $G(\varphi_0)$, but merely alternative upper bounds. With a very simple trial function, a complementary bound is closer to $G(\varphi_0)$ than is $G(\Phi)$.

1. INTRODUCTION

The nonlinear field equation

$$-\nabla^2\varphi + \varphi - \varphi^3 = 0 \tag{1}$$

is among those which have been suggested to describe extended elementary particles, and it also arises in a nonlinear electromagnetic theory.[1] In the (3-dimensional) spherically symmetric case with

$$\varphi = \varphi(r), \quad 0 \leq r < \infty, \tag{2}$$

it has been proved[2] that there exists a countably infinite set of analytic eigenfunctions $\{\varphi_n(r)\}$ which are characterized by a discrete set of initial values $\{\varphi_n(0)\}$. These eigenfunctions have the following properties:

$$\varphi_n(r) \text{ has } n \text{ zeros}, \quad n = 0, 1, 2, \cdots; \tag{3}$$

$$\varphi_n(r) \sim \text{a const} \times r^{-1} \exp(-r) \text{ for large } r; \tag{4}$$

$$\frac{d\varphi_n}{dr} = 0 \quad \text{when} \quad r = 0. \tag{5}$$

For the intervening intervals of $\varphi(0)$, other solutions of Eq. (1) exist which are asymptotic to ± 1 when r is large.

Approximate solutions for some of the eigenfunctions have been investigated by Schiff and his coworkers[3–5] using variational techniques. Equation (1) is the Euler equation for the Lagrangian functional

$$G(\Phi) = \tfrac{1}{2}\int [(\nabla\Phi)^2 + \Phi^2]\, d\mathbf{r} - \tfrac{1}{4}\int \Phi^4\, d\mathbf{r}, \tag{6}$$

which is the same as

$$G(\Phi) = \tfrac{1}{2}\int \Phi(-\nabla^2 + 1)\Phi\, d\mathbf{r} - \tfrac{1}{4}\int \Phi^4\, d\mathbf{r} \tag{7}$$

provided that Φ is well behaved and goes to zero fast enough at infinity. If we now suppose that φ is an eigenfunction and let

$$\Phi = \varphi + \delta\varphi \tag{8}$$

be an approximation to φ, then it follows that

$$G(\Phi) = G(\varphi) + \delta G(\varphi) + \delta^2 G(\varphi) + O(\delta\varphi^3), \tag{9}$$

[3] T. P. Coffey, J. Math. Phys. **10**, 426 (1969).
[4] J. Moser, Math. Ann. **169**, 136 (1967) [See also SIAM Rev. **8**, 145 (1966)].
[5] S. Sternberg, J. Math. Mech. **10**, 451 (1961).
[6] A. N. Kolmogorov, in the book by R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967), Appendix D.
[7] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964).
[8] J. Milnor, *Morse Theory* (Princeton U.P., Princeton, N.J., 1963).
[9] C. S. Gardner, Phys. Rev. **115**, 791 (1959).
[10] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).
[11] J. Berkowitz and C. S. Gardner, Commun. Pure Appl. Math. **12**, 501 (1959).
[12] T. P. Coffey, thesis, University of Michigan, 1966.
[13] E. A. Coddington and N. Levison, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[14] A. Halanay, *Differential Equations* (Academic, New York, 1966).
[15] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[16] C. L. Siegel, *Vorlesungen über Himmelsmechanik* (Springer, Berlin, 1956).
[17] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Extremum Principles for the Equation $\nabla^2\phi = \phi - \phi^3$ *

PETER D. ROBINSON†
Mathematics Research Center and Theoretical Chemistry Institute
University of Wisconsin, Madison, Wisconsin 53706

(Received 11 May 1970)

It has been conjectured that the Lagrangian functional

$$G(\Phi) = \int [\tfrac{1}{2}\Phi(-\nabla^2 + 1)\Phi - \tfrac{1}{4}\Phi^4]\, d\mathbf{r}$$

provides an upper bound for $G(\varphi_0)$, where φ_0 is the ground-state eigenfunction of the nonlinear field equation $-\nabla^2\varphi + \varphi - \varphi^3 = 0$, provided that $G(\Phi)$ is constrained to be stationary with respect to variations in the amplitude of Φ. In this paper we demonstrate that this conjecture is true. The effect of a stationary-scale constraint on $G(\Phi)$ is also shown to guarantee an upper bound. Complementary functionals arising from an Euler–Hamilton approach to the problem are investigated. Unfortunately, these do not (as in more favorable circumstances) provide lower bounds for $G(\varphi_0)$, but merely alternative upper bounds. With a very simple trial function, a complementary bound is closer to $G(\varphi_0)$ than is $G(\Phi)$.

1. INTRODUCTION

The nonlinear field equation

$$-\nabla^2\varphi + \varphi - \varphi^3 = 0 \tag{1}$$

is among those which have been suggested to describe extended elementary particles, and it also arises in a nonlinear electromagnetic theory.[1] In the (3-dimensional) spherically symmetric case with

$$\varphi = \varphi(r), \quad 0 \leq r < \infty, \tag{2}$$

it has been proved[2] that there exists a countably infinite set of analytic eigenfunctions $\{\varphi_n(r)\}$ which are characterized by a discrete set of initial values $\{\varphi_n(0)\}$. These eigenfunctions have the following properties:

$$\varphi_n(r) \text{ has } n \text{ zeros}, \quad n = 0, 1, 2, \cdots; \tag{3}$$

$$\varphi_n(r) \sim \text{a const} \times r^{-1} \exp(-r) \text{ for large } r; \tag{4}$$

$$\frac{d\varphi_n}{dr} = 0 \quad \text{when} \quad r = 0. \tag{5}$$

For the intervening intervals of $\varphi(0)$, other solutions of Eq. (1) exist which are asymptotic to ± 1 when r is large.

Approximate solutions for some of the eigenfunctions have been investigated by Schiff and his coworkers[3–5] using variational techniques. Equation (1) is the Euler equation for the Lagrangian functional

$$G(\Phi) = \tfrac{1}{2}\int [(\nabla\Phi)^2 + \Phi^2]\, d\mathbf{r} - \tfrac{1}{4}\int \Phi^4\, d\mathbf{r}, \tag{6}$$

which is the same as

$$G(\Phi) = \tfrac{1}{2}\int \Phi(-\nabla^2 + 1)\Phi\, d\mathbf{r} - \tfrac{1}{4}\int \Phi^4\, d\mathbf{r} \tag{7}$$

provided that Φ is well behaved and goes to zero fast enough at infinity. If we now suppose that φ is an eigenfunction and let

$$\Phi = \varphi + \delta\varphi \tag{8}$$

be an approximation to φ, then it follows that

$$G(\Phi) = G(\varphi) + \delta G(\varphi) + \delta^2 G(\varphi) + O(\delta\varphi^3), \tag{9}$$

where

$$G(\varphi) = \tfrac{1}{4}\int \varphi^4 \, d\mathbf{r}, \tag{10}$$

$$\delta G(\varphi) = 0, \tag{11}$$

and

$$\delta^2 G(\varphi) = \tfrac{1}{2}\int \delta\varphi(-\nabla^2 + 1 - 3\varphi^2)\delta\varphi \, d\mathbf{r}. \tag{12}$$

The stationary value $G(\varphi)$ represents the "energy" associated with the eigenfunction φ.

The sign of $\delta^2 G$ is not clear. It follows from Eqs. (12) and (1) that

$$\text{if} \quad \delta\varphi = \epsilon\varphi, \quad \delta^2 G = -\epsilon^2 \int \varphi^4 \, d\mathbf{r} < 0, \tag{13}$$

and so evidently there are some $\delta\varphi$ for which $\delta^2 G$ is negative. (Another choice which illustrates this is $\delta\varphi = \epsilon \, d\varphi/dr$.[6]) Yet the operators $-\nabla^2$ and 1 are positive, and so it is reasonable to suppose that there are also some $\delta\varphi$ for which $\delta^2 G$ is positive; such a $\delta\varphi$ would perhaps de-emphasize regions where $3\varphi^2 > 1$.

In Refs. 3 and 5 it is suggested that if Φ is constrained to satisfy the relation

$$\int [(\nabla\Phi)^2 + \Phi^2] \, d\mathbf{r} = \int \Phi^4 \, d\mathbf{r} \tag{14}$$

(which is one of the identities satisfied by the eigenfunctions), then for the "ground state" $\varphi = \varphi_0$ the Lagrangian functional furnishes an upper bound

$$G(\Phi) \geq G(\varphi_0). \tag{15}$$

The reason given is that if (14) holds, then from (6) it follows that $G(\Phi)$ reduces to

$$\mathcal{G}(\Phi) = \tfrac{1}{4}\int [(\nabla\Phi)^2 + \Phi^2] \, d\mathbf{r}, \tag{16}$$

which by its nature is a functional with a minimum value; this value can only be raised by constraints such as (14). In fact, subject to (14), $\mathcal{G}(\Phi)$ is indeed stationary around $\Phi = \varphi$, but the second-order variation is still equivalent to expression (12), where $\delta\varphi$ is now subject to conditions imposed on it by (14). As it is by no means obvious what the effect of these conditions is, the sign of $\delta^2 G$ remains uncertain.

However, the numerical evidence presented in Ref. 3 does support the conjecture that (14) implies (15). In the present paper we demonstrate that the sign of $\delta^2 G(\varphi_0)$ is settled by the magnitude of λ_1, the second smallest eigenvalue of the radial equation

$$(-\nabla^2 + 1)\theta = 3\lambda\phi_0^2\theta. \tag{17}$$

Subject to (14), $\delta^2 G(\varphi_0)$ is positive if $\lambda_1 > 1$. A good approximation to λ_1 is 1.28, and so the conjecture of Schiff and his coworkers is well substantiated. (We assume throughout that third- and higher-order variations can be neglected, so that the question of whether a functional is an extremum is settled by the sign of its second-order variation.)

The effect of an alternative constraint

$$2\int [(\nabla\Phi)^2 + 3\Phi^2] \, d\mathbf{r} = 3\int \Phi^4 \, d\mathbf{r} \tag{18}$$

upon $G(\Phi)$ is also considered; again, this guarantees an upper bound on $G(\varphi_0)$. In addition, complementary functionals arising from an Euler–Hamilton approach to the problem are investigated. These do not provide lower bounds for $G(\phi_0)$ as might be hoped, but merely alternative upper bounds. Using a simple trial function, we see that a complementary functional provides a closer upper bound than does $G(\Phi)$.

2. IMPLICITLY CONSTRAINED LAGRANGIANS

The conditions (14) and (18) respectively arise when the Lagrangian functional $G(\Phi)$ is made stationary for variations in A, the amplitude of Φ, and in α, a radial scale parameter.[1] To analyze the situation, let us work in terms of the positive functionals

$$X(\Phi) = \int (\nabla\Phi)^2 \, d\mathbf{r}, \quad Y(\Phi) = \int \Phi^2 \, d\mathbf{r},$$

$$Z(\Phi) = \int \Phi^4 \, d\mathbf{r}. \tag{19}$$

Then from (6)

$$G(\Phi(r)) = \tfrac{1}{2}X + \tfrac{1}{2}Y - \tfrac{1}{4}Z \tag{20}$$

so that

$$G(A\Phi(\alpha r)) = G(A, \alpha)$$
$$= \tfrac{1}{2}A^2(X/\alpha) + \tfrac{1}{2}A^2(Y/\alpha^3) - \tfrac{1}{4}A^4(Z/\alpha^3). \tag{21}$$

Thus $\partial G/\partial A$ vanishes when

$$A^2 = A_0^2 \text{ (say)} = \left(\frac{X}{\alpha} + \frac{Y}{\alpha^3}\right) \Big/ \left(\frac{Z}{\alpha^3}\right), \tag{22}$$

and $\partial G/\partial\alpha$ vanishes when

$$\alpha^2 = \alpha_0^2 \text{ (say)} = (3A^4 Z - 6A^2 Y)/2A^2 X. \tag{23}$$

Evidently, if $A\Phi(\alpha r)$ is treated as a new $\Phi(r)$ so that $A = \alpha = 1$, the stationary conditions (22) and (23) are precisely the constraints (14) and (18).

From (21) and (22) it follows that

$$G(A_0, \alpha) = \frac{1}{4}\left(\frac{X}{\alpha} + \frac{Y}{\alpha^3}\right)^2 \left(\frac{Z}{\alpha^3}\right)^{-1} = \tilde{G}(\Phi(\alpha r)) \quad \text{say,} \tag{24}$$

where

$$\tilde{G}(\Phi(r)) = \tfrac{1}{4}\left(\int [(\nabla\Phi)^2 + \Phi^2] \, d\mathbf{r}\right)^2 \left(\int \Phi^4 \, d\mathbf{r}\right)^{-1}. \tag{25}$$

The unrestricted functional $\tilde{G}(\Phi)$ is equivalent to the Lagrangian $G(\Phi)$ subjected to the stationary-amplitude condition (14). $\tilde{G}(\Phi)$ can be thought of as an implicitly constrained Lagrangian, and is independent of the amplitude of Φ.

In a similar way, from (21) and (23), we have

$$G(A, \alpha_0) = (\tfrac{2}{27})^{\frac{1}{2}}(A^2X)^{\frac{3}{2}}(A^4Z - 2A^2Y)^{-\frac{1}{2}} = G^\dagger\{A\Phi(r)\} \quad \text{say,} \tag{26}$$

which indicates that the unrestricted functional

$$G^\dagger(\Phi) = (\tfrac{2}{27})^{\frac{1}{2}}\left(\int(\nabla\Phi)^2\,d\mathbf{r}\right)^{\frac{3}{2}}\left(\int(\Phi^4 - 2\Phi^2)\,d\mathbf{r}\right)^{-\frac{1}{2}} \tag{27}$$

is equivalent to $G(\Phi)$ subjected to the stationary-scale condition (18). This second implicitly constrained Lagrangian $G^\dagger(\Phi)$ is independent of the radial scale parameter.

The doubly stationary $G(A_0, \alpha_0)$ takes the form

$$G(A_0, \alpha_0) = (16\,YX^3/27Z^2)^{\frac{1}{2}} = \tilde{G}(\Phi(\alpha_0 r)) = G^\dagger(A_0\Phi(r)), \tag{28}$$

which we can write logically as $\tilde{G}^\dagger(\Phi)$. This functional

$$\tilde{G}^\dagger(\Phi) = (\tfrac{1}{2}\tfrac{6}{7})^{\frac{1}{2}}\left(\int\Phi^2\,d\mathbf{r}\right)^{\frac{1}{2}}\left(\int(\nabla\Phi)^2\,d\mathbf{r}\right)^{\frac{3}{2}}\left(\int\Phi^4\,d\mathbf{r}\right)^{-1} \tag{29}$$

is equivalent to $G(\Phi)$ subjected to both of the constraints (14) and (18) and is independent of both A and α.

It follows from Eqs. (21)–(23) that at (A_0, α_0)

$$\frac{\partial^2 G}{\partial A^2} < 0, \quad \frac{\partial^2 G}{\partial \alpha^2} < 0, \quad \frac{\partial^2 G}{\partial A\,\partial\alpha} > \frac{\partial^2 G}{\partial A^2}\frac{\partial^2 G}{\partial \alpha^2}, \tag{30}$$

so that $G(A, \alpha)$ has a saddle point at (A_0, α_0). However, using (24) and (26), we can show that

$$\left[\frac{\partial^2}{\partial\alpha^2}G(A_0, \alpha)\right]_{\alpha=\alpha_0} > 0, \quad \left[\frac{\partial^2}{\partial A^2}G(A, \alpha_0)\right]_{A=A_0} > 0, \tag{31}$$

and so $G(A_0, \alpha)$ and $G(A, \alpha_0)$ each have minima at (A_0, α_0). There are no other relevant turning points, which implies that

$$\tilde{G}(\Phi(\alpha r)) \geq \tilde{G}(\Phi(\alpha_0 r)) = \tilde{G}^\dagger(\Phi) \tag{32}$$

and

$$G^\dagger(A\Phi(r)) \geq G^\dagger(A_0\Phi(r)) = \tilde{G}^\dagger(\Phi). \tag{33}$$

Thus, in particular, when $A = \alpha = 1$,

$$\tilde{G}(\Phi) \geq \tilde{G}^\dagger(\Phi) \quad \text{and} \quad G^\dagger(\Phi) \geq \tilde{G}^\dagger(\Phi), \tag{34}$$

and so $\tilde{G}(\Phi)$ and $G^\dagger(\Phi)$ are each possible maximizing functionals for $G(\varphi)$.

Of the three implicitly constrained Lagrangians, $\tilde{G}(\Phi)$ is the easiest to work with. In Sec. 3 below, we will show that

$$\tilde{G}(\Phi) \geq G(\varphi_0) \tag{35}$$

without any restriction on Φ other than it be close to φ_0. Since $\Phi(\alpha_0 r)$ is an allowed Φ, it will then follow from (32) that

$$\tilde{G}^\dagger(\Phi) \geq G(\varphi_0) \tag{36}$$

and finally from (34) and (36) that

$$G^\dagger(\Phi) \geq G(\varphi_0). \tag{37}$$

In practice, for a given form of trial function Φ, one would try to evaluate $\tilde{G}^\dagger$, thus obtaining a better bound than either $\tilde{G}$ or $G^\dagger$ in general.

3. JUSTIFICATION OF THE RESULT $\tilde{G}(\Phi) \geq G(\varphi_0)$

If we again suppose that φ is a radial eigenfunction and set

$$\Phi = \varphi + \delta\varphi \tag{38}$$

in expression (25) for $\tilde{G}(\Phi)$, we find that

$$\tilde{G}(\Phi) = \tilde{G}(\varphi) + \delta\tilde{G}(\varphi) + \delta^2\tilde{G}(\varphi) + O(\delta\varphi^3), \tag{39}$$

where

$$\tilde{G}(\varphi) = G(\varphi) = \tfrac{1}{4}\int\varphi^4\,d\mathbf{r}, \tag{40}$$

$$\delta\tilde{G}(\varphi) = \delta G(\varphi) = 0, \tag{41}$$

and

$$\delta^2\tilde{G}(\varphi) = \tfrac{1}{2}\int\delta\varphi(-\nabla^2 + 1 - 3\varphi^2)\delta\varphi\,d\mathbf{r} + \left(\int\varphi^3\delta\varphi\,d\mathbf{r}\right)^2\left(\int\varphi^4\,d\mathbf{r}\right)^{-1}. \tag{42}$$

The first term on the right of (42) looks the same as $\delta^2 G$ [expression (12)], but, if we were dealing directly with $\delta^2 G$, we should need to consider the effect of the constraint (14) on $\delta\varphi$. Here in (42) there is no restriction on $\delta\varphi$ (other than well-behavedness and being small enough at infinity); this is a merit of the implicitly constrained Lagrangian.

It is convenient to think of Eq. (42) in the form

$$\delta^2\tilde{G}(\varphi) = \int\delta\varphi\,\Omega\delta\varphi\,d\mathbf{r}\,, \tag{43}$$

where

$$\Omega = \tfrac{1}{2}(-\nabla^2 + 1 - 3\varphi^2) + \left(\int\varphi^4\,d\mathbf{r}\right)^{-1}|\varphi^3\rangle\langle\varphi^3|. \tag{44}$$

The operator $|\varphi^3\rangle\langle\varphi^3|$ is a positive-definite nonlocal operator with the property that for any χ

$$|\varphi^3\rangle\langle\varphi^3|\,\chi = \varphi^3\int\varphi^3\chi\,d\mathbf{r}; \tag{45}$$

the notation is Dirac's. We notice that, from (1),

$$\Omega\varphi = 0; \tag{46}$$

this is expected since $\tilde{G}$ is insensitive to amplitude change.

Let us investigate the consequences of $\delta^2\tilde{G}$ being negative for some $\delta\phi$. The only negative contribution to Ω is $-\frac{3}{2}\phi^2$, and ϕ^2 is finite and exponentially decreasing for large r. Thus $\delta^2\tilde{G}$ is certainly bounded below for acceptable $\delta\phi$. It follows that, since Ω is self-adjoint, there exists a smallest positive number β^2 for which

$$\int \delta\varphi(\Omega + \beta^2\varphi^2)\delta\varphi \, d\mathbf{r} \tag{47}$$

has a minimum value of zero, and this minimum value will occur when $\delta\varphi = \epsilon\theta$, say, where

$$(\Omega + \beta^2\varphi^2)\theta = 0. \tag{48}$$

Now multiply (48) on the left by φ and integrate over r-space. Then because β^2 is not zero and Ω is self-adjoint, it follows from (46) that

$$\int \varphi^3\theta \, d\mathbf{r} = 0. \tag{49}$$

Hence the nonlocal part $|\varphi^3\rangle\langle\varphi^3|$ of Ω annihilates θ, and from (44) and (48) we have simply

$$(-\nabla^2 + 1)\theta = 3\lambda\varphi^2\theta, \tag{50}$$

where

$$2\beta^2 = 3(1 - \lambda) > 0. \tag{51}$$

It is clear that λ must be positive, since $(-\nabla^2 + 1)$ is a positive operator and a zero λ would correspond to the unacceptable $\theta = r^{-1}\exp(-r)$. Also we know that the eigenvalue $\lambda = \frac{1}{3}$ corresponds to the irrelevant $\theta = \varphi$. The vital question is then, are there any other λ-eigenvalues in the interval

$$0 < \lambda < 1? \tag{52}$$

If there are, this line of argument does not help us. If there are not, then evidently there can be no $\delta\varphi$ for which $\delta^2\tilde{G}$ is negative, and we will have established the maximizing property of $\tilde{G}(\Phi)$.

We can answer this question in the simplest case when $\varphi = \varphi_0(r)$, the "ground state" radial eigenfunction (nothing we have said so far restricts us to any *particular* eigenfunction φ). It is sensible only to consider radial trial functions, so that $\theta = \theta(r)$ and Eq. (50) becomes

$$\left(-\frac{d^2}{dr^2} + 1\right)(r\theta) = 3\lambda\varphi_0^2 r\theta, \quad 0 \le r < \infty. \tag{53}$$

Since $\varphi_0(r)$ is unknown (otherwise we should not be bothering with approximate methods), an approximation to it has to be used in Eq. (53) to determine the λ-eigenvalues. This introduces a first-order error, but its effect is on $\delta^2\tilde{G}$ which is already second order, and so the net effect on $\tilde{G}$ is third order. We are already neglecting third-order terms, and so this step is justifiable as well as necessary.

A simple approximation to $\varphi_0(r)$ is the function[3]

$$C\exp(-\gamma r), \quad C = 4\sqrt{2}, \quad \gamma = \sqrt{3}, \tag{54}$$

the choice of C and γ being made to obey the constraints (14) and (18). With this approximation, the eigenfunctions of Eq. (53) are[7]

$$\theta_k(r) = r^{-1}J_{\gamma^{-1}}\{C\lambda_k^{\frac{1}{2}}\exp(-\gamma r)\}, \quad k = 0, 1, 2, \cdots, \tag{55}$$

where $J_{\gamma^{-1}}$ is the Bessel function of order γ^{-1}. It must vanish at $r = 0$, so that the equation determining λ_k is

$$J_{\gamma^{-1}}(C\lambda_k^{\frac{1}{2}}) = 0. \tag{56}$$

The lowest two values of λ_k turn out to be

$$\lambda_0 = 0.33024, \quad \lambda_1 = 1.27931. \tag{57}$$

The exact value of λ_0 should be $\frac{1}{3}$, corresponding to $\theta_0 = \varphi_0$. The fact that the error in λ_0 is only 1% testifies to the reasonableness of the approximation (54) to φ_0. One would hardly expect a 28% error in λ_1, and so it seems certain that there are no other λ-eigenvalues satisfying (52). We conclude that $\delta^2\tilde{G}(\varphi_0)$ is positive, so that $\tilde{G}(\Phi) \ge G(\varphi_0)$ if Φ is close to φ_0.

It seems unlikely that the vital question (52) can be answered simply for other eigenfunctions $\varphi_n(r)$. Even with the approximation $(a - br)\exp(-cr)$ for $\varphi_1(r)$, Eq. (50) becomes too complicated for further analytical progress. Thus we should have to resort to numerical work, which is probably not justified in this context.

4. COMPLEMENTARY FUNCTIONALS

An eigenfunction of Eq. (1) can be regarded as a solution $\Phi = \varphi$ of a more general equation of type

$$T^*T\Phi + f(\Phi) = 0 \quad \text{in } V, \tag{58}$$

subject to

$$\Phi = 0 \quad \text{on } \partial V, \tag{59}$$

where T is a linear operator and T^* is its adjoint defined by

$$\int_V UT\Phi \, d\mathbf{r} = \int_V (T^*U)\Phi \, d\mathbf{r}. \tag{60}$$

Here U is an arbitrary function in the space of $T\Phi$.

To identify Eqs. (1) and (58), we take V as the whole Euclidean 3-space and can choose either

$$T = \text{grad}, \quad T^* = -\text{div}, \quad f(\Phi) = \Phi - \Phi^3 \tag{61}$$

or

$$T^*T = (-\nabla^2 + 1), \quad f(\Phi) = -\Phi^3, \tag{62}$$

or possibly even make some choice intermediate between (61) and (62). Complementary variational principles have been developed[8–10] for equations like (58); they arise from the decomposition of (58) into the pair of generalized Euler–Hamilton equations

$$T\Phi = \frac{\partial H}{\partial U} = U, \tag{63}$$

$$T^*U = \frac{\partial H}{\partial \Phi} = -f(\Phi) \tag{64}$$

and from consideration of the generalized action functional when one or other of (63) and (64) is satisfied identically. If we set

$$L = T^*T \tag{65}$$

and

$$F(\Phi) = \int^{\Phi} f(\Phi')\, d\Phi', \tag{66}$$

then the complementary functionals

$$G(\Phi) = \int [\tfrac{1}{2}\Phi L\Phi + F(\Phi)]\, d\mathbf{r} \tag{67}$$

and

$$J(\Phi) = \int \{-\tfrac{1}{2}\Phi L\Phi + F[f^{-1}(-L\Phi)] + (L\Phi)f^{-1}(-L\Phi)\}\, d\mathbf{r} \tag{68}$$

are stationary for variations of Φ around φ and have the common stationary value

$$G(\varphi) = J(\varphi). \tag{69}$$

With either of the choices (61) or (62) (or an intermediate choice), the functional (67) is just the Lagrangian functional (7). Since

$$\delta^2 G = \tfrac{1}{2}\int \delta\varphi [L + f'(\varphi)]\delta\varphi\, d\mathbf{r} \tag{70}$$

and

$$\delta^2 J = -\tfrac{1}{2}\int (L\delta\varphi)\{L^{-1} + [f'(\varphi)]^{-1}\}(L\delta\varphi)\, d\mathbf{r}, \tag{71}$$

it is clear that G and J provide complementary upper and lower bounds to $G(\varphi)$ if $f'(\varphi)$ is nonnegative. For Eq. (1) the most positive choice of $f'(\varphi)$ is obtained from (61), i.e.,

$$f'(\varphi) = 1 - 3\varphi^2. \tag{72}$$

This can still be very negative, since[3] $\varphi_0(0) \approx 5$ and this is the smallest of the $\varphi_n(0)$. Thus, in circumstances when $\delta^2 G$ is positive with a partly negative $f'(\varphi)$, it is probably a question of the positive operator L dominating the $f'(\varphi)$. In such a case one would not expect L^{-1} to dominate $[f'(\varphi)]^{-1}$: rather the opposite, since for example $3 - 2 > 0$ but $3^{-1} - 2^{-1} < 0$. Hence in these circumstances we might anticipate that $J(\Phi)$ is also an upper bound to $G(\varphi)$. Precisely this situation does arise with linear equations.[11]

It is often convenient to remove the inverse functional f^{-1} from (68) by expressing everything in terms of a trial function χ specified by

$$L\Phi + f(\chi) = 0, \tag{73}$$

so that

$$J(\Phi) = J(-L^{-1}f(\chi)) = J[\chi] \quad \text{say}, \tag{74}$$

where

$$J[\chi] = \int [-\tfrac{1}{2}f(\chi)L^{-1}f(\chi) + F(\chi) - \chi f(\chi)]\, d\mathbf{r}. \tag{75}$$

It follows from (67) and (75) that

$$G(\chi) - J[\chi] = \tfrac{1}{2}\int (\Phi - \chi)L(\Phi - \chi)\, d\mathbf{r} \geq 0, \tag{76}$$

so that if $G(\chi)$ and $J[\chi]$ are each upper bounds, then $J[\chi]$ is the better one.

With Eq. (1), the simplest J-functional arises from the decomposition (62) and is

$$J[\chi] = \int (-\tfrac{1}{2}\chi^3 L^{-1}\chi^3 + \tfrac{3}{4}\chi^4)\, d\mathbf{r}, \quad L = -\nabla^2 + 1. \tag{77}$$

An investigation of this functional which parallels that of $G(\Phi)$ in Secs. 2 and 3 shows that the amplitude-optimized form

$$\tilde{J}[\chi] = \tfrac{1}{4}\left(\int \chi^4\, d\mathbf{r}\right)^3 \left(\int \chi^3 L^{-1}\chi^3\, d\mathbf{r}\right)^{-2} \tag{78}$$

is an upper bound to $G(\varphi)\,(= J(\varphi) = J[\varphi])$ in the case $\varphi = \varphi_0$. The proof again turns on the λ-eigenvalues of Eq. (53).

The decomposition (61) leads to the functional

$$J'[\chi] = \int (-\tfrac{1}{2}(\chi^3 - \chi)(-\nabla^2)^{-1}(\chi^3 - \chi) + \tfrac{3}{4}\chi^4 - \tfrac{1}{2}\chi^2)\, d\mathbf{r}, \tag{79}$$

which is more difficult to analyze.

Using the simple approximation for $\varphi_0(r)$,

$$\chi = Ae^{-\alpha r}, \tag{80}$$

we obtain

$$\tilde{G}^{\dagger}[\chi] = 1.540, \quad \alpha = \sqrt{3} = 1.7321, \tag{81}$$

$$\tilde{J}^{\dagger}[\chi] = 1.514, \quad \alpha = 1.733, \tag{82}$$

and the stationary value of $J'[\chi]$ is actually the same as $\tilde{J}^{\dagger}[\chi]$ to this accuracy, with the same optimum α-value. The stationary value calculated directly by numerical methods[3] is

$$G(\varphi_0) = J[\varphi_0] = 1.503. \tag{83}$$

Thus, in this situation, the complementary J-functional provides a better upper bound than the simpler G-functional. This also happens with linear equations.[11] We can merely regret that it does not provide a lower bound.

ACKNOWLEDGMENTS

I am most grateful to Professor S. Epstein and Professor B. Noble for helpful discussions and also to the Mathematics Research Center and Theoretical Chemistry Institute for their kind hospitality during the Fall of 1969 when this work was carried out.

* This work supported by the U.S. Army Mathematics Research Center, University of Wisconsin under contract No. DA-31-124-ARO-D-462 and by National Science Foundation Grant GP-12832.

† Permanent address: Mathematics Department, Bradford University, Bradford, England.

[1] H. Schiff, Proc. Roy. Soc. (London) **A269**, 277 (1962).
[2] R. Finkelstein, R. Lelevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[3] D. D. Betts, H. Schiff, and W. B. Strickfaden, J. Math. Phys. **4**, 334 (1963).
[4] J. W. Darewych and H. Schiff, J. Math. Phys. **8**, 1479 (1967).
[5] J. W. Darewych and H. Schiff, Can. J. Phys. **47**, 1419 (1969).
[6] R. H. Hobart, Proc. Phys. Soc. (London) **82**, 201 (1963).
[7] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 12.
[8] B. Noble, University of Wisconsin, Mathematics Research Center Report No. 473 (1964).
[9] A. M. Arthurs and P. D. Robinson, Proc. Cambridge Phil. Soc. **65**, 535 (1969).
[10] A. M. Arthurs, Proc. Cambridge Phil. Soc. **65**, 803 (1969).
[11] P. D. Robinson and A. M. Arthurs, J. Math. Phys. **9**, 1364 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

A Mathematical Property of Dirac Plane Waves

M. MURASKIN
University of North Dakota, Grand Forks, North Dakota

(Received 5 March 1970)

From Dirac plane waves, we can define a g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. It is found that Dirac plane waves satisfy, in addition to the Dirac equation, a set of nonlinear field equations involving these quantities. General $e^{\alpha}{}_i$ plane waves do not satisfy these equations.

I. INTRODUCTION

We here obtain some new mathematical properties of Dirac plane waves. Dirac plane waves define a g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. We find that Dirac plane waves satisfy, in addition to the Dirac equation, a set of nonlinear equations. We find that this property is not shared by general $e^{\alpha}{}_i$ plane waves.

II. RELATIONSHIP BETWEEN SPINORS AND g_{ij}, Γ^i_{jk}, $e^{\alpha}{}_i$

Starting with two 2-component spinors χ_A and ϕ_A, we define a set of 4 null vectors[1]

$$\begin{aligned} l_i &= \sigma_i^{A\dot{B}}\chi_A\chi_{\dot{B}}, \quad & n_i &= \sigma_i^{A\dot{B}}\phi_A\phi_{\dot{B}}, \\ m_i &= \sigma_i^{A\dot{B}}\chi_A\phi_{\dot{B}}, \quad & \bar{m}_i &= \sigma_i^{A\dot{B}}\phi_A\chi_{\dot{B}}. \end{aligned} \tag{1}$$

We then define $e^{\alpha}{}_i$ by

$$\begin{aligned} e^1{}_i &= (2)^{-\frac{1}{2}}(-m_i - \bar{m}_i), \\ e^2{}_i &= (2)^{-\frac{1}{2}}(i)^{-1}(\bar{m}_i - m_i), \\ e^3{}_i &= (2)^{-\frac{1}{2}}(n_i - l_i), \\ e^0{}_i &= (2)^{-\frac{1}{2}}(l_i + n_i). \end{aligned} \tag{2}$$

This implies

$$\begin{aligned} l_i &= (2)^{-\frac{1}{2}}(e^0{}_i - e^3{}_i), \quad & n_i &= (2)^{-\frac{1}{2}}(e^0{}_i + e^3{}_i), \\ m_i &= (2)^{-\frac{1}{2}}(-e^1{}_i - ie^2{}_i), \quad & \bar{m}_i &= (2)^{-\frac{1}{2}}(-e^1{}_i + ie^2{}_i). \end{aligned} \tag{3}$$

g_{ij} is defined by

$$g_{ij} = e^{\alpha}{}_i\, e^{\beta}{}_j\, g_{\alpha\beta} = l_i n_j + n_i l_j - m_i\bar{m}_j - \bar{m}_i m_j, \tag{4}$$

where $g_{\alpha\beta}$ is the Minkowski metric. Γ^i_{jk} is defined by

$$\Gamma^i_{jk} = e_{\alpha}{}^i \frac{\partial e^{\alpha}{}_j}{\partial x^k} = l^i \frac{\partial n_j}{\partial x^k} + n^i \frac{\partial l_j}{\partial x^k} - m^i \frac{\partial \bar{m}_j}{\partial x^k} - \bar{m}^i \frac{\partial m_j}{\partial x^k}. \tag{5}$$

Thus χ_A and ϕ_A define a set of functions g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. We notice, as a result of (4) and (5), that

$$\frac{\partial g_{ij}}{\partial x^k} - \Gamma^m_{ik}g_{mj} - \Gamma^m_{jk}g_{im} \equiv g_{ij;k} = 0 \tag{6}$$

is identically satisfied. We refer to $g_{ij;k}$ as the covariant derivative of g_{ij}.

and the stationary value of $J'[\chi]$ is actually the same as $\tilde{J}^{\dagger}[\chi]$ to this accuracy, with the same optimum α-value. The stationary value calculated directly by numerical methods[3] is

$$G(\varphi_0) = J[\varphi_0] = 1.503. \tag{83}$$

Thus, in this situation, the complementary J-functional provides a better upper bound than the simpler G-functional. This also happens with linear equations.[11] We can merely regret that it does not provide a lower bound.

ACKNOWLEDGMENTS

I am most grateful to Professor S. Epstein and Professor B. Noble for helpful discussions and also to the Mathematics Research Center and Theoretical Chemistry Institute for their kind hospitality during the Fall of 1969 when this work was carried out.

* This work supported by the U.S. Army Mathematics Research Center, University of Wisconsin under contract No. DA-31-124-ARO-D-462 and by National Science Foundation Grant GP-12832.

† Permanent address: Mathematics Department, Bradford University, Bradford, England.

[1] H. Schiff, Proc. Roy. Soc. (London) **A269**, 277 (1962).
[2] R. Finkelstein, R. Lelevier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
[3] D. D. Betts, H. Schiff, and W. B. Strickfaden, J. Math. Phys. **4**, 334 (1963).
[4] J. W. Darewych and H. Schiff, J. Math. Phys. **8**, 1479 (1967).
[5] J. W. Darewych and H. Schiff, Can. J. Phys. **47**, 1419 (1969).
[6] R. H. Hobart, Proc. Phys. Soc. (London) **82**, 201 (1963).
[7] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 12.
[8] B. Noble, University of Wisconsin, Mathematics Research Center Report No. 473 (1964).
[9] A. M. Arthurs and P. D. Robinson, Proc. Cambridge Phil. Soc. **65**, 535 (1969).
[10] A. M. Arthurs, Proc. Cambridge Phil. Soc. **65**, 803 (1969).
[11] P. D. Robinson and A. M. Arthurs, J. Math. Phys. **9**, 1364 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

A Mathematical Property of Dirac Plane Waves

M. MURASKIN
University of North Dakota, Grand Forks, North Dakota

(Received 5 March 1970)

From Dirac plane waves, we can define a g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. It is found that Dirac plane waves satisfy, in addition to the Dirac equation, a set of nonlinear field equations involving these quantities. General $e^{\alpha}{}_i$ plane waves do not satisfy these equations.

I. INTRODUCTION

We here obtain some new mathematical properties of Dirac plane waves. Dirac plane waves define a g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. We find that Dirac plane waves satisfy, in addition to the Dirac equation, a set of nonlinear equations. We find that this property is not shared by general $e^{\alpha}{}_i$ plane waves.

II. RELATIONSHIP BETWEEN SPINORS AND g_{ij}, Γ^i_{jk}, $e^{\alpha}{}_i$

Starting with two 2-component spinors χ_A and ϕ_A, we define a set of 4 null vectors[1]

$$\begin{aligned} l_i &= \sigma_i^{A\dot{B}}\chi_A\chi_{\dot{B}}, \quad & n_i &= \sigma_i^{A\dot{B}}\phi_A\phi_{\dot{B}}, \\ m_i &= \sigma_i^{A\dot{B}}\chi_A\phi_{\dot{B}}, \quad & \bar{m}_i &= \sigma_i^{A\dot{B}}\phi_A\chi_{\dot{B}}. \end{aligned} \tag{1}$$

We then define $e^{\alpha}{}_i$ by

$$\begin{aligned} e^1{}_i &= (2)^{-\frac{1}{2}}(-m_i - \bar{m}_i), \\ e^2{}_i &= (2)^{-\frac{1}{2}}(i)^{-1}(\bar{m}_i - m_i), \\ e^3{}_i &= (2)^{-\frac{1}{2}}(n_i - l_i), \\ e^0{}_i &= (2)^{-\frac{1}{2}}(l_i + n_i). \end{aligned} \tag{2}$$

This implies

$$\begin{aligned} l_i &= (2)^{-\frac{1}{2}}(e^0{}_i - e^3{}_i), \quad & n_i &= (2)^{-\frac{1}{2}}(e^0{}_i + e^3{}_i), \\ m_i &= (2)^{-\frac{1}{2}}(-e^1{}_i - ie^2{}_i), \quad & \bar{m}_i &= (2)^{-\frac{1}{2}}(-e^1{}_i + ie^2{}_i). \end{aligned} \tag{3}$$

g_{ij} is defined by

$$g_{ij} = e^{\alpha}{}_i\, e^{\beta}{}_j\, g_{\alpha\beta} = l_i n_j + n_i l_j - m_i\bar{m}_j - \bar{m}_i m_j, \tag{4}$$

where $g_{\alpha\beta}$ is the Minkowski metric. Γ^i_{jk} is defined by

$$\Gamma^i_{jk} = e_{\alpha}{}^i \frac{\partial e^{\alpha}{}_j}{\partial x^k} = l^i \frac{\partial n_j}{\partial x^k} + n^i \frac{\partial l_j}{\partial x^k} - m^i \frac{\partial \bar{m}_j}{\partial x^k} - \bar{m}^i \frac{\partial m_j}{\partial x^k}. \tag{5}$$

Thus χ_A and ϕ_A define a set of functions g_{ij}, Γ^i_{jk}, and $e^{\alpha}{}_i$. We notice, as a result of (4) and (5), that

$$\frac{\partial g_{ij}}{\partial x^k} - \Gamma^m_{ik}g_{mj} - \Gamma^m_{jk}g_{im} \equiv g_{ij;k} = 0 \tag{6}$$

is identically satisfied. We refer to $g_{ij;k}$ as the covariant derivative of g_{ij}.

Under constant Lorentz transformations, Γ^i_{jk} is a tensor function. Let us consider the equation[2-4]

$$\frac{\partial\Gamma^i_{jk}}{\partial x^l} + \Gamma^m_{jk}\Gamma^i_{ml} - \Gamma^i_{mk}\Gamma^m_{jl} - \Gamma^i_{jm}\Gamma^m_{kl} \equiv \Gamma^i_{jk;l} = 0. \quad (7)$$

We shall show that Dirac plane waves satisfy this equation. Furthermore, we shall show that the covariant derivatives of all functions constructed from g_{ij}, Γ^i_{jk}, $e^\alpha{}_i$, and ∂_k all vanish as a consequence of (6) and (7).

III. VANISHING OF ALL COVARIANT DERIVATIVES

We first consider $(\partial g_{ij}/\partial x^k)_{;m}$. We take the covariant derivative of (6). This gives

$$\left(\frac{\partial g_{ij}}{\partial x^k}\right)_{;m} - \Gamma^t_{ik;m}g_{tj} - \Gamma^t_{ik}g_{tj;m} - \Gamma^t_{jk;m}g_{it} - \Gamma^t_{jk}g_{it;m} = 0. \quad (8)$$

Then, using (7) and (6), we get

$$\left(\frac{\partial g_{ij}}{\partial x^k}\right)_{;m} = 0. \quad (9)$$

We next investigate $(\partial\Gamma^i_{jk}/\partial x^l)_{;m}$ by taking the covariant derivative of (7). Then, using (7), we see that

$$\left(\frac{\partial\Gamma^i_{jk}}{\partial x^l}\right)_{;m} = 0. \quad (10)$$

We next consider $(\partial^2 g_{ij}/\partial x^n\partial x^k)_{;m}$ by first taking $\partial/\partial x^n$ of (6) and then taking the covariant derivative of this equation. Then, using the previous results (9) and (10) and also (7) and (6), we find

$$\left(\frac{\partial^2 g_{ij}}{\partial x^n\partial x^k}\right)_{;m} = 0. \quad (11)$$

We may then look into $(\partial^2\Gamma^i_{jk}/\partial x^n\partial x^l)_{;m}$ by taking $\partial/\partial x^n$ of (7) and then taking the covariant derivative of this equation. Using (10) and (7), we get

$$\left(\frac{\partial^2\Gamma^i_{jk}}{\partial x^m\partial x^l}\right)_{;n} = 0. \quad (12)$$

Continuing on, in this manner, we can consider the covariant derivative of all the derivatives of g_{ij} and Γ^i_{jk}. Using the results for previously obtained lower-order derivatives, we then conclude that the covariant derivatives of all these functions are zero, thus proving the theorem.

For the 16 variables, we have, as a consequence of (5),

$$e^\alpha{}_{i;k} = \frac{\partial e^\alpha{}_i}{\partial x^k} - \Gamma^j_{ik}e^\alpha{}_j = 0. \quad (13)$$

Using the same procedure as above, we also find that the covariant derivatives of all functions constructed from $e^\alpha{}_i$ and ∂_k are zero as well.

IV. DIRAC PLANE WAVE SOLUTIONS OF THE FIELD EQUATIONS

Dirac plane wave solutions can be expressed in terms of χ_A and ϕ^A by means of

$$\psi = \tfrac{1}{2}\begin{pmatrix}\chi_1 + \phi^1\\ \chi_2 + \phi^2\\ \chi_1 - \phi^1\\ \chi_2 - \phi^2\end{pmatrix}. \quad (14)$$

We introduce the spin metric

$$\epsilon_{AB} = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} = \epsilon^{AB}. \quad (15)$$

The Dirac plane waves satisfy $|\chi_A\phi^A|^2 = 1$, and thus we have ($g^{(0)}_{ij}$ is the Minkowski metric)

$$g_{ij} = g^{(0)}_{ij} \quad (16)$$

for all four plane wave solutions.

We write down χ_A and ϕ_A for the four plane wave solutions[5] {where $N = (E + m/2m)^{\frac{1}{2}}$ and $\exp(ip_ix^i) = \exp[i(Et - \mathbf{p}\cdot\mathbf{x})]$}. For the first plane wave solution we have

$$\begin{aligned}\chi_1 &= N\left(1 + \frac{p_3}{E+m}\right)e^{-ip_ix^i}, & \phi_1 &= N\left(\frac{p_1 - ip_2}{E+m}\right)e^{ip_ix^i},\\ \chi_2 &= N\left(\frac{p_1 + ip_2}{E+m}\right)e^{-ip_ix^i}, & \phi_2 &= N\left(1 - \frac{p_3}{E+m}\right)e^{ip_ix^i}.\end{aligned} \quad (17)$$

For the second solution we have

$$\begin{aligned}\chi_1 &= N\left(\frac{p_1 - ip_2}{E+m}\right)e^{-ip_ix^i}, & \phi_1 &= -N\left(1 + \frac{p_3}{E+m}\right)e^{ip_ix^i},\\ \chi_2 &= N\left(1 - \frac{p_3}{E+m}\right)e^{-ip_ix^i}, & \phi_2 &= -N\left(\frac{p_1 + ip_2}{E+m}\right)e^{ip_ix^i}.\end{aligned} \quad (18)$$

For the third solution we have

$$\chi_1 = N\left(1 + \frac{p_3}{E+m}\right)e^{ip_ix^i},$$
$$\phi_1 = -N\left(\frac{p_1 - ip_2}{E+m}\right)e^{-ip_ix^i},$$
$$\chi_2 = N\left(\frac{p_1 + ip_2}{E+m}\right)e^{ip_ix^i}, \tag{19}$$
$$\phi_2 = -N\left(1 - \frac{p_3}{E+m}\right)e^{-ip_ix^i}.$$

For the fourth solution we have

$$\chi_1 = N\left(\frac{p_1 - ip_2}{E+m}\right)e^{ip_ix^i},$$
$$\phi_1 = N\left(1 + \frac{p_3}{E+m}\right)e^{-ip_ix^i},$$
$$\chi_2 = N\left(1 - \frac{p_3}{E+m}\right)e^{ip_ix^i}, \tag{20}$$
$$\phi_2 = N\left(\frac{p_1 + ip_2}{E+m}\right)e^{-ip_ix^i}.$$

From χ_A and ϕ_A we construct l_i, n_i, m_i, and $\bar{m}_i$, by means of (1). We shall work in the Lorentz frame where $p_1 = p_2 = 0$ (our equations are all covariant under Lorentz transformations, and this frame makes things simpler). From (5) we then obtain for the nonvanishing Γ^i_{jk} the following. For the first solution (17) we get

$$\Gamma^1_{23} = -2p_3, \quad \Gamma^1_{20} = 2E,$$
$$\Gamma^2_{13} = 2p_3, \quad \Gamma^2_{10} = -2E. \tag{21}$$

For the second solution (18) we get

$$\Gamma^1_{23} = 2p_3, \quad \Gamma^1_{20} = -2E,$$
$$\Gamma^2_{13} = -2p_3, \quad \Gamma^2_{10} = 2E. \tag{22}$$

For the third solution (19), we get the same as (22) and, for the fourth solution (20), we get the same as (21). These constant Γ^i_{jk} correspond to l_i, n_i; m_i, and $\bar{m}_i$ that have an x dependence of the type exp $(\pm 2ip_ix^i)$. We find that (7) is identically satisfied by (21) and (22).

Associated with (17) we have these nonvanishing $e^\alpha{}_i$:

$$e^1{}_1 = \cos(2p_ix^i), \quad e^1{}_2 = \sin(2p_ix^i),$$
$$e^2{}_2 = \cos(2p_ix^i), \quad e^2{}_1 = -\sin(2p_ix^i),$$
$$e^3{}_3 = E/m, \quad e^0{}_3 = -p_3/m, \tag{23}$$
$$e^0{}_0 = E/m, \quad e^3{}_0 = -p_3/m.$$

From (18) we get

$$e^1{}_1 = -\cos(2p_ix^i), \quad e^1{}_2 = \sin(2p_ix^i),$$
$$e^2{}_2 = \cos(2p_ix^i), \quad e^2{}_1 = \sin(2p_ix^i),$$
$$e^3{}_3 = -E/m, \quad e^0{}_3 = -p_3/m, \tag{24}$$
$$e^0{}_0 = E/m, \quad e^3{}_0 = p_3/m.$$

From (19) we get

$$e^1{}_1 = -\cos(2p_ix^i), \quad e^1{}_2 = \sin(2p_ix^i),$$
$$e^2{}_2 = -\cos(2p_ix^i), \quad e^2{}_1 = -\sin(2p_ix^i),$$
$$e^3{}_3 = E/m, \quad e^0{}_3 = -p_3/m, \tag{25}$$
$$e^0{}_0 = E/m, \quad e^3{}_0 = -p_3/m.$$

From (20) we get

$$e^1{}_1 = \cos(2p_ix^i), \quad e^1{}_2 = \sin(2p_ix^i),$$
$$e^2{}_2 = -\cos(2p_ix^i), \quad e^2{}_1 = \sin(2p_ix^i),$$
$$e^3{}_3 = -E/m, \quad e^0{}_3 = -p_3/m, \tag{26}$$
$$e^0{}_0 = E/m, \quad e^3{}_0 = p_3/m.$$

We have shown that the Eq. (7) does admit Dirac plane waves as solutions.

V. NONDIRAC PLANE WAVES

It is an easy matter to show that not all the plane wave solutions to ($\Box = g^{ij(0)}\partial_i\partial_j$)

$$\Box e^\alpha{}_i = -m^2 e^\alpha{}_i \tag{27}$$

give a solution to (7) and (6). We can see this, for example, by taking (no summation on i index)

$$e^\alpha{}_i = \delta^\alpha_i (g_{ii}\epsilon_i)^{\frac{1}{2}}, \tag{28}$$

where $\epsilon_i = (1, -1, -1, -1)$.[6] From (4), this gives a diagonal g_{ij},

$$g_{ij} = g^{(0)}_{ij} g_{ii}. \tag{29}$$

From (27) we get

$$\Box(g_{ii}\epsilon_i)^{\frac{1}{2}} = -m^2(g_{ii}\epsilon_i)^{\frac{1}{2}}. \tag{30}$$

A plane wave solution to (30) may be taken as

$$g_{11} = -a^2 e^{2ip_ix^i}, \quad g_{33} = -c^2 e^{2ip_ix^i},$$
$$g_{22} = -b^2 e^{2ip_ix^i}, \quad g_{00} = d^2 e^{2ip_ix^i}, \tag{31}$$

where a, b, c, and d are constants, and it is understood that we want the real part of (31). From (28) and (5) we get

$$\Gamma^i_{jk} = \delta^i_j \frac{\partial \ln (g_{jj}\epsilon_j)^{\frac{1}{2}}}{\partial x^k}. \tag{32}$$

Inserting this in the field equation (7), we get (no summation over j or k)

$$\frac{\partial \ln (g_{jj}\epsilon_j)^{\frac{1}{2}}}{\partial x^l \partial x^k} = \frac{\partial \ln (g_{jj}\epsilon_j)^{\frac{1}{2}}}{\partial x^k} \frac{\partial \ln (g_{kk}\epsilon_k)^{\frac{1}{2}}}{\partial x^l}. \quad (33)$$

Using (31), we see that (33) is not satisfied since the left-hand side is zero, but the right-hand side is not.

Thus, the Dirac plane waves have a special significance as compared to other plane waves, as far as (7) is concerned.

VI. THE EQUATION $\Lambda^i_{jk;l} = 0$

It has been our approach, as set forth in our previous papers, to consider only linear coordinate transformations, and, thus, Γ^i_{jk} is a tensor function. Our use of the term covariant derivatives is meant as an abbreviation for certain combinations of terms that keep on recurring and which have the same formal structure as the covariant derivatives in generally covariant theories. In this section, we point out that our results have greater generality than this.

From (16), we see that $g_{ij} = g^{(0)}_{ij}$, where $g^{(0)}_{ij}$ is the Minkowski metric. Thus, all the Christoffels are zero. We may now form the object (Ricci coefficients),

$$\Lambda^i_{jk} = \Gamma^i_{jk} - \left\{ \begin{matrix} i \\ jk \end{matrix} \right\}. \quad (34)$$

Λ^i_{jk} is a tensor under general coordinate transformations.In our coordinate system $\Lambda^i_{jk} = \Gamma^i_{jk}$. Thus, in the coordinate system with which we are working, we have that[7]

$$\Lambda^i_{jk;l} = 0 \quad (35)$$

is satisfied. Since this is a tensor equation, it therefore holds in all coordinate systems.

Thus, we conclude that Dirac plane waves also satisfy the equation $\Lambda^i_{jk;l} = 0$, which is covariant under general coordinate transformations.

[1] H. S. Ruse, Proc. Roy. Soc. (Edinburgh) **57**, 97 (1937).
[2] M. Muraskin, Ann. Phys. **59**, 27 (1970). See also Erratum for correction of misprints.
[3] M. Muraskin and T. Clark, Ann. Phys. **59**, 19 (1970).
[4] M. Muraskin and B. Ring, University of North Dakota, preprint.
[5] J. Bjorken and S. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), p. 30.
[6] C. Pellegrini and J. Plebański, Mat. Fys. Skr. Dan. Vid. Selsk. **2**, 1 (1963).
[7] The author is grateful to the referee for pointing out this result.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Tidal Gravitational Accelerations near an Arbitrary Timelike Geodesic in Schwarzschild Space

J. D. FINLEY III
The University of New Mexico, Albuquerque, New Mexico 87106

(Received 27 February 1970)

It is suggested that a "physical" definition of a singularity in a space–time manifold might be that it is a point where the relative accelerations of nearby timelike geodesics become infinite. Along an arbitrary timelike geodesic in Schwarzschild space, we construct an orthonormal tetrad of 4-vectors which are used to define "elevator coordinates" in a neighborhood of the geodesic. We use these coordinates to determine the tidal gravitational accelerations near the geodesic, and we point out that these accelerations are finite (and continuous) at $r = 2m$, the "Schwarzschild surface," although they are unbounded as r approaches 0.

1. INTRODUCTION

There has been considerable discussion about the definition of a singularity in general relativity.[1,2] No one proposal is as yet accepted by everyone, perhaps because they aim at different notions. One possible "physical" definition of a singularity is obtained by considering a timelike geodesic (freely falling observer) passing through a given point P on the space–time manifold. While in the neighborhood of the point P, we imagine the observer measuring the accelerations (relative to himself) of nearby points also in free fall. If these accelerations become infinite at P, we may say that the manifold is singular there, at least in the sense that a real physical observer (larger than a point) passing through P will be torn completely apart.[3]

We will now consider this approach to singularities for the Schwarzschild solution to the Einstein field equations. (The Schwarzschild solution is particularly appropriate since the question of singularities probably originated with it. It is well known[4] that, although a component of the metric tensor becomes infinite at the "Schwarzschild surface," $r = 2m$, this surface is not a genuine singularity of the manifold, but only arises due to a poor choice of coordinates. Kruskal,[5] as well as others,[4] have introduced new coordinate systems which eliminate this apparent singularity.) For *all* timelike geodesics a natural coordinate system is constructed, used by an observer moving along that geodesic.[3,6] We then compute the components of Riemann's tensor in this coordinate system, which are related to the relative accelerations of nearby points in free fall. We find, as expected from the work of Kruskal,[5] that these accelerations remain finite at $r = 2m$, independently of how it is approached, but do become infinite at $r = 0$. In addition to the viewpoint about singularities expressed by these results, there are possible applications to gravitational collapse and small neutron stars. (Both subjects occur frequently in current literature.)

2. AN ORTHONORMAL TETRAD CARRIED ALONG BY AN ARBITRARY OBSERVER

We use the standard Schwarzschild solution to the Einstein equations in free space, which is static and spherically symmetric, with the line element

$$\begin{aligned}(ds)^2 &= g_{\mu\nu}x^\mu x^\nu \\ &= (1 - 2m/r)^{-1}(dr)^2 + r^2[(d\theta)^2 + \sin^2\theta(d\phi)^2] \\ &\quad - (1 - 2m/r)(dt)^2, \end{aligned} \tag{2.1}$$

in Schwarzschild coordinates,[7] where m is the mass of the central body which we take to have negligible size. The equations for a timelike geodesic may be written as

$$x^{\mu''} + \begin{Bmatrix} \mu \\ \nu\ \lambda \end{Bmatrix} x^{\nu'}x^{\lambda'} = 0, \tag{2.2}$$

where the prime stands for the total derivative with respect to proper time τ, with $(d\tau)^2 = -(ds)^2$, which is positive for a timelike curve. The solution to these equations always lies in a single 3-plane, and the coordinates may be chosen in such a way that this 3-plane is specified by the requirement $\theta = \frac{1}{2}\pi$. The general solution is then well known, having been worked out first by Hagihara[8] in 1931. The orbit equation for a test particle following a timelike geodesic is given by (for nonradial orbits)

$$r = 2m/\{v[\tfrac{1}{2}(\phi + \delta)]\}, \tag{2.3}$$

$$v(\psi) = \wp(\psi) + \tfrac{1}{3}, \tag{2.4}$$

where δ is a constant of integration fixing the origin of ϕ. The Weierstrass elliptic function[9] $\wp(\psi)$ is a doubly periodic, meromorphic function with a double pole at $\psi = 0 + 2n\omega + 2m\omega'$, where ω and ω' are the two periods and n and m are any integers.

[We fix the phase of the periods by the requirement $\mathrm{Im}\,(\omega'/\omega) > 0$.] The other variables, t and τ, are determined by the equations

$$\phi' = B/r^2 = 2m/\beta r^2, \tag{2.5a}$$

$$t' = A/(1 - 2m/r), \tag{2.5b}$$

where A and B are constants of integration given by the total energy per unit mass of our test particle and the angular momentum per unit mass, respectively, with $\beta = 2m/B$ as a dimensionless parameter.[10] Depending on the values of A and B, there are various types of solutions corresponding to different types of orbits.[8] There are no orbits which approach the center nearer than $r = 2m$ and then return to large distances without first going through the center, $r = 0$. We will therefore be mostly interested in those initial conditions (values of A and B) which allow our observer to penetrate closer than $r = 2m$, having originated at some fairly large distance.

We imagine that our observer (traveling along a particular timelike geodesic) carries with him a clock, measuring his proper time, and has the ability to make local measurements. We then can construct[11] a system of "elevator coordinates" y^α, such that $y^4 = \tau$, the proper time of the observer. We define these coordinates by the use of a standard orthonormal tetrad of vectors with components $V^\mu{}_{(\alpha)}$, where the index (α) picks out the four different vectors. We carry these vectors (which may be thought of as the "measuring rods and clock" of the observer) along the geodesic by parallel transport. We may then, at least in a neighborhood of the geodesic, define a natural coordinate system y^α—our elevator coordinates—for the observer by the equations

$$x^\mu = x^\mu(\tau) + V^\mu_{(a)}(\tau)y^a - \frac{1}{2}\left\{\begin{matrix}\mu\\ \nu\ \lambda\end{matrix}\right\}\Big|_\tau V^\nu_{(a)}(\tau)V^\lambda_{(b)}(\tau)y^a y^b \tag{2.6}$$

and $\tau = y^4$, where the $x^\mu(\tau)$, the tetrad components $V^\mu{}_{(a)}$, and the Christöffel symbols are to be evaluated on the geodesic as functions of τ. Therefore, we have that[11]

$$\frac{\partial x^\mu}{\partial y^\alpha} = V^\mu_{(\alpha)} + O(y^b) \tag{2.7}$$

and

$$\bar{g}_{\alpha\beta} = \eta_{\alpha\beta} + O[(y^b)^2], \tag{2.8}$$

where we indicate tensors taken with respect to the elevator coordinates by an overbar. It is clear that the geodesic itself is given by the coordinates $y^\alpha = (0, 0, 0, \tau)$.

The $V^\mu{}_{(\alpha)}$ are determined by the equations[11]

$$V^{\mu'}{}_{(\alpha)} + \left\{\begin{matrix}\mu\\ \nu\ \lambda\end{matrix}\right\}\Big|_\tau V^\nu{}_{(\alpha)}x^{\lambda'} = 0, \tag{2.9}$$

subject to the constraints

$$g_{\mu\nu}V^\mu{}_{(\alpha)}V^\nu{}_{(\beta)} = \eta_{\alpha\beta} \quad \text{and} \quad V^\mu{}_{(4)} = x^{\mu'}. \tag{2.10}$$

The general solution to these equations is given in the Appendix. We shall be interested here only in those solutions that satisfy certain initial conditions appropriate to the case under consideration.

We desire to study the observations made by a freely falling observer falling toward $r = 0$ (the center of gravitational attraction) from a considerable distance away. This requires[8] $A^2 \leqslant 1$ (i.e., originally the kinetic energy was less than the absolute value of the potential energy) and that the discriminant Δ of the elliptic function be negative. This implies[9] that $\omega = \omega'^*$ is complex; it is therefore most convenient to work with $\omega_2 \equiv \omega + \omega'$, which is real. To obtain $dr/d\tau < 0$, we must let ψ, the argument of the elliptic function, increase from $-\omega_2$ to 0 along the real axis. During this motion, r decreases from a maximum value $r_2 > 2m$ (which may be infinite) to 0, monotonically. We may therefore construct initial conditions on the $V^\mu{}_{(\alpha)}$ as if they were "measuring rods and clock" in an essentially flat space. Suitable initial conditions would be to have each of $V_{(1)}$ through $V_{(3)}$ point in different spatial directions and $V_{(4)}$ to be timelike. To maintain the orthonormality properties, therefore, we may take

$$\begin{aligned}
V_{(1)}(\phi = 0) &= (1 - v_2)^{\frac{1}{2}}(1, 0, 0, 0),\\
V_{(2)}(\phi = 0) &= (v_2/2m)(0, 1, 0, 0),\\
V_{(3)}(\phi = 0) &= v_2(1 - v_2)^{-\frac{1}{2}}(0, 0, A/2m, 1/\beta),\\
V_{(4)}(\phi = 0) &= (0, 0, v_2^2/2m\beta, A/(1 - v_2)),
\end{aligned} \tag{2.11}$$

where $v_2 = 2m/r_2$ is the original value of v (at $\phi = 0$). With these initial conditions we have (see the Appendix)

$$\begin{aligned}
V_{(1)}(\psi) &= [(v^2 + \beta^2)^{-\frac{1}{2}}/2\beta](-vv' \sin\xi\\
&\qquad + 2A\beta^2 \cos\xi, 0, (v^2 + \beta^2)(v/m)\sin\xi,\\
&\qquad \beta(2Av \sin\xi - v' \cos\xi)/(1 - v)),\\
V_{(2)}(\psi) &= (0, v/2m, 0, 0),\\
V_{(3)}(\psi) &= [(v^2 + \beta^2)^{-\frac{1}{2}}/2\beta](-vv' \cos\xi - 2A\beta^2 \sin\xi,\\
&\qquad 0, (v/m)(v^2 + \beta^2)\cos\xi,\\
&\qquad \beta(2Av \cos\xi + v' \sin\xi)/(1 - v)),\\
V_{(4)}(\psi) &= (2\beta)^{-1}(-v', 0, v^2/m, 2A\beta/(1 - v)),
\end{aligned} \tag{2.12}$$

where ξ is a real-valued function of ψ which runs from 0 to some maximum, dependent on the initial conditions, as the particle moves from r_2 to 0, and $v' = dv/d\psi$. (See the Appendix for more details.)

3. TIDAL GRAVITATIONAL ACCELERATIONS SEEN BY AN ARBITRARY OBSERVER

The observer, who is at the center of the elevator, finds that there is an apparent force acting on nearby particles; that is, nearby particles have nonzero accelerations relative to the center of the elevator. If a nearby particle has the elevator coordinates (y^a, τ), then it is well known[12] that

$$\frac{d^2}{d\tau^2} y^a|_\tau = \bar{R}^a{}_{4b4}|_\tau \, y^b, \tag{3.1}$$

where the $\bar{R}^a{}_{4b4}$ are components of Riemann's tensor in the elevator coordinate system. Intuitively, our observer will interpret this relative acceleration as a "force" acting on nearby particles. Whatever the interpretation, however, there will be a divergence (or convergence) of the world lines of nearby particles and the world line of our observer. It is proposed here that there is no "singularity" in a space–time unless the relative accelerations of nearby particles, all traveling on timelike geodesics, becomes infinite. (In pictorial terms, the observer, who, of course, is larger than a point, is torn asunder at a singularity.) We must then merely calculate Riemann's tensor in the elevator coordinate system. Using Eq. (2.8), we have that

$$\bar{R}_{\alpha\beta\gamma\delta}|_\tau = R_{\mu\nu\eta\lambda} V^\mu_{(\alpha)} V^\nu_{(\beta)} V^\eta_{(\gamma)} V^\lambda_{(\delta)}|_\tau \tag{3.2}$$

and

$$\bar{R}^\alpha{}_{\beta\gamma\delta}|_\tau = \eta^{\alpha\epsilon} \bar{R}_{\epsilon\beta\gamma\delta}|_\tau . \tag{3.3}$$

It is a lengthy but straightforward task to determine $\bar{R}^\alpha{}_{\beta\gamma\delta}$.[13] We need only the components

$$(\bar{R}^a{}_{4b4}) = \frac{v^3}{8m^2} \begin{pmatrix} 3(1 + v^2/\beta^2)\cos^2\xi - 1, & 0, & -3(1 + v^2/\beta^2)\sin\xi\cos\xi \\ 0, & -1 - 3v^2/\beta^2, & 0 \\ -3(1 + v^2/\beta^2)\sin\xi\cos\xi, & 0, & 3(1 + v^2/\beta^2)\sin^2\xi - 1 \end{pmatrix}. \tag{3.4}$$

We recall that $v = 2m/r$ so that $0 \leqslant v_2 \leq v \leq \infty$. Clearly every factor in $\bar{R}^a{}_{4b4}$ is well behaved at all finite values of v, except possibly $\xi = \xi[\psi(v)]$. But ξ is always real, so that $|\sin\xi|$ and $|\cos\xi|$ stay between 0 and 1. We see, then, that Riemann's tensor in elevator coordinates is finite for all values of r, except $r = 0$ $(v = \infty)$, for any timelike geodesic. In particular, at $r = 2m$ $(v = 1)$, $\xi[\psi(1)]$ is well defined, real, and finite, so that, as our observer passes $r = 2m$ on his way to the center, he notices no sudden change in the relative accelerations of particles near him. It is also worthwhile to point out that the eigenvalues of $\bar{R}^a{}_{4b4}$ (considered as a 3×3 matrix) are just $v^3(2 + 3v^2/\beta^2)/8m^2$, $-v^3(1 + 3v^2/\beta^2)/8m^2$, and $-v^3/8m^2$, independent of the angle ξ.

4. CONCLUSION

In a classical analysis, we may replace the relative accelerations due to tidal gravitational forces by "apparent forces." Then[12]

$$\begin{aligned} m\bar{R}^a{}_{4b4} y^b &= m \frac{d^2}{d\tau^2} y^a \\ &= F^a(y) - F^a(0) \\ &= y^b \frac{\partial}{\partial y^b} F^a + O[(y^c)^2]. \end{aligned} \tag{4.1}$$

This implies that

$$m\bar{R}^a{}_{4b4}|_\tau = \frac{\partial}{\partial y^b} F^a|_\tau . \tag{4.2}$$

We have constructed elevator coordinates for an observer on an *arbitrary* timelike geodesic and determined the form of Riemann's tensor in those coordinates. We use the "apparent force" argument above and our notion that a singularity in space–time causes infinite accelerations of nearby particles relative to a timelike observer passing through the singularity. It is seen that there are no singular points in Schwarzschild space except the origin. This is, of course, well known, but the above proof by means of an arbitrary timelike geodesic (physical observer) is new. This proof also is meant as an example of a method for detecting "physical singularities" in other space–times than this one.

At $r = 0$, $\xi(0)$ is a finite real number (see the Appendix), while v becomes infinite, so that the relative accelerations become infinite and any physical (nonpoint) particle is torn completely asunder. Therefore it seems questionable to attempt to identify the two pieces of Kruskal space in such a way that material particles rebound from the origin.[14]

Additional aspects of the Schwarzschild space near $r = 0$ are being studied, particularly the question of the physical interpretation of the various known extensions of the manifold. Extensions of this method

to other solutions of the Einstein equations are also being studied.

APPENDIX: SOLUTION OF THE TETRAD EQUATIONS

With the usual values for the Christöffel symbols in Schwarzschild coordinates,[10] Eq. (2.9) becomes

$$V^{1'}_{(\alpha)} - (v^2/4m)(1-v)^{-1}V^1_{(\alpha)}r' - (v/\beta)(1-v)V^3_{(\alpha)} + (A/4m)v^2V^4_{(\alpha)} = 0, \quad \text{(A1)}$$

$$V^{2'}_{(\alpha)} + (r'/r)V^2_{(\alpha)} = 0, \quad \text{(A2)}$$

$$V^{3'}_{(\alpha)} + (v^3/4m^2\beta)V^1_{(\alpha)} + (v/2m)V^3_{(\alpha)}r' = 0, \quad \text{(A3)}$$

$$V^{4'}_{(\alpha)} + (v^2/4m)(1-v)^{-1}V^4_{(\alpha)}r' + (Av^2/4m)(1-v)^{-2}V^1_{(\alpha)} = 0. \quad \text{(A4)}$$

Equation (A2) has the immediate solution

$$V^2_{(\alpha)}(\psi) = (2m)^{-1}D_{(\alpha)}v(\psi). \quad \text{(A5)}$$

Now we use the orthonormality conditions [Eq. (2.10)] to obtain (remembering that Latin indices vary from 1 to 3)

$$AV^4_{(a)} = BV^3_{(a)} + (1-v)^{-1}V^1_{(a)}r'. \quad \text{(A6)}$$

Combining this with Eqs. (A1) and (A3), we obtain [via Eq. (2.5a)]

$$V^{1'}_{(a)} = (v/\beta)(1-3v/2)V^3_{(a)} = (1-3v/2)(rV^3_{(a)})\phi' \quad \text{(A7)}$$

and

$$(rV^3_{(a)})' = -(v^2/2m\beta)V^1_{(a)} = -V^1_{(a)}\phi', \quad \text{(A8)}$$

from which [using Eq. (2.4)] we have Lamé's equation [with $x = rV^3_{(a)}$ and $\psi = \frac{1}{2}(\phi + \delta)$]:

$$\frac{d^2}{d\psi^2}x = [6\wp(\psi) - 2]x. \quad \text{(A9)}$$

The solutions are well known[15]; they are the Lamé functions $\Lambda_1(\psi)$ and $\Lambda_2(\psi)$, which can be written in terms of functions related to the Weierstrass elliptic function $\wp$. Let $\zeta(\psi)$ and $\sigma(\psi)$ be defined by the equations[9]

$$\zeta(\psi) = \psi^{-1} + \int_0^\psi du[u^{-2} - \wp(u)], \quad \text{(A10)}$$

$$\sigma(\psi) = \psi \exp\left(\int_0^\psi du[\zeta(u) - u^{-1}]\right). \quad \text{(A11)}$$

The functions ζ and σ are quasiperiodic, and ζ is a meromorphic function with a single pole at every period point, while σ is an entire function which vanishes at every period point. Then linearly independent solutions of Lamé's equation are

$$\sigma(\psi \pm a)\sigma(\psi \pm b)\sigma^{-2}(\psi)e^{\psi[\zeta(a)+\zeta(b)]},$$

where $\wp(a) + \wp(b) = -\frac{2}{3}$ and $\wp'(a) + \wp'(b) = 0$. Solutions for these constraint equations in terms of A and β are given by

$$\wp(b) = -\tfrac{1}{3} - i\beta, \quad \wp'(b) = -2A\beta, \quad a = -b^*. \quad \text{(A12)}$$

It is convenient to normalize these functions so that

$$\Lambda_1(\psi) = \Lambda_2^*(\psi^*) = \frac{\sigma(\psi+b)\sigma(\psi-b^*)}{|\sigma(b)|^2\,\sigma^2(\psi)}\,e^{-2i\psi\,\mathrm{Im}\,[\zeta(b)]}. \quad \text{(A13)}$$

From the standard "addition theorem" for the σ functions[9] and Eqs. (2.4), (A11), and (A12), we have that

$$\Lambda_2(\psi)\Lambda_1'(\psi) = v v' - 2iA\beta^2 \quad \text{(A14)}$$

and

$$\Lambda_1(\psi)\Lambda_2(\psi) = v^2(\psi) + \beta^2. \quad \text{(A15)}$$

From Eqs. (A6), (A8), (A9), and (A14), we can complete the orthonormal tetrad [along with Eq. (A5)]:

$$\begin{aligned} V^1_{(a)} &= -\tfrac{1}{2}(v^2+\beta^2)^{-1}[vv'(C_{(a)}\Lambda_1 + \bar{C}_{(a)}\Lambda_2) - 2iA\beta^2(C_{(a)}\Lambda_1 - \bar{C}_{(a)}\Lambda_2)], \\ V^3_{(a)} &= (C_{(a)}\Lambda_1 + \bar{C}_{(a)}\Lambda_2)(v/2m), \\ V^4_{(a)} &= \tfrac{1}{2}[(1-v)(v^2+\beta^2)]^{-1} \times [2Av(C_{(a)}\Lambda_1 + \bar{C}_{(a)}\Lambda_2) - iv'(C_{(a)}\Lambda_1 - \bar{C}_{(a)}\Lambda_2)]. \end{aligned} \quad \text{(A16)}$$

The nine constants in Eqs. (A5) and (A16) are still subject to the orthonormality conditions [Eq. (2.10)]. These imply the following relations between them (only six of which are independent):

$$\begin{aligned} 1 - (D_{(a)})^2 &= 4\beta^2 C_{(a)}\bar{C}_{(a)}, \\ -D_{(a)}D_{(b)} &= 2\beta^2(C_{(a)}\bar{C}_{(b)} + \bar{C}_{(a)}C_{(b)}), \quad a \neq b. \end{aligned} \quad \text{(A17)}$$

Since we are considering only real ψ, we can write

$$\Lambda_1(\psi) = \Lambda_2^*(\psi) = [v^2(\psi) + \beta^2]^{\frac{1}{2}}e^{i[\xi(\psi)-\chi]}, \quad \text{(A18)}$$

where

$$e^{2i[\xi(\psi)-\chi]} = \frac{\sigma(\psi+b)\sigma(\psi-b^*)}{\sigma(\psi+b^*)\sigma(\psi-b)}\,e^{-4i\psi\,\mathrm{Im}[\xi(b)]} \quad \text{(A19)}$$

and

$$\tfrac{1}{2}\chi = \eta_2\,\mathrm{Im}\,(b) - \omega_2\,\mathrm{Im}\,[\zeta(b)], \quad \eta_2 \equiv \zeta(\omega_2). \quad \text{(A20)}$$

This choice of the constant phase χ and our initial conditions makes the final answer depend only on $\xi(\psi)$, which is real for real ψ and is normalized so that $\xi(-\omega_2) = 0$ and $\xi(0) \triangleq \chi$. [It should be pointed out that b is always complex and Im (b) is not equal to a multiple of a period, so that $\sigma(\psi \pm b)$ can never

vanish for real ψ.] Since v, v', and $V^{\mu}_{(\alpha)}$ must all be real, we have that the $D_{(a)}$ are real and $C_{(a)} = \bar{C}^*_{(a)}$. Fitting to our initial conditions [Eq. (2.9)], we find

$$D_{(a)} = \delta^2_a, \quad 2\beta C_{(3)} = 2i\beta C_{(1)} = e^{+i\chi}, \quad C_{(2)} = 0. \tag{A21}$$

Insertion of these values in Eq. (A16) yields the desired tetrad solutions given in Eq. (2.12).

[1] R. Geroch, Ann. Phys. (N.Y.) **48**, 526 (1968); *Battelle Rencontres*, edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968), Chap. VIII.

[2] See also R. Penrose, Phys. Rev. Letters **14**, 57 (1965); S. W. Hawking, Proc. Roy. Soc. (London) **A294**, 511 (1966); **A295**, 490 (1966); **A300**, 187 (1967); P. Hájiček, Helv. Phys. Acta **42**, 808 (1969).

[3] This has been pointed out, in the case of radial geodesics only, in unpublished lecture notes by C. W. Misner, K. S. Thorne, and J. A. Wheeler, where a similar definition of a singularity is used.

[4] G. Lemaitre, Ann. Soc. Sci. Bruxelles **A53**, 51 (1933) is the first place this is clearly pointed out. Several other extensions of the manifold without coordinate singularities (except at $r = 0$) have been given. For instance, see J. L. Synge, Proc. Roy. Irish Acad. **A53**, 83 (1950); D. Finkelstein, Phys. Rev. **110**, 965 (1958); or P. I. Fomin, Zh. Eksp. Teor. Fiz. **54**, 905 (1968) [Sov. Phys.—JETP **27**, 483 (1968)].

[5] The maximal analytic completion of the Schwarzschild manifold was given by M. D. Kruskal [Phys. Rev. **119**, 1743 (1960)].

[6] R. W. Fuller and J. A. Wheeler [Phys. Rev. **128**, 919 (1962)] study the properties of *radial* geodesics in connection with these problems. It would seem that no one has previously used arbitrary (nonradial) timelike geodesics.

[7] We use a metric which has signature +2; Greek indices take values from 1 to 4 and are summed according to the Einstein summation convention, while Latin indices take only the values 1 to 3. We use $\eta_{\alpha\beta}$ for the metric tensor for flat space-time, which has diagonal elements (1, 1, 1, −1). We also use units such that $c = 1 = G$.

[8] Y. Hagihara, Japan. J. Astron. Geophys. **8**, 67 (1931). Along the same lines, see also A. F. Bogorodskii, *Uravneniya polya Einshteina i· ikh primenenie v astronomii* (Kiev University, 1962); B. Mielnik and J. Plebanski, Acta Phys. Polon. **21**, 239 (1962).

[9] For an excellent review of the properties of the Weierstrass elliptic function and related functions, see H. Hancock, *Lectures on the Theory of Elliptic Functions* (Dover, New York, 1958).

[10] H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1968), pp. 234–38.

[11] E. Fermi, Rend. Acad. Nazl. Lincei **31**, 21, 51, 101 (1922); H. P. Robertson and T. W. Noonan, Ref. 10, pp. 203–8.

[12] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), p. 162.

[13] To do the calculation, one needs the components of Riemann's tensor in the Schwarzschild coordinate system $R_{\mu\nu\lambda\eta}$, the components of V^{μ}_{α} given by Eq. (2.12), and the relation between v', v, A, and β determined by the Weierstrass differential equation

$$\tfrac{1}{4}(v')^2 + (1 - v)(v^2 + \beta^2) = (A\beta)^2$$

(see Ref. 9). The components of Riemann's tensor are most conveniently given by

$$R_{\mu\nu\lambda\eta} = g_{\mu\rho}g_{\nu\sigma}R^{\rho\sigma}{}_{\lambda\eta} \quad \text{with} \quad R^{14}{}_{14} = R^{23}{}_{23} = -2m/r^3$$
$$= -v^3/4\,m^2 = -2R^{12}{}_{12} = -2R^{13}{}_{13} = -2R^{24}{}_{24} = -2R^{34}{}_{34}.$$

[14] W. Rindler, Phys. Rev. Letters **15**, 1001 (1965); W. Israel, Nature **211**, 466 (1966); Phys. Rev. **143**, 1016 (1966).

[15] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1965), 4th ed., pp. 554 and 572.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Field-Theoretic Description of Massless Particles with Higher Spin and Definite Parity. II. Half-Integer Spin

T. CUKIERDA
Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Cybulskiego 36, Poland

(Received 13 January 1970; Revised Manuscript Received 26 March 1970)

The $S + \frac{1}{2}$ equivalent theories describing a free massive field with half-integer spin S in the limit $m \to 0$ are investigated. It is shown that for $m = 0$ they are *not* equivalent and describe different massless particles, with helicities $\mathcal{H} = \pm\frac{1}{2}, \pm\frac{3}{2}, \cdots, \pm S$. The inequivalence is caused by the possibility of having different gauge transformations for different massless spinor-tensors $G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\Lambda}$.

1. INTRODUCTION

In the previous paper[1] it was shown that if we put in the $S + 1$ equivalent covariant formulations of the theory of a free massive particle with integer spin[2] S, the value of mass m equal to zero, all the formulations describe different $S + 1$ massless particles. This result is caused by the fact that, in the case $m = 0$, besides the notion of Poincaré invariance, one can also introduce the gauge transformations that do not change the field equations. Inequivalent formulations obtained after putting $m = 0$ are caused by different possibilities of choice of the gauge transformations.

The case of arbitrary half-integer spin $S = \frac{3}{2}, \frac{5}{2}, \cdots$ is considered in this note. We use $S + \frac{1}{2}$ equivalent formulations of the theory of a massive particle with half-integer spin S, which were first introduced by the author. This set of $S + \frac{1}{2}$ equivalent equations are given in Sec. 2. As the fundamental fields we use the spinor-tensors $G^{(k,S)}_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_{S-\frac{1}{2}}\nu_{S-\frac{1}{2}}]\Lambda}$. Any such spinor-tensor is symmetric with respect to any permutations $\sigma_i \leftrightarrow \sigma_j$ and $[\mu_m\nu_m] \leftrightarrow [\mu_n\nu_n]$ and is

vanish for real ψ.] Since v, v', and $V^{\mu}_{(\alpha)}$ must all be real, we have that the $D_{(a)}$ are real and $C_{(a)} = \bar{C}^*_{(a)}$. Fitting to our initial conditions [Eq. (2.9)], we find

$$D_{(a)} = \delta^2_a, \quad 2\beta C_{(3)} = 2i\beta C_{(1)} = e^{+i\chi}, \quad C_{(2)} = 0. \tag{A21}$$

Insertion of these values in Eq. (A16) yields the desired tetrad solutions given in Eq. (2.12).

[1] R. Geroch, Ann. Phys. (N.Y.) **48**, 526 (1968); *Battelle Rencontres*, edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968), Chap. VIII.

[2] See also R. Penrose, Phys. Rev. Letters **14**, 57 (1965); S. W. Hawking, Proc. Roy. Soc. (London) **A294**, 511 (1966); **A295**, 490 (1966); **A300**, 187 (1967); P. Hájiček, Helv. Phys. Acta **42**, 808 (1969).

[3] This has been pointed out, in the case of radial geodesics only, in unpublished lecture notes by C. W. Misner, K. S. Thorne, and J. A. Wheeler, where a similar definition of a singularity is used.

[4] G. Lemaitre, Ann. Soc. Sci. Bruxelles **A53**, 51 (1933) is the first place this is clearly pointed out. Several other extensions of the manifold without coordinate singularities (except at $r = 0$) have been given. For instance, see J. L. Synge, Proc. Roy. Irish Acad. **A53**, 83 (1950); D. Finkelstein, Phys. Rev. **110**, 965 (1958); or P. I. Fomin, Zh. Eksp. Teor. Fiz. **54**, 905 (1968) [Sov. Phys.—JETP **27**, 483 (1968)].

[5] The maximal analytic completion of the Schwarzschild manifold was given by M. D. Kruskal [Phys. Rev. **119**, 1743 (1960)].

[6] R. W. Fuller and J. A. Wheeler [Phys. Rev. **128**, 919 (1962)] study the properties of *radial* geodesics in connection with these problems. It would seem that no one has previously used arbitrary (nonradial) timelike geodesics.

[7] We use a metric which has signature +2; Greek indices take values from 1 to 4 and are summed according to the Einstein summation convention, while Latin indices take only the values 1 to 3. We use $\eta_{\alpha\beta}$ for the metric tensor for flat space-time, which has diagonal elements (1, 1, 1, −1). We also use units such that $c = 1 = G$.

[8] Y. Hagihara, Japan. J. Astron. Geophys. **8**, 67 (1931). Along the same lines, see also A. F. Bogorodskii, *Uravneniya polya Einshteina i· ikh primenenie v astronomii* (Kiev University, 1962); B. Mielnik and J. Plebanski, Acta Phys. Polon. **21**, 239 (1962).

[9] For an excellent review of the properties of the Weierstrass elliptic function and related functions, see H. Hancock, *Lectures on the Theory of Elliptic Functions* (Dover, New York, 1958).

[10] H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1968), pp. 234–38.

[11] E. Fermi, Rend. Acad. Nazl. Lincei **31**, 21, 51, 101 (1922); H. P. Robertson and T. W. Noonan, Ref. 10, pp. 203–8.

[12] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), p. 162.

[13] To do the calculation, one needs the components of Riemann's tensor in the Schwarzschild coordinate system $R_{\mu\nu\lambda\eta}$, the components of V^{μ}_{α} given by Eq. (2.12), and the relation between v', v, A, and β determined by the Weierstrass differential equation

$$\tfrac{1}{4}(v')^2 + (1 - v)(v^2 + \beta^2) = (A\beta)^2$$

(see Ref. 9). The components of Riemann's tensor are most conveniently given by

$$R_{\mu\nu\lambda\eta} = g_{\mu\rho}g_{\nu\sigma}R^{\rho\sigma}{}_{\lambda\eta} \quad \text{with} \quad R^{14}{}_{14} = R^{23}{}_{23} = -2m/r^3$$
$$= -v^3/4\,m^2 = -2R^{12}{}_{12} = -2R^{13}{}_{13} = -2R^{24}{}_{24} = -2R^{34}{}_{34}.$$

[14] W. Rindler, Phys. Rev. Letters **15**, 1001 (1965); W. Israel, Nature **211**, 466 (1966); Phys. Rev. **143**, 1016 (1966).

[15] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1965), 4th ed., pp. 554 and 572.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Field-Theoretic Description of Massless Particles with Higher Spin and Definite Parity. II. Half-Integer Spin

T. CUKIERDA
Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Cybulskiego 36, Poland

(Received 13 January 1970; Revised Manuscript Received 26 March 1970)

The $S + \frac{1}{2}$ equivalent theories describing a free massive field with half-integer spin S in the limit $m \to 0$ are investigated. It is shown that for $m = 0$ they are *not* equivalent and describe different massless particles, with helicities $\mathcal{H} = \pm\frac{1}{2}, \pm\frac{3}{2}, \cdots, \pm S$. The inequivalence is caused by the possibility of having different gauge transformations for different massless spinor-tensors $G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\Lambda}$.

1. INTRODUCTION

In the previous paper[1] it was shown that if we put in the $S + 1$ equivalent covariant formulations of the theory of a free massive particle with integer spin[2] S, the value of mass m equal to zero, all the formulations describe different $S + 1$ massless particles. This result is caused by the fact that, in the case $m = 0$, besides the notion of Poincaré invariance, one can also introduce the gauge transformations that do not change the field equations. Inequivalent formulations obtained after putting $m = 0$ are caused by different possibilities of choice of the gauge transformations.

The case of arbitrary half-integer spin $S = \frac{3}{2}, \frac{5}{2}, \cdots$ is considered in this note. We use $S + \frac{1}{2}$ equivalent formulations of the theory of a massive particle with half-integer spin S, which were first introduced by the author. This set of $S + \frac{1}{2}$ equivalent equations are given in Sec. 2. As the fundamental fields we use the spinor-tensors $G^{(k,S)}_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_{S-\frac{1}{2}}\nu_{S-\frac{1}{2}}]\Lambda}$. Any such spinor-tensor is symmetric with respect to any permutations $\sigma_i \leftrightarrow \sigma_j$ and $[\mu_m\nu_m] \leftrightarrow [\mu_n\nu_n]$ and is

antisymmetric with respect to the indices in the square bracket. Index A has four values and, using Weyl's notation, we get

$$G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]A} \equiv \begin{pmatrix} G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\alpha} \\ G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{\alpha}} \end{pmatrix} \equiv \begin{pmatrix} G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]1} \\ G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]2} \\ G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{1}} \\ G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{2}} \end{pmatrix}, \quad (1.1)$$

i.e., with respect to the spinor indices, $G^{(k,S)}$ transforms as Dirac's 4-spinor.

The main result is obtained in Sec. 3, where the case $m = 0$ is investigated and the gauge transformations are defined. It is shown explicitly that in the massless case we are led to $S + \frac{1}{2}$ different theories. We see, therefore, that the result obtained by Ogievetski and Polubarinov[3] for the spin-1 case can be generalized to any finite-dimensional irreducible representation of the Lorentz group.

2. ARBITRARY HALF-INTEGER SPIN: MASSIVE CASE, $m \neq 0$

In this section we show how to generalize the Rarita–Schwinger formalism to the case when, as the fundamental field, the spinor–tensors $G^{(k,s)}_{\cdots\sigma\cdots[\mu\nu]A}$ are used.

The field $G^{(k,S)}$ describing massive particle with spin S should satisfy the Dirac equation

$$(i\partial_\rho\gamma^\rho + m)G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0 \quad (2.1)$$

and the following six subsidiary conditions:

$$\partial^\sigma G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (2.2)$$

$$\partial^\mu G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (2.3)$$

$$\gamma^\sigma G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (2.4)$$

$$G^{(k,S)}_{\cdots\sigma[\sigma\rho]\cdots[\mu\nu]} = 0, \quad (2.5)$$

$$\gamma^\mu\gamma^\nu\gamma^\rho G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu][\rho\tau]} = 0. \quad (2.6a)$$

Because Eq. (2.6a) can be written only for $k < S - \frac{3}{2}$ if $k = S - \frac{3}{2}$, we use the following equation:

$$\partial^\mu\gamma^\nu G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0. \quad (2.6b)$$

We shall now discuss the set of Eqs. (2.1)–(2.6) in the rest system with the spacelike components of the 4-momentum vector of the particles equal to zero; i.e., we assume that $p = (0, 0, 0, m)$. We hold that (2.1)–(2.4) can be written in momentum space in our particular coordinate system as follows:

$$G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{\alpha}}(p) = 0, \quad (2.7)$$

$$G^{(k,S)}_{\cdots0\cdots[\mu\nu]\alpha}(p) = 0, \quad (2.8)$$

$$G^{(k,S)}_{\cdots\sigma\cdots[0\nu]\alpha}(p) = 0, \quad (2.9)$$

$$G^{(k,S)}_{\cdots1\cdots[\mu\nu]2}(p) - iG^{(k,S)}_{\cdots2\cdots[\mu\nu]2}(p) + G^{(k,S)}_{\cdots3\cdots[\mu\nu]1}(p) = 0,$$
$$G^{(k,S)}_{\cdots1\cdots[\mu\nu]1}(p) + iG^{(k,S)}_{\cdots2\cdots[\mu\nu]1}(p) + G^{(k,S)}_{\cdots3\cdots[\mu\nu]2}(p) = 0. \quad (2.10)$$

From the Eq. (2.5) we obtain

$$G^{(k,S)}_{\cdots2\cdots[12]1}(p) - G^{(k,S)}_{\cdots1\cdots[13]2}(p) + iG^{(k,S)}_{\cdots2\cdots[13]2}(p) = 0,$$
$$G^{(k,S)}_{\cdots2\cdots[12]2}(p) + G^{(k,S)}_{\cdots1\cdots[13]1}(p) + iG^{(k,S)}_{\cdots2\cdots[13]1}(p) = 0. \quad (2.11)$$

Equations (2.6a) and (2.6b) give, for arbitrary k,

$$G^{(k,S)}_{\cdots\sigma\cdots[23]1}(p) - iG^{(k,S)}_{\cdots\sigma\cdots[13]1}(p) - G^{(k,S)}_{\cdots\sigma\cdots[12]2}(p) = 0,$$
$$G^{(k,S)}_{\cdots\sigma\cdots[23]2}(p) + iG^{(k,S)}_{\cdots\sigma\cdots[13]2}(p) + G^{(k,S)}_{\cdots\sigma\cdots[12]1}(p) = 0. \quad (2.12)$$

On the basis of the relations (2.7)–(2.12), one can easily see that the set of Eqs. (2.1)–(2.6) has, for $p_0 = m$, only $2S + 1$ linearly independent solutions that describe massive particle with a definite half-integer spin S.

3. ARBITRARY HALF-INTEGER SPIN: MASSLESS CASE, $m = 0$

Let us put $m = 0$ in Eqs. (2.1)–(2.6). We obtain for the massless case the following set of equations of motion:

$$\partial_\rho\gamma^\rho G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (3.1)$$

$$\partial^\sigma G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (3.2)$$

$$\partial^\mu G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (3.3)$$

$$\gamma^\sigma G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0, \quad (3.4)$$

$$G^{(k,S)}_{\cdots\sigma\cdots[\sigma\nu]} = 0, \quad (3.5)$$

$$\gamma^\mu\gamma^\nu\gamma^\rho G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu][\rho\tau]} = 0, \quad (3.6a)$$

and for $k = S - \frac{3}{2}$,

$$\partial^\mu\gamma^\nu G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} = 0. \quad (3.6b)$$

Choosing $p = (|\mathbf{p}|, 0, 0, |\mathbf{p}|)$, from (3.1)–(3.4) we obtain

$$G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]1}(p) = -G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{1}}(p),$$
$$G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]2}(p) = G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]\dot{2}}(p), \quad (3.7)$$

$$G^{(k,S)}_{\cdots3\cdots[\mu\nu]A}(p) = G^{(k,S)}_{\cdots0\cdots[\mu\nu]A}(p), \quad (3.8)$$

$$G^{(k,S)}_{\cdots\sigma\cdots[3\nu]A}(p) = G^{(k,S)}_{\cdots\sigma\cdots[0\nu]A}(p), \quad (3.9)$$

and

$$G^{(k,S)}_{\cdots1\cdots[\mu\nu]1}(p) = -iG^{(k,S)}_{\cdots2\cdots[\mu\nu]1}(p),$$
$$G^{(k,S)}_{\cdots1\cdots[\mu\nu]2}(p) = iG^{(k,S)}_{\cdots2\cdots[\mu\nu]2}(p). \quad (3.10)$$

Taking into account (3.8) and (3.9), Eq. (3.5) may be reduced to the following form:

$$G^{(k,S)}_{\cdots i\cdots[iv]A}(p) = 0, \quad i = 1, 2, \quad (3.11)$$

from which it follows that

$$G^{(k,S)}_{\cdots i\cdots[12]A}(p) = 0, \quad i = 1, 2, \quad (3.12)$$

and

$$G^{(k,S)}_{\cdots 1\cdots[13]_A}(p) + G^{(k,S)}_{\cdots 2\cdots[23]_A}(p) = 0. \quad (3.13)$$

Finally, using (3.8) and (3.9), we see that Eq. (3.6a) gives

$$\gamma^i G^{(k,S)}_{\cdots\sigma\cdots[12][iv]}(p) = 0, \quad i = 1, 2, \quad (3.14)$$

or

$$G^{(k,S)}_{\cdots\sigma\cdots[12][12]_A}(p) = 0, \quad (3.15)$$

$$G^{(k,S)}_{\cdots\sigma\cdots[12][13]1}(p) = -iG^{(k,S)}_{\cdots\sigma\cdots[12][23]1}(p),$$
$$G^{(k,S)}_{\cdots\sigma\cdots[12][13]2}(p) = iG^{(k,S)}_{\cdots\sigma\cdots[12][23]2}(p). \quad (3.16)$$

Equation (3.6b) is satisfied identically.

From these considerations we see that the following components of the field $G^{(k,s)}$ are linearly independent:

$$G^{(k,S)}_{3\cdots 3[12][13]\cdots[13]\alpha}(p), \quad (3.17)$$

$$G^{(k,S)}_{3\cdots 3[r_{k+1}q_{k+1}]\cdots[r_{s-\frac{1}{2}}q_{S-\frac{1}{2}}]\alpha}(p), \quad [rq] = [13], [12], \quad (3.18)$$

and

$$G^{(k,S)}_{1\sigma_2\cdots\sigma_k[13]\cdots[13]\alpha}(p), \quad \sigma = 1, 3. \quad (3.19)$$

Other components are identically equal to zero or can be expressed as linear combinations of the components (3.17)–(3.19). The components (3.17)–(3.19) are characterized by different helicities and describe several massless particles; some of them, however, are nonphysical, because they can be eliminated by means of the gauge transformations, leaving the field equations invariant. One can introduce the gauge transformations in such a way that only two components with defined helicity remain. These gauge transformations are essentially different for the case $k = S - \frac{1}{2}$ and $k < S - \frac{1}{2}$. In the first case (if $k = S - \frac{1}{2}$, then $G^{(S-\frac{1}{2},S)} \equiv G^{(S-\frac{1}{2})}$), they are defined as follows:

$$G^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}} \to G^{(S-\frac{1}{2})'}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}} = G^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}} + \delta G^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}},$$
$$\delta G^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}} = \sum_{\substack{\text{perm}\\(\sigma_i\sigma_j)}} \partial_{\sigma_i}\Lambda^{(S-\frac{3}{2})}_{\sigma_1\cdots\sigma_j\cdots\sigma_{S-\frac{1}{2}}}, \quad (3.20)$$

where

$$\partial_\rho\gamma^\rho\Lambda^{(S-\frac{3}{2})}_{\sigma_1\cdots\sigma_{S-\frac{3}{2}}} = 0,$$
$$\partial^{\sigma_k}\Lambda^{(S-\frac{3}{2})}_{\sigma_1\cdots\sigma_k\cdots\sigma_{S-\frac{3}{2}}} = 0, \quad (3.21)$$
$$\gamma^{\sigma_k}\Lambda^{(S-\frac{3}{2})}_{\sigma_1\cdots\sigma_k\cdots\sigma_{S-\frac{3}{2}}} = 0.$$

Because in the rest system

$$\delta G^{(S-\frac{1}{2})}_{\underbrace{3\cdots 3}_{r}\,1\cdots 1,A}(p) = r\,|\mathbf{p}|\,\Lambda_{\underbrace{3\cdots 3}_{r-1}\,1\cdots 1,A}(p), \quad (3.22)$$

we see that only two components $G^{(S-\frac{1}{2})}_{1\cdots 1,\alpha}(p)$ are gauge invariant and describe the massless particle with helicity $\mathcal{H} = \pm S$.

In the second case ($k < S - \frac{1}{2}$), the spinor-tensor $\Lambda^{(S-\frac{1}{2})}$ is used as the gauged field. Now the gauge transformation has the form

$$G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} \to G^{(k,S)'}_{\cdots\sigma\cdots[\mu\nu]} = G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]} + \delta G^{(k,S)}_{\cdots\sigma\cdots[\mu\nu]},$$
$$\delta G^{(k,S)}_{\sigma_1\cdots\sigma_k[\mu_{k+1}\nu_{k+1}]\cdots[\mu_{S-\frac{1}{2}}\nu_{S-\frac{1}{2}}]}$$
$$= \sum_{\text{perm}} (-1)^n\partial_{\mu_{k+1}}\cdots\partial_{\mu_{S-\frac{1}{2}}}\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_k\nu_{k+1}\cdots\nu_{S-\frac{1}{2}}}, \quad (3.23)$$

where $\sum_{\text{perm}}$ denotes sum over all permutations of pairs $\nu_{k+i} \leftrightarrow \mu_{k+i}$ and n describes the number of such permutations. The field $\Lambda^{(S-\frac{1}{2})}$ must satisfy the following conditions:

$$\partial_\rho\gamma^\rho\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{S-\frac{1}{2}}} = 0, \quad (3.24)$$

$$\partial^{\sigma_k}\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_k\cdots\sigma_{S-\frac{1}{2}}} = 0, \quad (3.25)$$

$$\gamma^{\sigma_k}\sum_{\text{perm}} (-1)^n\partial_{\mu_{k+1}}\cdots\partial_{\mu_{S-\frac{1}{2}}}\Lambda_{\sigma_1\cdots\sigma_k\nu_{k+1}\cdots\nu_{S-\frac{1}{2}}} = 0. \quad (3.26)$$

Equations (3.24) and (3.25) are identical with (3.1) and (3.2); however, Eq. (3.26) gives

$$\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{k-1}i_{k+1}\cdots i_{S-\frac{1}{2}},1}(p) = -i\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{k-1}2i_{k+1}\cdots i_{S-\frac{1}{2}},1}(p),$$
$$\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{k-1}1i_{k+1}\cdots i_{S-\frac{1}{2}},2}(p) = i\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_{k-1}2i_{k+1}\cdots i_{S-\frac{1}{2}},2}(p),$$
$$i = 1, 2. \quad (3.27)$$

With respect to the gauge transformations defined in such a way, only the following two components with helicity $\mathcal{H} = \pm(S - k - \frac{3}{2})$ are gauge invariant:

$$G^{(k,S)}_{3\cdots 3[12][13]\cdots[13]\alpha}(p).$$

Because

$$\delta G^{(k,S)}_{\sigma_1\cdots\sigma_k[i_{k+1}3]\cdots[i_{S-\frac{1}{2}}3]}(p) = |\mathbf{p}|^{S-k-\frac{1}{2}}\Lambda^{(S-\frac{1}{2})}_{\sigma_1\cdots\sigma_k i_{k+1}\cdots i_{S-\frac{1}{2}}}(p), \quad (3.28)$$

the remaining components (3.18) and (3.19) may be eliminated by suitable choice of the gauge function $\Lambda^{(S-\frac{1}{2})}$. It has been explicitly demonstrated, therefore, that the gauge transformations are able to restrict the number of nonvanishing components to a pair with two opposite helicities.

4. CONCLUSIONS

One can summarize our results in the following three points:

(i) There exist $S + \frac{1}{2}$ equivalent theories describing massive particles with half-integer spin S. In these theories the spinor–tensors $G^{(k,S)}$ are used as the fundamental fields.

(ii) In the case $m = 0$, one can introduce in these theories the gauge transformations in such a way that they describe $S + \frac{1}{2}$ different massless particles with helicities $\mathcal{H} = \pm\frac{1}{2}, \pm\frac{3}{2}, \cdots, \pm S$.

(iii) One can conclude that there exist an infinite number of ways of introducing the free theory of

massless particles with given helicity $\mathcal{H}$ by means of the spinor–tensors $G^{(\mathcal{H}-\frac{1}{2})}_{\sigma_1\cdots\sigma_{\mathcal{H}-\frac{1}{2}}}$ and

$$G^{(r-\mathcal{H}-1,r)}_{\sigma_1\cdots\sigma_{r-\mathcal{H}-1}[\mu_{r-\mathcal{H}}{}^{\nu}{}_{r-\mathcal{H}}\cdots[\mu_{r-\frac{1}{2}}\nu_{r-\frac{1}{2}}]}$$

where $r = \mathcal{H}+1, \mathcal{H}+2, \cdots$.

Comparing these three statements with the results obtained in Ref. 1, one can say that there exists a similarity between theories of the particles with integer spin and half-integer spin. Such similarity is clear on the grounds of the representation theory of the Lorentz group and the wavefunction formalisms for higher-spin particles.

Our investigations in Ref. 1 and in this paper are restricted only to the classical field equations. The limit $m = 0$ in quantum theory is much more complicated because of the problem of spurious divergences, coming from the factors in the Green's functions m^{-2k}. These problems will be studied in subsequent work.

ACKNOWLEDGMENT

I am grateful to Dr. J. Lukierski for his interest and stimulating conversations.

[1] T. Cukierda and J. Lukierski, J. Math. Phys. **11**, 46 (1970).
[2] M. Fierz, Helv. Phys. Acta **12**, 1 (1939).
[3] W. I. Ogievetskiĭ and I. W. Polubarinov, Yad. Fiz. 216 (1966) [Sov. J. Nucl. Phys. **4**, 156 (1967)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Wigner Coefficients for the Group $E(2)$

H. P. W. GOTTLIEB
Department of Applied Mathematics and Theoretical Physics
University of Cambridge, England

(Received 29 December 1969; Revised Manuscript Received 29 April 1970)

The Wigner coefficients for the faithful unitary representations of the 2-dimensional Euclidean group are derived from two identities involving Bessel functions. Since the multiplicity in the decomposition of a direct product is two, we find two sets of coefficients which are real and mutually orthogonal and symmetric and antisymmetric, respectively, under interchange of constituent representations. Properties of the coefficients at the ends of the decomposition spectrum are discussed.

I. INTRODUCTION

There is increasing interest in the role which the 2-dimensional Euclidean group $E(2)$, the group of rotations and translations in the plane, plays in elementary particle physics. As the isomorph of the little group for lightlike particles in the Wigner classification of irreducible representations of the Poincaré group,[1] its faithful representations, with infinite spin spectrum[2] ("indefinite helicity"), do not appear to be realized as free particles in nature. However, in the formalism of Toller,[3] the amplitude for the case of zero momentum transfer square (as distinct from null momentum transfer) should be expansible in terms of representation functions of $E(2)$. A study of the $E(2)$ representations seems essential for an appreciation of the transition from timelike to spacelike momentum transfer. The group $E(2)$ also occurs naturally in the study of the infinite-momentum frame limit of kinematics.[4]

In order to gain a deeper understanding of the nature of the representations of $E(2)$, we here derive the Wigner or generalized Clebsch–Gordan coefficients. These may be deduced from some well-known identities for Bessel functions, the main difficulty being caused by the multiplicity of irreducible representations in the direct product. Two sets of coefficients which are orthogonal and can be chosen to be real are required.

II. REPRESENTATIONS OF $E(2)$ AND BESSEL FUNCTIONS

The faithful unitary irreducible representations of $E(2)$ may be realized over a discrete basis and written in the form[2]

$$D^{X\lambda}_{m,n}(\mathbf{b}, \phi) = e^{-i(n+\lambda)\phi}e^{i(n-m)\beta}(i)^{n-m}J_{m-n}(bX), \quad (1)$$

where ϕ is the angle of rotation in the plane, $\mathbf{b}$ is the subsequent displacement vector with $b = |\mathbf{b}|$ and $\beta = \arg(\mathbf{b})$, X^2 (where $X > 0$) is the value of the Casimir operator specifying the irreducible representation, $\lambda = 0(\frac{1}{2})$ for single- (double-) valued representations, m and n are integers running from $-\infty$ to $+\infty$, and J_n is the Bessel function of order n.

massless particles with given helicity $\mathcal{H}$ by means of the spinor–tensors $G^{(\mathcal{H}-\frac{1}{2})}_{\sigma_1\cdots\sigma_{\mathcal{H}-\frac{1}{2}}}$ and

$$G^{(r-\mathcal{H}-1,r)}_{\sigma_1\cdots\sigma_{r-\mathcal{H}-1}[\mu_{r-\mathcal{H}}{}^{\nu}{}_{r-\mathcal{H}}\cdots[\mu_{r-\frac{1}{2}}\nu_{r-\frac{1}{2}}]}$$

where $r = \mathcal{H}+1, \mathcal{H}+2, \cdots$.

Comparing these three statements with the results obtained in Ref. 1, one can say that there exists a similarity between theories of the particles with integer spin and half-integer spin. Such similarity is clear on the grounds of the representation theory of the Lorentz group and the wavefunction formalisms for higher-spin particles.

Our investigations in Ref. 1 and in this paper are restricted only to the classical field equations. The limit $m = 0$ in quantum theory is much more complicated because of the problem of spurious divergences, coming from the factors in the Green's functions m^{-2k}. These problems will be studied in subsequent work.

ACKNOWLEDGMENT

I am grateful to Dr. J. Lukierski for his interest and stimulating conversations.

[1] T. Cukierda and J. Lukierski, J. Math. Phys. **11**, 46 (1970).
[2] M. Fierz, Helv. Phys. Acta **12**, 1 (1939).
[3] W. I. Ogievetskiĭ and I. W. Polubarinov, Yad. Fiz. 216 (1966) [Sov. J. Nucl. Phys. **4**, 156 (1967)].

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Wigner Coefficients for the Group $E(2)$

H. P. W. GOTTLIEB
Department of Applied Mathematics and Theoretical Physics
University of Cambridge, England

(Received 29 December 1969; Revised Manuscript Received 29 April 1970)

The Wigner coefficients for the faithful unitary representations of the 2-dimensional Euclidean group are derived from two identities involving Bessel functions. Since the multiplicity in the decomposition of a direct product is two, we find two sets of coefficients which are real and mutually orthogonal and symmetric and antisymmetric, respectively, under interchange of constituent representations. Properties of the coefficients at the ends of the decomposition spectrum are discussed.

I. INTRODUCTION

There is increasing interest in the role which the 2-dimensional Euclidean group $E(2)$, the group of rotations and translations in the plane, plays in elementary particle physics. As the isomorph of the little group for lightlike particles in the Wigner classification of irreducible representations of the Poincaré group,[1] its faithful representations, with infinite spin spectrum[2] ("indefinite helicity"), do not appear to be realized as free particles in nature. However, in the formalism of Toller,[3] the amplitude for the case of zero momentum transfer square (as distinct from null momentum transfer) should be expansible in terms of representation functions of $E(2)$. A study of the $E(2)$ representations seems essential for an appreciation of the transition from timelike to spacelike momentum transfer. The group $E(2)$ also occurs naturally in the study of the infinite-momentum frame limit of kinematics.[4]

In order to gain a deeper understanding of the nature of the representations of $E(2)$, we here derive the Wigner or generalized Clebsch–Gordan coefficients. These may be deduced from some well-known identities for Bessel functions, the main difficulty being caused by the multiplicity of irreducible representations in the direct product. Two sets of coefficients which are orthogonal and can be chosen to be real are required.

II. REPRESENTATIONS OF $E(2)$ AND BESSEL FUNCTIONS

The faithful unitary irreducible representations of $E(2)$ may be realized over a discrete basis and written in the form[2]

$$D^{X\lambda}_{m,n}(\mathbf{b}, \phi) = e^{-i(n+\lambda)\phi}e^{i(n-m)\beta}(i)^{n-m}J_{m-n}(bX), \quad (1)$$

where ϕ is the angle of rotation in the plane, $\mathbf{b}$ is the subsequent displacement vector with $b = |\mathbf{b}|$ and $\beta = \arg(\mathbf{b})$, X^2 (where $X > 0$) is the value of the Casimir operator specifying the irreducible representation, $\lambda = 0(\frac{1}{2})$ for single- (double-) valued representations, m and n are integers running from $-\infty$ to $+\infty$, and J_n is the Bessel function of order n.

The Clebsch–Gordan series for $E(2)$ has been obtained by Mackey[5]: In the decomposition of the direct product of two representations specified by X' and X'', each representation in the range $|X' - X''|$, $(X' + X'')$ appears *twice*. An extra parameter $\omega = 1, 2$ is therefore introduced into the definition of the Wigner coefficients to distinguish between functions transforming under similar representations. If the basis functions of representations are realized over some space $H = \{h\}$, the defining relations for the Wigner coefficients in terms of basis functions may be written as

$$\Psi^{X'\lambda'}_{n'}(h')\Psi^{X''\lambda''}_{n''}(h'') = \sum_{\omega}\sum_{n}\sum_{\lambda}\$_{X} c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''}\Psi^{\omega X\lambda;X'\lambda'X''\lambda''}_{n}(h', h''), \quad (2)$$

where $\Psi^{X'\lambda'}_{n'}$ and $\Psi^{\omega X\lambda;X'\lambda'X''\lambda''}_{n}$ are, respectively, the single-particle and 2-particle irreducible basis functions; states within a given basis are completely specified by the discrete parameter n.

In terms of the representations themselves, the standard definition is

$$D^{X'\lambda'}_{m',n'}(g)D^{X''\lambda''}_{m'',n''}(g) = \sum_{m}\sum_{n}\sum_{\omega}\sum_{\lambda}\$_{X} c^{\omega X\lambda X'\lambda' X''\lambda''}_{m\ m'\ m''} * D^{X\lambda}_{m,n}(g)c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''}, \quad (3)$$

where g is a group element.

In these expressions, the integral over representation space, i.e., the Plancherel measure for the group $E(2)$, is given by

$$\$_{X} = \int X\,dX = \int dW, \qquad W = \tfrac{1}{2}X^2,$$

where W is actually the variable[6] for which orthogonality holds in the conventional form, because of the following integral relation for Bessel functions of a given integer order:

$$\int_0^\infty bJ_n(bX)J_n(bX')\,db = X^{-1}\delta(X - X') = \delta(W - W'), \quad (4)$$

where $W' = \frac{1}{2}(X')^2$.

The orthogonality and completeness relations for the Wigner coefficients are

$$\sum_{n}\sum_{\omega}\sum_{\lambda}\$_{X} c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''}c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ \bar{n}'\ \bar{n}''} * = \delta_{n',\bar{n}'}\delta_{n'',\bar{n}''} \quad (5)$$

and

$$\sum_{n'}\sum_{n''} c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''}c^{\bar{\omega}\bar{X}\bar{\lambda}X'\lambda' X''\lambda''}_{\bar{n}\ n'\ n''} * = \delta_{\omega,\bar{\omega}}\delta_{n,\bar{n}}\delta_{\lambda,\bar{\lambda}}\delta(W - \overline{W}) \quad (6)$$

if $(X\lambda)$ appears in the decomposition of the product of the representations $(X'\lambda')$ and $(X''\lambda'')$.

These imply the converse relation to (3), viz.,

$$\sum_{m'}\sum_{m''}\sum_{n'}\sum_{n''} c^{\bar{\omega}\bar{X}\bar{\lambda}X'\lambda' X''\lambda''}_{\bar{m}\ m'\ m''}c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''} * \times D^{X'\lambda'}_{m',n'}(g)D^{X''\lambda''}_{m'',n''}(g) = D^{X\lambda}_{\bar{m},n}(g)\delta_{\bar{\omega},\omega}\delta_{\bar{\lambda},\lambda}\delta(\overline{W} - W) \quad (7)$$

if $(X\lambda)$ appears in the product of $(X'\lambda')$ and $(X''\lambda'')$, and also the converse relation to (2), which may be taken as defining the 2-particle irreducible basis functions

$$\Psi^{\omega X\lambda;X'\lambda'X''\lambda''}_{n}(h', h'') = \sum_{n'}\sum_{n''} c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''} * \Psi^{X'\lambda'}_{n'}(h')\Psi^{X''\lambda''}_{n''}(h''). \quad (8)$$

The identification of the Wigner coefficients of $E(2)$ is facilitated by two formulas involving Bessel functions. A formula of Dixon and Ferrar [Ref. 7, p. 205, Eq. (6.41)], for the case of integer-order Bessel functions, may be written in the form

$$J_{n'}(X')J_{n''}(X'') = \int_0^\pi \frac{d\alpha}{\pi}(-1)^{n'}\exp i[n'\alpha + (n' + n'')\gamma]J_{n'+n''}(X), \quad (9)$$

where

$$X = +[(X')^2 + (X'')^2 - 2X'X''\cos\alpha]^{\frac{1}{2}} \quad (10)$$

and γ is defined as the angle opposite X' in the triangle formed by X, X', and X'' (and so α is the angle between X' and X''). This holds for all integers n' and n'' and all positive X' and X''. The case $X' = X''$ is Neumann's formula [Ref. 7, p. 205, Eq. (7.11)].

The Graf addition formula for Bessel functions of integer order, given by Watson,[8] may be written in the form[9]

$$J_n(X) = \sum_{n'=-\infty}^{\infty}(-1)^{n'}\exp i(n'\alpha + n\gamma)J_{n'}(X')J_{n-n'}(X''), \quad (11)$$

with X again given by (10).

The geometrical meaning of the symbols is made clear by Fig. 1 (cf. Watson[8]). Note that

$$(X'' - X'e^{-i\alpha})/X = e^{i\gamma}$$

and

$$X\,dX = X'X''\sin\alpha\,d\alpha = 2A\,d\alpha,$$

where the area of the triangle formed by X, X', and X'' is

$$A = \tfrac{1}{2}X'X''\sin\alpha.$$

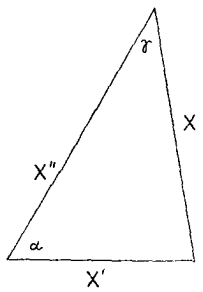


FIG. 1. Geometrical significance of the symbols used in the text.

III. DETERMINATION OF THE WIGNER COEFFICIENTS

We shall utilize the fact that a multiple of the elements of a fixed row of a representation forms (under right multiplication by a group element) a basis for that representation. Equation (1) will therefore be used to define basis functions.

For single-particle basis functions, we choose a row $m = M$, where M is an arbitrary integer which remains fixed for the rest of the discussion. Thus we define[10]

$$\begin{aligned}\Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) &= (8\pi^2)^{-\frac{1}{2}} D^{X'\lambda'}_{M,n'}(\mathbf{b}, \phi) \\ &= (8\pi^2)^{-\frac{1}{2}} e^{-i(n'+\lambda')\phi} e^{i(n'-M)\beta} (i)^{(M-n')} J_{n'-M}(bX'),\end{aligned} \tag{12}$$

where X' specifies the representation under which the function transforms irreducibly, λ' specifies the valuedness, and $n' + \lambda'$ is the helicity component which labels the states within the basis. [Compare this with the case of the rotation group,[11] where the basis functions $Y_{ln}(\theta, \phi)$ may be obtained from the representation functions $D^{l}_{m,n}(\alpha, \theta, \phi)$ by choosing the row $m = 0$ and introducing an appropriate normalization factor.] These basis functions are normalized over the space of ϕ, β, and b, with ϕ going from 0 to 4π to give orthogonality when the case of double-valued representations is included. Equation (4) is used to obtain

$$\int_0^{4\pi} d\phi \int_0^{2\pi} d\beta \int_0^{\infty} b\, db \Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) \Psi^{X''\lambda''}_{n''}(\phi, \mathbf{b})^* = \delta_{n',n''}\delta_{\lambda',\lambda''}\delta(W' - W''). \tag{13}$$

The "2-particle" basis functions defined by Eq. (8) and depending on two independent sets of variables will correspondingly contain a normalization factor $(8\pi^2)^{-1}$.

Now, using the real part of (9), we find

$$\begin{aligned}&\Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) \Psi^{X''\lambda''}_{n''}(\phi, \mathbf{b}) \\ &\quad = \int_{|X'-X''|}^{(X'+X'')} \frac{X\, dX}{2\pi A} (-1)^{(n'-M)} \\ &\qquad \times \cos[(n' - M)\alpha + (n' + n'' - 2M)\gamma] \\ &\qquad \times (8\pi^2)^{-1} e^{-i(n'+n''+\lambda'+\lambda'')\phi} e^{i(n'+n''-2M)\beta} \\ &\qquad \times (i)^{(2M-n'-n'')} J_{n'+n''-2M}(bX).\end{aligned} \tag{14}$$

Apparently, what we are seeing here is part of the Clebsch–Gordan decomposition corresponding to Eq. (2) for the particular case $h' = h'' = (\phi, \mathbf{b})$. The Wigner coefficients depend only on representation and basis labels and are, of course, independent of the particular variables over which the basis functions are realized. All values of X in the known decomposition range appear in (14). Moreover, since we know that the decomposition has multiplicity two, we deduce from the above that the coefficient c^1, say, and the function $\Psi^1(h', h') = \Psi^1((\phi, \mathbf{b}), (\phi, \mathbf{b}))$ appear in (14) while the function $\Psi^2(h', h') = \Psi^2((\phi, \mathbf{b}), (\phi, \mathbf{b})) = 0$ and so does not appear in the particular case $h' = h''$. Since, in general, $\Psi^2(h', h'')$ is defined over a 6-dimensional space, its vanishing over a 3-dimensional subspace $h' = h''$ does not contradict normalization. We must therefore be able to write (14) in the form of Eq. (2), i.e., as

$$\begin{aligned}&\Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) \Psi^{X''\lambda''}_{n''}(\phi, \mathbf{b}) \\ &\quad = \sum_{\omega} \sum_{n} \sum_{\lambda} \int X\, dX c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''} \\ &\qquad \times \Psi^{\omega X\lambda; X'\lambda' X''\lambda''}_{n}((\phi, \mathbf{b}), (\phi, \mathbf{b})).\end{aligned} \tag{15}$$

This method of procedure will subsequently be justified.

The real part of Eq. (11) may be used to write

$$\begin{aligned}&(8\pi^2)^{-1} e^{-i(N+\lambda'+\lambda'')\phi} e^{i(N-2M)\beta} (i)^{(2M-N)} J_{N-2M}(bX) \\ &\quad = \sum_{n'} (-1)^{(n'-M)} \cos[(n' - M)\alpha + (N - 2M)\gamma] \\ &\qquad \times \Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) \Psi^{X''\lambda''}_{N-n'}(\phi, \mathbf{b}).\end{aligned} \tag{16}$$

This is evidently the converse relation to (14), corresponding to (8) for the case $\omega = 1$, and we must therefore be able to write it in the same form as (8), i.e., as

$$\begin{aligned}&\Psi^{1X\lambda; X'\lambda' X''\lambda''}_{n}((\phi, \mathbf{b}), (\phi, \mathbf{b})) \\ &\quad = \sum_{n'} \sum_{n''} c^{1X\lambda X'\lambda' X''\lambda''}_{n\ n'\ n''} * \Psi^{X'\lambda'}_{n'}(\phi, \mathbf{b}) \Psi^{X''\lambda''}_{n''}(\phi, \mathbf{b}).\end{aligned} \tag{17}$$

Since the representation (1) is diagonal in helicity (i.e., with $\mathbf{b} = \mathbf{0}$), the Wigner coefficients must contain a factor $\delta_{n+\lambda, n'+\lambda'+n''+\lambda''}$, and the necessary valuedness of the product of two representations

implies a further factor $\delta_{\lambda,\lambda'+\lambda''//1}$, where $\lambda' + \lambda''//1 = (\lambda' + \lambda'')(\text{mod } 1)$. Finally, the Clebsch-Gordan series implies a coefficient factor $\Delta(X, X', X'') = 1$ or 0, according to whether or not it is possible to form a triangle from the three quantities X, X', and X''.

Comparison of (17) with (16) and of (15) with (14) then yields

$$\Psi^{2}_{n}{}^{X\lambda;X'\lambda'X''\lambda''}((\phi, \mathbf{b}), (\phi, \mathbf{b})) = 0, \tag{18}$$

$$\begin{aligned}\Psi^{1}_{n}{}^{X\lambda;X'\lambda'X''\lambda''}&((\phi, \mathbf{b}), (\phi, \mathbf{b}))\\ &= (2\pi A)^{-\frac{1}{2}}(8\pi^2)^{-1}e^{-i(n+\lambda)\phi}e^{i(n-2M+\lambda-\lambda'-\lambda'')\beta}\\ &\quad\times (i)^{(2M-n-\lambda+\lambda'+\lambda'')}J_{n-2M+\lambda-\lambda'-\lambda''}(bX)\\ &= (2\pi A)^{-\frac{1}{2}}(8\pi^2)^{-1}D^{X\lambda}_{2M-\lambda+\lambda'+\lambda'',n}(\mathbf{b}, \phi),\end{aligned} \tag{19}$$

where

$$\begin{aligned}\lambda' + \lambda'' - \lambda = 0 \quad &\text{if } (\lambda', \lambda'') = (0, 0), (\tfrac{1}{2}, 0), (0, \tfrac{1}{2}),\\ = +1 \quad &\text{if } (\lambda', \lambda'') = (\tfrac{1}{2}, \tfrac{1}{2}),\end{aligned}$$

and

$$\begin{aligned}c^{1}_{n\ n'\ n''}{}^{X\lambda X'\lambda' X''\lambda''}&\\ &= (-1)^{n'-M}(2\pi A)^{-\frac{1}{2}}\\ &\quad\times \cos[(n' - M)\alpha + (n' + n'' - 2M)\gamma]\\ &\quad\times \delta_{n+\lambda,n'+\lambda'+n''+\lambda''}\delta_{\lambda,\lambda'+\lambda''//1}\Delta(X, X', X''),\end{aligned} \tag{20}$$

where by convention we have adopted the positive square root in the factor $(2\pi A)^{-\frac{1}{2}}$.

The identification of Ψ^1 with one set of standard 2-particle basis functions corresponds to the choice of real Wigner coefficients c^1. Furthermore, the c^1 are symmetric under interchange of $(X'\lambda'n')$ and $(X''\lambda''n'')$ (under which γ by definition changes value), so we see that by (8) the Ψ^1 must be symmetric under interchange of $(X'\lambda', h')$ and $(X''\lambda'', h'')$, as is indeed the case for (19) in which $h' = h''$ $[= (\phi, \mathbf{b})]$. The known multiplicity of the decomposition and the orthogonality and completeness relations (5) and (6) imply the existence of another set of coefficients which, with choice of positive over-all sign, are evidently

$$\begin{aligned}c^{2}_{n\ n'\ n''}{}^{X\lambda X'\lambda' X''\lambda''}&\\ &= (-1)^{n'-M}(2\pi A)^{-\frac{1}{2}}\\ &\quad\times \sin[(n' - M)\alpha + (n' + n'' - 2M)\gamma]\\ &\quad\times \delta_{n+\lambda,n'+\lambda'+n''+\lambda''}\delta_{\lambda,\lambda'+\lambda''//1}\Delta(X, X', X'').\end{aligned} \tag{21}$$

Together with the c^1 of (20), these do satisfy (5) and (6). The coefficients c^2 are antisymmetric under interchange of $(X'\lambda'n')$ and $(X''\lambda''n'')$, and thus correspond to 2-particle basis functions $\Psi^2(h', h'')$ which are antisymmetric under interchange of constituent representations and, thus not inconsistently, vanish when $h' = h''$ $[= (\phi, \mathbf{b})]$ as given by (18). The final verification that the coefficients c^1 and c^2 are given by (20) and (21) is that, by using the real part of (9), we may prove[12] that these coefficients do satisfy (3) for the representations (1).

The 2-particle irreducible basis functions $\Psi^\omega(h', h'')$ are now, in general, given explicitly by (8), together with (20), (21), and (12), although they only have simple forms in the case $h' = h''$. Thus, use of the real part of the identity (11) yields (16) and, hence, (19) as the case of (8) when $h' = h''$ $[= (\phi, \mathbf{b})]$ and $\omega = 1$, and use of the imaginary part of (11) yields (18) when $h' = h''$ and $\omega = 2$. The symmetry or antisymmetry properties of the Ψ^ω are explicitly confirmed by the definition (8) and the properties of the coefficients c^ω. This is in accord with the customary preference for constructing symmetric and antisymmetric functions and coefficients for a group whose representation decomposition series has multiplicity two.

Because of the completeness of the coefficients (20) and (21), given by (6), and the orthonormality of the single-particle basis functions (12) given by (13), the functions Ψ^ω given by (8) are orthonormal according to

$$\begin{aligned}\int_0^{4\pi} d\phi' \int_0^{4\pi} d\phi'' \int_0^{2\pi} d\beta' \int_0^{2\pi} d\beta'' \int_0^{\infty} b'\, db' \int_0^{\infty} b''\, db''&\\ \times \Psi^{\omega}_{n}{}^{X\lambda;X'\lambda'X''\lambda''}((\phi', \mathbf{b}'), (\phi'', \mathbf{b}''))&\\ \times \Psi^{\bar{\omega}}_{\bar{n}}{}^{\bar{X}\bar{\lambda};\bar{X}'\bar{\lambda}'\bar{X}''\bar{\lambda}''}((\phi', \mathbf{b}'), (\phi'', \mathbf{b}''))^*&\\ = \delta_{\omega,\bar{\omega}}\, \delta_{n,\bar{n}}\delta_{\lambda,\bar{\lambda}}\delta(W - \bar{W})&\\ \times \delta_{\lambda',\bar{\lambda}'}\delta_{\lambda'',\bar{\lambda}''}\delta(W' - \bar{W}')\delta(W'' - \bar{W}'')&\end{aligned} \tag{22}$$

when $(X\lambda)$ appears in the decomposition of the product of the representations $(X'\lambda')$ and $(X''\lambda'')$.

IV. COMMENTS ON THE METHOD

It might be thought that a similar method could be used to derive the Clebsch-Gordan coefficients for the rotation group $O(3)$. However, the essence of our derivation above is the symmetry in b and X of the Bessel-function part of the representation (1) for which the addition formula (11), with argument (bX), is relevant both when X is fixed and b takes two values (as used in an explicit demonstration of the representation property) and also when b is fixed and X takes two values (as used above in the derivation of the Wigner coefficients). In the case of the rotation group, the addition theorem for spherical harmonics which is used to demonstrate the representation property cannot be used for a fixed value of the group (angle) parameter and two different values of the Casimir (spin) operator.

It would be possible in principle to proceed from a comparison of (9) with (3) and of (11) with (7),

without mentioning the explicit form of single-particle basis functions. The imposition of symmetry–antisymmetry and reality conditions and choice of over-all signs still do not, however, determine the coefficients uniquely, as we have seen; M can take any (fixed) integer value. Although this nonuniqueness might seem to be a consequence of the multiplicity two,[12] it is, in fact, simply a result of the form of the dependence of the representations (1) and the coefficients (20) and (21) on the helicity indices, which assume an infinite range:

$$D^{X\lambda}_{m+\mu,n+\mu}(\mathbf{b}, \phi) = e^{-i\mu\phi} D^{X\lambda}_{m,n}(\mathbf{b}, \phi), \tag{23}$$

$$[c^{\omega X\lambda\ X'\lambda'\ X''\lambda''}_{n+2\mu,n'+\mu,n''+\mu}]_{M=\mu} = [c^{\omega X\lambda X'\lambda' X''\lambda''}_{n\ \ n'\ \ n''}]_{M=0}, \tag{24}$$

$$m, n, n', n'', \mu = -\infty, \cdots, +\infty \text{ (integers)}.$$

Use of only (23) and (24) shows that, if (3) is satisfied by c^ω with $M = 0$, it is satisfied by c^ω with $M = \mu$, for any fixed integer μ. Proceeding from Eqs. (3) and (7), which involve the representations themselves and have a quadratic dependence on the coefficients, would not give any particular reason for a standard choice of coefficients, and the existence of an infinite family of sets of coefficients (corresponding to different integer M) all having the symmetry–antisymmetry properties might escape notice. Our method used above, which proceeds via basis functions obtained from a row of the representations, immediately yields that family and suggests the possibility of a relation of the type (24).

In general, a 2-particle basis function defined by (8) for a set of coefficients satisfying (3), (5), and (6) and for any set of single-particle basis functions which transform under the representations appearing in (3) has the correct irreducible transformation property. There is, in principle, no need, therefore, to use the same value of M in the single-particle basis functions (12) as in the coefficients (20) and (21). However, the sum (8) will not, in general, be simplifiable. Our method leads in a natural way to a convention for choosing coefficients and basis functions; the choice of the same value of M in the $\Psi^{X'\lambda'}_{n'}$ and the c^ω corresponds to a choice which gives the simple formula (19), i.e., to a formula for the 2-particle basis functions $\Psi^\omega(h', h'')$ which, for the case $h' = h''$ in particular, is the simplest.

V. PROPERTIES OF THE COEFFICIENTS

The coefficients given by (20) and (21), for fixed M, satisfy Eq. (3) defining the Wigner coefficients for the representations (1). In addition, all the orthogonality and completeness relations (5) and (6), expressing the unitarity of the transformations effected by the coefficients, are satisfied, and hence (7) also is satisfied.

We have therefore obtained the Wigner coefficients (20) and (21) for the faithful unitary irreducible representations of $E(2)$ given by (1). Single-particle basis functions may be realized by (12) and normalized according to (13), and the 2-particle irreducible basis functions defined by (8) [of which (18) and (19) are then particular cases] are then normalized according to (22). The coefficients have been chosen to be real, and the two sets of coefficients (and 2-particle basis functions) appropriate to the known multiplicity two of the direct product decomposition correspond to symmetry and antisymmetry under interchange of constituent representations.

Some symmetry relations are

$$c^{\frac{1}{2}X\lambda X''\lambda'' X'\lambda'}_{\ n\ \ n''\ \ n'} = \pm c^{\frac{1}{2}X\lambda X'\lambda' X''\lambda''}_{\ n\ \ n'\ \ n''}, \tag{25}$$

$$c^{\frac{1}{2}\ X\ \ \lambda X'\lambda' X''\lambda''}_{(-n)\ \ n'\ \ n''} = \pm c^{\frac{1}{2}X\lambda}_{n+4M+2(\lambda'+\lambda''-\lambda)\ 2M-n'\ 2M-n''}, \tag{26}$$

$$c^{\frac{1}{2}X''\lambda'' X\lambda X'\lambda'}_{\ n''\ \ n\ \ n'} = \pm(-1)^{n'-M} c^{\frac{1}{2}X\lambda\ \ X'\lambda'\ \ X''\lambda''}_{(n+M+\lambda'+\lambda''-\lambda)\ (2M-n')\ (n''-M+\lambda''-\lambda-\lambda')}, \tag{27}$$

$$\Psi^{\frac{1}{2}X\lambda;X'\lambda' X''\lambda''}_{n}((\phi', \mathbf{b}'), (\phi'', \mathbf{b}'')) = \pm\Psi^{\frac{1}{2}X\lambda;X''\lambda'' X'\lambda'}_{n}((\phi'', \mathbf{b}''), (\phi', \mathbf{b}')). \tag{28}$$

A recurrence relation is

$$c^{\omega X\lambda\ \ X'\lambda'\ \ X''\lambda''}_{(n+2)\ (n'+1)\ (n''+1)} = \sum_{\bar{\omega}=1}^{2} R_{\omega,\bar{\omega}}(\alpha + 2\gamma + \pi) c^{\bar{\omega}X\lambda X'\lambda' X''\lambda''}_{\ n\ \ n'\ \ n''}, \tag{29}$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Finally, we note that coefficients for different values of M are obtained from each other by a rotation, in the 2-dimensional subspace labeled by ω, which preserves the symmetry–antisymmetry properties. For instance,[13]

$$[c^{\omega X\lambda X'\lambda' X''\lambda''}_{\ n\ \ n'\ \ n''}]_{M=\mu} = \sum_{\bar{\omega}} R_{\omega,\bar{\omega}}(-\mu(\alpha + 2\gamma + \pi))[c^{\bar{\omega}X\lambda X'\lambda' X''\lambda''}_{\ n\ \ n'\ \ n''}]_{M=0}. \tag{30}$$

For a standard set of coefficients we may choose $M = 0$; this results in the simplest behavior under cyclic reordering of representation labels (27).

VI. DISCUSSION

The law for the decomposition of a direct product shows that the Casimir operators combine somewhat like the spins of the $O(3)$ case, with values of X

appearing between the difference and sum of the original two values. However, unlike the $O(3)$ spin case where a discrete spectrum occurs, for $E(2)$ a continuum of representations appears, with multiplicity two. In general, every component (allowed by angular momentum conservation) of representations specified by every value of the Casimir operator occurring in the decomposition will eventually appear with nonzero coefficient, although it may appear as only one of the symmetric or antisymmetric types. If, in a particular instance, the trigonometric function in a coefficient of one type vanishes, then the trigonometric function in the corresponding coefficient of the other type will be unity.

The most interesting behavior of the coefficients occurs at the two ends of the decomposition spectrum, where the limiting cases may be analyzed using the geometry of Fig. 1 and also the fact that

$$A = \tfrac{1}{4}(X + X' + X'')^{\frac{1}{2}}(X' + X'' - X)^{\frac{1}{2}} \times [X - (X' - X'')]^{\frac{1}{2}}[X - (X'' - X')]^{\frac{1}{2}}.$$

As X tends to $X' + X''$, the (moduli of the) symmetric-type coefficients c^1 tend to infinity like an inverse fourth root, while the antisymmetric-type coefficients c^2 tend to zero. If $X' \neq X''$, the same holds as $X \to |X' - X''|$. Thus, if $X' \neq X''$, the "stretch" cases are infinitely more likely to occur than intermediate values which are always finite, and this suggests that a first approximation to a decomposition of a direct product might consider that only these two cases occur and are of the symmetric type.

When $X' = X''$ and $n' + n''$ is even, as $X \to |X' - X''| = 0$ the coefficients c^1 tend to infinity like an inverse square root, i.e., faster than for any of the previous cases, and the coefficients c^2 tend to zero; but if $n' + n''$ is odd, it is the antisymmetric-type coefficients c^2 which tend to infinity like an inverse square root while the coefficients c^1 tend to zero (since $\gamma \to \frac{1}{2}\pi$). This suggests that, in the product of two similar representations, the representations most likely to occur may be regarded in the limit as having $X = 0$, leading to two infinitely reducible representations consisting of 1-dimensional irreducible, but unfaithful, representations of $E(2)$ which are just the standard (faithful, irreducible) helicity representations of $O(2)$. There are then effectively two degenerate infinite towers of helicity representations of $O(2)$. The towers differ in their symmetry properties and each contains helicity values differing by two; together they yield every (integer or half-integer) helicity just once. If the constituent representations are both single valued, then $n' + n'' = n$, and the even-helicity components have the symmetric property under interchange of constituent representations, while the odd-helicity components are of the antisymmetric type. The same may be said when the resultant representations are double valued, provided that $\frac{1}{2}$ is subtracted from the helicity values. If both the constituent representations are double valued, then it is the odd-helicity components which are of the symmetric type and the even-helicity components which are of the antisymmetric type.

ACKNOWLEDGMENTS

I should like to thank Dr. A. J Macfarlane, P. Goddard, and A. White for useful discussions. I am grateful to the Commissioners for the Exhibition of 1851 for the award of an Overseas Scholarship.

[1] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[2] E. P. Wigner, "The Application of Group Theory to the Special Functions of Mathematical Physics," lecture notes, Princeton (1955).

[3] M. Toller, Nuovo Cimento **54A**, 295 (1968).

[4] H. Bacry and N. P. Chang, Ann. Phys. (N.Y.) **47**, 407 (1968).

[5] G. W. Mackey, Ann. Math. **55**, 101 (1952).

[6] W is $(-\frac{1}{2})$ times the value of the standard Casimir operator for an irreducible representation of the Poincaré group corresponding to zero mass and infinite spin spectrum.

[7] A. L. Dixon and W. L. Ferrar, Quart. J. Math. (Oxford) **4**, 193 (1933).

[8] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., London, 1966), 2nd ed., pp. 359–61.

[9] For real X and integer n, $J_n(X)$ is real, and

$$J_{-n}(X) = (-1)^n J_n(X), \quad J_n(0) = \delta_{n,0}.$$

[10] These *functions*, with states (for a given representation) specified by n, are realized over a space (ϕ, **b**) and should not be compared with, for instance, the case of representations of the Poincaré group, whose *kets* are $|n, p\rangle$, where both n and p, with $p^2 = (\text{mass})^2$, label the states within the irreducible representation; **b** is not equivalent to p and bears no relation to the Casimir operator of the group, nor does it label states.

[11] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).

[12] As well as being useful in Eq. (2), it is convenient that the identity (9) can also be used in (3); this is a result of the necessity of including *two* orthogonal sets of coefficients c^1, c^2, and the identity

$$\cos(x - z)\cos(y - z) + \sin(x - z)\sin(y - z) = \cos(x - y).$$

[13] $(\alpha + 2\gamma)$ simply changes sign (mod 2π) under interchange of X' and X''.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Statistical Group Theory and the Distribution of Angular Momentum States*

J. G. CLEARY AND B. G. WYBOURNE
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 29 April 1970)

The problem of developing relations for the statistical distribution of the angular momentum states of an electron configuration l^N, where l and N are large, has been considered. If $D(L)$ is the number of times the orbital angular momentum L occurs, then, using the theory of partitions and groups, we find that the numbers $D(L)$ are approximately distributed with respect to L according to the Wigner-type form

$$D(L) = A(L + \tfrac{1}{2}) \exp [-(L + \tfrac{1}{2})^2/2\sigma^2].$$

A number of examples are examined.

I. INTRODUCTION

In recent years there has been considerable interest in the statistical description of the average properties of atomic and nuclear spectra, especially of nuclear-level densities and the distribution of energy-level spacings.[1] Bloch[2] has discussed the distribution of angular momenta among the nuclear levels and established that the number N_J of levels with angular momentum J and energy up to a given value is

$$N_J \sim (2J + 1) \exp [-(J + \tfrac{1}{2})^2/2\sigma^2], \qquad (1)$$

which is the form of the usual Wigner distribution.[3] Bloch's result was developed from an examination of nuclear-level density models.

In this paper we wish to consider the determination of the distribution of the number of states $D(L)$ as a function of the total orbital angular momentum L associated with the states of maximal spin of a configuration l^N of N electrons occupying equivalent orbitals of angular momenta $x = l$, where l is an integer. For $x = j$, a half-integer, we obtain the jj coupling analog of this problem.

Bloch's derivation of Eq. (1) involved a detailed study of the nuclear-level density problem together with entropy considerations. In our problem the Pauli exclusion principle assumes a fundamental importance because we wish to restrict our attention to the occupation of a single electron shell, and thus Bloch's result cannot be simply translated to the solution of our problem. Rather, we must choose a different line of attack.

Our work originally arose from the empirical observation that, if the irreducible representation Γ^λ_G of a Lie group G, where λ is the highest weight of the representation, is decomposed into the irreducible representations $\mathfrak{D}^J$ of the subgroup R_3 such that

$$\Gamma^\lambda_G = \sum_J g_{\lambda J} \mathfrak{D}^J, \qquad (2)$$

where $g_{\lambda J}$ is the number of times $\mathfrak{D}^J$ appears in the decomposition, then the numbers $g_{\lambda J}$ tend, under certain conditions, to be distributed with respect to the maximal weight J of $\mathfrak{D}^J$ according to a distribution analogous to that given in Eq. (1).

The problem of decomposing irreducible representations of a higher Lie group into those of the 3-dimensional group R_3 is, of course, common in the enumeration of the angular momentum states in the LS and jj coupling of identical fermions in atomic spectroscopy.

The above observations suggest that the discussion of the statistical properties of atomic and nuclear spectra might in some cases be developed from the vantage point of group theory. Here, we propose to develop some of the relevant mathematical tools and, in particular, to establish the Wigner form as an approximation relevant to the problem at hand with the hope of stimulating the further development of a statistical group theory.

In Sec. II we recast the problem of the enumeration of the angular momentum states of a configuration of identical fermions in terms of the plethysm of S functions. This then allows us, in Sec. III, to obtain a connection between the problem of enumerating angular momentum states and that of enumerating partitions in ordinary number theory. Having demonstrated the relevance of the theory of partitions to the problem at hand, we then investigate in Sec. IV methods of obtaining polynomial coefficients that give an approximate representation of the relevant partitions. This development leads to the need to consider three basic types of partitions, Q, P, and R, which correspond to differently classified partitions. These are defined in Sec. IV, and relationships between them investigated briefly in Sec. V. The Wigner form as an approximate solution to our problem is developed in Sec. VI using the previous results. A brief discussion of the extension of the

problem to states of nonmaximal multiplicity is given in Sec. VII, which establishes the Wigner form once again. A detailed example of the application of our methods to the states of the configuration i^5, $l = 6$, is given in Sec. VIII, while in Sec. IX we consider the estimation of the quantities A and σ associated with the Wigner form. In Sec. X we apply these results to the case of the configuration t^{13}, $l = 14$.

II. ENUMERATION OF ANGULAR MOMENTUM STATES

The group-theoretical classification of the angular momentum states of identical fermions is well known[4] in atomic and nuclear spectroscopy, and here we simply recast the familiar theory in terms of the plethysm of S functions.[5] In LS coupling the orbital angular-momentum states of the N-particle configuration l^N are classified using the chain of groups

$$U_{2l+1} \rightarrow R_{2l+1} \rightarrow R_3, \tag{3}$$

while in the jj-coupled configuration j^N the chain of groups

$$U_{2j+1} \rightarrow Sp_{2j+1} \rightarrow R_3 \tag{4}$$

is appropriate.

If under the restriction $G \rightarrow R_3$ the character $[\![1]\!]$ of the vector representation $\Gamma^{[1]}$ decomposes as

$$[\![1]\!] \rightarrow \sum_J g_J[J], \tag{5}$$

then the decomposition of the character $[\![\lambda]\!]$ of the representation $\Gamma^{[\lambda]}$ of G will contain all the terms that arise in the plethysm[6,7]

$$\left(\sum_J g_J[J]\right) \otimes [\![\lambda]\!] = \sum g_{\lambda J'}[J'], \tag{6}$$

where $[J]$ is the character of the representation D^J of R_3. Methods for evaluating the plethysms of Eq. (7) in terms of S functions have been discussed elsewhere.[8–10]

In the case of electrons, the orbital states $[L]$ of spin S and of the configuration l^N are just those that arise in the decomposition under $U_{2l+1} \rightarrow R_3$ of the unitary group character $\{2^{N-2S}1^{2S}\}$, which is equivalent to the similarly designated S function constructed on the characteristic roots of the unitary matrices of rank $2l + 1$. This S function may be written in terms of the elementary symmetry functions a_r to give

$$\{2^{N-2S}1^{2S}\} = \{1^N\}\{1^{N-2S}\} - \{1^{N+1}\}\{1^{N-2S-1}\}. \tag{7}$$

The orbital states associated with maximum S will be just those arising in the reduction of the unitary character $\{1^N\}$.

If under the group restriction

$$U_{2l+1} \rightarrow R_3 \tag{8}$$

we have $\{1\} \rightarrow [l]$, then it follows from Eqs. (6) and (7) that the orbital states $[L]$ associated with spin S in the l^N configurations are just those terms arising in the plethysm

$$\begin{aligned}
&[l] \otimes \{2^{N-2S}1^{2S}\} \\
&\quad = ([l] \otimes \{1^N\})([l] \otimes \{1^{N-2S}\}) \\
&\qquad - ([l] \otimes \{1^{N+1}\})([l] \otimes \{1^{N-2S-1}\}) \\
&\quad = ([(2l - N + 1)/2] \otimes \{N\}) \\
&\qquad \times ([(2l - N + 2S + 1)/2] \otimes \{N - 2S\}) \\
&\qquad - ([(2l - N)/2] \otimes \{N + 1\}) \\
&\qquad \times ([(2l - N + 2S)/2] \otimes \{N - 2S - 1\}),
\end{aligned} \tag{9}$$

where the last line follows from Hermite's reciprocity principle.[5] Equation (9) suggests that any discussion of the statistical distribution of orbital angular-momentum states will require a knowledge of the distribution of the terms arising in the basic plethysm

$$[p] \otimes \{k\} = \sum_L g_{pkL}[L], \tag{10}$$

where p is an integer or half-integer and k is an integer.

For nucleon configurations involving both protons and neutrons, it is necessary also to consider the isotopic spin T which necessitates the examination of the reduction of S functions associated with Young tableaux of up to four columns, and the analogs of Eq. (9) are somewhat more complex.

The angular momentum states $[J]$ of a jj-coupled configuration j^N of electrons are just those terms arising in the plethysm

$$[j] \otimes \{1^N\} = [(2j - N + 1)/2] \otimes \{N\}, \tag{11}$$

assuming that under $U_{2j+1} \rightarrow R_3$ we have $\{1\} \rightarrow [j]$. Again the basic plethysm is of the form

$$[p] \otimes \{k\} = \sum_J g_{pkJ}[J]. \tag{12}$$

Equations (10) and (12) indicate that a statistical group theory of angular momentum states must place considerable emphasis on the terms arising in the plethysm

$$[p] \otimes \{k\} = \sum_r g_{pkr}[r], \tag{13}$$

where k is an integer and p and r are integers or half-integers.

III. ANGULAR MOMENTUM STATES AND PARTITIONS

Having established the relevance of the plethysm of S functions to the traditional spectroscopic problem of the enumeration of angular momentum states, we now translate the plethysm given in Eq. (13) into partition functions familiar in number theory.[11] The terms arising in the plethysm given by Eq. (13) may be evaluated by noting a theorem due to Littlewood,[7] which states that, if

$$[p] \times \{k\} = \sum g_{pkr}[r], \tag{14}$$

then g_{pkr} is the coefficient of ρ^{-r} in the expansion of

$$\rho^{-pk}(1-\rho)\prod_{i=1}^{k}[(1-\rho^{2p+i})/(1-\rho^{i})]. \tag{15}$$

The evaluation of the coefficients of ρ^{-r} in the above expansion is a very tedious, though elementary, task and is well suited to machine computation. Even plethysms such as $[8] \otimes \{13\}$, where the coefficient of ρ^{-24} is 30 598, may be rapidly evaluated by a moderately fast computer. A plot of the coefficients of ρ^{-r} versus r for $[8] \otimes \{13\}$ is displayed in Fig. 1. This plot can be considered as representing the distribution of the orbital states $[L]$ associated with the maximum multiplicity of the configuration l^{13}, $l = 14$, but we postpone detailed discussion until later.

While Eq. (14) is well adapted to machine calculation, it is not in a form suitable for describing the distribution of the coefficients of ρ^{-r} as p and k become very large.

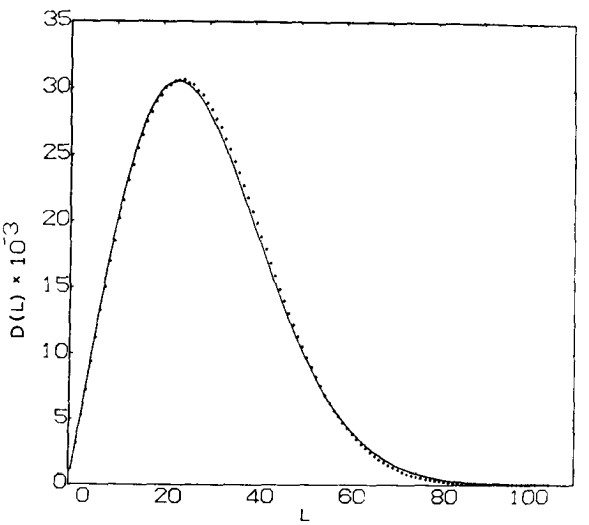


FIG. 1. Distribution $D(L)$ of the orbital states associated with maximum multiplicity for l^{13}. The dotted points are from the explicit calculation of the values of $D(L)$ for the integers L, while the smooth curve is the Wigner-type distribution, $\sigma = 23.2$ and $A = 2170$.

From number theory and the theory of partitions,[11,12] it is well known that the coefficient of x^q in

$$\prod_{i=1}^{k}((1-x^{m+i})/(1-x^{i}))$$

is, in Cayley's notation,[12] just $P_m^k(q)$, which is the number of ordered partitions of q made up of the sum of m terms with the elements $0, 1, 2, \cdots, k$. The ordering is in ascending magnitude of the elements in going from left to right. The elements of the partition are all positive integers, and we define

$$P_k^m(0) = P_k^m(1) = 1, \quad k \geq 1. \tag{16}$$

We may now express the coefficient g_{pkr} of Eq. (14) as

$$g_{pkr} = P_k^{2p}(x) - P_k^{2p}(x-1), \tag{17}$$

where $x = kp - r$. This result forms the key to analyzing the statistical distribution of orbital states $[L]$, with maximum multiplicity, for the case of LS coupling or of the total angular momentum states $[J]$ for jj coupling.

Equation (17) possesses a number of obvious properties which readily demonstrate well-known results. Thus, the maximum value of r is pk; in this case, $g_{pkr} = 1$, and it follows from Eq. (16) that, for the penultimate value of r, we have $g_{pkr} = 0$. If $2p \geq k$, then the coefficient g_{pkr} becomes

$$g_{pkr} = P_k(x) - P_k(x-1) = P_k^*(x), \tag{18}$$

where $P_k(x)$ is the number of ordered partitions of x with no part greater than k and where $P_k^*(x)$ is the number of partitions of x with no part less than k.

While mathematicians have devoted considerable attention to the asymptotic form of $P(x)$, the number of ordered partitions of n without restriction, very little has been reported on the asymptotic behavior of the coefficients defined by Eqs. (17) and (18). In the case of Eq. (18), it is not difficult to establish the special results:

$$P_2^*(x) = 1, \quad x \text{ even}, \qquad = 0, \quad x \text{ odd}, \quad x \geq 2; \tag{19a}$$

$$P_3^*(x) = t + 1 \begin{cases} t = \text{c.}\,(\tfrac{1}{6}x), & x \text{ even}, \\ t = \text{c.}\,[\tfrac{1}{6}(x-3)], & x \text{ odd}. \end{cases} \tag{19b}$$

These two results suggest that the distributions for the odd and even values of x are distinct, as is indeed found by empirical observation. The extension of Eq. (18) to cases with $k > 3$ leads to unwieldy formulas of increasing complexity, and for this reason an alternative method of approximating g_{pkr} by using a series of polynomials to generate the partition numbers has been developed.

IV. POLYNOMIAL COEFFICIENTS AND PARTITION NUMBERS

We now consider the problem of representing the partition numbers associated with Eq. (17) in terms of polynomial coefficients, which are then used to provide an approximation for the distribution of the number g_{pkr}. To do this, we must first develop a systematic notation to provide a connection between the representation of ordered and unordered partitions, taking care to distinguish this notation from that commonly used for S functions and the labeling of group representations.

We commence the development of our notation by first considering a few examples which will be used to illustrate the general notation described later.

The number of ordered partitions of the positive integer x into m parts may be designated as $(1^m)_Q(x)$. If a partition of x has at least two members equal, the number of such partitions is written as $(1^{m-2}2^1)_Q(x)$. The subscripted letter Q is used to designate this type of partition. In this case, we have

$$x = a_{11} + \cdots + a_{1m-2} + 2a_{21}, \tag{20}$$

where the a_{ij} are the set of integers constituting a particular partition of x and where the coefficients of the a_{ij} give the number of times they occur in the partition. For example, if we let $m = 5$ in Eq. (20), two sets A satisfying this relationship would be

$$6 = 0 + 0 + 0 + 2.3$$

or

$$7 = 0 + 2 + 3 + 2.1,$$

where the corresponding partitions are

$$0 \quad 0 \quad 0 \quad 3 \quad 3$$

and

$$0 \quad 1 \quad 1 \quad 2 \quad 3.$$

Similarly, if a partition is a member of the set $(1^{m-4}2^2)_Q(x)$, then it has at least two pairs of equal members. A typical example for $m = 5$ would be $0 + 0 + 1 + 1 + 2$.

The above notation may be made quite general by letting

$$(n_1^{l_1} n_2^{l_2} \cdots n_q^{l_q})_Q(x)$$

denote the number of ordered selections of a_{ij}, (A), such that

$$\begin{aligned} x = {} & n_1(a_{11} + a_{12} + \cdots + a_{1l_1}) \\ & + n_2(a_{21} + \cdots + a_{2l_2}) \\ & + \cdots + n_q(a_{q1} + \cdots + a_{ql_q}). \end{aligned} \tag{21}$$

Now it will be noticed that

$$(1^m)_Q(x) = P^m(x).$$

If the a_{ij} are restricted to $a_{ij} \leq k$, then

$$(1^m)_Q(x) = P_k^m(x). \tag{22}$$

Later, we will not wish to distinguish between the various x's and k's and will write

$$(n_1^{l_1} \cdots n_q^{l_q})_Q$$

for the number of ordered selections satisfying Eq. (21) for any given x or k.

The unordered sets of a_{ij} satisfying Eq. (21) will be designated by $(n_1^{l_1} \cdots n_q^{l_q})_P$. In a given set A, it is possible that two or more a_{ij} are equal without necessarily having equivalent subscripts. If we restrict ourselves to the case where all a_{ij} are ordered and distinct, we have a new set of solutions of Eq. (21) (a subset of the more general case) which we will write as $(n_1^{l_1} \cdots n_q^{l_q})_R$. We shall later refer to arbitrary representations of $(n_1^{l_1} \cdots n_q^{l_q})_Q$, $(n_1^{l_1} \cdots n_q^{l_q})_P$, and $(n_1^{l_1} \cdots n_q^{l_q})_R$ as Q, P, and R representations, respectively. The utility of this rotation is that, as we shall show, we can express the Q and P partitions in terms of the R partitions, thus permitting a solution of the Q partitions in terms of the P partitions.

The P partition $(n_1^{l_1} \cdots n_q^{l_q})_P(x)$ with $a_{ij} \leq k$ is just the coefficient of z^x in the product

$$\prod_{i=1}^{q} (1 + z^{n_i} \cdots z^{n_i k})^{l_i}, \tag{23}$$

and it is from an approximation of the coefficients of this product that an approximation of the numbers of the partition emerges.

V. RELATIONSHIPS BETWEEN Q-, P-, AND R-TYPE PARTITIONS

We now develop relationships between the Q-, P-, and R-type partitions which will later be used to analyze angular momentum distributions. We may readily establish the typical relationships

$$\begin{aligned} (1^4)_Q &= (1^4)_R + (1^2 2)_R + (2^2)_R + (13)_R + (4)_R, \\ (1^2 2)_Q &= (1^2 2)_R + (2^2)_R + (13)_R + (4)_R, \\ (2^2)_Q &= (2^2)_R + (4)_R, \\ (13)_Q &= (13)_R + (4)_R, \\ (4)_Q &= (4)_R. \end{aligned} \tag{24}$$

The general method of obtaining these relationships is to select differing groups of a_{ij} as being equal and to make no distinction between members in which the first subscripts are equal. The general problem of establishing these relationships is by no means

TABLE I. Representation of Q-type partitions in terms of P-type partitions.

	M	N	MN^{-1}
$m=2$	$\begin{pmatrix}1&1\\0&1\end{pmatrix}$	$\begin{pmatrix}2&1\\0&1\end{pmatrix}$	$\frac{1}{2!}\begin{pmatrix}1&1\\0&2\end{pmatrix}$
	$(1^2)_Q = \frac{1}{2!}[(1^2)_P + (2^1)_P]$		
$m=3$	$\begin{pmatrix}1&1&1\\&1&1\\&&1\end{pmatrix}$	$\begin{pmatrix}6&3&1\\&1&1\\&&1\end{pmatrix}$	$\frac{1}{3!}\begin{pmatrix}1&3&2\\&6&0\\&&6\end{pmatrix}$
	$(1^3)_Q = \frac{1}{3!}[(1^3)_P + 3(1^12^1)_P + 2(3^1)_P]$		
$m=4$	$\begin{pmatrix}1&1&1&1&1\\&1&1&1&1\\&&1&0&1\\&&&1&1\\&&&&1\end{pmatrix}$	$\begin{pmatrix}24&12&6&4&1\\&2&2&2&1\\&&2&0&1\\&&&1&1\\&&&&1\end{pmatrix}$	$\frac{1}{4!}\begin{pmatrix}1&6&3&8&6\\&12&0&0&12\\&&12&0&12\\&&&24&0\\&&&&24\end{pmatrix}$
	$(1^4)_Q = \frac{1}{4!}[(1^4)_P + 6(1^22^1)_P + 3(2^2)_P + 8(1^13^1)_P + 6(4^1)_P]$		

trivial especially because, in higher cases, coefficients greater than one appear. Examples of this are

$$(1^22^2)_Q = (1^22^2)_R + (2^3)_R + (123)_R + (3^2)_R + (1^24)_R + 2(24)_R + (15)_R + (6)_R,$$

$$(1^22^24)_Q = (1^22^24)_R + (2^34)_R + (1234)_R + (3^24)_R + (1^24^2)_R + 2(24^2)_R + (12^25)_R + (235)_R + 2(145)_R + (5^2)_R + (1^226)_R + 2(2^26)_R + (136)_R + 3(46)_R + (127)_R + (37)_R + (1^28)_R + 2(28)_R + (19)_R + (10)_R.$$

Relationships, such as given in Eq. (24), between the Q- and R-type partitions may be written in matrix form by treating the various Q and R partitions as components of vectors $\mathbf{Q}$ and $\mathbf{R}$, where the ordering of the components is standard, as in Eq. (24). Thus, we have

$$\mathbf{Q} = M\mathbf{R}, \tag{25}$$

where, in the case of Eq. (24), the matrix M is

$$\begin{pmatrix}1&1&1&1&1\\&1&1&1&1\\&&1&0&1\\0&&&1&1\\&&&&1\end{pmatrix}.$$

By considering the foregoing in terms of set theory, it can be shown that the transformation matrix M always exists and is upper triangular.

The problem of expressing P partitions in terms of R partitions is a somewhat more difficult problem, but the result is similar to that of Eq. (25). We may generally write in vector form

$$\mathbf{P} = N\mathbf{R}, \tag{26}$$

where N is a transformation matrix similar to that of the M matrix of Eq. (25), namely, N is upper triangular with no zero terms on the diagonal. Thus, the inverse matrix N^{-1} exists, and we may write

$$\mathbf{Q} = MN^{-1}\mathbf{P}, \tag{27}$$

which allows us to relate the Q partitions to the P partitions. However, since we are only interested here in terms of the form $(1^m)_Q$, the enumeration of MN^{-1} is not in general necessary. Solutions for $m = 2$ to 4 are given in Table I.

VI. ANGULAR MOMENTA DISTRIBUTION FORMULA

We now show that the coefficients of the polynomials have, in the limit, a normal distribution; as a consequence, the numbers g_{pkr} follow a Wigner-type distribution with respect to r.

The coefficients of polynomials of the form

$$(1 + x + \cdots + x^m)^n$$

tend to a normal distribution as $n \to \infty$. For $m = 1$ this is, of course, the well-known binomial case. Proof for higher values of m may be readily obtained by

statistical methods. Thus, we will assume a distribution of the form $A \exp([-(x-\mu)^2/2\sigma^2])$ for the terms in the expansion of $(1^m)_Q$ as $m \to \infty$. Using the notation of Eq. (23), we obtain

$$\begin{aligned}\mu &= \tfrac{1}{2}k \sum n_i l_i \\ &= \tfrac{1}{2}qk. \end{aligned} \tag{28}$$

Recalling Eq. (17),

$$g_{pkr} = P_k^{2p}(x) - P_k^{2p}(x-1),$$

and, substituting for x, we obtain

$$g_{pkr} = P_k^{2p}(kp - r) - P_k^{2p}(kp - r - 1). \tag{29}$$

But $q = 2p$ and hence $\mu = kp$, so that if we let

$$G_k^{2p}(r) = P_k^{2p}(kp - r), \tag{30}$$

then $G(r)$ will in general be a sum of normal distributions with mean zero and

$$g_{pkr} = G_k^{2p}(r) - G_k^{2p}(r+1). \tag{31}$$

Using the finite-difference operator ∇, defined by

$$\nabla F(x) \equiv F(x) - F(x+1), \tag{32}$$

we have

$$g_{pkr} = \nabla G_k^{2p}(r). \tag{33}$$

Now

$$G(r) = \sum A_i \exp\left(\frac{-x^2}{2\sigma_i^2}\right),$$

and, remembering that ∇ is linear, we will consider

$$\begin{aligned}\nabla\left[\exp\left(\frac{-x^2}{2\sigma^2}\right)\right] &= \exp\left(\frac{-r^2}{2\sigma^2}\right) - \exp\left(\frac{-(r+1)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{-(r+\frac{1}{2})^2}{2\sigma^2}\right)\left[\exp\left(\frac{r+\frac{1}{4}}{2\sigma^2}\right)\right. \\ &\quad \left. - \exp\left(\frac{-(r+\frac{1}{4})}{2\sigma^2}\right)\right].\end{aligned}$$

Upon expansion, we find that

$$\begin{aligned}\nabla\left[\exp\left(\frac{-x^2}{2\sigma^2}\right)\right] &= \frac{(r+\frac{1}{2})}{\sigma^2}\exp\left(\frac{-(r+\frac{1}{2})^2}{2\sigma^2}\right) \\ &\quad \times \left[1 - \frac{1}{4\sigma^2}\cdots\right],\end{aligned}$$

but $\sigma \to \infty$ as $m \to \infty$, and hence

$$\nabla\left[\exp\left(\frac{-x^2}{2\sigma^2}\right)\right] \approx \frac{r+\frac{1}{2}}{\sigma^2}\exp\left(\frac{-(r+\frac{1}{2})^2}{2\sigma^2}\right), \tag{34}$$

which is just the usual form of the Wigner distribution.

Thus, the distribution g_{pkr} is a sum of Wigner-type distributions and not just a single Wigner distribution, enabling us to write

$$\nabla G_m^{2p}(r) = \sum A_i(r+\tfrac{1}{2})\exp\left(\frac{-(r+\frac{1}{2})^2}{2\sigma_i^2}\right). \tag{35}$$

VII. STATES OF NONMAXIMUM MULTIPLICITY

If, as we would hope, states of nonmaximum multiplicity are also, in the limit, representable in terms of a sum of Wigner-type distributions, then it will suffice to show that the multiplication of functions $G(r)$ by using angular momentum multiplication preserves the character of the distribution since the states of nonmaximum multiplicity can be expressed as sums of products of the states of maximum multiplicity.

Remembering the usual rule for angular momentum products,

$$\begin{aligned}L \times L' = (L + L') + (L + L' - 1) \\ + \cdots + |L - L'|,\end{aligned} \tag{36}$$

we can write the product of two distributions as

$$\begin{aligned}(G_1 \times G_2)(r) &= \sum_{i=0}^{\infty} G(i) \sum_{k=r-i}^{r+i} G(k), \quad r > i, \\ &= \sum_{i=0}^{\infty} G(i) \sum_{k=i-r}^{r+i} G(k), \quad r < i.\end{aligned} \tag{37}$$

But, if we extend $G(r)$ by defining

$$G(-r-1) = G(r), \tag{38}$$

we have

$$(G_1 \times G_2)(r) = \sum_{i=0}^{\infty} G(i) \sum_{k=r-i}^{r+i} G(k), \quad \forall\, r, i. \tag{39}$$

Now

$$\begin{aligned}&\nabla(G_1 \times G_2) \\ &\quad = \sum_{i=0}^{\infty} G_1(i)\left(\sum_{k=r-i}^{r+i} G_2(k) - \sum_{k=r-i+1}^{r+i+1} G_2(k)\right) \\ &\quad = \sum_{i=0}^{\infty}[G_1(i)G_2(r-i) - G_2(i)G(r+i+1)].\end{aligned} \tag{40}$$

If we let $i' = -i - 1$ and make use of Eq. (38), then

$$\nabla(G_1 \times G_2) = \sum_{i=-\infty}^{\infty} G_1(i)G_2(r-i). \tag{41}$$

But this is the usual polynomial product, and, if we denote this by

$$(G_1 \vee G_2)(r) = \sum_{i=-\infty}^{\infty} G_1(i)G_2(r-i), \tag{42}$$

we have

$$G_1 \times G_2 = \nabla^{-1}(G_1 \vee G_2). \tag{43}$$

Using the linearity of ∇, we obtain

$$G_1 \times G_2 = \nabla\,[(\nabla^{-1}G_1) \vee (\nabla^{-1}G_2)].$$

But $\nabla^{-1}G$ is a normal distribution which is reproduced under polynomial multiplication, so that we may write

$$G_1 \times G_2 = \nabla N, \tag{44}$$

where N is a normal distribution and, of course, ∇N is the Wigner-type distribution.

Thus, the Wigner-type distribution is reproduced under normal angular momentum multiplication. This is an aesthetically very pleasing result and we feel that it argues strongly for consideration of the distribution of the g_{pkr} in terms of the Wigner distribution.

VIII. EXAMPLE OF i^5

We now calculate a specific example of the foregoing theory, and in the process we demonstrate one of the more outstanding deviations from the Wigner-type distribution. Let us consider the states of maximum multiplicity of the i^5 configuration which is sufficiently simple to be readily accessible numerically, but large enough to exhibit the main features. Recall Eq. (17):

$$g_{pkr} = P_k^{2p}(x) - P_k^{2p}(x-1), \quad x = kp - r.$$

Normally we write

$$2p = 2l + 1 - N \tag{45a}$$

and

$$k = N, \tag{45b}$$

where N is the number of electrons with orbital angular momentum l.

However, since we deal here only with states of maximum multiplicity, we may simplify our calculations by recalling the $\frac{1}{4}$-shell symmetry of these states and considering only the cases where

$$2p = N \tag{46a}$$

and

$$k = 2l + 1 - N. \tag{46b}$$

Thus, for i^5, from Eqs. (17) and (32), we have

$$D(L) = P_8^5(20 - L). \tag{47}$$

From the solution of the matrix MN^{-1} for $q = 5$, we obtain the relationship

$$(1^5)_Q = \frac{1}{5!}[(1^5)_P + 10(1^3 2^1)_P + 15(1^1 2^2)_P + 20(1^2 3^1)_P + 20(2^1 3^1)_P + 30(1^1 4^1)_P + 24(5^1)_P]. \tag{48}$$

Equations (47) and (48) then lead to

$$D(L) = \tfrac{1}{120}\nabla(1^5)_P + \tfrac{1}{12}\nabla(1^3 2^1)_P + \tfrac{1}{8}\nabla(1^1 2^2)_P + \tfrac{1}{6}\nabla(1^2 3^1)_P + \tfrac{1}{6}\nabla(2^1 3^1)_P + \tfrac{1}{4}\nabla(1^1 4^1)_P + \tfrac{1}{5}\nabla(5^1)_P, \tag{49}$$

with $a_{ij} \leq 8$ [see Eq. (21)].

The distribution $D(L)$ for the terms of maximum multiplicity in i^5 is plotted in Fig. 2. It will be noted

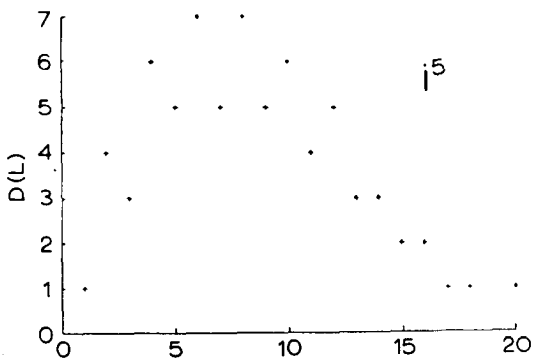


FIG. 2. Plot of $D(L)$ vs L for the terms of i^5 with maximum multiplicity.

that the graph appears to split into two, one for *even* values of L being consistently less than that for *odd* values of L. This type of deviation from the Wigner form tends to be common and can be explained in terms of the P partitions making up the various distributions.

Examination of the individual terms in Eq. (49) reveals that most of the "splitting" is caused by the term $\nabla(1^1 2^2)_P$ and a small contribution from $\nabla(1^3 2^1)_P$. For any general term of the form $\nabla(1^m 2^n)_P$, we have

$$\nabla(1^m 2^n)_P = (1^m)_P \vee \nabla(2^n)_P. \tag{50}$$

However,

$$[\nabla(2^n)_P](2x+1) = -[\nabla(2^n)_P](2x),$$

so that all the terms $\nabla(1^m 2^n)_P$ tend to have large differences between their odd and even values and thus give the very common splitting effect. Similar effects for terms of the form $\nabla(1^m 3^n)_P$, etc., occur but are less prominent because, in general, the terms $\nabla(1^q)_P$, $\nabla(1^{q-2}2^1)_P$, and $\nabla(1^{q-4}2^2)_P$ tend to predominate in any given distribution. Furthermore, many of the terms tend mutually to cancel their respective splittings. These effects become less important as q becomes large since the terms $\nabla(1^{q-2}2^1)_P$, etc., tend to become smoothed.

IX. ESTIMATION OF A AND σ

The results of the previous sections have established that, for sufficiently large l and N, the orbital angular momenta L of the states of maximum multiplicity of an electron configuration l^N are distributed according to the Wigner-type distribution

$$D(L) = \sum_i A_i(L + \tfrac{1}{2}) \exp[-(L+\tfrac{1}{2})^2/2\sigma_i^2] \tag{51}$$

or, to a lesser approximation, as a single Wigner distribution

$$D(L) = A(L + \tfrac{1}{2}) \exp[-(L+\tfrac{1}{2})^2/2\sigma^2]. \tag{52}$$

We now proceed to estimate the values of A and σ. Two procedures are possible: Either we can obtain the A_i and σ_i of the various polynomials associated with the P-type partitions which sum to give the g_{pkr} or we can obtain the value of σ from the value of L for which $D(L)$ is a maximum and then relate σ to A. The first method requires the solution of the MN^{-1} matrix for the requisite values of q and becomes a complex, though interesting, problem for large values of q. The estimation is straightforward though tedious. In the case of i^5 we find

$$\begin{aligned}(1^5)_P:&\quad \sigma_1 = 5.75,\quad A_1 = 1.44,\\ (1^32^1)_P:&\quad \sigma_2 = 6.85,\quad A_2 = 0.44,\\ (1^12^2)_P:&\quad \sigma_3 = 7.75,\quad A_3 = 0.055,\end{aligned}$$

which is in reasonable agreement with the observed position of the maximum at $L = 7$.

For t^{13} we have $q = 13$ and the above method becomes extremely tedious. We shall simply estimate A and σ, assuming a single distribution. We first note that the orbital states associated with the states of maximum multiplicity in l^N span the $\{1^N\}$ representation of U_{2l+1}, and hence from dimensional considerations we have

$$\begin{aligned}\sum (2L+1)D(L) &= \binom{2l+1}{N}\\ &\approx \int_0^\infty (2L+1)A(L+\tfrac{1}{2})\exp\left[-(L+\tfrac{1}{2})^2/2\sigma^2\right]dL\\ &= A\sigma^3(2\pi)^{\frac{1}{2}},\end{aligned} \tag{53}$$

which provides an important relationship between A and σ. Likewise, it is a simple matter to show that

$$\langle 2L+1\rangle \approx (2\pi)^{\frac{1}{2}}\sigma. \tag{54}$$

X. EXAMPLE OF t^{13}

Explicit calculation of the orbital angular momentum states associated with maximum multiplicity in t^{13} gave the points plotted in Fig. 1, from which we deduce

$$\sigma = 23.2 \quad \text{and} \quad A = 2170, \tag{55}$$

the value of A being obtained via Eq. (53). A plot of the Wigner-type distribution (Eq. 52) using these values has been superimposed on the points of Fig. 1 and shows a remarkably good fit. Use of Eq. (54) gives $\langle 2L+1\rangle = 58.2$, which may be compared with the calculated value of 58.25.

XI. CONCLUSION

We have shown, using some extensions of the theory of partitions and groups, that for $l \gg 0$ the distribution of the orbital states in the LS-coupling scheme of a configuration l^N of equivalent electrons may be approximated by a Wigner-type distribution. Exactly the same analysis may be carried out for the jj-coupling scheme with identical conclusions. Similar results should be obtained for the coupling of nucleons in the nuclear case.

A number of problems remain to be resolved. It would clearly be desirable to obtain expressions for A and σ directly in terms of the number of particles N and their orbital momenta l. This problem is closely akin to the elusive problem of solving the Clebsch–Gordan series in an analytic sense. A number of interesting aspects of the theory of partitions that could impinge on this problem remain to be fully investigated. In this sense our present work must be regarded as a preliminary attack on the problem of examining some statistical aspects of group theory and the many-body problem.

ACKNOWLEDGMENTS

We wish to take the opportunity of thanking P. H. Butler for providing the numerical data upon which this paper is based. Conversations and correspondence with Dr. N. Rosenzweig of Argonne National Laboratories have been both encouraging and helpful. One of us (J. G. C.) also acknowledges a Research Assistantship grant from the University of Canterbury.

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR Grant Number 1275-67.

[1] C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic, New York, 1965).
[2] C. Bloch, Phys. Rev. **93**, 1094 (1954).
[3] E. P. Wigner, Oak Ridge Natl. Lab. Rept. ORNL-2309, (1957), p. 67 (reprinted in Ref. 1).
[4] B. R. Judd, *Operator Techniques in Atomic Spectroscopy*, (McGraw-Hill, New York, 1962).
[5] B. G. Wybourne, J. Math. Phys. **10**, 467 (1969).
[6] B. G. Wybourne and P. H. Butler, J. Phys. **30**, 181 (1969).
[7] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U.P., London, 1950), 2nd ed.
[8] P. R. Smith and B. G. Wybourne, J. Math. Phys. **8**, 2434 (1967); **9**, 1040 (1968).
[9] P. H. Butler and B. G. Wybourne, J. Phys. **30**, 795 (1969).
[10] B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970).
[11] E. Grosswald, *Topics from the Theory of Numbers* (Macmillan, New York, 1966).
[12] P. A. Macmahon, *Combinatory Analysis* (Cambridge U.P., London, 1915), Vols. I and II.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Probabilistic Interpretation of the Classical Scattering Cross Section

DONALD G. BABBITT*
University of California, Los Angeles, Department of Mathematics, California 90024

(Received 22 June 1970)

The probabilistic interpretation of the classical scattering cross section is discussed in a mathematically rigorous framework. In particular, extensive use is made of the notion of a Poisson process based on a plane and on a sphere. The case of Rutherford scattering is given as a detailed illustration.

1. INTRODUCTION

A casual reading of the standard treatments of classical 1-particle potential scattering indicate the differential cross section and the total cross section have (at least vaguely) a probabilistic interpretation. For example, see Newton, Ref. 1, Chap. 5. It is our intention in this paper to make explicit the probabilistic interpretation of these observables. A similar interpretation should hold in the more important quantum mechanical case.

In Sec. 2 we define the notion of a scattering system which includes the notion of a potential scattering system as a special case. We feel our more general definition isolates the essential features of classical elastic scattering.[2] In Sec. 3 we make precise the notion of an infinite beam of particles with a fixed velocity v but with uniformly distributed impact parameters. Once this is done, we shall see that the cross-section measure of a scattering system S in the direction v is a measure $\sigma(v, \cdot)$ on S^2, the 2-sphere in $\mathbb{R}^3$. In particular, for $\Sigma \subseteq S^2$, $\sigma(v, \Sigma)$ will turn out to be the expected number of particles in the beam which are scattered in the direction of Σ divided by the beam intensity. If $\sigma(v, \cdot)$ has a Radon–Nikodym derivative $d\sigma(v, \cdot)/d\Omega$ with respect to the canonical measure Ω on S^2, we call $d\sigma(v, \cdot)/d\Omega$ the differential cross section of S in the direction v. The integral

$$\int_{(S^2 - v/\|v\|)} \frac{d\sigma(v, \cdot)}{d\Omega}\, d\Omega$$

is called the total cross section (perhaps infinite), and it is the expected number of particles in the beam which are scattered, divided by the beam intensity.

In Sec. 4 we prove an approximation theorem which we feel justifies our definition of an incoming beam of particles. The proof of the theorem is rather trivial and only requires results from a junior level probability course.

Finally, in Sec. 5 we work out the example of Rutherford scattering in our framework.

2. CLASSICAL SCATTERING SYSTEM

Throughout the remainder of the paper "ex" will stand for either "in" or "out." Let U_{ex} be a Borel subset in $\mathbb{R}^3 \times \mathbb{R}^3$ possessing the following properties: (a) If $(x_{\text{ex}}, v_{\text{ex}}) \in U_{\text{ex}}$, then $v_{\text{ex}} \neq 0$; (b) if $(x_{\text{ex}}, v_{\text{ex}}) \in U_{\text{ex}}$, then $(x_{\text{ex}} - v_{\text{ex}}t_0, v_{\text{ex}}) \in U_{\text{ex}}$ for all $t_0 \in \mathbb{R}$. Let $\mathbb{A}^{\text{ex}} = \{X_{\text{ex}}: \mathbb{R} \to \mathbb{R}^3 \mid X_{\text{ex}}(t) = x_{\text{ex}} + v_{\text{ex}}t, \quad t \in \mathbb{R}, \text{ for some } (x_{\text{ex}} v_{\text{ex}}) \in U_{\text{ex}}\}$. We think of the first $\mathbb{R}^3$ in $\mathbb{R}^3 \times \mathbb{R}^3$ as physical space, the second $\mathbb{R}^3$ as the space of velocities and elements of $\mathbb{A}^{\text{in}}$ and $\mathbb{A}^{\text{out}}$ as the incoming and outgoing asymptotes of the system we wish to study. Typical examples of U_{ex} would be $\mathbb{R}^3 \times (\mathbb{R}^3 - \{0\})$ itself and

$$\{(x, v) \in \mathbb{R}^3 \times (\mathbb{R}^3 - \{0\}) \mid x + tv \neq 0 \text{ for } \forall\, t\}.$$

In particular, if U_{in} is such that it is contained in the first set and contains the second set we say that $\mathbb{A}^{\text{in}}$ is *standard.*

Note that we have a natural action of "time" on $\mathbb{A}^{\text{ex}}$ defined by $T_{t_0}X_{\text{ex}}(t) \equiv X_{\text{ex}}(t - t)$, for $t, t_0 \in \mathbb{R}$. Note also that $\mathbb{A}^{\text{ex}}$ has a Borel structure it inherited from U_{ex}.

Definition: A (classical 1-particle) scattering system consists of a triple $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ where $\mathbb{A}^{\text{ex}}$ is as above and $S: \mathbb{A}^{\text{in}} \to \mathbb{A}^{\text{out}}$ is a bijective Borel mapping which commutes with the time translations $\{T_t: -\infty < -\infty\}$. Elements of $\mathbb{A}^{\text{in}}$ are called the "incoming asymptotes" of the system $\mathbb{A}^{\text{out}}$, the "outgoing asymptotes" of the system, and S is called the *scattering operator* of the system. $S(X_{\text{in}})$ is interpreted as the outgoing asymptote of a particle when X_{in} is its incoming asymptote. Thus S is the analog of the S matrix in quantum mechanics. When no confusion should arise, we shall use S to designate the scattering system.

The fact that S commutes with time translation essentially says the scattering system is autonomous. Note that S has the following nice feature. Let $v \in \mathbb{R}^3$ be such that there exists $x \in \mathbb{R}^3 \ni (x, v) \in U_{\text{in}}$

and let P_v be the plane in $\mathbb{R}^3$ which is perpendicular to v and which passes through O ($\in R^3$). Finally let

$$\mathbb{A}_v^{\text{in}} \equiv \{X_{\text{in}} \in \mathbb{A}^{\text{in}} \mid X_{\text{in}}(t) = x_{\text{in}} + vt, \forall\, t \text{ and } x_{\text{in}} \in P_v\}$$

and $S_v \equiv S|\mathbb{A}_{v}^{\text{in}}$. Observe that S is completely determined when we know the $\{S_v\}$ where v is described above.

To obtain a general example of a scattering system, consider the dynamical system governed by the differential equation in $\mathbb{R}^3$:

$$m\frac{d^2x}{dt^2} = F(x, \dot{x}), \tag{2.1}$$

where F is C' on $\mathbb{R}^3 \times \mathbb{R}^3$. Let $\mathbb{A}^{\text{in}}$ be the set of affine mappings $x_{\text{in}}: \mathbb{R} \to \mathbb{R}^3: t \rightsquigarrow x_{\text{in}} + v_{\text{in}}t$ with the following properties: (a) $v_{\text{in}} \neq 0$; (b) there exists a unique solution $x(\cdot)$ of (2.1) which satisfies

$$\lim_{t\to-\infty} \|X_{\text{in}}(t) - x(t)\| = 0,$$

$$\lim_{t\to-\infty} \left\| v_{\text{in}} - \frac{dx}{dt}(t) \right\| = 0;$$

(c) there exists an affine mapping

$$X_{\text{out}}: \mathbb{R} \to \mathbb{R}^3: t \rightsquigarrow x_{\text{out}} + v_{\text{out}}t$$

such that

$$\lim_{t\to+\infty} \|X_{\text{out}}(t) - x(t)\| = 0, \tag{2.2}$$

$$\lim_{t\to+\infty} \left\| v_{\text{out}} - \frac{dx}{dt} \right\| = 0. \tag{2.3}$$

Note that X_{out} is uniquely determined by (2.2) and (2.3). Let $\mathbb{A}^{\text{out}}$ be the set of X_{out} described by (c) and let $S: \mathbb{A}^{\text{in}} \to \mathbb{A}^{\text{out}}$ be defined by $S(X_{\text{in}}) = X_{\text{out}}$. It is clear that S is bijective and that it commutes with time translation. It is true, but technically messy to show, that $\mathbb{A}^{\text{in}}$ and $\mathbb{A}^{\text{out}}$ inherit a natural Borel structure from $\mathbb{R}^3 \times \mathbb{R}^3$ and S is Borel measurable. Thus $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ is a scattering system.

In general, it would be difficult to determine $\mathbb{A}^{\text{in}}$ and $\mathbb{A}^{\text{out}}$ from (2.1). However, in the case of the Coulomb or gravitational problem, $\mathbb{A}^{\text{in}}$ is standard and $\mathbb{A}^{\text{out}}$ and S can be explicitly computed (see Sec. 5). In fact, most reasonable scattering systems will be such that $\mathbb{A}^{\text{in}}$ is standard.

3. THE CROSS-SECTION OF A SCATTERING SYSTEM AND ITS PROBABILITIES INTERPRETATION

One of the main reasons why one considers cross sections of a scattering system $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ is that they are experimentally more accessible than the scattering operator S. Roughly speaking, cross sections tell us *the directions into which parallel beams of incoming particles are scattered by S.* Thus our first task is to define in a mathematically precise way the notion of a beam of incoming particles.

Let $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ be a scattering system. We shall say that the system is *standard* if $\mathbb{A}^{\text{in}}$ is standard. For $0 \neq v \in \mathbb{R}^3$, let P_v be the plane perpendicular to v and passing through O. The *impact set* $\mathbf{I}_v$ of the scattering system for the velocity v is defined as follows:

$$\mathbf{I}_v = \{x \in P_v \mid X: R \to R^3: t \rightsquigarrow x + tv \text{ is an element of } \mathbb{A}^{\text{in}}\}.$$

If $\mathbb{A}^{\text{in}}$ is standard, $\mathbf{I}_v = P_v$ or $P_v - \{0\}$.

Definition: Let $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ be a standard scattering system. *An (infinite homogeneous) incoming beam of particles with velocity v and intensity I* is a Poisson process $Y_{v,I} = Y$ on $\mathbf{I}_v$ with measure $I \cdot \lambda$, where λ is a (2-dimensional) Lebesque measure on $\mathbf{I}_v$. Recall that a Poisson process with measure ν on a measurable space $(\mathfrak{M}, \mathfrak{B})$ is a function Y from $\mathfrak{B}$ to a space of nonnegative integer-valued random variables such that: (a) $Y(B)$, $B \in \mathfrak{B}$, is a Poisson random variable with mean $\nu(B)$; (b) if B_1, $B_2 \in \mathfrak{B}$ are disjoint, then $Y(B_1)$ and $Y(B_2)$ are independent.[3]

Remark: It is not necessary to assume that the scattering system is standard in order to define the notion of a beam. We make this assumption for two reasons: (1) Most physically interesting scattering systems including Rutherford scattering systems (see Sec. 5) are standard; (2) the definition of a beam when the impact set is not P_v or $P_v - \{0\}$ is more complicated to state, and thus the essential idea of a beam may be obscured by the added complexity.

Interpretation of Y: Observe first of all that $\mathbb{A}_v^{\text{in}}$ and $\mathbf{I}_v$ can be canonically identified. Since $Y(\cdot)$ will be an atomic measure on $\mathbf{I}_v$ with probability one, supp $[Y(\cdot)]$ will be a subset of $\mathbf{I}_v \cong \mathbb{A}_v^{\text{in}}$ which is interpreted as the set of incoming asymptotes of the particles in the beam. In particular, if B is a measurable subset of $\mathbf{I}_v$, $Y(B)$ tells us the number of particles in the beam which have impact parameters in B.

We are now in a position to define the scattered beam of direction. Let S^2 be the 2-sphere in $\mathbb{R}^3$ and let Ω be the canonical measure or S^2. Define $d_{\text{out}}: \mathbb{A}^{\text{out}} \to S^2$ by

$$d_{\text{out}}(X_{\text{out}}) = v_{\text{out}}/\|v_{\text{out}}\|,$$

where $X_{\text{out}}(t) = x_{\text{out}} + v_{\text{out}}t$. Thus $d_{\text{out}}(X_{\text{out}})$ is the *outgoing direction* of X_{out}.

Definition: Let $Y_{v,I} = Y$ be an incoming beam of particles with velocity v and intensity I. Let $S_v = d_{\text{out}} \circ S|\mathbb{A}_v^{\text{in}}$. The Poisson process N_v on S^2 is called the *scattered beam of directions* where N_v is defined by

$$N_v(\Omega) \equiv Y[S_v^{-1}(\Omega)]$$

for Ω a measurable subset of S^2. N_v is a Poisson process on S^2 with measure $I \cdot S_v(\lambda)$ where $S_v(\lambda)$ is defined by

$$S_v(\lambda)[\Omega] = \lambda[S_v^{-1}(\Omega)].$$

$N_v(\Omega)$ is interpreted as the number of particles in the incoming beam Y which are scattered (by S) in the direction Ω. The *cross section* (*measure*) of S w.r.t. v is defined to be $S_v(\lambda) \equiv \sigma(v, \cdot)$. Thus $\sigma(v, \Omega)$, $\Omega \subseteq S^2$, is the expected number of particles in the incoming beam Y which are scattered in the direction Ω, divided by the intensity I of the incoming beam.

If $\sigma(v, \cdot)$ is absolutely continuous w.r.t. Ω on $S^2 - \{v/\|v\|\}$, then the Radon–Nikodym derivative of $\sigma(v, \cdot)$ w.r.t. Ω, denoted by $d\sigma(V, \cdot)/d\Omega$, is the *differential cross section of S w.r.t. v*. The integral

$$\sigma(v) \equiv \int_{S^2} \frac{d\sigma(v, \cdot)}{d\Omega}\, d\Omega$$

is called the *total cross section of S w.r.t. v*. It may be infinite.

In most applications we also have spherical symmetry about the scattering center (which is O in our case.) We shall indicate how this can be incorporated in our framework. Let $SO(3)$ denote the connected component of the orthogonal group for $\mathbb{R}^3$. We have a natural action of $SO(3)$ on the set of affine mappings of $\mathbb{R}$ into R^3. In fact, if $R \in SO(3)$ and

$$X: \mathbb{R} \to \mathbb{R}^3: t \rightsquigarrow x + vt,$$

then

$$RX: R \to \mathbb{R}^3: t \rightsquigarrow Rx + (Rv)t.$$

We say the scattering system $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ is spherically symmetric (about the scattering center) if $\mathbb{A}^{\text{in}}$ and $\mathbb{A}^{\text{out}}$ are left invariant by $SO(3)$ and

$$RS = SR$$

for all $R \in SO(3)$.

Theorem: If $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ is a standard spherically symmetric scattering system, then

$$R\sigma(v, \cdot) = \sigma(Rv, \cdot) \tag{3.1}$$

for all $R \in SO(\epsilon)$, where $R\mu$, μ a measure as S^2, is defined by

$$R\mu(\Sigma) \equiv (R^{-1}\Sigma)$$

for Σ a measurable subset of S^2. In particular, if $Rv = v$, then

$$R\sigma(v, \cdot) = \sigma(v, \cdot),$$

i.e., σ is invariant under the action of such R.

Proof: Let Σ be a measurable subset of S^2. Then

$$\begin{aligned} R\sigma(v, \Sigma) &= \sigma(v, R^{-1}\Sigma) = \lambda(S_v^{-1}(R^{-1}\Sigma)) \\ &= \lambda\{(RS_v)^{-1}(\Sigma)\} = \lambda(S_{Rv}^{-1}(\Sigma)) \\ &= \sigma(Rv, \Sigma). \end{aligned}$$

Since Σ was arbitrary, this implies (3.1).

Corollary: Let $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ be as in the above theorem and assume that the differential cross section exists. Then $\sigma(v)$ depends only on $\|v\|$ and thus only on the "energy" $E = \frac{1}{2}\|v\|^2$. In this case we write $\sigma(E)$. (We are assuming mass has been normalized to be one.)

Proof: The proof follows directly from (3.1) and the fact that Ω is invariant under the action of $SO(3)$.

Remark: If we choose spherical coordinates (θ, φ), $-\pi < \varphi < \pi$, $0 \leq \theta \leq \pi$, on S^2 such that $(0, 0)$ is the *forward direction* $v/\|v\|$, then the coordinatized differential cross section $d\sigma(v; \theta, \varphi)/d\Omega$ in the *spherically symmetric case* will depend only on $E = \frac{1}{2}\|v\|^2$ and θ. We thus write $d\sigma(E; \theta)/d\Omega$.

Note that the spherical coordinates we choose on S^2 depend on the forward direction and thus the function $d\sigma(E, \theta)/d\Omega$ only tells us how much the incoming particles were deflected from its initial path by the scatterer.

4. AN APPROXIMATION THEOREM

Let $(\mathbb{A}^{\text{in}}, \mathbb{A}^{\text{out}}, S)$ be a standard scattering system and let v ($\neq 0$) R^3 be fixed. Let $Y_{v,I}$ be an incoming beam of particles. Let $\mathbb{D}_n$ be a nested sequence of disks in $\mathbf{I}_v$ such that

$$\lim_{n\to\infty} \frac{n}{\lambda(\mathbb{D}_n)} \equiv \lim_{n\to\infty} I_n = I.$$

Observe that $\bigcup \mathbb{D}_n = \mathbf{I}_v$. We can associate a *finite beam of incoming particles* with each $\mathbb{D}_n$ in the following way. Let $Y_n^1, \cdots, Y_n^n$ be uniformly distributed random variables with values in $\mathbb{D}_n$ which, if you recall, can be thought of as a subset of $\mathbb{A}_v^{\text{in}}$. It is clear that the collection $\{Y_n^1, \cdots, Y_n^n\}$ may be thought of as a finite incoming beam with impact parameters in $\mathbb{D}_n$ and intensity I_n. We also have a *finite scattered beam of directions* associated with $\mathbb{D}_n$. In particular, let Σ be a measurable subset of S^2 and let

$$N_n(v, \Sigma) = \sum_{j=1}^{n} \chi_\Sigma(S_v(Y_n^j)),$$

where χ_Σ is the indicator of Σ. Thus $N_n(v, \Sigma)$ is the number of particles in the incoming beam $\{Y_n^1, \cdots, Y_n^n\}$ which are scattered in the direction Σ. Let $\sigma_n(v, \Sigma) \equiv I_n^{-1}E(N_n(v, \Sigma))$, where $E(\ )$ denotes the expectation functional.

Theorem: For each measurable $\Sigma \subseteq S^2$, $N_n(v, \Sigma)$ converges in distribution to $N(v, \Sigma)$, where $N(v, \Sigma)$ is defined in Sec. 3. Moreover,

$$\lim_{n\to\infty} \sigma_n(v, \Sigma) = \sigma(v, \Sigma).$$

Thus, for sufficiently large n, $(\{Y_n^1, \cdots, Y_n^n\}, N_n(v, \cdot), \sigma_n(v, \cdot))$ is a "good" approximation to

$$(Y_{v,n}, N(v, \cdot), \sigma(v, \cdot)).$$

Remark: Note that $N(v, \Sigma)$ may be infinite, but that does not matter.

Proof of the theorem: Observe that $N_n(v, \Sigma)$ is the sum of independent Bernoulli random variables with

$$P_n \equiv E\{\chi_\Sigma(S_v(Y_n^j))\} = \frac{\lambda(S_v^{-1}(\Sigma) \cap \mathbb{D}_n)}{\lambda(\mathbb{D}_n)}.$$

Thus

$$\lim_{n\to\infty} nP_n = I \cdot \lambda(S_v^{-1}(\Sigma)) = I \cdot \sigma(v, \Sigma), \qquad (4.1)$$

which may be infinite. Since $nP_n = I_n\sigma_n(v, \Sigma)$, we see that

$$\lim_{n\to\infty} \sigma_n(v, \Sigma) = \sigma(v, \Sigma).$$

From the Poisson approximation to the sum of Bernoulli random variables,[4] we conclude that $N_n(v, \Sigma)$ converges to a Poisson random variable with mean $I \cdot \sigma(v, \Sigma)$, i.e., $N_n(v, \Sigma)$ converges in distribution to $N(v, \Sigma)$. This completes the proof of the theorem.

In a realistic situation we actually would have a finite beam with a large number of particles n. Thus, looking at the above theorem in another way, we see that $\{Y_{v,I}, N(v, \cdot), \sigma(v, \cdot)\}$ is a good approximation to $(\{Y_n^1, \cdots, Y_n^n\}, N_n(v, \cdot), \sigma_n(v, \cdot))$, and, since $\{Y_{v,I}, N(v, \cdot), \sigma(v, \cdot)\}$ is analytically more tractable than $(\{Y_n^1, \cdots, Y_n^n\}, N_n(v, \cdot), \sigma_n(v, \cdot))$, we use it as the theoretical model which is compared with experiment.

5. RUTHERFORD SCATTERING[5]

Consider the dynamical system governed by the following differential equation in $\mathbb{R}^3$:

$$m\frac{d^2x}{dt^2} = -\alpha \operatorname{grad}(V),$$

where $V = \|x\|^{-1} \equiv r^{-1}$, $m > 0$, and $\alpha \in \mathbb{R}$. For example, $\alpha = Z_1Z_2e^2$, and thus αr^{-1} is the Coulomb potential when the particles being scattered have finite point mass m and charge Z_1e and the "scatterer," located at O, has infinite point mass and point charge Z_2e. When Z_1 and Z_2 have opposite sign, i.e., when the two particles attract one another, $\mathbb{A}^{\text{in}} = \mathbb{A}^{\text{out}} =$ the set of nonconstant affine mappings from $\mathbb{R}$ to $\mathbb{R}^3$ which do not pass through O. When Z_1 and Z_2 have the same sign, i.e., when the two particles repulse one another, $\mathbb{A}^{\text{in}} = \mathbb{A}^{\text{out}} =$ the set of all nonconstant affine mappings from $\mathbb{R}$ to $\mathbb{R}^3$.

To be explicit, we shall take $\alpha > 0$. Choose $v\ (\neq 0) \in \mathbb{R}^3$. In the impact plane $\mathbf{I}_v = P_v$, choose polar coordinates (b, φ) with the origin of the coordinate system being O. The choice of polar axis is immaterial since we have rotational symmetry in $\mathbf{I}_v$. The *impact parameter* of an incoming asymptote, $X: \mathbb{R} \to \mathbb{R}^3: t \rightsquigarrow x_0 + vt \in \mathbb{A}_v^{\text{in}}$, is just the radial coordinate of x_0 in $\mathbf{I}_v$. Let (θ, φ), $0 \leq \theta \leq \pi$, $-\pi \leq \varphi \leq \pi$, be the usual spherical coordinates in S^2 such that the *forward direction* $v/\|v\|$ has coordinates $(0, 0)$ and the direction of the polar axis in $\mathbf{I}_v$ has coordinates $(\frac{1}{2}\pi, 0)$. θ is called the scattering angle.

In Newton,[1] Chap. 5, it is essentially shown that if $b > 0$ is the impact parameter of $X \in \mathbb{A}_v^{\text{in}}$, then

$$S_v(X) = (0, \varphi), \qquad (5.1)$$

where θ, $0 < \theta \leq \pi$, is the solution of

$$b = (\alpha/m\,\|v\|^2)\cot(\tfrac{1}{2}\theta). \qquad (5.2)$$

Thus, for Σ a measurable subset of S^2 contained in the complement of the forward direction, we have that

$$\sigma(v, \Sigma) = \lambda(S_v^{-1}(\Omega)) = \int_{S_v^{-1}(\Omega)} b\,db\,d\varphi.$$

Making the following change of coordinates,

$$b = (\alpha/m\,\|v\|^2)\cot(\tfrac{1}{2}\theta), \quad \varphi = \varphi,$$

and letting $\Sigma = \{(\theta, \varphi): 0 < \theta < \pi, (\theta, \varphi) \in \Sigma\}$, we obtain

$$\begin{aligned}
\sigma(v, \Sigma) &= \int_{S_v^{-1}(\Sigma)} b\,db\,d\varphi \\
&= \int_{\Sigma'} \frac{1}{2}\left(\frac{\alpha}{m\,\|v\|^2}\right)^2 \frac{\cos(\frac{1}{2}\theta)}{\sin(\frac{1}{2}\theta)} \frac{1}{2} \frac{1}{\sin(\frac{1}{2}\theta)^2}\,d\theta\,d\varphi \\
&= \int_{\Sigma'} \frac{1}{4}\left(\frac{\alpha}{m\,\|v\|^2}\right)^2 \frac{\sin\theta}{[\sin(\frac{1}{2}\theta)]^4}\,d\theta\,d\varphi \\
&= \int_{\Sigma} \frac{1}{4}\left(\frac{\alpha}{m\,\|v\|^2}\right)^2 \frac{1}{[\sin(\frac{1}{2}\theta)]^4}\,d\Omega.
\end{aligned}$$

The Radon–Nikodym derivative of $\sigma(v, \cdot)$ w.r.t. Ω, i.e., the differential cross section of S w.r.t. v, is

$$\frac{d\sigma}{d\Omega}(v; \theta, \varphi) = \frac{1}{4}\left(\frac{\alpha}{m\,\|v\|^2}\right)^2 \left(\frac{1}{\sin(\frac{1}{2}\theta)}\right)^4.$$

This same formula holds when $\alpha < 0$, but (5.1) has to be modified slightly in the φ coordinate and the sign on the left-hand side of (5.2) must be changed.

ACKNOWLEDGMENT

I would like to thank Professor T. Liggett for pointing out the existence and properties of complete random measures.

* This work was supported by GP-6727.

[1] G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[2] The lecture notes of W. Hunziker, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. XA, have influenced us in this section and in Sec. 3.

[3] For a discussion of such processes, see J. F. C. Kingman, Pacific J. Math. **21**, 59 (1967).

[4] W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1965), 2nd ed., p. 142.

[5] We are adopting some of Chap. 5 in Newton (see Ref. 1) for our purposes.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

New Approach to the Noether Theorems

BENJAMIN F. PLYBON
Department of Mathematics, Miami University, Oxford, Ohio 45056

(Received 7 July 1970)

The Noether theorems were derived by Noether for n-dimensional Euclidean spaces, but they have been used by many writers in relativistic theories where the geometry is not Euclidean. We give a derivation of the Noether theorems, assuming only a Riemannian space and following the method used by Noether as closely as possible. This requires new definitions of total variations for fields and integrals since a covariant total variation for tensor fields is required. The results have been applied to electromagnetic fields.

1. INTRODUCTION

In 1918, Noether[1] published two theorems on invariance of the integral in a variational problem in which she showed that invariance with respect to a group of transformations implies existence of certain identities. Since that time the Noether theorems have been applied to various physical systems to study the correspondence between symmetries of the system and existence of conservation laws.[2] The Noether theorems were derived for Euclidean spaces, and so, strictly speaking, they should only be used for problems set in Euclidean spaces. Despite this fact, many authors use the Noether theorems in relativistic theories where the geometry is not Euclidean. Arguments intended to justify such applications have been given by Anderson,[3] Trautman,[4] and others, but these arguments do not produce a true mathematical extension of the Noether theorems to Riemannian spaces in the opinion of this author.

In this paper it is shown that such an extension can be accomplished. To do so, one must carefully distinguish between point transformations and coordinate transformations.[5] In connection with this, we find that the role of the metric tensor must be clarified. Finally, this attempt to parallel the original derivation of the Noether theorems requires some new ideas about variations of vector or tensor fields. This suggests new possibilities for variational principles in relativistic theories. Some results on variational principles will be given in a later paper.

2. REVIEW OF THE NOETHER THEOREMS

In the following discussion we shall consider an integral

$$W[u] = \int_\Omega dx L(x^\alpha, u^i, \partial_\alpha u^i) \tag{2.1}$$

over a region Ω in R^4. Greek indices $\alpha, \beta, \gamma, \cdots$ have the range 1, 2, 3, 4, and Latin indices $i, j, k, \cdots$ have the range $1, \cdots, n$. Functions $u^i(x^\alpha)$ are real valued with continuous second derivations.

Infinitesimal transformations

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha, \tag{2.2}$$

$$\bar{u}^i(\bar{x}^\alpha) = u^i(x^\alpha) + \Delta u^i \tag{2.3}$$

are introduced. The total variation Δu^i is related to local variation

$$\delta u^i = \bar{u}^i(x^\alpha) - u^i(x^\alpha) \tag{2.4}$$

by the equation

$$\Delta u^i = \delta u^i + \partial_\alpha u^i \Delta x^\alpha. \tag{2.5}$$

The usual summation convention is employed in (2.5).

The total variation of integral (2.1) is defined by

$$\Delta W = \int_{\bar{\Omega}} d\bar{x} L\left(\bar{x}^\alpha, \bar{u}^i, \frac{\partial \bar{u}^i}{\partial \bar{x}^\alpha}\right) - \int_\Omega dx L\left(x^\alpha, u^i, \frac{\partial u^i}{\partial x^\alpha}\right). \tag{2.6}$$

This same formula holds when $\alpha < 0$, but (5.1) has to be modified slightly in the φ coordinate and the sign on the left-hand side of (5.2) must be changed.

ACKNOWLEDGMENT

I would like to thank Professor T. Liggett for pointing out the existence and properties of complete random measures.

* This work was supported by GP-6727.

[1] G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[2] The lecture notes of W. Hunziker, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. XA, have influenced us in this section and in Sec. 3.

[3] For a discussion of such processes, see J. F. C. Kingman, Pacific J. Math. **21**, 59 (1967).

[4] W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1965), 2nd ed., p. 142.

[5] We are adopting some of Chap. 5 in Newton (see Ref. 1) for our purposes.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

New Approach to the Noether Theorems

BENJAMIN F. PLYBON
Department of Mathematics, Miami University, Oxford, Ohio 45056

(Received 7 July 1970)

The Noether theorems were derived by Noether for n-dimensional Euclidean spaces, but they have been used by many writers in relativistic theories where the geometry is not Euclidean. We give a derivation of the Noether theorems, assuming only a Riemannian space and following the method used by Noether as closely as possible. This requires new definitions of total variations for fields and integrals since a covariant total variation for tensor fields is required. The results have been applied to electromagnetic fields.

1. INTRODUCTION

In 1918, Noether[1] published two theorems on invariance of the integral in a variational problem in which she showed that invariance with respect to a group of transformations implies existence of certain identities. Since that time the Noether theorems have been applied to various physical systems to study the correspondence between symmetries of the system and existence of conservation laws.[2] The Noether theorems were derived for Euclidean spaces, and so, strictly speaking, they should only be used for problems set in Euclidean spaces. Despite this fact, many authors use the Noether theorems in relativistic theories where the geometry is not Euclidean. Arguments intended to justify such applications have been given by Anderson,[3] Trautman,[4] and others, but these arguments do not produce a true mathematical extension of the Noether theorems to Riemannian spaces in the opinion of this author.

In this paper it is shown that such an extension can be accomplished. To do so, one must carefully distinguish between point transformations and coordinate transformations.[5] In connection with this, we find that the role of the metric tensor must be clarified. Finally, this attempt to parallel the original derivation of the Noether theorems requires some new ideas about variations of vector or tensor fields. This suggests new possibilities for variational principles in relativistic theories. Some results on variational principles will be given in a later paper.

2. REVIEW OF THE NOETHER THEOREMS

In the following discussion we shall consider an integral

$$W[u] = \int_\Omega dx L(x^\alpha, u^i, \partial_\alpha u^i) \tag{2.1}$$

over a region Ω in R^4. Greek indices $\alpha, \beta, \gamma, \cdots$ have the range 1, 2, 3, 4, and Latin indices $i, j, k, \cdots$ have the range $1, \cdots, n$. Functions $u^i(x^\alpha)$ are real valued with continuous second derivations.

Infinitesimal transformations

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha, \tag{2.2}$$

$$\bar{u}^i(\bar{x}^\alpha) = u^i(x^\alpha) + \Delta u^i \tag{2.3}$$

are introduced. The total variation Δu^i is related to local variation

$$\delta u^i = \bar{u}^i(x^\alpha) - u^i(x^\alpha) \tag{2.4}$$

by the equation

$$\Delta u^i = \delta u^i + \partial_\alpha u^i \Delta x^\alpha. \tag{2.5}$$

The usual summation convention is employed in (2.5).

The total variation of integral (2.1) is defined by

$$\Delta W = \int_{\bar{\Omega}} d\bar{x} L\left(\bar{x}^\alpha, \bar{u}^i, \frac{\partial \bar{u}^i}{\partial \bar{x}^\alpha}\right) - \int_\Omega dx L\left(x^\alpha, u^i, \frac{\partial u^i}{\partial x^\alpha}\right). \tag{2.6}$$

It follows from (2.6), by use of (2.2), (2.3), and (2.5), that

$$\Delta W = \int_\Omega dx \left[\frac{\delta L}{\delta u^i} \delta u^i + \frac{\partial}{\partial x^\alpha} \left(L \Delta x^\alpha + \frac{\partial L}{\partial(\partial_\alpha u^i)} \delta u^i \right) \right], \tag{2.7}$$

where

$$\frac{\delta L}{\delta u^i} = \frac{\partial L}{\partial u^i} - \frac{\partial}{\partial x^\alpha} \left(\frac{\partial L}{\partial(\partial_\alpha u^i)} \right). \tag{2.8}$$

If we define

$$T^\alpha_{\cdot\beta} = L \delta^\alpha_\beta - \frac{\partial L}{\partial(\partial_\alpha u^i)} \partial_\beta u^i, \tag{2.9}$$

then (2.7) can be put in the form

$$\Delta W = \int_\Omega dx \frac{\delta L}{\delta u^i} (\Delta u^i - \partial_\beta u^i \Delta x^\beta) + \int_\Omega dx \partial_\alpha \left(T^\alpha_{\cdot\beta} \Delta x^\beta + \frac{\partial L}{\partial(\partial_\alpha u^i)} \Delta u^i \right). \tag{2.10}$$

The Noether theorems follow from (2.10).

Let indices $A, B, C, \cdots$ have the range $1, \cdots, N$ in the following theorems:

Theorem 1: Suppose functions $\xi^\alpha_A(x^\alpha, u^i, \partial_\alpha u^i)$, $\omega^\kappa_A(x^\alpha, u^i, \partial_\alpha u^i)$, and $\gamma^\alpha_A(x^\alpha, u^i, \partial_\alpha u^i)$ exist such that for arbitrary parameters $\epsilon^1, \cdots, \epsilon^N$ we have

$$\Delta W = \int_\Omega dx \partial_\alpha (\epsilon^A \gamma^\alpha_A), \tag{2.11}$$

when $\Delta x^\alpha = \epsilon^A \xi^\alpha_A$ and $\Delta u^i = \epsilon^A \omega^i_A$. Then N relations exist of the form

$$(\xi^\alpha_A \partial_\alpha u^i - \omega^i_A) \frac{\delta L}{\delta u^i} = \partial_\alpha \left(T^\alpha_{\cdot\beta} \xi^\beta_A + \frac{\partial L}{\partial(\partial_\alpha u^i)} \omega^i_A - \gamma^\alpha_A \right). \tag{2.12}$$

Theorem 2: Suppose linear operators Λ^α_A, Υ^α_A, and Θ^i_A exist such that

$$\Delta W = \int_\Omega dx \partial_\alpha (\Lambda^\alpha_A [\Delta \phi^A]) \tag{2.13}$$

for arbitrary functions $\Delta\phi^A$, which vanish on the boundary of Ω, when

$$\Delta x^\alpha = \Upsilon^\alpha_A [\Delta \phi^A] \quad \text{and} \quad \Delta u^i = \Theta^i_A [\Delta \phi^A].$$

Then N relations exist of the form

$$\Theta^{*i}_A \left[\frac{\delta L}{\delta u^i} \right] - \Upsilon^{*\alpha}_A \left[\partial_\alpha u^k \frac{\delta L}{\delta u^k} \right] = 0. \tag{2.14}$$

Operators Θ^{*i}_A and $\Upsilon^{*\alpha}_A$ are adjoints of operators Θ^i_A and Υ^α_A, respectively.

The Noether theorems and the variational formula (2.7) will now be extended to Riemannian spaces to produce covariant forms of (2.12) and (2.14). The purpose of the following derivation is to show that a covariant form of (2.7) can be produced by arguments very similar to those used in producing (2.7), with some interesting changes in the concept of a total variation.

3. COVARIANT DIFFERENTIALS AND VARIATIONS

We shall assume X is a 4-dimensional Riemannian manifold with points $x = (x^0, x^1, x^2, x^3) = (x^\alpha)$ and metric $g_{\alpha\beta}$. Functions $\Gamma^\nu_{\lambda\rho}$ are the Christoffel symbols corresponding to metric $g_{\alpha\beta}$. At each point $x \in X$ we have a tangent space with basis $\{\mathbf{I}_\alpha(x)\}$. Basis systems at (x^α) and $(x^\alpha + dx^\alpha)$ are related by

$$\mathbf{I}_\lambda(x^\alpha + dx^\alpha) = \mathbf{I}_\lambda(x^\alpha) + \Gamma^\nu_{\lambda\rho}(x^\alpha)\, dx^\rho \mathbf{I}_\nu(x^\alpha) + o(dx^\alpha)\mathbf{I}_\lambda(x^\alpha).$$

To first-order terms in dx^α, we have

$$\mathbf{I}_\lambda(x^\alpha + dx^\alpha) = \mathbf{I}_\lambda(x^\alpha) + \Gamma^\nu_{\lambda\rho}(x^\alpha)\, dx^\rho \mathbf{I}_\nu(x^\alpha). \tag{3.1}$$

In the following calculations only first-order terms in differentials or variations will be retained. Assume

$$g = |\det(g_{\alpha\beta})| \neq 0 \quad \text{on } X.$$

Each point $x \in X$ has coordinates (x^α) in some coordinate system. No coordinate transformations will be introduced in the following. We shall employ point transformations such as (2.2) above, but such transformations must not be confused with coordinate transformations. This distinction is important if the variation Δx^α to be used is to be correctly interpreted in the following.

Let $\mathbf{A}$ be a vector field on X. Then

$$\mathbf{A}(x) = A^\alpha(x)\mathbf{I}_\alpha(x),$$

and the covariant derivative of $\mathbf{A}$ is given by

$$\nabla_\beta A^\alpha = \partial_\beta A^\alpha + \Gamma^\alpha_{\beta\lambda} A^\lambda.$$

It is important to observe that

$$\mathbf{A}(x + \Delta x) - \mathbf{A}(x) = \nabla_\beta A^\alpha \Delta x^\beta \mathbf{I}_\alpha(x). \tag{3.2}$$

Hence, differentials $\nabla A^\alpha = \nabla_\beta A^\alpha \Delta x^\beta$ represent components of a *vector increment*

$$\Delta \mathbf{A} = \mathbf{A}(x + \Delta x) - \mathbf{A}(x).$$

The differentials $dA^\lambda = \partial_\beta A^\lambda\, dx^\beta$ represent changes in component functions A^α as we move from point (x^α) to point $(x^\alpha + dx^\alpha)$. These two kinds of differentials are related by

$$\nabla A^\alpha = dA^\alpha + \Gamma^\alpha_{\beta\lambda} A^\lambda\, dx^\beta.$$

Hence $\nabla A^\alpha = dA^\alpha$ in flat spaces, but these differentials are generally different.

Assume that **A** is a vector field on X and infinitesimal transformations

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha, \tag{3.3}$$

$$\bar{A}^\alpha(\bar{x}) = A^\alpha(x) + \Delta(A^\alpha) \tag{3.4}$$

are given. Equation (3.3) defines a point transformation in X. We shall interpret (3.4) as introducing a new vector field **B** on X, where

$$\mathbf{B}(\bar{x}) = \bar{A}^\alpha(\bar{x})\mathbf{I}_\alpha(\bar{x}). \tag{3.5}$$

Then

$$\mathbf{B}(x) - \mathbf{A}(x) = \delta\mathbf{A}(x) \tag{3.6}$$

is a *vector* local variation. We may express (3.6) in terms of components as

$$\bar{A}^\alpha(x) - A^\alpha(x) = \delta A^\alpha(x)$$

if δA^α is defined by $\delta\mathbf{A}(x) = \delta A^\alpha(x)\mathbf{I}_\alpha(x)$. Hence, δA^α may be interpreted either as local variations of components A^α or as components of the vector variation $\delta\mathbf{A}$.

We now define a vector total variation by

$$\Delta\mathbf{A} = \mathbf{B}(x + \Delta x) - \mathbf{A}(x).$$

Adding and subtracting $\mathbf{B}(x)$, we find

$$\Delta\mathbf{A} = \nabla B^\lambda \mathbf{I}_\lambda(x) + \delta A^\alpha \mathbf{I}_\alpha(x).$$

Up to first-order terms in δA^α and Δx^α, we have $\nabla B^\lambda = \nabla A^\lambda$ so

$$\Delta\mathbf{A} = (\nabla A^\lambda + \delta A^\lambda)\mathbf{I}_\lambda(x).$$

Define components of $\Delta\mathbf{A}$ by $\Delta\mathbf{A} = (\Delta A)^\lambda \mathbf{I}_\lambda(x)$. Then

$$(\Delta A)^\lambda = \delta A^\lambda + \nabla_\alpha A^\lambda \Delta x^\alpha. \tag{3.7}$$

For $\Delta(A^\lambda)$ as defined by (3.4) we have

$$\Delta(A^\lambda) = \delta A^\lambda + \partial_\alpha A^\lambda \Delta x^\alpha. \tag{3.8}$$

Hence $(\Delta A)^\lambda = \Delta(A^\lambda)$ in flat spaces, but in general these total variations are different. The total variations $\Delta(A^\lambda)$ are not components of a vector, and so they are not covariant quantities. The variations $(\Delta A)^\lambda$ are covariant, however.

Similarly, for higher-order tensors, we define total variations by rules such as

$$(\Delta A)_{\mu\nu} = \delta A_{\mu\nu} + \nabla_\lambda A_{\mu\nu} \Delta x^\lambda$$

when variation $\bar{A}_{\mu\nu}(\bar{x}) = A_{\mu\nu}(x) + \Delta(A_{\mu\nu})$ is given.

4. VARIATION OF AN INTEGRAL

Assume **A** is a vector field on X and $L(x, \mathbf{A}, \nabla\mathbf{A})$ is a scalar function of point x, vector **A**, and covariant derivative $\nabla\mathbf{A}$. Integral

$$W[\mathbf{A}] = \int_\Omega dx g^{\frac{1}{2}} L(x, \mathbf{A}, \nabla\mathbf{A}) \tag{4.1}$$

will be considered. Assume that infinitesimal transformations (3.3) and (3.4) have been given and variations of **A** and $\nabla\mathbf{A}$ have been defined as in Sec. 3. We define the total variation of W by

$$\Delta W = \int_{\bar{\Omega}} d\bar{x}[g(\bar{x})]^{\frac{1}{2}} L(\bar{x}, \mathbf{B}, \nabla\mathbf{B}) - \int_\Omega dx[g(x)]^{\frac{1}{2}} L(x, \mathbf{A}, \nabla\mathbf{A}). \tag{4.2}$$

It is *assumed* that $\delta g_{\alpha\beta} = 0$. We shall now derive a formula for ΔW to replace (2.7).

Assume $g > 0$ on X for convenience. Then

$$[g(\bar{x})]^{\frac{1}{2}} = [g(x)]^{\frac{1}{2}} + \tfrac{1}{2} g^{-\frac{1}{2}} \frac{\partial g}{\partial x^\lambda} \Delta x^\lambda. \tag{4.3}$$

Using $\partial g/\partial x^\lambda = g g^{\alpha\beta} \partial_\lambda g_{\alpha\beta}$, we find

$$[g(\bar{x})]^{\frac{1}{2}} = g^{\frac{1}{2}}(1 + \tfrac{1}{2} g^{\alpha\beta} \partial_\lambda g_{\alpha\beta} \Delta x^\lambda).$$

Since $g^{\alpha\beta}\partial_\lambda g_{\alpha\beta} = 2\Gamma^\alpha_{\alpha\lambda}$, then

$$[g(\bar{x})]^{\frac{1}{2}} = g^{\frac{1}{2}}(1 + \Gamma^\alpha_{\alpha\lambda} \Delta x^\lambda). \tag{4.4}$$

Now we use $d\bar{x} = dx(1 + \partial_\alpha \Delta x^\alpha)$ and the result (4.4) in (4.2) with $L(\bar{x}, \mathbf{B}, \nabla\mathbf{B}) = L + \Delta L$. Then

$$\Delta W = \int_\Omega dx(1 + \partial_\alpha \Delta x^\alpha)(1 + \Gamma^\alpha_{\alpha\lambda} \Delta x^\lambda) g^{\frac{1}{2}}(L + \Delta L) - \int_\Omega dx g^{\frac{1}{2}} L.$$

The equation simplifies to

$$\Delta W = \int_\Omega dx g^{\frac{1}{2}}(L \nabla_\alpha \Delta x^\alpha + \Delta L). \tag{4.5}$$

Now,

$$\Delta L = L(\bar{x}, \mathbf{B}, \nabla\mathbf{B}) - L(x, \mathbf{A}, \nabla\mathbf{A}).$$

So,

$$\Delta L = \frac{\partial L}{\partial x^\alpha} \Delta x^\alpha + \frac{\partial L}{\partial A^\alpha}(\delta A^\alpha + \nabla_\beta A^\alpha \Delta x^\beta) + \frac{\partial L}{\partial(\nabla_\alpha A^\beta)}(\delta(\nabla_\alpha A^\beta) + \nabla_\lambda \nabla_\alpha A^\beta \Delta x^\lambda).$$

Thus,

$$\Delta L = \nabla_\lambda L \Delta x^\lambda + \frac{\delta L}{\delta A^\alpha} \delta A^\alpha + \nabla_\alpha\left(\frac{\partial L}{\partial(\nabla_\alpha A^\beta)} \delta A^\beta\right), \tag{4.6}$$

where

$$\nabla_\lambda L = \frac{\partial L}{\partial x^\lambda} + \frac{\partial L}{\partial A^\alpha} \nabla_\lambda A^\alpha + \frac{\partial L}{\partial(\nabla_\alpha A^\beta)} \nabla_\lambda(\nabla_\alpha A^\beta)$$

and

$$\frac{\delta L}{\delta A^\alpha} = \frac{\partial L}{\partial A^\alpha} - \nabla_\beta\left(\frac{\partial L}{\partial(\nabla_\beta A^\alpha)}\right). \tag{4.7}$$

Substitution of (4.6) in Eq. (4.5) yields, after some

rearrangement,

$$\Delta W = \int_\Omega dx g^{\frac{1}{2}} \frac{\delta L}{\delta A^\alpha} \delta A^\alpha + \int_\Omega dx g^{\frac{1}{2}} \nabla_\alpha \left(L \Delta x^\alpha + \frac{\partial L}{\partial(\nabla_\alpha A^\beta)} \delta A^\beta \right). \quad (4.8)$$

Equation (4.8) is clearly the covariant equivalent of (2.7) and could have been guessed in advance. It is interesting, however, that a derivation of (4.8) along lines similar to those leading to (2.7) requires some new ideas about variations of vector or tensor fields.

5. COVARIANT NOETHER THEOREMS

From (4.8) new forms of the Noether theorems follow easily. First observe that again we may define a tensor

$$T^\alpha_{\cdot\lambda} = L\delta^\alpha_\lambda - \frac{\partial L}{\partial(\nabla_\alpha A^\beta)} \nabla_\lambda A^\beta$$

and replace (4.8) by

$$\Delta W = \int_\Omega dx g^{\frac{1}{2}} \frac{\delta L}{\delta A^\alpha} \delta A^\alpha + \int_\Omega dx g^{\frac{1}{2}} \nabla_\alpha \left(T^\alpha_{\cdot\lambda} \Delta x^\lambda + \frac{\partial L}{\partial(\nabla_\alpha A^\beta)} (\Delta A)^\beta \right). \quad (5.1)$$

Theorem 3: Suppose $3r$ vectors $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_r$, $\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_r$, $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_r$ exist at each point $x \in X$ such that for arbitrary real numbers $\epsilon^1, \cdots, \epsilon^r$, if

$$\Delta \mathbf{x} = \epsilon^k \boldsymbol{\xi}_k \quad \text{and} \quad \Delta \mathbf{A} = \epsilon^k \boldsymbol{\eta}_k,$$

then

$$\Delta W = \int_\Omega dx g^{\frac{1}{2}} \operatorname{div}(\epsilon^k \boldsymbol{\gamma}_k).$$

Let $\boldsymbol{\xi}_k = \xi^\lambda_k \mathbf{I}_\lambda(x)$, $\boldsymbol{\eta}_k = \boldsymbol{\eta}^\alpha_k \mathbf{I}_\alpha(x)$, and $\boldsymbol{\gamma}_k = \gamma^\alpha_k \mathbf{I}_\alpha(x)$. Then r identities exist of the form

$$\frac{\delta L}{\delta A^\beta} (\nabla_\lambda A^\beta \xi^\lambda_k - \eta^\beta_k) = \nabla_\alpha \left(T^\alpha_{\cdot\lambda} \xi^\lambda_k + \frac{\partial L}{\partial(\nabla_\alpha A^\beta)} \eta^\beta_k - \gamma^\alpha_k \right). \quad (5.2)$$

Proof: The proof follows immediately from Eq. (5.1) by using the hypotheses that

$$\Delta x^\lambda = \xi^\lambda_k \epsilon^k, \quad (\Delta A)^\beta = \eta^\beta_k \epsilon^k$$

and arbitrariness of parameters ϵ^k and region Ω.

Theorem 4: Assume linear differential operators Λ^α_k, Υ^α_k, and Θ^α_k, $k = 1, \cdots, r$, exist such that for arbitrary functions $\Delta\phi^1(x), \cdots, \Delta\phi^r(x)$ we have

$$\Delta W = \int_\Omega dx g^{\frac{1}{2}} \nabla_\alpha (\Lambda^\alpha_k [\Delta\phi^k])$$

when

$$\Delta x^\alpha = \Upsilon^\alpha_k [\Delta\phi^k] \quad \text{and} \quad (\Delta A)^\alpha = \Theta^\alpha_k [\Delta\phi^k].$$

Then r identities exist of the form

$$\Theta^{*\lambda}_k \left[\frac{\delta L}{\delta A^\lambda} \right] - \Upsilon^{*\beta}_k \left[\frac{\delta L}{\delta A^\lambda} \nabla_\beta A^\lambda \right] = 0, \quad (5.3)$$

where $\Theta^{*\lambda}_k$ denotes the adjoint of Θ^λ_k.

Proof: The proof follows from (5.1) using the hypotheses given for variations Δx^α and $(\Delta A)^\alpha$.

It would be reasonable at this point to present applications of Theorems 3 and 4. One such application has been made by the author[6] to electromagnetic fields. The calculations are lengthy, and the result is essentially that of Bessel-Hagen,[2] with the expected result that the symmetry group involved depends upon the choice of metric $g_{\alpha\beta}$. The results are similar to conclusions given by Fock,[7] although his conclusions did not stem from the Noether theorems.

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor S. Drobot of the Ohio State University for his many helpful suggestions and valuable criticism of the work in this paper.

[1] E. Noether, Nachr. Ges. Wiss. Göttingen **2**, 235 (1918); E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[2] E. Bessel-Hagen, Math. Ann. **84**, 258 (1921).
[3] J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), Chap. 4, pp. 88–94.
[4] A. Trautman, F. A. E. Pirani, and H. Bondi, *Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, N.J., 1965), Vol. I, p. 172.
[5] T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 443 (1962).
[6] B. F. Plybon, *Conservation Laws for Electromagnetic Fields*, University Microfilms, 1968 (unpublished doctoral dissertation).
[7] V. Fock, *The Theory of Space, Time and Gravitation* (Macmillan, New York, 1964), Chap. 4, pp. 175–78.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

High-Temperature Expansion of the Spin-$\frac{1}{2}$ XY Model

M. HOWARD LEE*
Theoretical Physics Institute, University of Alberta, Edmonton, Canada

(Received 6 January 1969)

The high-temperature expansion of the spin-$\frac{1}{2}$ XY model is shown to be particularly simple when the Hamiltonian is written in the second-quantization form. By aid of a few simple rules, the partition function and susceptibility are easily evaluated to high orders.

I. INTRODUCTION

We present here a method particularly well suited for obtaining the critical properties of the spin-$\frac{1}{2}$ XY model by high-temperature expansions. The high temperature expansion of the XY model has been previously approached by others[1] as an anisotropic limit of the Heisenberg model. This approach entails, however, some unexpected complications, and as a result the critical properties have not been established until recently.

The method we propose is to write the Hamiltonian of the XY model in terms of the usual raising and lowering operators. The high-temperature expansion of the partition function then gives rise to graphs or clusters in the usual way. But by having the Hamiltonian in the second-quantization form, we have reduced number of graphs considerably and the trace becomes quite trivial to evaluate. Further, we can obtain the fluctuation in the long-range order and the susceptibility in the perpendicular direction directly from the partition function series.[2a]

In our first paper we considered an application of this method to the cluster expansion of the Jastrow function found in the models of liquid ^{4}He.[2b]

II. XY MODEL

The XY model is a limiting form of the anisotropic Heisenberg model and is usually written as

$$\mathcal{H} = -2J \sum_{(mn)} (S_m^x S_n^x + S_m^y S_n^y), \tag{1}$$

where by (mn) we mean that the sum is over the nearest-neighbor pairs only. The physical relevance of the XY model as a model for liquid helium near lambda point or for certain ferromagnets near the critical temperature is already discussed elsewhere.[3,4] We shall restrict our discussion here to the case of spin-$\frac{1}{2}$ only.

Consider the spin operators

$$S_j^{\pm} = S_j^x \pm iS_j^y \quad \text{and} \quad S_j^+ S_j^- = S_j^z + \tfrac{1}{2}.$$

The commutation relations for $S_j^{\pm}$ are very simple. If $S_i^{\pm}$ and $S_j^{\pm}$ are on the same site of a lattice (i.e., $i = j$), then they act as Fermi operators. If $S_i^{\pm}$ and $S_j^{\pm}$ are not on the same site of a lattice (i.e., $i \neq j$), then they act as Bose operators. In terms of these operators the Hamiltonian has the following simple form:

$$\mathcal{H} = -J \sum_{(mn)} S_m^+ S_n^-. \tag{2}$$

For the high-temperature expansion it will be found more convenient to work with another set of operators defined as

$$N_i \equiv S_i^+ S_i^- = \begin{pmatrix} 10 \\ 00 \end{pmatrix}_i, \quad N_i^\dagger \equiv S_i^- S_i^+ = \begin{pmatrix} 00 \\ 01 \end{pmatrix}_i. \tag{3}$$

Some of the more useful algebraic properties of N and $N^\dagger$ are:

(i) N and $N^\dagger$ are idempotent.
(ii) N and $N^\dagger$ are orthogonal to each other.
(iii) All the N's and $N^\dagger$'s commute.
(iv) The trace of both N and $N^\dagger$ is unity.

These four properties will be found most useful in generating the high-temperature expansion graphs of the partition function and in evaluating the successive traces.

III. PARTITION FUNCTION

We shall write the partition function in the customary way

$$\begin{aligned} Z &= \mathrm{Tr}\,[1 - \beta\mathcal{H} + \beta^2\mathcal{H}^2/2 - \cdots] \\ &= \mathrm{Tr}\left[1 + \sum_{n=1}^{\infty} K^n P^n / n!\right], \end{aligned} \tag{4}$$

where

$$K = J/K_B T \quad \text{and} \quad P = \sum_{(mn)} S_m^+ S_n^-.$$

The trace is over the spin space I,

$$I = 1 \times 1 \times 1 \times \cdots \times 1,$$

where 1 is the unit matrix of rank 2. To obtain physical quantities like specific heat, entropy, we need, however,

$$\ln Z = N \ln 2 + \sum_{n=1}^{\infty} \frac{K^n \,\mathrm{Tr}^*\, p^n}{n!}, \tag{5}$$

where[5]

$$\mathrm{Tr}^* = 2^{-N} \times O(N) \quad \text{in Tr } P^n.$$

As in the high-temperature expansions of the Ising and Heisenberg models,[6] Tr P^n will generate a variety of connected and disconnected graphs or clusters. But,unlike the Ising and Heisenberg models, the XY model will have only a subset of the total graphs. In drawing a graph, it will be convenient to assign a direction to (as if a forward momentum):

$$S_m^+ S_n^- \equiv m \rightarrow n.$$

The line between the two lattice points m and n is by definition the nearest-neighbor distance. Since $S_i^{\pm}$ are themselves traceless, all terms in Tr P^n will vanish except those which have $S_i^{\pm}, S_j^{\pm}, S_k^{\pm}, \cdots$ paired up in an *even* number of ways with $S_i^{\mp}, S_j^{\mp}, S_k^{\mp}, \cdots$ (i.e., the nonvanishing are those which can be expressed entirely by a combination of N's and $N^{\dagger}$'s). A few nonvanishing terms are graphically represented in Fig. 1. A few words about graphology: The assignment of the arrows shows that, for every arrow pointing towards a vertex, there is another pointing away from it if the total momentum is conserved at the vertex. In fact, nonzero graphs always satisfy this requirement (and once this point is understood, the arrows can be altogether suppressed). It will be then apparent from the rule of the "momentum conservation" that the only allowed graphs in the high-temperature expansion of the partition functions are polygons.[7]

We shall now show the matrix elements of a few low-order graphs. The matrix elements for Fig. 1(A) are

$$S_i^+ S_j^- \cdot S_j^+ S_i^- = N_i N_j^{\dagger}$$

whose trace (in the sense of our definition) is 2^{-2}. The matrix elements for Fig. 1(B) are

$$S_i^+ S_j^- \cdot S_j^+ S_k^- \cdot S_k^+ S_i^- = N_i N_j^{\dagger} N_k,$$

and

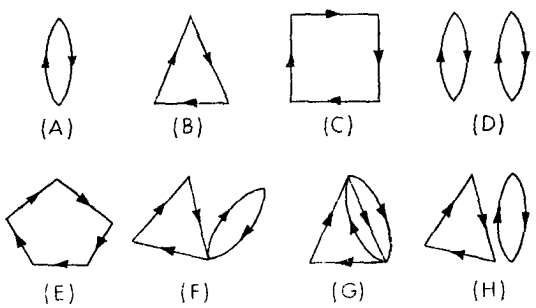


FIG. 1. Z graphs.

$$S_i^+ S_j^- \cdot S_k^+ S_i^- \cdot S_j^+ S_k^- = N_i N_j^{\dagger} N_k^{\dagger}.$$

Since ij, jk, ki are constrained to be nearest-neighbor pairs, $N_i N_j^{\dagger} N_k^{\dagger}$ and $N_i N_j^{\dagger} N_k$ must correspond to an equilateral triangle. $N_i N_j^{\dagger} N_k^{\dagger}$ and $N_i N_j N_k^{\dagger}$ both have an identical trace value of 2^{-3}. It will be seen that in all cases the trace will produce a simple factor in the form of 2^{-v},where v is the number of the vertices of a graph.

The matrix elements for Fig. 1(C) are

$$N_i N_j N_k N_1^{\dagger} + 4 N_i N_j N_k^{\dagger} N_1^{\dagger} + N_i N_j^{\dagger} N_k^{\dagger} N_1^{\dagger}.$$

The coefficients 1, 4, and 1 represent the number of ways a square can be formed by the combinations of N's and $N^{\dagger}$'s. We shall term these numbers "occurrence factors." The occurrence factors are quite evidently combinatorial factors associated with some particular combinations of N's and $N^{\dagger}$'s for a given graph.[8] For most of the graphs their occurrence factors can be very simply obtained. In Appendix A we give a few examples. Rules for drawing the graphs contributing to the partition function are thus all self-evident. We shall write them down here without proof:

(i) The nonzero terms of Tr P^n are those which correspond to polygons.

(ii) Those polygons, furthermore, must have an *even* number of lines at all the vertices.

(iii) At each vertex, momentum is conserved (for every arrow pointing towards a vertex, there must be one pointing away from it).

(iv) The trace of a graph yields a factor 2^{-v}, where v is the number of the vertices of the graph.

According to these rules, then, it is not possible to include, for example, the graph $G(11)$ in 5th order where by $G(11)$ we mean the graph number 11 listed in Table I. The same shadow graph can, however, occur at 6th order (see Table III).

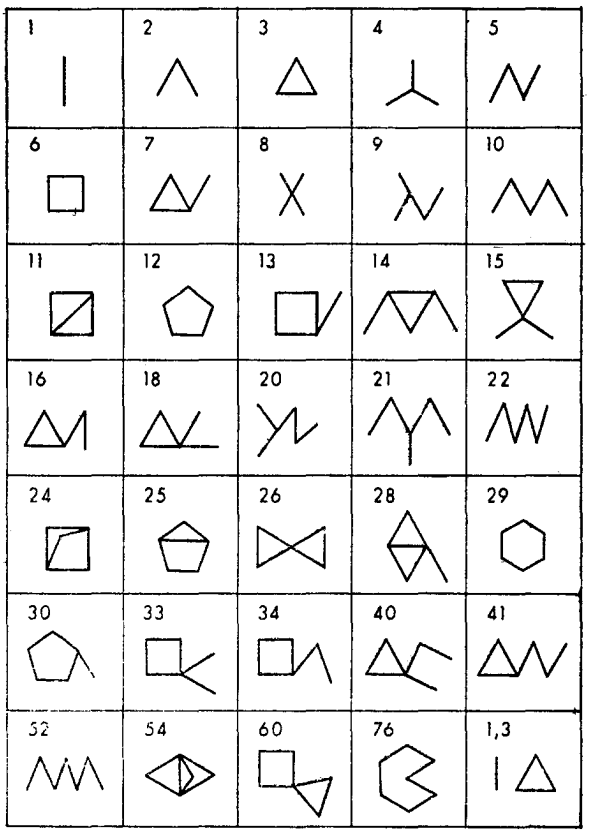
In Tr P^n, the trace and the sums contained in the P's are interchangeable. Hence we can write, for example,

$$\mathrm{Tr}^* \sum_{\Delta} N_i N_j N_k^{\dagger} = \sum_{\Delta} \mathrm{Tr}^* N_i N_j N_k^{\dagger} = 2^{-3} N q q',$$

where q is the coordination number and q' is the number of ways a third point can be placed to complete an equilateral triangle, given that there is a pair of nearest-neighbor points in the lattice ($q' = 4$ for fcc). It is customary to express the result of such sums in terms of the lattice constants of Domb and Sykes.[9] In this instance,

$$N q q' = 6 N p_3,$$

TABLE I. Shadow graphs are enumerated according to the classification of Baker *et al.* (see Ref. 9).



where p_3 is the lattice constant for triangle and the numerical factor 6 is the random-walk factor for the three-sided polygon. These numbers, which will enable us to express our results in terms of the lattice constants, are tabulated in Table II. They are independent of the lattice of a system and are quite systematic.

Finally, the partition function is

$$\ln Z - N \ln 2 = (K^2/2!)\,\mathrm{Tr}^*\,P^2 + (K^3/3!)\,\mathrm{Tr}^*\,P^3 + \cdots, \quad (6)$$

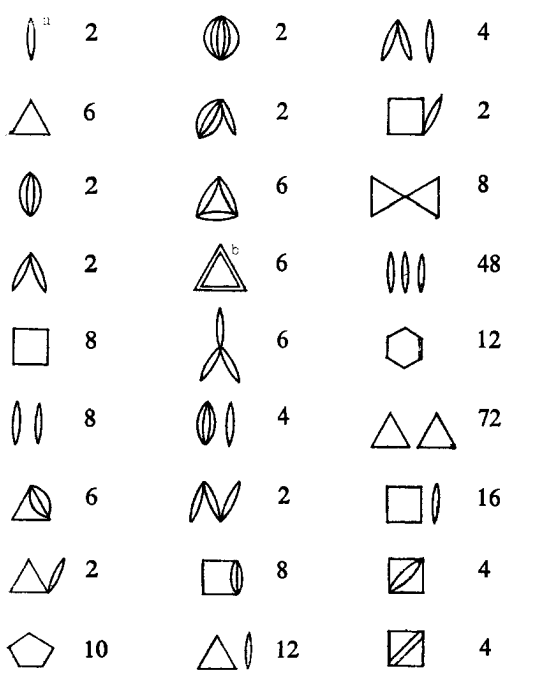
where, in Table III, terms which contribute to the above expansion are shown with the graphs[10] and the occurrence factors. It would appear that representing a graph by the combinations of N's and $N^\dagger$'s is unnecessary, since the final trace depends only on the number of the vertices of a graph. For the partition function the total sum of the occurrence factors is, in fact, sufficient. It will be shown later, however, that, for physical observables like the perpendicular susceptibility, the knowledge of the decomposition is quite necessary.

The high-temperature expansion of the partition function for the general lattice is given below:

$$\begin{aligned}
&\ln Z - N \ln 2 \\
&= (K^2/2!)[0.5\{G(1)\}] + (K^3/3!)[1.5\{G(3)\}] \\
&\quad + (K^4/4!)[0.5\{G(1)\} + \{G(2)\} + 3\{G(6)\} \\
&\quad + 1.5\{G(1,1)\}] + (K^5/5!)[7.5\{G(3)\} + 5\{G(7)\} \\
&\quad + 7.5\{G(12)\} + 7.5\{G(1,3)\}] + (K^6/6!)[0.5\{G(1)\} \\
&\quad + 3\{G(2)\} + 6\{G(3)\} + 4.5\{G(4)\} + 5.25\{G(5)\} \\
&\quad + 30\{G(6)\} + 16.5\{G(11)\} + 15\{G(13)\} \\
&\quad + 30\{G(26)\} + 22.5\{G(29)\} + 7.5\{G(1,1)\} \\
&\quad + 7.5\{G(1,2)\} + 22.4\{G(1,4)\} + 45\{G(3,3)\} \\
&\quad + 11.25\{G(1,1,1)\}] + (K^7/7!)[31.5\{G(3)\} \\
&\quad + 57.75\{G(7)\} + 77\{G(11)\} + 131.25\{G(12)\} \\
&\quad + 35\{G(14)\} + 31.5\{G(15)\} + 36.75\{G(16)\} \\
&\quad + 57.75\{G(25)\} + 52.5\{G(30)\} + 105\{G(54)\} \\
&\quad + 105\{G(60)\} + 78.75\{G(76)\} + 105\{G(1,3)\} \\
&\quad + 78.75\{G(1,12)\} + 52.5\{G(2,3)\} \\
&\quad + 157.5\{G(3,6)\} + 78.75\{G(1,1,3)\}] + \cdots, \quad (7)
\end{aligned}$$

where by $\{G(n)\}$ we mean the lattice constant of the connected graph $G(n)$ listed in Table I and by $\{G(n,m)\}$ the lattice constant of the disconnected graphs $G(n)$ and $G(m)$.

TABLE II. These numbers connect the directed graphs of the XY model and the corresponding shadow graphs.



[a] The curved lines indicate that the arrows are in the opposite directions.

[b] The parallel lines indicate that the arrows are in the same direction.

TABLE III. Occurrence factors of the XY model.

Order	Graphs	Occurrence factors
2		$NN^\dagger$
3		$NNN^\dagger + NN^\dagger N^\dagger$
4		$NN^\dagger$
		$2(NNN^\dagger + NN^\dagger N^\dagger)$
		$NNNN^\dagger + 4NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger$
		$3(NN^\dagger \cdot NN^\dagger)$
5		$5(NNN^\dagger + NN^\dagger N^\dagger)$
		$5(NNNN^\dagger + 6NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$NNNNN^\dagger + 11NNNN^\dagger N^\dagger + 11NNN^\dagger N^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger N^\dagger$
		$10(NNN^\dagger + NN^\dagger N^\dagger) \cdot NN^\dagger$
6		$NN^\dagger$
		$6(NNN^\dagger + NN^\dagger N^\dagger)$
		$3(NNN^\dagger + NN^\dagger N^\dagger)$
		$NNN^\dagger + NN^\dagger N^\dagger$
6		$2(NNNN^\dagger + 4NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$30NN^\dagger \cdot NN^\dagger$
		$3(NNNN^\dagger + 12NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$6(NNNN^\dagger + 8NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$6(NNNN^\dagger + 5NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$3(NNNN^\dagger + 6NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger)$
		$30(NNN^\dagger + NN^\dagger N^\dagger) \cdot NN^\dagger$
		$6(NNNNN^\dagger + 19NNNN^\dagger N^\dagger + 19NNN^\dagger N^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger N^\dagger)$
		$3(NNNN^\dagger + 19NNNN^\dagger N^\dagger + 19NNN^\dagger N^\dagger N^\dagger + 1NN^\dagger N^\dagger N^\dagger N^\dagger)$
		$15(NN^\dagger \cdot NN^\dagger \cdot NN^\dagger)$
		$NNNNNN^\dagger + 26NNNNN^\dagger N^\dagger + 66NNNN^\dagger N^\dagger N^\dagger + 26NNN^\dagger N^\dagger N^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger N^\dagger N^\dagger$
		$10(NNN^\dagger + NN^\dagger N^\dagger) \cdot (NNN^\dagger + NN^\dagger N^\dagger)$
		$15(NNNN^\dagger + 4NNN^\dagger N^\dagger + NN^\dagger N^\dagger N^\dagger) \cdot NN^\dagger$

IV. FLUCTUATION

For the Ising and Heisenberg models one usually computes the parallel susceptibility, from which the critical temperature is precisely estimated. In the XY model, as in the Ising model,[11] one has the possibility of calculating the susceptibility both in the parallel and perpendicular direction. It may be suspect that in the perpendicular directions, where the spins are only weakly correlated, the susceptibility may not even be singular at the critical temperature. It will be, thus, more advantageous from the point of high-temperature expansions to study the susceptibility in the parallel direction.

The susceptibility in the parallel direction is

$$\chi_{\parallel} = K_B T \frac{\partial^2}{\partial H^2} \ln Z_{\parallel}|_{H=0}, \tag{8}$$

where

$$Z_{\parallel} = \mathrm{Tr} \exp -\beta\left(\mathcal{H} - H \sum_i S_i^x\right).$$

But since the interaction term and the Zeeman term do not commute, it is very difficult to evaluate the high-temperature expansion of the above equation. There is, however, another physical quantity, the fluctuation in the long-range order, which overcomes this difficulty and thus can be more easily calculated. When both terms in the Hamiltonian commute, the fluctuation is identical to the susceptibility.[12]

The fluctuation Y may be defined as

$$Y = \sum_{r=1}^{N} \langle S_0^x S_r^x \rangle = \tfrac{1}{4} + \sum_{r \neq 0}^{N} \langle S_0^x S_r^x \rangle$$
$$= \frac{1}{4} + \frac{1}{N} \sum_{rt}' \langle S_r^x S_t^x \rangle, \qquad (9)$$

where by the prime on the sum we exclude the diagonal terms $r = t$ and

$$\langle AB \rangle \equiv Z^{-1} \operatorname{Tr} AB \exp(-\beta \mathcal{H}).$$

In terms of the raising and lowering operators, we can write

$$Y = \frac{1}{4}\left[1 + \frac{2}{N} \sum_{rt}' \langle S_r^+ S_t^- \rangle\right]. \qquad (10)$$

$\langle S_r^+ S_t^+ \rangle$ and $\langle S_r^- S_t^- \rangle$ are necessarily zero since $\mathcal{H}$ does not contain such terms.

Now if we define

$$\mathfrak{Z} = \operatorname{Tr} \exp K(P + 2\lambda R), \qquad (11)$$

where

$$R = \sum_{rt}' S_r^+ S_t^-,$$

then

$$Y = \frac{1}{4}\left[1 + \frac{1}{N}\frac{1}{K}\frac{\partial}{\partial \lambda} \ln \mathfrak{Z}|_{\lambda=0}\right]. \qquad (12)$$

Thus, $\ln \mathfrak{Z}$ can be expanded as in the partition function,

$$\frac{1}{NK}\frac{\partial}{\partial \lambda} \ln \mathfrak{Z}|_{\lambda=0} = \frac{2}{N} \sum_{n=1}^{\infty} \frac{K^n}{n!} \operatorname{Tr}^* P^n R. \qquad (13)$$

The above expression looks very similar to the expression (5) of the partition function. The relationship between the two is as follows: Consider $\operatorname{Tr} P^{n+1}$ of the partition function. If we relax the restriction on one of the sums contained in P's (where the sums are limited to the nearest-neighbor pairs), so that now one sum is over any pairs, then we obtain precisely $\operatorname{Tr}(P^n R)$ of the fluctuation. Thus, to obtain graphs of nth order of the fluctuation,[13] it is sufficient to consider only the graphs of $(n+1)$th order of the partition function and to transform one edge of each graph to any arbitrary length. For an n-sided graph (i.e., nth-order Z graph) there are n ways of transforming it, and the occurrence factor associated with the Z graph is distributed equally among the n new graphs (some of the graphs may be the same). The trace factor is still unaffected since it depends only on the number of, not the length between, the vertices of a graph. The following examples will illustrate the above transformation.

Example 1: Consider the graph 1(A), which is the 2nd-order partition function graph. To get the corresponding first-order fluctuation graph, we are to make one line arbitrary in length. There are two ways this can be done. But in this case it is not possible to do so, since each is constrained by the other to be strictly of the nearest-neighbor distance. Thus we have for first-order fluctuation exactly the same partition function graph with the same occurrence factor [see Fig. 2(A)].

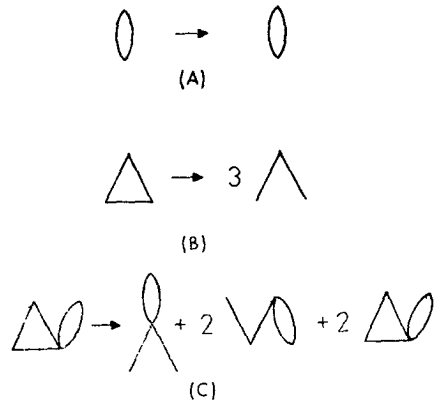


FIG. 2. The process by which Y graphs are generated from Z graphs.

Example 2: Consider Fig. 1(B), which is the third-order partition function graph. The second-order fluctuation graph can be obtained from this by making one side of the triangle arbitrary in length. There are 3 ways this can be done, but each of which produces the same chain of 2 lines as shown in Fig. 2(B). Since the chain of 2 lines is the only new graph formed, it will "inherit" the entire occurrence factor of the triangle of the partition function.

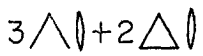
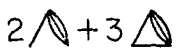
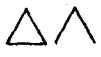
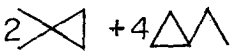
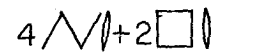
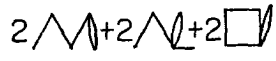
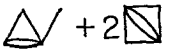
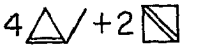
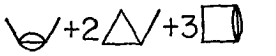
Example 3: Consider Fig. 1(F) which is 5th-order Z graph . This generates 4th-order Y graphs as shown in Fig. 2(C). The coefficients 1, 2, 2 indicate the number of ways the new graphs can be made from the Z graph. Now if the occurrence factor of the Z graph is x, then the occurrence factors for the Y graphs are, in the order of appearance, $\frac{1}{5}x$, $\frac{2}{5}x$, and $\frac{2}{5}x$.

Thus from these examples the following points will be self-evident:

(i) Y graphs as well as their occurrence factors can be generated directly from Z graphs.

(ii) There are many more graphs in the fluctuation than in the corresponding order of the partition function. (See Table IV.) Therefore, it is reasonable to expect that the behavior of the fluctuation series is much more satisfactory than that of the specific-heat series. (Table V shows the comparison in the numbers of graphs.)

TABLE IV. Fluctuation graphs are obtained directly from the Z graphs by the elongation process explained in the text.

Partition Function		Fluctuation	
Order	Z Graphs	Order	Y Graphs
2		1	
3		2	
4[a]		3[a]	
5		4	
			
			
			
			I
6[a]		5[a]	
			
			
			
			
			
			
			
			II

[a] Graphs which are unchanged are omitted.

(iii) Y graphs are more often open graphs and, since open graphs have much larger values of the lattice constants, we may also expect that the fluctuation series to settle down more quickly than the specific-heat series.

For the general lattice, the fluctuation series up to 6th order is

TABLE V. The number of graphs. The two numbers in a parenthesis represent the numbers connected and disconnected graphs, respectively.

Order	Partition Function	Fluctuation
1	0(0, 0)	1(1, 0)
2	1(1, 0)	1(1, 0)
3	1(1, 0)	4(3, 1)
4	4(3, 1)	8(6, 2)
5	4(3, 1)	25(19, 6)
6	17(12, 5)	64(49, 15)
7	24(17, 7)	193(148, 45)
8	81(55, 26)	587(464, 123)
9	159(114, 45)	

$$
\begin{aligned}
2Y = \tfrac{1}{2} &+ K[0.5\{G(1)\}] + (K^2/2!)[0.5\{G(2)\}] \\
&+ (K^3/3!)[0.5\{G(1)\} + \{G(2)\} + 0.75\{G(5)\} \\
&+ 1.5\{G(1,1)\}] + (K^4/4!)[\{G(2)\} + 4.5\{G(3)\} \\
&+ 3\{G(4)\} + 2\{G(5)\} + 2\{G(7)\} + 1.5\{G(10)\} \\
&+ 1.5\{G(1,2)\} + 3\{G(1,3)\}] \\
&+ (K^5/5!)[0.5\{G(1)\} + 3\{G(2)\} + 6\{G(3)\} \\
&+ 4.5\{G(4)\} + 9\{G(5)\} + 30\{G(6)\} \\
&+ 5.5\{G(7)\} + 5\{G(9)\} + 5\{G(10)\} \\
&+ 5.5\{G(11)\} + 5\{G(13)\} + 5\{G(15)\} \\
&+ 5\{G(16)\} + 3.75\{G(22)\} + 7.5\{G(1,1)\} \\
&+ 7.5\{G(1,2)\} + 3.75\{G(1,5)\} + 7.5\{G(1,6)\} \\
&+ 7.5\{G(2,3)\} + 11.25\{G(1,1,1)\}] \\
&+ (K^6/6!)[2\{G(2)\} + 25.5\{G(3)\} + 16.5\{G(4)\} \\
&+ 12.5\{G(5)\} + 22\{G(6)\} + 50.75\{G(7)\} \\
&+ 27\{G(8)\} + 19.75\{G(9)\} + 30.5\{G(10)\} \\
&+ 44\{G(11)\} + 56.25\{G(12)\} + 8.25\{G(13)\} \\
&+ 24.125\{G(14)\} + 18\{G(15)\} + 29.25\{G(16)\} \\
&+ 15\{G(20)\} + 7.5\{G(21)\} + 15\{G(22)\} \\
&+ 15\{G(24)\} + 16.5\{G(25)\} + 30\{G(28)\} \\
&+ 15\{G(30)\} + 15\{G(33)\} + 15\{G(34)\} \\
&+ 15\{G(40)\} + 15\{G(41)\} + 11.25\{G(52)\} \\
&+ 11.25\{G(1,2)\} + 71.25\{G(1,3)\} \\
&+ 22.5\{G(1,4)\} + 15\{G(1,5)\} + 30\{G(1,7)\} \\
&+ 11.25\{G(1,10)\} + 22.5\{G(1,12)\} \\
&+ 7.5\{G(2,2)\} + 30\{G(2,3)\} + 22.5\{G(2,6)\} \\
&+ 22.5\{G(3,5)\} + 11.25\{G(1,1,2)\} \\
&+ 45\{G(1,1,3)\}] + \cdots . \qquad (14)
\end{aligned}
$$

V. THE SUSCEPTIBILITY IN THE PERPENDICULAR DIRECTION

One can also obtain the susceptibility in the perpendicular direction from the partition function series. The susceptibility is defined in the usual way

$$\chi_\perp = K_B T \frac{\partial^2}{\partial H^2} \ln Z_\perp,$$

where

$$Z_\perp = \mathrm{Tr} \exp -\beta\left(\mathcal{H} - H \sum_r S_r^z\right).$$

Since the interaction term and the Zeeman term commute, we can write the susceptibility per spin as

$$K_B T \chi_\perp = \frac{1}{N} \sum_{n=0}^{\infty} \frac{K^n}{n!} \mathrm{Tr}^* P^n Q^2, \tag{15}$$

where

$$Q = \sum_{r=1}^{N} S_r^+ S_r^- = \sum_r N_r. \tag{16}$$

We shall regard $S_r^+ S_r^-$ as a "bubble" (loosely in the sense of an electron in a Fermi sea) and graphically denote it as shown in Fig. 5(A). Since Q does not connect any pair of points in the lattice, a bubble or a combination of bubbles is the only kind of graph that Q can generate. The effect of Q in $\mathrm{Tr}\, P^n Q^2$ is thus to decorate each of the partition function graphs ($\mathrm{Tr}\, P^n$) with one or two bubbles on or off the site of a graph on the lattice. The presence of Q, however, considerably complicates the trace of a graph, which is now no longer simply related to the number of the vertices of a graph as is in the trace of the partition function graphs and the fluctuation graphs.

We can write

$$Q^2 = \sum_r N_r + \sum_{rt}{}' N_r N_t. \tag{17}$$

Hence, to obtain $\mathrm{Tr}\, P^n Q^2$, we must decorate each of the partition function graphs twice, first with one bubble and then with two bubbles. When bubbles are decorated on the site of a graph [see Figs. 3(A), 3(B)], then the trace is still simply 2^{-v}. But when bubbles are decorated off the site of a graph [see Figs. 3(C), 3(D)], there seems to be no simple relation for the trace. In Appendix B we shall consider a few simple examples.

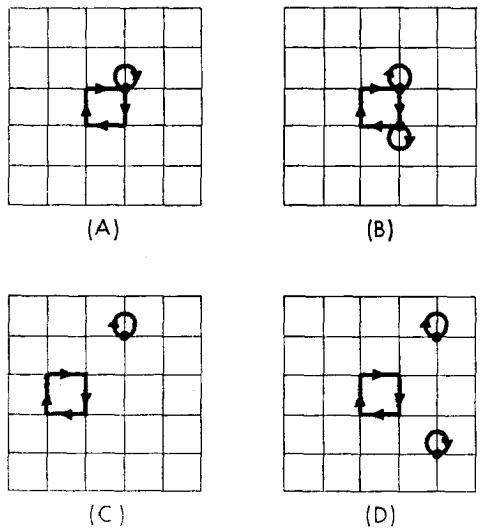


FIG. 3. A square is decorated by one and two bubbles on and off the site of the simple square lattice.

It will be clear from these examples in Appendix B that the trace of a decorated graph $\mathrm{Tr}\, P^n Q^2$, apart from the lattice constant and the occurrence factor both of which arise from P^n, depends solely on the number of N's and $N^\dagger$'s that a graph has.[14] Thus the effect of decoration is, in this sense, again independent of a graph and a lattice constant. That is, $N N N^\dagger N^\dagger$, for example, which refers to $G(6)$ or $G(11)$, yields an identical coefficient ($-2^{-4}x\{G(6)\}$ and $-2^{-4}y\{G(11)\}$, where x and y are their respective occurrence factors). Hence one can conveniently tabulate the coefficients of traces of all possible combination of N's and $N^\dagger$'s which can make up the Z graphs and thereby avoid tedious effort of taking traces for each decorated graph. In Table VI we have tabulated the coefficients of the combinations of $7N$'s and $N^\dagger$'s.

TABLE VI. Trace of decorated graphs.

	Decorated Graphs	Trace[a]		Decorated Graphs	Trace[a]
2.1	$NN^\dagger$	$-\frac{1}{2}$	6.1	$NNNNNN^\dagger$	$\frac{5}{2}$
3.1	$NNN^\dagger$	$-\frac{1}{2}$	6.2	$NNNNN^\dagger N^\dagger$	$-\frac{1}{2}$
3.2	$NN^\dagger N^\dagger$	$-\frac{1}{2}$	6.3	$NNNN^\dagger N^\dagger N^\dagger$	$-\frac{3}{2}$
4.1	$NNNN^\dagger$	0	6.4	$NNN^\dagger N^\dagger N^\dagger N^\dagger$	$-\frac{1}{2}$
4.2	$NNN^\dagger N^\dagger$	-1	6.5	$NN^\dagger N^\dagger N^\dagger N^\dagger N^\dagger$	$\frac{5}{2}$
4.3	$NN^\dagger N^\dagger N^\dagger$	0	7.1	$NNNNNNN^\dagger$	$\frac{3}{2}$
5.1	$NNNNN^\dagger$	1	7.2	$NNNNNN^\dagger N^\dagger$	$\frac{1}{2}$
5.2	$NNNN^\dagger N^\dagger$	-1	7.3	$NNNNN^\dagger N^\dagger N^\dagger$	$-\frac{3}{2}$
5.3	$NNN^\dagger N^\dagger N^\dagger$	-1	7.4	$NNNN^\dagger N^\dagger N^\dagger N^\dagger$	$-\frac{3}{2}$
5.4	$NN^\dagger N^\dagger N^\dagger N^\dagger$	1	7.5	$NNN^\dagger N^\dagger N^\dagger N^\dagger N^\dagger$	$\frac{1}{2}$
			7.6	$NN^\dagger N^\dagger N^\dagger N^\dagger N^\dagger N^\dagger$	$\frac{9}{2}$

[a] To be multiplied by 2^{-n}, where n is the number of N's and $N^\dagger$'s of a graph, which is the same as the number of the vertices of a graph.

The susceptibility per spin in the perpendicular direction up to the seventh term for the general lattice is given as follows[15]:

$$
\begin{aligned}
&4K_BT\chi_\perp \\
&= 1 - (K^2/2!)[\{G(1)\}] - (K^3/3!)[3\{G(3)\}] \\
&\quad - (K^4/4!)[\{G(1)\} + 2\{G(2)\} + 8\{G(6)\} \\
&\quad + 6\{G(1,1)\}] - (K^5/5!)[15\{G(3)\} + 15\{G(7)\} \\
&\quad + 25\{G(12)\} + 30\{G(1,3)\}] - (K^6/6!)[\{G(1)\} \\
&\quad + 6\{G(2)\} + 12\{G(3)\} + 3\{G(4)\} + 18\{G(5)\} \\
&\quad + 96\{G(6)\} + 48\{G(11)\} + 5\{G(13)\} + 36\{G(26)\} \\
&\quad + 90\{G(29)\} + 30\{G(1,1)\} + 30\{G(1,2)\} \\
&\quad + 7\{G(1,6)\} + 18\{G(3,3)\} + 67.5\{G(1,1,1)\}] \\
&\quad + (K^7/7!)[53\{G(3)\} + 175\{G(7)\} + 280\{G(11)\} \\
&\quad + 490\{G(12)\} + 133\{G(14)\} + 112\{G(15)\} \\
&\quad + 140\{G(16)\} + 1008\{G(25)\} + 448\{G(30)\} \\
&\quad + 399\{G(54)\} + 441\{G(60)\} + 367.5\{G(76)\} \\
&\quad + 420\{G(1,3)\} + 37.5\{G(1,7)\} + 420\{G(1,12)\} \\
&\quad + 105\{G(2,3)\} + 735\{G(3,6)\} \\
&\quad + 472.5\{G(1,1,3)\}] - \cdots . \qquad (18)
\end{aligned}
$$

ACKNOWLEDGMENT

I am most grateful to Professor D. D. Betts for suggesting this problem and for many helpful discussions with him.

APPENDIX A: OCCURRENCE FACTORS

Through the following examples we show how the occurrence factors are determined (the graphs are identified in Fig. 4).

1. G(ij)

Define $A = S_i^+ S_j^-$ and $A^\dagger = S_j^+ S_i^-$, where sum on i and j are implied. From the commutation relations of $S_j^\pm$, we see that $A\,A = A^\dagger\,A^\dagger = 0$. The matrix elements for this graph are $A\,A^\dagger\,A\,A^\dagger = N\,N^\dagger$. Since this is the only possible representation, the occurrence factor is unity. In fact, for all the two-vertex polygons the occurrence factor is always 1.

2. G(ijk)

Define $A = S_i^+ S_j^-$, $A^\dagger = S_j^+ S_i^-$, $B = S_j^+ S_k^-$, and $B^\dagger = S_k^+ S_j^-$. From the commutation relations of $S_j^\pm$,

$$A\,B = A^\dagger\,B^\dagger = A^\dagger\,B\,A^\dagger = A\,B^\dagger\,A = 0.$$

The matrix elements in terms of A and B are

$$
\begin{aligned}
&A\,A^\dagger\,A\,A^\dagger\,B\,B^\dagger + A^\dagger\,A\,A^\dagger\,B\,B^\dagger\,A + A\,A^\dagger\,B\,B^\dagger\,A\,A^\dagger \\
&\qquad + A^\dagger\,B\,B^\dagger\,A\,A^\dagger\,A + B\,B^\dagger\,A\,A^\dagger\,A\,A^\dagger \\
&\qquad + B^\dagger\,A\,A^\dagger\,A\,A^\dagger\,B.
\end{aligned}
$$

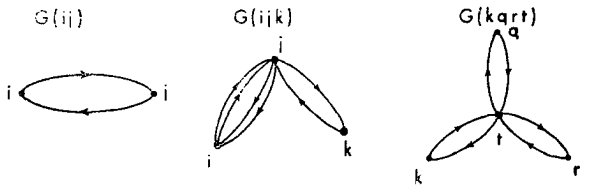


FIG. 4. Graphs whose occurrence factors are illustrated in Appendix A.

Now, since A and B are inequivalent, one can produce 6 more terms by exchanging the roles of A and B. Hence we obtain for the occurrence factor which is $6(N\,N\,N^\dagger + N\,N^\dagger\,N^\dagger)$.

3. G(kqrt)

Define $A = S_k^+ S_t^-$, $B = S_q^+ S_t^-$, and $C = S_r^+ S_t^-$, with $A^\dagger\,B^\dagger\,C^\dagger$ defined analogously as before. Again

$$A\,B = B\,C = A\,C = A^\dagger\,B^\dagger = B^\dagger\,C^\dagger = A^\dagger\,C^\dagger = 0.$$

The matrix elements can be generated by permuting $A^\dagger\,B^\dagger\,C^\dagger$ in $A\,A^\dagger\,B\,B^\dagger\,C\,C^\dagger$ with $A\,B\,C$ fixed and by permuting $A^\dagger\,B^\dagger\,C^\dagger$ in $A^\dagger\,A\,B^\dagger\,B\,C^\dagger\,C$ with $A\,B\,C$ fixed. The result of this permutation gives the occurrence factor which is

$$2(N\,N\,N\,N^\dagger + 4N\,N\,N^\dagger\,N^\dagger + N\,N^\dagger\,N^\dagger\,N^\dagger).$$

4. The Occurrence Factors for Simple Polygons

The occurrence factor for n-sided simple polygon (by a simple polygon we mean a graph with 2 lines at every vertex) can be obtained recursively from the occurrence factor for $(n-1)$-sided simple polygon. This recursion relation is quite simple and can be used to obtain the occurrence factors for all simple polygons. We shall illustrate this relation by showing the occurrence factor for $G(6)$ obtained from $G(3)$ whose occurrence factor is assumed known. Write $(N_iN_jN_k^\dagger + N_iN_j^\dagger N_k^\dagger)\cdot N_mN_n^\dagger$. If m or n happens to be the same as i, j, or k, then, N being idempotent, we reduce the numbers of N's by one from 5 to 4, which is just the required number of N's for $G(6)$. Recalling that $N_iN_j^\dagger = 0$, we find that the result of this reduction in all possible ways is

$$N\,N\,N\,N^\dagger + 4N\,N\,N^\dagger\,N^\dagger + N\,N^\dagger\,N^\dagger\,N^\dagger.$$

This is the occurrence factor for $G(6)$.

APPENDIX B: THE TRACE OF A DECORATED GRAPH

The trace of a graph decorated by one or two bubbles can be simply obtained. The following two examples will illustrate the simplicity.

(1) The trace of a lattice which is decorated by one and two bubbles [see Figs.5(A) and 5(B)]:

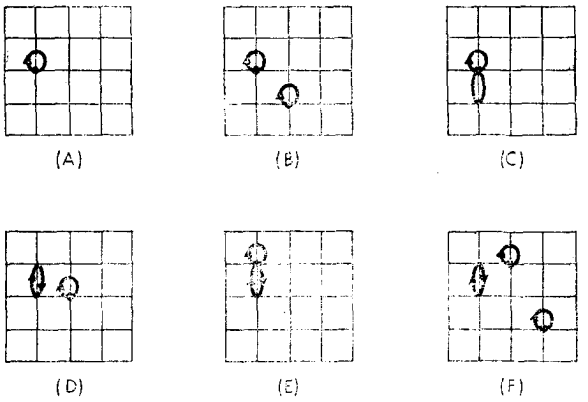


FIG. 5. (A), (B) The simple square lattice is decorated by one and two bubbles. (C)–(F) A two-sided polygon is decorated by one and two bubbles on and off the site of the lattice.

$$\mathrm{Tr}^* Q^2 = \mathrm{Tr}^* \sum_r N_r + \mathrm{Tr}^* \sum_{rt}{}' N_r N_t = \tfrac{1}{2}N - \tfrac{1}{4}N.$$

The second term results from

$$2^{-N}\,\mathrm{Tr} \sum_{rt}{}' N_r N_t = 2^{-2}N(N-1).$$

Hence,

$$\mathrm{Tr}^* \sum_{rt}{}' N_r N_t = -\tfrac{1}{4}N.$$

(2) The trace of the 2-sided polygon decorated by one and two bubbles [see Fig. 5(C), 5(D), 5(E), 5(F)]:

$$\mathrm{Tr}^* P^2Q^2 = \mathrm{Tr}^* \sum_{(mn)} N_m N_n^\dagger \Big\{ \sum_r N_r + \sum_{rt}{}' N_r N_t \Big\}.$$

(i) One bubble:

$$\mathrm{Tr}^* \sum_{(mn)} N_m N_n^\dagger \sum_r N_r = \tfrac{1}{4}Nq - \tfrac{1}{4}Nq = 0.$$

(ii) Two bubbles:

$$\mathrm{Tr}^* \sum_{(mn)} N_m N_n^\dagger \sum_{rt}{}' N_r N_t = -\tfrac{1}{2}Nq + \tfrac{3}{8}Nq = -\tfrac{1}{8}Nq.$$

Hence,

$$\mathrm{Tr}^* P^2Q^2 = -\tfrac{1}{8}Nq.$$

---

* Address after September 1, 1969: Department of Physics, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

[1] T. Obukata, I. Ono, and T. Oguchi, J. Phys. Soc. (Japan) **23**, 516 (1967); K. Pirnie (unpublished).

[2a] D. D. Betts and M. H. Lee, Phys. Rev. Letters **20**, 1507 (1968). M. H. Lee, J. Math. Phys. **10**, 1813 (1969).

[3] T. Matsubara and H. Matsuda, Progr. Theoret. Phys. (Kyoto) **16**, 416 (1956).

[4] R. Witlock and P. Zilsel, Phys. Rev. **13**, 2409 (1963); M. E. Fisher, Rept. Progr. Phys. **30**, 615 (1967), in particular, see Sec. 4.3.

[5] By $O(N)$ we mean a term proportional to N.

[6] See, for example, C. Domb, Advan. in Phys. **9**, 149 (1960).

[7] We shall regard Fig. 1(A) as a two-sided polygon.

[8] We shall henceforth suppress the suffixes of N's and place a corresponding graph along side. There will be no ambiguities.

[9] C. Domb and M. F. Sykes, Proc. Roy. Soc. (London) **A240**, 214 (1957); C. Domb, Advan. Phys. **9**, 149 (1960); more complete information on the lattice constants may be found in G. A. Baker, H. E. Gilbert, J. Eve, and G. S. Rushbrooke, "A Date Compendium of Linear Graphs with Application to the Heisenberg Model," BNL50053(T-460), 1967.

[10] We shall call them Z graphs.

[11] M. E. Fisher, J. Math. Phys. **4**, 124 (1963).

[12] M. E. Fisher, *Lectures in Theoretical Physics, Vol. VII* (University of Colorado Press, Boulder, 1964), p. 1; R. B. Griffiths, Phys. Rev. **152**, 240 (1966).

[13] We shall call them Y graphs.

[14] Recall that in the partition function the effect of the trace is to give the coefficient 2^{-v}. Thus the two different graphs, but with the same number of the vertices, for example, have an identical coefficient.

[15] Pirnie has obtained the first five terms of this series for fcc and Obukata *et al.* have obtained the first seven terms for linear chain, simple quadratic, and simple cubic lattices (see Ref. 1). Our series is in exact agreement.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Application of Regge Calculus to the Schwarzschild and Reissner–Nordstrøm Geometries at the Moment of Time Symmetry*

CHEUK-YIN WONG†
Princeton University, Princeton, New Jersey
and
Oak Ridge National Laboratory,‡ Oak Ridge, Tennessee 37830

(Received 9 July 1970)

Applied to 3-dimensional space, Regge calculus approximates a curved space by a collection of tetrahedrons or other simple solid blocks. Within each block the geometry is Euclidean. Curvature is idealized as concentrated at the edge common to two or more of these solids. We specialize to a static geometry endowed with spherical symmetry and to a radial electric field produced by the flux of electric lines of force trapped in a throat connecting two quasi-Euclidean regions of space. The one relevant Einstein field equation takes the form

$$\sum_{\substack{\text{all edges which}\\ \text{meet at a given}\\ \text{vertex}}} \begin{pmatrix}\text{length of edge}\\ \text{of prism}\end{pmatrix}\begin{pmatrix}\text{deficit between (1) sum of}\\ \text{dihedral angles which meet}\\ \text{at that edge and (2) normal}\\ \text{value of } 2\pi\end{pmatrix} = \begin{pmatrix}\text{factor proportional to}\\ \text{square of electric field}\end{pmatrix}.$$

In method (1) the space is decomposed into shells separated from one another by icosahedral surfaces, all having a common center. Method (2) is even simpler: Space is decomposed into successive spherical shells of area $4\pi\rho_i^2$ separated by a proper distance d. Regge calculus gives a recurrence relation relating the dimensions of the successive shells. The approximate geometries calculated by methods (1) and (2) are compared with the well-known exact Schwarzschild and Reissner–Nordstrøm geometries. Errors range from roughly 10% down to less than 1%, depending upon the method of analysis, the quantity under analysis, and the fineness of the subdivision.

I. INTRODUCTION AND SUMMARY

We apply Regge calculus[1] to a 3-dimensional problem of general relativity where the answer is already known (1) to assess the accuracy of this kind of skeleton analysis, (2) to display the features of the geometry, and (3) to pave the way for some later applications of Regge calculus to truly 4-dimensional problems, problems of geometrodynamics where the 3-dimensional geometry changes with time. Such a 3-dimensional problem can be readily found in a time-symmetric initial value problem[2] possessing spherical symmetry. The features of the time-symmetric initial value geometries of Schwarzschild (mass, no charge) and of Reissner–Nordstrøm (mass and charge) are already well known.[2,3] They possess a throat of minimum area connecting two regions of space which are asymptotically flat. Moreover, these geometries are symmetric with respect to reflection in the throat. This is to say, the 3-geometry $^{(3)}\mathcal{G}$ extended forward in the direction of the "upper Euclidean" region from the throat is identical to the $^{(3)}\mathcal{G}$ extended backward in the "lower Euclidean" region. By symmetry, we therefore know that the extrinsic curvature of the throat $^{(2)}\mathcal{G}$ vanishes. For this reason it is enough to specify the radius of the 2-sphere located at the throat in order to have a well-defined boundary condition. With this boundary condition, Regge calculus allows one to integrate step by step from the throat so that the complete $^{(3)}\mathcal{G}$ can be traced out. The procedure employed in this paper, where one starts from the throat of the wormhole and generates the entire 3-geometry, is analogous to that envisaged for future, more sophisticated, applications of Regge calculus where one thinks of starting from a time-symmetric 3-geometry and finding by integration the entire 4-geometry.

Two methods have been used in the present work to decompose into Euclidean blocks (with "rattle" at the edges of joining) a 3-geometry possessing spherical symmetry. In the first, the "icosahedral method," successive spherical cross sections throughout the geometry are approximated by icosahedral surfaces, as an icosahedron is the regular polyhedron with the largest number of identical faces. Were we applying the Regge analysis to a problem of no special symmetry, we would be well advised to take the building blocks of the 3-dimensional space to be 3-simplexes (tetrahedrons). However, the symmetry of the present problem suggests the use of prisms, such as shown in Fig. 27 of Ref. 4, rather than tetrahedrons, to fill in the space between two successive icosahedral surfaces. Twenty blocks are needed for

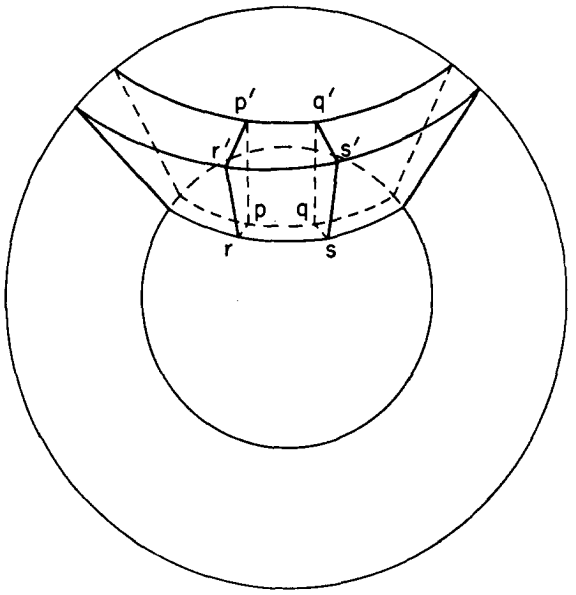


FIG. 1. Decomposition of the 3-geometry by the continuum method. Each surface is divided into very large numbers of very small rhombuses, such as $pqrs$ lying on the inner surface and $p'q'r's'$ lying on the outer surface. Joining the corresponding points p to p', q to q', r to r', and s to s' gives the typical building block with which the 3-space of the shell can be constructed. To insure spherical symmetry of the outer surface after construction and to maintain the "rigidity" of the building block, we demand (1) that the base and summit rhombus of each block be parallel to each other, (2) that all the surfaces of the block be plane surfaces, and (3) that the lengths pp', qq', rr', and ss' be all equal to d.

each "shell." It is sufficient to give the three edge lengths a_i, γa_i, and a_{i+1} in order to determine completely the geometry of such a block. The "continuum method" treats each spherical cross section throughout the geometry as a polyhedral surface with a very large number of very small rhombus faces. The shell between two spherical surfaces separated by a small proper distance d is broken up into blocks which have parallel rhombus bases as shown in Fig. 1.

Numerical calculation was carried out with an IBM 650 and later an IBM 7094 computer at Princeton. In the region where the space is quite flat, the error of approximating a sphere by an icosahedron shows up (Table I) in the icosahedral method, but there is no such error in the "continuum method"; there is only an error (Table II) in that method because of the finite separation d between successive spherical surfaces.

II. THE SCHWARZSCHILD AND REISSNER-NORDSTRØM GEOMETRY AT THE MOMENT OF TIME SYMMETRY

Acceptable initial value data on an initial spacelike hypersurface σ, in our case of empty space, comprise the following: First, divergence-free electric and magnetic fields on the hypersurface,

$$\nabla \cdot \mathbf{H} = 0, \tag{1}$$

$$\nabla \cdot \mathbf{E} = 0; \tag{2}$$

second, a suitably regular initial geometry $g_{ik}(x^1, x^2, x^3)$ on this hypersurface; and third, an imbedding of this hypersurface in the enveloping and yet-to-be constructed 4-geometry, for which the tensor K_{ik} of "extrinsic curvature"[5] will satisfy the four Fourès-Bruhat initial value conditions. Three of these conditions have to do with the Poynting flux and will be automatically fulfilled in our special problem (zero magnetic field and zero extrinsic curvature). The fourth reads

$$\begin{pmatrix}\text{intrinsic}\\ \text{scalar}\\ \text{curvature}\\ \text{invariant}\\ \text{of } \sigma\end{pmatrix} - \begin{pmatrix}\text{extrinsic}\\ \text{curvature}\\ \text{invariant}\\ \text{of } \sigma\end{pmatrix} = 2\begin{pmatrix}8\pi \text{ times}\\ \text{energy}\\ \text{density}\\ \text{on } \sigma\end{pmatrix}. \tag{3}$$

Here and throughout the paper we use the geometrized quantities which are related to the conventional quantities ("conv") by

$$\text{mass:} \quad m = (G/c^2)m_{\text{conv}}, \tag{4}$$

$$\text{charge:} \quad q = (G^{\frac{1}{2}}/c^2)q_{\text{conv}}, \tag{5}$$

$$\text{electric field:} \quad \mathbf{e} = (G^{\frac{1}{2}}/c^2)\mathbf{e}_{\text{conv}}, \tag{6}$$

$$\text{energy density:} \quad \begin{pmatrix}\text{energy}\\ \text{density}\end{pmatrix} = (G/c^4)\begin{pmatrix}\text{energy}\\ \text{density}\end{pmatrix}_{\text{conv}}. \tag{7}$$

Our problem is simple because the initial value problem is time symmetric.[2] The extrinsic curvature vanishes on the 3-geometry of time symmetry. Equation (3) is simplified to

$${}^{(3)}R = 2\begin{pmatrix}8\pi \text{ times the energy}\\ \text{density on this}\\ \text{3-space}\end{pmatrix}. \tag{8}$$

The most natural problem to solve within the framework of Eqs. (1), (2), (3), and (8)—as a test of Regge calculus—is one where the answer is known. The simplest examples of this kind possess spherical symmetry. They alone are considered here. The electric field is

$$e = \frac{(\text{flux constant})}{(\text{proper surface area})} = \frac{4\pi q}{A}. \tag{9}$$

The charge q is envisaged as a measure of the number of electric lines of force trapped in the topology of space ("wormhole interpretation of electric charge").[3]

From the electric field we find the density of electromagnetic energy. Thereupon we find that the initial value equation (8) takes the form

$$^{(3)}R = 32\pi^2 q^2/A^2. \tag{10}$$

The metric of a 3-space of spherical symmetry can always be written

$$ds^2 = \psi^4(dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\varphi^2). \tag{11}$$

Consequently, the initial value condition is given by[6]

$$8\psi^{-5}\nabla^2\psi = 2q^2/(\psi^8 r^4). \tag{12}$$

The solution is[7]

$$\psi = [(1 + m/2r)^2 - (q/2r)^2]^{\frac{1}{2}}. \tag{13}$$

Here one of the constants of integration is chosen to be unity ("scale normalization"). The other is set equal to the mass.

The proper area of a spherical surface with coordinate radius r is

$$\begin{pmatrix}\text{proper area}\\ \text{of sphere of}\\ \text{coordinate } r\end{pmatrix} = 4\pi r^2[(1 + m/2r)^2 - (q/2r)^2]^2. \tag{14}$$

This expression has a minimum at $r = \frac{1}{2}(m^2 - q^2)^{\frac{1}{2}}$ corresponding to the throat of the wormhole. The proper area at the throat is given by

$$\begin{pmatrix}\text{minimum surface}\\ \text{area at throat}\end{pmatrix} = 4\pi[(m^2 - q^2)^{\frac{1}{2}} + m]^2. \tag{15}$$

The 3-geometry is completely symmetrical with respect to an interchange of the "upper" and "lower" Euclidean spaces, as appears most quickly on making the coordinate substitution

$$r = (m^2 - q^2)/4r'.$$

Three quantitative features of the geometry $^{(3)}\mathcal{G}$ are desired in order to provide points of comparison with the approximate geometry to be obtained from the Regge calculus: (1) the proper volume contained between $^{(2)}\mathcal{G}$ and a spherical (or icosahedral) surface of coordinate radius r, (2) the geodesic distance between the throat $^{(2)}\mathcal{G}$ and this surface, and (3) the rotation per unit area of a unit vector if transported parallel on this surface along a loop perpendicular to the radial direction. These relations are most conveniently expressed in parametric form in terms of the dimensionless quantity

$$u = 2r/(1 - \alpha^2)^{\frac{1}{2}}m, \tag{16}$$

where we have introduced

$$\alpha = |q|/m\ (\leq 1). \tag{17}$$

In terms of this parameter we have

$$(\text{radial coordinate } r) = \tfrac{1}{2}(1 - \alpha^2)^{\frac{1}{2}}mu, \tag{18}$$

$$\begin{pmatrix}\text{proper area of}\\ \text{throat } ^{(2)}\mathcal{G}\end{pmatrix} = 4\pi[1 + (1 - \alpha^2)^{\frac{1}{2}}]^2m^2, \tag{19}$$

$$\begin{pmatrix}\text{proper area of}\\ \text{sperical surface with}\\ \text{coordinate } u\end{pmatrix} = \pi(1 - \alpha^2)[u + u^{-1} + 2(1 - \alpha^2)^{-\frac{1}{2}}]^2m^2, \tag{20}$$

$$\begin{pmatrix}\text{proper volume between}\\ ^{(2)}\mathcal{G} \text{ and the spherical}\\ \text{surface with coordinate } u\end{pmatrix} = \pi(1 - \alpha^2)^{\frac{3}{2}}\tfrac{1}{6}\{u^3 - u^{-3} + 9(1 - \alpha^2)^{-\frac{1}{2}}(u^2 - u^{-2}) + 9[1 + 4(1 - \alpha^2)^{-1}](u - u^{-1}) + 3(20 - 12\alpha^2)(1 - \alpha^2)^{-\frac{3}{2}}\ln u\}m^3, \tag{21}$$

$$\begin{pmatrix}\text{proper distance measured}\\ \text{outward from } ^{(2)}\mathcal{G} \text{ to a}\\ \text{spherical surface with}\\ \text{coordinate } u\end{pmatrix} = (1 - \alpha^2)^{\frac{1}{2}}\tfrac{1}{2}[u - u^{-1} + 2(1 - \alpha^2)^{-\frac{1}{2}}\ln u]m. \tag{22}$$

Turning from distances to curvature, we consider a vector $\boldsymbol{\theta}$ pointing in the direction of increasing θ and a vector $\boldsymbol{\varphi}$ pointing in the direction of increasing φ. We use them to define an elementary loop perpendicular to the direction of increasing r. We take a unit vector that lies in the plane of this loop and carry it around the loop by parallel transport. The amount of rotation that this vector undergoes, divided by the size of the loop, defines a "rotation factor," with the value

$$\theta_{23} \equiv \begin{pmatrix}\text{rotation of unit vector}\\ \text{per unit of area in}\\ \text{plane of } \theta,\ \varphi \text{ loop}\end{pmatrix} = \frac{R^i{}_{jkl}\varphi^j\theta^k\varphi^l\theta_i}{g_{mn}\theta^m\varphi^n g_{pq}\theta^p\varphi^q} = g^{22}g^{33}R_{2323}. \tag{23}$$

In deriving the last equation we have assumed that the 2 and 3 axes are aligned along the directions of increasing θ and φ. From (23) we find

$$\theta_{23} = 16(1 - \alpha^2)^{-\frac{3}{2}}[u + u^{-1} + 2(1 - \alpha^2)^{-\frac{1}{2}}]^{-3} \times [1 - \alpha^2(1 - \alpha^2)^{-\frac{1}{2}} \times (u + u^{-1} + 2(1 - \alpha^2)^{-\frac{1}{2}})]m^{-2}. \tag{24}$$

Each of the above is written in such a way as to express the symmetry or antisymmetry with respect to the exchange of

$$u \to u^{-1},$$

which is the operation for the exchange of "upper" and "lower" Euclidean spaces.

In Regge calculus the initial value equation of general relativity for a time-symmetric 3-geometry translates[4] to

$$\sum_{\substack{\text{bones that meet}\\ \text{at the vertex } V_K\\ \text{enclosed by the}\\ \text{volume } V}} \frac{l_b \epsilon_b}{V} = \frac{32\pi^2 q^2}{A^2}. \tag{25}$$

This set of equations, plus an initial 2-surface of reflectional symmetry (between "upper" and "lower" Euclidean space) and spherical symmetry, is enough to allow one to calculate the whole 3-geometry.

A. Icosahedral Method

Travel outward in the Reissner–Nordstrøm geometry, starting at the throat. Then one encounters successive spherical surfaces, with successively increasing radii. These surfaces can be approximated as closely as one desires by polyhedral surfaces of sufficiently many faces. Let one limit attention to regular polyhedrons. Among regular polyhedrons none has more faces than the icosahedron, with its 20 triangles, 12 vertices, and 30 edges. Accordingly, we divide all space up into "shells" located between successive icosahedral surfaces. We further chop up each individual shell into 20 triangular prisms. The base of each prism sits on the inner icosahedral boundary of the shell, and the summit touches the outer icosahedral boundary.

Inside of each prism the geometry is Euclidean. Any curvature comes from the "angle of rattle" between these prismoidal blocks. Each triangular block can be uniquely specified by three lengths: an edge of the lower triangle, $(A_iB_i) = a_i$; an edge of the upper triangle, $(A_{i+1}B_{i+1}) = a_{i+1}$; and an edge connecting the two, $(A_iA_{i+1}) = b_i$ (Fig. 2).

If the building blocks are fitted together according to Euclidean geometry, these three lengths are not all independent. Any two of them, a_i and b_i for example, directly fix the third:

$$\begin{aligned} a_{i+1} &= a_i + [3 - 4\cos^2(\tfrac{1}{5}\pi)]b_i/\sin(\tfrac{1}{5}\pi) \\ &= a_i + 1.050 b_i \quad \text{(Euclidean case)}. \end{aligned} \tag{26}$$

In our case the blocks, though still individually Euclidean, do not fit together into any over-all Euclidean space, and (26) does not apply. In actuality, we do not even want to leave the slant height b_i of the prism as a free parameter. We want to have a purely automatic way of going from the characteristic dimension a_i of one icosahedral surface to the characteristic dimension a_{i+1} of the next icosahedral surface. Consequently, we fix b_i instead of leaving it as a freely disposable parameter. We give it the value

$$b_i = \gamma a_i. \tag{27}$$

To the "proportion factor" γ we assign a small value when we want to decompose space into thin prisms and a big value when we are content with a smaller number of thicker slices. Thus, with $\gamma = 0.1$ (the value adopted in the present calculations) and a characteristic dimension for the tenth icosahedron of a_{10}, we have for the characteristic dimensions of any other icosahedron

$$a_i = (1 + 0.1050)^{i-10} a_{10}.$$

In other words, an expansion factor of 1.1050 carries the family of icosahedrons into itself. No such simple

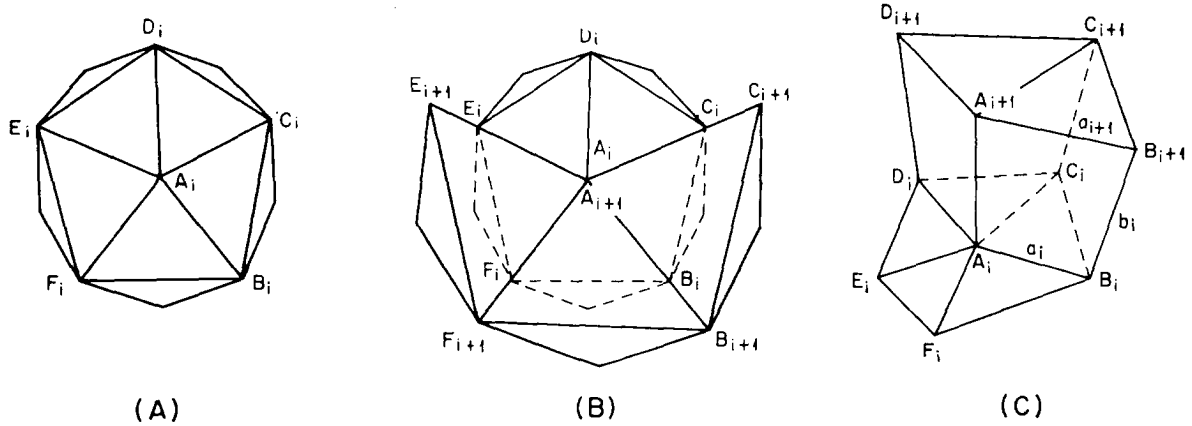


FIG. 2. Decomposition of the 3-geometry by the icosahedral method (see also Ref. 4). (A) The icosahedral surface $A_iB_iC_iD_iE_iF_i\cdots$ is used to approximate a spherical surface of the same area. (B) The 3-dimensional space is broken up into shells separated by icosahedral surfaces. (C) Between the surfaces $A_iB_iC_iD_iE_iF_i\cdots$ and $A_{i+1}B_{i+1}C_{i+1}D_{i+1}E_{i+1}F\cdots$ the space is fitted with 20 triangular blocks of the kind $A_iB_iC_iA_{i+1}B_{i+1}C_{i+1}$. We choose the planes $A_iB_iC_i$ and $A_{i+1}B_{i+1}C_{i+1}$ to be parallel and to have equal connecting edges A_iA_{i+1}, B_iB_{i+1}, and C_iC_{i+1}. Thus, the three lengths a_i, b_i, and a_{i+1} uniquely specify the geometry between the two surfaces.

relations obtain when we turn from flat space to the approximation of curved space by Euclidean blocks. The family of icosahedrons, instead of running from indefinitely small dimensions to indefinitely large ones, with a factor $(1 + 1.050\gamma)$ from one to the next, have now a smallest member, "the icosahedron of the throat," $i = 0$. Not only are icosahedra of positive i values larger than this special one, so also are icosahedra of negative i values (reflection symmetry $i \to -i$).

To fix the dimensions of the smallest icosahedron, we identify its *area* with the area of the throat as given by the exact analysis; thus, we write

$$\begin{pmatrix}\text{area of smallest}\\ \text{icosahedron}\end{pmatrix} = 5(\sqrt{3})a_0^2 = 4\pi[1 + (1 - \alpha^2)^{\frac{1}{2}}]^2 m^2 = \begin{pmatrix}\text{area of minimal}\\ \text{2-sphere}\end{pmatrix}. \quad (28)$$

We find $a_0 = 2.4092m$ for the Schwarzschild case ($\alpha = q/m = 0$) and $a_0 = 1.9273m$ for the case $\alpha = q/m = 0.8$ (a little less than the upper limit $\alpha = 1$ at which the Reissner–Nordstrøm geometry ceases to be regular). Whatever the actual value of a_0, it is convenient in the following to measure all quantities in terms of a_0 as a unit of length ("icosahedral units"). Thus the geometrized charge q, which has the dimensions of a length, is expressed in terms of a_0 as follows:

$$q = \alpha m = \alpha(m/a_0)a_0 = (\alpha/\kappa)a_0. \quad (29)$$

Here the conversion factor has the value

$$\kappa = a_0/m = \begin{cases}2.4092 & \text{for } \alpha = 0\\ 1.9273 & \text{for } \alpha = 0.8.\end{cases} \quad (30)$$

The departure of the dimensions of successive icosahedrons from a simple monotonic geometrical sequence arises from the curvature of space. This curvature cannot be seen or evaluated simply by an inspection of the dimensions of the prisms in a single icosahedral shell. It requires for its determination the dimensions of the blocks in two immediately succeeding icosahedrons. More specifically, focus attention on the vertex A_i in Fig. 2(C). We can get the curvature at this point only if we know all the dimensions of all the blocks that meet at this point: five prisms in the inner shell and five prisms in the outer one. Conversely, knowing the curvature (from the density of electric field energy) and knowing the dimensions

$$\left.\begin{matrix}a_{i-1}\\ b_{i-1} = \gamma a_{i-1}\\ a_i\end{matrix}\right\} \text{of the typical prism in the inner shell}$$

and

$$\left.\begin{matrix}a_i\\ b_i = \gamma a_i\end{matrix}\right\} \text{of the typical prism in the outer shell,}$$

we can determine the remaining dimension a_{i+1}. We are then ready to proceed to the next shell.

Knowing the characteristic dimension a_{i+1} of the $(i + 1)$th icosahedron, we can find its surface area A. Thence we find the electric field $4\pi q/A$, the density of electric energy at this new location, and the curvature there. This information puts us in a position to find a_{i+2}. So the calculation proceeds, step by step, in recursive style, as far out as one cares to go.

Denote by $V(a_i, b_i, a_{i+1})$ the volume of a triangular prism with edge lengths a_i and a_{i+1} on the triangles and slant height b_i. From (25) we find the formula to determine the characteristic dimension of the "new" icosahedron from the dimensions of the two preceding "old" icosahedrons:

$$5a_i\epsilon(a_i) + b_i\epsilon(b_i) + b_{i-1}\epsilon(b_{i-1}) = 32\pi^2(\alpha/\kappa)^2[\tfrac{5}{3}V(a_i, \tfrac{1}{2}b_i, \tfrac{1}{2}a_i + \tfrac{1}{2}a_{i+1}) + \tfrac{5}{3}V(\tfrac{1}{2}a_i + \tfrac{1}{2}a_{i-1}, \tfrac{1}{2}b_{i-1}, a_i)]/[5(\sqrt{3})a_i^2]^2, \quad (31)$$

where the formula for the angles is given in the Appendix.

This recursive procedure for going from a_{i-1} and a_i to a_{i+1} requires a slight amendment to make it apply to the very start of the calculation. In the beginning one knows a_0, but not a_{-1}, and one wants to calculate a_1. One has simply to identify a_{-1} with a_1 and b_{-1} with b_1 ("symmetry with respect to reflection at the throat"). Then in this special situation, as in the general case, (31) supplied a single equation to determine a single unknown. The calculations were made by the well-known method of repeated trial and error (Newton's method; rapid convergence). The results of the calculation appear in Table I.

B. *Continuum Method*

In the continuum method, a 3-geometry possessing spherical symmetry is divided into spherical shells separated by spherical surfaces. A surface is conveniently identified by its proper area. Consequently, it is appropriate to introduce a "characteristic dimension" (Schwarzschild radial coordinate) ρ for each surface such that the proper area of that spherical surface is given by $4\pi\rho^2$. Each spherical surface throughout the geometry is further subdivided into surface area elements as follows:

(1) Divide the polar angle π into equal intervals $\Delta\theta$ so that bands with widths $\rho\Delta\theta$ are obtained.

(2) In each band, divide the total azimuthal angle of 2π into equal intervals $\Delta\varphi$ (Figs. 1 and 3).

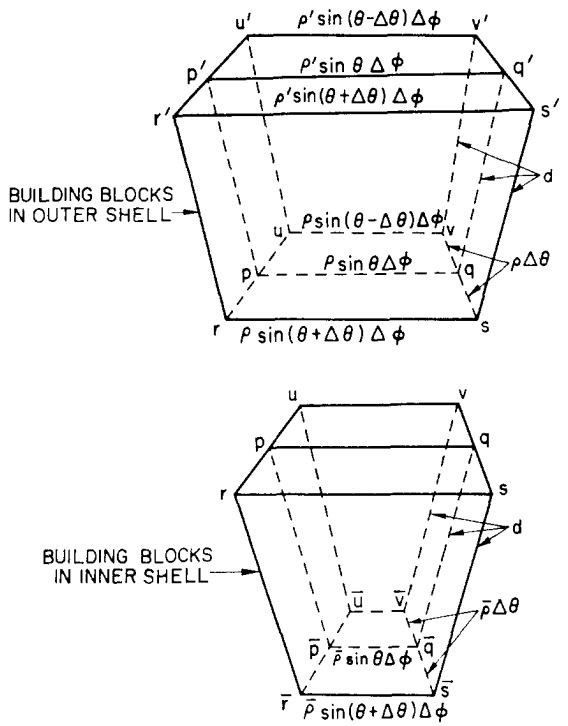


FIG. 3. In the continuum method, the 3-dimensional space between one spherical surface and the next is broken up into building blocks such as $pqrsp'q'r's'$. Shown here are surface area elements in three successive spherical surfaces whose areas starting from the bottom are determined by the quantities $\bar{\rho}$, ρ, and ρ', respectively. They are separated by proper distance d. Two building blocks extend the 3-geometry outward from $pqrsuv$ to $p'q'r's'u'v'$ and two inward to $\bar{p}\bar{q}\bar{r}\bar{s}\bar{u}\bar{v}$. The lengths of all the edges are as indicated. The 3-geometry in these successive spherical surfaces is uniquely specified by $\bar{\rho}$, ρ, ρ', and d.

The shell contained between one spherical surface and the next is broken up in this way into building blocks. The spherical symmetry so "rigidifies" these blocks that they need not be taken to be tetrahedrons. The typical building block is shown in Fig. 3. We build it as follows:

(1) Take a typical surface element $pqrs$ and join it to the next corresponding element $p'q'r's'$ with the connecting edges pp', qq', rr', and ss' all equal to one another. The common length d of these edges would be equal to $\Delta\rho$ if we were dealing with Euclidean geometry—but we are not!

(2) Demand that all the faces of the block be plane surfaces.

(3) Demand further that the planes $pqrs$ and $p'q'r's'$ be parallel to each other.

In order to specify the 3-geometry between one spherical surface and the next, three quantities are needed: the "characteristic dimensions" ρ and ρ' of the two surfaces and the "thickness" d between them. But these quantities in and by themselves do not determine curvature. Focus attention on the vertex q in Fig. 3. We can get the curvature at this point only if we know all the dimensions of all the blocks that meet at this point: four blocks in the inner shell and four in the outer one. For simplicity, we give the thickness of the inner and outer shells equal proper values, d. In that case, the dimensions of all the blocks meeting at the vertex q can be specified by the proper distance d and the characteristic dimensions $\bar{\rho}$, ρ, and ρ' of three successive spheres. Six bones meet at this vertex, qv, qs, qq', $q\bar{q}$, qp, and one more bone having the same characteristics as qp. We need to specify the quantities l_b, ϵ_b, V, and A for this vertex. The length l_b and the deficit angle ϵ_b of the bones are, respectively, $\rho\Delta\theta$ and $\epsilon(qv)$ for the bone qv, $\rho\Delta\theta$ and $\epsilon(qs)$ for the bone qs, d and $\epsilon(qq')$ for the bone qq', d and $\epsilon(q\bar{q})$ for the bone $q\bar{q}'$, and $\rho\sin\theta\Delta\varphi$ and $\epsilon(pq)$ for the bone pq. The fundamental equation of relativity becomes

$$\frac{d[\epsilon(qq') + \epsilon(q\bar{q})] + \rho\Delta\theta[\epsilon(qs) + \epsilon(qv)] + 2\rho\sin\theta\Delta\varphi\epsilon(qp)}{\rho^2 d\sin\theta\Delta\theta\Delta\varphi} = \frac{2q^2}{\rho^4}. \tag{32}$$

Write the deficit angles explicitly in terms of the lengths. Then take the limit when $\Delta\theta$ and $\Delta\varphi$ approach zero. After some simplifications, we finally obtain the following equation:

$$(\rho' - \rho)^2 + 4\rho(\rho' - \rho) + (\rho - \bar{\rho})^2 - 4\rho(\rho - \bar{\rho}) - 2d^2(1 - q^2/\rho^2) = 0. \tag{33}$$

From this quadratic equation in ρ', we can get ρ' as a function of ρ and $\bar{\rho}$. In writing the answer, we replace ρ' by ρ_{i+1}, ρ by ρ_i, and $\bar{\rho}$ by ρ_{i-1} to obtain

$$\rho_{i+1} = -\rho_i + [4\rho_i^2 - 4\rho_i(\rho_{i-1} - \rho_i) + 2d^2(1 - q^2/\rho_i^2) - (\rho_i - \rho_{i-1})^2]^{\frac{1}{2}}, \tag{34}$$

where the other solution of the equation is discarded because it gives a negative ρ' (or ρ_{i+1}) value.

At the start of the calculation, we identify ρ_{-1} with ρ_1 by "symmetry with respect to reflection at the throat." Then in this special case, as in the general situation, (34) supplies a single equation to determine a single unknown, the characteristic dimension of the first sphere. We fix the dimensions of the smallest sphere by identifying its *area* with the area of the throat as given by the exact analysis. Thus, we write

$$\rho_0^2 = [1 + (1 - \alpha^2)^{\frac{1}{2}}]^2 m^2 \tag{35}$$

TABLE I. Schwarzschild and Reissner–Nordstrøm geometry in Regge calculus, icosahedral method. a_i, edge length of triangular prism; b_i, its height is taken to be $b_i = \gamma a_i$ with $\gamma = 0.1$; $\epsilon(b_i)$, deficit angle associated with b_i; $(a_{i+1} - a_i)/b_i$, a measure of the increment of surface area relative to the increase in radial separation.

	$q = 0$			$q = 0.8m$		
Number of the surface	a_i	$\epsilon(b_i)$ (rad)	$\frac{a_{i+1} - a_i}{b_i}$	a_i	$\epsilon(b_i)$ (rad)	$\frac{a_{i+1} - a_i}{b_i}$
0	1.000	1.046	0.0036	1.0000	1.047	0.0027
1	1.0036	1.039	0.0108	1.0027	1.042	0.0082
2	1.0145	1.024	0.0179	1.0109	1.034	0.0134
3	1.0326	1.003	0.0247	1.0245	1.021	0.0188
5	1.0911	0.943	0.0374	1.0689	0.985	0.0291
10	1.3820	0.735	—	1.2926	0.838	—
—	—	—	—	—	—	—
40	18.9143	0.0527	1.0310	15.3019	0.0790	1.0140
41	20.8644	0.0477	1.0330	16.8634	0.0718	1.0234
48	41.6396	0.0239	1.0423	33.4996	0.0363	1.0375
Euclidean	a_i	0	1.050	a_i	0	1.050

and find $\rho_0 = 2m$ for the Schwarzschild case ($\alpha = q/m = 0$) and $\rho_0 = 1.6m$ for the case $\alpha = 0.8$ (Reissner–Nordstrøm geometry).

III. RESULTS

The 3-geometry generated by the icosahedral method is listed at selected points in Table I. We have used a value of $\gamma = b_i/a_i = 0.1$. The quantity $\epsilon(b_i)$ in Table I is the deficit angle of the bone b_i in radians; it measures the deviation of the geometry from flatness.

Many features of the geometry can be observed. Near the throat the base and summit of the triangular block have edges only slightly different from each other. Hinging on the bone b are five dihedral angles each of which is a little larger than $\frac{1}{3}\pi$. Therefore the deficit angle of b_i is close to $\frac{1}{3}\pi$ near the throat. The space near the throat is indeed far from being flat!

Far away from the throat we approach Euclidean geometry. In flat Euclidean space the edges a_i, b_i, and a_{i+1} (regardless whether i has the value 48 or any other value!) satisfy

$$(a_{i+1} - a_i)/b_i = 1.050.$$

It is seen from Table I that the quantity $(a_{i+1} - a_i)/b_i$ at the end of the 48th construction is very close to this Euclidean value. Consequently, space far away from the throat is asymptotically flat. This is further indicated by the smallness of the deficit angle $\epsilon(b_i)$ in this region.

The expression (34) obtained for the continuum method can be considered already as the solution for the 3-geometry under consideration since it is an *explicit* expression relating the surface areas and their separations. However, in order to exhibit the features of the geometry and to compare the results with the

TABLE II. Schwarzschild and Reissner–Nordstrøm 3-geometries in the continuum method according to Regge calculus. $4\pi\rho_i^2$ is the area of the ith spherical surface. Sum of all the separations d starting from the throat gives the geodesic distance measured from the throat up to that surface. θ_{23} measures how much the space deviates from flat space. Units: ρ_i and d in units of m; θ_{23} in rad/m^2.

	$q = 0$			$q = 0.08m$		
Number of the surface	ρ_i	Sum of d starting from throat	θ_{23}	ρ_i	Sum of d starting from throat	θ_{23}
0	2.0000	0.0		1.6000	0.0	
100	2.0013	0.1	0.2495	1.6012	0.1	0.3898
200	2.0050	0.2	0.2481	1.6047	0.2	0.3875
900	2.0998	0.9	0.2160	1.6939	0.9	0.3339
1000	2.1228	1.0	0.2091	1.7156	1.0	0.3223
—	—	—	—	—	—	—
29 500	26.5857	29.5	0.000106	26.0822	29.5	0.000111
30 000	27.0667	30.0	0.000102	26.5631	30.0	0.000105

TABLE III. Absolute values of errors in the icosahedral method and the continuum method.

Discrepancy, Regge calculus vs Ricci calculus		At throat in per cent	At infinity in per cent	At worst place, in per cent	Location u of this worst place
Volume from throat to a given surface	icosahedron method	1	8.63	8.63	∞
	continuum method	0.1	0.005	0.02	10
Geodesic distance from throat to given surface	icosahedron method	< 0.01	2.90	2.90	∞
	continuum method	1.00	0.003	1.00	1
Rotation factor θ_{23}	icosahedron method	0.1	9.0	9.0	∞
	continuum method	0.001	0.01	0.05	5

exact solution, one proceeds to generate the 3-geometry step by step with a computer.

In our calculation, the separation between successive shells is set equal to $d = 0.001m$. We use double precision in the machine calculations. The results of the calculations are listed for selected points in Table II. The important features of the geometry can be readily observed.

To compare the triangulation quantitatively with the corresponding exact curved geometry, we identify each spherically symmetric surface with a coordinate u by equating their surface areas. The accuracies for both the Schwarzschild and Reissner–Nordstrøm geometries are very close so that it is not necessary to distinguish the two cases. The discrepancies in our approximate geometry versus the exact solution is listed in Table III for both the icosahedral method and the continuum method. Except for the measurement of geodesic distance from the throat, the continuum method gives very small fractional error in all the quantities. However, the error for the icosahedral method is particularly large for the asymptotically flat region. This arises because there is a significant degree of mismatch when an icosahedron is used to represent a sphere.

In conclusion, the Regge skeleton calculus offers a workable way to determine—or specify—an initial 3-geometry. There is every reason to believe that it will prove equally useful as a way to trace out the dynamical evolution of geometry with time in accordance with Einstein's equations. Thus one can treat situations which in practical terms are beyond the reach of analytical methods. For example, when a strong gravitational pulse of nearly spherical form implodes, the geometry will become very strongly curved. In consequence, the outgoing pulse may differ drastically in pulse shape from the ingoing pulse; or a gravitational geon, stable against the loss of individual gravitons, may undergo collective gravitational collapse; or a model universe itself may undergo gravitational collapse.[4] The Regge calculus puts these and scores of other problems within the reach of analysis. Any desired level of accuracy can be obtained by sufficiently fine subdivision of the space-time region under consideration. Finally, the analysis offers—by way of its numbered building blocks—a practical way of *displaying* the results of such calculations. Therefore, the skeleton calculus, with its universal workability, makes one recognize more than ever the truly dynamic character of geometry and the wealth of problems opened up to examination by Einstein's theory.

ACKNOWLEDGMENTS

The author wishes to thank Professor J. A. Wheeler for suggesting this problem, for his guidance,and for contributions to the wording of the present revised text, many paragraphs of which are in Professor Wheeler's own words. He also wishes to thank Professor C. W. Misner for stimulating discussions.

APPENDIX: FORMULAS FOR THE COSINE FUNCTION BETWEEN HYPERPLANES

Formulas for the sine function of angles between hyperplanes have been given in detail by Wheeler.[4] The evaluation of another trigonometric function enables one to determine uniquely the quadrants in which the angles lie. The cosine of the dihedral angle between the planes pqr and pqs is given by[8]

$$\cos(pqr, pqs) = -D(pqr, pqs)/[D(pqr)D(pqs)]^{\frac{1}{2}}, \tag{A1}$$

where

$$D(pqr, pqs) = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & qp^2 & rp^2 \\ 1 & pq^2 & 0 & rq^2 \\ 1 & ps^2 & qs^2 & rs^2 \end{vmatrix} \tag{A2}$$

and

$$D(pqr) \equiv D(pqr, pqr). \tag{A3}$$

Here $pq^2 = qp^2$ is the square of the distance between points p and q.

The cosine of the hyperangle between the hyperplanes *pqrs* and *pqrt* is given by

$$\cos(pqrs, pqrt) = \Delta(pqrs, pqrt)/[\Delta(pqrs)\Delta(pqrt)]^{\frac{1}{2}}, \tag{A4}$$

where

$$\Delta(pqrs, pqrt) = D(pqr)D(pqs, pqt) - D(pqr, pqt)D(pqr, pqs)$$

and

$$\Delta(pqrs) = \Delta(pqrs, pqrs).$$

The cosine of the hyperangle between the hyperplanes *pqrst* and *pqrsu* in a 5-dimensional Euclidean space is

$$\cos(pqrst, pqrsu) = \Delta(pqrst, pqrsu)/[\Delta(pqrst)\Delta(pqrsu)]^{\frac{1}{2}}, \tag{A5}$$

where

$$\Delta(pqrst, pqrsu) = \Delta(pqrs)\Delta(pqrt, (pqru) - \Delta(pqrs, pqru)\Delta(pqrs, pqrt)$$

and

$$\Delta(pqrst) = \Delta(pqrst, pqrst).$$

* Research jointly sponsored by Princeton University, and by the U.S. Atomic Energy Commission under contract with the Union Carbide Corp.

† Based in part on the A.B. senior thesis of the author, Princeton University, 1961.

‡ Present address.

[1] T. Regge, Nuovo Cimento **17**, 558 (1961).

[2] J. Weber and J. A. Wheeler, Rev. Mod. Phys. **29**, 509 (1957); D. Brill, Ann. Phys. (N.Y.) **7**, 466 (1959); H. Araki, *ibid.* 456 (1959); J. A. Wheeler, Rev. Mod. Phys. **33**, 63 (1961).

[3] J. A. Wheeler, Phys. Rev. **97**, 511 (1955); see also J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).

[4] J. A. Wheeler, in *Relativity, Groups, and Topology*, edited by C. DeWitt and B. DeWitt, (Gordon and Breach, New York, 1964).

[5] Y. Fourès-Bruhat, Acta. Math. **88**, 141 (1952); J. Ratl. Mech. Anal. **4**, 951 (1956).

[6] See, e.g., L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J. 1926), p. 90.

[7] C. Misner, Phys. Rev. **118**, 1110 (1960).

[8] See, e.g., L. M. Blumenthal, *A Modern View of Geometry* (Freeman, San Francisco, 1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Introduction to Multiple Asymptotic Series with an Application to Elastic Scattering

KENNETH D. SHERE*
Department of Mathematics, University of Nebraska, Lincoln, Nebraska

(Received 26 June 1970)

The concept of asymptotic power series is extended to multiple series of the form

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{m,n}x^{-n}e^{-\lambda mx}.$$

This theory is then applied to elastic scattering from the Yukawa potential. An asymptotic series for the solution at large distances is obtained which is valid for arbitrary wavenumber and angular momentum.

1. INTRODUCTION

Several areas of mathematical physics give rise to differential equations whose coefficients are expandable in a double series in terms of $x^{-n}e^{-m\lambda x}$ ($n, m = 0, 1, 2, \cdots$; λ is a complex constant with positive real part). An example for $\lambda = 1$ occurs in the study of acoustic gravity waves.[1] Other examples [Ref. 2, pp. 1067 and 1082; Ref. 3, Appendix B and Eq. (23)] occur in the study of elastic scattering. In particular, the elastic scattering of particles from a Yukawa potential is governed by the radial Schrödinger equation

$$\frac{d^2u_q}{dr^2} + \{k^2 + [q(q+1)/r^2] - \mu e^{-\lambda r}/\lambda r\}u_q = 0. \tag{1.1}$$

If an asymptotic power series solution of (1.1) as $r \to \infty$ were obtained according to the usual theory of Poincaré, it would be observed that the Yukawa potential $e^{-\lambda r}/\lambda r \sim O\{r^{-n}\}$ and concluded that the asymptotic behavior of (1.1) is the same as the asymptotic behavior of

$$\frac{d^2u_q}{dr^2} + \{k^2 + [q(q+1)/r^2]\}u_q = 0. \tag{1.2}$$

The general solution of (1.2) can be given as a linear combination of spherical Bessel functions of the first and second kind.

In Sec. 2, multiple asymptotic power series are defined and the basic properties of asymptotics are established. In Sec. 3, this theory is applied to (1.1),

The cosine of the hyperangle between the hyperplanes *pqrs* and *pqrt* is given by

$$\cos(pqrs, pqrt) = \Delta(pqrs, pqrt)/[\Delta(pqrs)\Delta(pqrt)]^{\frac{1}{2}}, \tag{A4}$$

where

$$\Delta(pqrs, pqrt) = D(pqr)D(pqs, pqt) - D(pqr, pqt)D(pqr, pqs)$$

and

$$\Delta(pqrs) = \Delta(pqrs, pqrs).$$

The cosine of the hyperangle between the hyperplanes *pqrst* and *pqrsu* in a 5-dimensional Euclidean space is

$$\cos(pqrst, pqrsu) = \Delta(pqrst, pqrsu)/[\Delta(pqrst)\Delta(pqrsu)]^{\frac{1}{2}}, \tag{A5}$$

where

$$\Delta(pqrst, pqrsu) = \Delta(pqrs)\Delta(pqrt, (pqru) - \Delta(pqrs, pqru)\Delta(pqrs, pqrt)$$

and

$$\Delta(pqrst) = \Delta(pqrst, pqrst).$$

* Research jointly sponsored by Princeton University, and by the U.S. Atomic Energy Commission under contract with the Union Carbide Corp.

† Based in part on the A.B. senior thesis of the author, Princeton University, 1961.

‡ Present address.

[1] T. Regge, Nuovo Cimento **17**, 558 (1961).

[2] J. Weber and J. A. Wheeler, Rev. Mod. Phys. **29**, 509 (1957); D. Brill, Ann. Phys. (N.Y.) **7**, 466 (1959); H. Araki, *ibid.* 456 (1959); J. A. Wheeler, Rev. Mod. Phys. **33**, 63 (1961).

[3] J. A. Wheeler, Phys. Rev. **97**, 511 (1955); see also J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).

[4] J. A. Wheeler, in *Relativity, Groups, and Topology*, edited by C. DeWitt and B. DeWitt, (Gordon and Breach, New York, 1964).

[5] Y. Fourès-Bruhat, Acta. Math. **88**, 141 (1952); J. Ratl. Mech. Anal. **4**, 951 (1956).

[6] See, e.g., L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J. 1926), p. 90.

[7] C. Misner, Phys. Rev. **118**, 1110 (1960).

[8] See, e.g., L. M. Blumenthal, *A Modern View of Geometry* (Freeman, San Francisco, 1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Introduction to Multiple Asymptotic Series with an Application to Elastic Scattering

KENNETH D. SHERE*
Department of Mathematics, University of Nebraska, Lincoln, Nebraska

(Received 26 June 1970)

The concept of asymptotic power series is extended to multiple series of the form

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{m,n}x^{-n}e^{-\lambda mx}.$$

This theory is then applied to elastic scattering from the Yukawa potential. An asymptotic series for the solution at large distances is obtained which is valid for arbitrary wavenumber and angular momentum.

1. INTRODUCTION

Several areas of mathematical physics give rise to differential equations whose coefficients are expandable in a double series in terms of $x^{-n}e^{-m\lambda x}$ ($n, m = 0, 1, 2, \cdots$; λ is a complex constant with positive real part). An example for $\lambda = 1$ occurs in the study of acoustic gravity waves.[1] Other examples [Ref. 2, pp. 1067 and 1082; Ref. 3, Appendix B and Eq. (23)] occur in the study of elastic scattering. In particular, the elastic scattering of particles from a Yukawa potential is governed by the radial Schrödinger equation

$$\frac{d^2u_q}{dr^2} + \{k^2 + [q(q+1)/r^2] - \mu e^{-\lambda r}/\lambda r\}u_q = 0. \tag{1.1}$$

If an asymptotic power series solution of (1.1) as $r \to \infty$ were obtained according to the usual theory of Poincaré, it would be observed that the Yukawa potential $e^{-\lambda r}/\lambda r \sim O\{r^{-n}\}$ and concluded that the asymptotic behavior of (1.1) is the same as the asymptotic behavior of

$$\frac{d^2u_q}{dr^2} + \{k^2 + [q(q+1)/r^2]\}u_q = 0. \tag{1.2}$$

The general solution of (1.2) can be given as a linear combination of spherical Bessel functions of the first and second kind.

In Sec. 2, multiple asymptotic power series are defined and the basic properties of asymptotics are established. In Sec. 3, this theory is applied to (1.1),

and it is shown that the terms neglected by the usual Poincaré theory have a large effect at finite points, especially for small λ.

2. DEFINITION AND PROPERTIES

Let S be a sector of infinity of the complex plane centered on the real line with central angle $2\alpha \in (0, \frac{1}{2}\pi)$. Also suppose that $(x_0, \infty) \subset S$ for some positive number x_0. The function $f(x)$ defined on S is asymptotic to the series

$$\sum_{n=0}^{\infty} a_n x^{-n} e^{-\lambda m x}$$

in the Poincaré sense as $x \to \infty$ in S if

$$f(x) = \sum_{n=0}^{N} a_n x^{-n} e^{-\lambda m x} + o(x^{-N} e^{-\lambda m x})$$

as $x \to \infty$ in S for each fixed N. We denote this by

$$f(x) \sim \sum_{n=0}^{\infty} a_n x^{-n} e^{-\lambda m x} \quad (x^{-n} e^{-\lambda m x}\colon n = 0, 1, \cdots).$$

It is convenient to define the class of functions $Am(\lambda) = [f(x) \mid f(x)$ *has an asymptotic expansion in* S *as* $|x| \to \infty$ *with respect to* $(x^{-n} e^{-\lambda m x}\colon n = 0, 1, 2, \cdots)]$.

Define P to be the set of p-tuples over the nonnegative integers and impose on P the reverse lexicographic order $\leq$, i.e., if $N = (n_1, \cdots, n_p)$ and $M = (m_1, \cdots, m_p)$ and q is the largest integer such that $m_i \neq n_i$, then $M < N$ whenever $m_q < n_q$ and $M = N$ if q does not exist.

Definition 2.1: The sequence of functions $[\phi_N(x)\colon N \in P]$ defined on an unbounded region R of the complex plane is called an asymptotic sequence (or multiple asymptotic sequence) if $\phi_N = o(\phi_M)$ as $|x| \to \infty$ in R whenever $M < N$ uniformly in $N \in P$.

In this article we restrict our attention to "power" sequences of the form $(x^{-n} e^{-\lambda m x}\colon n, m = 0, 1, 2, \cdots)$ in the region $R \equiv S$. Extension of the following work to more general regions R and to asymptotic sequences of analytic function $\{\phi_N\}$ in S, having the property that $|\phi_N| > 0$ for all $x \in R$, $|x| > x_0$ for some x_0, is straightforward.

Suppose that the not necessarily convergent, formal double sum

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} \tag{2.1}$$

is given where $\{a_{m,n}\}$ is a sequence of constants. Define

$$f_m(x) \equiv \sum_{n=0}^{\infty} a_{mn} x^{-n} (1 - \exp(-f_{mn} x^2)) e^{-\lambda m x} \tag{2.2}$$

where f_{mn} is a nonnegative constant

$$\begin{aligned} f_{mn} &\equiv (\tfrac{1}{2}) \min\{1, 1/|a_{mn} n!\, m!|\}\colon && a_{mn} \neq 0, \\ &\equiv 0\colon && a_{mn} = 0. \end{aligned} \tag{2.3}$$

It can be shown by using Ritt's method[4] that the series $e^{\lambda m x} f_m(x)$ converges uniformly and absolutely in S (the proof of this statement is contained in the demonstration of Theorem 2.4 below). It has been shown by Ritt that

$$f_m(x) \sim \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} \quad (x^{-n} e^{-\lambda m x}\colon n = 0, 1, 2, \cdots).$$

Definition 2.2: A function $F(x)$ defined in S is said to be asymptotic to (2.1) as $|x| \to \infty$ in S, and is denoted

$$F(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} \tag{2.4}$$

if:

(a) $\mathcal{F}_m \equiv \left[f(x)\colon f(x) \sim \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} (x^{-n} e^{-\lambda m x})\right]$ for each fixed $m = 0, 1, 2, \cdots$;

(b) $\mathcal{F}_m^* \equiv [f(x) \in \mathcal{F}_m\colon f(x) - f_m(x) \sim o(x^{-n} e^{-\lambda \mu x})$ for each fixed $\mu = m, m+1, \cdots]$;

(c) $F(x) \in \mathcal{F}_0$;

(d) $F(x) - \sum_{\mu=0}^{m} f_\mu^* \in \mathcal{F}_{m+1}$ for any $f_\mu^* \in \mathcal{F}_\mu^*$.

The asymptotic symbols of (a) and (b) are taken in the Poincaré sense. The dependence on the order of summation of (2.4) is consistent with the dependence on the order of summation in the usual theory.

The following theorems and remarks show that Definition 2.2 is consistent with the properties normally expected of asymptotic series.

Theorem 2.1: If (2.1) is a convergent sum and

$$F(x) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x},$$

then

$$F(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x}.$$

Proof: Set

$$f_m^* = \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x}$$

and observe that $f_m^* \in \mathcal{F}_m^*$. Clearly, $F \in \mathcal{F}_0$ and

$$F(x) - \sum_{\mu=1}^{m} f_m^*(x) = \sum_{\mu=m+1}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda \mu x} \in \mathcal{F}_{m+1}.$$

The standard addition and uniqueness theorems are straightforward. The multiplication theorem is proven below:

Theorem 2.2: Let

$$F(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x}$$

and

$$G(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} b_{mn} x^{-n} e^{-\lambda m x},$$

then

$$H(x) \equiv F(x)G(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_{mn} x^{-n} e^{-\lambda m x},$$

where

$$c_{mn} \equiv \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} a_{m-\mu, n-\nu} b_{\mu\nu}.$$

Proof: By the usual theory $H(x) \in \mathcal{H}_0$. Since the series for $f_0(x)$, $g_0(x)$, and $h_0(x)$ are uniformly convergent,

$$\begin{aligned} f_0(x)g_0(x) &= \sum_{n=0}^{\infty} \Bigg(\sum_{\nu=0}^{n} a_{0,n-\nu} b_{0,\nu} \{1 - \exp(-f_{0,n-\nu} x^2) \\ &\quad - \exp(-g_{0,\nu} x^2) + \exp[-(f_{0,n-\nu} + g_{0,\nu})x^2]\} \Bigg) x^{-n} \\ &= h_0(x) + o[\exp(-\tfrac{1}{2}x^2)]. \end{aligned}$$

Hence $f_0(x)g_0(x) \in \mathcal{H}_0^*$.

Since $F(x) - f_0(x)$ and $G(x) - g_0(x) \in A_1$ and since $f_0(x)$ and $G(x) \in A_0$, $[F(x) - f_0(x)]G(x)$ and $f_0(x)[G(x) - g_0(x)] \in A_1$ (cf. Ref. 5, p. 18). Consequently,

$$\begin{aligned} &[F(x) - f_0(x)]G(x) + f_0(x)[G(x) - g_0(x)] \\ &\qquad = H(x) - f_0(x)g_0(x) \in A_1. \end{aligned} \tag{2.5}$$

By substituting into (2.5) the appropriate asymptotic expansions and performing the necessary calculations, it is seen that $H(x) - f_0(x)g_0(x) \in \mathcal{H}_1$.

Similarly it is observed that $f_\mu(x)g_{m-\mu}(x) \in A_m$ and $\sum_{\mu=0}^{m} f_\mu(x)g_{m-\mu}(x) \in \mathcal{H}_m^*$. Condition (d) has been demonstrated for $m = 0$. Assume that this condition is satisfied for $m = 1, 2, 3, \cdots, M-1$. Furthermore, we note that

$$f_\mu\left(G - \sum_{m=0}^{M-\mu} g_m\right) \in A_{m+1} \tag{2.6a}$$

and

$$\left(F - \sum_{m=0}^{M} f_m\right) G \in A_{M+1}. \tag{2.6b}$$

Consequently,

$$\begin{aligned} &\left(F - \sum_{m=0}^{M} f_m\right) G - \sum_{\mu=0}^{M} f_\mu \left(G - \sum_{m=0}^{M-\mu} g_m\right) \\ &\qquad = H(x) - \sum_{\mu=0}^{M} \sum_{m=0}^{M-\mu} f_\mu g_m \\ &\qquad = H(x) - \sum_{m=0}^{M} \sum_{\mu=0}^{m} f_\mu g_{m-\mu} \in A_{M+1}. \end{aligned} \tag{2.7}$$

By substituting the appropriate asymptotic series into (2.7) [e.g.,

$$\sum_{n=0}^{\infty} b_{0,n} x^{-n} \text{ for } G$$

and

$$\sum_{n=0}^{\infty} a_{M+1,n} x^{-n} e^{-\lambda(M+1)x} \quad \text{for} \quad F - \sum_{m=0}^{M} f_m \Bigg]$$

and performing the necessary calculations, which are valid because these operations are being done for "usual" asymptotic series, it is seen that

$$H(x) - \sum_{m=0}^{M} \sum_{\mu=0}^{m} f_\mu g_{m-\mu} \in \mathcal{H}_{M+1}.$$

By induction the theorem is concluded.

Theorem 2.3: Let $f(x)$ be differentiable in S and suppose that

$$f(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} \quad \text{as} \quad |x| \to \infty \text{ in } S;$$

then

$$\begin{aligned} h(x) \equiv f'(x) &\sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} [-\lambda m a_{mn} - (n-1)a_{m,n-1}] x^{-n} e^{-\lambda m x} \end{aligned}$$

as $|x| \to \infty$ in S, where $a_{m,-1} \equiv 0$.

Proof: Ritt demonstrated that the series for $[e^{\lambda m x} f_m(x)]'$ is uniformly convergent whenever the series for $\exp(\lambda m x) f_m(x)$ converges uniformly. Hence

$$f_m'(x) \sim \sum_{n=0}^{\infty} [-\lambda m a_{mn} - (n-1)a_{m,n-1}] x^{-n} e^{-\lambda m x}$$

and $f_m'(x) \in \mathcal{H}_m^*$. For each fixed m,

$$f(x) - \sum_{\mu=0}^{m} f_\mu(x) \sim f_{m+1} \in A_{m+1}, \tag{2.8}$$

so that

$$f'(x) - \sum_{\mu=0}^{m} f_\mu'(x) \sim f_{m+1}' \in A_{m+1}. \tag{2.9}$$

Since $f_{m+1}' \in \mathcal{H}_{m+1}$, we are done. In this proof we use the fact that S is not a ray (i.e., $\alpha \neq 0$). This is necessary for (2.9) to follow from (2.8) (Ref. 6, p. 38).

Having demonstrated the primary algebraic and analytic properties, it remains for us to demonstrate existence.

Theorem 2.4: Given an arbitrary sequence of numbers $(a_{mn}: m, n = 0, 1, 2, 3, \cdots)$, there is an analytic function $f(x)$ in S so that

$$f(x) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} \quad \text{as} \quad |x| \to \infty \text{ in } S. \tag{2.10}$$

Proof (Ritt's Method): Define

$$u_m(x) \equiv \sum_{n=0}^{\infty} (a_{mn} n!\, m!/n!\, m!) x^{-n} [1 - \exp(-f_{mn} x^2)], \tag{2.11}$$

where f_{mn} is defined in (2.3). Using the inequality $|A(1 - e^{-B})| < 2$ if $|B| < 1$ and $|B| \cdot |A| < 1$ or if $\mathcal{R}eB > 0$ and $|B| \cdot |A| < 1$, we have

$$|x^{-n} a_{mn} m!\, n! \, [1 - \exp(-f_{mn} x^2)]| < 2 |x|^{-n+2}.$$

Thus

$$|u_m(x)| < \sum_{n=0}^{\infty} |x|^{-n+2}/n!\, m! = (|x|^2/m!) e^{1/|x|}.$$

Let

$$f(x) \equiv \sum_{m=0}^{\infty} u_m(x) e^{-\lambda m x}$$

and observe that

$$\begin{aligned} |f(x)| &\leq \sum_{m=0}^{\infty} |u_m(x)| \, |\exp(-\lambda m x)| \\ &< |x|^2 \, e^{1/|x|} \sum_{m=0}^{\infty} |e^{-\lambda m x}|/m! \\ &\leq |x|^2 \exp(e^{|\lambda x|} + 1/|x|). \end{aligned}$$

Consequently,

$$f(x) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} x^{-n} e^{-\lambda m x} [1 - \exp(-f_{mn} x^2)] \tag{2.12}$$

is an absolutely and uniformly convergent series in S. It is obvious [from $f_m(x) = u_m(x) \exp(-\lambda m x)$] that the asymptotic behavior (2.10) of $f(x)$ follows. Furthermore, $f(x)$ is analytic in S because the limit of a uniformly convergent sequence of analytic functions is analytic.

3. ELASTIC SCATTERING FROM A YUKAWA POTENTIAL

The elastic scattering of particles is governed by the Schrödinger equation

$$\frac{\hbar^2}{2m} \Delta\Phi + i\hbar \frac{\partial \Phi}{\partial t} - V(r)\Phi = 0, \tag{3.1}$$

where Φ is a wavefunction, $\hbar$ is Planck's constant, m is the particle mass, and $V(r)$ is the potential energy.[7] The Yukawa potential is

$$V(r) \equiv (\hbar^2/2m) U(r) \equiv \lambda(\hbar^2/2m\mu r) e^{-\mu r}, \tag{3.2}$$

where λ/μ characterizes the depth and μ the range of the potential. For a stationary state of energy E, define $\Phi \equiv e^{-iEt/\hbar}\psi$. From (3.1),

$$(\hbar^2/2m)\Delta\psi + (E - V)\psi = 0$$

or

$$\Delta\psi + (k^2 - U)\psi = 0, \tag{3.3}$$

where $k \equiv (2mE/\hbar^2)^{\frac{1}{2}}$ is the wavenumber.

The quantity E is the total energy and (3.3) is recognized as a reduced wave equation (cf. Ref. 8, pp. 313ff) in spherical coordinates. Separating variables (cf. Ref. 2, p. 1067; Ref. 9, p. 216), we have

$$\psi(r, \theta, \phi) = S(\theta, \phi) u_q(r)/r,$$

where, in terms of spherical harmonics,

$$\begin{aligned} S_{pq}(\theta, \phi) = A_{pq}(\cos p\phi) P_q^p(\cos\theta) \\ + B_{pq}(\sin p\phi) P_q^p(\cos\theta). \end{aligned}$$

Also $u_q(r)$ must satisfy the radial Schrödinger equation

$$\frac{d^2 u_q}{dr^2} + [k^2 - q(q+1) r^{-2} - U(r)] u_q = 0. \tag{3.4}$$

Expand $u_q(r)$ in terms of the asymptotic sequence $(r^{-n} e^{-\mu m r}: n, m = 0, 1, 2, \cdots)$ as $|r| \to \infty$ in S, i.e.,

$$u_q(r) \sim \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{q,m,n} r^{-n} e^{-\mu m r}, \tag{3.5}$$

where $a_{q,m,n}$ is a complex constant. Substituting (3.5) into (3.4) and proceeding with the calculations, we obtain, upon setting the coefficients of $r^{-n} e^{-\mu m r}$ to zero, the recurrence relation

$$\begin{aligned} (\mu^2 m^2 + k^2) a_{q,m,n} + 2\mu m(n-1) a_{q,m,n-1} \\ + [(n-1)(n-2) - q(q+1)] a_{q,m,n-2} \\ - (\lambda/\mu) a_{q,m-1,n-1} = 0, \end{aligned} \tag{3.6}$$

where $a_{q,-1,n} \equiv a_{q,m,-1} \equiv a_{q,m,-2} \equiv 0$.

Upon inspection of (3.6), it is seen that no nontrivial solution of (3.4) admits an asymptotic expansion of the form (3.5) unless $k = \pm i\mu\tilde{m}$ for some $m = \tilde{m}$. These values of k are branch points of the scattering matrix (Ref. 10, pp. 362ff) in the complex

k plane. Some results for these values of k are

$$a_{q,m,n} = 0 \quad \text{for} \quad m < \tilde{m} \quad \text{and all } n, q, \tag{3.7a}$$

$$a_{q,\tilde{m},n} = \{[n(n-1) - q(q+1)] \times [(n-1)(n-2) - q(q+1)] \cdots q(q+1)a_{q,\tilde{m},0}\}/(2\mu\tilde{m})^n n! \quad \text{for} \quad n = 1, 2, \cdots, \tag{3.7b}$$

for q a natural number and $n \geq q+1$, $a_{q,\tilde{m},n} = 0$,

$$\tag{3.7c}$$

$$a_{q,m,0} = 0 \quad \text{for all} \quad m > \tilde{m}, \tag{3.7d}$$

$$a_{q,\tilde{m}+p,1} = \lambda^p a_{q,\tilde{m},0}/\mu^{3p}(2\tilde{m}+1)(4\tilde{m}+4)\cdots(2p\tilde{m}+p^2) \quad \text{for} \quad p = 1, 2, \cdots. \tag{3.7e}$$

Note that $a_{q,\tilde{m},0}$ is a constant determined by some side condition; for example,

$$a_{q,\tilde{m},0} = \lim_{\substack{r\to\infty \\ r\in S}} u_q(r) \exp(\mu\tilde{m}r).$$

Since $\tilde{m} \geq 1$, we note that $u_q(r)$ satisfies the radiation condition $du_q(r)/dr - iku_q(r) = o(1/r^2)$ (Ref. 8, p. 316).

To obtain asymptotic formulas for other values of k, define

$$w_q(r) \equiv e^{-ikr}u_q(r). \tag{3.8}$$

Equation (3.4) becomes

$$d^2w_q/dr^2 + 2ik\,dw_q/dr - [q(q+1) + (\lambda/\mu r)e^{-\mu r}]w_q = 0. \tag{3.9}$$

Expanding w_q,

$$w_q(r) \sim \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} b_{q,m,n}r^{-n}e^{-\mu mr},$$

we obtain the recurrence relation

$$\mu m(\mu m - 2ki)b_{q,m,n} + 2(\mu m - ik)(n-1)b_{q,m,n-1} + [(n-1)(n-2) - q(q+1)]b_{q,m,n-2} - (\lambda/\mu)b_{q,m-1,n-1} = 0. \tag{3.10}$$

Whenever $ik \neq \mu m$ for all m, (3.10) has a nontrivial solution. Results for the exceptional case are given above. Some results for $k \neq -\frac{1}{2}i\mu m$ for all m are given below; when $k = -\frac{1}{2}i\mu\tilde{m}$ and $\tilde{m}$ is odd, similar results are obtainable.

Then

$$b_{q,0,n} = -[n(n-1) - q(q+1)] \times [(n-1)(n-2) - q(q+1)] \cdots q(q+1)] \times b_{q,0,0}/(2k)^n n!, \quad n = 1, 2, 3, \cdots. \tag{3.11a}$$

If q is an integer, $b_{q,0,n} = 0$ for $n = q+1, q+2, \cdots$,

$$b_{q,m,0} = 0 \quad \text{for} \quad m = 1, 2, \cdots, \tag{3.11b}$$

$$b_{q,m,1} = (\lambda/\mu)^m b_{q,0,0}/m!\,(\mu - 2ki)(2\mu - 2ki)\cdots(\mu m - 2ki), \quad \text{for} \quad m = 1, 2, 3, \cdots. \tag{3.11c}$$

From (3.8) we see that $u_q(r)$ satisfies the radiation condition when $\operatorname{Im} k > 0$. We note that

$$\lim_{r\to\infty} w_q(r) = b_{q,0,0}.$$

Observe that $b_{q,0,0} = 1$ is the Jost solution (cf. Ref. 11, p. 373). One of the nice features of this asymptotic series is that it demonstrates explicitly the manner in which the Yukawa potential influences the scattering; this influence is felt only in terms of $m \geq 1$. The terms with $m = 0$ are essentially an asymptotic representation of spherical Bessel functions.

In terms of incoming and outgoing waves we can write a formal solution of (3.4) as

$$u_q(r) = A[e^{-i(kr-\frac{1}{2}\pi q)}f_q(-k, r) + S_q^{(k)}e^{i(kr-\frac{1}{2}\pi q)}f_q(k, r)], \tag{3.12}$$

where $A = b_{q,0,0}$ and $f_q(+k, r) = e^{-\frac{1}{2}i\pi q}w_q(r)/b_{q,0,0}$. When solutions of the form (3.12) are obtained for regions of the origin, $\{S_q^{(k)}\}$ is the S matrix whose value is determined by the boundary condition at $r = 0$. In this expansion $\{S_q^{(k)}\}$ can be obtained by matching (3.12) to another symptotic solution which is valid near the origin.

ACKNOWLEDGMENTS

Thanks are extended to Professor P. Finkler and Professor J. Macek for their comments on Sec. 3 and to Professor A. Zechmann for his general comments.

* Present address: U.S. Naval Ordnance Laboratory, Mathematical Analysis Division, Silver Spring, Maryland 20910.

[1] K. Shere and S. Bowhill, "Gravity Waves in a Viscous Atmosphere," University of Illinois Aeronomy Report No. 31, 1969.
[2] P. Morse and H. Feshbach, *Methods of Theoretical Physics, Part II* (McGraw-Hill, New York, 1953).
[3] K. Omidvar, *Theory of the 2s and 2p Excitation of the Hydrogen Atom Induced by Electron Impact* (NASA TN D-2145, 1964).
[4] J. Ritt, Ann. Math. **18**, 18 (1916).
[5] A. Erdélyi, *Asymptotic Expansions* (Dover, New York, 1956).
[6] W. Wasow, *Asymptotic Expansions for Ordinary Differential Equations* (Interscience, New York, 1965).
[7] A. Baz', Ya. Zel'dovich, and A. Perelomov, *Scattering, Reactions, and Decay in Nonrelativistic Quantum Mechanics* (Israel Program for Scientific Translations, NASA TT F-510, 1969).
[8] R. Courant and D. Hilbert, *Methods of Mathematical Physics II* (Interscience, New York, 1962).
[9] J. Dettman, *Mathematical Methods in Physics and Engineering* (McGraw-Hill, New York, 1969), 2nd ed.
[10] D. Bates, Ed., *Quantum Theory. I. Elements* (Academic, New York, 1961).
[11] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Asymptotic Distribution of Eigenvalues of the Kernel in the Kirkwood-Riseman Integral Equation

SEIJI UKAI
Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Japan

(Received 2 June 1970)

An asymptotic formula for the eigenvalue is presented for the kernel appearing in the integral equation of the Kirkwood–Riseman type. The method used here provides a procedure of successive iteration for solving the inhomogeneous integral equation as well as the eigenvalue equation.

I. INTRODUCTION

In the present paper we discuss the asymptotic distribution of eigenvalues of the integral operator

$$K\varphi = \int_{-1}^{1} \frac{\varphi(x')}{|x - x'|^{\alpha}}\, dx', \quad |x| \leq 1, \tag{1}$$

for $0 < \alpha < 1$. This integral operator appears in the Kirkwood–Riseman equation[1]

$$\varphi(x) = f(x) + \lambda \int_{-1}^{1} \frac{\varphi(x')}{|x - x'|^{\alpha}}\, dx', \tag{2}$$

with negative λ, in connection with the theory of intrinsic viscosities and diffusion coefficients of flexible macromolecules.

The method used here is based on the observations that the integral kernel can be written as

$$\frac{1}{|x - x'|^{\alpha}} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha-1} e^{-|x-x'|t}\, dt, \tag{3}$$

where $\Gamma(\alpha)$ denotes the gamma function, and that $\exp(-|x - x'|\, t)$ is the Green's function of a second-order differential operator. As a consequence, Eq. (2) can be transformed into an integro-differential equation.

This kind of transformation has been used successfully in the theory of neutron transport equation for critical media by Mitsis[2] and Gibbs.[3] These authors have solved corresponding integro-differential equations by Case's singular eigenfunction expansion method[4] and, using the theory of singular integral equations,[5] have obtained, for expansion coefficients, Fredholm's integral equations of the second kind whose Neumann series are convergent.

Case's method is applicable also to our case, but we will take the Laplace transformation technique discussed in Ref. 6, which gives the final Fredholm integral equation more directly than Case's method. The integral equation is inhomogeneous even when we consider the eigenvalue problem of K—that is, even for the case $f(x) \equiv 0$. It will be found that the corresponding Neumann series converges for all real values of λ except for negative λ of small magnitude with $0 < \alpha < \alpha^*$, $\alpha^* \doteqdot 0.27$. The truncation of the Neumann series, therefore, gives approximate solutions both for the eigenvalue equation and for the inhomogeneous equation for almost all real values of λ.

We will consider the operator K in $L^2(-1, 1)$. Then K is compact[7] and positive[8] and hence has a denumerable set of positive eigenvalues λ_n, $n = 1, 2, 3, \cdots$, tending to infinity. [Here we call λ an eigenvalue when the equation $\varphi = \lambda K\varphi$ has nontrivial solutions, $\varphi \in L^2(-1, 1)$. Thus λ is the inverse of an eigenvalue in the ordinary sense.] Carlemann[9] has proved that the series

$$\sum_{n=1}^{\infty} \lambda_n^{-1/(1-\alpha)} \tag{4}$$

is divergent. The main aim of this paper is to show that

$$\lambda_n = \frac{\Gamma(\alpha) \sin[\frac{1}{2}(1-\alpha)\pi]}{\pi} \left(\frac{n - 1}{2}\, \pi + \frac{1 + \alpha}{8}\, \pi\right)^{1-\alpha} + O\left(\frac{1}{n^{1+\alpha}}\right), \tag{5}$$

for large n.

II. TRANSFORMATION OF THE EQUATION

Let us first define

$$u(x, t) = \int_{-1}^{1} e^{-|x-x'|t} \varphi(x')\, dx', \tag{6}$$

where $|x| \leq 1$ and $t \geq 0$. Since $\phi(x) \in L^2(-1, 1)$, $u(x, t)$ is twice differentiable for almost all $x \in (-1, 1)$ and infinitely many times differentiable with respect to t. Furthermore,

$$u_0(x) = \int_0^{\infty} t^{\alpha-1} u(x, t)\, dt \tag{7}$$

exists for almost all $x \in (-1, 1)$ and is equal to $K\varphi \in L^2(-1, 1)$ since the order of integration can be converted. Consequently, Eq. (2) reduces to

$$\varphi(x) = f(x) + [\lambda/\Gamma(\alpha)] u_0(x). \tag{8}$$

On the other hand, differentiating Eq. (6) twice with respect to x, we obtain

$$u'(x,t) = -t\left(e^{-xt}\int_{-1}^{x} e^{x't}\varphi(x')\,dx' - e^{xt}\int_{x}^{1} e^{-x't}\varphi(x')\,dx'\right) \quad (9)$$

and

$$u''(x,t) = t^2u(x,t) - 2t\varphi(x), \quad (10)$$

the last of which holds for almost all $x \in (-1, 1)$. Substituting Eq. (8) into Eq. (10), we get the integro-differential equation

$$u''(x,t) - t^2u(x,t) = -2tf(x) - \frac{2\lambda t}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}u(x,t)\,dt. \quad (11)$$

The boundary condition for this equation follows from Eqs. (6) and (9) as

$$u'(1,t) + tu(1,t) = 0, \quad (12a)$$

$$u'(-1,t) - tu(-1,t) = 0. \quad (12b)$$

Thus we should seek the solution $u(x,t)$ of Eq. (11) that satisfies the boundary conditions (12) and possesses the properties mentioned above. For this purpose, we Laplace-transform Eq. (11) assuming that $u(x,t) \equiv 0$ and $f(x) \equiv 0$ for $|x| > 1$. This procedure has been developed[6] for the 1-dimensional neutron transport equation.

Define the Laplace transforms

$$\tilde{u}(z,t) = \int_{-1}^{1} e^{-zx}u(x,t)\,dx, \quad (13)$$

$$\tilde{f}(z) = \int_{-1}^{1} e^{-zx}f(x)\,dx, \quad (14)$$

$$\tilde{u}_0(z) = \int_{-1}^{1} e^{-zx}u_0(x)\,dx = \int_0^\infty t^{\alpha-1}\tilde{u}(z,t)\,dt. \quad (15)$$

The last step in Eq. (15) is justified since $u_0 = K\varphi \in L^2(-1, 1)$ and the interchange of integration order is permitted. From Eq. (11), we now obtain

$$(z^2 - t^2)\tilde{u}(z,t) = -2t\tilde{f}(z) - [2\lambda t/\Gamma(\alpha)]\tilde{u}_0(z) + e^z[u'(-1,t) + zu(-1,t)] - e^{-z}[u'(1,t) + zu(1,t)], \quad (16)$$

whence, substituting the boundary conditions (12), we have

$$\tilde{u}(z,t) = -\frac{2t}{z^2 - t^2}\left(\tilde{f}(z) + \frac{\lambda}{\Gamma(\alpha)}\tilde{u}_0(z)\right) - e^{-z}\frac{u(1,t)}{z+t} + e^{z}\frac{u(-1,t)}{z-t} \quad (17)$$

and, in view of Eq. (15),

$$\tilde{u}_0(z) = -\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(z) + \frac{1}{\Lambda(z)}\left\{\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(z) - e^{-z}\Psi_{+1}(-z) - e^{z}\Psi_{-1}(z)\right\}, \quad (18)$$

where we have defined

$$\Lambda(z) = 1 - \frac{2\lambda}{\Gamma(\alpha)}\int_0^\infty \frac{t^\alpha}{t^2 - z^2}\,dt, \quad (19)$$

$$\Psi_{\pm1}(z) = \int_0^\infty \frac{t^{\alpha-1}u(\pm1,t)}{t - z}\,dt. \quad (20)$$

Equation (17), when coupled with Eq. (18), gives the Laplace transform of a solution of Eq. (11). To complete the solution, it is required to determine $u(\pm1, t)$ and subsequently $\Psi_{\pm1}(z)$.

III. INTEGRAL EQUATIONS FOR $\Psi_{\pm1}$

First of all, we shall examine the properties of the functions $\Lambda(z)$ and $\Psi_{\pm1}(z)$. The integral in $\Lambda(z)$ is calculated explicitly in Appendix A as

$$2\int_0^\infty \frac{t^\alpha}{t^2 - z^2}\,dt = \frac{\pi}{\sin[\frac{1}{2}(1-\alpha)\pi]} \times \begin{cases} e^{\frac{1}{2}\pi i(1-\alpha)}z^{\alpha-1}, & 0 < \arg z < \pi, \\ e^{\frac{3}{2}\pi i(1-\alpha)}z^{\alpha-1}, & \pi < \arg z < 2\pi. \end{cases} \quad (21)$$

Hence $\Lambda(z)$ is holomorphic in the complex z plane cut along the real axis. When z approaches a point t on the real axis from the above (+) and from the below (−) in the cut plane, $\Lambda(z)$ has definite limits $\Lambda^{\pm}(t)$ given by

$$\Lambda^{\pm}(t) = 1 - \frac{\lambda\pi}{\Gamma(\alpha)\sin[\frac{1}{2}(1-\alpha)\pi]} \times \begin{cases} e^{\pm\frac{1}{2}\pi i(1-\alpha)}t^{\alpha-1}, & t > 0, \\ e^{\mp\frac{1}{2}\pi i(1-\alpha)}|t|^{\alpha-1}, & t < 0. \end{cases} \quad (22)$$

Thus $\Lambda(z)$ is a sectionally holomorphic function with the line of discontinuity on the real axis. Furthermore, let us note that if $\lambda > 0$, $\Lambda(z)$ has two simple zeros $\pm z_0$ in the cut plane with

$$z_0 = \nu i, \quad \nu = \left(\frac{\lambda\pi}{\Gamma(\alpha)\sin[\frac{1}{2}(1-\alpha)\pi]}\right)^{1/(1-\alpha)}, \quad (23)$$

while it has no zeros if $\lambda < 0$.

The functions $\Psi_{\pm1}(z)$ are ordinary Cauchy integrals whose properties are discussed extensively in Ref. 5. They are sectionally holomorphic with the line of

discontinuity on the real positive axis and have definite limits $\Psi^{\pm}_{\pm 1}(t)$ for $z \to t \in (0, \infty)$. These follow from the fact that $u(\pm 1, t)$ possess continuous derivatives.

From what has been stated, we can now conclude that $\tilde{u}_0(z)$ given by Eq. (18) has singularity of discontinuity on the real axis and, if $\lambda > 0$, two simple poles at $z = \pm z_0$, while $\tilde{u}(z, t)$ defined by Eq. (17) has, in addition, two simple poles at $z = \pm t$. In the remaining region of the complex plane, they are analytic.

As seen from Eqs. (13) and (15), however, these two functions should, by definition, be entire functions of z. Consequently, the singularities mentioned above should be removable. The discontinuity on the real axis disappears if $\tilde{u}_0(z)$ of Eq. (18) satisfies the condition that $\tilde{u}_0^+(t) = \tilde{u}_0^-(t)$, $-\infty < t < \infty$, that is, if the equations

$$\frac{1}{\Lambda^+(t)}\left(\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(t) - e^{-t}\Psi^+_{+1}(-t) - e^{t}\Psi^+_{-1}(t)\right) = \frac{1}{\Lambda^-(t)}\left(\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(t) - e^{-t}\Psi^-_{+1}(-t) - e^{t}\Psi^-_{-1}(t)\right), \quad t > 0, \quad (24a)$$

$$\frac{1}{\Lambda^+(t)}\left(\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(t) - e^{-t}\Psi^+_{+1}(-t) - e^{t}\Psi^+_{-1}(t)\right) = \frac{1}{\Lambda^-(t)}\left(\frac{\Gamma(\alpha)}{\lambda}\tilde{f}(t) - e^{-t}\Psi^-_{+1}(-t) - e^{t}\Psi^-_{-1}(t)\right), \quad t < 0, \quad (24b)$$

hold. The simple poles at $z = \pm z_0$ become removable if the numerator in the right-hand side of Eq. (18) vanishes, namely, if

$$[\Gamma(\alpha)/\lambda]\tilde{f}(\pm z_0) - e^{\mp z_0}\Psi^-_{+1}(\mp z_0) - e^{\pm z_0}\Psi^-_{-1}(\pm z_0) = 0. \quad (25)$$

Of course, Eq. (25) is not needed in the case $\lambda < 0$. The simple poles at $z = \pm t$ of $\tilde{u}(z, t)$ necessarily vanish due to Eqs. (24). The proof of this assertion can be carried out in a manner similar to that in Ref. 6 and, therefore, is not given here.

Since $\overline{\Lambda^+(t)} = \Lambda^-(t)$, we may write

$$\Lambda^{\pm}(t) = \gamma(t)e^{\pm i\theta(t)}, \quad \gamma(t) = |\Lambda^{\pm}(t)|,$$
$$\theta(t) = \frac{1}{2i}\ln\frac{\Lambda^+(t)}{\Lambda^-(t)} = \tan^{-1}\frac{-\kappa \sin[\frac{1}{2}(1-\alpha)\pi]}{t^{1-\alpha} - \kappa\cos[\frac{1}{2}(1-\alpha)\pi]}, \quad (26)$$
$$\kappa = \frac{\lambda\pi}{\Gamma(\alpha)\sin[\frac{1}{2}(1-\alpha)\pi]},$$

for $t \geq 0$. Here we choose the branch of $\tan^{-1}$ such that $\theta(t)$ varies continuously in $0 < t < +\infty$ and tends to zero as t tends to infinity. Hence, $\theta(t) \lesseqgtr 0$ for $\lambda \gtreqless 0$. Now we can rewrite Eq. (24a) as

$$\Psi^{+}_{-1}(t) - e^{2i\theta(t)}\Psi^{-}_{-1}(t) = 2ie^{i\theta(t)}\sin\theta(t)\, e^{-t}\{e^{-t}\Psi_{+1}(-t) - [\Gamma(\alpha)/\lambda]\tilde{f}(t)\}, \quad t > 0, \quad (27a)$$

while, taking Eq. (22) into account and changing the variable $t \to -t$, we have, from Eq. (24b),

$$\Psi^{+}_{+1}(t) - e^{2i\theta(t)}\Psi^{-}_{+1}(t) = 2ie^{i\theta(t)}\sin\theta(t)\, e^{-t}\{e^{-t}\Psi_{-1}(-t) - [\Gamma(\alpha)/\lambda]\tilde{f}(-t)\}, \quad t > 0. \quad (27b)$$

Equations (27) turn to be the so-called inhomogeneous Hilbert problem if their right-hand sides are assumed to be known. In order to solve the problems, a particular solution of the corresponding homogeneous Hilbert problem

$$X^+(t) = e^{2i\theta(t)}X^-(t), \quad 0 < t < \infty, \quad (28)$$

is required. The solution which is suitable for our purpose[5] is

$$X(z) = Q(z)\exp\Gamma_0(z), \quad (29)$$

where

$$\Gamma_0(z) = \frac{1}{\pi}\int_0^{\infty}\frac{\theta(t)}{t - z}\,dt, \quad (30)$$

$$Q(z) = \begin{cases} z^{-1}, & \lambda > 0, \\ 1, & \lambda < 0. \end{cases} \quad (31)$$

Due to our choice of the branch of $\tan^{-1}$, $\theta(t) = O(t^{\alpha-1})$ for $t \gg 1$, so that $\Gamma_0(z)$ and subsequently $X(z)$ exist and are holomorphic in the plane cut along the real positive axis. The function $X(z)$ has no zeros in the cut plane, and, for $|z| \gg 1$,

$$X(z) = \begin{cases} O(|z|^{-1}), & \lambda > 0, \\ O(1), & \lambda < 0. \end{cases} \quad (32)$$

Near $z = t = 0$,

$$|X(z)| \leq \frac{\text{const.}}{|z|^{\frac{1}{2}(1-\alpha)}}, \quad |X^{\pm}(t)| \leq \frac{\text{const.}}{|t|^{\frac{1}{2}(1-\alpha)}}. \quad (33)$$

Moreover, it is useful to note the relation

$$X(z)X(-z) = \begin{cases} \dfrac{\Lambda(z)}{z_0^2 - z^2}, & \lambda > 0, \\ \Lambda(z), & \lambda < 0, \end{cases} \quad (34)$$

which is proved in Appendix B.

Substituting Eq. (28) into Eqs. (27) and using Plemelj's formula, we can obtain the solution of Eq. (27) in a usual manner.[5] For later convenience, we will write the solutions in the following forms:

$$\Psi_{\pm 1}(z) = \frac{X(z)}{\pi}\int_0^\infty \frac{h(t)e^{-t}}{t-z}\,\psi_{\mp 1}(t)\,dt + G_{\pm 1}(z) + a_{\pm 1}X(z)H(\lambda), \quad (35)$$

in which we have defined

$$\psi_{\pm 1}(t) = \left(\frac{e^{i\theta(t)}\sin\theta(t)}{X(-t)X^+(t)}\right)^{\frac{1}{2}} e^{-t}\Psi_{\pm 1}(-t), \quad (36)$$

$$h(t) = \left(\frac{X(-t)e^{i\theta(t)}\sin\theta(t)}{X^+(t)}\right)^{\frac{1}{2}}, \quad (37)$$

$$G_{\pm 1}(z) = -\frac{\Gamma(\alpha)X(z)}{\lambda\pi}\int_0^\infty \frac{e^{i\theta(t)}\sin\theta(t)}{X^+(t)}\,e^{-t}\tilde{f}(\mp t)\,\frac{dt}{t-z}, \quad (38)$$

while $a_{\pm 1}$ are constants to be determined later and

$$H(\lambda) = \begin{cases} 1, & \lambda > 0, \\ 0, & \lambda < 0, \end{cases} \quad (39)$$

that is, if $\lambda < 0$, the last term in Eqs. (35) vanishes.

From Eqs. (35) with $z = -t$, we now have a set of integral equations for $\Psi_{\pm 1}(-t)$ or, by the definitions of Eq. (36), for $\psi_{\pm 1}(t)$, which we write simply as

$$\psi_{\pm 1} = S\psi_{\mp 1} + g_{\pm 1}. \quad (40)$$

Here S denotes the integral operator defined as

$$S\psi = \frac{1}{\pi}\int_0^\infty \frac{h(t)h(t')e^{-(t+t')}}{t+t'}\,\psi(t')\,dt', \quad 0 \le t < \infty, \quad (41)$$

and

$$g_{\pm 1}(t) = h(t)\{G_{\pm 1}(-t)[X(-t)]^{-\frac{1}{2}} + a_{\pm 1}H(\lambda)\}e^{-t}. \quad (42)$$

In spite of its appearance, $h(t) \ge 0$ for $t \ge 0$ (see Appendix C), whence S is a symmetric operator.

Equations (40) have been derived from Eqs. (24). Now we should show that the converse also holds. For this, we shall first prove that Eqs. (40) possess unique solutions for given $g_{\pm 1}(t)$. Note that $g_{\pm 1}(t)$ contain unknown constants $a_{\pm 1}$.

It is well known that if the norm of the operator S is less than unity, Eqs. (40) have a unique solution which can be expressed in Neumann series. Indeed, Appendix C shows that S is less than 1 in norm as an operator in $L^2(0, \infty)$,

$$\|S\|_{L^2(0,\infty)} < 1, \quad (43)$$

for all real λ excepting the case $\lambda^* < \lambda < 0$ with $0 < \alpha < \alpha^*$, where $\alpha^* \doteqdot 0.27$ and λ^* is a negative constant depending only on α. The values of λ^* for some α are given at the end of the Appendix C. We will not consider the exceptional case in the following: We have not been able to prove Eq. (43) for this case.

Now note that $g(t) \in L^2(0, \infty)$. Hence the successive iteration defined by

$$\psi_{\pm 1}^{(0)} = 0, \quad \psi_{\pm 1}^{(i)} = S\psi_{\mp 1}^{(i-1)} + g_{\pm 1}, \quad i = 1, 2, 3, \cdots, \quad (44)$$

converges in $L^2(0, \infty)$, and $\psi_{\pm 1}^{(i)}(t)$ provides the ith-order approximation of the solutions $\psi_{\pm 1}(t)$.

The convergence in the L^2 sense suffices for all our purposes. Thus, let us define $\Psi_{\pm 1}^{(i)}(z)$ by Eq. (35) with $\psi_{\pm 1}^{(i)}$ in place of $\psi_{\pm 1}$. Since $\psi_{\pm 1}^{(i)} \in L^2(0, \infty)$ and since if z is not on the real positive axis, $h(t)e^{-t}/(t-z) \in L^2(0, \infty)$ for t and is holomorphic for z, then $\Psi_{\pm 1}^{(i)}(z)$ exist and are holomorphic in the plane cut along the real positive axis. Moreover, $\psi_{\pm 1}^{(i)} \to \psi_{\pm 1}$ in $L^2(0, \infty)$, so that $\Psi_{\pm 1}^{(i)}(z)$ converges uniformly for z in the cut plane. Hence the limits $\Psi_{\pm 1}(z)$ is also holomorphic in the cut plane. On the other hand, it is easy to see that the function $v(t)$ defined by $v = S\psi$ with any $\psi \in L^2(0, \infty)$ is continuous and has continuous derivatives in $\delta \le t < \infty$, where δ is an arbitrary positive number, and, by Schwartz' inequality,

$$|v(t)| \le C\,\|\psi\|_{L^2(0,\infty)}, \quad \delta \le t < \infty, \quad (45)$$

where C is a constant depending only on δ, λ, and α. Similar inequalities hold also for derivatives of $v(t)$. Thus $\psi_{\pm 1}^{(i)}(t)$ is continuously differentiable and, by Eq. (45), uniformly convergent in $0 < t < \infty$. It then follows that the limits $\Psi_{\pm 1}(z)$ of $\Psi_{\pm 1}^{(i)}(z)$ possess definite values $\Psi_{\pm 1}^{\pm}(t)$ for $t > 0$ and that, in virtue of Eqs. (40), Eqs. (24) are satisfied. Consequently, we can now conclude that Eqs. (24) and (40) are equivalent.

Finally we should point out that the unknown constants $a_{\pm 1}$ which appear for $\lambda > 0$ can be determined by Eq. (25) if $\psi_{\pm 1}$ are known. In fact, substitution of Eqs. (35) into Eqs. (25) gives rise to a set of algebraic equation for $a_{\pm 1}$.

IV. APPROXIMATE SOLUTIONS OF EQ. (2)

The solution of Eq. (2) can be found if the inverse transform of $\tilde{u}_0(z)$ is known. Since $\tilde{u}_0(z)$ is now an entire function, we can choose any path parallel to the imaginary axis as the integration path in the inverse Laplace transformation. For simplicity, we take the

imaginary axis. Then, in view of Eq. (8), we get

$$\varphi(x) = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{-iR}^{iR} e^{xz} \frac{1}{\Lambda(z)} \times \left(\tilde{f}(z) - \frac{\lambda}{\Gamma(\alpha)} [e^{-z}\Psi_{+1}(-z) + e^{z}\Psi_{-1}(z)] \right) dz, \tag{46}$$

where the limit is to be understood in the L^2 sense since $\varphi \in L^2(-1, 1)$.

A more convenient form of the solution can be found by the deformation of the integration path. This is carried out in Appendix D for the case when $\tilde{f}(z)$ can be decomposed as $\tilde{f}(z) = \tilde{f}_1(z) + \tilde{f}_2(z)$ in which $\tilde{f}_1(z)$ and $\tilde{f}_2(z)$ are holomorphic with the possible exception of a simple pole at $z = 0$ and, for $z \to \infty$,

$$\tilde{f}_1(z) = O\left(\frac{e^{-\mathrm{Re}\, z}}{|z|}\right), \quad \mathrm{Re}\, z > 0,$$

$$\tilde{f}_2(z) = O\left(\frac{e^{\mathrm{Re}\, z}}{|z|}\right), \quad \mathrm{Re}\, z < 0. \tag{47}$$

Then we can have

$$\varphi(x) = \frac{1}{\pi} \int_0^\infty \frac{\sin\theta(t)}{r(t)} [e^{xt}A_1(t) + e^{-xt}A_2(t)]\, dt + (a_1 e^{z_0 t} + a_2 e^{-z_0 x})H(\lambda), \tag{48a}$$

with

$$A_1(t) = \frac{\lambda}{\Gamma(\alpha)} \frac{X(-t)}{h(-t)} \psi_{+1}(t) - \tilde{f}_1(t),$$

$$A_2(t) = \frac{\lambda}{\Gamma(\alpha)} \frac{X(-t)}{h(-t)} \psi_{-1}(t) - \tilde{f}_2(-t), \tag{48b}$$

$$a_1 = \frac{1}{2z_0 X(z_0)X(-z_0)} \left(\frac{\lambda}{\Gamma(\alpha)} e^{-z_0}\Psi_{+1}(-z_0) - \tilde{f}_1(z_0) \right),$$

$$a_2 = \frac{1}{2z_0 X(z_0)X(-z_0)} \left(\frac{\lambda}{\Gamma(\alpha)} e^{z_0}\Psi_{-1}(-z_0) - \tilde{f}_2(-z_0) \right). \tag{48c}$$

Equation (48a) is valid for $|x| < 1$.

It is clear that if $\psi_{\pm1}(t)$ are replaced by $\psi_{\pm1}^{(i)}(t)$, Eq. (47) gives the ith-order approximation of $\varphi(x)$.

In the Kirkwood–Riseman equation, $\lambda < 0$, and the case $f(x) = 1$ is of interest since $\varphi(x)$ is related to the translational diffusion coefficient D:

$$\int_{-1}^{1} \varphi(x)\, dx = c/D, \tag{49}$$

where c is a physical quantity specific to the fluid (for details, see Ref. 1). In this case we see that

$$\tilde{f}(z) = (e^z - e^{-z})/z, \tag{50}$$

whence we have

$$\tilde{f}_1(z) = \tilde{f}_2(-z) = -e^{-z}/z. \tag{51}$$

Consequently, the zeroth-order approximation ($\psi_{\pm1} = \psi_{\pm1}^{(0)} = 0$) yields

$$\int_{-1}^{1} \varphi^{(0)}(x)\, dx = \frac{4}{\pi} \int_0^\infty \frac{\sin\theta(t)}{t^2 r(t)} e^{-t} \sinh t\, dt, \tag{52}$$

while in the first-order approximation ($\psi_{\pm1} = \psi_{\pm1}^{(1)} = g_{\pm1}$),

$$\int_{-1}^{1} \varphi^{(1)}(x)\, dx = \int_{-1}^{1} \varphi^{(0)}(x)\, dx + \frac{2\lambda}{\pi\Gamma(\alpha)} \int_0^\infty \frac{\sin\theta(t)}{th(t)} \frac{X(-t)}{h(t)} [g_{+1}(t) + g_{-1}(t)]\, dt, \tag{53}$$

where $g_{\pm1}$ are given by Eq. (42) with $a_{\pm1} = 0$.

V. ASYMPTOTIC DISTRIBUTION OF EIGENVALUES

Let us consider the eigenvalue equation $\phi = \lambda K\phi$: We should put $f(x) = 0$. Since K is a positive operator, we have only to consider the case $\lambda > 0$. Moreover we can easily see that the eigenfunctions $\phi(x)$ are either even or odd functions, so that the set of integral equations (40) reduces to a single equation

$$\psi = \pm S\psi + ag, \tag{54}$$

in which

$$\psi = \psi_{+1} = \pm\psi_{-1}, \quad a = a_+ = \pm a_-, \tag{55}$$

$$g(t) = h(t)e^{-t}. \tag{56}$$

Here the signs + and − correspond to the case of even and odd eigenfunctions, respectively.

Obviously, Eq. (54) has the solution of the form

$$\psi = a\psi_0, \tag{57}$$

where ψ_0 is the solution of Eq. (54) for $a = 1$. Then Eq. (25) becomes, by use of Eqs. (35),

$$e^{2z_0} = \mp \frac{X(-z_0)}{X(z_0)} \frac{1 \pm I(z_0)}{1 \pm I(-z_0)}, \tag{58}$$

$$I(z_0) = \int_0^\infty \frac{h(t)e^{-t}}{t + z_0} \psi_0(t)\, dt. \tag{59}$$

This equation serves to determine the eigenvalues of K.

In Appendix E we have shown that

$$X(-z_0)/X(z_0) = -\exp[\tfrac{1}{2}(1 + \alpha)\pi i], \tag{60}$$

being independent of z_0. Further, note that $z_0 = vi$ is purely imaginary and that ψ_0 is real. The latter follows from the fact that S is symmetric and g is real. Hence $\overline{I_0(z_0)} = I_0(-z_0)$ holds. Therefore, Eq.

TABLE I. Numerical calculation of $\lambda_n^{(0)}$ and $\lambda_n^{(1)}$.

	$n=1$	$n=3$	$n=5$	$n=7$	$n=9$
$\alpha=0.1$					
$\lambda^{(0)}$	0.1405(+1)	0.9410(+1)	0.1660(+2)	0.2345(+2)	0.3008(+2)
$\lambda^{(1)}$	0.1016(+1)	0.9329(+1)	0.1656(+2)	0.2342(+2)	0.3006(+2)
%	27.6[a]	0.86	0.27	0.13	0.076
$\alpha=0.3$					
$\lambda^{(0)}$	0.5230	0.2101(+1)	0.3244(+1)	0.4233(+1)	0.5131(+1)
$\lambda^{(1)}$	0.4413	0.2089(+1)	0.3238(+1)	0.4229(+1)	0.5128(+1)
%	16.7	0.59	0.19	0.089	0.053
$\alpha=0.5$					
$\lambda^{(0)}$	0.3062	0.7706	0.1046(+1)	$0.1262_4(+1)$	$0.1447_0(+1)$
$\lambda^{(1)}$	0.2784	0.7678	0.1045(+1)	$0.1261_8(+1)$	$0.1446_5(+1)$
%	9.07	0.35	0.11	0.053	0.031
$\alpha=0.7$					
$\lambda^{(0)}$	0.1662	0.2802	0.3356	0.3753	0.4071
$\lambda^{(1)}$	0.1594	0.2797	0.3354	0.3752	0.4070
%	4.06	0.17	0.052	0.026	0.014
$\alpha=0.9$					
$\lambda^{(0)}$	0.5168(−1)	0.6095(−1)	0.6467(−1)	$0.6710_4(-1)$	$0.6893_4(-1)$
$\lambda^{(1)}$	0.5116(−1)	0.6092(−1)	0.6466(−1)	$0.6710_0(-1)$	$0.6893_2(-1)$
%	1.00	0.043	0.013	0.006	0.003

[a] $(\lambda^{(0)}-\lambda^{(1)})\times 100/\lambda^{(0)}$.

(58) can be rewritten as

$$\nu=\tfrac{1}{2}(n-1)\pi+\tfrac{1}{8}(1+\alpha)\pi-\chi(\nu),\qquad n=1,2,3,\cdots, \tag{61}$$

where n is odd (even) for even (odd) eigenfunctions while

$$\chi(\nu)=-\arg\,[1\pm I(z_0)]=\tan^{-1}\left[\left(\pm\frac{\nu}{\pi}\int_0^\infty\frac{h(t)e^{-t}\psi_0(t)}{t^2+\nu^2}\,dt\right)\times\left(1\pm\frac{1}{\pi}\int_0^\infty\frac{th(t)e^{-t}\psi_0(t)}{t^2+\nu^2}\,dt\right)^{-1}\right], \tag{62}$$

in which the branch of $\tan^{-1}$ is chosen such that $|\chi(\nu)|<\frac{1}{2}\pi$. Appendix F shows that for $\nu\gg 1$

$$|\chi(\nu)|\le O(\nu^{-1}),\qquad\left|\frac{d}{d\nu}\chi(\nu)\right|\le O(\nu^{-1}). \tag{63}$$

This implies, with $|\chi(\nu)|<\frac{1}{2}\pi$ taken into account, that Eq. (61) has one and only one solution ν_n in the interval $(0,\infty)$ for large n, which takes the form

$$\nu_n=\tfrac{1}{2}(n-1)\pi+\tfrac{1}{8}(1+\alpha)\pi+O(n^{-1}). \tag{64}$$

Substituting this into Eq. (23), we now obtain an asymptotic formula of eigenvalue of the operator K:

$$\lambda_n=\frac{\Gamma(\alpha)\sin[\frac{1}{2}(1-\alpha)\pi]}{\pi}\times\left(\frac{n-1}{2}\pi+\frac{1+\alpha}{8}\pi\right)^{1-\alpha}+O\left(\frac{1}{n^{1+\alpha}}\right). \tag{65}$$

Since $\chi(\nu)=0$ if $\psi_0=0$, the first term on the right-hand side of Eq. (65) gives the zeroth-order approximate eigenvalue $\lambda_n^{(0)}$. The remaining term gives the correction for $\lambda_n^{(0)}$. Numerical computation of the first-order approximate eigenvalue $\lambda_n^{(1)}$, which is given by putting $\psi_0=g$ in Eq. (62) and solving Eq. (61) numerically, shows that the correction is within 1% for $n\ge 3$ when $0.1\le\alpha\le 0.9$ (see Table I).

APPENDIX A: PROOF OF EQ. (21)

Putting $u=t^2$, we get

$$2\int_0^\infty\frac{t^\alpha}{t^2-z^2}\,dt=\int_0^\infty\frac{u^{\frac{1}{2}(\alpha-1)}}{u-z^2}\,du. \tag{A1}$$

Consider the contour integral

$$\int_C\frac{\xi^{\frac{1}{2}(\alpha-1)}}{\xi-s}\,d\xi, \tag{A2}$$

where $\xi^{\frac{1}{2}(\alpha-1)}$ is a single-valued function with $0<\arg\xi<2\pi$ in the ξ plane cut along the real positive axis, and C denotes the contour consisting of two circles $|\xi|=R$ and $|\xi|=\delta$, $R>\delta$, and of two straight lines $\delta<\xi<R$ connecting the two circles above and below the cut. Suppose that s is not on the cut. Then, for sufficiently large R and small δ, we have, by the calculus of residues,

$$\int_C\frac{\xi^{\frac{1}{2}(\alpha-1)}}{\xi-s}\,d\xi=2\pi i s^{\frac{1}{2}(\alpha-1)}. \tag{A3}$$

The integrals on the circles vanish in the limit $R \to \infty$, $\delta \to 0$, and the sum of two integrals along the cut reduces to

$$(1 - e^{\pi i(\alpha-1)}) \int_0^\infty \frac{u^{\frac{1}{2}(\alpha-1)}}{u - s}\, du, \tag{A4}$$

since $\xi^{\frac{1}{2}(\alpha-1)} = u^{\frac{1}{2}(\alpha-1)} e^{\pi i(\alpha-1)}$, $u > 0$, below the cut. Thus Eq. (21) follows with $s = z^2$ for $0 < \arg z < \pi$ and $s = z^2 e^{-2\pi i}$ for $\pi < \arg z < 2\pi$.

APPENDIX B: PROOF OF EQ. (34)

Equations (26) and (30) give

$$\Gamma_0(z) + \Gamma_0(-z) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\ln [\Lambda^+(t)/\Lambda^-(t)]}{t - z}\, dt. \tag{B1}$$

(i) $\lambda > 0$: Recall that the function $\Lambda(z)$ is holomorphic in the plane cut along the real axis and possesses two simple zeros $\pm z_0$ on the imaginary axis. Define the function

$$\Lambda_0(z) = z^2 \Lambda(z)/(z^2 - z_0^2). \tag{B2}$$

This is also holomorphic in the cut plane but has no zeros, and thus

$$\Lambda_0(z) \to 1 \tag{B3}$$

uniformly for $z \to \infty$. Since $\Lambda^+(t)/\Lambda^-(t) = \Lambda_0^+(t)/\Lambda_0^-(t)$ on the cut and since $\Lambda^\pm(t) \to 1$ for $t \to \pm\infty$, Eq. (B1) becomes

$$\frac{1}{2\pi i} \int_{\Sigma_+} \frac{\ln \Lambda_0(\xi)}{\xi - z}\, d\xi - \frac{1}{2\pi i} \int_{\Sigma_-} \frac{\ln \Lambda_0(\xi)}{\xi - z}\, d\xi, \tag{B4}$$

where $\Sigma_+(\Sigma_-)$ represents the straight line $-\infty < \xi < \infty$ above (below) the cut. In view of Eq. (B3), $\ln \Lambda_0(z) \to 0$ for $z \to \infty$, so that the path $\Sigma_+(\Sigma_-)$ can be closed by adding a large semicircle in the upper (lower) half-plane, and the integral of Eq. (B4) is equal to the residue at $\xi = z$, namely, to $\ln \Lambda_0(z)$. Hence

$$\begin{aligned} X(z)X(-z) &= -z^{-2} \exp [\Gamma_0(z) + \Gamma_0(-z)] \\ &= \Lambda(z)/(z_0^2 - z^2). \end{aligned} \tag{B5}$$

(ii) $\lambda < 0$: In this case we can take $\Lambda_0(z) = \Lambda(z)$, whence Eq. (34) is obvious.

APPENDIX C: PROOF OF EQ. (43)

The function $h(t)$ of Eq. (37) is positive and bounded for $t \geq 0$. In particular, we can have

$$0 \leq h(t) \leq 1 - \epsilon, \quad \begin{cases} \lambda > 0, & 0 < \alpha < 1, \\ \lambda < 0, & \alpha^* < \alpha < 1, \\ & \alpha^* \doteq 0.27, \end{cases} \tag{C1}$$

where ϵ is a small positive number which does not depend on the values of λ and t. The proof will be given at the end of this appendix.

Now we shall evaluate the inner product

$$\begin{aligned} &(S\varphi, \psi)_{L^2(0,\infty)} \\ &\quad = \frac{1}{\pi} \int_0^\infty \int_0^\infty \frac{h(t)h(t')e^{-(t+t')}\varphi(t)\overline{\psi(t')}}{t + t'}\, dt\, dt', \end{aligned} \tag{C2}$$

with arbitrary $\varphi, \psi \in L^2(0, \infty)$. In view of Eq. (C1),

$$|(S\varphi, \psi)| \leq \frac{(1 - \epsilon)^2}{\pi} \int_0^\infty \int_0^\infty \frac{|\varphi(t)|\, |\psi(t')|}{t + t'}\, dt\, dt'. \tag{C3}$$

Writing the integrand on the right-hand side as

$$|\varphi(t)|\, (t + t')^{-\frac{1}{2}} (t/t')^{\frac{1}{4}}\, |\psi(t')|\, (t + t')^{-\frac{1}{2}} (t'/t)^{\frac{1}{4}}, \tag{C4}$$

and applying Schwartz' inequality, we get

$$\begin{aligned} |(S\varphi, \psi)| &\leq \frac{(1 - \epsilon)^2}{\pi} \left[\int_0^\infty |\varphi(t)|^2 \int_0^\infty \frac{1}{t + t'} \left(\frac{t}{t'}\right)^{\frac{1}{2}} dt'\, dt \right]^{\frac{1}{2}} \\ &\quad \times \left[\int_0^\infty |\psi(t)|^2 \int_0^\infty \frac{1}{t + t'} \left(\frac{t}{t'}\right)^{\frac{1}{2}} dt'\, dt \right]^{\frac{1}{2}} \\ &\leq (1 - \epsilon)^2 \, \|\varphi\|_{L^2(0,\infty)} \, \|\psi\|_{L^2(0,\infty)}, \end{aligned} \tag{C5}$$

where we have used the relation

$$\int_0^\infty \frac{1}{t + t'} \left(\frac{t}{t'}\right)^{\frac{1}{2}} dt' = \int_0^\infty \frac{u^{-\frac{1}{2}}}{1 + u}\, du = 2 \int_0^\infty \frac{du}{1 + u^2} = \pi. \tag{C6}$$

Equation (C5) assures the validity of Eq. (43) within the restrictions imposed on Eq. (C1) for the ranges of λ and α.

There remains to prove Eq. (C1). First note that

$$\begin{aligned} h^2(t) &= \frac{e^{i\theta(t)} \sin \theta(t) X(-t)}{X^+(t)} \\ &= \frac{Q(-t)}{Q(t)} \sin \theta(t) \exp Z(t), \\ Z(t) &= -\frac{2t}{\pi} \int_0^\infty \frac{\theta(t')}{t'^2 - t^2}\, dt', \end{aligned} \tag{C7}$$

which follows from the definition of $X(z)$ in Eq. (29). Due to our choice of the branch for $\theta(t)$, $\sin \theta(t) \lesseqgtr 0$ for $\lambda \gtreqless 0$ while $Q(-t)/Q(t) = \mp 1$ for $\lambda \gtreqless 0$, as seen from Eq. (31). Hence

$$h^2(t) = |\sin \theta(t)| \exp Z(t), \tag{C8}$$

which implies $h(t) \geq 0$. Further, integrating by parts, we can obtain

$$Z(t) = -\frac{1}{\pi} \int_0^\infty \theta'(t') \ln \left| \frac{t + t'}{t - t'} \right| dt'. \tag{C9}$$

(i) $\lambda > 0$. Because of our choice of branch, $\theta'(t) \geq 0$ for all $t \geq 0$. Hence $Z(t) \leq 0$. $Z(t) = 0$ occurs only at

$t = 0$ and $t = \infty$, but $|\sin \theta(0)| = \sin [\frac{1}{2}(1 + \alpha)\pi] < 1$ and $\sin \theta(\infty) = 0$. Moreover, $h(t)$ is continuous. Thus Eq. (C1) holds for all values of $\lambda > 0$ and $0 < \alpha < 1$.

(ii) $\lambda < 0$. In this case $Z(t) \geq 0$ since $\theta'(t) \leq 0$. It is convenient to change the variables

$$t = |\kappa|^\beta u^\beta, \quad t' = |\kappa|^\beta u'^\beta, \quad \beta = (1 - \alpha)^{-1},$$
$$\kappa = \lambda\pi/\Gamma(\alpha) \sin [\tfrac{1}{2}(1 - \alpha)\pi] < 0, \qquad \text{(C10)}$$
$$h(t) = h_0(u), \quad Z(t) = Z_0(u).$$

Then $h_0(u)$ and $Z_0(u)$ are independent of λ (or κ), and we have

$$Z_0(u) = \frac{1}{\pi} \int_0^\infty \frac{\sin [\frac{1}{2}(1 - \alpha)\pi]}{\{u' + \cos [\frac{1}{2}(1 - \alpha)\pi]\}^2 + \{\sin [\frac{1}{2}(1 - \alpha)\pi]\}^2} \times \ln \left| \frac{u^\beta + u'^\beta}{u^\beta - u'^\beta} \right| du'$$
$$\leq \frac{\sin [\frac{1}{2}(1 - \alpha)\pi]}{\pi} \int_0^\infty \frac{1}{u^2 + 1} \ln \left| \frac{u^\beta + u'^\beta}{u^\beta + u'^\beta} \right| du'$$
$$= \frac{\sin [\frac{1}{2}(1 - \alpha)\pi]}{\pi} Y(u). \qquad \text{(C11)}$$

Divide the integral $Y(u)$ for $0 \leq u' < u$ and $u \leq u' < \infty$ and put $x = u'/u$ and $x = u/u'$, respectively. Then

$$Y(u) = \int_0^1 \left(\frac{u}{u^2 + x^2} + \frac{u}{1 + x^2u^2} \right) \ln \left| \frac{1 + x^\beta}{1 - x^\beta} \right| dx. \qquad \text{(C12)}$$

Consider first the case $0 \leq u \leq 1$. It is readily proved that

$$\frac{u}{u^2 + x^2} \leq \frac{u}{2xu} = \frac{1}{2x}, \quad \frac{1}{1 + x^2u^2} \leq \frac{1}{1 + x^2} \leq \frac{1}{2x}, \qquad 0 \leq x \leq 1, \quad 0 \leq u \leq 1, \qquad \text{(C13)}$$

and

$$\int_0^1 \frac{1}{2x} \ln \left| \frac{1 + x^\beta}{1 - x^\beta} \right| dx = \tfrac{1}{2}(1 - \alpha) \int_0^1 \frac{1}{x} \ln \left| \frac{1 + x}{1 - x} \right| dx = \tfrac{1}{8}(1 - \alpha)\pi^2, \qquad \text{(C14)}$$

$$\int_0^1 \frac{1}{1 + x^2} \ln (1 + x^\beta)\, dx \leq \int_0^1 \frac{1}{1 + x^2} \ln (1 + x)\, dx = \tfrac{1}{8}\pi \ln 2, \qquad \text{(C15)}$$

$$\int_0^1 \frac{1}{2x} \ln \frac{1}{|1 - x^\beta|} dx = \tfrac{1}{2}(1 - \alpha) \int_0^1 \frac{1}{x} \ln \frac{1}{1 - x} dx = \tfrac{1}{12}(1 - \alpha)\pi^2, \qquad \text{(C16)}$$

whence

$$Y(u) \leq (\tfrac{1}{8} \ln 2 + \tfrac{5}{24}\pi)\pi. \qquad \text{(C17)}$$

For $1 \leq u < \infty$, we get, instead of Eq. (C13),

$$\frac{u}{u^2 + x^2} \leq \frac{1}{1 + x^2} \leq \frac{1}{2x}, \quad \frac{u}{1 + u^2x^2} \leq \frac{u}{2ux} = \frac{1}{2x}, \qquad 0 \leq x \leq 1, \quad 1 \leq u < \infty. \qquad \text{(C18)}$$

Therefore, Eq. (C17) applies also. Consequently, we can have

$$h^2(t) \leq \sin [\tfrac{1}{2}(1 - \alpha)\pi] \times \exp \{\sin [\tfrac{1}{2}(1 - \alpha)\pi][\tfrac{1}{8} \ln 2 + (1 - \alpha)\tfrac{5}{24}\pi]\}, \qquad t \geq 0, \qquad \text{(C19)}$$

for all $\lambda < 0$ and $0 < \alpha < 1$. This estimate justifies Eq. (C1) with $\alpha^* \doteqdot 0.45$. A more precise estimate should be found to improve the lower bound α^*. Instead, however, numerical computation of $h(t)$ has been made. The result is $\alpha^* \doteqdot 0.27$.

For $0 < \alpha < \alpha^*$, $h(t) \geq 1$, actually, for some values of t and, therefore, Eq. (C1) is violated. However, by examining the proof of Eq. (C6), we can easily see that Eq. (43) is valid only if

$$h(t)e^{-t} = h_0(u)e^{-|\kappa|^\beta u^\beta} < 1 \qquad \text{(C20)}$$

holds for $t \geq 0$. Since $h_0(u)$ is independent of λ and is bounded [see Eq. (C19)] and, moreover, since the set of points u such that $h(u) \geq 1$ is bounded and strictly bounded away from the point $u = 0$, which follows from the fact that $Z_0(u) = 0$ only at $u = 0$ and ∞ while $\sin \theta(0) = \sin [\frac{1}{2}(1 - \alpha)\pi] < 1$ and $h_0(u) \to 0$ as $u \to \infty$, then we see, from the second form of Eq. (C20), that the inequality holds in Eq. (C20) for sufficiently large $|\lambda|$, that is, for $\lambda < \lambda^*$ with negative constant λ^* depending on α. From the numerical computation, $-\lambda^* = 7.0$, 3.1, 1.3, and 0.3 have been found for $\alpha = 0.1$, 0.15, 0.2, and 0.25, respectively.

APPENDIX D: PROOF OF EQ. (48)

Suppose first that $\lambda < 0$. Then the integral in Eq. (46) can be written term by term as

$$\int_{-iR}^{iR} e^{xz} \frac{f(z)}{\Lambda(z)} dz - \frac{\lambda}{\Gamma(\alpha)} \int_{-iR}^{iR} e^{xz} \frac{e^{-z}\Psi_+(-z)}{\Lambda(z)} dz - \frac{\lambda}{\Gamma(\alpha)} \int_{-iR}^{iR} e^{xz} \frac{e^{z}\Psi_{-1}(z)}{\Lambda(z)} dz, \qquad \text{(D1)}$$

since each integral by itself has no singularities on the path of integration. Assume that the decomposition of $\tilde{f}(z)$ stated in the text is possible. Then the first integral of the above can be divided again into two parts containing $\tilde{f}_1(z)$ and $\tilde{f}_2(z)$, respectively. Each integrand has no singularities also in the plane cut

along the real axis. Hence Cauchy's theorem verifies that

$$\int_{i\delta}^{iR} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz = \int_{C_{R^+}} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz + \int_0^R e^{xt}\frac{\tilde{f}_1(t)}{\Lambda^+(t)}\,dt - \int_{C_{\delta^+}} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz, \tag{D2}$$

$$\int_{-iR}^{-i\delta} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz = \int_{C_{R^-}} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz - \int_0^R e^{xt}\frac{\tilde{f}_1(t)}{\Lambda^-(t)}\,dt - \int_{C_{\delta^-}} e^{xz}\frac{\tilde{f}_1(z)}{\Lambda(z)}\,dz, \tag{D3}$$

where $C_{R^+}(C_{R^-})$ and $C_{\delta^+}(C_{\delta^-})$ denote the arcs $|z| = R$ and $|z| = \delta$ with $0 < \arg z < \frac{1}{2}\pi$ $(-\frac{1}{2}\pi < \arg z < 0)$. In view of Eq. (48) and well-known Jordan's lemma, the integrals on $C_{R^\pm}$ vanish in the limit $R \to \infty$ for $x < 1$, while the integrals on $C_{\delta^\pm} \to 0$ as $\delta \to 0$ since $\tilde{f}_1(z)/\Lambda(z) = O(|z|^{-\alpha})$ near $z = 0$. Thus we see, using Eq. (26), that

$$\int_{-i\infty}^{i\infty} e^{xz}\frac{\tilde{f}_2(z)}{\Lambda(z)}\,dz = \int_0^\infty e^{xt}\left(\frac{1}{\Lambda^+(t)} - \frac{1}{\Lambda^-(t)}\right)\tilde{f}_2(t)\,dt = -2i\int_0^\infty e^{xt}\frac{\sin\theta(t)}{\gamma(t)}\tilde{f}_2(t)\,dt. \tag{D4}$$

Similarly, with the aid of integration paths lying in the left half-plane $\operatorname{Re} z < 0$, we obtain for $x > -1$

$$\int_{-i\infty}^{i\infty} e^{xz}\frac{\tilde{f}_2(z)}{\Lambda(z)}\,dz = -\int_{-\infty}^0 e^{xt}\left(\frac{1}{\Lambda^+(t)} - \frac{1}{\Lambda^-(t)}\right)\tilde{f}_2(t)\,dt = -2i\int_0^\infty e^{-xt}\frac{\sin\theta(t)}{\gamma(t)}\tilde{f}_2(-t)\,dt. \tag{D5}$$

As for the second integral of Eq. (D1), we observe that $e^{-z}\Psi_{+1}(-z)$ possesses entirely the same properties as $\tilde{f}_1(z)$ in the right half-plane $\operatorname{Re} z > 0$. Hence the deformation used for Eq. (D4) is valid. Finally, note that $e^z\Psi_{-1}(z)$ resembles $f_2(z)$ in nature, for $\operatorname{Re} z < 0$, which completes the proof of Eq. (48) for $\lambda < 0$.

In the case $\lambda > 0$, we have only to add the contributions from the simple poles at $z = \pm z_0$. The residue calculation will be greatly simplified if one makes use of Eq. (34).

APPENDIX E: PROOF OF EQ. (60)

Making the transformation $t = \nu u$, we get

$$-\frac{2z_0}{\pi}\int_0^\infty \frac{\theta(t')}{t^2 - z_0^2}\,dt = -\frac{2i}{\pi}\int_0^\infty \frac{\theta_0(u)}{u^2+1}\,du,$$

$$\theta_0(u) = \tan\frac{-\sin[\frac{1}{2}(1-\alpha)\pi]}{u^{1-\alpha} - \cos[\frac{1}{2}(1-\alpha)\pi]}. \tag{E1}$$

Divide the integral for $0 \le u < 1$ and for $1 \le u < \infty$, and change the variable u to u^{-1} in the latter integral. Then Eq. (E1) becomes

$$-\frac{2i}{\pi}\int_0^1 \frac{\theta_0(u) + \theta_0(u^{-1})}{u^2+1}\,du. \tag{E2}$$

On the other hand, it is easy to show that

$$\theta_0(u) + \theta_0(u^{-1}) = -\tfrac{1}{2}(1+\alpha)\pi. \tag{E3}$$

Thus the integral of Eq. (E1) is identical with

$$\tfrac{1}{4}(1+\alpha)\pi i, \tag{E4}$$

irrespective of the value of z_0. This together with the definition of $X(z)$ by Eq. (29) completes the proof.

APPENDIX F: PROOF OF EQ. (63)

Putting $t = \nu u$, $h(t) = h_0(u)$, and $\psi_0(t) = \varphi_0(u)$ in Eq. (59), we get

$$I(z_0) \equiv I_0(\nu) = \int_0^\infty \frac{h_0(u)e^{-\nu u}\varphi_0(u)}{u+i}\,du. \tag{F1}$$

Evidently $h_0(u)$ is independent of ν, and $\varphi_0(u)$ is a solution of the equation

$$\varphi_0 = \pm S_0\varphi_0 + g_0, \tag{F2}$$

$$g_0(u) = h_0(u)e^{-\nu u},$$

$$S_0\varphi_0 = \int_0^\infty \frac{h_0(u)h_0(u')e^{-\nu(u+u')}}{u+u'}\,\varphi_0(u')\,du'. \tag{F3}$$

By definition, $0 \le h_0(u) \le 1 - \epsilon$ follows from Appendix C so that $\|S_0\|_{L^2(0,\infty)} \le 1 - \epsilon$, $\|g_0\| \le \nu^{-\frac{1}{2}}$, and from Eq. (F2)

$$\|\varphi_0\| = \|(1 \mp S_0)^{-1}g_0\| \le (1 - \|S_0\|)^{-1}\,\|g_0\| \le (\epsilon\nu^{\frac{1}{2}})^{-1}. \tag{F4}$$

Hence the first part of Eq. (63) follows from the definition of $\chi(\nu)$ and the fact that

$$|I(z_0)| = |I_0(\nu)| \le \int_0^\infty h_0(u)e^{-\nu u}\,|\varphi_0(u)|\,du \le \|g_0\|\,\|\varphi_0\| \le (\epsilon\nu)^{-1}. \tag{F5}$$

Now differentiate Eqs. (F1) and (F2) with respect to ν:

$$I_0'(\nu) = -\int_0^\infty \frac{ug_0(u)\varphi_0(u)}{u+i}\,du + \int_0^\infty \frac{g_0(u)\varphi_0'(u)}{u+i}\,du, \tag{F6}$$

$$\varphi_0' = \pm S_0\varphi_0' + f_0,$$

$$f_0(u) = -g_0(u)\left(\pm\int_0^\infty g_0(u)\varphi_0(u)\,du + u\right). \tag{F7}$$

Clearly, $\|f_0\| \leq \|g_0\|^2 \|\varphi_0\| + \|ug_0\| = O(\nu^{-\frac{3}{2}})$, whence also $\|\varphi_0'\| = O(\nu^{-\frac{3}{2}})$ and, thereby,

$$|I'(\nu)| \leq \int_0^\infty g_0(u)\,|\varphi_0(u)|\,du + \int_0^\infty g_0(u)|\varphi_0'(u)|\,du = O(\nu^{-1}).$$

This and Eq. (F5) then assure the second estimate of Eq. (63).

[1] J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[2] G. J. Mitsis, Argonne National Laboratory, Argonne, Illinois, Report No. ANL-6787, 1963.
[3] A. G. Gibbs, J. Math. Phys. **10**, 875 (1969).
[4] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[5] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[6] S. Ukai, J. Math. Anal. Appl. **27**, 323 (1969).
[7] A. C. Zaanen, *Linear Analysis* (Noordhoff, Groningen, The Netherlands, 1953).
[8] P. L. Auer and C. S. Gardner, J. Chem. Phys. **23**, 1545 (1955).
[9] T. Carleman, Arkiv Mat. Astron. Fysik **13**, 1 (1918). (See also Ref. 7, p. 549.)

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Relaxation Spectrum of Phonons: A Solvable Model

FRANCISCO H. CLARO AND GREGORY H. WANNIER
Physics Department, University of Oregon, Eugene, Oregon

(Received 15 July 1970)

The approach towards equilibrium of a slightly disturbed phonon gas is analyzed in terms of its normal decay modes. It is shown that a simplified model permits a complete solution of the linearized phonon Boltzmann equation. The model considers 3-phonon normal collisions as the relaxing process and assumes a nondispersive isotropic elastic continuum. In analogy with the Maxwellian molecule model, the coupling term is assumed independent of the frequency. The spectrum of relaxation rates found has no discrete modes except for one trivial solution of rate zero. Above this value there is a gap and then a continuum extending to infinity.

I. INTRODUCTION

Most theories describing the relaxation towards equilibrium of a given statistical system consider the case of exponentially decaying configurations. These configurations, also called "normal decay modes," are convenient to treat because they give the otherwise integro-differential problem the form of an eigenvalue equation. It is important for this purpose to assume that the system is left alone to decay. The operator in the eigenvalue equation is a nonlinear integral operator usually called the collision operator. This difficult nonlinearity may be avoided by considering only small departures from equilibrium and keeping only linear terms. However, this simplification usually does not render the problem soluble. On the contrary, the mathematics remains quite complex and most attempts to solve such problems have been unsuccessful.

To make the eigenvalue equation explicitly soluble, some rather drastic assumptions about the collisions have been made in some cases. For example, such an approach has been taken in the theories of dilute gases and neutron transport. In the former case the molecules are assumed to repel each other as the inverse fifth power of the distance, thus making the collision rate of an individual molecule independent of its velocity.[1] Molecules obeying this power law are called Maxwellian molecules, after J. C. Maxwell who first studied them.[2] In the latter case, the interactions of a neutron with the medium were assumed, on the average, not to change its speed. For both cases the full spectrum has been given in detail and the corresponding eigenfunctions have been found.[3] Although these assumptions are unrealistic, they provide a framework for the study of the structure of solutions to the actual eigenvalue equation, and sometimes they may even be used as an approximation to a particular real physical situation.

Our purpose in this paper is to apply a similar approach to a phonon gas under somewhat restricted conditions and to solve for the spectrum and eigenfunctions of the resulting collision operator. We will see that this solution can be obtained if the 3-phonon matrix elements are assumed independent of wavenumber. Phonons obeying this law will be called "Maxwellian phonons" in what follows, owing to the similarity of this model with the Maxwellian molecule model mentioned above.

Clearly, $\|f_0\| \leq \|g_0\|^2 \|\varphi_0\| + \|ug_0\| = O(\nu^{-\frac{3}{2}})$, whence also $\|\varphi_0'\| = O(\nu^{-\frac{3}{2}})$ and, thereby,

$$|I'(\nu)| \leq \int_0^\infty g_0(u)\,|\varphi_0(u)|\,du + \int_0^\infty g_0(u)|\varphi_0'(u)|\,du = O(\nu^{-1}).$$

This and Eq. (F5) then assure the second estimate of Eq. (63).

[1] J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[2] G. J. Mitsis, Argonne National Laboratory, Argonne, Illinois, Report No. ANL-6787, 1963.
[3] A. G. Gibbs, J. Math. Phys. **10**, 875 (1969).
[4] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[5] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[6] S. Ukai, J. Math. Anal. Appl. **27**, 323 (1969).
[7] A. C. Zaanen, *Linear Analysis* (Noordhoff, Groningen, The Netherlands, 1953).
[8] P. L. Auer and C. S. Gardner, J. Chem. Phys. **23**, 1545 (1955).
[9] T. Carleman, Arkiv Mat. Astron. Fysik **13**, 1 (1918). (See also Ref. 7, p. 549.)

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Relaxation Spectrum of Phonons: A Solvable Model

FRANCISCO H. CLARO AND GREGORY H. WANNIER
Physics Department, University of Oregon, Eugene, Oregon

(Received 15 July 1970)

The approach towards equilibrium of a slightly disturbed phonon gas is analyzed in terms of its normal decay modes. It is shown that a simplified model permits a complete solution of the linearized phonon Boltzmann equation. The model considers 3-phonon normal collisions as the relaxing process and assumes a nondispersive isotropic elastic continuum. In analogy with the Maxwellian molecule model, the coupling term is assumed independent of the frequency. The spectrum of relaxation rates found has no discrete modes except for one trivial solution of rate zero. Above this value there is a gap and then a continuum extending to infinity.

I. INTRODUCTION

Most theories describing the relaxation towards equilibrium of a given statistical system consider the case of exponentially decaying configurations. These configurations, also called "normal decay modes," are convenient to treat because they give the otherwise integro-differential problem the form of an eigenvalue equation. It is important for this purpose to assume that the system is left alone to decay. The operator in the eigenvalue equation is a nonlinear integral operator usually called the collision operator. This difficult nonlinearity may be avoided by considering only small departures from equilibrium and keeping only linear terms. However, this simplification usually does not render the problem soluble. On the contrary, the mathematics remains quite complex and most attempts to solve such problems have been unsuccessful.

To make the eigenvalue equation explicitly soluble, some rather drastic assumptions about the collisions have been made in some cases. For example, such an approach has been taken in the theories of dilute gases and neutron transport. In the former case the molecules are assumed to repel each other as the inverse fifth power of the distance, thus making the collision rate of an individual molecule independent of its velocity.[1] Molecules obeying this power law are called Maxwellian molecules, after J. C. Maxwell who first studied them.[2] In the latter case, the interactions of a neutron with the medium were assumed, on the average, not to change its speed. For both cases the full spectrum has been given in detail and the corresponding eigenfunctions have been found.[3] Although these assumptions are unrealistic, they provide a framework for the study of the structure of solutions to the actual eigenvalue equation, and sometimes they may even be used as an approximation to a particular real physical situation.

Our purpose in this paper is to apply a similar approach to a phonon gas under somewhat restricted conditions and to solve for the spectrum and eigenfunctions of the resulting collision operator. We will see that this solution can be obtained if the 3-phonon matrix elements are assumed independent of wavenumber. Phonons obeying this law will be called "Maxwellian phonons" in what follows, owing to the similarity of this model with the Maxwellian molecule model mentioned above.

II. THE MAXWELLIAN PHONON MODEL

The situation we wish to describe is that of a phonon gas that is initially slightly perturbed from thermodynamic equilibrium. We then watch the system decay and analyze the change in occupation number with time as a function of wave vector.

Different processes can contribute to the decay of the system. Among them, most important are imperfection and boundary scattering, and phonon–phonon interactions. Of these effects the first two can be neglected if we assume a large and pure system at not too low temperatures. We are left with phonon–phonon interactions which appear in the formalism as terms of order higher than the second in the Taylor expansion of the potential energy of the system. Thus, they are a consequence of anharmonicity. In our treatment, we will consider only 3-phonon interactions that conserve wave vector. These interactions appear to be predominant at temperatures much lower than the Debye temperatures and at long wavelengths.[4] If $N(\mathbf{q})$ is the occupation number for wave vector $\mathbf{q}$, the total transition probability per unit time is given, with these assumptions, by[5]

$$\frac{dN(\mathbf{q})}{dt} = CN(\mathbf{q}), \tag{1}$$

where C is the nonlinear collision operator as defined by the expression

$$\begin{aligned} CN(\mathbf{q}) = \frac{V}{4\pi h}\Bigg[& \int d\mathbf{q}' N(\mathbf{q})N(\mathbf{q}')N(\mathbf{q}+\mathbf{q}') \\ & \times \left(\frac{1}{N(\mathbf{q})} + \frac{1}{N(\mathbf{q}')} + \frac{1}{N(\mathbf{q})N(\mathbf{q}')} - \frac{1}{N(\mathbf{q}+\mathbf{q}')}\right) \\ & \times |M(\mathbf{q}+\mathbf{q}';\mathbf{q},\mathbf{q}')|^2\, \delta[\epsilon(\mathbf{q}) - \epsilon(\mathbf{q}') - \epsilon(\mathbf{q}+\mathbf{q}')] \\ & + \tfrac{1}{2}\int d\mathbf{q}' N(\mathbf{q})N(\mathbf{q}')N(\mathbf{q}-\mathbf{q}') \\ & \times \left(\frac{1}{N(\mathbf{q})} - \frac{1}{N(\mathbf{q}')} - \frac{1}{N(\mathbf{q}-\mathbf{q}')} - \frac{1}{N(\mathbf{q}')N(\mathbf{q}-\mathbf{q}')}\right) \\ & \times |M(\mathbf{q};\mathbf{q}',\mathbf{q}-\mathbf{q}')|^2\, \delta[\epsilon(\mathbf{q}') + \epsilon(\mathbf{q}-\mathbf{q}') - \epsilon(\mathbf{q})]\Bigg]. \end{aligned}$$

Here, V is the volume of the system and h is Planck's constant. $M(\mathbf{q}_1;\mathbf{q}_2,\mathbf{q}_3)$ is the matrix element for the process in which a phonon of wave vector $\mathbf{q}_1$ goes into a phonon of wave vector $\mathbf{q}_2$ and a phonon of wave vector $\mathbf{q}_3$, as given by the third-order term in the expansion of the potential energy in terms of creation and annihilation operators. The Dirac δ function in the integrands express conservation of energy for the allowed transitions. Note that a time-independent solution of (1) is the Bose–Einstein distribution, as can easily be checked. Polarization has been ignored in this expression.

Further progress towards the solution of (1) requires the specification of the dispersion law $\epsilon(\mathbf{q})$. Since we are mainly interested in long wavelengths, we will assume that the medium is nondispersive, a behavior known to be followed for small wave vectors. We also assume that the medium is elastically isotropic. The conservation laws are satisfied with these two assumptions only when collisions are collinear, a condition that will make the integral in (1) one dimensional. These assumptions are quite restrictive, and we adopt them solely for the purpose of simplifying the equations.

The next step is to linearize the collision operator. This is done by introducing the perturbed Bose distribution

$$N(\mathbf{q}) = \left[[1 + \nu(\mathbf{q})e^{-\lambda t}] \exp\left(\frac{\epsilon(\mathbf{q})}{kT}\right) - 1\right]^{-1} \tag{2}$$

and, upon substitution in (1), retaining only terms of first order in the ν's. We have thus introduced normal modes, since the small perturbation of statistical equilibrium depends exponentially on the time. λ is the relaxation rate of a particular decay normal mode. Since the medium is isotropic, we expect to find no preferred direction for the collisions. This is actually the case, as it is easily shown by writing $\nu(\mathbf{q})$ as a function of the magnitude of the argument times a spherical harmonic. The angular function gets factored out after integration of the delta functions in (1), thus making the eigenvalues independent of the angular quantum numbers. We then get for the radial part $\rho(q)$

$$\begin{aligned} \lambda\rho(q) = \frac{V}{8\pi\hbar^2 c}\Bigg(& \int_0^\infty dq' \, |M(q+q';q,q')|^2 \\ & \times \frac{q'(q+q')}{q} \frac{\sinh sq}{\sinh sq' \sinh s(q+q')} \\ & \times [\rho(q) + \rho(q') - \rho(q+q')] \\ & + \tfrac{1}{2}\int_0^q dq' \, |M(q;q',q-q')|^2 \\ & \times \frac{q'(q-q')}{q} \frac{\sinh sq}{\sinh sq' \sinh s(q-q')} \\ & \times [\rho(q) - \rho(q') - \rho(q-q')]\Bigg), \end{aligned}$$

where $s = \hbar c/2kT$ and q and q' are the magnitudes of the corresponding vectors, k is Boltzmann's constant, and c the velocity of sound in the medium.

In the last expression the right-hand side can be made a single integral if the range of the variable q'

is extended to negative values. Together with this extension of the variable, we define $|M(q_1; q_2, q_3)|^2$ to be an even function of its arguments and $\rho(q)$ to be odd. This last definition is, in fact, a condition on the eigenfunctions that they vanish at the origin. It is seen from (2) that this condition avoids the appearance of a chemical potential. We also use the fact that for collinear processes M is symmetric with respect to interchange of its arguments. All this and some rearrangements give the equation the form

$$\lambda\rho(q) = \frac{V}{16\pi\hbar^2 c}\int_{-\infty}^{\infty} dq'\,|M(q, q', q-q')|^2 \times \frac{q'(q-q')}{q}\frac{\sinh sq}{\sinh sq' \sinh s(q-q')} \times [\rho(q) - \rho(q') - \rho(q-q')].$$

Because of the symmetry of the kernel, the last two terms in the square bracket are equivalent, and we can give this expression the more convenient form

$$\mu\varphi(q) = \Gamma(q)\varphi(q) - 2\int_{-\infty}^{\infty} K(q, q')\varphi(q')\,dq', \quad (3)$$

with

$$K(q, q') = s^2\,|M(q, q', q-q')|^2\frac{q-q'}{\sinh s(q-q')},$$

$$\Gamma(q) = \int_{-\infty}^{\infty} K(q, q')\frac{q'}{\sinh sq'}\frac{\sinh sq}{q}\,dq',$$

and

$$\varphi(q) = \frac{q\rho(q)}{\sinh sq}, \quad \mu = \frac{4\pi\hbar^4 c^3}{V}\frac{\lambda}{(kT)^2}.$$

This equation has been published previously by one of the authors.[6] The integral operator on the right is a nonnegative linear transformation whose eigenvalues can be chosen all real. This constitutes an H theorem for the system since it means that all modes decay. The eigenvalues thus lie on the real line between zero and infinity. The eigenfunctions are orthogonal to each other.

The detailed structure of the spectrum depends on the choice of $|M(q_1, q_2, q_3)|^2$. For long wavelengths, this quantity is proportional to its arguments.[7] This makes (3) quite hard to solve since the assumption of M being an even function of the arguments introduces absolute values in the kernel. To solve Eq. (3), we shall then assume that $|M(q_1, q_2, q_3)|^2$ has a constant average value that we shall denote by $\langle M^2\rangle$. The coupling between individual phonons is thus made independent of their energy, in analogy with Maxwellian molecules. This is not a perfect analogy since our assumption does not give a frequency-independent collision rate. Nevertheless, for the sake of giving a name to our model, phonons allowing this assumption are then called Maxwellian phonons. We shall see in the next section that their normal decay spectrum can be analyzed in full detail.

III. RELAXATION RATES AND RELAXATION MODES

An immediate solution of (3) independent of the form of the interaction is $\rho(q) = q$. The orthogonality of this solution to all others implies that the energy content of each normal mode is zero, a statement of conservation of energy for the system. This solution does not relax since it corresponds to the eigenvalue $\lambda = 0$. It merely modifies the temperature of the system, as is readily seen by substituting the eigenfunction in expression (2). In our case the eigenvalue zero is expected to be nondegenerate since, under the particular assumptions adopted, conservation of momentum coincides with conservation of energy.

For Maxwellian phonons, as defined earlier, (3) reduces to

$$\bar{\mu}\varphi(q) = \tfrac{1}{3}(\pi^2 + s^2q^2)\varphi(q) - 2s^2\int_{-\infty}^{\infty}\frac{q-q'}{\sinh s(q-q')}\varphi(q')\,dq', \quad (4)$$

where

$$\bar{\mu} = \mu/\langle M^2\rangle.$$

We have performed the integration in the multiplier $\Gamma(q)$ and transferred a constant to the left-hand side. Taking the Fourier transform of (4), we get

$$s^2\frac{d^2\psi}{dx^2} + \left(\frac{3\pi^2}{\cosh^2(\pi x/2s)} + (3\bar{\mu} - \pi^2)\right)\psi = 0, \quad (5)$$

with

$$\psi(x) = \int_{-\infty}^{\infty}\varphi(q)\sin qx\,dq.$$

This is a Schrödinger-type differential equation for a shallow symmetric potential well. Since the potential is an even function, the solutions to this equation can be classified into functions of even and odd parity. Since parity is preserved under Fourier transformation, we need only the odd solutions of (5), given by

$$\psi_\alpha(x) = \sinh\left(\frac{\pi x}{2s}\right)\cosh^{-3}\left(\frac{\pi x}{2s}\right) \times {}_2F_1\left(-1 + i\alpha, -1 - i\alpha; \tfrac{3}{2}; -\sinh^2\left(\frac{\pi x}{2s}\right)\right); \quad (6)$$

${}_2F_1$ is the hypergeometric function and

$$\alpha = \pi^{-1}(3\bar{\mu} - \pi^2)^{\frac{1}{2}}.$$

A solution that vanishes at infinity is found if the series is terminated after the first term. This corresponds to $\alpha = i$, or $\lambda = 0$. The Fourier transform of this solution is

$$\varphi_i(q) = \left(\frac{15s^5}{\pi^4}\right)^{\frac{1}{2}} \frac{q^2}{\sinh sq}, \tag{7a}$$

which is the same solution found by inspection of Eq. (3). This is the only solution belonging to the discrete spectrum. Hence, we conclude that there is no relaxing normal mode in the discrete spectrum and that the nonrelaxing solution that corresponds to energy–momentum conservation is unique.

The continuum spectrum extends from $\alpha = 0$ to infinity. In this region the solutions (6) are all real, and they oscillate for large values of $|x|$. Their Fourier transform is given by

$$\varphi_\alpha(q) = \left(\frac{2\pi}{s}(1 + 4\alpha^2)(9 + 4\alpha^2)(1 + \alpha^2)\right)^{-\frac{1}{2}} \times [R_\alpha(q) - R_\alpha(-q)], \tag{7b}$$

where

$$R_\alpha(q) = P\frac{3}{\pi^2}\frac{5s^2q^2 - \pi^2(1 + \alpha^2)}{\sinh(sq - \pi\alpha)} - \pi\alpha(11 - 4\alpha^2)\delta(sq - \pi\alpha).$$

In this expression P stands for the Cauchy principal value. The functions are normalized in the space of (5), where they form a complete set for odd square-integrable functions.[8] They are defined for all real values of q, but since they are odd, no generality is lost if the definition is restricted to the nonnegative part of the real axis.

All the functions (7b) have a pole for $q = (\pi/s)\alpha$, except for the special case of $\alpha = \frac{1}{2}$. For this singular value of q, the unbounded multiplier $\Gamma(q)$ in Eq. 4 becomes equal to the corresponding eigenvalue $\bar{\mu}$.

IV. CONCLUDING REMARKS

(7a) and (7b) may appear hard to work with. However, in the Fourier transform space of (5) the mathematics is quite simple. This should allow us to expand a small departure from equilibrium that vanishes at the origin, but is otherwise arbitrary, in terms of the complete set of solutions of (5). The exponential time factor can then be inserted in the expansion and finally the whole expression transformed back to wavenumber space. In this way and under the restrictions imposed by the model, a time-dependent solution to (1) can be found for given initial conditions.

The way in which $\langle M^2 \rangle$ should be estimated in an application is still obscure, because the assumption of its constancy does not agree with the facts. For long wavelengths it involves an average over a product of wavenumbers that will probably have to be estimated on purely physical grounds.

Note added in the proof: The expression (6) can be given a closed form. Known identities for hypergeometric functions permit changing the first argument of ${}_2F_1$ to -3. It thereby becomes a cubic polynomial. Apart from a factor, the expression (6) then reads

$$\psi_\alpha(x) = \cos(\alpha\pi x/s)[\tfrac{5}{2}\tanh^3(\pi x/2s) - (\tfrac{3}{2} + 4\alpha^2)\tanh(\pi x/2s)] + \sin(\alpha\pi x/s)[5\alpha\tanh^2(\pi x/2s) - \tfrac{4}{3}\alpha(1 + \alpha^3)].$$

ACKNOWLEDGMENTS

The authors wish to thank Dr. D. Benin, Professor J. W. McClure, and Professor R. Barrar for very inspiring discussions.

[1] G. H. Wannier, *Statistical Physics* (Wiley, New York, 1966), p. 405; G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (American Mathematical Society, Providence, R.I., 1963), p. 84.

[2] J. C. Maxwell, *Collected Papers* (Cambridge U.P., New York, 1890), Vol. II, p. 41.

[3] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967), Chap. 4. See also Ref. 1, Sec. 18-5.

[4] I. Pomeranchuk, J. Phys. (USSR) **4**, 259 (1941); **6**, 237 (1942); Phys. Rev. **60**, 820 (1941).

[5] See, for example, R. E. Peierls, *Quantum Theory of Solids* (Clarendon, Oxford, 1955), p. 48.

[6] G. H. Wannier, Bull. Am. Phys. Soc. **14**, 303 (1969).

[7] C. Herring, Phys. Rev. **95**, 954 (1954).

[8] E. C. Titchmarsh, *Eigenfunction Expansions* (Clarendon, Oxford, 1946), p. 121.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Dimensionality of Space-Time

K. H. MARIWALLA
Department of Physics, Arya-Mehr University of Technology, Tehran, Iran

(Received 26 May 1970)

It is shown that, for a local Poincaré covariant quantum field theory in which the usual relation between the *TCP* symmetry and spin-statistic holds, the dimensionality of space-time should be even. Further plausibility arguments are presented in favor of the observed dimensionality, on basis of the following factors: (1) simplicity, (2) relativistic local quantum field theory, (3) general relativity, (4) the nature of some observed interactions, (5) classical mechanical force concept.

INTRODUCTION

Recently Rosen[1] has attempted to show that if one requires the existence of a *TCP* operation for an n-dimensional Dirac equation, then n can only be even. The basic point of his argument is the following. Since the Dirac equation corresponds to a double-valued representation of a (pseudo-)orthogonal group, one must use elements of a Clifford algebra. The irreducible set of Clifford matrices of dimension 2^m have a maximum subset of $2m + 1$ anticommuting matrices. These could be used for either O_{2m} or O_{2m+1}. Now the *TCP* operation changes the sign of all the $2m$ or $2m + 1$ coordinates in the respective cases; hence, in order for the *TCP* symmetry to exist for a Dirac type equation, one must have one matrix which anticommutes in the respective cases with $2m$ or $2m + 1$ anticommuting Clifford matrices. This is clearly impossible for O_{2m+1}. One could, of course, argue that, if one uses higher-dimensional Clifford matrices, one would have large number of anticommuting matrices and hence *TCP* operation is possible even for odd dimensions. However, this argument is fallacious, as the set of such Clifford matrices is reducible. One can in this case by a unitary transformation reduce the higher-dimensional equation into the direct sum of the lowest-dimensional equation with itself. To avoid any confusion on this point, we note that in the general proof of the spin-statistics theorem[2] one requires that the fields belong to irreducible, finite-dimensional real representations of the covering group of $L_+^\uparrow$ (the fields themselves do not have to be real). These are given by $D^{(l,k)} \oplus D^{(k,l)}$ and $D^{(k,k)}$. Considered over the field of complex numbers, the first of these are not irreducible.[3] In particular, the four-component Dirac equation is irreducible only as a real representation. The requirement of reality of the representations is connected with the precise form of the assumption of locality (commutation or anticommutation for spacelike separations).

One can further clarify this point by considering the proof of the *TCP* theorem in the abstract field theory approach, where it becomes trivially clear that the dimensionality should be even. In this approach the important point is that the operation *TCP* is continuously connected to the identity in $L_+(4, c)$, thereby connecting the otherwise disjoint real groups $L_+^\uparrow(4, R)$ and $L_+^\downarrow(4, R)$. For this connectivity, it is essential that the operation *TCP* should have "determinant" $+1$, which is not the case for odd dimensions. The group $L_+(n, c)$ is indeed connected for both even and odd n. To see the essential difference between the even and odd dimensions, let us decompose $L(n, c)$ into its constituents for the two cases. For even dimensions,

$$L(n, c) = L_+(n, c) + PL_+(n, c),$$
$$L_+(n, R) = L_+^\uparrow(n, R) + PTL_+^\uparrow(n, R),$$

where P and T are the diagonal matrices with components

$$P = (-1, -1, \cdots, -1; +1),$$
$$T = (1, 1, \cdots, 1; -1).$$

For odd dimensions ($n = 2m + 1$),

$$L(n, c) = L_+(n, c) + \tau L_+(n, c),$$
$$L_+(n, R) = L_+^\uparrow(n, R) + \theta L_+^\uparrow(n, R).$$

θ and T are diagonal matrices with components

$$\theta = (1, 1, \cdots, 1, -1; -1),$$
$$\tau = (1, 1, \cdots, 1; -1).$$

The operation corresponding to parity consists in changing the sign of the $2m$ spatial coordinates. This can be accomplished continuously within $L_+^\uparrow(n, R)$ by performing rotations of magnitude π in m orthogonal, spatial planes. Parity is therefore not a discrete operation. The operation of changing the sign of all the coordinates is thus equivalent to a "parity rotation" followed by τ. The determinant of such a transformation is clearly -1, and it is clear[4] that this operation is not contained in $L_+(n, c)$. In fact, one

can go further to give plausibility arguments in favor of the dimensionality of space–time as four.

Suppose we admit that the dimensionality is even, $N = 2n$; then for the "spinor group" (covering group) we need elements of the Clifford algebra of dimension 2^N. Since N is even, one can represent the Clifford algebra of linear transformations on a vector space of dimension $d = 2^n$. The group of automorphisms of this space is $SL(\frac{1}{2}d; \mathbf{C})$; only for $N = 4$ does this exactly coincide with the N-dimensional Lorentz group. In any other case, there result considerable algebraic complications.

Since *TCP* is a discrete operation, there is another way of looking at the problem, by studying the embedding of the discrete symmetry *TCP* in some continuous group $G(TCP)$ and the connection of the latter with the Lorentz group. This problem we have studied elsewhere for the 4-dimensional Lorentz group[5]: We have shown there that if D is a discrete symmetry operator, $\{C^1, C_2, \cdots, C_N\}$ a "complete commuting set" of operators, and if M ($< N$) of these change sign under D, then D is continuously connected to the identity in the group $G(D, 4) = SU_2 \otimes SU_2 \otimes \cdots \otimes SU_2$ (M times). Actually, what was shown is that for such a D there are $2M$ possible "generators" Ω such that $D = \exp(i\pi\Omega)$ for any one of these Ω's; and the $2M$ Ω's together with the M commuting C's which change sign under D constitute a Lie algebra isomorphic to the Lie algebra of $SU_2 \otimes SU_2 \otimes \cdots \otimes SU_2$ (M times). We explicitly demonstrated this for the operators of parity, time reversal, charge conjugation, and *TCP*. In particular for the *TCP* operation in four dimensions, we find that only two of the complete commuting set of operators change sign so that *TCP* is continuously connected with respect to the group $G(TCP, 4) \sim O_4 \sim SU_2 \otimes SU_2$. Now if we take the complex Lorentz group, its maximal compact subgroup $M(4, c)$ is just O_4; so, in the 4-dimensional complex Lorentz group, *TCP* is continuously connected to the identity in its maximal compact subgroup, i.e., $M(4, c) \sim G(TCP, 4)$.

Now suppose that space–time is a $2n$-dimensional Minkowski space. Then, in the connected component of the complex Lorentz group $L_+(2n, c)$, the two disjoint subgroups $L_+^{\uparrow}(2n, R)$ and $L_+^{\downarrow}(2n, R)$ are connected continuously via the *TCP* operation. Since O_{2n}, $n \neq 2$, is not semisimple, from what we have said above about the connectivity properties of a discrete symmetry operation, it follows that the real group $G(TCP; 2n)$, in which *TCP* is connected continuously to the identity, is only a subgroup of the maximal compact subgroup $M(2n, \mathbf{c})$ of $L_+(2n, c)$. These two subgroups $G(TCP{:}2n)$ and $M(2n, \mathbf{c})$ coincide only if $n = 2$, i.e., the dimensionality of the space–time is four.

This may also be seen by the following construction. Let the space be of $2n$ dimensions, isotropic and homogeneous. Then the rank of the symmetry group is n. The sum of the squares of translation generators is a Casimir invariant. One can find $n - 1$ generators which can be simultaneously diagonalized with it and which together with a "charge operator" form a set of n-commuting operators which change sign under *TCP*; this essentially defines the *TCP* operation. It then follows from the connectivity property of a discrete symmetry operation that *TCP* is continuously connected to the identity with respect to the group (or one locally isomorphic to it) $G(TCP; 2n) \sim O_3 \otimes O_3 \otimes \cdots \otimes O_3$, n-fold, which has $3n$ generators. On the other hand, the maximal compact subgroup $M(2n, \mathbf{c})$ of $L_+(2n, c)$ has $n(2n - 1)$ generators. Therefore, in order for *TCP* to be continuously connected to the identity in $L_+(2n, c)$, we must have $3n \leq 2n^2 - n$, or $2n \leq n^2$. Equality obtains only for $n = 2$, i.e., the dimensionality of space–time is four. Thus from the point of view of economy of dimensions together with the *TCP*, the dimensionality of space–time should be just four.[4]

In the remaining part of this paper we shall summarize other plausibility arguments in support of the dimensionality of space–time as four.

1. STRENGTH OF THE FIELD EQUATIONS

Given a set of fields, and the field equations (as partial differential equations) they satisfy, the fields are determined to an extent, but (in general) not completely. Thus, there remain certain free data. To characterize the free data, Einstein[6] introduced the concept of the "coefficient of freedom" as follows. If we expand the fields in a Taylor series in the neighborhood of a point P, then the totality of its coefficients describe the field functions completely. Let N_n denote the number of nth-order coefficients, and let M_n be the number of conditions that the field equations or other constraints impose on the nth-order coefficients. The number of nth-order coefficients remaining free is

$$z = N_n - M_n = N_n(1 - M_m/N_n).$$

Let us expand:

$$1 - M_n/N_n = 1 + z_1/n + z_2/n^2 + \cdots.$$

Then z_1 is defined as the "coefficient of freedom." Einstein remarked on the fact that, for empty space, Maxwell equations and Einstein equations of general relativity z_1 have the same value 12 for the space–time

dimensionality of four, whereas the Klein–Gordon zero-mass scalar field leads to a different value. Penny[7] has shown that, of the two zero-mass neutrino equations of Dirac and Weyl, only the latter gives $z_1 = 12$ for the space–time dimensionality of four. Hence, if we make a virtue out of this to demand that in the absence of any other criteria the coefficient of freedom should have the same numerical value for various field equations describing physical phenomena, we are led to conclude that the space–time dimensionality is four.

2. VARIATIONAL PRINCIPLE AND UNIQUENESS OF THE EINSTEIN EQUATIONS

Suppose we assume that the gravitational phenomena are to be described in terms of the metrical properties of space–time. It is then *natural* to assume that the trajectory of a test particle is given by the geodesic equation; the Christoffel symbols therefore play the role of the field components of intensity. Under coordinate transformations these transform linearly but not homogeneously. As a result, by a mere coordinate transformation they can be made to vanish at a point. It is thus desirable to have field equations that are tensor equations. Since the Christoffel symbols are of the first order in $g_{\mu\nu}$ (the metric tensor), the Lagrangian should at most be of the second order in $g_{\mu\nu}$:

$$L = L(g_{\mu\nu}, g_{\mu\nu,\sigma}, g_{\mu\nu,\alpha\beta}).$$

The resulting Euler–Lagrange equations will be of the fourth order in general.[8] But for the lower dimensionalities the following holds[8]:

(a) $n = 4$; there is one third-order Euler–Lagrange equation;

(b) $n = 2, 3$; there are no third-order Euler–Lagrange equations;

(c) $n = 2, 3, 4$; the only second-order Euler–Lagrange equation are the Einstein equations.

Thus if we demand that the field equations be of second order, then $n \leq 4$. If $n \leq 3$, then the connection between curvature and gravitation envisaged in Einstein's theory breaks down. This may be seen as follows. For $n = 2$, the Riemann curvature tensor has just one independent component, and the space is both conformally and projectively flat.[9] If we consider $n = 3$, the Riemann curvature tensor has six independent components. In both cases, the contracted curvature tensor has the same number of components as the Riemann tensor. Hence the vanishing of the contracted curvature tensor is sufficient to insure that the space is flat. In case of the Einstein field equations, outside a material body we get

$$\text{contracted Riemann tensor} = 0.$$

For $n \leq 3$, this implies, from what we have said above, that the space (outside the material point) is flat and there is no gravitational field outside the material point. If we therefore accept Einstein's viewpoint on gravitation (connection between curvature and gravitation), the dimensionality of space–time should be at least four. We note parenthetically that if one considers the inverse-square law of force for electrostatic and gravito-static interactions as something basic, then the dimensionality of space has to be three (and of space–time four) if the field equations for the potentials are of the second order. For instance, if one considers the Schwarzschild type metric in a $(n+1)$-dimensional space–time ($\alpha_N = \sin\theta_{N-1}$)

$$dS^2 = f\,dt^2 - dr^2/f - r^2[d\theta_0^2 + \alpha_1^2\,d\theta_1^2 + (\alpha_1\alpha_2)^2 \times d\theta_2^2 + \cdots + (\alpha_1\cdots\alpha_{n-2})^2\,d\theta_{n-2}^2]$$

together with the Einstein field equations, one finds that

$$f - 1 \sim r^{-(n-2)}.$$

It is well known that attractive potentials $\phi(r) \sim r^{-m}$, for $m \geq 2$, do not give rise to well-behaved bound states. Such potentials are to be rejected if one requires that there exist bound states of matter.

3. EQUIVALENCE OF A GEODESIC IN FORCE-FREE CURVED SPACE AND FORCE EQUATION IN FLAT SPACE–TIME

There is an interesting connection between the concept of force and the concept of time. If one considers a homogeneous, isotropic space, then the sum of the squares of the translation generators is an invariant. In the presence of external "forces" this is no longer the case. For certain types of forces, however, the sum of the squares of the translation generators plus a certain "potential" function describes the development of the system in "time." The "potential" function then describes the "force" completely. An alternate way of looking at it is as follows.

Consider the motion of a "test particle" in a three-dimensional affinely connected space; one can represent it by the equation of a geodesic:

$$\frac{d^2X^i}{dS^2} + \Gamma^i_{jk}\frac{dX^j}{dS}\frac{dX^k}{dS} = 0. \tag{1}$$

If one further assumes the validity of the "Pythagorean theorem" so that dS is determined by

$$dS^2 = g_{ij}\,dX^i\,dX^j, \tag{2}$$

then Γ^i_{jk} are completely determined in terms of the g_{ij}. On the other hand, in mechanics one is used to considering the equations of motion in a $(3+1)$-dimensional space-time together with the concept of force:

$$\frac{d^2X^i}{dt^2} + \tilde{\Gamma}^i_{jk}\frac{dX^j}{dt}\frac{dX^k}{dt} = -\tilde{g}^{ij}\frac{\partial\phi(X_1, X_2, X_3)}{\partial X^j}, \quad (3)$$

where $\tilde{g}^{ij}$ is the metric tensor of a flat 3-dimensional space and t is the time parameter. More generally, one can consider the equations of motion in the form

$$\frac{d^2X^i}{dt^2} + \tilde{\Gamma}^i_{jk}\frac{dX^j}{dt}\frac{dX^k}{dt} = -\frac{\partial T^{ij}}{\partial X^j}, \quad (4)$$

where T^{ij} is a symmetric tensor. It is amusing to note that the system (3) can always be replaced by the system (1) and (2) by choosing for example[10]

$$\Gamma^i_{jk} = \tilde{\Gamma}^i_{jk} + \frac{\partial\phi}{\partial X^l}\tilde{g}^{li}\frac{\partial t}{\partial X^j}\frac{\partial t}{\partial X^k}, \quad (5)$$

irrespective of what field equations ϕ may satisfy. On the other hand, (4) cannot in general be replaced by the system (1) and (2) unless the dimensionality of space is three or less. The reason is that for three dimensions T^{ij} has six independent components which may be determined in terms of the six components of the Riemann tensor. This may be considered as an additional argument in favor of the dimensionality of space-time as four: viz., an arbitrary geodesic in a curved space assigned with a positive-definite metric can be looked upon as the motion of a particle under "force" in flat space-time and vice versa, provided that the dimensionality of space is three.

SUMMARY

It is shown that, for the validity of *TCP* theorem as usually understood, the dimensionality of space-time should be even. The following plausibility arguments are presented in favor of the dimensionality as four.

(1) The smallest subgroup of $L_+(2n, c)$ in which *TCP* is continuously connected to the identity is a subgroup of O_{2n}, the maximal compact subgroup of $L_+(2n, c)$ provided $2n \geq 4$. Hence, the requirement of economy of dimensions together with *TCP* lead to the space-time dimensionality as four.

(2) If one requires that the strength of the field equations describing natural phenomena (of zero mass fields) should be the same, then one finds that for Maxwell (electromagnetic), Einstein (gravitational), and Weyl (neutrino) equations this is so only if the dimensionality is four.

(3) If the connection between curvature and gravitation envisaged by Einstein holds and the field equations are of the second order and derivable from a variational principle, then the dimensionality is four.

(4) If the existence of bound states is an essential requirement and the field equations are of the second order, then the dimensionality of space-time is necessarily four.

(5) If the trajectory of a test particle in space is considered as a geodesic in curved space, assigned with a positive definite metric, then this description is equivalent to a particle under "force" (derivable from a symmetric stress tensor) in flat space-time provided the dimensionality of space is three.

Note: Since this paper was completed, there has appeared a paper by I. M. Freeman [Am. J. Phys. **37**, 1222 (1969)] on the dimensionality of space. This paper is an adaptation of an earlier work of W. Büchel [Physik. Bl. **19**, 547 (1963)], and gives reference to an important paper by F. R. Tangherlini [Nuovo Cimento **27**, 636 (1963)], which gives several reasons for the dimensionality of space as three.

ACKNOWLEDGMENTS

Most of the material reported above was presented at a Seminar at MATSCIENCE, Madras. I am indebted to the members of the Physics group at MATSCIENCE, in particular, Professor A. Ramakrishnan for comments. Thanks are due to Professor Riahi for a critical reading of the manuscript.

[1] S. P. Rosen, J. Math. Phys. **9**, 1593 (1968).

[2] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965), in particular, Chap. V.3; for the decomposition of complex Lorentz group in n-dimensions, see p. 20 and Appendix I.

[3] They are, however, irreducible with respect to $L_+^\uparrow$ extended by a discrete symmetry such as parity or charge conjugation.

[4] It should be emphasized that in the entire discussion above we have assumed that the space-time is Minkowskian, i.e., the signature is $\pm(n-2)$.

[5] K. H. Mariwalla, J. Math. Phys. **7**, 114 (1966). The results are based on a local quantum field theory, but are believed to be generally true.

[6] A. Einstein, *The Meaning of Relativity* (Methuen, London, 1956), 6th ed.

[7] R. Penny, J. Math. Phys. **6**, 1607 (1965).

[8] D. Lovelock, Arch. Ratl. Mech. Anal. **33**, 54 (1969).

[9] L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J.,1949).

[10] A. Trautmann, in *Lectures on General Relativity*, S. Deser and K. W. Ford, Eds. (Prentice-Hall, Englewood Cliffs, N.J., 1964), Vol. I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Stochastic Processes and Their Representations in Hilbert Space*

F. G. HALL AND R. E. COLLINS
Department of Physics, University of Houston, Houston, Texas 77004

(Received 8 April 1970)

Beginning with an intuitive consideration of sequences of measurements, we define a time-ordered event space representing the collection of all imaginable outcomes for measurement sequences. We then postulate the *generalized distributive relation* on the event space and examine the class of measurements for which this relation can be experimentally validated. The generalized distributive relation is shown to lead to a σ-additive conditional probability on the event space and to a *predictive* and *retrodictive* formalism for stochastic processes. We then show that this formalism has a predictive and a retrodictive representation in a separable Hilbert space $\mathcal{H}$, which has no counterpart in unitary quantum dynamics.

INTRODUCTION

A recent series of papers[1-6] has developed the idea that much of the formal mathematical structure of physical theory can be deduced directly from the statistical nature of experimental data. The present paper presents that portion of these studies which bears directly on the evolution of irreversible physical processes.

We begin the study of the evolution of a system by insisting that if we are to say we have observed the dynamic behavior of the system, then we must monitor the system by a sequence of time-documented measurements $\{M_0 \to M_1 \to M_2 \to \cdots \to M_L\}$.

With each of the measurements in the sequence, we associate in our mind a collection of *possible outcomes*, the collection being determined, of course, by the properties of the measuring apparatus. We may also associate a collection of possible outcomes with the entire measurement sequence.

We assume that all experimental data is statistical in nature, i.e., each outcome in the collection of possible outcomes is a random event. This assumption leads us to consider probability theory as a mathematical model for the kinematics of a system.

Since our imagination, at least for physical-experimental situations, seems to be conditioned by conventional logic, we will assume that a σ-algebra describes the collection of imaginable outcomes (event space) of a measurement and that the frequency of outcomes can be described by a σ-additive measure of unit norm whose domain is the σ-algebra.

This approach does not differ from conventional approaches except, as we will show, in the definition of the σ-algebra of possible outcomes for measurement sequence and the conditional probability defined on this σ-algebra.

We will show that an equivalence relation must be defined on the σ-algebra for the measurement sequences in order to obtain the predictive and retrodictive random walk formulation for stochastic processes. This equivalence relation, the *generalized distributive relation*, is empirical in nature and is not deducible from the logical structure of the mathematics describing the measurement sequence.

We will then show that the predictive and retrodictive random walk formulations for the dynamics of a physical system have representations in a separable Hilbert space $\mathcal{H}$, which differ considerably from the conventional quantum representation. It appears that the dynamical laws of conventional quantum theory are not the most general representation of the random walk formulation in $\mathcal{H}$.

THE MEASUREMENT

For the sake of clarity and brevity in the following discussions, we will begin by defining the measurement process.

We assume that an experimental situation may be completely described by a countable, functionally independent set of real-valued functions $(f_1, f_2, f_3, \cdots)$, which may be arbitrarily partitioned into two functionally independent sets; one set, a K-tuple $(f'_1, f'_2, \cdots, f'_K)$ describing the results of K simultaneous measurements, and one set $(f''_1, f''_2, f''_3, \cdots)$ describing the environment conditioning the measurement. (This simply states that we must be satisfied to determine a finite number of system properties.)

We suppose that a measurement is always limited to some finite resolution, and thus each of $f'_1, f'_2, \cdots, f'_K$ has a countable range $R_{f_1}, R_{f_2}, \cdots, R_{f_K}$, respectively. Since each of $f'_1, f'_2, \cdots, f'_K$ has a countable range, there exists a *countable collection*

$$\{(p_1, p_2, \cdots, p_K)\}_{p_1 \in R_{f_1}, p_2 \in R_{f_2}, \cdots, p_K \in R_{f_K}}$$

of K-tuples of real numbers (denoted $\{\hat{p}_k\}_{k=1,2,3,\ldots}$) which contains all possible K-tuples of real numbers in the range of $(f'_1, f'_2, \cdots, f'_K)$.

Such assumptions lead us to make the following definitions:

A *measurement* of a system is an operation performed on a system which assigns a *configuration* $\hat{p}_k \in \{\hat{p}_k\}_{k=1,2,3,\ldots}$ to the system.

The *spectrum* of a measurement is the collection of all possible configurations $\{\hat{p}_k\}_{k=1,2,3,\ldots}$. For example, if we are interested in the pressure and volume of a system, then a configuration assigned to the system is a 2-tuple of real numbers (P_i, V_i) in the range of the functions P and V, respectively.

We may now define the *event space* as the collection of all imaginable outcomes for a measurement. Let C_i denote the spectrum of a measurement process M_i. The event space $\{E_i(C_i)\}$ is the σ-algebra[7] of subsets of C_i. The motivations for such a choice for the event space are discussed in several texts[8,9]; arguments against such a choice have been discussed by Jauch.[10] We will assume the σ-algebra to be a valid representation since as we will see there seem to be many physical situations for which the σ-algebra is appropriate and yields results not obtainable by conventional quantum theory.

Here we will refer to the members of $\{E_i(C_i)\}$ as *events* and define the probability for an event as a σ-additive measure P of unit norm on $\{E_i(C_i)\}$. Such a function has the following properties:

(i) If $E \in \{E_i(C_i)\}$, $0 \leq P(E) \leq 1$;

(ii) $P(\phi) = 0$; ϕ is the null event corresponding to the empty set in $\{E_i(C_i)\}$;

(iii) $P(C_i) = 1$; $C_i = \bigcup_{k_i=1}^{\infty} (\hat{p}_{k_i})$ (set union is interpreted as logical or);

(iv) if $\{E_j\}_{j=1,2,3,\ldots}$ is a disjoint sequence of sets in $\{E_i(C_i)\}$, then

$$P\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} P(E_j).$$

There is a much wider agreement on the properties of P than the event space because of obvious physical interpretations. Axioms (i) and (ii) follow from the operational definition of probability. Axiom (iii) simply states that some value in the spectrum must be obtained as a result of M_i, and axiom (iv) is the mathematical statement of the familiar mutually exclusive rule in probability theory.

With this brief introduction we may now consider sequences of measurement operations.

SEQUENCES OF MEASUREMENTS

We wish now to consider the time-documented sequence of measurements $\{M_0 \rightarrow M_1 \rightarrow M_2 \rightarrow \cdots \rightarrow M_L\}$. By time documented we mean M_i occurs at t_i and, in case $i < j$, then $t_i < t_j$. Since each M_i has an associated event space $\{E_i(C_i)\}$, the collection of all imaginable outcomes for the ordered sequence $\{M_i\}_{i=1,L}$ is a physically meaningful notion; thus we proceed to define the event space $\{E(C)\}$ for $\{M_i\}_{i=1,L}$. Let C denote the Cartesian product space for the sequence of σ-algebras $\{\{E_0(C_0)\} \rightarrow \{E_1(C_1)\} \rightarrow \{E_2(C_2)\} \rightarrow \cdots \rightarrow \{E_L(C_L)\}\}$, i.e.,

$$C = \{E_0(C_0)\} \otimes \{E_1(C_1)\} \otimes \cdots \otimes \{E_L(C_L)\}. \quad (1)$$

$\{E(C)\}$ *is the event space for* $\{M_i\}_{i=1,L}$ *means* that E is an event in $\{E(C)\}$ only in case E is a subset of C.

That $\{E(C)\}$ contains the imaginable paths of outcomes for the measurement sequence can be seen from recognizing that $\{E(C)\}$ contains the collection $\{S_n\}$ of *simple paths* $\{(\hat{p}_{k_0} \rightarrow \hat{p}_{k_1} \rightarrow \cdots \rightarrow \hat{p}_{k_L})\}$ (which are read as "$\hat{p}_{k_0}$ occurred, then $\hat{p}_{k_1}$ occurred, then, $\cdots$, then $\hat{p}_{k_L}$ occurred"), the *compound paths* such as

$$\left\{\left(\bigcup_{k_0=1}^{l_0} \hat{p}_{k_0} \rightarrow \bigcup_{k_1=1}^{l_1} \hat{p}_{k_1} \rightarrow \cdots \rightarrow \bigcup_{k_L=1}^{l_L} \hat{p}_{k_L}\right)\right\},$$

and the unions and intersections of the compound paths, for example,

$$\left(\hat{p}_{k_0} \rightarrow \bigcup_{k_1=1}^{l_1} \hat{p}_{k_1} \rightarrow \cdots \rightarrow \hat{p}_{k_L}\right) \cup \left(\hat{p}_{k_0'} \rightarrow \hat{p}_{k_1'} \rightarrow \cdots \rightarrow \bigcup_{k_L=1}^{l_L} \hat{p}_{k_L}\right).$$

Notice that, in contrast to the usual route in probability theory,[8] we have not defined $\{E(C)\}$ to be the Cartesian product space of the σ-algebras $\{E_0(C_0)\} \otimes \{E_1(C_1)\} \otimes \cdots \otimes \{E_L(C_L)\}$. Such a choice is not the most general one since it requires that set operations in $\{E(C)\}$ be defined in terms of set operations in $\{E_i(C_i)\}$. For our definition of $\{E(C)\}$ we see that C does not form a σ-algebra since it contains no unions of members of C. However, by choosing an equivalence relation between members of C and the compliment of C in $\{E(C)\}$, one can "induce" a σ-algebra on C. As we shall see in the next section, such a choice is empirical and seems necessary in order to produce the stochastic process.

PROBABILITY ON $\{E(C)\}$

We now turn our attention to probability functions on $\{E(C)\}$ and, in particular, conditional probabilities. We will assume in the following discussions that the environment for the sequence $\{M_i\}_{i=1,L}$ is fixed and described by $\hat{q}$. We will tacitly require that all probability functions on $\{E(C)\}$ be conditioned by $\hat{q}$.

The unit norm condition for P on $\{E_i(C_i)\}$ is given by

$$P(C_i) = 1, \quad (2)$$

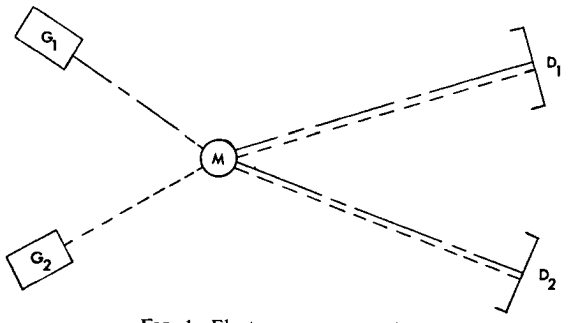


FIG. 1. Electron gun apparatus.

which was interpreted as the probability for some event to occur during M_i. In view of this, it would seem reasonable that, for the sequence of measurements,

$$P(C_0 \to C_1 \to \cdots \to C_L) = 1 \tag{3}$$

and, for the simple paths $\{S_n\}_{n=1,2,\cdots}$ in $\{E(C)\}$,

$$P\left(\bigcup_{n=1}^{\infty} S_n\right) = 1, \tag{4}$$

which is interpreted as *some simple path* must occur. In order for (3) and (4) to be true, we must postulate the following relation:

If

$$E_0 \in \{E_0(C_0)\}, \quad E_1 \in \{E_1(C_1)\}, \cdots,$$

both

$$E_i' \,\&\, E_i'' \in \{E_i(C_i)\}, \cdots, E_L \in \{E_L(C_L)\},$$

then

$$\begin{aligned}(E_0 \to E_1 \to \cdots \to E_i' \cup E_i'' \to \cdots \to E_L) &= (E_0 \to E_1 \to \cdots \to E_i' \to \cdots \to E_L) \\ &\cup (E_0 \to E_1 \to \cdots \to E_i'' \to \cdots \to E_L).\end{aligned} \tag{5}$$

We will also require the class of measurements that we are investigating to obey

$$\begin{aligned}P((\hat{p}_{k_0} \to \hat{p}_{k_1} \to \cdots \to \hat{p}_{k_i} \to \cdots \to \hat{p}_{k_L}) &\cap (\hat{p}_{k_0} \to \hat{p}_{k_1} \to \cdots \to \hat{p}_{k_i'} \to \cdots \to \hat{p}_{k_L})) \\ &= \begin{cases} 0, & k_i' \neq k_i \\ P(\hat{p}_{k_0} \to \hat{p}_{k_1} \to \cdots \to \hat{p}_{k_i} \to \cdots \to \hat{p}_{k_L}), & k_i' = k_i, \end{cases}\end{aligned} \tag{6}$$

which simply states that only one configuration may be obtained as the result of a measurement. Statements (4), (5), and (6) must be *a posteriori* in nature, not derivable from any *a priori* consideration. To clarify this point, consider the following measurement situation.

The schematic in Fig. 1 describes two electron guns G_1 and G_2 firing at a fixed target M. These electrons are scattered from M and detected at D_1 or D_2. The entire apparatus is placed in a cloud chamber so that the track of each electron can be monitored if desired. Such a device will serve to examine the generalized σ-algebra $\{E(C)\}$ and Eqs. (5) and (6).

Let M_0 denote detection of the firing of the guns, M_1 denote detection of scattering from the target, and M_2 denote detection at D_1 or D_2. We may now build $\{E(C)\}$ for the sequence $\{M_0 \to M_1 \to M_2\}$. The σ-algebras $\{E_0(C_0)\}$, $\{E_1(C_1)\}$, and $\{E_2(C_2)\}$ are given by

$$\begin{aligned}\{E_0(C_0)\} &= \{(G_1), (G_2), (G_1 \cup G_2), (G_1 \cap G_2), \varnothing\}, \\ \{E_1(C_1)\} &= \{M, \varnothing\}, \\ \{E_2(C_2)\} &= \{(D_1), (D_2), (D_1 \cup D_2), (D_1 \cap D_2), \varnothing\}.\end{aligned} \tag{7}$$

C as defined earlier is given by the Cartesian product space $\{E_0(C_0)\} \otimes \{E_1(C_1)\} \otimes \{E_2(C_2)\}$, and $\{E(C)\}$, the event space for $\{M_0 \to M_1 \to M_2\}$, is the σ-algebra of subsets of C.

If we form $\{E(C)\}$ by the prescription given above, we see that $\{E(C)\}$ contains events such as $(G_1 \to M \to D_1)$, $(G_1 \to M \to D_2)$, $(G_2 \to M \to D_1)$, and $(G_2 \to M \to D_2)$, the union of these $[(G_1 \to M \to D_1) \cup (G_1 \to M \to D_2) \cup (G_2 \to M \to D_1) \cup (G_2 \to M \to D_2)]$, and $(G_1 \cup G_2 \to M \to D_1 \cup D_2)$. It is quite natural to interpret each of the events in the collection $\{(G_i \to M \to D_j)\}$ as the event for a certain *simple path* to be observed in the cloud chamber. The union of these simple paths would, of course, be interpreted as the event for one or another of the *simple paths* to occur. However, the event $(G_1 \cup G_2 \to M \to D_1 \cup D_2)$ would appear to have no simple interpretation as an event independent of the events for simple paths. [The event $(G_1 \cup G_2 \to M \to D_1 \cup D_2)$ seems a likely candidate for a "superposition" event defined by Jauch[10] if the σ-algebraic structure of $\{E(C)\}$ is modified. This investigation will constitute another paper.]

We do see, however, that Eqs. (3) and (4) can be satisfied for $\{E(C)\}$ only in case Eq. (5) is valid on $\{E(C)\}$. Equation (5) defines the event $(G_1 \cup G_2 \to M \to D_1 \cup D_2)$ in terms of the simple paths in $\{E(C)\}$, i.e., by Eq. (5)

$$\begin{aligned}(G_1 \cup G_2 \to M \to D_1 \cup D_2) &= (G_1 \to M \to D_1 \cup D_2) \cup (G_2 \to M \to D_1 \cup D_2) \\ &= (G_1 \to M \to D_1) \cup (G_1 \to M \to D_2) \\ &\quad \cup (G_2 \to M \to D_1 \cup D_2) \\ &= (G_1 \to M \to D_1) \cup (G_1 \to M \to D_2) \\ &\quad \cup (G_2 \to M \to D_1) \cup (G_2 \to M \to D_2),\end{aligned} \tag{8}$$

and therefore the requirement for $\{E(C)\}$ that $P(C)$ is unity is consistent with Eqs. (3) and (4).

We will call Eq. (5) the *generalized distributive relation* of the set operation $\bullet$ with respect to the ordering operation $\to$. We see that this relation is

a posteriori in nature, i.e., it is *not* required by the structure of $\{E(C)\}$. Only when we require Eq. (3) or Eq. (4) to be valid must we require the generalized distributive relation. The validity of Eq. (4) can be tested only if each of the simple paths are observable; thus the generalized distributive relation is ultimately *a posteriori* in nature.

It should be evident that Eq. (5) "induces" a σ-algebra on C and thus reduces $\{E(C)\}$ to the conventional σ-algebra of simple paths. [If Eqs. (3) and (4) are to be consistent with the requirement $P(C) = 1$, then the generalized distributive relation must be valid for both union and intersection with respect to ordering.] We will see, however, that the generalized notation obtained from generalizing $\{E(C)\}$ leads to some new notions in stochastic processes.

Let us return to the experiment of Fig. 1, assuming that the generalized distributive relation is valid for this experiment. We see, in general, that the probability for $G_1 \cap G_2$ and $D_1 \cap D_2$ is nonzero. However, if we suppose that G_1 and G_2 never fire simultaneously and that D_1 and D_2 never detect simultaneously, then Eq. (6) is satisfied; thus we see that Eq. (6) is a requirement motivated by *a posteriori* knowledge.

From Eqs. (6) and (8) and the additive property of P, we see that

$$\begin{aligned} &P(G_1 \cup G_2 \to M \to D_1 \cup D_2) \\ &\quad = P(G_1 \to M \to D_1) + P(G_1 \to M \to D_2) \\ &\qquad + P(G_2 \to M \to D_1) + P(G_2 \to M \to D_2), \end{aligned} \tag{9}$$

from which we conclude that

$$\begin{aligned} &P(G_1 \cup G_2 \to M \to D_j) \\ &\quad = P(G_1 \to M \to D_j) + P(G_2 \to M \to D_j); \end{aligned} \tag{10}$$

thus we are provided with the definition

$$P(D_j) \triangleq P(C_0 \to C_1 \to D_j) = \sum_{i=1,2} P(G_i \to C_1 \to D_j) \tag{11}$$

for the *unconditional probability* to detect a particle at D_j. This definition may be generalized to an L-term measurement sequence, i.e., for the L-term measurement sequence $\{M_0 \to M_1 \to M_2 \to \cdots \to M_L\}$, the *unconditional probability* for a result $\hat{p}_{k_i}$ during M_i, $0 \leq i \leq L$, is given by

$$\begin{aligned} P(\hat{p}_{k_i}) &\triangleq P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_L) \\ &= P\Bigg(\bigcup_{k_0=1}^{\infty} \hat{p}_{k_0} \to \bigcup_{k_1=1}^{\infty} \hat{p}_{k_1} \to \cdots \to \hat{p}_{k_i} \\ &\qquad \to \bigcup_{k_{i+1}=1}^{\infty} \hat{p}_{k_{i+1}} \to \cdots \to \bigcup_{k_L=1}^{\infty} \hat{p}_{k_L}\Bigg). \end{aligned} \tag{12}$$

Thus, using the generalized distributive relation, the disjointness of the simple paths in $\{E(C)\}$, and the σ-additivity of P, we see that Eq. (12) may be written in the more familiar form

$$\begin{aligned} P(\hat{p}_{k_i}) = &\sum_{k_0=1}^{\infty} \sum_{k_1=1}^{\infty} \cdots \sum_{k_{i-1}=1}^{\infty} \sum_{k_{i+1}=1}^{\infty} \cdots \\ &\sum_{k_L=1}^{\infty} P(\hat{p}_{k_0} \to \hat{p}_{k_1} \to \cdots \to \hat{p}_{k_i} \to \cdots \to \hat{p}_{k_L}), \end{aligned} \tag{13}$$

that is, the unconditional probability for $\hat{p}_{k_i}$ is the sum of the probabilities of all simple paths containing $\hat{p}_{k_i}$.

With a suitable definition of *conditional probability*, Eq. (12) provides the general mathematical structure for a stochastic process. *Conditional probability* on $\{E(C)\}$ may be defined by analogy with the traditional definition. Conventionally, the probability for "E_i^l is observed if E_i^k is observed" is given by

$$P_C(E_i^l \mid E_i^k) \triangleq P(E_i^l \cap E_i^k)/P(E_i^k). \tag{14}$$

For the conventional event space, such a definition suffers from causal ambiguities; however, for the time-ordered event space such ambiguities disappear.

In addition to the simultaneous events of Eq. (14), we wish to consider the conditional probability for the time-separated events E_i^l and E_j^k, $i \neq j$. By analogy with the conventional definition (14), we define

$$\begin{aligned} P(E_j^l \mid E_i^k) &\triangleq P(C_0 \to C_1 \to \cdots \to E_j^l \to C_{j+1} \to \cdots \to C_L \mid C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_L) \\ &= \frac{P((C_0 \to C_1 \to \cdots \to E_j^l \to C_{j+1} \to \cdots \to C_L) \cap (C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_L))}{P(C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_L)}; \end{aligned} \tag{15}$$

we see that this is well defined, independent of the magnitude of i with respect to j. Let us examine this definition for the case where $i < j$ and the case where $i = j$.

When $i < j$, Eq. (15) becomes

$$\begin{aligned} P(E_j^l \mid E_i^k) = \frac{\begin{aligned} &P((C_0 \to C_1 \to \cdots \to C_i \to \cdots \to E_j^l \to C_{j+1} \to \cdots \to C_L) \\ &\quad \cap (C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_j \to \cdots \to C_L)) \end{aligned}}{P(C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_j \to \cdots \to C_L)}; \end{aligned} \tag{16}$$

thus $P(E_j^l \mid E_i^k)$; $i < j$ has the obvious interpretation "the conditional probability for the event E_j^l to occur at time t_j if E_i^k is known to have occurred at an earlier time t_i."

Now $P(E_i^k \mid E_j^l)$ is also well defined by Eq. (15). Let us examine the nature of this conditional probability. Equation (15) yields

$$P(E_i^k \mid E_j^l) = \frac{P((C_0 \to C_1 \to \cdots \to E_i^k \to C_{i+1} \to \cdots \to C_j \to \cdots \to C_L) \cap (C_0 \to C_1 \to \cdots \to E_j^l \to C_{j+1} \to \cdots \to C_L))}{P(C_0 \to C_1 \to \cdots \to C_i \to \cdots \to E_j^l \to C_{j+1} \to \cdots \to C_L)}, \tag{17}$$

which, in view of the nature of the sequenced event space, can only be interpreted as "the conditional probability for the event E_i^k to occur at t_i if E_j^l is known to have occurred at a later time t_j."

In case $i = j$, we see from Eqs. (16) and (17) that our definition of Eq. (15) is the analog of the conventional definition given by Eq. (14).

It is our claim, and we discuss this more fully in the sections to follow, that the sequenced formalism clearly distinguishes and defines both "types" of conditional probabilities as given in Eqs. (16) and (17). We will demonstrate that the conditional probability of Eq. (17) can be the "inverse" or "time-reversed" form of the conditional probability of Eq. (16) only in case the system follows a deterministic path through the measurement sequence. We also will see that $P(E_i^k \mid E_j^l)$, $i < j$, is definable only because of the *a posteriori* nature of the data from a measurement sequence.

We will postpone this discussion until we have more fully developed the stochastic equations describing the measurement sequence.

THE RANDOM WALK

Now that we have developed the definitions for conditional probability and unconditional probability, we are able to consider the measurement sequence as a generalized random walk problem. We will, in this section, develop the random walk equation which determines the probability for the statement, "the simple event $\hat{p}_{k_j}$ is the outcome of M_i, regardless of the outcomes of the rest of the measurements in the sequence," in terms of the conditional probabilities of $\hat{p}_{k_i}$ with respect to the outcomes of other measurements in the sequence.

We accomplish this by beginning with the definition in Eq. (12) of the unconditional probability. From this we may write

$$P(\hat{p}_{k_j}) = \sum_{k_i} P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L), \quad j > i. \tag{18}$$

Since the conditional probability is defined for each member of $\{E(C)\}$, we may write, from Eq. (15),

$$P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) = \frac{P((C_0 \to C_1 \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L) \cap (C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_L))}{P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_L)}. \tag{19}$$

Using the generalized distributive relation, we may reduce the numerator of Eq. (19) so that Eq. (19) becomes

$$P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) = \frac{P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L)}{P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_L)}. \tag{20}$$

Since the numerator of Eq. (20) is exactly the term inside the sum of Eq. (18), we may employ Eq. (20) to write Eq. (18) as

$$P(\hat{p}_{k_j}) = \sum_{k_i} P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_L). \tag{21}$$

Before we "expose" this as the random walk equation, let us consider the unconditional probability for $\hat{p}_{k_i}$. From Eq. (12), we may write

$$P(\hat{p}_{k_i}) = \sum_{k_j} P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L), \quad j > i, \tag{22}$$

and, as we saw in the development of Eq. (20), we may write from Eq. (15)

$$P(\hat{p}_{k_i} \mid \hat{p}_{k_j}) = \frac{P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L)}{P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L)}, \quad j > i, \tag{23}$$

which allows us to write Eq. (22) as

$$P(\hat{p}_{k_i}) = \sum_{k_j} P(\hat{p}_{k_i} \mid \hat{p}_{k_j}) P(C_0 \to C_1 \to \cdots \to \hat{p}_{k_j} \to C_{i+1} \to \cdots \to C_L), \quad j > i. \tag{24}$$

In the simplified notation provided by the definition of unconditional probability, Eq. (21) may be written as

$$P(\hat{p}_{k_j}) = \sum_{k_i} P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) P(\hat{p}_{k_i}), \quad j > i, \tag{25}$$

and Eq. (24) may be written

$$P(\hat{p}_{k_i}) = \sum_{k_j} P(\hat{p}_{k_i} \mid \hat{p}_{k_j}) P(\hat{p}_{k_j}), \quad j > i, \tag{26}$$

which we will name the *predictive random walk equation* and the *retrodictive random walk equation*, respectively. This is an obvious choice of terminology since Eq. (25) calculates probability distributions for events occurring at t_j in terms of the probability distributions for events occurring at an *earlier* time t_i and since Eq. (26) calculates probability distributions for events occurring at t_i in terms of the probability distributions for events occurring at a *later* time t_j.

We may go a step further in adapting our notation to the standard notation by defining the *predictive transition probability*

$$T_{k_j k_i} \triangleq P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) = P(\hat{p}_{k_j} \mid C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to C_j \to \cdots \to C_L) \tag{27}$$

and the *retrodictive transition probability*

$$T'_{k_i k_j} \triangleq P(\hat{p}_{k_i} \mid \hat{p}_{k_j}) = P(\hat{p}_{k_i} \mid C_0 \to C_1 \to \cdots \to C_i \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L), \tag{28}$$

so that the predictive random walk equation becomes

$$P(\hat{p}_{k_j}) = \sum_{k_i} T_{k_j k_i} P(\hat{p}_{k_i}) \tag{29}$$

and the retrodictive random walk equation becomes

$$P(\hat{p}_{k_i}) = \sum_{k_j} T'_{k_i k_j} P(\hat{p}_{k_j}). \tag{30}$$

We see from the preceding analysis that Eq. (29) is a generalized form of the conventional Markoff random walk equation. It is generalized in the sense that $T_{k_j k_i}$ is not Markoffian.

We also see that Eq. (30) is not at all conventional since it implies that if we know the probability set $\{P(\hat{p}_{k_j})\}$ at t_j and the set of retrodictive transition probabilities $\{T'_{k_j k_i}\}$, then we may calculate the probability set $\{P(\hat{p}_{k_i})\}$ even when $t_i < t_j$. Such a result is completely consistent with the *a posteriori* nature of data. We will discuss this property of data in the conclusion section of this paper.

PROPERTIES OF THE STOCHASTIC PROCESS

In this section we will examine the temporal behavior of the stochastic process in terms of prediction and retrodiction. This examination will clarify the relationship between the predictive dynamics and the retrodictive dynamics and will provide a foundation for our examination of the $\mathfrak{IC}$ representation of stochastic processes.

Each measurement pair M_i, M_j, $i < j$, in the measurement sequence $\{M_0 \to M_1 \to \cdots \to M_L\}$ defines a collection of predictive transition probabilities $\{T_{k_j k_i}\}$, a collection of retrodictive transition probabilities $\{T'_{k_i k_j}\}$, and a collection of *simultaneous conditional probabilities* $\{T_{k_i k_{i'}}\}$.

$\mathsf{T}(j, i)$ is the *predictive transition matrix* for the measurement pair M_i, M_j, $i < j$, means that $\mathsf{T}(j, i)$ is a matrix such that $T_{k_j k_i}$ is the k_jth-row and the k_ith-column element of $\mathsf{T}(j, i)$.

$\mathsf{T}'(i, j)$ is the *retrodictive transition matrix* for the measurement pair M_i, M_j, $i < j$, means that $\mathsf{T}'(i, j)$ is a matrix such that $T'_{k_i k_j}$ is the k_ith-row and the k_jth-column element of $\mathsf{T}'(i, j)$.

$\mathsf{T}(i, i)$ is the *simultaneous conditional probability matrix* for the measurement M_i means that $\mathsf{T}(i, i)$ is a matrix such that $T_{k_i k_{i'}}$ is the k_ith-row and the k'_ith-column element of $\mathsf{T}(i, i)$.

We see then that an L-term measurement sequence defines $\frac{1}{2}L(L+1)$ measurement pairs M_i, M_j, $i < j$, and thus defines $\frac{1}{2}L(L+1)$ retrodictive transition matrices, $\frac{1}{2}L(L+1)$ predictive transition matrices, and L simultaneous conditional probability matrices.

Let $\{\mathsf{T}(j, i)\}$ denote the collection of predictive transition matrices, $\{\mathsf{T}'(i, j)\}$ denote the collection of retrodictive transition matrices, and $\{\mathsf{T}(i, i)\}$ denote the collection of simultaneous conditional probability matrices. Let $\{\mathfrak{C}(j, i)\}$ denote the collection of members of $\{\mathsf{T}(j, i)\}$, $\{\mathsf{T}'(i, j)\}$, and $\{\mathsf{T}(i, i)\}$.

We will now investigate the conditions, if any, for the collections $\{\mathsf{T}(j,i)\}$, $\{\mathsf{T}'(i,j)\}$, and $\{\mathfrak{G}(j,i)\}$ to form either groups or semigroups with respect to matrix multiplication.

First, we note that Eq. (6) requires that the collection $\{\mathsf{T}(i,i)\}$ be the collection of unit matrices $\{\mathsf{I}_i\}$. In general, each member of $\{\mathsf{I}_i\}$ is of a different dimension, depending on the spectrum of M_i. In this investigation, we will assume that each spectrum is countably infinite, and thus each member of $\{\mathsf{I}_i\}$ will be of the same dimension.

It is not difficult to see that matrix multiplication between certain members of $\{\mathsf{T}(j,i)\}$ produces a transition matrix in $\{\mathsf{T}(j,i)\}$. To show this, we simply use Eq. (29) to write the following equation set:

$$\begin{aligned}
&P(\hat{p}_{k_1}) = \sum_{k_0} T_{k_1k_0}P(\hat{p}_{k_0}),\\
&P(\hat{p}_{k_2}) = \sum_{k_1} T_{k_2k_1}P(\hat{p}_{k_1}) = \sum_{k_0} T_{k_2k_0}P(\hat{p}_{k_0})\\
&\quad\cdot\\
&\quad\cdot\\
&\quad\cdot\\
&P(\hat{p}_{k_L}) = \sum_{k_{L-1}} T_{k_Lk_{L-1}}P(\hat{p}_{k_{L-1}}) = \sum_{k_{L-2}} T_{k_Lk_{L-2}}P(\hat{p}_{k_{L-2}})\\
&\qquad = \cdots = \sum_{k_1} T_{k_Lk_1}P(\hat{p}_{k_1}) = \sum_{k_0} T_{k_Lk_0}P(\hat{p}_{k_0}).
\end{aligned} \tag{31}$$

Substituting the first equation of the set (31) into the second equation in the set, we obtain

$$P(\hat{p}_{k_2}) = \sum_{k_0} P(\hat{p}_{k_0}) \sum_{k_1} T_{k_2k_1}T_{k_1k_0} = \sum_{k_0} P(\hat{p}_{k_0})T_{k_2k_0}, \tag{32}$$

which implies by comparison the Chapman–Kolmogorov relation[11]

$$T_{k_2k_0} = \sum_{k_1} T_{k_2k_1}T_{k_1k_0}. \tag{33}$$

This procedure may be repeated for the entire set (31) to obtain

$$T_{k_Lk_0} = \sum_{k_{L-1}} \sum_{k_{L-2}} \cdots \sum_{k_1} T_{k_Lk_{L-1}}T_{k_{L-1}k_{L-2}} \cdots T_{k_1k_0}. \tag{34}$$

Since $T_{k_Lk_0}$ is the k_Lth row and k_0th column of $\mathsf{T}(L,0)$, we see that Eq. (34) provides a *multiplication theorem* for transition matrices,

$$\mathsf{T}(L,0) = \mathsf{T}(L,L-1)\mathsf{T}(L-1,L-2)\cdots\mathsf{T}(1,0). \tag{35}$$

From the retrodictive equation (30), we may write an equation set similar to the equation set (31) and derive the multiplication theorem for the retrodictive transition matrices

$$\mathsf{T}'(0,L) = \mathsf{T}'(0,1)\mathsf{T}'(1,2)\cdots\mathsf{T}'(L-1,L). \tag{36}$$

In addition, Eqs. (29) and (30) can be combined for various integers i and j so that multiplication is defined between members of $\{\mathsf{T}(j,i)\}$ and $\{\mathsf{T}'(i,j)\}$. For example, consider the integers q, s, and t such that $0 \le q < s < t \le L$. Equations (29) and (30) then define the products

$$\begin{aligned}
\mathsf{T}'(q,s)\mathsf{T}'(s,t) &= \mathsf{T}'(q,t),\\
\mathsf{T}'(q,t)\mathsf{T}(t,s) &= \mathsf{T}'(q,s),\\
\mathsf{T}(s,q)\mathsf{T}'(q,t) &= \mathsf{T}'(s,t),\\
\mathsf{T}'(s,t)\mathsf{T}(t,q) &= \mathsf{T}(s,q),\\
\mathsf{T}(t,q)\mathsf{T}'(q,s) &= \mathsf{T}(t,s).
\end{aligned} \tag{37}$$

However, we also obtain from this process

$$\begin{aligned}
P(\hat{p}_{k_s}) &= \sum_{k_{s'}} P(\hat{p}_{k_{s'}}) \sum_{k_{t'}} T_{k_sk_{t'}}T_{k_{t'}k_{s'}}\\
&= \sum_{k_{s'}} M_{k_sk_{s'}}P(\hat{p}_{k_{s'}}),\\
P(\hat{p}_{k_t}) &= \sum_{k_{t'}} P(\hat{p}_{k_{t'}}) \sum_{k_{s'}} T_{k_tk_{s'}}T'_{k_{s'}k_{t'}}\\
&= \sum_{k_{t'}} M_{k_tk_{t'}}P(\hat{p}_{k_{t'}}).
\end{aligned} \tag{38}$$

Equations (38) define the matrix products

$$\begin{aligned}
\mathsf{M}(s,s) &= \mathsf{T}'(s,t)\mathsf{T}(t,s),\\
\mathsf{M}(t,t) &= \mathsf{T}(t,s)\mathsf{T}'(s,t).
\end{aligned} \tag{39}$$

The immediate inclination is to identify the collection $\{\mathsf{M}(i,i)\}_{i=0,L}$ as the collection $\{\mathsf{T}(i,i)\}$ of simultaneous conditional probability matrices. However, such an identification would require that

$$\begin{aligned}
\mathsf{M}(s,s) &= \mathsf{I}_s = \mathsf{T}'(s,t)\mathsf{T}(t,s),\\
\mathsf{M}(t,t) &= \mathsf{I}_t = \mathsf{T}(t,s)\mathsf{T}'(s,t),
\end{aligned} \tag{40}$$

and, if the dimension of I_s is the dimension of I_t, then Eqs. (40) imply that

$$\mathsf{T}'(s,t) = [\mathsf{T}(t,s)]^{-1}. \tag{41}$$

Wu[12] has shown, however, that since each member of $\mathsf{T}(t,s)$ is positive, then its inverse transition matrix $[\mathsf{T}(t,s)]^{-1}$ must have at least one *negative* member, unless, of course, $\mathsf{T}(t,s)$ has only one nonzero member. Since $\mathsf{T}'(s,t)$ is itself a transition matrix, Eq. (41) and thus Eqs. (40) can be satisfied only in case $\mathsf{T}(t,s)$ has only one nonzero member. [In this case $\mathsf{T}(t,s)$ would describe a *deterministic process*.] Thus we see that, in general, $\mathsf{M}(s,s)$ cannot be identified as the matrix $\mathsf{T}(s,s)$ of simultaneous conditional probabilities.

With multiplication defined in $\{\mathsf{T}(j,i)\}$ and $\{\mathsf{T}'(i,j)\}$, we may proceed to examine these collections as groups or semigroups.

Since $\{\mathsf{T}(j,i)\}$ can form a group only in case each member $\mathsf{T}(r,q) \in \{\mathsf{T}(j,i)\}$ has an inverse $[\mathsf{T}(r,q)]^{-1} \in \{\mathsf{T}(j,i)\}$, we see from the preceding arguments that neither $\{\mathsf{T}(j,i)\}$ nor $\{\mathsf{T}'(i,j)\}$ can form a group.

We also see that the collection $\{\mathcal{C}(j,i)\}$ cannot form a semigroup since the product M given by Eqs. (40) is not a member of $\{\mathcal{C}(j,i)\}$ unless, for each positive integer i such that $i \leq L$,

$$\mathsf{M}(i,i) = \mathsf{T}(i,i) = \mathsf{I}, \tag{42}$$

which, as we argued, is possible only for a deterministic system.

Let us now examine the conditions for $\{\mathsf{T}(j,i)\}$ and $\{\mathsf{T}'(i,j)\}$ to form semigroups. Suppose $\{\mathsf{T}(j,i)\}$ forms a semigroup. In this case, closed associative multiplication must be defined between each pair in $\{\mathsf{T}(j,i)\}$. We see from Eq. (35) that left multiplication of $\mathsf{T}(r,s)$ by $\mathsf{T}(q,r)$ yields $\mathsf{T}(q,s)$; thus the product $\mathsf{T}(q,r)\mathsf{T}(r,s)$ is a member of $\{\mathsf{T}(j,i)\}$ and the multiplication is closed. Since this multiplication is matrix multiplication, it is also associative.

We see, however, that multiplication of the two matrices $\mathsf{T}(p,q)\mathsf{T}(r,s)$ produces a transition matrix in $\{\mathsf{T}(j,i)\}$ only in case $q = r$ or $p = s$. This fact motivates us to define the following notion: Two transition matrices $\mathsf{T}(p,q)$ and $\mathsf{T}(r,s)$ are *adjacent* means either $p = s$ or $q = r$. It is clear, then, that if each pair of matrices in $\{\mathsf{T}(j,i)\}$ can be made adjacent, then $\{\mathsf{T}(j,i)\}$ will form a semigroup.

If each member of $\{\mathsf{T}(j,i)\}$ has the property that

$$\mathsf{T}(l,k) = \mathsf{T}(x,y) \quad \text{in case} \quad |l-k| = |x-y|, \tag{43}$$

then any two matrices $\mathsf{T}(p,q) \in \{\mathsf{T}(j,i)\}$ and $\mathsf{T}(r,s) \in \{\mathsf{T}(j,i)\}$ can be made adjacent simply by relabeling $\mathsf{T}(r,s)$ as $\mathsf{T}(q,s')$, where $\mathsf{T}(q,s') \in \{\mathsf{T}(j,i)\}$ and $|q-s'| = |r-s|$ so that

$$\mathsf{T}(p,q)\mathsf{T}(r,s) = \mathsf{T}(p,q)\mathsf{T}(q,s') = \mathsf{T}(p,s'). \tag{44}$$

Thus the collection $\{\mathsf{T}(j,i)\}$ can form a semigroup in case the matrices in the collection are all conformable and Eq. (43) is satisfied for each matrix in the collection. The same argument applies for the collection $\{\mathsf{T}'(i,j)\}$. If, in addition, we include the collection $\{\mathsf{T}(i,i)\}$ in $\{\mathsf{T}(j,i)\}$, we see that $\{\mathsf{T}(j,i)\}$ can form a *monoid* semigroup. The same argument applies for $\{\mathsf{T}'(i,j)\}$.

We see then that the predictive collection $\{\mathsf{T}(j,i)\}$ and the retrodictive collection $\{\mathsf{T}'(i,j)\}$ can each form a group only in case each member in $\{\mathsf{T}(j,i)\}$ and each member in $\{\mathsf{T}'(i,j)\}$ describes a deterministic system. However, each of $\{\mathsf{T}(j,i)\}$ and $\{\mathsf{T}'(i,j)\}$ can form a semigroup in case each member of $\{\mathsf{T}(j,i)\}$ and each member of $\{\mathsf{T}'(i,j)\}$ satisfies Eq. (43). Physically, Eq. (43) restricts the transition probabilities to be a function only of the number of measurements between M_i and M_j; this requires that each $\mathsf{T}(j,i)$ be a function only of the relative time difference between M_i and M_j. Thus Eq. (43) is analogous to the *quantum* requirement that $\mathsf{U}(t_2,t_1)$ be a function only of $|t_2 - t_1|$ if U is to be a member of the unitary group.

We also demonstrated that a retrodictive transition matrix is not the inverse of the corresponding *predictive* transition matrix. However, the equations resulting from the sequenced event space clearly define and distinguish between *retrodiction and prediction* and show that one may always predict or retrodict the stochastic process.

PROBABILITY FUNCTIONS IN l^2

In this section we will demonstrate that probabilities for simple paths in $\{E(C)\}$ may be represented as products of complex functions in l^2, the space of square summable sequences. From the isomorphism of l^2 to a separable Hilbert space $\mathcal{H}$, we deduce the existence of a continuous linear operator in $\mathcal{H}$ which corresponds to the transition probability of Eq. (27). Hilbert space representations for probabilities of simple paths in $\{E(C)\}$ are shown to be possible because of the positive-definite, unit norm and σ-additive properties of P.

Since $P(\hat{p}_{k_j})$ is positive definite, there exists a complex function α_{k_j} such that for each $\hat{p}_{k_j}$

$$P(\hat{p}_{k_j}) = \alpha_{k_j}^* \alpha_{k_j} \tag{45}$$

and the phase of α_{k_j} is *arbitrary*.

Using the unit norm property and the generalized distributive relation, we see that

$$\sum_{k_j} P(\hat{p}_{k_j}) = 1 = \sum_{k_j=1}^{\infty} \alpha_{k_j}^* \alpha_{k_j}. \tag{46}$$

Thus the sequence $\{\alpha_{k_j}\}_{k_j=1,2,\ldots}$ is square summable and is a member of l^2. If we now consider the vector $|\alpha(j)\rangle$ defined by

$$|\alpha(j)\rangle = \sum_{k_j=1}^{\infty} C_{k_j} |k_j\rangle, \tag{47}$$

where $\{|k_j\rangle\}_{k_j=1,2,\ldots}$ is an orthonormal basis for a separable Hilbert space $\mathcal{H}$, then $|\alpha(j)\rangle \in \mathcal{H}$ only in case $\{C_{k_j}\}$ is a square-summable sequence.[13] Thus, if we define C_{k_j} as

$$C_{k_j} \triangleq \alpha_{k_j} (\langle \alpha(j) \mid \alpha(j) \rangle)^{\frac{1}{2}} \tag{48}$$

we see that $\{C_{k_j}\}$ is square summable; therefore, $|\alpha(j)\rangle$ defined by

$$|\alpha(j)\rangle = (\langle \alpha(j) \mid \alpha(j) \rangle)^{\frac{1}{2}} \sum_{k_j} \alpha_{k_j} |k_j\rangle \tag{49}$$

is a member of $\mathcal{H}$. Thus we see that for each square-summable sequence $\{\alpha_{k_j}\}$ there exists a vector $|\alpha(j)\rangle \in \mathcal{H}$ such that each member of $\{\alpha_{k_j}\}$ has a representation in $\mathcal{H}$ given by

$$\alpha_{k_j} = \langle k_j \mid \alpha(j)\rangle/(\langle\alpha(j) \mid \alpha(j)\rangle)^{\frac{1}{2}}. \tag{50}$$

Thus we have established an $\mathcal{H}$ representation for each member in the collection $\{\alpha_{k_j}\}$ and therefore for $\{P(\hat{p}_{k_j})\}$.

Now let us examine the transition probability $T_{k_jk_i}$. Since $T_{k_jk_i}$ is positive, there exists a complex function for each k_j and k_i such that

$$T_{k_jk_i} = K^*_{k_jk_i}K_{k_jk_i}, \tag{51}$$

and, since $\{T_{k_jk_i}\}$ is singly stochastic, the sequence $\{K_{k_jk_i}\}_{k_j=1,2,\cdots}$ is square summable for each k_i. Therefore, there exists a countable orthonormal basis $\{|k_j\rangle\}$ and a member $|Q_{k_i}\rangle \in \mathcal{H}$ such that for each k_i

$$K_{k_jk_i} = \langle k_j \mid Q_{k_i}\rangle/(\langle Q_{k_i} \mid Q_{k_i}\rangle)^{\frac{1}{2}}. \tag{52}$$

We see from (51) and (52) that, for a given basis $\{|k_j\rangle\}$, each member of the countable collection $\{|Q_{k_i}\rangle\}$ is determined only to within a phase.

$K_{k_jk_i}$ may be written in a different form since we may associate with the collection $\{|Q_{k_i}\rangle\}$ an orthonormal basis $\{|k_i\rangle\}$ in $\mathcal{H}$ by an operator $\mathsf{K}(j, i)$ mapping $\mathcal{H}_i$ onto $\mathcal{H}_i$, i.e., for each k_i

$$|Q_{k_i}\rangle = \mathsf{K}(j, i)\,|k_i\rangle; \tag{53}$$

thus we may write (52) as

$$K_{k_jk_i} = \langle k_j|\,\mathsf{K}(j, i)\,|k_i\rangle/(\langle k_i|\,\mathsf{K}^+\mathsf{K}\,|k_i\rangle)^{\frac{1}{2}}. \tag{54}$$

With these representations for $T_{k_jk_i}$ and P, we may write the $\mathcal{H}$ representation for the predictive random-walk equation as

$$\frac{\langle k_j \mid \alpha(j)\rangle\langle\alpha(j) \mid k_j\rangle}{\langle\alpha(j) \mid \alpha(j)\rangle} = \sum_{ki} \frac{\langle k_j|\,\mathsf{K}(j, i)\,|k_i\rangle\langle k_i|\,\mathsf{K}^+(j, i)\,|k_j\rangle\langle k_i \mid \alpha(i)\rangle\langle\alpha(i) \mid k_i\rangle}{\langle k_i|\,\mathsf{K}^+(j, i)\mathsf{K}(j, i)\,|k_i\rangle\langle\alpha(i) \mid \alpha(i)\rangle}; \tag{55}$$

clearly, from this development, an $\mathcal{H}$ representation can be generated for the retrodictive equation (30). This equation would be given by

$$\frac{\langle k_i \mid \alpha(i)\rangle\langle\alpha(i) \mid k_i\rangle}{\langle\alpha(i) \mid \alpha(i)\rangle} = \sum_{kj} \frac{\langle k_i|\,\mathsf{K}'(i, j)\,|k_j\rangle\langle k_j|\,\mathsf{K}'^+(i, j)\,|k_i\rangle\langle k_j \mid \alpha(j)\rangle\langle\alpha(j) \mid k_j\rangle}{\langle k_j|\,\mathsf{K}'^+(i, j)\mathsf{K}(i, j)\,|k_j\rangle\langle\alpha(j) \mid \alpha(j)\rangle}, \tag{56}$$

where the operator $\mathsf{K}'(i, j)$ is constructed so that

$$T'_{k_ik_j} = \frac{\langle k_i \mid Q_{kj}\rangle\langle Q_{kj} \mid k_i\rangle}{\langle Q_{kj} \mid Q_{kj}\rangle} = \frac{\langle k_i|\,\mathsf{K}'(i, j)\,|k_j\rangle\langle k_j|\,\mathsf{K}'^+(i, j)\,|k_i\rangle}{\langle k_j|\,\mathsf{K}'^+(i, j)\mathsf{K}'(i, j)\,|k_j\rangle}, \tag{57}$$

the retrodictive transition probability, is reproduced. Thus we have established $\mathcal{H}$ representations for both the retrodictive and predictive random-walk equations.

RANDOM WALK AND TIME EVOLUTION IN $\mathcal{H}$

Now that we have established an $\mathcal{H}$ representation for the random walk equation, we may employ a phase choice theorem established in a previous paper[1] to establish another $\mathcal{H}$ representation for the random walk equation which will allow us to compare the dynamics of stochastic and quantum theory.

This theorem demonstrates the existence of choices for the phases of the sequence of products

$$\{K_{k_jk_i}\alpha_{k_i}\}_{k_i=1,2,\cdots}$$

such that Eq. (55) factors to yield (see Appendix A for this theorem and its connection here)

$$\frac{\langle k_j \mid \alpha(j)\rangle}{(\langle\alpha(j) \mid \alpha(j)\rangle)^{\frac{1}{2}}} = \sum_{ki} \frac{\langle k_j|\,\mathsf{K}(j, i)\,|k_i\rangle\langle k_i \mid \alpha(i)\rangle}{\langle k_i|\,\mathsf{K}^+\mathsf{K}\,|k_i\rangle^{\frac{1}{2}}(\langle\alpha(i) \mid \alpha(i)\rangle)^{\frac{1}{2}}}. \tag{58}$$

Equation (58) provides a very simple representation in $\mathcal{H}$ for the dynamics of classical probability theory; i.e., Eq. (58) may be written

$$|\alpha'(j)\rangle = \sum_{ki} \mathsf{K}(j, i)\,\frac{|k_i\rangle\langle k_i|}{a^j_{k_i}}\,|\alpha'(i)\rangle, \tag{59}$$

where

$$a^j_{k_i} \triangleq (\langle k_i|\,\mathsf{K}^+(j, i)\mathsf{K}(j, i)\,|k_i\rangle)^{\frac{1}{2}},$$
$$|\alpha'(j)\rangle = |\alpha(j)\rangle/(\langle\alpha(j) \mid \alpha(j)\rangle)^{\frac{1}{2}}. \tag{60}$$

We can further simplify by defining the operator $\mathsf{S}(j, i)$ as

$$\mathsf{S}(j, i) \triangleq \sum_{ki} \mathsf{K}(j, i)\,\frac{|k_i\rangle\langle k_i|}{a^j_{k_i}} \tag{61}$$

so that Eq. (59) becomes

$$|\alpha'(j)\rangle = \mathsf{S}(j, i)\,|\alpha'(i)\rangle, \tag{62}$$

and we see that in a similar manner we may construct this representation for the retrodictive case

$$|\alpha'(i)\rangle = \mathsf{S}'(i, j)\,|\alpha'(j)\rangle. \tag{63}$$

Equation (62) is similar in form to the evolution equation of quantum theory, although, as we will see in the discussion to follow, the stochastic operator $\mathsf{S}(j,i)$ differs strikingly from the quantum evolution operator $\mathsf{U}(t_j,t_i)$. In addition to Eq. (62), we have Eq. (63), the retrodictive evolution equation. No such formalism appears in conventional quantum theory.

Thus we see that, for the measurement sequence $\{M_0 \to M_1 \to \cdots \to M_L\}$, there exists a collection $\{\mathsf{S}(j,i)\}$ of $\frac{1}{2}L(L+1)$ predictive stochastic operators and a collection $\{\mathsf{S}'(i,j)\}$ of $\frac{1}{2}L(L+1)$ retrodictive stochastic operators. Let us now examine the properties of $\{\mathsf{S}'(i,j)\}$ and $\{\mathsf{S}(j,i)\}$.

First we see from Eq. (61) that

$$\langle k_j|\,\mathsf{S}(j,i)\,|k_i\rangle = \frac{\langle k_j|\,\mathsf{K}(j,i)\,|k_i\rangle}{(\langle k_i|\,\mathsf{K}^+(j,i)\mathsf{K}(j,i)\,|k_i\rangle)^{\frac{1}{2}}}. \quad (64)$$

If we multiply Eq. (64) by its complex conjugate and sum over all $|k_i\rangle$, then we obtain the isometric property for S,

$$\mathsf{S}^+(j,i)\mathsf{S}(j,i) = \mathsf{I}. \quad (65)$$

However, multiplying Eq. (64) from the right by its complex conjugate, we see that S is unitary ($\mathsf{S}^+\mathsf{S} = \mathsf{S}\mathsf{S}^+ = \mathsf{I}$) only in case K is unitary. Thus we see that S is automatically isometric by construction, but can be unitary only if K is unitary. This relationship of S to K, as we shall see, has important physical implications. In order to see these implications, we must explore the properties of the collections $\{\mathsf{S}(j,i)\}$ and $\{\mathsf{S}'(i,j)\}$.

The approach to the examination of $\{\mathsf{S}(j,i)\}$ and $\{\mathsf{S}'(i,j)\}$ will be almost identical to our earlier approach when we examined the collections $\{\mathsf{T}(j,i)\}$ and $\{\mathsf{T}'(i,j)\}$, and, not surprisingly, the results will be almost identical. The complex analogs to Eqs. (31) are by the phase choice theorem

$$\begin{aligned}
\alpha_{k_1} &= \sum_{k_0} \langle k_1|\,\mathsf{S}(1,0)\,|k_0\rangle\,\alpha_{k_0},\\
\alpha_{k_2} &= \sum_{k_1} \langle k_2|\,\mathsf{S}(2,1)\,|k_1\rangle\,\alpha_{k_1} = \sum_{k_0} \langle k_2|\,\mathsf{S}(2,0)\,|k_0\rangle\,\alpha_{k_0}\\
&\;\;\vdots\\
\alpha_{k_L} &= \sum_{k_{L-1}} \langle k_L|\,\mathsf{S}(L,L-1)\,|k_{L-1}\rangle\,\alpha_{k_{L-1}}\\
&= \sum_{k_{L-2}} \langle k_L|\,\mathsf{S}(L,L-2)\,|k_{L-2}\rangle\,\alpha_{k_{L-2}}\\
&= \cdots = \sum_{k_0} \langle k_L|\,\mathsf{S}(L,0)\,|k_0\rangle\,\alpha_{k_0}.
\end{aligned} \quad (66)$$

Substituting the first of Eqs. (66) into the second equation in the set and comparing, we obtain

$$\sum_{k_0}\alpha_{k_0}\sum_{k_1}\langle k_2|\,\mathsf{S}(2,1)\,|k_1\rangle\langle k_1|\,\mathsf{S}(1,0)\,|k_0\rangle = \sum_{k_0}\langle k_2|\,\mathsf{S}(2,0)\,|k_0\rangle\,\alpha_{k_0}, \quad (67)$$

so that we obtain the $\mathcal{K}$ representation of Eq. (33),

$$\langle k_2|\,\mathsf{S}(2,0)\,|k_0\rangle = \sum_{k_1}\langle k_2|\,\mathsf{S}(2,1)\,|k_1\rangle\langle k_1|\,\mathsf{S}(1,0)\,|k_0\rangle, \quad (68)$$

which implies the multiplication theorem

$$\mathsf{S}(2,0) = \mathsf{S}(2,1)\mathsf{S}(1,0). \quad (69)$$

This procedure may be repeated for the entire set (66) to obtain the general multiplication theorem for the stochastic operator set $\{\mathsf{S}(j;i)\}$, i.e.,

$$\begin{aligned}
\mathsf{S}(L,0) &= \mathsf{S}(L,L-1)\\
&\quad\times \mathsf{S}(L-1,L-2)\cdots\mathsf{S}(2,1)\mathsf{S}(1,0)
\end{aligned} \quad (70)$$

and similarly for the retrodictive set:

$$\begin{aligned}
\mathsf{S}'(0,L) &= \mathsf{S}'(0,1)\mathsf{S}'(1,2)\cdots\mathsf{S}'(L-2,L-1)\\
&\quad\times \mathsf{S}'(L-1,L).
\end{aligned} \quad (71)$$

In addition, we have the set $\{\mathsf{S}(i,i)\}$, which by Eq. (64) and the definition of $\{\mathsf{T}(i,i)\}$ is given by

$$\{\mathsf{S}(i,i)\} = \{\mathsf{I}_i\}. \quad (72)$$

Suppose $\{\mathsf{S}(j,i)\}$ forms a subset of a group. It must be true then that each member of $\{\mathsf{S}(j,i)\}$ has an inverse. We show in Appendix B that, in case $\mathsf{S}^{-1}(j,i)$ exists, then

$$P(\hat{p}_{k_i}) = \delta_{k_i k_{i'}}, \quad (73)$$

that is, the state of the system at M_i must be precisely determined. Consider the predictive random-walk equation in case $\mathsf{S}^{-1}(j,i)$ exists for each measurement pair in the sequence:

$$P(\hat{p}_{k_i}) = \sum_{k_{L-1}}\sum_{k_{L-2}}\cdots\sum_{k_0} T_{k_L k_{L-1}}T_{k_{L-1}k_{L-2}}\cdots T_{k_1 k_0}P(\hat{p}_{k_0}), \quad (74)$$

which by (73) must reduce to

$$P(\hat{p}_{k_i}) = T_{k_L k_{L-1}}\delta_{k_{L-1}k_{L-1'}}\delta_{k_{L-2}k_{L-2'}}\cdots\delta_{k_0 k_{0'}}. \quad (75)$$

Equation (75) is the random walk equation for a system which is *deterministic* from M_0 through M_{L-1}. We see from this that, in case $\{\mathsf{S}(j,i)\}$ is a subset of a group, then the members of $\{\mathsf{S}(j,i)\}$ cannot describe the most general class of stochastic processes. The same argument applies for $\{\mathsf{S}'(i,j)\}$.

Let $\{\mathbb{S}(j,i)\}$ denote the collection of members of $\{\mathsf{S}(j,i)\}$, $\{\mathsf{S}'(i,j)\}$, and $\{\mathsf{S}(i,i)\}$. As we did for the transition matrices, we may define multiplication between members of $\{\mathsf{S}(j,i)\}$ and $\{\mathsf{S}'(i,j)\}$ and show that, for t and s each a positive integer such that $t > s$,

$$\begin{aligned}
\alpha_{k_t} &= \sum_{k_{t'}}{}' \langle k_t|\,\mathsf{S}(t,s)\mathsf{S}'(s,t)\,|k_{t'}\rangle\,\alpha_{k_{t'}},\\
\alpha_{k_s} &= \sum_{k_{s'}}{}' \langle k_s|\,\mathsf{S}'(s,t)\mathsf{S}(t,s)\,|k_{s'}\rangle\,\alpha_{k_{s'}}.
\end{aligned} \quad (76)$$

Equations (76) are satisfied in case

$$\mathsf{S}(t,s)\mathsf{S}'(s,t) = \mathsf{S}'(t,t) = \mathsf{I},$$
$$\mathsf{S}'(s,t)\mathsf{S}(t,s) = \mathsf{S}(s,s) = \mathsf{I}, \qquad (77)$$

but can be satisfied, as could Eqs. (38), without the conditions imposed by Eqs. (77). In fact, if Eqs. (77) are required of each $\mathsf{S}(j,i)$ and each $\mathsf{S}'(i,j)$, then the system described by the collection $\{\mathcal{S}(j,i)\}$ would, by Eq. (75), be completely deterministic. In addition, we see that if $\{\mathcal{S}(j,i)\}$ is to form a semigroup, then Eqs. (77) must be satisfied if multiplication between $\mathsf{S}(j,i)$ and $\mathsf{S}'(j,i)$ is to be closed in $\{\mathcal{S}(j,i)\}$. Therefore, if $\{\mathcal{S}(j,i)\}$ forms a semigroup, then it must form a group, and this group must be a *unitary group* since each $\mathcal{S} \in \{\mathcal{S}(j,i)\}$ is isometric and has an inverse.

Now suppose that $\{\mathsf{S}(j,i)\}$ forms a semigroup. As with $\{\mathsf{T}(j,i)\}$, we must require that

$$\mathsf{S}(l,k) = \mathsf{S}(x,y), \quad |l-k| = |x-y|, \qquad (78)$$

that is, $\{\mathsf{S}(j,i)\}$ can form a semigroup only if each $\mathsf{S} \in \{\mathsf{S}(j,i)\}$ is a function of the relative time. The same argument applies for $\{\mathsf{S}'(i,j)\}$.

We are now in a position to fully appreciate the difference between stochastic dynamics and quantum dynamics. First we note that the stochastic evolution operator S is, in general, isometric while the quantum evolution operator U is always unitary.

We see that in case the collection of stochastic operators $\{\mathcal{S}(j,i)\}$ for the measurement sequence $\{M_0 \to M_1 \to \cdots \to M_L\}$ forms a unitary group, then a system must follow a *deterministic* path through the measurement sequence. We also see from Appendix B that in case each member of the collection $\{\mathsf{S}(j,i)\}$ has an inverse, then $\{\mathsf{S}(j,i)\}$ is a unitary collection and Eq. (75) implies that each measurement in the sequence, except the last, yields a unique result.

Since the quantum evolution operator U always has an inverse, we see that the quantum evolution equation, when subjected to the phase choice of Appendix A, can only describe evolution corresponding to Eq. (75). In case the quantum evolution operators form a unitary group, then unitary evolution in $\mathcal{K}$ can only describe a *deterministic stochastic process* when the phase choice is imposed. Thus we see that quantum dynamics, i.e., unitary evolution in $\mathcal{K}$, can never reproduce the random walk structure of stochastic processes.

QUANTUM AND STOCHASTIC DYNAMICS IN A SINGLE $\mathcal{K}$ REPRESENTATION

From the preceding section we see that quantum dynamics and stochastic dynamics in $\mathcal{K}$ are identical only in case the quantum evolution equation is subject to the phase choice of Appendix A and the stochastic operator S is unitary. However, if the quantum evolution equation is subject to the phase choice, then the peculiar probability structure produced by the "square" of this equation disappears; on the other hand, if the stochastic operator S is unitary, then the more general singly stochastic structure of the transition matrices of stochastic processes is restricted to the doubly stochastic structure of quantum theory. Furthermore, if the phase choice is imposed on unitary evolution in $\mathcal{K}$, then the ensuing dynamical model in $\mathcal{K}$ can reproduce only a special case, given by Eq. (75), of the random walk equation (29).

In view of this, it is interesting to note that Nelson[14,15] has derived the time-dependent Schrödinger equation from the diffusion equation. However, one may readily see from Chandrasekhar's[16] derivation of the diffusion equation that the diffusion format follows from the random walk equation (29) only in case $\mathsf{T}(j,i)$ is *doubly stochastic*.

Such a result emphasizes the peculiarity of the doubly stochastic "transition" matrix of quantum theory. The quantum "transition" matrix is clearly doubly stochastic since its elements are given by

$$T_{k_j k_i} \triangleq |\langle k_j|\, \mathsf{U}(t_j, t_i)\, |k_i\rangle|^2, \qquad (79)$$

and we see from this equation that since U is unitary,

$$\sum_{k_j} T_{k_j k_i} = \sum_{k_i} T_{k_j k_i} = 1. \qquad (80)$$

However, the stochastic representation with elements

$$T_{k_j k_i} = |\langle k_j|\, \mathsf{S}(j,i)\, |k_i\rangle|^2 \qquad (81)$$

is in general not doubly stochastic since in general S is only isometric and not unitary.

The above properties of the evolution equations and the transition matrices of quantum and stochastic dynamics provide the motivation for a more general mathematical structure in $\mathcal{K}$ which will include both stochastic and quantum dynamics as a special case. To do this, we simply hypothesize that each "state" of a physical system has a representation by a member of a separable Hilbert space $\mathcal{K}$ and that the dynamical evolution of the system is described by

$$|\alpha(t)\rangle = \mathsf{S}(t, t_0)\, |\alpha(t_0)\rangle, \qquad (82)$$

where S is, in general, isometric. The quantum dynamical description is given by a unitary S, and the stochastic dynamical description is given by applying the phase choice theorem to Eq. (82). In this way, we encompass both the peculiar probability structure

provided by quantum theory and the singly stochastic transition matrix of classical stochastic theory.

CONCLUSION

We have discussed in this paper a novel formulation for the σ-algebra of stochastic chains and have seen how the sequenced event space leads to the notions of both prediction and retrodiction in stochastic theory. We have shown also that the equations for stochastic dynamics have a representation in a separable Hilbert space $\mathcal{H}$ which, in general, is distinct from the conventional quantum representation in $\mathcal{H}$. The stochastic picture in $\mathcal{H}$ suggests a more general evolution picture in $\mathcal{H}$ which includes quantum evolution and stochastic evolution as special cases.

That retrodiction in stochastic theory is possible is not surprising and, in fact, is necessary when one considers the definitions upon which stochastic theory is built. For example, consider the measuring sequence $\{M_0 \to M_1 \to \cdots \to M_L\}$. Suppose we let N systems pass this sequence one at a time, so that a moving picture camera may record the configurations assigned to a system as it passes through the sequence. Let the ith frame on the film record the result of M_i. Then the passage of a single system through the L-term measurement sequence will be recorded on an L-frame strip of film, each frame containing the result of one measurement. Suppose we record each system's passage through the sequence until we obtain N L-frame strips of motion picture film. Suppose we mark the first frame of each strip to identify the direction of time passage for each strip. We may now place the N strips into a box and shuffle them. If the configuration of the environment is fixed for the N systems, then we may operationally define the unconditional probability, for some $\hat{p}_{k_i}$ during M_i, as the number of strips $n(\hat{p}_{k_i})$ which have the configuration $\hat{p}_{k_i}$ on the ith frame divided by the total number of strips, N, i.e.,

$$P(\hat{p}_{k_i}) = n(\hat{p}_{k_i})/N. \tag{83}$$

The unconditional probability for the sequence $(C_0 \to C_1 \to \cdots \to \hat{p}_{k_i} \to C_{i+1} \to \cdots \to \hat{p}_{k_j} \to C_{j+1} \to \cdots \to C_L)$ then is simply

$$P(\hat{p}_{k_i} \to \hat{p}_{k_j}) = n(\hat{p}_{k_i} \to \hat{p}_{k_j})/N, \tag{84}$$

and the predictive conditional probability is given by

$$P(\hat{p}_{k_j} \mid \hat{p}_{k_i}) = \frac{P(\hat{p}_{k_i} \to \hat{p}_{k_j})}{P(\hat{p}_{k_i})} = \frac{n(\hat{p}_{k_i} \to \hat{p}_{k_j})}{n(\hat{p}_{k_i})}, \quad i < j. \tag{85}$$

With these operational definitions, it is then absolutely reasonable to define the retrodictive conditional probability

$$P(\hat{p}_{k_i} \mid \hat{p}_{k_j}) = \frac{P(\hat{p}_{k_i} \to \hat{p}_{k_j})}{P(\hat{p}_{k_j})} = \frac{n(\hat{p}_{k_i} \to \hat{p}_{k_j})}{n(\hat{p}_{k_j})}, \tag{86}$$

which, as we see from our example, is not anticausal in nature but is a simple result of the *a posteriori* nature of the film data.

From the above example, we see that we may interpret the predictive and the retrodictive random-walk equations in the following way: The predictive random-walk equation will describe the diffusion of a drop of cream placed in a cup of coffee. If we film this process, then the retrodictive random-walk equation will describe the "reverse diffusion process" as it appears on a projection screen when the film is run in reverse. We saw, however, from the analysis of the transition matrices, that the retrodictive transition matrix is the inverse of the predictive transition matrix only for deterministic systems.

When the stochastic equations were cast into their respective $\mathcal{H}$ representations, we saw that the predictive evolution operator S and the retrodictive evolution operator S′ defined predictive and retrodictive evolution in $\mathcal{H}$. We saw that S and S′ are isometric, but that S′ is S^{-1} only for deterministic systems. Furthermore, we saw that, in contrast to conventional quantum theory, S is unitary only for systems described by Eq. (75). Thus we saw that the stochastic $\mathcal{H}$ representation is distinct from the quantum representation so that stochastic processes cannot be considered as a special case of quantum evolution.

We then postulated a mathematical structure [Eq. (82)] in $\mathcal{H}$ which would include both quantum evolution and stochastic evolution as special cases. No basis was given for such a structure, but it is envisioned that a more general definition of the event space $\{E(C)\}$ might well produce the more general postulated structure. Recall that we required $\{E(C)\}$ to be a σ-algebra and further imposed the generalized distributive relation on $\{E(C)\}$. It is hoped that a removal of the generalized distributive requirement, or a mathematical generalization of the σ-algebraic structure of $\{E(C)\}$, or both, will produce the more general evolution picture in $\mathcal{H}$.

APPENDIX A

Suppose that each of $\{T_{k_j k_i}\}_{k_i=1,2,\ldots}$ and

$$\{P(\hat{p}_{k_i})\}_{k_i=1,2,\cdots}$$

is a sequence of positive real numbers and that there exists a real number $P(\hat{p}_{k_j})$ such that

$$P(\hat{p}_{k_j}) = \sum_{k_i=1}^{\infty} T_{k_j k_i} P(\hat{p}_{k_i}). \tag{A1}$$

Then there exists a sequence of complex numbers $\{K_{k_jk_i}\}_{k_i=1,2,\cdots}$, a sequence of complex numbers $\{\alpha_{k_j}\}_{k_i=1,2,\cdots}$, and a complex number α'_{k_j} such that the following equations are consistent:

$$\alpha_{k_j} = \sum_{k_i=1}^{\infty} K_{k_jk_i}\alpha_{k_i}, \tag{A2}$$

$$P(\hat{p}_{k_j}) = \alpha^*_{k_j}\alpha_{k_j} \tag{A3}$$

and, for each positive integer k_i,

$$T_{k_jk_i} = K^*_{k_jk_i}K_{k_jk_i}, \tag{A4}$$

$$P(\hat{p}_{k_i}) = \alpha^*_{k_i}\alpha_{k_i}. \tag{A5}$$

This theorem thus states that phases for the sequence $\{K^{\ddagger}_{k_jk_i}\alpha_{k_i}\}_{k_i=1,2,\cdots}$ can be found such that the double sum formed from the square of equation (A2) reduces to a single sum of real numbers. Equations (55) and (58) are nothing more than the $\mathcal{H}$ representations of Eqs. (A1) and (A2), respectively.

APPENDIX B

Theorem: Suppose that S is a linear continuous operator such that S^{-1} exists and I is a collection of positive integers such that k_i belongs to I only in case $\alpha_{k_i} \neq 0$. Then the equations

$$\alpha^*_{k_j}\alpha_{k_j} = \sum_{k_i} \langle k_j|\, S(j,i)\,|k_i\rangle\langle k_i|\, S^+(j,i)\,|k_j\rangle\,\alpha^*_{k_i}\alpha_{k_i} \tag{B1}$$

and

$$\alpha_{k_j} = \sum \langle k_j|\, S(j,i)\,|k_i\rangle\,\alpha_{k_i} \tag{B2}$$

are consistent only in case I has only one member, i.e.,

$$\alpha_{k_i} = \delta_{k_ik_i'}. \tag{B3}$$

Proof: Substitution of (B2) into (B1) for α_{k_j} produces

$$\alpha^*_{k_j}\sum_{k_i\in I}\langle k_j|\, S(j,i)\,|k_i\rangle\,\alpha_{k_i} = \sum_{k_i\in I}\langle k_j|\, S(j,i)\,|k_i\rangle\langle k_i|\, S^+(j,i)\,|k_j\rangle\,\alpha^*_{k_i}\alpha_{k_i}. \tag{B4}$$

Rearranging, we obtain

$$\sum_{k_i\in I}(\alpha^*_{k_j} - \langle k_i|\, S^+(j,i)\,|k_j\rangle\,\alpha^*_{k_j})\alpha_{k_i}\langle k_j|\, S(j,i)\,|k_i\rangle = 0, \tag{B5}$$

which may be written

$$\sum_{k_i\in I}(\alpha^*_{k_j} - \langle k_i|\, S\ (j,i)\,|k_j\rangle\,\alpha^*_{k_i})S(j,i) = |0\rangle. \tag{B6}$$

If S^{-1} exists, the collection $\{S\,|k_i\rangle\}_{k_i\in I}$ is a linearly independent set so that (B6) is satisfied only in case

$$(\alpha^*_{k_j} - \langle k_i|\, S^+(j,i)\,|k_j\rangle\,\alpha^*_{k_i})\alpha_{k_i} = 0, \quad k_i \in I. \tag{B7}$$

Since α_{k_i} is nonzero for each k_i in I, (B7) is satisfied only in case

$$\alpha^*_{k_j} = \langle k_i|\, S^+(j,i)\,|k_j\rangle\,\alpha^*_{k_i}, \quad k_i \in I, \tag{B8}$$

or

$$\alpha_{k_j} = \langle k_j|\, S(j,i)\,|k_i\rangle\,\alpha_{k_j}, \quad k_i \in I. \tag{B9}$$

Thus we see that (B9) is consistent with (B2) only in case the set $\{\alpha_{k_i}\}$ has only one nonzero member, i.e., I has only one member.

Let us examine the implications of this in terms of probabilities. Since

$$P(\hat{p}_{k_i}) = |\langle k_i\,|\,\alpha(i)\rangle|^2 = \alpha^*_{k_i}\alpha_{k_i} = \delta_{k_ik_i'}, \tag{B10}$$

we see that the system must be in some initial state of M. We also see then that the random walk equation yields

$$P(\hat{p}_{k_j}) = P(\hat{p}_{k_j}\,|\,\hat{p}'_{k_i'})\delta_{k_ik_i'}. \tag{B11}$$

* Supported in part by a Frederick-Gardner Cottrell Grant-in-Aid from Research Corporation.

[1] R. E. Collins, Phys. Rev. **183**, 5,1081 (1969).
[2] R. E. Collins, in *Critical Review of Thermodynamics*, E. B. Stuart, Benjamin Gal-Or, and A. J. Brainard, Eds. (Mono, Baltimore, Md., 1970).
[3] R. E. Collins, Phys.Rev. D **1**, 2, 379 (1970).
[4] R. E. Collins, Phys.Rev. D **1**, 4, 658 (1970).
[5] R. E. Collins, Phys.Rev. D **1**, 4, 666 (1970).
[6] R. E. Collins and F. G. Hall, *Proceedings of the International Conference on Thermodynamics*, P. T. Landsberg, Ed., University College, Cardiff, Wales, (April 1970), in press.
[7] P. R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950).
[8] E. Feller, *Introduction to Probability Theory and Its Applications* (Wiley, New York, 1950), Vol. 2.
[9] M. Loeve, *Probability Theory* (Van Nostrand, New York, 1955).
[10] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[11] K. L. Chung, *Markov Chains with Stationary Transition Probabilities* (Springer-Verlag, Berlin, 1960).
[12] T.-Y. Wu, *Kenetic Equations of Gases and Plasmas* (Addison-Wesley, Reading, Mass., 1966).
[13] S. K. Berberian, *Introduction to Hilbert Space* (Oxford U.P., London, 1961).
[14] E. Nelson, Phys. Rev. **150**, 4 (1966).
[15] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, N.J., 1967).
[16] S. Chandrasekhar, Rev. Mod. Phys. **15**, 17 (1943).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Powers of a Matrix and the Generalized Lucas Polynomials

YEHIEL ILAMED
Soreq Nuclear Research Centre, Yavne, Israel

(Received 5 May 1970)

The functions $L_{nk}^{(r)}(\Phi_1, \cdots, \Phi_n)$ are defined by $X^r = \sum_{k=1}^{n} L_{nk}^{(r)} X^{n-k}$, where X is an indeterminate $n \times n$ matrix and $\Phi_1, \cdots, \Phi_n$ are the invariants of X (basic symmetric functions in the eigenvalues of X). In this paper the generalized Lucas polynomial $L_{n1}^{(r)}$ is expressed explicitly as a determinant of order $r - n + 1$ or as a ratio of two determinants of order n.

1. INTRODUCTION

The functions $L_{nk}^{(r)}(\Phi_1, \cdots, \Phi_n)$, $k = 1, \cdots, n$, are defined by

$$X^r = \sum_{k=1}^{n} L_{nk}^{(r)} X^{n-k}, \tag{1}$$

where X is an indeterminate $n \times n$ matrix and $\Phi_1, \cdots, \Phi_n$ are the coefficients of the characteristic equation of X,

$$X^n = \sum_{k=1}^{n} (-1)^{k+1} \Phi_k X^{n-k}. \tag{2}$$

The function $L_{n1}^{(r)}(\Phi_1, \cdots, \Phi_n)$ is called[1] the generalized Lucas polynomial of degree $r - n + 1$ in the variables $\Phi_1, \cdots, \Phi_n$ ($L_{n1}^{(r)}$ here is $U_r^{(n)}$ in Ref. 1). It has been shown[1] that $L_{n1}^{(r)}$ is determined by the recursion relation

$$L_{n1}^{(r)} = \sum_{k=1}^{n} (-1)^{k-1} \Phi_k L_{n1}^{(r-k)} \tag{3}$$

and the initial conditions

$$L_{n1}^{(0)} = L_{n1}^{(1)} = \cdots = L_{n1}^{(n-2)} = 0, \quad L_{n1}^{(n-1)} = 1. \tag{4}$$

Explicit expressions of $L_{31}^{(r)}$ for any r and tables of $L_{31}^{(r)}$, $L_{41}^{(r)}$, and $L_{51}^{(r)}$ for $r = n + 1, \cdots, n + 5$ were recently given.[1]

In this paper $L_{n1}^{(r)}(\Phi_1, \cdots, \Phi_n)$ is expressed explicitly as a determinant of order $r - n + 1$ or as a ratio of two determinants of order n.

2. $L_{n1}^{(r)}$ AS A DETERMINANT

Equation (3) suggests the following determinant of order r as an expression for $L_{n1}^{(r+n-1)}$:

$$L_{n1}^{(r+n-1)} = \det(-\tilde{H} + \Phi_1 I - \Phi_2 H + \cdots + (-1)^{n-1} \Phi_n H^{n-1}), \tag{5}$$

where $H = e_{21} + e_{32} + \cdots + e_{r(r-1)}$, $\tilde{H}$ is the transpose of H, I is the $r \times r$ identity matrix, and e_{ij} is an $r \times r$ matrix with 1 in the position ij and zeros elsewhere. The expansion of this determinant, Eq. (5), according to the first column, gives Eq. (3). The initial conditions are fulfilled by defining a determinant of zero or negative order by one or zero, respectively.

3. $L_{nk}^{(r)}$ AS A RATIO OF TWO DETERMINANTS

Let $x_1, \cdots, x_n$ denote the eigenvalues of the indeterminate $n \times n$ matrix X. It was shown[2] that $L_{nk}^{(r)}$ can be expressed as a ratio of two alternants of order n,

$$L_{nk}^{(r)} = a_{nk}^{(r)} / v_n, \tag{6}$$

where v_n is the Vandermonde determinant with rows $x_1^{i-1}, \cdots, x_n^{i-1}$, $i = 1, \cdots, n$, and $a_{nk}^{(r)}$ is obtained from v_n by substituting the row $x_1^r, \cdots, x_n^r$ for the row $x_1^{n-k}, \cdots, x_n^{n-k}$ in v_n. From this expression of $L_{nk}^{(r)}$, which assumes the knowledge of the eigenvalues of X, one obtains

$$L_{nk}^{(r)} = a_{nk}^{(r)} v_n / v_n^2, \tag{7}$$

which can be expressed[3] as a function of traces σ_i of X^i, since $(v_n^2)_{ij} = \sigma_{i+j-2}$, and the determinant $a_{n1}^{(r)} v_n$, for example, is obtained from v_n^2 by substituting the row $\sigma_r, \sigma_{r+1}, \cdots, \sigma_{r+n-1}$ for the nth row $\sigma_{n-1}\sigma_n \cdots \sigma_{2n-2}$ of v_n^2. The traces σ_i may be expressed as polynomials in the basic symmetric functions $\Phi_1, \cdots, \Phi_n$ ($\sigma_0 = n$).

4. REMARKS

(1) By multiplying Eq. (1) successively by $\phi_{n-j} X^j$, $j = 1, \cdots, n-1$, and using Eqs. (1) and (2), one obtains for all $L_{nk}^{(r)}$, $k = 1, 2, \cdots, n$, the same recursion relations as for $L_{n1}^{(r)}$ [Eq. (3)].

(2) The expression of $L_{n1}^{(r)}$ as a determinant is generally useful for analytical calculations. For numerical calculations this expression may be more difficult to use than the recurrence relation or the explicit polynomial expression (if known).

(3) The rational expression of $L_{nk}^{(r)}$ [Eq. (7)] is only formal. For $r \gg n$, it seems that it is easier to calculate $L_{n1}^{(r)}$ by using Eq. (7) instead of Eq. (5).

[1] R. Barakat and E. Baumann, J. Math. Phys. **10**, 1474 (1969).
[2] Y. Lehrer, Rend. Circ. Mat. Palermo **6**, 103 (1957).
[3] Y. Lehrer Ilamed, *Abstracts of International Congress of Mathematicians, Stockholm 1962* (Almqvist & Wiksells, Uppsala, 1962), p. 39.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Normalization of Certain Higher-Order Phase-Integral Approximations for Wavefunctions of Bound States in a Potential Well*

STAFFAN YNGVE
Department of Theoretical Physics, University of Uppsala, Uppsala, Sweden

(Received 8 July 1970)

Furry has given a formula for the normalization integral of a physically acceptable solution of the time-independent 1-dimensional Schrödinger equation corresponding to a bound state. Considering the wavefunction of a bound state in a single-well potential and using Furry's formula and the connection formulas discussed by N. Fröman, we derive the normalization factor for the higher-order phase-integral approximations introduced earlier by N. Fröman. For the special case of the first-order JWKB approximation, the approximate normalization factor thus obtained is equivalent to that given by Pauli and Furry.

1. FURRY'S GENERAL FORMULA FOR THE NORMALIZATION INTEGRAL

Consider the 1-dimensional differential equation

$$\frac{d^2\psi}{dz^2} + Q^2(z; E)\psi = 0, \tag{1}$$

where E is a real parameter and $Q^2(z; E)$ is real for real values x of z. If Q^2 is negative when $|z|$ is sufficiently large and z is real, this differential equation has solutions tending to zero as $z \to \pm\infty$ through real values only for certain discrete values of E, which will be called E_s, where s can take the values 0, 1, 2, $\cdots$. If Q^2 has certain properties, there corresponds to every eigenvalue E_s precisely one function $\psi(z; E_s)$ except for a factor, which is independent of z. Thus $\psi(z; E_s)$ is a solution of

$$\frac{d^2}{dz^2}\psi(z; E_s) + Q^2(z; E_s)\psi(z; E_s) = 0 \tag{2}$$

such that $\psi(z; E_s) \to 0$ as $z \to \pm\infty$ through real values.

For any real value of E (which may be different from E_s), we define $\psi_1(z; E)$ as a certain solution of

$$\frac{d^2}{dz^2}\psi_1(z; E) + Q^2(z; E)\psi_1(z; E) = 0, \tag{3}$$

[cf. Eq. (1)] which tends to zero as $z \to -\infty$ through real values. This solution is uniquely determined, except for a factor which is independent of z but may depend on E. We shall choose this factor such that

$$\psi_1(z; E_s) = \psi(z; E_s). \tag{4}$$

Multiplying (2) by $\psi_1^*(z; E)$ and the complex conjugate of (3) by $-\psi(z; E_s)$, considering real values of z, which we denote by x, and adding the resulting equations, we obtain

$$\frac{d}{dx}[\psi'(x; E_s)\psi_1^*(x; E) - \psi_1^{*\prime}(x; E)\psi(x; E_s)] + [Q^2(x; E_s) - Q^2(x; E)]\psi_1^*(x; E)\psi(x; E_s) = 0, \tag{5}$$

where the prime is used for indicating differentiation with respect to x. Integrating (5) from $-\infty$ to an arbitrarily fixed point x_0, differentiating the resulting formula with respect to E, and finally putting $E = E_s$ and using (4), we get

$$\int_{-\infty}^{x_0} |\psi(x; E_s)|^2 \frac{\partial}{\partial E_s} Q^2(x; E_s)\, dx = \left[\psi_1' \frac{\partial}{\partial E}\psi_1^* - \psi_1 \frac{\partial}{\partial E}\psi_1^{*\prime}\right]_{x=x_0; E=E_s}. \tag{6}$$

For any real value of E we define $\psi_2(z; E)$ as a solution of (1), which tends to zero as $z \to +\infty$ through real values and fulfills the condition

$$\psi_2(z; E_s) = \psi(z; E_s). \tag{7}$$

Similarly as above, we get

$$\int_{x_0}^{\infty} |\psi(x; E_s)|^2 \frac{\partial}{\partial E_s} Q^2(x; E_s)\, dx = \left[\psi_2 \frac{\partial}{\partial E}\psi_2^{*\prime} - \psi_2' \frac{\partial}{\partial E}\psi_2^*\right]_{x=x_0; E=E_s}. \tag{8}$$

For the case of the Schrödinger equation, which is of interest to us, we have

$$Q^2(z; E) = (2m/\hbar^2)[E - V(z)], \tag{9}$$

and the derivative of Q^2 with respect to E appearing in (6) and (8) reduces to $2m/\hbar^2$.

Equations (6) and (8) with Q^2 given by (9), which together give the normalization integral, were obtained by Furry,[1] and formulas similar to Furry's are used in the treatment on pp 102–10 in Ref. 2. Since the energy eigenvalues are nondegenerate, it is no restriction to choose the wavefunctions to be real, as Furry[1] did.

2. BOUND STATES IN A SINGLE-WELL POTENTIAL

We shall now assume that the function $Q^2(z; E)$, which appears in (1), is given by (9) and has precisely two zeros, t_1 and t_2, on the real axis. We further

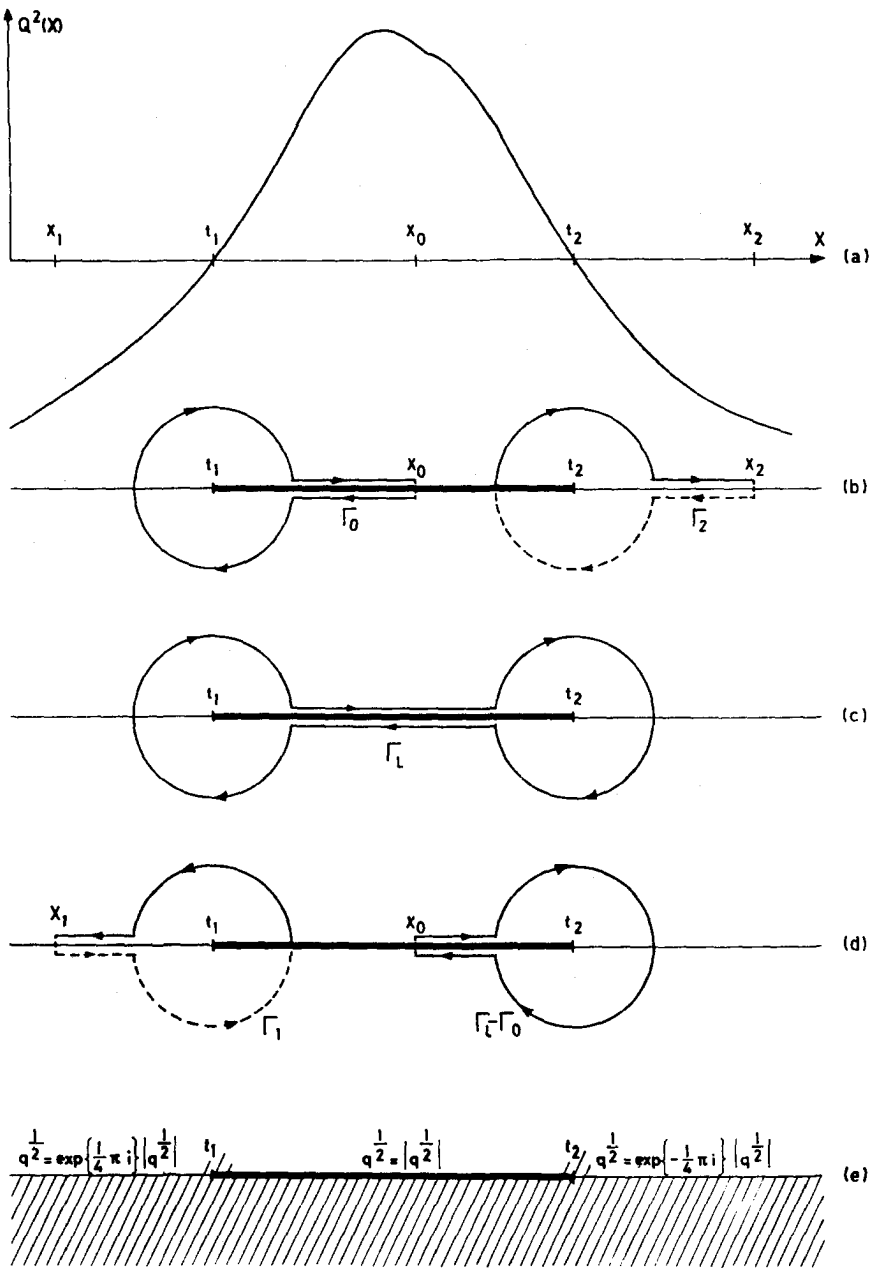


FIG. 1. (a) Qualitative behavior of $Q^2(x)$. (b), (c), (d) Contours of integration for obtaining $w(x)$. Cuts are indicated by heavy lines. The parts of the contours which lie on the second Riemann sheet are indicated by broken lines. (e) Phase of $q^{\frac{1}{2}}(z)$ on the first Riemann sheet on the upper lip of the cut along the real axis.

assume that these zeros are well separated and that no other zeros lie on or close to the real axis. It has been shown by N. Fröman[3] that two linearly independent approximate solutions of (1), called phase-integral approximations of order $2N+1$, can be written in the form $q^{-\frac{1}{2}}(z;E)\exp\{\pm iw(z;E)\}$, where

$$w(x_1;E) = w_1(x_1;E), \quad x_1 < t_1, \tag{10a}$$

$$w(x_0;E) = w_1(x_0;E) = L(E) - w_2(x_0;E), \quad t_1 < x_0 < t_2, \tag{10b}$$

$$w(x_2;E) = L(E) + w_2(x_2;E), \quad x_2 > t_2, \tag{10c}$$

with the definitions (cf. Fig. 1)

$$L(E) = \tfrac{1}{2}\int_{\Gamma_L} q(z;E)\,dz, \tag{11a}$$

$$w_1(x_1;E) = \tfrac{1}{2}\int_{\Gamma_1} q(z;E)\,dz, \tag{11b}$$

$$w_1(x_0;E) = \tfrac{1}{2}\int_{\Gamma_0} q(z;E)\,dz, \tag{11c}$$

$$w_2(x_0;E) = \tfrac{1}{2}\int_{\Gamma_L-\Gamma_0} q(z;E)\,dz = L(E) - w_1(x_0;E), \tag{11d}$$

$$w_2(x_2;E) = \tfrac{1}{2}\int_{\Gamma_2} q(z;E)\,dz, \tag{11e}$$

and

$$q(z; E) = Q(z; E)\sum_{n=0}^{N} Y_{2n}(z; E), \tag{12}$$

with the functions Y_{2n} given by the recursion formula (8) in Ref. 3 with $\lambda = 1$. It should be noted that in formulas (6) and (8) in Sec. 1 the point x_0 is arbitrary but that in this section the point x_0 is restricted to a classically allowed region. As in Ref. 3, we use two superposed Riemann sheets, which we cut along the real axis from t_1 to t_2 and join appropriately along the cut, getting a Riemann surface on which the functions $Q(z)$ and $q(z)$ are single valued. In formulas (11) the phase of $q^{\frac{1}{2}}(z)$ on the first Riemann sheet on the upper lip of the cut along the real axis is shown in Fig. 1.

The expressions for the first few functions $Y_{2n}(z; E)$ in (12) are

$$Y_0(z; E) = 1, \tag{13a}$$

$$Y_2(z; E) = \frac{\epsilon_0}{2}, \tag{13b}$$

$$Y_4(z; E) = -\frac{\epsilon_0^2}{8} - \frac{1}{8}\frac{d^2\epsilon_0}{d\zeta^2}, \tag{13c}$$

$$Y_6(z; E) = \frac{1}{32}\left[2\epsilon_0^3 - \left(\frac{d\epsilon_0}{d\zeta}\right)^2\right] + \frac{3}{32}\frac{d^2}{d\zeta^2}(\epsilon_0^2) + \frac{1}{32}\frac{d^4\epsilon_0}{d\zeta^4}, \tag{13d}$$

$$\begin{aligned} Y_8(z; E) = &-\frac{1}{128}\left[5\epsilon_0^4 - 10\epsilon_0\left(\frac{d\epsilon_0}{d\zeta}\right)^2 + \left(\frac{d^2\epsilon_0}{d\zeta^2}\right)^2\right] \\ &- \frac{1}{64}\frac{d^2}{d\zeta^2}\left[5\epsilon_0^3 - 3\left(\frac{d\epsilon_0}{d\zeta}\right)^2\right] \\ &- \frac{5}{128}\frac{d^4}{d\zeta^4}(\epsilon_0^2) - \frac{1}{128}\frac{d^6\epsilon_0}{d\zeta^6}, \end{aligned} \tag{13e}$$

with

$$\epsilon_0 = Q^{-\frac{3}{2}}\frac{d^2}{dz^2}Q^{-\frac{1}{2}} = \frac{1}{16Q^6}\left[5\left(\frac{d(Q^2)}{dz}\right)^2 - 4Q^2\frac{d^2(Q^2)}{dz^2}\right] \tag{14}$$

and

$$\zeta = \int^z Q(z; E)\, dz. \tag{15}$$

Formulas (13a), (13b), and (13c) were given in Ref. 3 whereas formulas (13d) and (13e) have been calculated by F. Karlsson and the present author.

According to Ref. 3 we have the approximate quantization condition

$$L(E_s) \approx (s + \tfrac{1}{2})\pi, \quad s = 0, 1, 2, \cdots, \tag{16}$$

and according to Ref. 4 we have the connection formulas

$$|q^{-\frac{1}{2}}(x_1)| \exp\left[-|w_1(x_1)|\right] \to 2\,|q^{-\frac{1}{2}}(x_0)| \cos\left[|w_1(x_0)| - \tfrac{1}{4}\pi\right], \tag{17a}$$

$$|q^{-\frac{1}{2}}(x_2)| \exp\left[-|w_2(x_2)|\right] \to 2\,|q^{-\frac{1}{2}}(x_0)| \cos\left[|w_2(x_0)| - \tfrac{1}{4}\pi\right]. \tag{17b}$$

Equations (16) and (17) are valid for any order $2N + 1$ of the phase-integral approximations used.

Remembering the definitions of $\psi_1(z; E)$ and $\psi_2(z; E)$ and using the connection formulas (17), we get the approximate formulas

$$\begin{cases} \psi_1(x_1; E) \approx C_1\,|q^{-\frac{1}{2}}(x_1; E)| \exp\left[-|w_1(x_1; E)|\right], \\ \qquad\qquad x_1 < t_1, \quad \text{(18a)} \\ \psi_1(x_0; E) \approx 2C_1\,|q^{-\frac{1}{2}}(x_0; E)| \cos\left[|w_1(x_0; E)| - \tfrac{1}{4}\pi\right], \\ \qquad\qquad t_1 < x_0 < t_2, \quad \text{(18b)} \end{cases}$$

$$\begin{cases} \psi_2(x_2; E) \approx C_2\,|q^{-\frac{1}{2}}(x_2; E)| \exp\left[-|w_2(x_2; E)|\right], \\ \qquad\qquad x_2 > t_2, \quad \text{(19a)} \\ \psi_2(x_0; E) \approx 2C_2\,|q^{-\frac{1}{2}}(x_0; E)| \cos\left[|w_2(x_0; E)| - \tfrac{1}{4}\pi\right], \\ \qquad\qquad t_1 < x_0 < t_2, \quad \text{(19b)} \end{cases}$$

where C_1 and C_2, which are independent of x but may depend on E, are to be determined such that (4) and (7) are fulfilled. Using (4), (7), and (16), we find from (11d), (18b), and (19b) that

$$C_2 = (-1)^s C_1 \tag{20}$$

when $E = E_s$. It is convenient to impose the condition (20) also when $E \neq E_s$. Substituting (18b) into (6), using (9), and remembering that the prime denotes differentiation with respect to x, we get after some calculations

$$\begin{aligned} &\int_{-\infty}^{x_0} |\psi(x; E_s)|^2\, dx \\ &\quad \approx \frac{2\hbar^2}{m}|C_1|^2\left\{\frac{\partial}{\partial E}w_1(x_0; E)\right. \\ &\qquad + \tfrac{1}{2}q^{-1}\frac{\partial q}{\partial E}\sin\left\{2[w_1(x_0; E) - \tfrac{1}{4}\pi)]\right\} \\ &\qquad - \tfrac{1}{2}q^{-3}q'\frac{\partial q}{\partial E}\cos^2\left[w_1(x_0; E) - \tfrac{1}{4}\pi\right] \\ &\qquad \left. + \tfrac{1}{2}q^{-2}\frac{\partial q'}{\partial E}\cos^2\left[w_1(x_0; E) - \tfrac{1}{4}\pi\right]\right\}_{E\approx E_s}. \end{aligned} \tag{21}$$

Substituting (19b) with (20) into (8) and using (9), we similarly get

$$\int_{x_0}^{\infty} |\psi(x; E_s)|^2\, dx \approx \frac{2\hbar^2}{m} |C_1|^2 \left\{ \frac{\partial}{\partial E} w_2(x_0; E) + \tfrac{1}{2} q^{-1} \frac{\partial q}{\partial E} \sin\{2[w_2(x_0; E) - \tfrac{1}{4}\pi]\} + \tfrac{1}{2} q^{-3} q' \frac{\partial q}{\partial E} \cos^2[w_2(x_0; E) - \tfrac{1}{4}\pi] - \tfrac{1}{2} q^{-2} \frac{\partial q'}{\partial E} \cos^2[w_2(x_0; E) - \tfrac{1}{4}\pi] \right\}_{E \approx E_s}. \quad (22)$$

Adding (21) and (22) and using (11d) and (16), we obtain

$$\int_{-\infty}^{\infty} |\psi(x; E_s)|^2\, dx \approx 2 |C_1|^2 \left[\frac{\hbar^2}{m} \frac{\partial}{\partial E} L(E) \right]_{E \approx E_s}. \quad (23)$$

Requiring that

$$\int_{-\infty}^{\infty} |\psi(x; E_s)|^2\, dx = 1, \quad (24)$$

we obtain from (23) and (11a) the following formula for the normalization factor C_1:

$$\frac{1}{|C_1|^2} \approx 2 \left[\frac{\hbar^2}{m} \frac{\partial}{\partial E} \frac{1}{2} \int_{\Gamma_L} q(z; E)\, dz \right]_{E \approx E_s}. \quad (25)$$

It is to be observed that formula (25) has been obtained by using only approximations and restrictions connected with the use of the phase-integral approximations and that (25) is valid for any order of these approximations used.

We shall now discuss the derivative

$$\frac{\partial}{\partial E} \frac{1}{2} \int_{\Gamma_L} q(z; E)\, dz.$$

Using (12) and (13), we obtain

$$\frac{1}{2} \int_{\Gamma_L} q(z; E)\, dz = \frac{1}{2} \int_{\Gamma_L} Q(z; E) \Big\{ \underset{(1)}{1} + \frac{\epsilon_0}{\underset{(3)}{2}} - \frac{\epsilon_0^2}{\underset{(5)}{8}} + \frac{1}{32} \Big[2\epsilon_0^3 - \underset{(7)}{\left(\frac{d\epsilon_0}{d\zeta} \right)^2} \Big] - \frac{1}{128} \Big[5\epsilon_0^4 - 10\epsilon_0 \underset{(9)}{\left(\frac{d\epsilon_0}{d\zeta} \right)^2} + \left(\frac{d^2\epsilon_0}{d\zeta^2} \right)^2 \Big] + \cdots \Big\} dz, \quad (26)$$

since the contour Γ_L is closed and integrals of total derivatives with respect to ζ therefore vanish. In (26) the numbers in parentheses below the terms indicate the contributions from successive orders of the approximation [cf. (12) and (13)]. Using the formula

$$\frac{\partial}{\partial E} \frac{d}{d\zeta} = \frac{d}{d\zeta} \frac{\partial}{\partial E} - \frac{m}{\hbar^2} Q^{-2} \frac{d}{d\zeta}, \quad (27)$$

which follows from (9) and (15), and defining

$$\eta = Q^2 \frac{\hbar^2}{m} \frac{\partial \epsilon_0}{\partial E} = -\frac{1}{8Q^6} \left[15 \left(\frac{d(Q^2)}{dz} \right)^2 - 8Q^2 \frac{d^2(Q^2)}{dz^2} \right], \quad (28)$$

we obtain from (26) after some partial integrations the formula

$$\frac{\hbar^2}{m} \frac{\partial}{\partial E} \frac{1}{2} \int_{\Gamma_L} q(z; E)\, dz = \frac{1}{2} \int_{\Gamma_L} Q^{-1}(z; E) \Big\{ \underset{(1)}{1} + \frac{(\epsilon_0 + \eta)}{\underset{(3)}{2}} - \frac{\epsilon_0(\epsilon_0 + 2\eta)}{\underset{(5)}{8}} + \frac{1}{32} \Big[\underset{(7)}{2\epsilon_0^2(\epsilon_0 + 3\eta)} + \left(\frac{d\epsilon_0}{d\zeta} \right)^2 + 2\eta \frac{d^2\epsilon_0}{d\zeta^2} \Big] - \frac{1}{128} \Big[5\epsilon_0^3(\epsilon_0 + 4\eta) + 10(\epsilon_0 + \eta) \underset{(9)}{\left(\frac{d\epsilon_0}{d\zeta} \right)^2} - \left(\frac{d^2\epsilon_0}{d\zeta^2} \right)^2 + 20\epsilon_0\eta \frac{d^2\epsilon_0}{d\zeta^2} + 2 \frac{d\epsilon_0}{d\zeta} \frac{d^3\epsilon_0}{d\zeta^3} + 2\eta \frac{d^4\epsilon_0}{d\zeta^4} \Big] \Big\} dz, \quad (29)$$

in which only terms up to the phase-integral approximations of ninth order are included.

It should be noted that (26) and (29) still hold if the function $V(z)$ in (9) is any analytic function and if Γ_L is replaced by any closed contour.

When (29) is substituted into (25), one obtains a formula for the normalization constant C_1. For the special case of the first-order JWKB approximation, (25) with (29) reduces to the normalization formula given by Pauli[5] and by Furry.[1]

ACKNOWLEDGMENTS

I am most grateful to Professor P. O. Fröman, who has suggested the problem and has given me much help in the course of the work. I also want to express my gratitude to Dr. N. Fröman, who has made several valuable comments on the manuscript.

* Work supported in part by the Swedish Natural Science Research Council.

[1] W. H. Furry, Phys. Rev. **71**, 366 (1947).

[2] J. Heading, *Introduction to Phase-Integral Methods*, Methuen's Monographs on Physical Subjects (Wiley, New York, 1962).

[3] N. Fröman, Arkiv Fysik **32**, 541 (1966).

[4] N. Fröman, Ann. Phys. (N.Y.) **61**, No. 2 (December, 1970) (to be published).

[5] W. Pauli, "Die allgemeinen Prinzipien der Wellenmechanik," *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Springer-Verlag, Berlin, 1933), Vol. XXIV/1, p. 173.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Conformal Charge Conjugation

ANATOLE P. HRISTEV
Faculty of Physics, University of Bucharest, Bucharest, Romania

(Received 4 September 1969; Revised Manuscript Received 29 April 1970)

Conformal spinor calculus for an arbitrary Euclidean space E_n, n even, $n = 2\nu$, is developed, and fundamental spin tensors for Clifford algebra Cl_{n+2} of the space of representation E_{n+2} are calculated. It is found that the *conformal* charge conjugate of a 2^ν-semispinor $\psi_c^\pm$ differs from the relativistic, conventional $\varphi_c^\pm$, by the permutation of the semispinors and the factor

$$\gamma_{n+1} = \frac{1}{n!}\frac{i^{t-\nu}}{|g|^{\frac{1}{2}}}\gamma_{[1}\cdots\gamma_{n]}, \quad t = \text{number of timelike dimensions of } E_n;$$

namely,

$$\psi_c^\pm = \mp\gamma_{n+1}\varphi_c^\mp \quad \text{for } \nu \text{ even, } t \text{ odd.}$$

This is related to the Pauli–Gürsey isospin group. The transformation laws for spinors and conjugate spinors under conformal group are studied. Conformal identities for matrix elements and bilinear covariants are indicated.

1. INTRODUCTION

The conformal group of the space–time especially was studied after the discovery of conformal invariance of the Maxwell equations.[1] It is possible that conformal field equations for other zero-rest-mass particles (graviton, neutrino) are valid. It is therefore important to investigate the conformal properties of the matter on the basis of group-theoretical method, independently of any field equations.[2] Some arguments for such an investigation are given in this introductory section.

Although the Dirac equation for electron is not invariant under conformal transformations, the commutation relations between the 15 Dirac–Pauli matrices are characteristic of the generators of the conformal group of the space–time or of the isomorphic pseudo-orthogonal group in six dimensions.

The conformal structure of the Dirac matrix algebra *as a whole* can be presented in a fully *conformal covariant form*, independently of any field equation.[3] Namely, the 15 Dirac matrices are unified in two antisymmetric conformal "matrix tensors"

$$\begin{aligned} &\mathsf{i}^\pm_{ab} = -\mathsf{i}^\pm_{ba}, && a, b = 1, \cdots, 6; \\ &\mathsf{i}^\pm_{jk} = -\tfrac{1}{2}i\gamma_{[j}\gamma_{k]}, && \mathsf{i}^\pm_{5k} = -i\gamma_5\gamma_k, \\ &\mathsf{i}^\pm_{6k} = \mp i\gamma_k, && \mathsf{i}^\pm_{65} = \mp i\gamma_5, \end{aligned} \tag{1}$$

$$\gamma_j\gamma_k + \gamma_k\gamma_j = 2g_{jk}, \quad \gamma_5 = \frac{1}{i}\frac{1}{4!}\frac{1}{|g|^{\frac{1}{2}}}\gamma_{[1}\cdots\gamma_{4]}, \quad g = \det g_{jk}. \tag{2}$$

The indices $\pm$ correspond to the two 4-dimensional irreducible spinor representations of the proper conformal group (the generators are $\mathsf{I}^\pm_{ab} = \frac{1}{2}i\mathsf{i}^\pm_{ab}$). If δ_{ab} is the metric tensor and ϵ^{abcdef} the antisymmetric unit pseudotensor of the Euclidean space of representation E_6 $(+++-+-)$, then the commutation relations are

$$[\mathsf{i}_{ab}, \mathsf{i}_{cd}] = 2i\delta_{|ac}\mathsf{i}_{bd|}, \quad \mathrm{Tr}\,\mathsf{i}_{ab} = 0, \tag{3}$$

where $|\cdots|$ is the sum over cyclic permutation of indices (we omit the indices $\pm$ in such obvious cases), and the *covariant* law of multiplication for all the 15 matrices is

$$\mathsf{i}_{ab}\mathsf{i}_{cd} = \delta_{a[c}\delta_{d]b} + i\delta_{|ac}\mathsf{i}_{bd|} \pm \tfrac{1}{2}\epsilon_{abcdef}\mathsf{i}^{ef}, \tag{4}$$

whence, in particular,

$$\mathsf{i}_{ab} = \pm(1/4!)\epsilon_{abcdef}\mathsf{i}^{cd}\mathsf{i}^{ef}, \quad 1 = \pm(1/6!)\epsilon_{abcdef}\mathsf{i}^{ab}\mathsf{i}^{cd}\mathsf{i}^{ef}, \tag{5}$$

$$\begin{aligned} \gamma_j\gamma_k &= g_{jk} + \tfrac{1}{2}i\epsilon_{jkmn}\gamma^m\gamma^n\gamma_5, \\ \gamma_{a'}\gamma_{b'} &= \delta_{a'b'} + (i/3!)\epsilon_{a'b'c'd'e'}\gamma^{c'}\gamma^{d'}\gamma^{e'}, \\ & \qquad a', \cdots = 1, \cdots, 5. \end{aligned} \tag{6}$$

The fundamental formula (4) permits a *conformal* derivation of bilinear relations for matrix elements:

$$\begin{aligned} &\tfrac{1}{2}(\mathsf{i}_{ab})^j_m(\mathsf{i}^{ab})^k_n + \delta^j_m\delta^k_n = 4\delta^k_m\delta^j_n, \\ &\delta^j_m(\mathsf{i}_{ab})^k_n + (\mathsf{i}_{ab})^j_m\delta^k_n \pm \epsilon_{abcdef}(\mathsf{i}^{cd})^j_m(\mathsf{i}^{ef})^k_n \\ &\qquad = 2\delta^j_n(\mathsf{i}_{ab})^k_m + 2(\mathsf{i}_{ab})^j_n\delta^k_m, \\ &\delta^j_m(\mathsf{i}_{ab})^k_n - (\mathsf{i}_{ab})^j_m\delta^k_n = \tfrac{1}{2}i(\mathsf{i}_{f[a})^j_n(\mathsf{i}^f{}_{b]})^k_m, \\ &\tfrac{1}{2}(\mathsf{i}_{f|a})^{|j}_m(\mathsf{i}^f{}_{b|})^{k]}_n = \delta_{ab}\delta^{|j}_m\delta^{k]}_n, \end{aligned} \tag{7}$$

where (anti)symmetrizations are made independently over the corresponding sort of indices. These conformal identities, based exclusively on the commutation relation (2), unify conformally the identities established by Pauli and Kofink,[4] and serve to derive

identities for bilinear covariants (Dirac-type tensors or de Broglie-type tensors in his fusion method). Some of these are indicated and discussed in the last section.

In the present work, conformal spinor calculus is developed, and conformal charge conjugation especially is studied. For generality, all considerations are made for an arbitrary Euclidean or pseudo-Euclidean n-dimensional space (n even, $n = 2\nu$). Thus we can explain, for example, the normalization factors and the symmetry properties of spin tensors. The principal application is, of course, to the 4-dimensional pseudo-Euclidean space–time ($n = 4$). However, the results can also be applied, for example, to the Lorentz group itself, if considered as locally isomorphic to the conformal group of a conformal Euclidean plane ($n = 2$).

For Minkowski space it follows that the conformal adjunct spinor coincides with the conventional (relativistic) adjunct spinor, but the *conformal charge conjugate* spinor differs from the conventional (Lorentz) charge conjugate spinor by permutation of the semispinors and by the *factor* γ_5. The product $\gamma_5\varphi_c$ has, therefore, a *conformal* meaning. This is just the combination which appears in the transformations of the Pauli–Gürsey isospin group (see below). The conformal charge conjugate spinor can also be obtained from the Lorentz charge conjugate spinor by inversion with respect to the "spacelike hypersphere" $g_{jk}x^jx^k - R^2 = 0$.

2. MATRICES

The conformal group of a conformal (pseudo-) Euclidean space H_n is locally isomorphic with the pseudo-orthogonal group R_{n+2} of a pseudo-Euclidean space of representation E_{n+2} with supplementary signature $(+\ -)$. The subalgebra of even aggregates of a Clifford algebra $\mathrm{Cl}_{2\nu+2}$, ν an integer, is isomorphic to $\mathrm{Cl}_{2\nu+1}$ which, in its turn, decomposes in a direct sum of two Clifford algebras, each isomorphic to $\mathrm{Cl}_{2\nu}$. The spinor representation of the group R_{n+2}, $n = 2\nu$, which for the proper (R^{p}_{n+2}) and "space–time-like" (R^{st}_{n+2}) components decomposes therefore in a direct sum, can be constructed in the following two simple ways, σ being Pauli matrices:

$$\beta_{a'} = \gamma_{a'} \times \sigma_1, \quad \beta_{n+2} = i \times \sigma_2, \quad \beta \overset{\mathrm{DEF}}{=} 1 \times \sigma_3, \tag{8a}$$

$$\beta_{a'} = i\gamma_{n+1}\gamma_{a'} \times \sigma_1, \quad \beta_{n+2} = i \times \sigma_2, \tag{8b}$$

$$a', b' = 1, \cdots, n, n+1$$

(σ_1 and σ_2 can be interchanged). Here γ_j, $j, k = 1, \cdots, n$, are 2^ν-order Dirac–Pauli-type matrices of the starting Euclidean space E_n:

$$\gamma_j\gamma_k + \gamma_k\gamma_j \equiv \{\gamma_j, \gamma_k\} = 2g_{jk}, \quad \{\gamma_{n+1}, \gamma_k\} = 0, \tag{9}$$

$$\gamma_{n+1} \overset{\mathrm{DEF}}{=} \frac{1}{n!}\frac{i^{t-\nu}}{|g|^{\frac{1}{2}}}\gamma_{[1}\cdots\gamma_{n]}, \quad \gamma_{n+1}^2 = 1, \tag{10}$$

where t is the number of timelike dimensions of E_n, $n = 2\nu$, and g_{jk} is the metric tensor of E_n, $g = \det g_{jk}$. The $2^{\nu+1}$-order Dirac–Pauli-type matrices β_a of the space of representation E_{n+2} consequently satisfy the commutation rules

$$\{\beta_a, \beta_b\} = 2\delta_{ab}, \quad \{\beta, \beta_a\} = 0, \tag{11}$$
$$a, b, \cdots = 1, \cdots, n, n+1, n+2,$$

with the metric tensor of the space of representation E_{n+2} being

$$(\delta_{ab}) = \left(\begin{array}{c|cc} g_{jk} & 0 & \\ \hline & 1 & 0 \\ 0 & 0 & -1 \end{array}\right), \quad \det\delta_{ab} = -\det g_{jk}. \tag{12}$$

Although the representation (8b) presents some advantages (e.g., for inversions, etc.),we shall use the representation (8a).

3. MORPHISMS OF $\mathrm{Cl}_{2\nu}$

Let the fundamental spin tensors of $\mathrm{Cl}_{2\nu}$ (which give the morphisms of this algebra) be[5]:

(a) $\mathsf{A} = \|A^{\mu\nu}\|$, $\lambda, \mu, \nu = 1, \cdots, 2^\nu$, which gives the anti-automorphism of $\mathrm{Cl}_{2\nu}$ of E_n, $n = 2\nu$ ($\sim$ means transposition),

$$\gamma_j = \mathsf{A}\tilde{\gamma}_j\mathsf{A}^{-1}, \quad (-1)^\nu\gamma_{n+1} = \mathsf{A}\tilde{\gamma}_{n+1}\mathsf{A}^{-1},$$
$$\tilde{\mathsf{A}} = (-1)^{\frac{1}{2}\nu(\nu-1)}\mathsf{A}; \tag{13}$$

(b) $\mathsf{B} = \|B^{\mu\nu}\|$,which gives the anti-automorphism

$$-\gamma_j = \mathsf{B}\tilde{\gamma}_j\mathsf{B}^{-1}, \quad (-1)^\nu\gamma_{n+1} = \mathsf{B}\tilde{\gamma}_{n+1}\mathsf{B}^{-1},$$
$$\tilde{\mathsf{B}} = (-1)^{\frac{1}{2}\nu(\nu+1)}\mathsf{B}, \tag{14}$$

$$\mathsf{A} = (-1)^\nu\gamma_{n+1}\mathsf{B} = \mathsf{B}\tilde{\gamma}_{n+1},$$
$$\mathsf{B} = (-1)^\nu\gamma_{n+1}\mathsf{A} = \mathsf{A}\tilde{\gamma}_{n+1}; \tag{15}$$

(c) $\mathsf{C} = \|C^{\mu}_{\dot{\nu}}\|$, which gives the pseudo-automorphism (dotted indices transform by means of the complex conjugate matrix of transformation; the dot denotes complex conjugation, and the asterisk, Hermitian conjugation)

$$\gamma_j = \mathsf{C}\dot{\gamma}_j\mathsf{C}^{-1}, \quad (-1)^{\nu-t}\gamma_{n+1} = \mathsf{C}\dot{\gamma}_{n+1}\mathsf{C}^{-1}, \tag{16}$$
$$\mathsf{C}\dot{\mathsf{C}} = \pm 1 \text{ according to}$$
$$\nu - t \equiv 0, 1 \text{ or } 2, 3 \pmod 4;$$

$$\text{(d) } \mathsf{D} = \|D^{\mu}_{\dot\nu}\|, \quad -\gamma_j = \mathsf{D}\dot\gamma_j\mathsf{D}^{-1}, \qquad (-1)^{\nu-t}\gamma_{n+1} = \mathsf{D}\dot\gamma_{n+1}\mathsf{D}^{-1}, \qquad \mathsf{DD} = \pm 1 \text{ according to } \nu - t \equiv 0, 3 \text{ or } 1, 2 \pmod 4, \tag{17}$$

$$\mathsf{D} = (-1)^{\nu}\gamma_{n+1}\mathsf{C} = (-1)^{t}\mathsf{C}\dot\gamma_{n+1}, \qquad \mathsf{C} = (-1)^{\nu}\gamma_{n+1}\mathsf{D} = (-1)^{t}\mathsf{D}\dot\gamma_{n+1}; \tag{18}$$

$$\text{(e) } \mathsf{E} = \|E_{\dot\mu\nu}\|, \quad \gamma_j^* = \mathsf{E}\gamma_j\mathsf{E}^{-1}, \qquad (-1)^{t}\gamma_{n+1}^* = \mathsf{E}\gamma_{n+1}\mathsf{E}^{-1}, \tag{19}$$

$$\tilde{\mathsf{E}} = \mathsf{A}^{-1}\mathsf{C} = \mathsf{B}^{-1}\mathsf{D}. \tag{20}$$

$$\text{(f) } \mathsf{F} = \|F_{\dot\mu\dot\nu}\|, \quad -\gamma_j^* = \mathsf{F}\gamma_j\mathsf{F}^{-1}, \qquad (-1)^{t}\gamma_{n+1}^* = \mathsf{F}\gamma_{n+1}\mathsf{F}^{-1}, \tag{21}$$

$$\tilde{\mathsf{F}} = \mathsf{B}^{-1}\mathsf{C} = \mathsf{A}^{-1}\mathsf{D}, \tag{22}$$

$$\mathsf{E} = \mathsf{F}\gamma_{n+1}, \quad \mathsf{F} = \mathsf{E}\gamma_{n+1}. \tag{23}$$

C (or D) can always be chosen so that E or F (sometimes both) become Hermitian.

For basis multivectors and pseudomultivectors, it follows that

$$\begin{aligned}(\gamma_{[j_1}\cdots\gamma_{j_k]})^* &= (-1)^{\frac{1}{2}k(k-1)}\mathsf{E}\gamma_{[j_1}\cdots\gamma_{j_k]}\mathsf{E}^{-1} \\ &= (-1)^{\frac{1}{2}k(k+1)}\mathsf{F}\gamma_{[j_1}\cdots\gamma_{j_k]}\mathsf{F}^{-1}, \\ (\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]})^* &= (-1)^{t+\frac{1}{2}k(k+1)}\mathsf{E}\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\mathsf{E}^{-1} \\ &= (-1)^{t+\frac{1}{2}k(k-1)}\mathsf{F}\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\mathsf{F}^{-1}.\end{aligned} \tag{24}$$

If E (or F) is Hermitian, one can write

$$(\mathsf{E}\gamma_{[j_1}\cdots\gamma_{j_k]})^* = (-1)^{\frac{1}{2}k(k-1)}\mathsf{E}\gamma_{[j_1}\cdots\gamma_{j_k]}, \quad \text{etc.}, \tag{24'}$$

i.e., the basis (pseudo)multivectors are E (or F)-(anti-) Hermitian.

4. ADJUNCT AND CONJUGATE SPINORS

We define the adjunct spinor, the conjugate, and "adjunct conjugate" spinors as follows:

$$\begin{aligned}&\bar\varphi \equiv \bar\varphi_F = \varphi^*\mathsf{F}, \quad \varphi_{\mathrm{c}} = \mathsf{C}\dot\varphi = \mathsf{B}\tilde{\bar\varphi}, \\ &\bar\varphi_{\mathrm{c}} = \varphi_{\mathrm{c}}^*\mathsf{F} = \pm\widetilde{\mathsf{B}^{-1}\varphi}, \quad (\varphi_{\mathrm{c}})_{\mathrm{c}} = \pm\varphi, \\ &\qquad \text{with } \pm \text{ according to } \nu - t \equiv 0, 1 \text{ or } 2, 3 \pmod 4.\end{aligned} \tag{25}$$

One can choose another definition, e.g.,

$$\begin{aligned}&\bar\varphi_E = \varphi^*\mathsf{E}, \quad \varphi_D = \mathsf{D}\dot\varphi = \mathsf{B}\tilde{\bar\varphi}_E = \mathsf{A}\tilde{\bar\varphi}_F, \\ &(\varphi_D)_D = \pm\varphi, \quad \text{with } \pm \text{ according to } \nu - t \equiv 0, 3 \text{ or } 1, 2 \pmod 4.\end{aligned} \tag{25'}$$

We shall choose, however, the usual definition (25).

5. ROTATIONS OF E_n

Under a (pseudo)rotation $R(c^j{}_k)$ of E_n, the spinors of the spinor space S_{2^ν} undergo the corresponding transformation

$$x'^j = c^j{}_k x^k, \quad \varphi' = \pm\mathsf{S}\varphi, \quad c^j{}_k\gamma_j = \mathsf{S}\gamma_k\mathsf{S}^{-1}, \qquad \operatorname{sgn}|c^j{}_k|\cdot\gamma_{n+1} = \mathsf{S}\gamma_{n+1}\mathsf{S}^{-1}. \tag{26}$$

The versors S can be normalized in different ways. We choose the *normalization for which* φ_{c} *transforms identically with*[6] φ:

$$\mathsf{SC}\dot{\mathsf{S}}^{-1} = \mathsf{C}, \quad \mathsf{SB}\tilde{\mathsf{S}} = \eta\mathsf{B}, \quad \eta \overset{\text{DEF}}{=} \begin{cases}\operatorname{sgn}|c^{t'}{}_t|, & t \text{ odd}, \\ \operatorname{sgn}|c^{s'}{}_s|, & t \text{ even}.\end{cases} \tag{27}$$

where t and t' are timelike indices, s and s' are spacelike indices of E_n, and $|c^{t'}{}_t|$ and $|c^{s'}{}_s|$ are determinants in a (pseudo-)orthonormalized basis of E_n. It follows for the rest of fundamental spin tensors that

$$\begin{aligned}&\mathsf{SA}\tilde{\mathsf{S}} = \operatorname{sgn}|c^j{}_k|\cdot\eta\mathsf{A}, \quad \mathsf{SD}\dot{\mathsf{S}}^{-1} = \operatorname{sgn}|c^j{}_k|\cdot\mathsf{D}, \\ &\mathsf{S}^* = \eta\mathsf{F}\mathsf{S}^{-1}\mathsf{F}^{-1} = \operatorname{sgn}|c^j{}_k|\cdot\eta\mathsf{E}\mathsf{S}^{-1}\mathsf{E}^{-1},\end{aligned} \tag{27'}$$

i.e., the versors S are A^{-1} (or B^{-1})-(pseudo-)(anti-) orthogonal and E (or F)-(pseudo-)(anti-)unitary.

In this normalization the spinors transform as follows:

$$\varphi' = \mathsf{S}\varphi, \quad \varphi'_{\mathrm{c}} = \mathsf{S}\varphi_{\mathrm{c}}, \quad \bar\varphi' = \eta\bar\varphi\mathsf{S}^{-1}, \quad \bar\varphi'_{\mathrm{c}} = \eta\bar\varphi_{\mathrm{c}}\mathsf{S}^{-1}. \tag{28}$$

6. BILINEAR COVARIANTS

The bilinear covariants and pseudocovariants (Dirac-type tensors) transform as follows ($^1\varphi$ and $^2\varphi$ are two arbitrary spinors):

$$\begin{aligned}p_{j_1\cdots j_k} &\equiv {}^{21}p_{j_1\cdots j_k} \\ &\overset{\text{DEF}}{=} i^{\frac{1}{2}k(k+1)}(k!)^{-1}\,{}^2\bar\varphi\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^1\varphi \\ &= \eta c^{l_1}{}_{j_1}\cdots c^{l_k}{}_{j_k}p'_{l_1\cdots l_k},\end{aligned} \tag{29}$$

$$\begin{aligned}\tilde p_{j_1\cdots j_k} &\equiv {}^{21}\tilde p_{j_1\cdots j_k} \\ &\overset{\text{DEF}}{=} i^{t+\frac{1}{2}k(k-1)}(k!)^{-1}\,{}^2\bar\varphi\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^1\varphi \\ &= \eta c^{l_1}{}_{j_1}\cdots c^{l_k}{}_{j_k}\tilde p'_{l_1\cdots l_k}, \\ {}^{21}p^*_{j_1\cdots j_k} &= {}^{12}p_{j_1\cdots j_k}, \quad {}^{21}\tilde p^*_{j_1\cdots j_k} = {}^{12}\tilde p_{j_1\cdots j_k};\end{aligned} \tag{29'}$$

and charge conjugate (pseudo)covariants transform as follows:

$$\begin{aligned}p^{\mathrm{c}}_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{\frac{1}{2}k(k+1)}(k!)^{-1}\,{}^2\bar\varphi_{\mathrm{c}}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^1\varphi_{\mathrm{c}} \\ &= (-1)^{\frac{1}{2}[\nu(\nu+1)+k(k+1)]}p^*_{j_1\cdots j_k}, \\ \tilde p^{\mathrm{c}}_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{t+\frac{1}{2}k(k-1)}(k!)^{-1}\,{}^2\bar\varphi_{\mathrm{c}}\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^1\varphi_{\mathrm{c}} \\ &= (-1)^{\frac{1}{2}[\nu(\nu-1)+k(k-1)]}\tilde p^*_{j_1\cdots j_k}.\end{aligned} \tag{30}$$

Further, we define the de Broglie-type (pseudo)-covariants as

$$\begin{aligned} {}^{21}q_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{\frac{1}{2}k(k+1)}(k!)^{-1}\,{}^{2}\bar{\varphi}_c\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^{1}\varphi \\ &= (-1)^{\frac{1}{2}[\nu(\nu+1)+k(k+1)]}\,{}^{12}q_{j_1\cdots j_k}, \\ {}^{21}\tilde{q}_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{t+\frac{1}{2}k(k-1)}(k!)^{-1}\,{}^{2}\bar{\varphi}_c\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^{1}\varphi \\ &= (-1)^{\frac{1}{2}[\nu(\nu-1)+k(k-1)]}\,{}^{12}\tilde{q}_{j_1\cdots j_k}, \end{aligned} \tag{31}$$

and the corresponding conjugate (pseudo)covariants as

$$\begin{aligned} {}^{21}q^c_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{\frac{1}{2}k(k+1)}(k!)^{-1}\,{}^{2}\bar{\varphi}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^{1}\varphi_c = {}^{12}q^*_{j_1\cdots j_k} \\ &= (-1)^{\frac{1}{2}[\nu(\nu+1)+k(k+1)]}\,{}^{12}q^c_{j_1\cdots j_k}, \\ {}^{21}\tilde{q}^c_{j_1\cdots j_k} &\overset{\text{DEF}}{=} i^{t+\frac{1}{2}k(k-1)}(k!)^{-1}\,{}^{2}\bar{\varphi}\gamma_{n+1}\gamma_{[j_1}\cdots\gamma_{j_k]}\,{}^{1}\varphi_c \\ &= {}^{12}\tilde{q}^*_{j_1\cdots j_k} = (-1)^{\frac{1}{2}[\nu(\nu-1)+k(k-1)]}\,{}^{12}\tilde{q}^c_{j_1\cdots j_k}. \end{aligned} \tag{32}$$

Finally, one defines in a similar way the "differential" (pseudo)covariants.

7. FUNDAMENTAL SPIN TENSORS

The generators of the group R_{n+2} are

$$\bar{\mathsf{I}}_{ab} = \tfrac{1}{4}[\beta_a, \beta_b] \equiv \tfrac{1}{4}\beta_{[a}\beta_{b]} = \mathsf{I}^{+}_{ab} \oplus \mathsf{I}^{-}_{ab}, \tag{33}$$

$$\mathsf{I}^{\pm}_{a'b'} = \tfrac{1}{4}\gamma_{[a'}\gamma_{b']}, \quad \mathsf{I}^{\pm}_{n+2,a'} = -\mathsf{I}^{\pm}_{a',n+2} = \pm\tfrac{1}{2}\gamma_{a'}, \tag{34}$$

$$a, b = 1, \cdots, n, n+1, n+2,$$
$$a', b' = 1, \cdots, n, n+1.$$

In formulas where the signs $\pm$ appear (as indices or in the sum), we take the superior and inferior signs separately. The inequivalent representations $\mathsf{I}^{\pm}_{ab}$ differ by the sign of all $\gamma_{a'}$, $a' = 1, \cdots, n, n+1$.[7]

The fundamental spin tensors for E_{n+2} are calculated in Table I. $\bar{\mathsf{C}}$ is chosen so that for t odd $\bar{\mathsf{F}}$ and $\bar{\mathsf{E}}$ are Hermitian if F is Hermitian. For t even, $\bar{\mathsf{E}}$ is Hermitian if E is Hermitian (then $\bar{\mathsf{F}}$ is anti-Hermitian). If we choose, however, $i\bar{\mathsf{C}}$ instead of $\bar{\mathsf{C}}$, then $i\bar{\mathsf{F}}$ is Hermitian and $i\bar{\mathsf{E}}$ anti-Hermitian if F is Hermitian. All properties (13)–(23) transcribed for E_{n+2} and β's (instead of E_n and γ's) are satisfied in the table.

The generators $\bar{\mathsf{I}}_{ab} = i\bar{\mathsf{M}}_{ab}$ are $\bar{\mathsf{F}}$ (or $\bar{\mathsf{E}}$)-anti-Hermitian; consequently, the operators $\bar{\mathsf{M}}_{ab}$ are $\bar{\mathsf{F}}$ (or $\bar{\mathsf{E}}$)-Hermitian and satisfy the following commutation rules:

$$[\bar{\mathsf{M}}_{ab}, \bar{\mathsf{M}}_{cd}] = i\delta_{|ac}\bar{\mathsf{M}}_{bd|}, \quad \bar{\mathsf{M}}_{ab} = \mathsf{M}^{+}_{ab} \oplus \mathsf{M}^{-}_{ab}, \tag{35}$$

$$[\mathsf{M}^{\pm}_{ab}, \mathsf{M}^{\pm}_{cd}] = i\delta_{|ac}\mathsf{M}^{\pm}_{bd|}. \tag{35'}$$

In the "nondiagonal" ("physical") metric δ_{ab},

$$(\delta_{ab}) = \left(\begin{array}{c|cc} g_{jk} & 0 & \\ \hline & 0 & -1 \\ 0 & -1 & 0 \end{array}\right), \quad \det\delta_{ab} = -\det g_{jk}, \tag{36}$$

$$a, b = 1, \cdots, n, 0, \infty,$$

the "momenta" M_{ab} give in E_n the angular momentum operator M_{jk}, the energy–momentum operator P_j, the special-conformal operator K_j, and the dilatation

TABLE I. The fundamental spin tensors for E_{n+2}.

	I t odd ν even ($\nu - t$ odd)	II t odd ν odd ($\nu - t$ even)	III t even ν even ($\nu - t$ even)	IV t even ν odd ($\nu - t$ odd)
$\bar{\mathsf{A}}$	$\begin{matrix}0 & \mathsf{A}\\ \mathsf{A} & 0\end{matrix}$	$\begin{matrix}\mathsf{B} & 0\\ 0 & -\mathsf{B}\end{matrix}$ ($\mathsf{B} = -\gamma_{n+1}\mathsf{A}$)	$\begin{matrix}0 & \mathsf{A}\\ \mathsf{A} & 0\end{matrix}$	$\begin{matrix}\mathsf{B} & 0\\ 0 & -\mathsf{B}\end{matrix}$ ($\mathsf{B} = -\gamma_{n+1}\mathsf{A}$)
$\bar{\mathsf{B}}$	$\begin{matrix}0 & -\mathsf{A}\\ \mathsf{A} & 0\end{matrix}$ ($\mathsf{A} = \gamma_{n+1}\mathsf{B}$)	$\begin{matrix}\mathsf{B} & 0\\ 0 & \mathsf{B}\end{matrix}$	$\begin{matrix}0 & -\mathsf{A}\\ \mathsf{A} & 0\end{matrix}$ ($\mathsf{A} = \gamma_{n+1}\mathsf{B}$)	$\begin{matrix}\mathsf{B} & 0\\ 0 & \mathsf{B}\end{matrix}$
$\bar{\mathsf{C}}$	$\begin{matrix}0 & -\mathsf{D}\\ \mathsf{D} & 0\end{matrix}$ ($\mathsf{D} = \gamma_{n+1}\mathsf{C}$)	$\begin{matrix}\mathsf{C} & 0\\ 0 & \mathsf{C}\end{matrix}$	$\begin{matrix}\mathsf{C} & 0\\ 0 & \mathsf{C}\end{matrix}$	$\begin{matrix}0 & \mathsf{D}\\ -\mathsf{D} & 0\end{matrix}$ ($\mathsf{D} = -\gamma_{n+1}\mathsf{C}$)
$\bar{\mathsf{D}}$	$\begin{matrix}0 & \mathsf{D}\\ \mathsf{D} & 0\end{matrix}$	$\begin{matrix}\mathsf{C} & 0\\ 0 & -\mathsf{C}\end{matrix}$ ($\mathsf{C} = -\gamma_{n+1}\mathsf{D}$)	$\begin{matrix}-\mathsf{C} & 0\\ 0 & \mathsf{C}\end{matrix}$ ($\mathsf{C} = \gamma_{n+1}\mathsf{D}$)	$\begin{matrix}0 & \mathsf{D}\\ \mathsf{D} & 0\end{matrix}$
$\bar{\mathsf{E}}$	$\begin{matrix}\mathsf{F} & 0\\ 0 & \mathsf{F}\end{matrix}$ ($\mathsf{F} = \mathsf{E}\gamma_{n+1}$)	$\begin{matrix}\mathsf{F} & 0\\ 0 & -\mathsf{F}\end{matrix}$ ($\mathsf{F} = \mathsf{E}\gamma_{n+1}$)	$\begin{matrix}0 & \mathsf{E}\\ \mathsf{E} & 0\end{matrix}$	$\begin{matrix}0 & \mathsf{E}\\ \mathsf{E} & 0\end{matrix}$
$\bar{\mathsf{F}}$	$\begin{matrix}\mathsf{F} & 0\\ 0 & \mathsf{F}\end{matrix}$	$\begin{matrix}\mathsf{F} & 0\\ 0 & \mathsf{F}\end{matrix}$	$\begin{matrix}0 & -\mathsf{E}\\ \mathsf{E} & 0\end{matrix}$ ($\mathsf{E} = \mathsf{F}\gamma_{n+1}$)	$\begin{matrix}0 & -\mathsf{E}\\ \mathsf{E} & 0\end{matrix}$ ($\mathsf{E} = \mathsf{F}\gamma_{n+1}$)

operator D:

$$\mathsf{M}^{\pm}_{jk} = -\tfrac{1}{4}i\gamma_{[j}\gamma_{k]},$$
$$\mathsf{P}^{\pm}_{j} = \mathsf{M}^{\pm}_{j0} = \mathsf{M}^{\pm}_{j,n+1} + \mathsf{M}^{\pm}_{j,n+2} = -\tfrac{1}{2}i\gamma_j(\gamma_{n+1} \mp 1),$$
$$\mathsf{K}^{\pm}_{j} = \mathsf{M}^{\pm}_{j\infty} = \tfrac{1}{2}(\mathsf{M}^{\pm}_{j,n+2} - \mathsf{M}^{\pm}_{j,n+1}) = \tfrac{1}{4}i\gamma_j(\gamma_{n+1} \pm 1),$$
$$\mathsf{D}^{\pm} = \mathsf{M}^{\pm}_{0\infty} = \mathsf{M}^{\pm}_{n+1,n+2} = \pm\tfrac{1}{2}i\gamma_{n+1}. \tag{37}$$

For the *anticommutator* of $\mathsf{M}^{\pm}_{ab}$, one gets

$$\{\mathsf{M}_{ab}, \mathsf{M}_{cd}\} = \tfrac{1}{2}\delta_{a[c}\delta_{d]b} \pm \frac{2^{\nu-2}i^{t-1}}{(n-2)!} \times \epsilon_{abcdf_1\cdots f_{n-2}}\mathsf{M}^{f_1f_2}\cdots\mathsf{M}^{f_{n-3}f_{n-2}}, \tag{38}$$

where ϵ is the Levi-Civita antisymmetric unit pseudotensor in E_{n+2}:

$$\epsilon^{a_1\cdots a_{n+2}} = |\det \delta_{ab}|^{-\frac{1}{2}} e^{a_1\cdots a_{n+2}},$$
$$e^{1\cdots n+2} = 1, \quad e_{1\cdots n+2} = (-1)^{t+1}.$$

8. CONFORMAL MATRIX IDENTITIES

For the conformal "matrix tensors" of arbitrary E_n (n even),

$$\mathsf{i}^{\pm}_{ab} = 2\mathsf{M}^{\pm}_{ab}, \quad [\mathsf{i}_{ab}, \mathsf{i}_{cd}] = 2i\delta_{|ac}\mathsf{i}_{bd|}, \tag{39}$$

we obtain the anticommutator

$$\{\mathsf{i}_{ab}, \mathsf{i}_{cd}\} = 2\delta_{a[c}\delta_{d]b} \pm \frac{2i^{t-1}}{(n-2)!}\epsilon_{abcdf_1\cdots f_{n-2}}\mathsf{i}^{f_1f_2}\cdots\mathsf{i}^{f_{n-3}f_{n-2}} \tag{40}$$

and the following consequences:

$$\mathsf{i}_{fa}\mathsf{i}^{f}{}_{b} = (n+1)\delta_{ab} + in\mathsf{i}_{ab}, \tag{41}$$

$$\mathsf{i}_{ab}\mathsf{i}^{ab} = 1, \quad \mathsf{i}^2_{ab} = \delta_{aa}\delta_{bb} - \delta^2_{ab} \quad \text{without summation}, \tag{42}$$

$$\tfrac{1}{2}\mathsf{i}_{ab}\mathsf{i}^{ab} = 4(\tfrac{1}{2}\mathsf{M}_{jk}\mathsf{M}^{jk} + \mathsf{P}_j\mathsf{K}^j + \mathsf{K}_j\mathsf{P}^j - \mathsf{D}^2) = \tfrac{1}{2}(n+1)(n+2), \tag{43}$$

$$\mathsf{i}_{ab} = \pm\frac{i^{t-1}}{n!}\epsilon_{abc_1\cdots c_n}\mathsf{i}^{c_1c_2}\cdots\mathsf{i}^{c_{n-1}c_n}, \tag{44}$$

$$1 = \pm\frac{i^{t-1}}{(n+2)!}\epsilon_{a_1\cdots a_{n+2}}\mathsf{i}^{a_1a_2}\cdots\mathsf{i}^{a_{n+1}a_{n+2}},$$

$$[(2k)!]^{-1}\mathsf{i}_{[a_1a_2}\cdots\mathsf{i}_{a_{2k-1}a_{2k}]} = \pm\frac{i^{t-1}}{(n+2-2k)!}\epsilon_{a_1\cdots a_{n+2}}\mathsf{i}^{a_{2k+1}a_{2k+2}}\cdots\mathsf{i}^{a_{n+1}a_{n+2}}, \tag{45}$$

which give, in particular for γ_j,

$$\gamma_j\gamma_k = g_{jk} - \frac{i^{t-\nu}}{(n-2)!}\gamma_{n+1} \times \epsilon_{jkl_1\cdots l_{n-2}}\gamma^{l_1}\cdots\gamma^{l_{n-2}},$$
$$\frac{1}{k!}\gamma_{[j_1}\cdots\gamma_{j_k]} = (-1)^{\frac{1}{2}k(k+1)}\frac{i^{t-\nu}}{(n-k)!}\gamma_{n+1} \times \epsilon_{j_1\cdots j_n}\gamma^{j_{k+1}}\cdots\gamma^{j_n} \tag{46}$$

and,for $\gamma_{a'}$, $a', b' = 1, \cdots, n, n+1$,

$$\gamma_{a'}\gamma_{b'} = \delta_{a'b'} - \frac{i^{t-\nu}}{(n-1)!} \times \epsilon_{a'b'c'_1\cdots c'_{n-1}}\gamma^{c'_1}\cdots\gamma^{c'_{n-1}},$$
$$\frac{1}{k!}\gamma_{[a'_1}\cdots\gamma_{a'_k]} = (-1)^{\frac{1}{2}k(k-1)}\frac{i^{t-\nu}}{(n-k)!} \times \epsilon_{a'_1\cdots a'_{n+1}}\gamma^{a'_{k+1}}\cdots\gamma^{a'_{n+1}}. \tag{47}$$

9. CONFORMAL CHARGE CONJUGATE

We adopt for E_{n+2} the same definition (25) and the same normalization (27) as for E_n (*mutatis mutandis*). We limit ourselves in the following considerations to the case I in the Table: *t odd, v even* (the case of the relativistic space–time E_4). Then the following results:

$$\psi = \begin{pmatrix}\psi^+\\ \psi^-\end{pmatrix}, \quad \psi^{\pm} = \varphi^{\pm}, \quad \bar{\psi} = \psi^*\bar{\mathsf{F}},$$

i.e., $\bar{\psi}^{\pm} = \bar{\varphi}^{\pm} = \varphi^{\pm *}\mathsf{F}$,

$$\psi^{\pm}_c = \mp\gamma_{n+1}\varphi^{\mp}_c, \quad \bar{\psi}^{\pm}_c = \pm\bar{\varphi}^{\mp}_c\gamma_{n+1}, \tag{48}$$

where ψ is $2^{\nu+1}$-spinor in E_{n+2}, $\psi^{\pm}$ are 2^{ν}-semi-spinors in E_{n+2}, and $\varphi^{\pm}$ are 2^{ν}-spinors in E_n. Therefore, the conformal adjunct $\bar{\psi}^{\pm}$ coincides with the "relativistic" one (i.e., with respect to the group R_n) $\bar{\varphi}^{\pm}$, but the *conformal charge conjugate* $\psi^{\pm}_c$ of a 2^{ν}-spinor of E_n *differs* from the usual, relativistic conjugate $\varphi^{\pm}_c = \mathsf{C}\dot{\varphi}^{\pm} = \mathsf{B}\tilde{\bar{\varphi}}^{\pm}$ *by the permutation of the spinors and by the factor* γ_{n+1}! This is just the combination which appears in the transformations of the Pauli–Gürsey isospin group[8]:

$$\varphi' = a\varphi + b\gamma_5\varphi_c,$$
$$\gamma_5\varphi'_c = -b^*\varphi + a^*\gamma_5\varphi_c, \quad aa^* + bb^* = 1. \tag{49}$$

Moreover, in the case $\nu - t \equiv 1 \pmod 4$ (the case of relativistic space–time E_4), the spinors

$$\Phi = \begin{pmatrix}\varphi^+\\ f\varphi^-_c\end{pmatrix} = \begin{pmatrix}\varphi^+\\ f\gamma_{n+1}\varphi^+_c\end{pmatrix}$$

or

$$\Phi = \begin{pmatrix}f\varphi^+_c\\ \varphi^-\end{pmatrix} = \begin{pmatrix}-f\gamma_{n+1}\varphi^-_c\\ \varphi^-\end{pmatrix}, \quad |f| = 1, \tag{50}$$

are reproduced by conformal charge conjugation:

$$\Phi_c = \bar{C}\dot{\Phi} = f^*\Phi. \tag{51}$$

For $f = 1$ the spinors (50) are *conformally self-conjugate*! In the case of relativistic space–time E_4, one gets

$$\Phi = \begin{pmatrix} \varphi^+ \\ \gamma_5\varphi_c^+ \end{pmatrix} \quad \text{or} \quad \Phi = \begin{pmatrix} -\gamma_5\varphi_c^- \\ \varphi^- \end{pmatrix}, \tag{50'}$$

and (49) can be written in the following form (U being a 2-order matrix):

$$\Phi' = \mathsf{U}\Phi \quad \text{or} \quad \Phi' = \dot{\mathsf{U}}\Phi, \qquad \text{where} \quad \mathsf{U}\mathsf{U}^* = 1, \quad \det \mathsf{U} = 1. \tag{49'}$$

10. PROPER CONFORMAL TRANSFORMATIONS

Let the conformal transformations in polispherical coordinates[9] be

$$\xi'^a = A^a{}_b\xi^b, \quad \delta_{ab}A^a{}_cA^b{}_d = \delta_{cd}, \tag{52}$$

$$\bar{\eta} \overset{\text{DEF}}{=} \operatorname{sgn}|A^{\sigma'}{}_{\sigma}|, \quad t \text{ odd (or } \operatorname{sgn}|A^{\tau'}{}_{\tau}| \text{ for } t \text{ even}), \tag{53}$$

with $|A^{\sigma'}{}_{\sigma}|$ and $|A^{\tau'}{}_{\tau}|$ being determinants with spacelike (σ and σ') and timelike (τ and τ') indices in a pseudo-orthonormalized basis in E_{n+2}. For the proper (C^p) and space–time-like (C^{st}) conformal transformations, we have, from Table I and from (27) and (27'),

$$\bar{\mathsf{S}} = \begin{pmatrix} \mathsf{S}^+ & 0 \\ 0 & \mathsf{S}^- \end{pmatrix}, \quad \gamma_{n+1}\mathsf{B}\tilde{\mathsf{S}}^{\pm} = \bar{\eta}\mathsf{S}^{-1\mp}\gamma_{n+1}\mathsf{B},$$

$$\mathsf{S}^{\pm}\gamma_{n+1}\mathsf{C} = \gamma_{n+1}\mathsf{C}\dot{\mathsf{S}}^{\mp}, \quad \mathsf{S}^{*\pm}\mathsf{F} = \bar{\eta}\mathsf{F}\mathsf{S}^{-1\pm}, \tag{54}$$

and, in particular,

$$\gamma_{n+1}\mathsf{B}\tilde{\mathsf{M}}^{\pm}_{ab} = -\mathsf{M}^{\mp}_{ab}\gamma_{n+1}\mathsf{B}, \quad \mathsf{M}^{\pm}_{ab}\gamma_{n+1}\mathsf{C} = -\gamma_{n+1}\mathsf{C}\dot{\mathsf{M}}^{\mp}_{ab},$$
$$\mathsf{M}^{*\pm}_{ab}\mathsf{F} = \mathsf{F}\mathsf{M}^{\pm}_{ab}, \quad A^c{}_aA^d{}_b\mathsf{M}^{\pm}_{cd} = \mathsf{S}^{\pm}\mathsf{M}^{\pm}_{ab}\mathsf{S}^{-1\pm}; \tag{55}$$

therefore each of the two representations is equivalent to the conjugate or complex representation of the other (in the known sense of the Lie group theory[10]), namely

$$\mathsf{M}^{\mp}_{ab} = -\gamma_{n+1}\mathsf{B}\tilde{\mathsf{M}}^{\pm}_{ab}(\gamma_{n+1}\mathsf{B})^{-1},$$
$$\mathsf{M}^{\pm}_{ab} = -\gamma_{n+1}\mathsf{C}\dot{\mathsf{M}}^{\mp}(\gamma_{n+1}\mathsf{C})^{-1}. \tag{56}$$

The transformation law for the 2^r-spinors is

$$\varphi'^{\pm} = \mathsf{S}^{\pm}\varphi^{\pm}, \quad \bar{\varphi}'^{\pm} = \bar{\eta}\bar{\varphi}^{\pm}\mathsf{S}^{-1\pm},$$
$$\varphi_c'^{\pm} = \gamma_{n+1}\mathsf{S}^{\mp}\gamma_{n+1}\varphi_c^{\pm}, \quad \bar{\varphi}_c'^{\pm} = \bar{\eta}\bar{\varphi}_c^{\pm}\gamma_{n+1}\mathsf{S}^{-1\mp}\gamma_{n+1}. \tag{57}$$

From the relations

$$\gamma_{n+1}\mathsf{M}^{\pm}_{jk}\gamma_{n+1} = \mathsf{M}^{\pm}_{jk}, \quad \gamma_{n+1}\mathsf{P}^{\pm}_j\gamma_{n+1} = \mathsf{K}^{\mp}_j,$$
$$\gamma_{n+1}\mathsf{D}^{\pm}\gamma_{n+1} = \mathsf{D}^{\pm}, \tag{58}$$

it is seen that, under the proper (pseudo-)orthogonal subgroup R^p_n, $\varphi^{\pm}_c$ (and $\bar{\varphi}^{\pm}_c$) transforms identically as $\varphi^{\pm}$ ($\bar{\varphi}^{\pm}$); but under the translations $\varphi^{\pm}_c$ transforms just as $\varphi^{\pm}$ under the special conformal transformations, and under the dilatations it transforms just as $\varphi^{\mp}$.

For the inversion of all coordinates ξ^a it follows that

$$\bar{\mathsf{S}} = i\beta, \quad \mathsf{S}^{\pm} = \pm i: \quad \varphi'^{\pm} = \pm i\varphi^{\pm}, \quad \bar{\varphi}'^{\pm} = \mp i\bar{\varphi}^{\pm},$$
$$\varphi_c'^{\pm} = \mp i\varphi_c^{\pm}, \quad \bar{\varphi}_c'^{\pm} = \pm i\bar{\varphi}_c^{\pm}. \tag{59}$$

Since the ξ^a are homogeneous, it must be assumed that the spinors $\varphi^{\pm}$, in addition to the simultaneous multiplication by -1 (as usual), can also be multiplied, one by $+i$ and the other by $-i$.

11. IMPROPER CONFORMAL TRANSFORMATIONS

For the improper conformal transformations C^s and C^t, it follows that

$$\bar{\mathsf{S}}' = \begin{pmatrix} 0 & \mathsf{S}'^{+} \\ \mathsf{S}'^{-} & 0 \end{pmatrix}, \quad \gamma_{n+1}\mathsf{B}\tilde{\mathsf{S}}'^{\pm} = -\bar{\eta}\mathsf{S}'^{-1\mp}\gamma_{n+1}\mathsf{B},$$
$$\mathsf{S}'^{\pm}\gamma_{n+1}\mathsf{C} = -\gamma_{n+1}\mathsf{C}\dot{\mathsf{S}}'^{\mp}, \quad \mathsf{S}'^{\pm *}\mathsf{F} = \bar{\eta}\mathsf{F}\mathsf{S}'^{-1\pm}, \tag{60}$$
$$A^c{}_aA^d{}_b\mathsf{M}^{\pm}_{cd} = \mathsf{S}'^{\pm}\mathsf{M}^{\mp}_{ab}\mathsf{S}'^{-1\pm},$$

and we obtain the following transformation laws for the 2^r-spinors:

$$\varphi'^{\pm} = S'^{\pm}\varphi^{\mp}, \quad \bar{\varphi}'^{\pm} = \bar{\eta}\bar{\varphi}^{\mp}S'^{-1\pm},$$
$$\varphi_c'^{\pm} = -\gamma_{n+1}S'^{\mp}\gamma_{n+1}\varphi_c^{\mp}, \quad \bar{\varphi}_c'^{\pm} = -\bar{\eta}\bar{\varphi}_c^{\mp}\gamma_{n+1}S'^{-1\mp}\gamma_{n+1} \tag{61}$$

(the spinors $\varphi^{\pm}$ turn into $\varphi^{\mp}$).

For the inversions I_a in ξ^a (pseudo-orthonormalized) (I_{n+1} and I_{n+2} are inversions with respect to the hyperspheres $g_{jk}x^jx^k \mp R^2 = 0$), one can choose

$$I_{a'}: \mathsf{S}'^{\pm} = \gamma_{a'} \quad \text{or} \quad \pm i\gamma_{a'}: \varphi'^{\pm} = \gamma_{a'}\varphi^{\mp} \quad \text{or} \quad \pm i\gamma_{a'}\varphi^{\mp},$$
$$a' = 1, \cdots, n, n+1, \tag{62}$$
$$I_{n+2}: \mathsf{S}'^{\pm} = \pm 1 \quad \text{or} \quad i: \varphi'^{\pm} = \pm\varphi^{\pm} \quad \text{or} \quad i\varphi^{\pm}.$$

The inversion I_{n+2} with respect to the "timelike hypersphere" $x^2 + R^2 = 0$ is remarkable; it transforms $\varphi^{\pm}$, $\bar{\varphi}^{\pm}$, $\varphi^{\pm}_c$, and $\bar{\varphi}^{\pm}_c$ into $\varphi^{\mp}$, $\bar{\varphi}^{\mp}$, etc. (except for a factor ± 1 or i).

For the total spacelike (P) and timelike (T) inversions we have, therefore, the equivalence

$$PT = \prod_j I_j: \mathsf{S}^{\pm} = i\gamma_{n+1} \sim I_{n+1}I_{n+2}: \mathsf{S}^{\pm} = \pm\gamma_{n+1}. \tag{63}$$

The conformal conjugation is equivalent (except for a factor i) to the relativistic one followed by the inversion I_{n+1}, with respect to the "spacelike hypersphere" $x^2 - R^2 = 0$.

12. CONFORMAL BILINEAR COVARIANTS

For the conformal "semi"-covariants (or "reduced" covariants)

$$p^{\pm}_{ab} \overset{\text{DEF}}{=} {}^2\bar{\varphi}^{\pm}\mathsf{i}^{\pm}_{ab}\,{}^1\varphi^{\pm}, \quad p^{\pm}_{0} \overset{\text{DEF}}{=} {}^2\bar{\varphi}^{\pm}\,{}^1\varphi^{\pm},$$
$$p^{\pm}_{a'b'} = -\tfrac{1}{2}i\,{}^2\bar{\varphi}^{\mp}\gamma_{[a'}\gamma_{b']}\,{}^1\varphi^{\pm}, \quad p^{\pm}_{a',n+2} = \pm i\,{}^2\bar{\varphi}^{\mp}\gamma_{a'}\,{}^1\varphi^{\pm},$$
$$a', b' = 1, \cdots, n, n+1, \tag{64}$$

$$q^{\pm}_{ab} \overset{\text{DEF}}{=} {}^2\bar{\psi}^{\pm}_{c}\mathsf{i}^{\pm}_{ab}\,{}^1\varphi^{\pm} = \pm\,{}^2\bar{\varphi}^{\mp}_{c}\gamma_{n+1}\mathsf{i}^{\pm}_{ab}\,{}^1\varphi^{\pm},$$
$$q^{\pm}_{0} \overset{\text{DEF}}{=} {}^2\bar{\psi}^{\pm}_{c}\,{}^1\varphi^{\pm} = \pm\,{}^2\bar{\varphi}^{\mp}_{c}\gamma_{n+1}\,{}^1\varphi^{\pm} \tag{65}$$

the following transformation laws are valid:

$$\begin{aligned}
&C^{\mathrm{p}}\colon\ p'^{\pm}_{0} = p^{\pm}_{0}, && p^{\pm}_{ab} = A^{c}{}_{a}A^{d}{}_{b}p'^{\pm}_{cd},\\
&C^{\mathrm{st}}\colon\ p'^{\pm}_{0} = -p^{\pm}_{0}, && p^{\pm}_{ab} = -A^{c}{}_{a}A^{d}{}_{b}p'^{\pm}_{cd},\\
&C^{\mathrm{s}}\colon\ p'^{\pm}_{0} = -p^{\mp}_{0}, && p^{\pm}_{ab} = -A^{c}{}_{a}A^{d}{}_{b}p'^{\mp}_{cd},\\
&C^{\mathrm{t}}\colon\ p'^{\pm}_{0} = p^{\mp}_{0}, && p^{\pm}_{ab} = A^{c}{}_{a}A^{d}{}_{b}p'^{\mp}_{cd};
\end{aligned} \tag{66}$$

there are identical laws for q_0 and q_{ab} (de Broglie-type tensors).

13. CONFORMAL TENSOR IDENTITIES

In the Dirac theory of electron, when ${}^1\varphi \equiv {}^2\varphi$, the conformal covariant p_{ab} unifies the relativistic tensors (densities): magnetoelectric moment $M_{jk} = p_{jk}$, "spin current" $S_j = p_{j5}$, electric current $C_j = p_{j6}$, and pseudoscalar $I_0 = p_{56}$ (except for the dimensional factors).

By contracting the matrix identities (7) with spinors, one gets conformal identities for spinors and bilinear covariants; for example,[11] we have

$$p_{fa}p^{f}{}_{b} = p_0^2\delta_{ab}, \quad p_0^2 = \tfrac{1}{6}p_{cd}p^{cd}, \tag{67a}$$
$$p_0 p_{ab} = \pm\tfrac{1}{8}\epsilon_{abcdef}p^{cd}p^{ef}, \quad a, b, \cdots = 1, \cdots, 6 \tag{67b}$$

(these identities are equivalent), and similar identities for q_{ab} and q_0. The "mixed" identities for Dirac–de Broglie-type covariants are

$$\tfrac{1}{2}p_{f[a}q^{f}{}_{b]} = p_0 q_0 \delta_{ab}, \tag{68a}$$
$$p_0 q_{ab} + p_{ab}q_0 = \pm\tfrac{1}{4}\epsilon_{abcdef}p^{cd}p^{ef}, \tag{68b}$$
$$p_0 q_{ab} - p_{ab}q_0 = \tfrac{1}{2}ip_{f[a}q^{f}{}_{b]}. \tag{68c}$$

In the quantum theory the operators p_{ab} and p_0 defined as normal products[12] verify, e.g., the remarkable (nonsingular) identity

$$\mathsf{p}_{fa'}\mathsf{p}^{f}{}_{b'} = \mathsf{p}_{fb'}\mathsf{p}^{f}{}_{a'} \equiv \mathsf{G}_{a'b'}, \quad a', b' = 1, \cdots, 5. \tag{69}$$

The identity (67a) gives a conformal Euclidean metric as a *quadri*linear form of spinors (or bilinear form in mesonic field p_{ab}) in the spirit of de Broglie's "méthode de fusion[13]" for gravitons:

$$G_{jk} \equiv p_0^2 g_{jk} = M_{lj}M^{l}{}_{k} + S_j S_k - C_j C_k$$
$$(G_{j5} \equiv M_{lj}S^{l} + C_j I_0). \tag{70}$$

These results give new arguments for a further investigation of the conformal properties of the matter.

[1] E. Cunningham, Proc. Math. Soc. (London) **8**, 77 (1910); H. Bateman, *ibid.* 223, 469 (1910). W. Thomson (Lord Kelvin) applied in 1845 inversions to problems of electrostatics in his method of electric images.

[2] Since the conformal transformations conserve the angle between curves in space-time, the conformal group could have an important role in analysis of angular relations in collision processes; see, e.g., A. Esteve and P. G. Sona, Nuovo Cimento **32**, 473 (1964). For other physical arguments see, e.g., R. L. Ingraham, Nuovo Cimento **9**, 886 (1952); **12**, 825 (1954); H. A. Kastrup, Ann. Physik **9**, 388 (1962).

[3] A. P. Hristev, *Travaux de la conference internationale de géometrie et topologie, Juin, 1958*, (Roumanian Academy), p. 203; Comun. Acad. Rep. Populare Române **11**, 755, 761 (1961); Bull. Math. (Bucharest) **4**, 23 (1961). An analogous unification of the 15 Dirac matrices was proposed by H. A. Buchdahl, Proc. Roy. Soc. (London) **A303**, 355 (1968).

[4] W. Pauli, Ann. Inst. Henri Poincaré **6**, 109 (1936); W. Kofink, Ann. Physik **38**, 421 (1940).

[5] See, e.g., J. A. Schouten, Indagationes Math. **11**, 597, 687, 938 (1949); R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935); P. K. Rashewsky, Usp. Mat. Nauk **10**, 3 (1955).

[6] Our normalization differs from that used in Refs. 5.

[7] See, e.g., H. A. Kastrup, Ann. Physik **9**, 388 (1962).

[8] W. Pauli, Nuovo Cimento **6**, 204 (1957); F. Gürsey, *ibid.* **7**, 411 (1958); W. Heisenberg and W. Pauli, "On the Isospin Group in the Theory of the Elementary Particles," Preprint, 1958.

[9] See, e.g., F. Klein, *Vorlesungen über höhere Geometrie* (Berlin, 1926), 3rd ed.; A. P. Hristev, Analele Univ. Bucharest (Math.) **13**, 167 (1964) and Rev. Roumaine Math. Pure Appl. (1970) (to be published).

[10] See, e.g., F. Gürsey, in *Relativity, Groups and Topology*, Lectures at the Summer Session for Theoretical Physics, University of Grenoble, Les Houches, 1963, C. De Witt and B. S. De Witt, Eds. (Gordon & Breach, New York, 1964); N. N. Bogoliubov, in *High Energy Physics and the Theory of Elementary Particles* (in Russian), Lectures at the International School on Theoretical Physics, Yalta, 1956, (Kiev, 1967).

[11] A. P. Hristev, Ref. 3; A. Popovici, A. P. Hristev, and J. Popovici, Compt. Rend. **257**, 3120 (1963). Our identities (67) unify conformally a series of identities established by W. Kofink, Ann. Physik **38**, 421, 436 (1940). The most general case of *four arbitrary* spinor fields is treated by A. P. Hristev, Rev. Roumaine Math. Pure Appl. **9**, 619 (1964). For differential conformal identities, see A. P. Hristev, Analele Univ. Bucharest (Math.) **14**, 51 (1965).

[12] A. P. Hristev, Rev. Roumaine Math. Pure Appl. **10**, 297 (1965).

[13] L. de Broglie, *Théorie générale des particules à spin (méthode de fusion)* (Gauthier-Villars, Paris, 1954), 2nd ed.; the most general metric tensor, quadrilinear in four arbitrary spinor fields, is given in A. P. Hristev, Rev. Roumaine Math. Pure Appl. **9**, 619 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Photon Spin Formalism

KENNETH J. EPSTEIN
2025 W. Granville Avenue, Chicago, Illinois 60645

(Received 23 October 1969; Revised Manuscript Received 17 April 1970)

A new approach to quantum electrodynamics is considered, in which photon spin terms are incorporated phenomenologically into the equations of classical electrodynamics in such a way as to yield quantized equations in the Heisenberg picture, for which relativistic requirements are satisfied. Photon momentum and energy analyses in the new formalism are then compared to the corresponding analyses in conventional field-theoretic formalisms, and questions regarding mathematical and physical consistency are discussed. Different notations are used to distinguish between physical and Hilbert space vectors, noting that the Hilbert space of the vector wavefunction has a 3-dimensional subspace resembling physical space, mathematically, in a way that allows interesting results, which can be described by two formalisms when auxiliary notation is necessary to analyze relationships involving both kinds of vectors.

1. INTRODUCTION

The subject of photon spin, *per se*, receives relatively little attention in discussions of electromagnetic field angular momentum,[1] and this may be due to the fact that quantum field theory attributes only a limited physical significance to the concept of intrinsic angular momentum for electromagnetic quanta, when they are treated as massless vector bosons.[2]

Experiment also yields no compelling evidence to suggest that photon spin should be emphasized more explicitly, in theory, than is usually the case when it is demonstrated to be derivable from its relative obscurity in the framework of quantum electrodynamics.

It is desired here, however, to introduce it in a way which is entirely direct and phenomenological, in the hope that the circumspection afforded by this different approach may serve to enhance the insight and perspective already made possible by the existing relativistic theories of neutral radiation fields.

2. GENERALIZATION OF CLASSICAL ELECTRODYNAMICS

Relativistic 4-tensors used here are expressed in a contravariant form, such that, for two 4-vectors a and b, the dot product (i.e., the scalar product) is $a \cdot g \cdot b$, where g is the 4×4 metric matrix. This formalism obviates the need for many subscripts and superscripts, and keeps the Riemannian geometrical aspects of the problem clear, because $a \cdot g \cdot b$ is a straightforward generalization of the 4-dimensional Euclidean dot product, to which it reduces if g is replaced by the 4×4 identity matrix. The 4-dyad ab is then a second-rank 4-tensor, with trace $a \cdot g \cdot b$, expressed in terms of the Riemannian dot product. $a = (a^\mu) = (a^0, \mathbf{a})$, using boldface letters for 3-space vectors, with $\mu = 0, 1, 2, 3$.

A second-rank 4-tensor

$$F = \|F^{\mu\nu}\| = \begin{pmatrix} f & \mathbf{f} \\ \mathbf{F} & \mathsf{F} \end{pmatrix},$$

using a 2×2 matrix array, in which $f = F^{00}$, $\mathbf{f}$ is a 3-space row vector, $\mathbf{F}$ is a 3-space column vector, and F is a 3×3 matrix, expressed in 3-space dyadic form. Superscripts μ and ν are standard relativistic notation for contravariant tensors, although contravariance, as opposed to covariance, need not be emphasized explicitly in this context.

In the remainder of this problem, g is assumed to be the Lorentz metric of special relativity, because gravitational effects are treated as negligible, at least to the extent of allowing g to be locally diagonalizable to the Lorentzian form, for all phenomena under consideration, so that general relativity, *per se*, is not part of the subject of this discussion.

For the electromagnetic field,

$$F = \partial A - (\partial A)' = \left(\begin{array}{c|c} 0 & -\mathbf{E} \\ \hline \mathbf{E} & \mathsf{I} \times \mathbf{B} \end{array}\right), \tag{1}$$

using contravariant 4-gradient operator ∂, 4-potential $A = (\varphi, \mathbf{A})$, electric field strength vector $\mathbf{E}$, magnetic flux density vector $\mathbf{B}$, and 3-space identity dyadic I, with $(\partial\mathbf{A})'$ defined as the transpose of ∂A and $\mathsf{I} \times \mathbf{B}$ as the cross product of I and $\mathbf{B}$, we see that the dynamical equations are

$$\partial \cdot g \cdot F = \nabla_4 \cdot F = 4\pi c^{-1} J, \tag{2}$$

where 4-current $J = (c\rho, \mathbf{J})$ and covariant 4-gradient $\nabla_4 = (\partial_0, \nabla)$ in terms of 3-space gradient vector ∇, electric current density $\mathbf{J}$, electric charge density ρ, scalar potential φ, and vector potential $\mathbf{A}$. c is the speed of light in a vacuum.

Using the field relations $\mathbf{E} = -\partial_0\mathbf{A} - \nabla\varphi$ and $\mathbf{B} = \nabla \times \mathbf{A}$, with 4th rank 4-index ϵ [whose components are $+1$ when their indices are an even permutation of (0, 1, 2, 3) and are -1 for odd permutations, and zero if any indices are repeated], we have the second-rank 4-tensor

$$G = \partial \cdot g \cdot \epsilon \cdot g \cdot A = \left(\begin{array}{c|c} 0 & -\mathbf{B} \\ \hline \mathbf{B} & -\mathsf{I} \times \mathbf{E} \end{array}\right), \tag{3}$$

the dual of F, obtained from it by a duality transformation.[3]

The trace of the symmetric tensor $F \cdot g \cdot F$ is

$$\mathrm{Tr}\,(F \cdot g \cdot F) = -\mathrm{Tr}\,(G \cdot g \cdot G) = 2(E^2 - B^2), \tag{4}$$

the well-known relationship yielding the classical electromagnetic Lagrangian density $(8\pi)^{-1}(E^2 - B^2)$. The classical electromagnetic 4-tensor of energy–momentum[2] may be obtained as the traceless symmetric tensor

$$(8\pi)^{-1}(F \cdot g \cdot F + G \cdot g \cdot G) = (4\pi)^{-1}\left(\begin{array}{c|c} \tfrac{1}{2}(E^2 + B^2) & \mathbf{E} \times \mathbf{B} \\ \hline \mathbf{E} \times \mathbf{B} & \tfrac{1}{2}(E^2 + B^2)\mathsf{I} - \mathbf{EE} - \mathbf{BB} \end{array}\right). \tag{5}$$

Assume that the correct energy–momentum tensor Γ is

$$\Gamma = (8\pi)^{-1}(F \cdot g \cdot F + G \cdot g \cdot G) + \tfrac{1}{2}cQ = \left(\begin{array}{c|c} U & c\boldsymbol{\Gamma} \\ \hline c^{-1}\boldsymbol{\mathfrak{E}} & \mathit{\Gamma} \end{array}\right), \tag{6}$$

where Q is a quantum term for the quantized electromagnetic field, with energy density U, momentum density $\boldsymbol{\Gamma}$, energy current density vector $\boldsymbol{\mathfrak{E}}$, and momentum current density dyadic $\mathit{\Gamma}$.

We assume that the particle component of the electrodynamic problem is a Dirac electron field, with spinor wavefunction ψ, having Hermitian conjugate $\psi^\dagger$, so that $J = e\psi^\dagger v\psi$ for electron charge e and $v = (c, \mathbf{v})$; then, using velocity operator $\mathbf{v} = c\boldsymbol{\alpha}$ with $\boldsymbol{\alpha} = \alpha\boldsymbol{\sigma}$ (in terms of Pauli vector $\boldsymbol{\sigma}$, anticommuting α and β of the Dirac theory) and introducing particle 4-momentum $P = (P_0, \mathbf{P}) = p - ec^{-1}A$, where canonical 4-momentum $p = i\hbar\partial$ [so that $cP_0 = i\hbar(\partial/\partial t) - e\varphi$ and $\mathbf{P} = -i\hbar\nabla - ec^{-1}\mathbf{A}$, using $i^2 = -1$ and $2\pi\hbar =$ Planck's constant], we may express the 4-tensor of energy–momentum for the particle field as

$$\mathfrak{T} = \tfrac{1}{2}\psi^\dagger(vP\psi) + \tfrac{1}{2}(vP\psi)^\dagger\psi = \left(\begin{array}{c|c} \mathfrak{U} & c\boldsymbol{\mathfrak{T}} \\ \hline c^{-1}\boldsymbol{\Pi} & \overset{\approx}{\mathfrak{T}} \end{array}\right), \tag{7}$$

where the energy density is

$$\mathfrak{U} = \tfrac{1}{2}c\psi^\dagger P_0\psi + \tfrac{1}{2}c(P_0\psi)^\dagger\psi \tag{8a}$$

$$= \tfrac{1}{2}\psi^\dagger(\mathbf{v} \cdot \mathbf{P} + \beta mc^2)\psi + \tfrac{1}{2}[(\mathbf{v} \cdot \mathbf{P} + \beta mc^2)\psi]^\dagger\psi, \tag{8b}$$

using the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = (\mathbf{v} \cdot \mathbf{P} + \beta mc^2 + e\varphi)\psi, \tag{9}$$

and where $\mathrm{Tr}\,(\mathfrak{T}) = \psi^\dagger\beta\psi mc^2$, m being the rest mass of the electron. The momentum density is

$$\boldsymbol{\mathfrak{T}} = \tfrac{1}{2}\psi^\dagger\mathbf{P}\psi + \tfrac{1}{2}(\mathbf{P}\psi)^\dagger\psi. \tag{10}$$

The energy-current density is

$$\boldsymbol{\Pi} = \tfrac{1}{2}c\psi^\dagger\mathbf{v}P_0\psi + \tfrac{1}{2}c(\mathbf{v}P_0\psi)^\dagger\psi \tag{11a}$$

$$= c^2\boldsymbol{\mathfrak{T}} + c^2\nabla x(\tfrac{1}{2}\psi^\dagger\hbar\boldsymbol{\sigma}\psi), \tag{11b}$$

using Eq. (9), and the properties of $\boldsymbol{\sigma}$, to derive Eq. (11b). $\frac{1}{2}\psi^\dagger\hbar\boldsymbol{\sigma}\psi$, occurring in Eq. (11b), is the electron-spin angular momentum density. The momentum current density dyadic is

$$\overset{\approx}{\mathfrak{T}} = \tfrac{1}{2}\psi^\dagger(\mathbf{v}\mathbf{P}\psi) + \tfrac{1}{2}(\mathbf{v}\mathbf{P}\psi)^\dagger\psi. \tag{12}$$

$\mathfrak{T}$ and Γ satisfy the dynamical equations

$$\nabla_4 \cdot \mathfrak{T} = -\nabla_4 \cdot \Gamma = -c^{-1}J \cdot g \cdot F \equiv \mathcal{F} = (\mathcal{F}_0, \boldsymbol{\mathcal{F}}) = (c^{-1}\mathbf{J} \cdot \mathbf{E}, \rho\mathbf{E} + c^{-1}\mathbf{J} \times \mathbf{B}), \tag{13}$$

yielding the energy and momentum equations for the electromagnetic and particle fields in terms of the 4-vector $\mathcal{F}$, for which $c\mathcal{F}_0 = \mathbf{J} \cdot \mathbf{E}$ is the power density transferred by interaction of the electric field with the particles, and the 3-space component $\boldsymbol{\mathcal{F}}$ is the Lorentz force density.

Equations (5), (6), and (13) indicate that $\nabla_4 \cdot Q = 0$, because the classical part of Γ already satisfies Eq. (13). However, if the vector cross product is taken between the 3-space position vector $\mathbf{r}$ and the momentum equations, then Eq. (13) yields the angular-momentum equations,

$$\frac{\partial}{\partial t}(\mathbf{r} \times \boldsymbol{\mathfrak{T}}) + \nabla \cdot (-\overset{\approx}{\mathfrak{T}} \times \mathbf{r}) - \mathfrak{T}_v = -\frac{\partial}{\partial t}(\mathbf{r} \times \boldsymbol{\Gamma}) - \nabla \cdot (-\mathit{\Gamma} \times \mathbf{r}) + \boldsymbol{\Gamma}_v = \mathbf{r} \times \boldsymbol{\mathcal{F}}, \tag{14}$$

where $\mathfrak{T}_v$ is "the vector of $\overset{\approx}{\mathfrak{T}}$" defining the vector of a 3-space dyadic to be that 3-vector which is obtained by replacing all of its dyads by their corresponding cross products (so that the vector of $\mathbf{ab}$ is $\mathbf{a} \times \mathbf{b}$). $\boldsymbol{\Gamma}_v = \frac{1}{2}c\mathbf{Q}_v$, because Eqs. (5) and (6) indicate that

the classical part of Γ is symmetric, and the vector of a symmetric dyadic vanishes.

Equations (9) and (12) and the properties of $\boldsymbol{\sigma}$ yield

$$\mathfrak{T}_v = \tfrac{1}{2}\psi^\dagger \mathbf{v} \times \mathbf{P}\psi + \tfrac{1}{2}(\mathbf{v} \times \mathbf{P}\psi)^\dagger \psi$$
$$= -\frac{\partial}{\partial t}(\tfrac{1}{2}\psi^\dagger \hbar\boldsymbol{\sigma}\psi) - \nabla(\tfrac{1}{2}c\hbar\psi^\dagger\alpha\psi), \qquad (15)$$

indicating that $\mathfrak{T}_v$ adds the spin component to the particle angular momentum, in Eq. (14), assuming that the current density dyadic of $\boldsymbol{\sigma}$ is $\frac{1}{2}\psi^\dagger \mathbf{v}\boldsymbol{\sigma}\psi + \frac{1}{2}(\mathbf{v}\boldsymbol{\sigma}\psi)^\dagger\psi = c\mathsf{I}\psi^\dagger\alpha\psi$, using $\mathbf{v} = c\boldsymbol{\alpha}\boldsymbol{\sigma}$, and $\boldsymbol{\sigma}\boldsymbol{\sigma} + (\boldsymbol{\sigma}\boldsymbol{\sigma})^\dagger = 2\mathsf{I}$.

The $\boldsymbol{\Gamma}_v$ must account for photon spin in Eq. (14). Assuming a photon spin angular momentum density $\mathbf{S}$ with symmetric current density dyadic $c\tilde{\tilde{\mathbf{S}}}$, it is necessary that

$$\tfrac{1}{2}\mathbf{Q}_v = -\partial_0 \mathbf{S} - \nabla \cdot \tilde{\tilde{\mathbf{S}}}. \qquad (16)$$

Equation (16), for the inclusion of the photon spin component in Eq. (14), is satisfied if

$$Q = \left(\begin{array}{c|c} 0 & -\nabla \times \mathbf{S} \\ \hline \nabla \times \mathbf{S} & \mathsf{I} \times (\partial_0 \mathbf{S} + \nabla \mathfrak{C}) - 2\nabla \times \tilde{\tilde{\mathbf{S}}} \end{array}\right), \qquad (17)$$

where $\mathfrak{C}$ is the trace of $\tilde{\tilde{\mathbf{S}}}$. Equations (5), (6), and (17) also give

$$c^{-2}\mathfrak{E} - \boldsymbol{\Gamma} = \nabla \times \mathbf{S}, \qquad (18)$$

consistent with Eq. (11b), so that the corresponding densities, including the total densities (in the total energy–momentum tensor $\mathfrak{T} + \Gamma$), satisfy similar relationships. As the electron spin density is related to the electron spin vector $\frac{1}{2}\boldsymbol{\sigma}$, so it is assumed that the photon spin density is related to the photon spin vector $\mathbf{s}$, satisfying $\mathbf{s} \times \mathbf{s} = i\mathbf{s}$ and $\mathbf{s} \cdot \mathbf{s} \equiv \mathbf{s}^2 = s(s+1) = 2$ for magnitude $s = 1$. $\mathbf{s}$ may be related to the 3-space triadic $\mathsf{I} \times \mathsf{I}$ by using the representation in Cartesian unit vectors $\hat{e}_i$, with $\mathsf{I} = \hat{e}_i\hat{e}_i$, summed over subscripts $i = 1, 2, 3$, and such that

$$\mathbf{s} = s_i\hat{e}_i = \left(\begin{array}{c|c|c} 0 & -i\hat{e}_3 & i\hat{e}_2 \\ \hline i\hat{e}_3 & 0 & -i\hat{e}_1 \\ \hline -i\hat{e}_2 & i\hat{e}_1 & 0 \end{array}\right) = \|\mathbf{s}_{ij}\|, \qquad (19)$$

where $\mathbf{s}_{ij} = -i\hat{e}_i \times \hat{e}_j$, allowing the triadic form

$$\hat{e}_i\mathbf{s}_{ij}\hat{e}_j = -i\hat{e}_i(\hat{e}_i \times \hat{e}_j)\hat{e}_j = -i\mathsf{I} \times \mathsf{I}. \qquad (20)$$

3. PHOTON WAVEFUNCTION

$\mathbf{S}$ and $\mathbf{s}$ may be related by postulating a photon wavefunction

$$\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{pmatrix}$$

in 3-component column matrix form with Hermitian conjugate $\Psi^\dagger = (\Psi_1^\dagger, \Psi_2^\dagger, \Psi_3^\dagger)$ in row matrix form. The Hilbert space problem may also be treated on a Cartesian vector basis, as in Eqs. (19) and (20), with $\boldsymbol{\Psi} = \Psi_i\hat{e}_i$ in 3-space vector notation and $\boldsymbol{\Psi}^\dagger = \boldsymbol{\Psi}_i^\dagger\hat{e}_i$, assuming real $\hat{e}_i$, so that $\hat{e}_i^\dagger = \hat{e}_i$. With this formalism,

$$\mathbf{S} = \hbar\Psi^\dagger \mathbf{s}\Psi = -i\hbar\boldsymbol{\Psi}^\dagger \times \boldsymbol{\Psi}, \qquad (21)$$

thus associating the 3-space vector $\boldsymbol{\Psi}^\dagger \times \boldsymbol{\Psi}$ with the intrinsic spin angular momentum of the electromagnetic field and thus giving another example of the occurrence of the cross product in problems involving angular momentum—although, of course, it cannot be inferred *a priori* that this vector formalism should play a role in relation to quantum spin.[4]

It is also assumed that there is a photon velocity operator $\mathbf{V}$ such that

$$c\tilde{\tilde{\mathbf{S}}} = \tfrac{1}{2}\Psi^\dagger \mathbf{V}\mathbf{s}\Psi + \tfrac{1}{2}(\mathbf{V}\mathbf{s}\Psi)^\dagger\Psi = c\tilde{\tilde{\mathbf{S}}}', \qquad (22)$$

assuming that $\tilde{\tilde{\mathbf{S}}}$ is a symmetric dyadic equal to its transpose $\tilde{\tilde{\mathbf{S}}}'$.

For photons in free space (not interacting with any other fields), a simple Schrödinger equation may be sought of the form

$$H\Psi = i\hbar\frac{\partial\Psi}{\partial t}, \qquad (23a)$$

equivalent to

$$\mathsf{H} \cdot \boldsymbol{\Psi} = i\hbar\frac{\partial\boldsymbol{\Psi}}{\partial t}, \qquad (23b)$$

assuming Hamiltonian operator H, with dyadic form H. For this case (free photons), H should be a generalization of the classical relativistic Hamiltonian cp, for particles of zero mass, and momentum vector $\mathbf{p}$ (with magnitude p). Noting that the photon spin vector $\mathbf{s}$ satisfies the identities

$$(\mathbf{ss})^\dagger = (\mathbf{ss})' \quad \text{and} \quad (\mathbf{ss})' - (\mathbf{ss}) = i\mathsf{I} \times \mathbf{s} = i\mathbf{s} \times \mathsf{I},$$

and assuming the simplification (for the free field case) of letting $\nabla \cdot \boldsymbol{\Psi} = 0$, the remaining mathematical requirements are met by

$$H = c\mathbf{s} \cdot \mathbf{p} \quad \text{or} \quad \mathsf{H} = ic\mathsf{I} \times \mathbf{p} = c\hbar\mathsf{I} \times \nabla,$$

which yield $c\nabla \times \boldsymbol{\Psi} = i(\partial\boldsymbol{\Psi}/\partial t)$, the same equation satisfied by the complex vector $\mathbf{X} \equiv \mathbf{E} + i\mathbf{B}$ when $J = 0$.

In this formalism $\mathbf{V} = c\mathbf{s}$, and Eq. (22) yields

$$\tilde{\tilde{\mathbf{S}}} = \tfrac{1}{2}\Psi^\dagger[\mathbf{ss} + (\mathbf{ss})']\Psi = \mathsf{I}\boldsymbol{\Psi}^\dagger \cdot \boldsymbol{\Psi}$$
$$- \tfrac{1}{2}\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} - \tfrac{1}{2}(\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi})', \qquad (24)$$

using $(\Psi^\dagger\Psi)'$ rather than $\Psi\Psi^\dagger$ because the components of Ψ are operators in the fully quantized field theory, and Eq. (24) is written in such a way as to preserve "normal ordering."

$\mathbf{V}$ yields eigenvalues $\pm c$ (corresponding to transverse modes) and eigenvalue 0 (corresponding to longituidnal modes),[5] while the field (assumed solenoidal) is restricted to transverse modes, for which the total energy equals the total volume integrals

$$\int d^3x\Psi^\dagger H\Psi = (8\pi)^{-1}\int d^3x\mathbf{X}^\dagger\cdot\mathbf{X} \qquad (25)$$

taken over the entire volume of the radiation. Equation (25),together with the fact that $\mathbf{X}$ satisfies Eq. (23), may be interpreted by hypothesizing a Hermitian operator $\mathcal{O}$, such that $\mathcal{O}^2 = 8\pi H$ and $X = \mathcal{O}\Psi$, equivalent to $\mathbf{X} = \tilde{\tilde{\mathcal{O}}}\cdot\Psi$, assuming that X is the column matrix representation of $\mathbf{X}$ and $\mathcal{O}$ has dyadic form $\tilde{\tilde{\mathcal{O}}}$.

4. PHOTON MOMENTUM

For an isolated electrodynamic system (interactions with other systems or fields assumed negligible), the total physical momentum is $\int d^3x(\mathcal{G} + \boldsymbol{\Gamma})$, integrated over the entire volume of the system and summed over the electron and electromagnetic components. The total canonical momentum of the system is

$$\int d^3x(\psi^\dagger\mathbf{p}\psi + \Psi^\dagger\mathbf{p}\Psi),$$

using canonical momentum operator $\mathbf{p} = -i\hbar\nabla$. For an isolated system, it might be supposed, axiomatically, that the total physical momentum is equal to the total canonical momentum, an assumption which yields

$$\int d^3x\boldsymbol{\Gamma} = \int d^3x(\Psi^\dagger\mathbf{p}\Psi + c^{-1}\rho\mathbf{A}), \qquad (26)$$

which, noting that $\boldsymbol{\Gamma} = (4\pi c)^{-1}\mathbf{E}\times\mathbf{B} - \frac{1}{2}\nabla\times\mathbf{S}$, gives a relationship between the electromagnetic wavefunction and field vectors, and can be analyzed further by letting $\Psi = \Psi_0 + \Psi_1$ and $\mathbf{X} = \mathbf{X}_0 + \mathbf{X}_1$, with $\nabla\cdot\Psi_1 = 0 = \nabla\times\Psi_0$, $\mathbf{X}_1 = \nabla\times\mathbf{A}_1$, and $\mathbf{X}_0 = -\nabla\varphi_0$. It is possible, in general, to let $\nabla\cdot\mathbf{A}_1 = 0$, $\varphi_0^\dagger = \varphi_0$, and $\Psi_0^\dagger = \Psi_0$, so that, introducing Ψ_0 as the column matrix form of Ψ_0, it follows that

$$\int d^3x\Psi_0^\dagger\mathbf{p}\Psi_0 = 0,$$

indicating the absence of momentum in longitudinal photon modes.

In Coulomb gauge ($\nabla\cdot\mathbf{A} = 0$), the electromagnetic scalar potential φ may be identified with φ_0, and $\mathbf{X}_1 = i\nabla\times\mathbf{A} - c^{-1}(\partial\mathbf{A}/\partial t)$ with $2i\mathbf{A} = \mathbf{A}_1 - \mathbf{A}_1^\dagger$.

Of the various possible approaches to Eq. (26), the axiomatic approach stated here seems to be the simplest, and the equation simplifies further if

$$\int d^3x(\nabla\times\mathbf{S}) = 0.$$

The axiom leading to Eq. (26) may be generalized to the statement

$$\int d^3x(\psi^\dagger p\psi + \Psi^\dagger p\Psi) = \int d^3x(c^{-1}\mathcal{U} + c^{-1}U, \mathcal{G} + \boldsymbol{\Gamma}) \qquad (27)$$

using canonical 4-momentum operator $p = i\hbar\partial$, so that the 3-space momentum relationship is the same, and the energy relationship follows from the Schrödinger equation with the assumption that the total Hamiltonian is equal to the total energy for an isolated system.

Thus, for an isolated physical system, the Schrödinger equation may be generalized and stated axiomatically as the condition that the only allowed states are those for which the eigenvalues of the total canonical 4-momentum operator are equal to the eigenvalues of the total physical 4-momentum (i.e., energy–momentum 4-vector) operator (assuming that the allowed eigenfunctions are the state vectors, or kets, spanning the Hilbert space of the quantum system in the completely quantized treatment of the dynamical problem).

On this basis, the dynamical variables are operators whose time dependence here results from the fact that the mathematical problem is being treated in the Heisenberg representation, which largely obviates the problem of explicit consideration of the state vector in this context, owing to the constancy of the total system ket.[6]

5. PHOTON INTERACTION ENERGY

The simple Hamiltonian of Eq. (23), for the free-field case, does not generalize in any obvious way for nonvanishing J, and mathematical problems related to this (the photon wavefunction and the Hamiltonian) have been discussed elsewhere in other contexts.[7]

The problem is approached here by hypothesizing a phenomenological photon energy operator $\Upsilon = K + \mathcal{U}$, where $K = c\mathbf{s}\cdot\mathbf{p}$ represents photon kinetic energy, so that Υ reduces to H when $\mathcal{U}$ vanishes.

$\mathcal{U}$ must then account for the potential energy of interaction of photons with the charged particles.

Here it is further assumed that additional Hermitian operators may be defined in such a way that $\mathbf{p} = p\hat{p} = \hat{p}p$, where $p \equiv |\mathbf{p}|$ is defined so that Hermitian operator p, acting on an eigenfunction of $\mathbf{p}$, multiplies it by the magnitude (absolute value) of the momentum for that state and $\hat{p}$, acting on this eigenfunction, multiplies it by the unit vector in the direction of the momentum. In this mathematical formalism, then, we may let gradient vector $\nabla = |\nabla|\,\hat{\nabla}$, where $p = \hbar\,|\nabla|$ and $\hat{\nabla} = i\hat{p}$.

Now we define the Hermitian operator $\mathfrak{Z} \equiv \hat{p} \cdot \mathbf{s}$, with eigenvalues ± 1 and 0. Then $K = cp\mathfrak{Z}$, and the remaining mathematical requirements may be met by letting $\mathfrak{V} = (1 - \mathfrak{Z}^2)cp$. For transverse photon modes, $\mathfrak{Z}^2 = 1$ and $\mathfrak{V}$ vanishes; it accounts for the potential energy (associated with longitudinal modes) through the relation

$$\begin{aligned}\int d^3x \Psi^\dagger \mathfrak{V} \Psi &= \int d^3x \Psi_0^\dagger cp \Psi_0 \\ &= (8\pi)^{-1} \int d^3x (\nabla \varphi_0)^2 \\ &= \tfrac{1}{2} \int d^3x \rho \varphi_0, \end{aligned} \tag{28}$$

assuming integration over the entire volume of the system and assuming that Ψ_0 and X_0 are connected by an appropriate relationship, such as the $\mathfrak{O}$ transformation of Eq. (25). The operators defined in this section may be used in the construction of such transformations and in the analysis following Eq. (26). The dyadic representation of $\mathfrak{V}$ may be expressed as $\tilde{\tilde{\mathfrak{V}}} = cp\hat{p}$.

For $\mathfrak{Z} = -1$, the mathematical occurrence of negative eigenvalues of K may appear to give some basis for speculation about the physical meaning of antiphotons; and while photon–antiphoton annihilation, with creation of photons, may be regarded as a meaningless process in which there is not really any interaction, it can be noted that antiphotons, conjugate to photons of a given circular polarization, may be regarded as photons of the opposite polarization.[8]

For establishing the total photon energy relationship

$$\int d^3x \Psi^\dagger \Upsilon \Psi = (8\pi)^{-1} \int d^3x \mathbf{X}^\dagger \cdot \mathbf{X} \tag{29}$$

and seeking an operator Ω such that $\Omega^2 = 8\pi\Upsilon = 8\pi cp(1 + \mathfrak{Z} - \mathfrak{Z}^2)$, negative Υ may result in mathematical problems which can be resolved by using different representations of the square root of -1. Thus, let $i^2 = j^2 = -1$, assuming that $ij = ji$ with $i^\dagger = -i$ and $j^\dagger = j$, so that Ω may be treated as formally Hermitian by substituting j for i wherever necessary for that purpose.

In this treatment, $\mathfrak{Z}$ may be interpreted as the photon helicity operator, and the square root of $\mathfrak{Z}$ may be regarded as a Hermitian operator with eigenvalues 1, 0, and j.

If K is generalized to $\zeta cp\mathfrak{Z}$, where $\zeta = \pm 1$, negative energy can be prevented by forbidding negative $\zeta\mathfrak{Z}$ for physical reasons (unlike the situation for Fermi–Dirac fields).

The mathematical methods introduced here are not unique or exhaustive, and other approaches to the problematical aspects of the discussion should be possible. The problem of negative $\mathfrak{O}^2$ or Ω^2 can be avoided[8] by transforming H and Υ by constant additive factors of indeterminate magnitude, adjustable to prevent occurrence of negative eigenvalues within the range of physical validity or application of the formulas (an example of unitary transformation of eigenfunctions or base vectors in Hilbert space, with a time-dependent transformation yielding the new Hamiltonian, whose eigenvalues are displaced by a constant additive factor, in this case), without altering the physics.

Implicit in this analysis is the importance of the physical theory being invariant under the operations of the Lorentz transformation group, and the 4-tensor formalism has been employed far enough (explicitly in the derivations) to ensure that this principle is not neglected in the results. This (a special case of the principle of relativity) may be extended to general relativity by defining generalized Lorentz transformation $\mathfrak{L}$ such that $\mathfrak{L}^\dagger \mathfrak{G} \mathfrak{L} = \mathfrak{L} \mathfrak{G} \mathfrak{L}^\dagger = \mathfrak{G}$, for general metric operator $\mathfrak{G} = SgS^\dagger$, equivalent to the Lorentz metric g by unitary similarity transformation S, yielding $\mathfrak{L} = S\lambda S^\dagger$, with $\lambda^\dagger g \lambda = \lambda g \lambda^\dagger = g$, indicating the equivalence of $\mathfrak{L}$ to a Lorentz transformation λ of special relativity and illustrating the principle of local Lorentz covariance (for regions in which S may be regarded as constant).

Equation (21) may also be generalized by letting $\Psi = \mathcal{R}\mathfrak{W}$, in terms of unitary transformation $\mathcal{R}$ and transformed wavefunction $\mathfrak{W}$, yielding $\mathbf{S} = -i\hbar \mathfrak{W}^\dagger \cdot (\tilde{\tilde{\mathcal{R}}}^{-1} \times \tilde{\tilde{\mathcal{R}}}) \cdot \mathfrak{W}$, where $\tilde{\tilde{\mathcal{R}}}^{-1}$ is the dyadic form of $\mathcal{R}^{-1} = \mathcal{R}^\dagger$. The spin representation is $\mathcal{R}^\dagger \mathbf{s} \mathcal{R} \Leftrightarrow i^{-1} \tilde{\tilde{\mathcal{R}}}^{-1} \times \tilde{\tilde{\mathcal{R}}}$.

6. CONCLUSIONS

Equations (6) and (17) satisfy the angular momentum Eq. (14), as well as the standard energy–momentum Eq. (13), in a manner which preserves the over-all interpretation of the 4-dimensional energy–momentum tensor of the electromagnetic field,[2] in spite of its alteration from the classical form of Eq. (5).

Thus, the phenomenological formalism here satisfies

the relativistic requirement for a Lorentz covariant theory, although the approach (to the formulation of quantum electrodynamics) differs considerably from the manifestly covariant formulations of quantum field theory.

Physical considerations here tend more than usual to be manifestly consistent with the correspondence principle, by reliance on the formal similarity between the Heisenberg formulation of quantum mechanics and the classical formulation of dynamics—although the main emphasis, of course, is on the differences (in particular, the spin terms), which form the crux of the problem under consideration.

The photon wavefunction Ψ, however, is treated like a particle wavefunction, including first- and second-quantization formalisms, and this differs from the usual quantization procedure for the electromagnetic field in that it relies less on classical theory. Relationships between the wavefunction method (useful in this problem) and the field vector method (more useful in general) are expressed in integral form in Eqs. (25)–(27). Mathematical and physical problems confronting the wavefunction formulation have been discussed previously.[7]

APPENDIX: PHOTON SPIN EIGENSTATES

Interpretation of $-i\mathbf{l}\times\mathbf{l}$ as a photon spin operator leads to the interpretation of $i\hat{p}\times\mathbf{l}=\mathbf{l}\times\hat{\nabla}$ as photon helicity operator, with eigenfunctions $\hat{u}(\hat{k},\mathfrak{Z},\Theta)=\hat{e}(\hat{k},\mathfrak{Z})\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)$ satisfying $\hat{\nabla}\times\hat{u}=\mathfrak{Z}\hat{u}$, and $\hat{\nabla}\times\hat{u}^{\dagger}=\mathfrak{Z}\hat{u}^{\dagger}$, through use of the complex unit vector $\hat{e}(\hat{k},\mathfrak{Z})=2^{-\frac{1}{2}}(\hat{e}_1+i\mathfrak{Z}\hat{e}_2)$, where $\hat{e}_1$, $\hat{e}_2$, and $\hat{k}$ form a right-handed set of real orthonormal vectors, such that $\hat{e}_1\times\hat{e}_2=\hat{k}$, and wave vector $\mathbf{k}=c^{-1}\omega\hat{k}$, in terms of angular frequency ω, and $\Theta\equiv\omega t-\mathbf{k}\cdot\mathbf{r}$. $\hat{e}(\hat{k},\mathfrak{Z})$ satisfies $i\hat{k}\times\hat{e}=\mathfrak{Z}\hat{e}$, and is normalized in the sense that $\hat{e}^{\dagger}\cdot\hat{e}=1$, although $\hat{e}\cdot\hat{e}=0$.

The photon spin vector is $-i\hat{u}^{\dagger}\times\hat{u}=-i\hat{e}^{\dagger}\times\hat{e}=\mathfrak{Z}\hat{k}$, for which the second-order homogeneous electromagnetic field equations also have the real solution $2^{-\frac{1}{2}}(\hat{u}^{\dagger}+\hat{u})=\hat{e}_1\cos\Theta+\mathfrak{Z}\hat{e}_2\sin\Theta$, representing a circularly polarized mode, with right-handed polarization corresponding to $\mathfrak{Z}=1$ and left-handed polarization corresponding to $\mathfrak{Z}=-1$, thus indicating how the photon spin aligns with the constant angular velocity of the rotating polarization vector. Longitudinal photon solutions ($\mathfrak{Z}=0$) do not exist.

For integration over system volume V, the functions $\mathbf{u}_j=V^{-\frac{1}{2}}\hat{u}_j$, where subscript $j\Leftrightarrow(\mathbf{k},\mathfrak{Z})$, satisfy

$$\int d^3x\mathbf{u}_j^{\dagger}\cdot\mathbf{u}_l=\delta_{jl}$$

and $\int d^3x\mathbf{u}_j^{\dagger}\times\mathbf{u}_l=i\mathfrak{Z}\hat{k}\delta_{jl}$, for Kronecker δ_{jl}.

Introducing annihilation operators a_j, with corresponding creation operators $a_j^{\dagger}$, yielding number operators $(a^{\dagger}a)_j$ for photon mode j, and,using summation convention over subscripts j and l, we have $\Psi=a_j\mathbf{u}_j$, $\mathbf{S}=-i\hbar a_j^{\dagger}a_l\mathbf{u}_j^{\dagger}\times\mathbf{u}_l$, and $\mathbf{A}=c(h/\omega)^{\frac{1}{2}}\times(a_j\mathbf{u}_j+a_j^{\dagger}\mathbf{u}_j^{\dagger})$, for quantized transverse modes (replacing summation by integration, where necessary, as volume $V\to\infty$).

Regarding relativistic requirements on constructs used here, it might be well to note that some fault has been found with the covariance of the standard formulation of quantum electrodynamics,[9] so that the findings of relativistic invariance here or claims of physically possible results must be tempered by these considerations.

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), p. 77; N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959), p. 57; P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, England, 1958), 4th ed. (revised 1967); W. Heitler, *The Quantum Theory of Radiation* (Clarendon, Oxford, England, 1954), 3rd ed., p. 59; A. Messiah, *Quantum Mechanics* (Wiley, New York, 1962), Vol. II, p. 1022; L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed.; J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962), p. 90.

[2] A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* (Interscience, New York, 1965); B. Kurşunoğlu, *Modern Quantum Theory* (Miller Freeman, San Francisco, 1962).

[3] Cf. J. A. Wheeler, Ref. 1, and D. Zwanziger, Phys. Rev. **176**, 1489 (1968). This is mentioned incidentally here, without introducing magnetic monopoles. Cf. R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), and R. Penny, J. Math. Phys. **5**, 1431 (1964).

[4] Cf. L. I. Schiff, Ref. 1, p. 198. $\mathbf{S}=-(\Psi^{\dagger}\cdot\mathbf{s})\Psi=$ column matrix form of $\mathbf{s}$. Formally, $\mathbf{s}=-i\hat{l}'\times\hat{l}$, representing $\mathbf{l}$ as $\hat{l}=(\hat{e}_1,\hat{e}_2,\hat{e}_3)$ and using a combination of Cartesian and matrix notation, with $\hat{l}'$ denoting the transpose of $\hat{l}$. One formalism suffices for many problems, but both are used here because additional notation is required.

[5] H. A. Kramers, *Quantum Mechanics* (Dover, New York, 1964), p. 418; P. I. Richards, *Manual of Mathematical Physics* (Pergamon, New York, 1959), p. 176; W. Heitler, Ref. 1, p. 45; L. I. Schiff, Ref. 1, p. 524; B. Kurşunoğlu, Ref. 2, p. 143.

[6] P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Massachusetts, 1965), p. 407; S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1961), p. 650.

[7] B. Kurşunoğlu, Ref. 2; S. S. Schweber, Ref. 6, p. 117; T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).

[8] Cf. B. Kurşunoğlu, Ref. 2, p. 311; D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B241 (1964); D. Shay and R. H. Good, Jr., *ibid.* **179**, 1410 (1969); T. J. Nelson and R. H. Good, Jr., *ibid.* **179**, 1445 (1969); Y. Frishman and C. Itzykson, *ibid.* **180**, 1556 (1969).

[9] K. Haller and L. F. Landovitz, Phys. Rev. Letters **22**, 245 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Unitary Irreducible Representations of the Groups $SO_0(n, 1)$

FRITZ SCHWARZ
Center for Particle Theory, Department of Physics
The University of Texas, Austin, Texas 78712

(Received 20 July 1970)

By $SO_0(n, 1)$ we denote the identity component of the group $O(n, 1)$, by $so(n, 1)$, the Lie algebra of $SO_0(n, 1)$. We determine all those representations of the Lie algebras $so(n, 1)$ which can be extended to a unitary irreducible representation of the group and give explicit expressions for the generators. The general results are specialized to the cases $n = 2$, 3, 4, and 5.

1. INTRODUCTION

We denote by $O(n, 1)$ the group of linear homogeneous transformations of the real $(n + 1)$-dimensional vector space which leave the quadratic form

$$x_1^2 + x_2^2 + \cdots + x_n^2 - x_{n+1}^2$$

invariant. By $SO_0(n, 1)$ we denote the identity component of this group and by $so(n, 1)$ the Lie algebra of $SO_0(n, 1)$. The group $SO_0(n, 1)$ contains $n(n + 1)/2$ subgroups of the type

$$g_{ij}(t) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos t & \cdots & \sin t & & \\ & & \vdots & & \vdots & & \\ & & -\sin t & \cdots & \cos t & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \qquad 1 \leq i < j \leq n,$$

and n subgroups of the type

$$g_{i,n+1}(t) = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \cosh t & \cdots & \sinh t & \\ & & \vdots & & \vdots & \\ & & \sinh t & \cdots & \cosh t & \end{bmatrix}, \qquad 1 \leq i \leq n.$$

The matrix $g_{ij}(t)$ corresponds to a rotation in the (x_i, x_j) plane, and the matrix $g_{i,n+1}(t)$ to a hyperbolic rotation in the (x_i, x_{n+1}) plane. The basis elements A_{ij} of the Lie algebra $so(n, 1)$ are defined by

$$A_{ij} = \frac{d}{dt} g_{ij}(t)\Big|_{t=0}, \tag{1}$$

and they obey the commutation relations

$$[A_{ij}, A_{kl}] = g_{jk}A_{il} + g_{il}A_{jk} - g_{ik}A_{jl} - g_{jl}A_{ik}, \tag{2}$$

with $g_{ii} = +1$ for $1 \leq i \leq n$, $g_{n+1,n+1} = -1$, and $g_{ij} = 0$ for $i \neq j$. In a unitary irreducible representation (UIR) of $SO_0(n, 1)$, the generators A_{ij} are anti-Hermitian:

$$A_{ij}^+ = -A_{ij}. \tag{3}$$

Throughout this paper we use the same symbol for a generator A_{ij} and the corresponding operator acting in a representation space, instead of the more rigorous notation $\rho(A_{ij})$, which is used, for example, in Ref. 1. By exactly the same arguments as in this reference (pp. 13, 14), it follows that the representations of the Lie algebras $so(n, 1)$ which we determine can be extended to UIR's of the group $SO_0(n, 1)$. The equation on page 14 of Ref. 1 reads in our case

$$\Delta = \sum_{1 \leq i < j \leq n+1} A_{ij}^2 = -S_2 + 2 \sum_{1 \leq i < j \leq n} A_{ij}^2,$$

where

$$S_2 = \sum_{1 \leq i < j \leq n} A_{ij}^2 - \sum_{1 \leq i \leq n} A_{i,n+1}^2$$

is the second-order Casimir operator of the Lie algebra $so(n, 1)$ and Δ is the Nelson operator of $so(n, 1)$.

From relations(2) it is easy to see that a representation of $so(n, 1)$ is completely determined if we know the operators $A_{12}, A_{23}, \cdots, A_{n,n+1}$, because the other operators A_{ij} can be expressed through them by relations(2). The generators A_{ij} with $1 \leq i < j \leq n$ are a basis for the Lie algebra $so(n)$ of the n-dimensional rotation group, the irreducible representations of which have been determined in Ref. 2. Therefore, the problem of determining the UIR's of $so(n, 1)$ reduces to determining the generator $A_{n,n+1}$. It turns out that this can be done in a straightforward way by using the same methods as in Ref. 2. This is essentially a consequence of two facts: We reduce

with respect to the maximal compact subgroup $SO(n)$, so that within a UIR of $SO_0(n, 1)$ a state is completely labeled by discrete indices only. Further, in a UIR of $SO_0(n, 1)$ an irreducible representation of $SO(n)$ occurs either with multiplicity 1 or not at all. This has been shown by Dixmier.[3]

We give in Sec. 2 the results of Ref. 2 in a form that is especially suited for our purpose. In Sec. 3 we determine the operator $A_{n,n+1}$ and give a complete classification of the UIR's of $SO_0(n, 1)$. In Sec. 4 we specialize the general expressions from Sec. 3 to the cases $n = 2$, 3, 4, and 5, and collect the results in Tables III–VI. In the Appendix we compare the cases $n = 2$, 3, and 4 with the work of other authors. Results similar to ours have been given in Refs. 4 and 5.

2. THE IRREDUCIBLE REPRESENTATIONS OF THE GROUPS $SO(n)$

We give in this section the results of Ref. 2 with some slight modifications of the notation due to Hirai[4] and Pang.[6] The generators A_{ij} with $1 \leq i < j \leq n$ form a basis of $so(n)$ and obey the commutation relations

$$[A_{ij}, A_{kl}] = \delta_{jk}A_{il} + \delta_{il}A_{jk} - \delta_{ik}A_{jl} - \delta_{jl}A_{ik}. \quad (4)$$

There are some characteristic differences for $n = 2p$ even or $n = 2p + 1$ odd. In either case an irreducible representation is determined by a set of p numbers m_{ij}, all integer or half-integer at the same time. We denote a vector in a representation space by $|m_{ij}\rangle$, where m_{ij} is an abbreviation for a complete set of labels which determine an irreducible representation and specify each vector within the representation space uniquely. For $n = 2p$ the complete scheme is

$$|m_{ij}\rangle = \begin{pmatrix} m_{2p,1} & m_{2p,2} & \cdots & m_{2p,p-1} & m_{2p,p} \\ m_{2p-1,1} & m_{2p-1,2} & \cdots & m_{2p-1,p-1} & \\ m_{2p-2,1} & m_{2p-2,2} & \cdots & m_{2p-2,p-1} & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ m_{41} & m_{42} & & & \\ m_{31} & & & & \\ m_{21} & & & & \end{pmatrix}, \quad (5)$$

and for $n = 2p + 1$

$$|m_{ij}\rangle = \begin{pmatrix} m_{2p+1,1} & m_{2p+1,2} & \cdots & m_{2p+1,p-1} & m_{2p+1,p} \\ m_{2p,1} & m_{2p,2} & \cdots & m_{2p,p-1} & m_{2p,p} \\ m_{2p-1,1} & m_{2p-1,2} & \cdots & m_{2p-1,p-1} & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ m_{41} & m_{42} & & & \\ m_{31} & & & & \\ m_{21} & & & & \end{pmatrix}. \quad (6)$$

The first lines in (5) and (6) determine an irreducible representation of $so(n)$, and the other labels specify a vector within a representation space. All the m_{ij} are integer or half-integer at the same time. The first index in m_{ij} has always the same numerical value as the dimension of a rotation group an irreducible representation of which it specifies. The indices m_{ij} obey the following conditions:

$$\begin{aligned} -m_{2k+1,1} \leq m_{2k,1} \leq m_{2k+1,1} \leq \cdots \leq m_{2k+1,k-1} \\ \leq m_{2k,k} \leq m_{2k+1,k}, \\ |m_{2k,1}| \leq m_{2k-1,1} \leq m_{2k,2} \leq \cdots \leq m_{2k,k-1} \\ \leq m_{2k-1,k-1} \leq m_{2k,k}. \end{aligned} \quad (7)$$

The index k goes from 1 to $p - 1$ or p for n even or odd, respectively. The generators $A_{i,i+1}$ for $1 \leq i \leq n - 1$ are given by

$$\begin{aligned} A_{2k,2k+1}|l_{ij}\rangle = \sum_{j=1}^{k} A(l_{2k,j})\,|l_{2k,j} + 1\rangle \\ - \sum_{j=1}^{k} A(l_{2k,j} - 1)\,|l_{2k,j} - 1\rangle, \end{aligned} \quad (8)$$

$$\begin{aligned} A_{2k-1,2k}|l_{ij}\rangle = \sum_{j=1}^{k-1} B(l_{2k-1,j})\,|l_{2k-1,j} + 1\rangle \\ - \sum_{j=1}^{k-1} B(l_{2k-1,j} - 1)\,|l_{2k-1,j} - 1\rangle \\ + iC_{2k}\,|l_{ij}\rangle. \end{aligned} \quad (9)$$

The matrix elements are

$$\begin{aligned} A(l_{2k,j}) = \frac{1}{2}\left(\prod_{r=1}^{k-1}[(l_{2k-1,r} - \tfrac{1}{2})^2 - (l_{2k,j} + \tfrac{1}{2})^2]\right)^{\frac{1}{2}} \left(\prod_{r=1}^{k}[(l_{2k+1,r} - \tfrac{1}{2})^2 - (l_{2k,j} + \tfrac{1}{2})^2]\right)^{\frac{1}{2}} \\ \times \left(\prod_{\substack{r=1 \\ r \neq j}}^{k-1}[l_{2k,r}^2 - l_{2k,j}^2][l_{2k,r}^2 - (l_{2k,j} + 1)^2]\right)^{-\frac{1}{2}}, \end{aligned} \quad (10)$$

$$B(l_{2k-1,j}) = \left(\prod_{r=1}^{k-1}(l_{2k-2,r}^2 - l_{2k-1,j}^2)\right)^{\frac{1}{2}}\left(\prod_{r=1}^{k}(l_{2k,r}^2 - l_{2k-1,j}^2)\right)^{\frac{1}{2}}$$
$$\times\left(l_{2k-1,j}^2(4l_{2k-1,j}^2 - 1)\prod_{\substack{r\neq j\\ r=1}}^{k-1}(l_{2k-1,r}^2 - l_{2k-1,j}^2)[(l_{2k-1,r}-1)^2 - l_{2k-1,j}^2]\right)^{-\frac{1}{2}}, \tag{11}$$

$$C_{2k} = \left(\prod_{r=1}^{k-1} l_{2k-2,r}\prod_{r=1}^{k} l_{2k,r}\right)\left(\prod_{r=1}^{k-1} l_{2k-1,r}(l_{2k-1,r}-1)\right)^{-1}. \tag{12}$$

The indices l_{ij} are connected with the m_{ij} through

$$l_{2k+1,i} = m_{2k+1,i} + i, \tag{13}$$
$$l_{2k,i} = m_{2k,i} + i - 1. \tag{14}$$

If the representations of $so(n)$ are already known, the determination of the representations of $so(n+1)$ is reduced to the problem of specifying the action of $A_{n,n+1}$. It can be shown that the commutation relations of $so(n+1)$ are equivalent to those of $so(n)$ and the additional commutation relations

$$[A_{i,i+1}, A_{n,n+1}] = 0, \quad \text{for} \quad 1 \leq i \leq n-2, \tag{15}$$
$$[A_{n-1,n}, [A_{n,n+1}, A_{n-1,n}]] = A_{n,n+1}, \tag{16}$$
$$[A_{n,n+1}, [A_{n,n+1}, A_{n-1,n}]] = -A_{n-1,n}. \tag{17}$$

The result of these commutation relations are again generators (8) or (9) with the matrix elements (10)–(12); however, as a result of the commutation relations alone, the uppermost line in (5) or (6) respectively consists of complex numbers z_{ij}. They are restricted to the labels m_{ij} by the requirement of antihermiticity. In the next section we see how this fact can be exploited to allow us to determine the UIR's of $SO_0(n, 1)$ very easily.

3. THE UIR's OF $SO_0(n, 1)$

In this main section we determine an explicit expression for the generator $A_{n,n+1}$ of the Lie algebra $so(n, 1)$ and give a complete classification of the UIR's. As far as possible, we want to use the results of the last section. Therefore, it is convenient to define a new generator $B_{n,n+1}$, which is connected to $A_{n,n+1}$ by

$$A_{n,n+1} = iB_{n,n+1}. \tag{18}$$

It is easy to see that the generators $A_{i,i+1}$, with $i = 1 \cdots n-1$, and $B_{n,n+1}$ together obey the commutation relations of the Lie algebra $so(n+1)$, e.g., the A_{ij} with $1 \leq i < j \leq n$ obey relations (4) and $B_{n,n+1}$ satisfies relations (15)–(17), with $A_{n,n+1}$ replaced by $B_{n,n+1}$. However, in a UIR of $SO_0(n, 1)$ we have now, instead of (3),

$$B_{n,n+1} = B_{n,n+1}^{+}. \tag{19}$$

From these considerations it follows that we can take from the last section all those results which were derived using only the commutation relations, and then we have to impose as an additional requirement the hermiticity of $B_{n,n+1}$. Clearly, we have to distinguish again the cases n even or odd, and we begin with $n = 2p$ even. According to what we said at the end of the last section, we get, for $B_{n,n+1}$, from the commutation relations, an expression which is given by the right-hand side of (8) for $k = p$. However, in the matrix elements (10) the labels $l_{2p+1,j}$ are not defined by (13), but we now have

$$l_{2p+1,j} = z_{2p+1,j} + j, \quad j = 1 \cdots p, \tag{20}$$

where the $z_{ij} = x_{ij} + iy_{ij}$ are complex numbers. They are strongly restricted by the requirement that the $so(n)$ labels $m_{2p,j}, j = 1 \cdots p$, obey the conditions

$$|m_{2p,1}| \leq m_{2p,2} \leq \cdots \leq m_{2p,p-1} \leq m_{2p,p} \tag{21}$$

and by the hermiticity of $B_{n,n+1}$. Equation (21) means that, for $j = 1 \cdots p-1$, the labels $m_{2p,j}$ obey the condition

$$m_{2p,j}^{\max} = m_{2p,j+1}^{\min}, \quad j = 1 \cdots p-1.$$

For this to be true, we need

$$A(m_{2p,j}^{\max}) = A(m_{2p,j+1}^{\min} - 1) = 0 \tag{22}$$

for $j = 1 \cdots p-1$. Equation (22) gives $p-1$ conditions for the $z_{2p+1,j}$; we choose $j = 1 \cdots p-1$ and get

$$(z_{2p+1,j} + j - \tfrac{1}{2})^2 = (m_{2p,j}^{\max} + j - \tfrac{1}{2})^2$$
$$= (m_{2p,j+1}^{\min} + j - \tfrac{1}{2})^2. \tag{23}$$

From (23) it follows that $z_{2p+1,j}$, with $j = 1 \cdots p-1$, must be integer or half-integer, together with the $so(n)$-labels $m_{2p,j}$, i.e., $z_{2p+1,j} = m_{2p+1,j} = m_{2p,j}^{\max} = m_{2p,j+1}^{\min}$ with the conditions

$$|m_{2p,1}| \leq m_{2p+1,1} \leq m_{2p,2} \leq m_{2p+1,2}$$
$$= \cdots \leq m_{2p+1,p-1} \leq m_{2p,p} \tag{24}$$

and $m_{2p+1,j} = 0, \tfrac{1}{2}, 1, \cdots, j = 1 \cdots p-1$. For the remaining constant $z_{2p+1,p}$ the only requirement is that $B_{n,n+1}$ be Hermitian, i.e., that all the matrix elements $A(l_{2p,j})$ be purely imaginary. For this to be true, the expressions under the square root in (10) have to be real and negative. The condition of reality

restricts $z_{2p+1,p}$ to one of the two possibilities:

$$z_{2p+1,p} + p - \tfrac{1}{2} = iy_{2p+1,p}, \tag{25a}$$

$$z_{2p+1,p} + p - \tfrac{1}{2} = x_{2p+1,p} + p - \tfrac{1}{2}. \tag{25b}$$

The requirement that the square root in (10) has to be negative means

$$(z_{2p+1,p} + p - \tfrac{1}{2})^2 = (m_{2p,j} + j - \tfrac{1}{2})^2 \tag{26}$$

for $j = 1, \cdots, p$. These p inequalities are a consequence of the single inequality (26) with $j = k + 1$, where k is the same as in the equation $m_{2p+1,1} = \cdots = m_{2p+1,k} = 0$, because in this case $A(l_{2p,j}) \equiv 0$ for $1 \leq j \leq k$. The origin of Eq. (26) can be seen as follows: The denominator in the matrix elements $A(l_{2p,j})$ is always positive, so that the whole numerator must be negative. Now it is easy to see that the expression

$$\prod_{r=1}^{p-1} [(l_{2p-1,r} - \tfrac{1}{2})^2 - (l_{2p,j} + \tfrac{1}{2})^2] \times [(l_{2p+1,r} - \tfrac{1}{2})^2 - (l_{2p,j} + \tfrac{1}{2})^2]$$

is always positive, because $m_{2p\pm1,j} \leq m_{2p,j+1} \leq m_{2p\pm1,j+1}$, and consequently the whole expression under the square root of (10) is negative if (26) is fulfilled. We consider now the condition (26) for the two possibilities (25) separately. In the case (25a) the inequality (26) is satisfied for arbitrary real $y_{2p+1,p}$; however, to avoid having the same representation occur more than once, we make the restriction $0 < y_{2p+1,p}$. The equality $y_{2p+1,p} = 0$ is excluded because then $A(l_{2p,1}) = 0$ for $l_{2p,1} = m_{2p,1} = -\frac{1}{2}$ and $A(l_{2p,1} - 1) = 0$ for $m_{2p,1} = +\frac{1}{2}$. This case will be included later. Here we have no further restrictions for the $m_{2p+1,j}$, with $1 \leq j \leq p - 1$, and the $so(n)$ content is given by (24). We call the representations of this series $D(m_{2p+1,1} \cdots m_{2p+1,p-1}, iy_{2p+1,p})$. In the case (25b) different possibilities occur. If $x_{2p+1,p} = m_{2p+1,p}$ is integer or half-integer, together with the $so(n)$-labels, it can happen that the smallest $so(n)$ label $m_{2p,1}$ is further restricted because, for $m_{2p+1,p} + p = \pm m_{2p,1}$, we have $A(m_{2p,1} - 1) = A(m_{2p,1}) = 0$. That means that in this case the restrictions for $m_{2p,1}$ are, instead of (24), $m_{2p+1,p} + p \leq m_{2p,1} \leq m_{2p+1,1}$, with $m_{2p+1,p} + p$ and $m_{2p+1,j}$, $j = 1 \cdots p - 1$, equal to $\frac{1}{2}, 1, \frac{3}{2}, \cdots$. We call the UIR's of this type $D^{\pm}(m_{2p+1,1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$ according to the sign of $m_{2p,1}$.

Let us now assume that $m_{2p+1,j} \neq 0$ and is integer for $1 \leq j \leq p - 1$. Then in (26) we have $j = 1$, and evidently the smallest value of the right-hand side is $\frac{1}{4}$, so that (26) reads

$$(x_{2p+1,p} + p - \tfrac{1}{2})^2 \leq \tfrac{1}{4}. \tag{27}$$

To avoid having the same representation occur more than once, we have the stronger restriction

$$0 \leq x_{2p+1,p} + p - \tfrac{1}{2} \leq \tfrac{1}{2}. \tag{28}$$

If on the right side the "less than" sign is valid, the $so(n)$ content is given by (24), all m_{ij} integers, and we call these representations $D^0(m_{2p+1,1} \cdots m_{2p+1,p-1}; x_{2p+1,p})$. If instead the "equals" sign is valid, the matrix element $A(l_{2p,1}) = A(l_{2p,1} - 1) = 0$ for $m_{2p,1} = 0$ so that the $so(n)$ content is given by

$$m_{2p,1} = 0 < m_{2p+1,1} \leq m_{2p,2}, \tag{29}$$

and the rest is taken from Eq. (24). We call these representations $D^0(m_{2p+1,1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$. There is a last case left. If $m_{2p+1,j} = 0$ for $1 \leq j \leq k$, the inequality (26) is

$$(x_{2p+1,p} + p - \tfrac{1}{2})^2 \leq (k + \tfrac{1}{2})^2. \tag{30}$$

By the same arguments, following Eq. (27) we have in this case

$$0 \leq x_{2p+1,p} + p - \tfrac{1}{2} \leq k + \tfrac{1}{2}. \tag{31}$$

Completely analogous to the case (28), there are two series of representation: for the "less than" sign at the right side of (31) the representations

$$D^k(m_{2p+1,k+1} \cdots m_{2p+1,p-1}; x_{2p+1,p}),$$

with the $so(n)$ content (24) and $1 \leq k \leq p - 1$, and for the "equals" sign the representations

$$D^k(m_{2p+1,k+1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$$

with the $so(n)$ content

$$m_{2p,1} = \cdots = m_{2p,k+1} = 0 < m_{2p+1,k+1} \leq \cdots \leq m_{2p,p}, \tag{32}$$

with $1 \leq k \leq p - 2$.

Now let $n = 2p + 1$ be odd. Then the generator $B_{n,n+1}$ is given by (9) with $k = p + 1$ and the matrix elements (11) and (12). The whole discussion is similar to the case of n even, and so we give only the major steps. The irreducible representations of $so(n)$ are determined by the labels $m_{2p+1,j}$, $1 \leq j \leq p$, with the condition

$$0 \leq m_{2p+1,1} \leq m_{2p+1,2} \leq \cdots \leq m_{2p+1,p}. \tag{33}$$

In the matrix elements $B(l_{2p+1,j})$ and C_{2p+2}, there occur $p + 1$ complex constants $z_{2p+2,j}$ which are connected to the $l_{2p+2,j}$ through

$$l_{2p+2,j} = z_{2p+2,j} + j - 1. \tag{34}$$

The inequalities $m^{\min}_{2p+1,1} \leq m_{2p+1,1}$ and $m^{\max}_{2p+1,j} = m^{\min}_{2p+1,j+1}$, for $1 \leq j \leq p - 1$, give p conditions which restrict the $z_{2p+2,j}$ for $1 \leq j \leq p$ to integral or half-integral values which fulfill

$$|m_{2p+2,1}| \leq m_{2p+1,1} \leq m_{2p+2,2} \leq \cdots \leq m_{2p+2,p} \leq m_{2p+1,p}. \tag{35}$$

The last constant $z_{2p+2,p+1}$ is restricted by the hermiticity of $B_{2p+1,2p+2}$. With the same arguments as those for n even, we get the following two possibilities, which are analogous to (25),

$$z_{2p+2,p+1} + p = iy_{2p+2,p+1}, \tag{36a}$$

$$z_{2p+2,p+1} + p = x_{2p+2,p+1} + p, \tag{36b}$$

and the following inequality which is analogous to (26),

$$(z_{2p+2,p+1} + p)^2 \leq (m_{2p+2,k} + k)^2. \tag{37}$$

Again, k is the highest value of the second index j in $m_{2p+2,j}$ for which $m_{2p+2,j} = 0$. If all $m_{2p+2,j} \neq 0$, then $k = 0$.

Now from the hermiticity of $B_{2p+1,2p+2}$ it follows that for $C_{2p+2} \neq 0$ only case (36a) can occur. The same arguments as for n even give the restriction $0 \leq y_{2p+2,p+1}$. The $m_{2p+2,j}$, $1 \leq j \leq p$, have the range $|m_{2p+1,1}|$, $m_{2p+1,j} = 0, \frac{1}{2}, 1 \cdots$. The $so(n)$ content is given by (35). We denote the representations of this class by $D(m_{2p+2,1} \cdots m_{2p+2,p}; iy_{2p+2,p+1})$. For (36b) to be allowed, it is necessary that $m_{2p+2,1} = 0$. It may happen that some more of the $m_{2p+2,j} = 0$, with $1 \leq j \leq k \leq p$. Then we get from (37)

$$0 < x_{2p+2,p+1} + p \leq k, \tag{38}$$

and from (35) it follows that

$$m_{2p+1,1} = \cdots = m_{2p+1,k-1} = 0.$$

The equality $x_{2p+2,p+1} + p = 0$ is excluded because this case is already contained in the class

$$D(m_{2p+2,1} \cdots m_{2p+2,p}; iy_{2p+2,p+1}).$$

If the "less than" sign at the right side of (38) is valid, we have no further restrictions and denote these representations by $D^k(m_{2p+2,k+1} \cdots m_{2p+2,p}; x_{2p+2,p+1})$ with $1 \leq k \leq p$. If the "equals" sign is valid, we have $x_{2p+2,p+1} + p = k = m_{2p+2,p+1} + p$. We denote these representations by $D^k(m_{2p+2,k+1} \cdots m_{2p+2,p}; m_{2p+2,p+1})$. Now $1 \leq k \leq p-1$ and, in addition, $m_{2p+1,k} = 0$. The results of this paragraph are collected in Tables I and II; the identity representation has been omitted in these tables.

4. SOME SPECIAL CASES

In this section we specialize the results of Sec. 3 to the cases $n = 2, 3, 4$, and 5. We give explicit expressions for the generators $A_{i,i+1}$, $i = 1 \cdots n$; the classification of the UIR's is given in Tables III–VI. In the coordinate systems the coordinates of each dot represent an irreducible representation of $so(n)$. In the Appendix we compare our results with those of other authors. The generators $A_{n,n+1}$ and $B_{n,n+1}$ are always connected through $A_{n,n+1} = iB_{n,n+1}$.

(a) $n = 2$: In a UIR a state is completely labeled by

$$|m_{ij}\rangle = \begin{pmatrix} z_{31} \\ m_{21} \end{pmatrix}. \tag{39}$$

TABLE I. The complete list of UIR's of $SO_0(2p, 1)$.

Representation and conditions for $m_{2p+1,1} \cdots m_{2p+1,p-1}$ and $z_{2p+1,p}$	$SO(2p)$ content
$D(m_{2p+1,1} \cdots m_{2p+1,p-1}; iy_{2p+1,p})$ $m_{2p+1,j} = 0, \frac{1}{2}, 1, \cdots$ for $1 \leq j \leq p-1$ $z_{2p+1,p} + p - \frac{1}{2} = iy_{2p+1,p}$; $0 < y_{2p+1,p}$	$\lvert m_{2p,1}\rvert \leq m_{2p+1,1} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$
$D^0(m_{2p+1,1} \cdots m_{2p+1,p-1}; x_{2p+1,p})$ $m_{2p+1,j} = 1, 2, 3, \cdots$ for $1 \leq j \leq p-1$ $z_{2p+1,p} = x_{2p+1,p}$; $0 \leq x_{2p+1,p} + p - \frac{1}{2} < \frac{1}{2}$	$\lvert m_{2p,1}\rvert \leq m_{2p+1,1} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$
$D^0(m_{2p+1,1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$ $m_{2p+1,j} = 1, 2, 3, \cdots$ for $1 \leq j \leq p-1$ $z_{2p+1,p} = m_{2p+1,p}$; $m_{2p+1,p} + p = 1$	$m_{2p,1} = 0$ $m_{2p+1,1} \leq m_{2p,2} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$
$D^k(m_{2p+1,k+1} \cdots m_{2p+1,p-1}; x_{2p+1,p})$; $1 \leq k \leq p-1$ $m_{2p+1,j} = \begin{cases} 0 \text{ for } 1 \leq j \leq k \\ 1, 2, 3, \cdots \text{ for } k+1 \leq j \leq p-1 \end{cases}$ $z_{2p+1,p} = x_{2p+1,p}$; $0 \leq x_{2p+1,p} + p - \frac{1}{2} < k + \frac{1}{2}$	$m_{2p,j} = 0$ for $1 \leq j \leq k$ $0 \leq m_{2p,k+1} \leq m_{2p+1,k+1} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$
$D^k(m_{2p+1,k+1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$, $1 \leq k \leq p-2$ $m_{2p+1,j} = \begin{cases} 0 \text{ for } 1 \leq j \leq k \\ 1, 2, 3, \cdots \text{ for } k+1 \leq j \leq p-1 \end{cases}$ $z_{2p+1,p} = m_{2p+1,p}$; $m_{2p+1,p} + p = k + 1$	$m_{2p,j} = 0$ for $1 \leq j \leq k+1$ $m_{2p+1,k+1} \leq m_{2p,k+2} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$
$D^{\pm}(m_{2p+1,1} \cdots m_{2p+1,p-1}; m_{2p+1,p})$ $m_{2p+1,j} = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ for $1 \leq j \leq p-1$ $z_{2p+1,p} = m_{2p+1,p}$; $m_{2p+1,p} + p = \frac{1}{2}, 1, \frac{3}{2} \cdots$	$m_{2p+1,p} \leq \pm m_{2p,1} \leq m_{2p+1,1} \leq \cdots \leq m_{2p+1,p-1} \leq m_{2p,p}$

TABLE II. The complete list of UIR's of $SO_0(2p+1, 1)$.

Representation and conditions for $m_{2p+2,1} \cdots m_{2p+2,p}$ and $z_{2p+2,p+1}$	$SO(2p+1)$ content
$D(m_{2p+2,1} \cdots m_{2p+2,p}; iy_{2p+2,p+1})$ $m_{2p+2,j} = 0, \frac{1}{2}, 1, \cdots$ for $1 \leq j \leq p$ $z_{2p+2,p+1} + p = iy_{2p+2,p+1}; 0 \leq y_{2p+2,p+1}$	$\lvert m_{2p+2,1}\rvert \leq m_{2p+1,1} \leq \cdots \leq m_{2p+2,p} \leq m_{2p+1,p}$
$D^k(m_{2p+2,k+1} \cdots m_{2p+2,p}; x_{2p+2,p+1}), 1 \leq k \leq p$ $m_{2p+2,j} = \begin{cases} 0 \text{ for } 1 \leq j \leq k \\ 1, 2, 3, \cdots \text{ for } k+1 \leq j \leq p \end{cases}$ $z_{2p+2,p+1} = x_{2p+2,p+1}; 0 < x_{2p+2,p+1} + p < k$	$m_{2p+1,j} = 0$ for $1 \leq j \leq k-1$ $0 \leq m_{2p+1,k} \leq m_{2p+2,k+1} \leq \cdots \leq m_{2p+2,p} \leq m_{2p+1,p}$
$D^k(m_{2p+2,k+1} \cdots m_{2p+2,p}; m_{2p+2,p+1}), 1 \leq k \leq p-1$ $m_{2p+2,j} = \begin{cases} 0 \text{ for } 1 \leq j \leq k \\ 1, 2, 3, \cdots \text{ for } k+1 \leq j \leq p \end{cases}$ $z_{2p+2,p+1} = m_{2p+2,p+1}; m_{2p+2,p+1} + p = k$	$m_{2p+1,j} = 0$ for $1 \leq j \leq k$ $0 \leq m_{2p+2,k} \leq m_{2p+2,k+1} \leq \cdots \leq m_{2p+2,p} \leq m_{2p+1,p}$

TABLE III. The UIR's of $SO_0(2, 1)$. The generator is given by Eqs. (41) and (42).

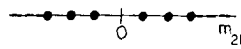
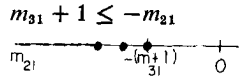
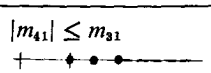
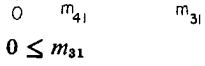
Representations	Conditions for z_{31}	$SO(2)$ content
$D(iy_{31})$	$z_{31} + \frac{1}{2} = iy_{31}; 0 < y_{31}$	$-\infty < m_{21} < +\infty$, integer or half-integer
$D^0(x_{31})$	$z_{31} = x_{31}, 0 \leq x_{31} + \frac{1}{2} < \frac{1}{2}$ $m_{21} = 0, \pm 1, \pm 2, \cdots$	
$D^+(m_{31})$	$z_{31} = m_{31}; m_{31} + 1 = \frac{1}{2}, 1, \frac{3}{2} \cdots$	$m_{31} + 1 \leq m_{21}$ 
$D^-(m_{31})$	Same as for $D^+(m_{31})$	$m_{31} + 1 \leq -m_{21}$ 

TABLE IV. The UIR's of $SO_0(3, 1)$. The generator B_{34} is given by Eqs. (45)–(47).

Representations	Conditions for m_{41} & z_{42}	$SO(3)$ content
$D(m_{41}, iy_{42})$	$\lvert m_{41}\rvert = 0, \frac{1}{2}, 1 \cdots$ $z_{42} + 1 = iy_{42}; 0 \leq y_{42}$	$\lvert m_{41}\rvert \leq m_{31}$ 
$D^1(x_{42})$	$m_{41} = 0$ $z_{42} = x_{42}; 0 < x_{42} + 1 < 1$	$0 \leq m_{31}$ 

From Secs. 2 and 3 we get

$$A_{12} |m_{21}\rangle = im_{21} |m_{21}\rangle, \tag{40}$$

$$B_{23} |m_{21}\rangle = A(m_{21}) |m_{21} + 1\rangle - A(m_{21} - 1) |m_{21} - 1\rangle. \tag{41}$$

There is only one matrix element

$$A(m_{21}) = \tfrac{1}{2}[(z_{31} + \tfrac{1}{2})^2 - (m_{21} + \tfrac{1}{2})^2]^{\frac{1}{2}}. \tag{42}$$

In Table III the classification of the UIR's of $SO_0(2, 1)$ is given, and in the Appendix these results are compared with those of Bargmann.[7]

(b) $n = 3$: A state $|m_{ij}\rangle$ in a UIR is completely specified by the labels

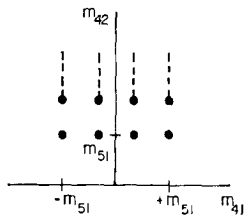
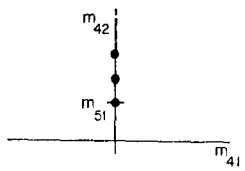
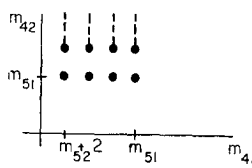
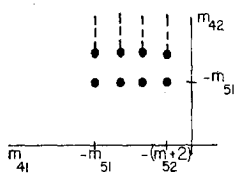
$$|m_{ij}\rangle = \begin{pmatrix} m_{41} & z_{42} \\ m_{31} & \\ m_{21} & \end{pmatrix}. \tag{43}$$

The generator A_{12} is given by (40) and, for the generators A_{23} and B_{34}, we get from Secs. 2 and 3, respectively,

$$A_{23} |m_{21}\rangle = A(m_{21}) |m_{21} + 1\rangle - A(m_{21} - 1) |m_{21} - 1\rangle, \tag{44}$$

$$B_{34} |m_{31}\rangle = B(m_{31}) |m_{31} + 1\rangle + iC_2 |m_{31}\rangle - B(m_{31} - 1) |m_{31} - 1\rangle. \tag{45}$$

TABLE V. The UIR's of $SO_0(4, 1)$. The generator B_{45} is given by Eqs. (50)–(52).

Representations	Conditions for m_{51} & z_{52}	$SO(4)$ content
$D(m_{51}, iy_{52})$	$m_{51} = 0, \frac{1}{2}, 1, \cdots$ $z_{52} + \frac{3}{2} = iy_{52}, 0 < y_{52}$	$\lvert m_{41}\rvert \leq m_{51} \leq m_{42}$ 
$D^0(m_{51}, x_{52})$	$m_{51} = 1, 2, 3, \cdots$ $z_{52} = x_{52}, 0 \leq x_{52} + \frac{3}{2} < \frac{1}{2}$	
$D^1(x_{52})$	$m_{51} = 0$ $z_{52} = x_{52}; 0 \leq x_{52} + \frac{3}{2} < \frac{3}{2}$	
$D^0(m_{51}m_{52})$	$m_{51} = 1, 2, 3, \cdots$ $z_{52} = m_{52}, m_{52} + 2 = 1$	$m_{41} = 0, m_{51} \leq m_{42}$ 
$D^+(m_{51}, m_{52})$	$m_{51} = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ $z_{52} = m_{52}, m_{52} + 2 = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ $1 \leq m_{52} + 2 \leq m_{51}$ or $\frac{1}{2} \leq m_{52} + 2 \leq m_{51}$ for m_{51}, m_{52} integer or half-integer, respectively	$m_{52} + 2 \leq m_{41} \leq m_{51} \leq m_{42}$ 
$D^-(m_{51}, m_{52})$	Same as for $D^+(m_{51}, m_{52})$	$m_{52} + 2 \leq -m_{41} \leq m_{51} \leq m_{42}$ 

The matrix element $A(m_{21})$ is the same as (42), except that z_{31} has to be replaced by m_{31} with the condition $|m_{21}| \leq m_{31}$. The other matrix elements are

$$B(m_{31}) = [m_{21}^2 - (m_{31} + 1)^2]^{\frac{1}{2}} \left(\frac{[m_{41}^2 - (m_{31} + 1)^2][(z_{42} + 1)^2 - (m_{31} + 1)^2]}{(m_{31} + 1)^2[4(m_{31} + 1)^2 - 1]} \right)^{\frac{1}{2}}, \tag{46}$$

$$C_2 = \frac{m_{21} \cdot m_{41}(z_{42} + 1)}{(m_{31} + 1)m_{31}}. \tag{47}$$

In Table IV the UIR's are classified, and in the Appendix the connection with the notation of Gel'fand *et al.*[8] is given.

(c) $n = 4$: A complete labeling for a state $|m_{ij}\rangle$ in a UIR is

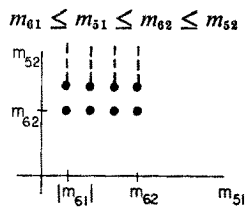
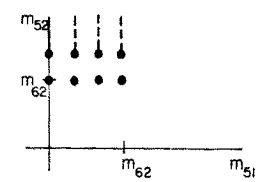
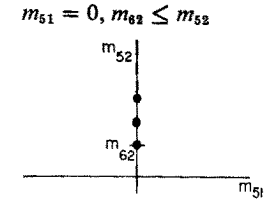
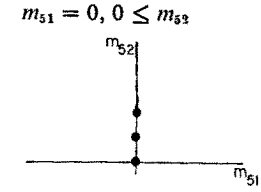
$$|m_{ij}\rangle = \begin{pmatrix} m_{51} & z_{52} \\ m_{41} & m_{42} \\ m_{31} & \\ m_{21} & \end{pmatrix}. \tag{48}$$

The generators A_{12} and A_{23} are given by (40) and (44), respectively; the other generators are

$$A_{34} |m_{31}\rangle = B(m_{31}) |m_{31} + 1\rangle + iC_2 |m_{31}\rangle - B(m_{31} - 1) |m_{31} - 1\rangle, \tag{49}$$

$$\begin{aligned} B_{45} |m_{41}m_{42}\rangle = {} & A(m_{41}) |m_{41} + 1, m_{42}\rangle \\ & + A(m_{42}) |m_{41}, m_{42} + 1\rangle \\ & - A(m_{41} - 1) |m_{41} - 1, m_{42}\rangle \\ & - A(m_{42} - 1) |m_{41}, m_{42} - 1\rangle. \end{aligned} \tag{50}$$

TABLE VI. The UIR's of $SO_0(5, 1)$. The generator is given by Eqs. (55)–(58).

Representations	Conditions for m_{61}, m_{62}, & z_{63}	$SO(5)$ content
$D(m_{61}, m_{62}, iy_{63})$	$m_{61}, m_{62} = 0, \frac{1}{2}, 1, \cdots$ $0 \leq m_{61} \leq m_{62}$ or $\frac{1}{2} \leq m_{61} \leq m_{62}$ for m_{61}, m_{62} integer or half-integer respectively $z_{63} + 2 = iy_{63}, 0 \leq y_{63}$	$m_{61} \leq m_{51} \leq m_{62} \leq m_{52}$ 
$D^1(m_{62}, x_{63})$	$m_{61} = 0; m_{62} = 1, 2, 3, \cdots$ $z_{63} = x_{63}; 0 < x_{63} + 2 < 1$	$0 \leq m_{51} \leq m_{62} \leq m_{56}$ 
$D^1(m_{62}, m_{63})$	$m_{61} = 0; m_{62} = 1, 2, 3, \cdots$ $z_{63} = m_{63}$	$m_{51} = 0, m_{62} \leq m_{52}$ 
$D^1(x_{63})$	$m_{61} = m_{62} = 0$ $z_{63} = x_{63}; 0 < x_{63} + 2 < 2$	$m_{51} = 0, 0 \leq m_{52}$ 

The expressions $B(m_{31})$ and C_2 are the same as (46) and (47), with z_{42} replaced by m_{42} with the condition $|m_{41}| \leq m_{31} \leq m_{42}$. The matrix elements $A(m_{41})$ and $A(m_{42})$ are

$$A(m_{41}) = \tfrac{1}{2}[(m_{31} + \tfrac{1}{2})^2 - (m_{41} + \tfrac{1}{2})^2]^{\frac{1}{2}} \left(\frac{[(m_{51} + \frac{1}{2})^2 - (m_{41} + \frac{1}{2})^2][(z_{52} + \frac{3}{2})^2 - (m_{41} + \frac{1}{2})^2]}{[(m_{42} + 1)^2 - m_{41}^2][(m_{42} + 1)^2 - (m_{41} + 1)^2]} \right)^{\frac{1}{2}}, \tag{51}$$

$$A(m_{42}) = \tfrac{1}{2}[(m_{31} + \tfrac{1}{2})^2 - (m_{42} + \tfrac{3}{2})^2]^{\frac{1}{2}} \left(\frac{[(m_{51} + \frac{1}{2})^2 - (m_{42} + \frac{3}{2})^2][(z_{52} + \frac{3}{2})^2 - (m_{42} + \frac{3}{2})^2]}{[m_{41}^2 - (m_{42} + 1)^2][m_{41}^2 - (m_{42} + 2)^2]} \right)^{\frac{1}{2}}. \tag{52}$$

The UIR's of $SO_0(4, 1)$ are classified in Table V, and in the Appendix we compare these results with those of Dixmier[1] and Strom.[9]

(d) $n = 5$: A state $|m_{ij}\rangle$ is completely labeled by

$$|m_{ij}\rangle = \begin{pmatrix} m_{61} & m_{62} & z_{63} \\ m_{51} & m_{52} & \\ m_{41} & m_{42} & \\ m_{31} & & \\ m_{21} & & \end{pmatrix}. \tag{53}$$

The generators A_{12}, A_{23}, and A_{34} are already known [(40), (44), and (49)]. The remaining two generators are

$$\begin{aligned} A_{45}\,|m_{41}m_{42}\rangle &= A(m_{41})\,|m_{41} + 1, m_{42}\rangle \\ &\quad + A(m_{42})\,|m_{41}, m_{42} + 1\rangle \\ &\quad - A(m_{41} - 1)\,|m_{41} - 1, m_{42}\rangle \\ &\quad - A(m_{42} - 1)\,|m_{41}, m_{42} - 1\rangle, \end{aligned} \tag{54}$$

$$\begin{aligned} B_{56}\,|m_{51}m_{52}\rangle &= B(m_{51})\,|m_{51} + 1, m_{52}\rangle \\ &\quad + B(m_{52})\,|m_{51}m_{52} + 1\rangle \\ &\quad + iC_4\,|m_{51}m_{52}\rangle - B(m_{51} - 1) \\ &\quad \times |m_{51} - 1, m_{52}\rangle - B(m_{52} - 1) \\ &\quad \times |m_{51}m_{52} - 1\rangle. \end{aligned} \tag{55}$$

We get $A(m_{41})$ and $A(m_{42})$ by replacing z_{52} through

m_{52} in (51) and (52) with the condition $|m_{41}| \leq m_{51} \leq m_{42} \leq m_{52}$. The other matrix elements are

$$B(m_{51}) = \{[m_{41}^2 - (m_{51}+1)^2][(m_{42}+1)^2 - (m_{51}+1)^2]\}^{\frac{1}{2}} \times \left(\frac{[m_{61}^2 - (m_{51}+1)^2][(m_{62}+1)^2 - (m_{51}+1)^2][(z_{63}+2)^2 - (m_{51}+1)^2]}{(m_{51}+1)^2[4(m_{51}+1)^2 - 1][(m_{52}+2)^2 - (m_{51}+1)^2][(m_{52}+1)^2 - (m_{51}+1)^2]}\right)^{\frac{1}{2}}, \quad (56)$$

$$B(m_{52}) = \{[m_{41}^2 - (m_{52}+2)^2][(m_{42}+1)^2 - (m_{52}+2)^2]\}^{\frac{1}{2}} \times \left(\frac{[m_{61}^2 - (m_{52}+2)^2][(m_{62}+1)^2 - (m_{52}+2)^2][(z_{63}+2)^2 - (m_{52}+2)^2]}{(m_{52}+2)^2[4(m_{52}+2)^2 - 1][(m_{51}+1)^2 - (m_{52}+2)^2][m_{51}^2 - (m_{52}+2)^2]}\right)^{\frac{1}{2}}, \quad (57)$$

$$C_4 = \frac{m_{41}(m_{42}+1)m_{61}(m_{62}+1)(z_{63}+2)}{(m_{51}+1)m_{51}(m_{52}+2)(m_{52}+1)}. \quad (58)$$

The classification of the UIR's of $SO_0(5, 1)$ is given in Table VI.

ACKNOWLEDGMENTS

The author thanks Dr. S. Strom and L. Jaffe for interesting discussions and carefully reading the manuscript. He further acknowledges some valuable comments of Dr. A. Böhm.

APPENDIX

In this appendix we compare our special cases of Sec. 4 with the results of some other authors, for $n = 2$ with Bargmann,[7] for $n = 3$ with Gel'fand *et al.*,[8] for $n = 4$ with Dixmier[1] and Strom.[9]

(a) $n = 2$: Our generators are related to those of Bargmann[7] through $A_{12} = iH_0$ and $A_{23} = iH_2$, e.g., $B_{23} = H_2$. Our choice of phases is such that $\eta_m \equiv 1$. The connection between the notation in Ref. 7 and for ours is given by the following diagram.

Notation in Ref. 7	q	k	m
Our notation	$\frac{1}{4} - (z_{31} + \frac{1}{2})^2$	$m_{31} + 1$	m_{21}

Our series $D^0(x_{31})$ is the part of C_q^0 in Ref. 7 for which $0 < q \leq \frac{1}{4}$, and our $D(iy_{31})$ is the union of the remaining part at C_q^0 with $\frac{1}{4} < q$ and the series $C_q^{\frac{1}{2}}$. Our series $D^{\pm}(m_{31})$ are directly related to the cases $D_k^{\pm}$ of Ref. 7.

(b) $n = 3$: It is easy to see that Gel'fand *et al.*[8] denote our B_{34} by F_3. The labels are connected in the following way.

Notation in Ref. 8	m	l	l_0	l_1
Our notation	m_{21}	m_{31}	m_{41}	y_{42}

Evidently $D(m_{41}, iy_{42})$ is the main series and $D^1(x_{42})$ the supplementary series.

(c) $n = 4$: In this case we compare our results with those of Strom[9] because he calculates from the generators of the Weyl basis, which is given in Dixmier,[1] the generator $P_0 = iA_{45}$. The connection with our notation is given by the following.

Notation of Strom[9]	m	j	n	l	r	q	σ
Our notation	m_{21}	m_{31}	m_{41}	$m_{42} + 1$	m_{51}	$m_{52} + 2$	$\frac{1}{4} - (z_{52} + \frac{3}{2})^2$

The representations $\pi_{r,q}^{\pm}$ of Strom correspond to our $D^{\pm}(m_{51}, m_{52})$ and $\pi_{r,0}$ to our $D^0(m_{51}, m_{52})$. The representations $\nu_{r,\sigma}$, with $r = \frac{1}{2}, 1, \frac{3}{2} \cdots$ and $\frac{1}{4} < \sigma$, correspond to our $D(m_{51}, iy_{52})$, and those for $r = 1, 2, 3, \cdots$ and $0 < \sigma \leq \frac{1}{4}$ to our $D^0(m_{51}, x_{52})$. The class $\nu_{0,\sigma}$ with $-2 < \sigma \leq \frac{1}{4}$ is our $D^1(x_{52})$, and the remaining part of the series $\nu_{0,\sigma}$ for $\frac{1}{4} < \sigma$ is our $D^0(m_{51}, m_{52})$ for $m_{51} = 0$.

[1] J. Dixmier, Bull. Soc. Math. (France) **89**, 9 (1961).
[2] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSR **71**, 1017 (1950).
[3] J. Dixmier, Compt. Rend. **250**, 3263 (1960).
[4] T. Hirai, Proc. Japan Acad. **38**, 83, 258 (1962).
[5] U. Ottoson, Commun. Math. Phys. **8**, 228 (1968).
[6] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[7] V. Bargmann, Ann. Math. **48**, 586 (1947).
[8] I. M. Gel'fand, R. A. Minlos, and Z. Y. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963).
[9] S. Strom, Arkiv Fysik **30**, 455 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Distributions and Their Hermite Expansions*

BARRY SIMON†
Fine Hall, Princeton University, Princeton, New Jersey 08540

(Received 1 April 1970)

We present a self-contained treatment of the technical parts of distribution theory needed in quantum field theory. The treatment is particularly suited for physicists since an absolute minimum of abstract functional analysis is used: In fact, only the Baire category theorem is needed. The simple nature of some proofs depends on extensive use of the expansion of a distribution as a sum of harmonic oscillator wavefunctions. While this Hermite expansion is not new, the fact that it provides elementary proofs of several theorems does appear to be new.

1. INTRODUCTION AND NOTATION

Schwartz's theory of tempered distributions is basic to the Gårding–Wightman axiomatation of relativistic quantum field theory.[1–3] Field theory requires technical results from distribution theory and not merely the "classical" differential calculus and Fourier analysis of distributions—in particular, the kernel (or nuclear) theorem is needed to define the Wightman functions as distributions in many variables (Ref. 2, p. 106). The purpose of this article is to present a proof of the kernel theorem particularly suited for the physicist—not only is a minimum amount of real analysis used, but the basic tool is the harmonic oscillator wavefunctions, a familiar friend to any physicist.

The approach we use also provides a simple proof of the regularity theorem and several other results mentioned in Streater and Wightman.[2] By adding short sections on the Baire category theorem and on convergence in $\mathcal{D}'$, we are able to provide a complete treatment of the distribution theory used in Ref. 2. We have thus used the discovery of simple proofs of the kernel and regularity theorems to present a general pedagogic presentation to the reader who wishes to study axiomatic field theory without an extensive detour into the functional analysis texts.

Because we will be dealing with several sets of infinitely many norms and with objects in many-dimensional real spaces, an extensive set of notational conventions seems imperative. The letters s and l will refer to the dimension of the underlying real space. $\mathcal{S}(\mathbb{R}^l)$ and $\mathcal{S}^{(l)}$ will be used interchangeably for the functions of rapid decrease in $\mathbb{R}^l$. The letters m, n, α, and β will be used to refer to multi-indices, i.e., l-tuples of nonnegative integers $m = (m_1, \cdots, m_l)$. We adopt the standard notation

$$m! = m_1!\, m_2! \cdots m_l!,$$
$$|m| = m_1 + m_2 + \cdots + m_l,$$
$$(m)^n = m_1^{n_1} \cdots m_l^{n_l},$$
$$m + 1 = (m_1 + 1, \cdots, m_l + 1),$$
$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_l^{\alpha_l}},$$
$$x^\alpha = x_1^{\alpha_1} \cdots x_l^{\alpha_l}.$$

In one place, we will slightly abuse this notation as follows: We will have l-tuples m and n and s-tuples α and β and will write

$$(m + 1)^n(\alpha + 1)^\beta = ((m, \alpha) + 1)^{(n,\beta)}. \quad (1.1)$$

The letters r, s, i, and j will be used for nonnegative integers. We will use the letters a, b, and c to refer to multisequences, i.e., $a = a_n$ where n runs over N^l, that is, all multi-indices of l entries. We will, of course, use other symbols and, while we will use them systematically, we will not introduce them at this point.

We will suppose that the reader is familiar with the basic notions of distribution theory, as presented, for example, in Ref. 2 (for additional background, see Refs. 4 and 5 on an elementary level and 6–10 on a more advanced level).

In Sec. 2, we present without proofs the basic identification of $\mathcal{S}$ and $\mathcal{S}'$ with sequence spaces, an identification which is basic to the simple proofs we present in Secs. 3 and 4. We return in Sec. 5 to the proof of the identifications of Sec. 2. After a section on miscellaneous results which follow from the sequence approach, we derive the remaining distribution theory needed for (axiomatic) quantum field theory as consequences of one general result—the Baire category theorem.

2. THE n-REPRESENTATION

Functions in $\mathcal{S}(\mathbb{R})$ are in L^2 and thus have expansions $\sum a_n\phi_n$, where the ϕ_n are the harmonic oscillator wavefunctions

$$\phi_n(x) = \pi^{-\frac{1}{4}}2^{-\frac{1}{2}n}(n!)^{-\frac{1}{2}}e^{\frac{1}{2}x^2}\left(\frac{d}{dx}\right)^n e^{-x^2}.$$

The simple proofs in Secs. 3 and 4 depend on the characterization of the Hermite coefficients of functions in $\mathcal{S}$:

Theorem 1: Suppose that $f \in \mathcal{S}$ and

$$a_n = \int \phi_n(x) f(x)\, dx.$$

Then, for any m

$$\sum |a_n|^2 (n+1)^m \equiv \|a\|_m^2 < \infty.$$

Conversely, if $\|a\|_m < \infty$ for all m, then $\sum a_n \phi_n$ converges (in the topology of $\mathcal{S}$) to a function in $\mathcal{S}$.

This theorem, which we prove in Sec. 5, establishes an isomorphism between $\mathcal{S}$ and a sequence space. We will call the representation of $\phi \in \mathcal{S}$, as a sequence a_n, the n-space representation.

Not only do the a_n that arise from functions in $\mathcal{S}$ have a simple description, but they also provide a simple form for the notion of convergence in $\mathcal{S}$. $\mathcal{S}$ has convergence defined by an infinite set of norms $\| \ \|_k$, specifically the norms

$$\|f\|_{r,s,\infty,\Sigma} = \sum_{\substack{|\alpha| \le r \\ |\beta| \le s}} \|x^\alpha D^\beta f\|_\infty,$$

where

$$\|g\|_\infty = \sup_x |g(x)|.$$

If x is any countably normed space, one says $x_n \to x$ if $\|x_n - x\|_k \to 0$ for each fixed k. Equivalently, convergence can be described by the metric

$$\rho(f, g) = \sum_k 2^{-k} \min(1, \|f - g\|_k).$$

If x is given two sets of norms $\| \ \|_i$ and $\| \ \|_j$, we say the sets are equivalent if and only if, for any i, there is a C and $j_1, \cdots, j_r$ so that

$$\|f\|_i \le C(\|f\|_{j_1} + \cdots + \|f\|_{j_r})$$

and, for any j, there is a D and $i_1, \cdots, i_s$ so that $\|f\|_j \le D(\|f\|_{i_1} + \cdots + \|f\|_{i_j})$. It is easy to see equivalent sets of norms provide identical notions of convergence, open set, etc., and that "equivalent" is an equivalence relation.

For example, the norms $\| \ \|_{r,s,\infty,\Sigma}$ on $\mathcal{S}$ are "equivalent" to the norms

$$\|f\|_{\alpha,\beta,\infty} = \|x^\alpha D^\beta f\|_\infty.$$

More to the point:

Theorem 2: For $f \in \mathcal{S}$ define

$$\|f\|_m = (\sum (n+1)^m |a_n|^2)^{\frac{1}{2}},$$

where $\{a_n\}$ are the Hermite coefficients for f. The norms $\| \ \|_m$ and $\| \ \|_{r,s,\infty,\Sigma}$ are equivalent.

This characterization of convergence in $\mathcal{S}$ (also proven in Sec. 5) allows us to find an n-space representation of distributions $T \in \mathcal{S}'$. Let us first point out a useful property of the $\| \ \|_m$.

Definition: A countable family of norms $\| \ \|_k$ is called *directed* if for any finite set $k_1, \cdots, k_r$ there is a k and a C so that $\|f\|_{k_1} + \cdots + \|f\|_{k_r} \le C \|f\|_k$.

The families $\{\| \ \|_{r,s,\infty,\Sigma}\}$ and $\{\| \ \|_m\}$ are directed but the family $\{\| \ \|_{\alpha,\beta,\infty}\}$ is not.

Directed families are very useful because they provide a simple description of open sets and continuous functionals. If one looks at the metric ρ, it is not hard to see that, for any family of countable norms, every neighborhood of 0 contains a canonical neighborhood of the form

$$\{x \mid \|x\|_{k_1} \le A_1, \cdots, \|x\|_{k_r} \le A_r\}.$$

If, in addition, the family of norms is directed, every canonical neighborhood contains a simpler neighborhood

$$\{x \mid \|x\|_k \le A\}.$$

Finally, using the fact that the inverse image of $\{z \mid |z| < 1\}$ under a continuous linear functional is open, one finds:

Lemma 1: A linear map $T: X \to \mathbb{C}$ with X a countably normed space *with a directed family of norms* $\{\| \ \|_j\}$ is continuous if and only if $\exists\ C, k$ such that

$$|Tx| \le C \|x\|_k.$$

This fact and the directed nature of the $\| \ \|_m$ allows us to prove:

Theorem 3: Suppose that $T \in \mathcal{S}'(\mathbb{R})$. Let $b_n = T(\phi_n)$. Then $|b_n| \le C(1+n)^m$ for some C and m, and $T(f) = \sum a_n b_n$ if a_n is the n-space representative of f. Conversely, if $|b_n| \le C(1+n)^m$, then $f \to \sum a_n b_n$ defines a tempered distribution.

Proof: Since $T \in \mathcal{S}'$ and $\| \ \|_m$ is directed, $|Tf| \le C \|f\|_m$ for some m. But $\|\phi_n\|_m = (1+n)^{\frac{1}{2}m}$ so that $|Tf| \le C(1+n)^{\frac{1}{2}m} \le C(1+n)^m$. To complete the proof of the first half of the theorem, we use Theorem 1, which tells us $\sum a_n \phi_n$ converges in $\mathcal{S}$ to f. For the

converse, we merely compute the following:

$$\begin{aligned}|\sum a_n b_n|^2 &\leq [\sum |a_n|^2 (n+1)^{2m+2}] \\ &\quad \times [\sum |b_n|^2 (n+1)^{-2m-2}] \\ &\leq C \|f\|_{2m+2}^2 \sum (n+1)^{-2} \\ &\leq \tfrac{1}{6} C \pi^2 \|f\|_{2m+2}^2,\end{aligned}$$

so that $f \to \sum a_n b_n$ is a continuous linear functional on $\mathcal{S}$. QED

We remark that, while we have stated the results for $\mathcal{S}(\mathbb{R})$ and $\mathcal{S}'(\mathbb{R})$, identical results hold for $\mathcal{S}(\mathbb{R}^l)$ and $\mathcal{S}'(\mathbb{R}^l)$. We need only interpret n and m as multi-indices, and

$$\phi_n(x) = \phi_{n_1}(x_1) \cdots \phi_{n_l}(x_l).$$

To summarize, we have seen that, in the n-representation, $\mathcal{S}$ represents just the sequences of fast falloff and $\mathcal{S}'$ represents just the sequences of polynomial growth.

3. THE REGULARITY THEOREM

The regularity theorem for tempered distributions says that any tempered distribution is the derivative of a continuous function of polynomial growth. The usual proof (Ref. 6, pp. 239–43) uses the Hahn–Banach and Reisz–Markov theorems plus a detailed analysis of tempered measures. It might seem a little strange that a theorem that never mentions measures needs measure theory in its proof. In fact, it does not: Using the n-space realization, we present a scandalously elementary proof of this theorem. This proof is a distant relative of the proof given by Zerner.[11]

The basic idea behind the proof is that we expect $\frac{1}{2}[-(d^2/dx^2) + x^2 + 1]$ to act as multiplication by $n+1$ in the n space. In fact:

Lemma 2: Let $T \in \mathcal{S}'(\mathbb{R})$ have Hermite coefficients $b_n = T(\phi_n)$. Then $2^{-m}[-(d^2/dx^2) + x^2 + 1]^m T$ has Hermite coefficients $(n+1)^m b_n$.

Proof:

$$\begin{aligned}2^{-m}\left(-\frac{d^2}{dx^2} + x^2 + 1\right)^m T(\phi_n) &= T\left[2^{-m}\left(-\frac{d^2}{dx^2} + x^2 + 1\right)^m \phi_n\right] \\ &= (n+1)^m T(\phi_n).\end{aligned}$$

The second input to the proof is that $\sum a_n \phi_n$ is "nice" if a_n falls off fast enough. This follows from:

Lemma 3: $\|\phi_n\|_\infty \leq C(n+1)^M$ for some C and M (independent of n).

Proof: By Theorem 2 and the directed nature of the $\| \ \|_m$, $\|f\|_\infty \leq C \|f\|_M$ for some C and M. QED

Remarks: (1) The arithmetic of Sec. 5 actually shows that we can take $M = \frac{1}{2}$. (2) Detailed studies of the generating function for the ϕ_n show that $\|\phi_n\|_\infty \sim C(n+1)^{-\frac{1}{12}}$ as $n \to \infty$.[12]

We are thus ready to prove the regularity theorem.

Theorem 4: Suppose that $T \in \mathcal{S}'$. Then $\exists\, m$ and a continuous bounded function f such that

$$T = \left(-\frac{d^2}{dx^2} + x^2 + 1\right)^m f.$$

Proof: Let b_n be the n-space representative of T. Then $|b_n| \leq C(n+1)^k$ for some k. Let $m = k + M + 2$, where M is given in Lemma 3. Let $a_n = (n+1)^{-m} b_n$. Then

$$\sum a_n \|\phi_n\|_\infty < \infty$$

by Lemma 3, and so $\sum a_n \phi_n$ converges in L^∞ norm (and thus in $\mathcal{S}'$) to a bounded continuous function, say $2^m f$. By Lemma 2,

$$T = \left(-\frac{d^2}{dx^2} + x^2 + 1\right)^m f. \qquad \text{QED}$$

It is now straightforward to obtain the alternate form $T = D^k F$, where F is of polynomial growth. By using multi-indices, we can prove the regularity theorem for $\mathcal{S}(\mathbb{R}^l)$.

4. THE KERNEL THEOREM

Most proofs of the kernel theorem for $\mathcal{S}$ rely heavily on the theory of "nuclear" spaces (see Ref. 10, p. 530 or Ref. 13, pp. 73–84). We present here a proof of the kernel theorem on $\mathcal{S}$ which relies only on the n-space representation. As we will discuss in Sec. 8, this is a relative of existing proofs for $\mathcal{D}$.

In its "normal" form, the kernel theorem is a statement about separately continuous bilinear functionals. We divide it into two parts: that any separately continuous functional is jointly continuous and that jointly continuous functionals have the requisite form. In this section, we consider only the latter part. This part is the crucial half of the kernel theorem—in particular, the kernel theorem fails to hold for, say, $L^2(\mathbb{R}^l)$ because the analog of this half breaks down. We will prove the other part of the kernel theorem in Sec. 7.

Let us first establish the form we will need for joint continuity.

Definition: A bilinear map $B(x, y)$ from pairs $x \in X$, $y \in Y$ into $\mathbb{C}$ is jointly continuous if it is continuous as a map of $X \times Y$ into $\mathbb{C}$, i.e., if and only if for any ϵ, x_0, and y_0 there are neighborhoods N of x_0 and M of y_0 such that $x \in N$, $y \in M$ implies $|B(x, y) - B(x_0, y_0)| < \epsilon$.

Lemma 4: Let X and Y be two countably normed spaces with directed families of norms $\{\| \ \|_r\}$ and $\{\| \ \|_\lambda\}$. Let B be a bilinear form on $X \times Y$. Then the following are equivalent:

(a) B is jointly continuous.

(b) B is jointly continuous at $(0, 0)$.

(c) If $x_n \to 0$, $y_n \to 0$, then $B(x_n, y_n) \to 0$.

(d) For some r, λ, and C

$$|B(x, y)| \leq C \|x\|_r \|y\|_\lambda .$$

Proof: (a) $\Rightarrow$ (b) $\Rightarrow$ (d) $\Rightarrow$ (a) and (c) $\Rightarrow$ (b) can be proven by "standard" methods such as those used for linear functionals in Banach space. We only remark that (c) $\Rightarrow$ (b) depends essentially on the fact that we are in metric spaces where the open sets are describable in terms of *sequential* convergence (for example, the analogous result is false for $\mathcal{D}$). (b) $\Rightarrow$ (d) depends on the fact that the norms are directed. QED

Theorem 5: Let B be jointly continuous bilinear functional on $\mathcal{S}(\mathbb{R}^l) \times \mathcal{S}(\mathbb{R}^s)$. Then there is a unique distribution T in $\mathcal{S}'(\mathbb{R}^{l+s})$ so that

$$B(f, g) = T(f \otimes g),$$

where

$$(f \otimes g)(x, y) = f(x)g(y).$$

Proof: Let C, m, and β be chosen such that

$$|B(f, g)| \leq C \|f\|_m \|g\|_\beta . \tag{4.1}$$

Suppose that

$$t_{(n;\alpha)} = B(\phi_n, \phi_\alpha), \quad n \in N^l, \quad \alpha \in N^s.$$

Since B is jointly continuous and $f = \sum a_n \phi_n$ and $g = \sum b_\alpha \phi_\alpha$, we have that $B(f, g) = \sum t_{n\alpha} a_n b_\alpha$. On the other hand, by (4.1),

$$\begin{aligned} |t_{n,\alpha}| \leq \|\phi_n\|_m \|\phi_\alpha\|_\beta &= (n + 1)^m(\alpha + 1)^\beta \\ &= [(n, \alpha) + 1]^{(m,\beta)}. \end{aligned}$$

Thus the sequence $t_{(n,\alpha)}$ defines an element $\sum t_{(n,\alpha)} \phi_{(n,\alpha)}$ of $\mathcal{S}'(\mathbb{R}^{s+l})$,

$$T(h) = \sum t_{n\alpha} c_{(n,\alpha)},$$

where

$$h = \sum c_{(n,\alpha)} \phi_{(n,\alpha)} .$$

Since $f \otimes g$ has the Hermite coefficients $a_n b_\alpha$, we have that

$$T(f \otimes g) = \sum_{n,\alpha} t_{n_\alpha} a_n b_\alpha = B(f, g).$$

This proves existence. Since T is completely determined by the $T(\phi_{(n,\alpha)})$ (its Hermite coefficients) and we must have

$$T(\phi_{(n,\alpha)}) = T(\phi_n \otimes \phi_\alpha) = B(\phi_n, \phi_\alpha) = t_{n\alpha},$$

T is unique. QED

Theorem 6: Let M be a jointly continuous multilinear functional on $\mathcal{S}(\mathbb{R}^{l_1}) \times \cdots \times \mathcal{S}(\mathbb{R}^{l_r})$. Then there is a unique distribution T in $\mathcal{S}'(\mathbb{R}^{l_1+\cdots+l_r})$ such that

$$M(f_1, \cdots, f_r) = T(f_1 \otimes \cdots \otimes f_r).$$

Proof: The proof is analogous to Theorem 5.

5. PROOFS OF THEOREMS 1 AND 2

We prove Theorems 1 and 2 through a sequence of lemmas.

Lemma 5: Suppose that $f \in \mathcal{S}(\mathbb{R})$ and $a_n = \langle f, \phi_n \rangle$. Then

$$\sum |a_n|^2 (n + 1)^m < \infty$$

for all m.

Proof: Since $f \in D[(p^2 + x^2 + 1)^m]$ for all m, $\sum |a_n|^2 (n + 1)^m = 2^{-m} \langle f, (p^2 + x^2 + 1)^m f \rangle < \infty$ for all m.

To complete the proofs of Theorems 1 and 2, we must first establish the equivalence of the $\| \ \|_{r,s,\infty,\Sigma}$ and $\| \ \|_m$. We do this by establishing the equivalence of each of these families of norms with several families of intermediate norms. First we show the $\| \ \|_{\alpha,\beta,\infty}$ are equivalent to the norms

$$\|f\|_{\alpha,\beta,2} = \|x^\alpha D^\beta f\|_2,$$

with

$$\|f\|_2^2 = \int |f(x)|^2 \, dx.$$

Lemma 6:

$$\|f\|_2 \leq \pi^{\frac{1}{2}}(\|f\|_\infty + \|xf\|_\infty),$$

so that

$$\|f\|_{\alpha,\beta,2} \leq \pi^{\frac{1}{2}}(\|f\|_{\alpha,\beta,\infty} + \|f\|_{\alpha+1,\beta,\infty}).$$

Proof:

$$\begin{aligned} \|f\|_2^2 &= \int_{-\infty}^{\infty} dx(1 + x^2)^{-1}[(1 + x^2)\,|f|^2] \\ &\leq \|(1 + x^2)\,|f|^2\|_\infty \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} \\ &\leq \pi(\|f\|_\infty + \|xf\|_\infty)^2. \end{aligned}$$ QED

To bound the $\| \ \|_\infty$'s by the $\| \ \|_2$ norms, the above

trick does not work. However, the Fourier transform "reverses" the ordering of L^p spaces, explicitly:

Lemma 7: Let

$$\hat{f}(p) = (2\pi)^{-\frac{1}{2}} \int e^{-ipx} f(x)\, dx.$$

Then, for $f \in \mathcal{S}$,

$$\|f\|_2 = \|\hat{f}\|_2, \quad \|f\|_\infty \leq (2\pi)^{-\frac{1}{2}} \|\hat{f}\|_1.$$

Proof: The L^2 inequality is, of course, well known to any physicist (see, e.g., Ref. 14, pp. 355–62). The L^∞ inequality is trivial. QED

Lemma 8:

$$\|\hat{f}\|_1 \leq \pi^{\frac{1}{2}}(\|\hat{f}\|_2 + \|p\hat{f}\|_2)$$

so that (by Lemma 7)

$$\|f\|_\infty \leq 2^{-\frac{1}{2}}(\|f\|_2 + \|Df\|_2)$$

and

$$\|f\|_{\alpha,\beta,\infty} \leq 2^{-\frac{1}{2}}(\|f\|_{\alpha,\beta,2} + \|f\|_{\alpha,\beta+1,2} + \alpha \|f\|_{\alpha-1,\beta,2}).$$

Proof: As in the proof of Lemma 6,

$$\begin{aligned}\|\hat{f}\|_1 &= \int_{-\infty}^{\infty} dp[(1+p^2)^{\frac{1}{2}} |\hat{f}|][(1+p^2)^{-\frac{1}{2}}] \\ &\leq \left[\int_{-\infty}^{\infty} dp(1+p^2)^{-1}\right]^{\frac{1}{2}} \left(\int dp\, |\hat{f}|^2 + |p\hat{f}|^2\right)^{\frac{1}{2}} \\ &\leq \pi^{\frac{1}{2}}(\|\hat{f}\|_2 + \|p\hat{f}\|_2).\end{aligned}$$ QED

N.B.: (1) Lemma 8 is known as a Sobolev inequality in the mathematics literature. (2) Thus,

$$\begin{aligned}\|\phi_n\|_\infty &\leq \|\phi_n\|_2 + \|p\phi_n\|_2 \\ &= 1 + 2^{-\frac{1}{2}}[n^{\frac{1}{2}} + (n+1)^{\frac{1}{2}}] \\ &\leq 3(n+1)^{\frac{1}{2}},\end{aligned}$$

as stated in Sec. 2.

We thus see that the $\| \ \|_{\alpha,\beta,\infty}$ and $\| \ \|_{\alpha,\beta,2}$ families are equivalent. Finally:

Lemma 9: The $\| \ \|_{r,s,\infty,\Sigma}$ and $\| \ \|_m$ norms are equivalent.

Proof: All we need prove is that the $\| \ \|_m$ and the $\| \ \|_{\alpha,\beta,2}$ norms are equivalent because we already have proven the $\| \ \|_{\alpha,\beta,2}$ and $\| \ \|_{r,s,\infty,\Sigma}$ equivalent. Let $\eta^\dagger$ and η be the usual creation and annihilation operators. Since η, $\eta^\dagger$ are linear combinations of x, p, and vice versa, and since any polynomial in x's and p's is equal to a polynomial with only $x^\alpha D^\beta$ terms, the $\| \ \|_{\alpha,\beta,2}$ norms are equivalent to the norms $\|(\eta^\#)^k f\|_2$, where $(\eta^\#)^k$ is a generic symbol for a monomial of degree k in η and $\eta^\dagger$. Since $\|f\|_m = \|(\eta\eta^\dagger)^m f\|_2$, the $\| \ \|_m$ are a subset of the $\|(\eta^\#)^k \cdot \|_2$ norms. But it is easy to see that $\|(\eta^\#)^k f\|_2 \leq (2k)^k \|f\|_m$ with $m = k$ (crude estimate) so that the $\| \ \|_m$ norms are equivalent to the $\|(\eta^\#)^m \cdot \|_2$. This completes the proof. QED

To complete this section, we need only show that $\|a\|_m < \infty$ implies that $\sum a_n \phi_n$ converges in $\mathcal{S}$. This is a consequence of the equivalence of the norms and the fact that $\mathcal{S}$ is complete. (For a proof of this last fact, see Ref. 10, pp. 92–94.)

6. OTHER THEOREMS IN THE n-REPRESENTATION

In this section we point out several theorems whose proofs are also simple in the n-representation.

Theorem 7: $\mathcal{S}$ is separable; i.e., it has a countable dense set.

Proof: Since $\sum_{n<N} a_n \phi_n$ converges to f in $\mathcal{S}$ if $a_n = \langle \phi_n, f \rangle$, the finite linear combinations of the ϕ_n with rational coefficients are dense in $\mathcal{S}$ and are countable. QED

Theorem 8: $\mathcal{S}$ is dense in $\mathcal{S}'$ in the weak topology on $\mathcal{S}'$.

Proof: If $b_n = T(\phi_n)$, $\sum_{n<N} b_n \phi_n \to T$ in the weak topology on $\mathcal{S}'$. But $\sum_{n<N} b_n \phi_n \in \mathcal{S}$. QED

The next result is a little surprising:

Theorem 9[15]*:* For any l, $\mathcal{S}^{(l)}$ and $\mathcal{S}^{(1)}$ are isomorphic as topological vector spaces. Thus, for any s and l, $\mathcal{S}(\mathbb{R}^l)$ and $\mathcal{S}(\mathbb{R}^s)$ are isomorphic.

Proof: We prove the result for $l = 2$. The proof is similar for $l > 2$. Consider the map u of N^2 onto N by $u(0,0) = 0$, $u(1,0) = 1$, $u(0,1) = 2$; $u(2,0) = 3, \cdots$; i.e.,

$$u(r,s) = \tfrac{1}{2}(r+s)(r+s+1) + s.$$

We map $\mathcal{S}^{(1)}$ onto $\mathcal{S}^{(2)}$ by $(F(a))_{r,s} = a_{u(r,s)}$. Because $u(r,s)$ obeys the relations

$$r \leq u(r,s), \quad s \leq u(r,s),$$
$$u(r,s) + 1 \leq (r+1)^2(s+1)^2,$$

we immediately have

$$\sum_{n=1}^{\infty} (n+1)^m |a_n|^2 \leq \sum_{r,s=1}^{\infty} (r+1)^{2m}(s+1)^{2m} |F(a)_{r,s}|^2$$

and

$$\sum_{r,s=1}^{\infty} (r+1)^{m_1}(s+1)^{m_2} |F(a)_{r,s}|^2 \leq \sum_{n=1}^{\infty} (n+1)^{m_1 m_2} |a_n|^2.$$

Thus the norms $a \to \|F(a)\|_{m_1,m_2}$ and the $\| \ \|_m$ are equivalent. QED

This identity of $\mathcal{S}^{(l)}$ and $\mathcal{S}^{(s)}$ is not so useful as one might think at first. It says that we only have to prove theorems for $\mathcal{S}^{(1)}$ *if the theorem only refers to the "internal" structure of* $\mathcal{S}$. However, theorems like the regularity and kernel theorem refer to "external" structure, i.e., the realization of some distributions as functions and the map of $\mathcal{S}^{(l_1)} \times \mathcal{S}^{(l_2)}$ into $\mathcal{S}^{(l_1+l_2)}$.

7. THE BAIRE CATEGORY THEOREM AND APPLICATIONS

There are four results mentioned in Ref. 2 which we have not yet proven:

(1) The completeness of $\mathcal{S}'$.
(2) The nature of bounded sets in $\mathcal{S}$ (equivalence of weak and norm boundedness).
(3) The fact that separate continuity implies joint continuity for bilinear forms.
(4) The uniform convergence on bounded sets of ordinary distributions.

One is able to prove (1)–(3) from one abstract principle (Theorem 10); we will also be able to prove a weak form of (4) sufficient for the application in Ref. 2. The material in this section is rather standard. We only present it here because it is usually difficult to cull only the results needed for Ref. 2 from the texts.

Theorem 10 (Baire Category Theorem): Let X be a complete metric space and suppose that

$$X = \bigcup_{i=1}^{\infty} A_i.$$

Then some i, $\bar{A}_i$ has a nonempty interior.

Proof: The argument is quite simple. See Ref. 16.

As a simple consequence:

Theorem 11 (Principle of Uniform Boundedness): Let X be a countably normed space with $\| \ \|_r$ a directed sequence of norms. Let $\mathcal{F}$ be a set in X', the dual of X. If $\{F(f) \mid F \in \mathcal{F}\}$ is bounded for each $f \in X$, then there is a C and an r so that, for all f and all $F \in \mathcal{F}$,

$$|F(f)| \leq C \|f\|_r.$$

Proof: Suppose that $S_N = \{f \in X \mid |F(f)| < N$ for all $F \in \mathcal{F}\}$. Then each S_N is closed and, by the hypothesis of the theorem, $X = \bigcup S_N$. Thus, for some N, S_N has a nonempty interior. Therefore there exist an N, r, f_0, and ϵ such that $\|g - f_0\|_k \leq \epsilon$ implies $g \in S_N$. Suppose that $a = \sup |F(f_0)|$. Then $\|h\|_r < \epsilon$ and $F \in \mathcal{F}$ imply that

$$|F(h)| \leq |F(f_0 + h)| + |F(f_0)| \leq N + a.$$

Therefore,

$$|F(h)| \leq [(N + a)/\epsilon] \|h\|_r \quad \text{for all } h \text{ and } F \in \mathcal{F}.$$

QED

Corollary 1: $\mathcal{S}'$ is weakly sequentially complete.

Proof: Let T_n be a weak Cauchy sequence of tempered distributions; i.e., for each $f \in \mathcal{S}$, let $T_n(f)$ be a Cauchy sequence of numbers. Since this is Cauchy, $\lim T_n(f) = T(f)$ as $n \to \infty$ exists. T defined this way is linear. We must only show it is continuous. But since $\lim T_n(f)$ as $n \to \infty$ exists, $\{T_n(f)\}_{n=1,2,\ldots}$ is bounded. Thus, by Theorem 11, for some C and m, $|T_n(f)| \leq C \|f\|_m$. Therefore, $|T(f)| \leq C \|f\|_m$, and so T is continuous.

Corollary 2: Let B be a separately continuous bilinear form on $\mathcal{S}(\mathbb{R}^{l_1}) \times \mathcal{S}(\mathbb{R}^{l_2})$. Then it is jointly continuous.

Proof: Let $f_n \to 0, g_n \to 0$, where $f_n \in \mathcal{S}^{(l_1)}, g_n \in \mathcal{S}^{(l_2)}$. We need only show that $B(f_n, g_n) \to 0$. Let $F_n(g) = B(f_n, g)$. By continuity for fixed f, $F_n \in \mathcal{S}^{(l_2)'}$. By continuity for fixed g, $F_n(g) \to 0$ for each g, and thus $\{F_n(g)\}$ is bounded for each g. Thus, for some C and m, $|F_n(g)| \leq C \|g\|_m$ for all n. Since $g_n \to 0$ in $\mathcal{S}^{(l_2)}$, $\|g_n\|_m \to 0$. QED

Corollary 3: Let M be a separately continuous r-linear form on $\mathcal{S}(\mathbb{R}^{l_1}) \times \cdots \times \mathcal{S}(\mathbb{R}^{l_r})$. Then it is jointly continuous.

Proof: We use induction on r. $r = 2$ has been proved in Corollary 2. Assuming the results for $r = R$, we let $M(f^{(1)}, \cdots, f^{(R+1)})$ be given, and let $f_n^{(j)}$ be sequences in $\mathcal{S}^{(l_j)}$ with $f_n^{(j)} \to 0$. For each $g \in \mathcal{S}^{(l_{R+1})}$, $M(-, -, \cdots, g)$ is jointly continuous as an R-linear form by the induction hypothesis, and thus

$$M(f_n^{(1)}, \cdots, f_n^{(R)}, g) \to 0.$$

Proceeding as in the proof of Corollary 2, we see that

$$M(f_n^{(1)}, \cdots, f_n^{(R+1)}) \to 0$$

so that M is jointly continuous. QED

We can also discuss bounded sets by using Theorems 10 and 11.

Theorem 12: For a set $A \subset \mathcal{S}$, the following are equivalent:

(a) For any neighborhood N of 0, there is a real number λ with $\lambda A \subset N$.
(b) For each m, $\{\|f\|_m \mid f \in A\}$ is bounded.
(c) For each $F \in \mathcal{S}'$, $\{F(f) \mid f \in A\}$ is bounded.

Proof: (a) $\Leftrightarrow$ (b) is quite simple, as is (b) $\Rightarrow$ (c). To prove (c) $\Rightarrow$ (b), we proceed as follows: For each $g \in L^2(\mathbb{R}^s)$ and fixed $(\eta^\#)^m$,

$$\left\{ \int (\eta^\#)^m \tilde{g}(f) = \int \tilde{g}[(\eta^\#)^m f] \,\middle|\, f \in A \right\}$$

is bounded. By Theorem 11 with $X = L^2$, the $\{\|(\eta^\#)^m f\|_2 \,|\, f \in A\}$ are bounded. Thus (b) follows. QED

Definition: A set $A \subset \mathcal{S}$ obeying the conditions in Theorem 12 is called bounded.

Theorem 13: for a set $B \subset \mathcal{S}'$, the following are equivalent:

(a) For any $f \in \mathcal{S}$, $\{F(f) \mid F \in B\}$ is bounded.

(b) There is a C and an m such that, for all $F \in B$, $f \in \mathcal{S}$, $|F(f)| \leq C \|f\|_m$.

(c) For any bounded set $A \subset \mathcal{S}$, $\{F(f) \mid F \in B, f \in A\}$ is bounded.

Proof: (c) $\Rightarrow$ (a) is easy. (a) $\Rightarrow$ (b) is Theorem 12. (b) $\Rightarrow$ (c) is proven as follows: Given (b) and a bounded set A, $\sup\{\|f\|_m f \in A\} = k < \infty$. Thus, for any $f \in A$, $F \in B$, $|F(f)| \leq Ck$. QED

Definition: A set $B \subset \mathcal{S}'$ obeying the conditions in Theorem 13 is called bounded.

Theorem 14: A sequence $f_n \to f$ in the $\| \ \|$-topology on $\mathcal{S}$ if and only if, for any bounded set $B \subset \mathcal{S}'$, $F(f_n) \to F(f)$ uniformly for $F \in B$.

Proof: One direction of the proof (only if) is a simple consequence of Theorem 13(b). The other direction is an interesting exercise; since this theorem is purely motivational, we do not provide a complete proof.

Theorem 14 suggests the following definition:

Definition: A sequence of distribution F_n is said to converge strongly to F if and only if, for any bounded subset $A \subset \mathcal{S}$, $F_n(f) \to F(f)$ uniformly for $f \in A$.

The analog for $\mathcal{S}$ of statement (4) at the beginning of the section is the following theorem, which we will not prove.

Theorem 15: A sequence of distributions F_n converges strongly to F if and only if it converges weakly.

Remark: In Theorem 14, we could replace "sequence" by the more general notion of net necessary for the complete description of a topology by convergence. However, the word sequence is essential in Theorem 15 and cannot be replaced by net.

Theorem 15 is implied by two other statements.

Theorem 16: Suppose that $F_n \to F$ weakly with F_n, $F \in \mathcal{S}'$. Let $A \subset \mathcal{S}$ be a compact subset. Then $F_n(f) \to F(f)$ uniformly for $f \in A$.

Theorem 17: If $A \subset \mathcal{S}$ is bounded, then A is compact.

We will not prove Theorem 17, but Theorem 16 will follow from results in Sec. 8.

8. A THEOREM FOR ORDINARY DISTRIBUTIONS

In Ref. 2, Wightman and Streater state and use the analog of Theorem 15 for $\mathcal{D}$. Actually, one only needs a weak analog of Theorem 16 in his application, and we will prove this weak form in this section and show that it suffices in the application.

Lemma 10: Let X be a countably normed space and suppose that F_n, $F \in X'$, the dual of X. Suppose that $F_n(x) \to F(x)$ for each x in X. Let A be a compact subset of X. Then $F_n(x) \to F(x)$ uniformly for $x \in A$; that is, given ϵ, we can find N such that $n > N$ and $x \in A$ implies $|F_n(x) - F(x)| < \epsilon$.

Proof: Let $\| \ \|_r$ be a directed sequence of norms for X. By Theorem 11, there is a C and an r such that $|F_n(x)| \leq C\|x\|_r$ and $|F(x)| \leq C\|x\|_r$. For each $x \in A$, let $B_x = \{y \mid \|x - y\|_r < \epsilon/3C\}$. The $\{B_x \mid x \in A\}$ cover A and so, by compactness, we find $x_1, \cdots, x_m$ such that $\bigcup_{i=1}^m B_{x_i} \supset A$. Since $F_n(x_i) \to F(x_i)$, we can find N such that $n > N$ implies $|F_n(x_i) - F(x_i)| < \epsilon/3$ for $i = 1, \cdots, m$. Let $x \in A$. Find i such that $x \in B_{x_i}$, i.e., $\|x - x_i\|_r < \epsilon/3G$. Then, for any n,

$$|F_n(x) - F_n(x_i)| < \epsilon/3 \quad \text{and} \quad |F(x) - F(x_i)| < \epsilon/3.$$

Thus, if $n > N$,

$$|F(x) - F_n(x)| \leq |F(x) - F(x_i)| + |F(x_i) - F_n(x_i)| + |F_n(x_i) - F_n(x)| < \epsilon.$$

This proves the result since ϵ is arbitrary.

Remarks: (1) The proof of Lemma 10 is really a classical equicontinuity argument. (2) Thus Theorem 16 is proven.

Theorem 18: Let $A \subset \mathcal{D}(\mathcal{O})$ be a family of functions such that

(i) A is compact in $\mathcal{D}(\mathcal{O})$.

(ii) For some fixed compact $C \subset \mathcal{O}$, $f \in A$ implies supp $f \subset C$. Let $F_n, F \in \mathcal{D}(\mathcal{O})'$ and let $F_n(g) \to F_n(g)$ for all $g \in \mathcal{D}$. Then $F_n(f) \to F(f)$ uniformly for $f \in A$.

Remarks: (1) It is not difficult to prove that (i) implies (ii), but in the application (ii) can be directly verified. (2) One can actually show that any closed bounded set in $\mathcal{D}$ obeys (i) and (ii), and so the analog of Theorem 17 follows.

Proof of 18: Let $\mathcal{D}_c(\mathcal{O})$ be the subspace of $\mathcal{D}(\mathcal{O})$ of functions with support in C. The topology of $\mathcal{D}(\mathcal{O})$ restricted to $\mathcal{D}_c(\mathcal{O})$ is described by the norms $\|f\|_\alpha = \|D^\alpha f\|_\infty$, so that $\mathcal{D}_c(\mathcal{O})$ is a countably normed space. Since the F_n and F are continuous on $\mathcal{D}(\mathcal{O})$, they are continuous on $\mathcal{D}_c(\mathcal{O})$, and thus we are in the conditions of Lemma 10. QED

The application in Ref. 2 of the uniform convergence idea is to the case (Ref. 2, p. 83) $A = \{f_{-x}\}$ where $f \in \mathcal{D}(\mathcal{O})$ and x lies in some small compact, so small that we can assume (ii) without thought. To verify (i) is easy: The map $x \to f_{-x}$ is continuous, and so the image of a compact set of x is compact in $\mathcal{D}(\mathcal{O})$.

9. RELATION TO OTHER APPROACHES

The crucial element in the proofs of Secs. 5 and 6 is the realization of $\mathcal{S}$ as a sequence space and the realization of topology in terms of the norms $\| \ \|_m$. The systematic use of Hermite expansions goes back at least as far as Weiner.[17] Our realization of $\mathcal{S}$ is certainly not new; there is a short discussion of it in Schwartz's book (Ref. 6, pp. 271–83). A set of norms closely related to the $\| \ \|_m$ is implicit in Schwartz and a similar set of norms is discussed by Kristensen *et al.*[18] The proof of the kernel theorem in their norms

$$\|a\|_{r,K}^2 = \sum_n (|n| + l)^r |c_n|^2$$

is not as direct as in the $\| \ \|_m$ since the multiplicative property (1.1) avoids messy arithmetic. The only "new" result which we can possibly claim is the fact that the n-space realization of $\mathcal{S}$ provides a simple proof of the nuclear theorem—but this proof is clearly related to the various proofs of the nuclear theorem for $\mathcal{D}$ which depend on Fourier series (Ref. 13, pp. 11–18; Refs. 19 and 20); in fact, our proof must be the "analogous proof for $\mathcal{S}$" alluded to by Gel'fand and Vilenkin (Ref. 13, p. 19). However, for the student of Ref. 2 faced with the statement "there does not seem to be an analogous elementary proof available for $\mathcal{S}$" (p. 43), it seems useful to have the details spelled out.

It is interesting to notice the close connection with Bargmann's beautiful and complete treatment of tempered distributions.[21] He realizes $\mathcal{S}$ as a family of entire functions and finds that $\mathcal{S}'$ can also be so realized. Up to the factors of $\sqrt{n!}$ the Taylor coefficients for his entire functions are just the Hermite coefficients of the elements of $\mathcal{S}$ and $\mathcal{S}'$. His Hilbert spaces F_r are just the multisequences with $\|a\|_r < \infty$ (although his inner product is not quite that given by $\| \ \|_r$ and he has r run over the all reals). Bargmann's results that $\mathcal{S}$ is "essentially" $\bigcap_{k=-\infty}^{+\infty} \mathcal{F}_k$ and $\mathcal{S}'$ is essentially $\bigcup_{k=-\infty}^{+\infty} \mathcal{F}_k$ (Ref. 21, p. 4) is evident from our Theorems 1 and 3. Bargmann's proofs of the regularity and kernel theorems (Ref. 20, pp. 70 and 68) are more or less our proofs in a complex function theoretic guise. In one sense, then, our simple proof is based on the observation that *for these two theorems* Bargmann's proofs do not require the elaborate constructions he uses. However, the treatment of the wide array of problems he considers uses analytic function theory (particularly variants of the maximum modulus principle) in an essential way. [Perhaps the relation of our approach to Bargmann's can be illustrated by remarking that it is identical to the relation of Schwinger's creation operator treatment of angular momentum,[22] to Bargmann's approach[24] for $SU(2)$.]

To the reader who wishes to use this note as a jumping off point for a more detailed study of tempered distributions, we can recommend Bargmann's approach most emphatically. Alternately, sequence spaces have been studied extensively by Köthe.[25,26]

We should also mention to the student of axiomatic field theory that, while he can avoid delving into the theory of nuclear spaces in studying Ref. 2, Jaffe's important work on "strictly localizable fields"[27] introduces a large class of test function spaces for which the kernel theorem is needed and for which the Hermite expansion method does not work.

ACKNOWLEDGMENTS

It is a pleasure to thank E. Nelson, R. Powers, and M. Reed for valuable discussion. They bear much of the credit (but none of the responsibility!) for the genesis of this note and for several simplifications.

* This research partially sponsored by the Air Force Office of Scientific Research under Contract AF 49(638)1545.

† N.S.F. Predoctoral Fellow, 1966–1969.

[1] A. S. Wightman and L. Gårding, Arkiv Fysik **28**, 129 (1965).

[2] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).

[3] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[4] M. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions* (Cambridge U.P., New York, 1958).

[5] I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1967), Vol. I.
[6] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966), rev. ed.
[7] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I.
[8] H. Bremmerman, *Distributions, Complex Variables and Fourier Transforms* (Addison-Wesley, Reading, Mass., 1965).
[9] J. Gårding and J. Lions, Nuovo Cimento Suppl. **14**, 9 (1959).
[10] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[11] M. Zerner, "Morceaux choisis de théorie des distributions," in *Cargése Lecture Notes, 1964, Vol. 2: High Energy Electromagnetic Interactions and Field Theory*, edited by M. Lévy (Gordon and Breach, New York, 1967).
[12] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, New York, 1959), rev. ed., p. 240.
[13] I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. IV.
[14] L. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1968).
[15] M. Reed, private communication.
[16] K. Yoshida, *Functional Analysis* (Academic, New York, 1965), p. 11.
[17] N. Weiner, Acta Math. **55**, 117 (1930), and *The Fourier Integral* (Cambridge U.P., Cambridge, 1933), pp. 51–71. I should like to thank Professor R. Jost for bringing these references to my attention.
[18] P. Kristensen, L. Mejlbo, and E. T. Poulsen, Commun. Math. Phys. **1**, 175 (1965), esp. p. 184.
[19] L. Ehrenpreis, Proc. Am. Math. Soc. **7**, 713 (1956).
[20] H. Gask, Math. Scand. **8**, 327 (1960).
[21] V. Bargmann, Commun. Pure Appl. Math. **20**, 1 (1967).
[22] J. Schwinger, U.S. Atomic Energy Commission, NYO-3071, 1952, reprinted in Ref. 23.
[23] L. Biedenharn and H. Van Dam, Eds., *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[24] V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962), reprinted in Ref. 23.
[25] G. Köthe, Math. Z. **51**, 317 (1948).
[26] G. Köthe, *Topologische Lineare Raume* (Springer-Verlag, Berlin, 1960), esp. pp. 422–26.
[27] A. Jaffe, Phys. Rev. **158**, 1454 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Can Local Gauge Transformations Be Implemented?

G. F. DELL'ANTONIO*†
Institute for Advanced Study, Princeton, New Jersey 08540

(Received 25 May 1970)

One studies conditions under which a gauge transformation can be implemented by a unitary operator in some representations of the canonical (anti-) commutation relations; an application is then given to local gauge transformations in field theory.

INTRODUCTION

This paper originates from an attempt to understand the source of the difficulties which one faces in constructing generators of "canonical" local transformations of relativistic fields. We shall be mainly concerned with gauge transformation of the second type, but most of what will be said goes over almost verbatim to, e.g., "internal symmetry" groups.

The fact that such generators cannot be constructed in otherwise simple cases has been known for some time,[1] although domain problems plague the nonexistence proofs (these generators, when they exist, are expected to be unbounded operators). We shall discuss the existence of a continuous group of unitary operators that induce the group of transformations considered, a problem equivalent to the previous one, via Stone's theorem. We shall consider only some representations of the canonical fields, selected for having a structure particularly well suited for our purposes, and probably of not much physical interest,[2] and we shall show that in most of them (in a sense to be made precise later) such a weakly continuous group of unitary operators cannot be found.

Also, while the nonexistence of the generators of gauge groups (equal-time currents) as bona fide operators may cast some doubts on their formal manipulations, meaningful results can be obtained by giving them a meaning as bilinear forms.[1] Our results have no bearing on such an approach.

The content of the paper is as follows: In Sec. 1 we pose the problem and fix our notation, and actually generalize the previous setup in a rather natural way. Section 2 will be devoted to the solution of the generalized problem. In Sec. 3 the case of relativistic free fields will be considered, in the light of the preceding results, and the corresponding statement about local "charges" will be explicated.

1. THE PROBLEM

A. Canonical Anticommutation Relations

We shall start posing our problem in the case of canonical anticommutation relations (CAR's).

Let $\{a_i, a_i^*\}$, $i = 1, \cdots, n$, be a countable set of

[5] I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1967), Vol. I.
[6] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966), rev. ed.
[7] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I.
[8] H. Bremmerman, *Distributions, Complex Variables and Fourier Transforms* (Addison-Wesley, Reading, Mass., 1965).
[9] J. Gårding and J. Lions, Nuovo Cimento Suppl. **14**, 9 (1959).
[10] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[11] M. Zerner, "Morceaux choisis de théorie des distributions," in *Cargése Lecture Notes, 1964, Vol. 2: High Energy Electromagnetic Interactions and Field Theory*, edited by M. Lévy (Gordon and Breach, New York, 1967).
[12] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, New York, 1959), rev. ed., p. 240.
[13] I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. IV.
[14] L. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1968).
[15] M. Reed, private communication.
[16] K. Yoshida, *Functional Analysis* (Academic, New York, 1965), p. 11.
[17] N. Weiner, Acta Math. **55**, 117 (1930), and *The Fourier Integral* (Cambridge U.P., Cambridge, 1933), pp. 51–71. I should like to thank Professor R. Jost for bringing these references to my attention.
[18] P. Kristensen, L. Mejlbo, and E. T. Poulsen, Commun. Math. Phys. **1**, 175 (1965), esp. p. 184.
[19] L. Ehrenpreis, Proc. Am. Math. Soc. **7**, 713 (1956).
[20] H. Gask, Math. Scand. **8**, 327 (1960).
[21] V. Bargmann, Commun. Pure Appl. Math. **20**, 1 (1967).
[22] J. Schwinger, U.S. Atomic Energy Commission, NYO-3071, 1952, reprinted in Ref. 23.
[23] L. Biedenharn and H. Van Dam, Eds., *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[24] V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962), reprinted in Ref. 23.
[25] G. Köthe, Math. Z. **51**, 317 (1948).
[26] G. Köthe, *Topologische Lineare Raume* (Springer-Verlag, Berlin, 1960), esp. pp. 422–26.
[27] A. Jaffe, Phys. Rev. **158**, 1454 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Can Local Gauge Transformations Be Implemented?

G. F. DELL'ANTONIO*†
Institute for Advanced Study, Princeton, New Jersey 08540

(Received 25 May 1970)

One studies conditions under which a gauge transformation can be implemented by a unitary operator in some representations of the canonical (anti-) commutation relations; an application is then given to local gauge transformations in field theory.

INTRODUCTION

This paper originates from an attempt to understand the source of the difficulties which one faces in constructing generators of "canonical" local transformations of relativistic fields. We shall be mainly concerned with gauge transformation of the second type, but most of what will be said goes over almost verbatim to, e.g., "internal symmetry" groups.

The fact that such generators cannot be constructed in otherwise simple cases has been known for some time,[1] although domain problems plague the nonexistence proofs (these generators, when they exist, are expected to be unbounded operators). We shall discuss the existence of a continuous group of unitary operators that induce the group of transformations considered, a problem equivalent to the previous one, via Stone's theorem. We shall consider only some representations of the canonical fields, selected for having a structure particularly well suited for our purposes, and probably of not much physical interest,[2] and we shall show that in most of them (in a sense to be made precise later) such a weakly continuous group of unitary operators cannot be found.

Also, while the nonexistence of the generators of gauge groups (equal-time currents) as bona fide operators may cast some doubts on their formal manipulations, meaningful results can be obtained by giving them a meaning as bilinear forms.[1] Our results have no bearing on such an approach.

The content of the paper is as follows: In Sec. 1 we pose the problem and fix our notation, and actually generalize the previous setup in a rather natural way. Section 2 will be devoted to the solution of the generalized problem. In Sec. 3 the case of relativistic free fields will be considered, in the light of the preceding results, and the corresponding statement about local "charges" will be explicated.

1. THE PROBLEM

A. Canonical Anticommutation Relations

We shall start posing our problem in the case of canonical anticommutation relations (CAR's).

Let $\{a_i, a_i^*\}$, $i = 1, \cdots, n$, be a countable set of

operators on a Hilbert space $\mathcal{H}$, satisfying the relations

$$a_i a_j^* + a_j^* a_i = I\delta_{ij}, \quad a_i a_j + a_j a_i = 0, \tag{1}$$

where A^* is the adjoint of A, I the identity in $\mathfrak{L}(\mathcal{H})$, and δ_{ij} is the Kronecker delta.

Let $\hat{\mathcal{A}}$ be the smallest (concrete) C^*-algebra containing all a_i and a_i^*; we shall denote by $\mathcal{A}$ the algebra $\hat{\mathcal{A}}$ considered as an abstract algebra. If $\{\lambda_n\}$, $n = 1, 2, \cdots$, is a sequence of real numbers, the mapping τ_θ defined for each θ by

$$a_n \to a_n e^{i\lambda_n\theta}, \quad a_n^* \to a_n^* e^{-i\lambda_n\theta} \tag{2}$$

defines an automorphism of $\hat{\mathcal{A}}$ and therefore of $\mathcal{A}$.

We are interested in the following question: In which representation π of $\mathcal{A}$ can one find a (strongly) continuous group of unitary operators U such that, for all a_n,

$$e^{i\lambda_n\theta}\pi(a_n) = U_\theta \pi(a_n) U_\theta^{-1}, \quad \forall\, \theta, \tag{3}$$

where $\pi(a_n)$ is the representative of a_n.

One could alternatively consider the free Fermi system[3] over[4] $\mathfrak{L}^2(\mathbb{R}^3, \mu)$ generated by

$$\mathfrak{L}^2(\mathbb{R}^3) \ni f \to a(f), \quad a^*(f).$$

[It is then part of the definition that $f \to a(f)$ is linear and affine and that $a(f)a^*(g) + a^*(g)a(f) = (g,f)$, where $(\cdot,\cdot)$ denotes scalar product in $\mathfrak{L}^2(\mathbb{R}^3)$].

The automorphisms of interest would then be the ones induced by a one-parameter continuous group of unitary operators on $\mathfrak{L}^2(\mathbb{R}^3)$, with generator Γ having totally discrete spectrum.

This is actually the setup with an interpretation in "current algebra" terminology; the correspondence with the previous discrete scheme is given by $a_n \leftrightarrow a(g_n)$, where $g_n \in \mathfrak{L}^2(\mathbb{R}^3)$ are the eigenfunctions of Γ and $\Gamma g_n = \lambda_n g_n$.

B. Tensor Product Representations of $\mathcal{A}$

We shall only consider a special type of representations of the algebra $\mathcal{A}$, the "tensor-product" ones,[5] whose defining properties will be outlined presently. This will enable us to discuss the existence of "local generators" in the simple but instructive case of free fields; a similar analysis can presumably be performed for other representations, such as the exponential ones,[6] so as to accommodate also cases of interacting fields; this is, however, beyond the scope of the present paper. Let the $\mathcal{H}_i$, $i = 1, 2, \cdots$, be 2-dimensional Hilbert spaces and $\Omega_i \equiv (\alpha_i, \beta_i)$ be a vector in $\mathcal{H}_i$ of norm one ($|\alpha_i|^2 + |\beta_i|^2 = 1$).

Consider, in the Cartesian product $\prod_{i=1,2,\cdots} \mathcal{H}_i$ (the set of sequences $\{\xi_i\}$, $\xi_i \in \mathcal{H}_i$), the subset Σ defined by: $\xi_i \neq \Omega_i$ only for a finite number of indices.

On Σ a (strictly positive) bilinear form is defined by

$$(\{\xi_i\} \mid \{\eta_i\}) = \prod_{i=1,\cdots} (\xi_i, \eta_i),$$

where (ξ_i, η_i) is the scalar product in $\mathcal{H}_i$ and by convention

$$\prod_{i=1,\cdots} \lambda_i = 1 \quad \text{if} \quad \lambda_i = 1, \quad \forall\, i.$$

The completion of Σ in the norm defined by $(\cdot\, 1\, \cdot)$ is called tensor product of the $\mathcal{H}_i$'s relative to the vector $\{\Omega_i\}$ and is denoted by

$$\mathcal{H}_\Omega \equiv \overset{\{\Omega_i\}}{\otimes} \mathcal{H}_i.$$

Let $\hat{a}_i$, $\hat{a}_i^*$, and $\hat{N}_i$ be the operators on $\mathcal{H}_i$ defined by (in matrix notation) $a_i \leftrightarrow \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, $a_i^* \leftrightarrow \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, and $N_i \leftrightarrow \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$; then the operators defined on $\mathcal{H}_\Omega$ by

$$a_i\{\xi_j\} \equiv \{e^{i\pi N_1}\xi_1, \cdots, e^{i\pi N_{i-1}}\xi_{i-1}, \hat{a}_i\xi_i, \xi_{i+1}, \cdots\},$$
$$a_i\{\xi_j\} \equiv \{e^{+i\pi N_1}\xi_1, \cdots, e^{i\pi N_{i-1}}\xi_{i-1}, \hat{a}_i^*\xi_i, \xi_{i+1}, \cdots\}$$

satisfy

$$a_i a_k^* + a_k^* a_i = \delta_{ik}, \quad a_i a_j + a_j a_i = 0$$

and provide therefore a representation of the CAR on $\otimes^\Omega \mathcal{H}_i$, called the tensor product representation relative to $\Omega \equiv \{\Omega_i\} \equiv (\Omega_1, \Omega_2, \cdots)$.

For different choices of $\{\Omega_i\}$ one has, in general, different representations; necessary and sufficient conditions for their equivalence are known,[7] and we could formulate our problem in that language.

2. SOLUTION OF THE GENERALIZED PROBLEM

A. Unitary Operators

We shall now determine conditions for the existence of unitary operators with the properties

$$U_\theta a_k U_\theta^{-1} = e^{i\theta\lambda_k} a_k, \quad k = 1, \cdots, n.$$

Suppose that U_θ exists. Let $V_{\theta,k}$ be defined on $\mathcal{H}_k$ by

$$V_{\theta,k}\begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} = \begin{pmatrix} \alpha_k \\ e^{i\lambda_k\theta}\beta_k \end{pmatrix}.$$

With $\mathcal{H}_N \equiv \otimes_{i=1}^N \mathcal{H}_i$, one can show that

$$\mathcal{H}^{\{\Omega_k\}} = \mathcal{H}_N \otimes \mathcal{H}^{\Omega^1}, \quad \Omega^1 = \{\Omega_k, k \geq N\}. \tag{4}$$

Let $V_\theta^{(N)}$ be defined by

$$V^{(N)}\{\xi_k\} = \{\xi_k^1\}, \quad \xi_k^1 = V_{\theta,k}\xi_k,$$
$$\text{for} \quad k \leq N, \quad \xi_k^1 = \xi_k, \quad k > N. \tag{5}$$

Then $U_\theta^{-1}V_\theta^{(N)}$ is of the form $I \otimes W_\theta^{(N)}$ relative to the decomposition (4), and one has

$$|U_\theta\Omega, \Omega)| \leq |(V_\theta^N\Omega_N, \Omega_N)|$$
$$= \prod_{k=1}^N [1 - 2x_k(1 - x_k)(1 - \cos\lambda_k\theta)],$$

where $\Omega_N = \{\Omega_k, k \leq N\}$ and $x_k = |\alpha_k|^2 = 1 - |\beta_k|^2$, $0 \leq x_k \leq 1$. Similarly, let $\xi \equiv \{\xi_k\} \in \mathcal{H}^\Omega$ be such that $\exists\, k_0 > 0$ for which $\xi_k = \Omega_k$ if $k > k_0$ (vectors of this form are dense in $\mathcal{H}^\Omega$ by construction).

Arguing as before, one proves

$$|(U_\theta \xi, \xi)| \leq \prod_{k=1}^{k_0} [1 - 2x_k(1 - x_k)(1 - \cos \lambda_k \theta)]. \quad (6)$$

It follows that a necessary and sufficient condition for the existence of U_θ in the given representation is

$$\sum_k x_k(1 - x_k)(1 - \cos \theta\lambda_k) < \infty. \quad (7)$$

Indeed, if (6) were not satisfied, one would have $(\xi, U_\theta \xi) = 0$ for ξ in a dense subset of $\mathcal{H}$, which is impossible if U_θ is unitary (and therefore bounded). On the other hand, if (6) is satisfied, one can easily prove that U_θ exists and is, in fact, the strong limit of V_θ^N, as defined in Eq. (5), when $N \to \infty$.

Clearly, inequality (7) is satisfied for all θ and all choices of $\{\lambda_\theta\}$ if

$$\sum_k x_k(1 - x_k) < \infty. \quad (8)$$

Tensor product representations relative to vectors Ω for which (8) is satisfied are called discrete in the terminology of Ref. 8; in such representations the automorphisms defined by (2) are always implemented by a unitary operator, and, in fact, U_θ is strongly continuous in θ. We shall therefore concentrate on the problem of the existence of the unitary operator U_θ in representations which are not discrete. The following result is easily established.

Proposition 1: If the (tensor product) representation is not discrete and if $\{\lambda_k\}$ has a finite number of limit points, among which there is neither 0 nor ∞, then (7) can be satisfied for at most one value of θ in the open interval 0–2π.

The proof follows immediately from the following remark: If $\Lambda_1, \cdots, \Lambda_s$ are the limit points of $\{\lambda_k\}$, let I^p, $p = 1, \cdots, s$, be defined by

$$I^p: \{k \in Z \mid |\lambda_k - \Lambda_p| < \epsilon\},$$

where Z are the integers and 2ϵ is chosen smaller than

$$\inf_{\substack{(i,j) \\ i \neq j}} |\Lambda_i - \Lambda_j|.$$

Then for at least one value of p, $1 \leq p \leq s$, one has

$$\sum_{k \in I_p} x_k(1 - x_k) = \infty.$$

By similar arguments, it is easily established that if $\{\lambda_k\}$ does not have 0 and ∞ as limit points and the representation is not a discrete one, one can find a neighborhood $\mathcal{N}_0$ of $\theta = 0$ such that U_θ does not exist for $\theta \in \mathcal{N}_0$, $\theta \neq 0$. We now consider the general case in which $\{\lambda_k\}$ can have any limit point [e.g., $\lambda_k \equiv k$, in which case (2) could be interpreted as translations in a discrete space, see Ref. 9].

Given the representation and the sequence $\{\lambda_k\}$, let $\mathcal{N}$ be the set of values of θ for which U_θ exists. By Schwartz's inequality it is easily established that $\mathcal{N}$ is closed over the field of integers Z, i.e.,

$$\varphi \in \mathcal{N}, \quad \psi \in \mathcal{N} \Rightarrow m\varphi + n\psi \in \mathcal{N}, \quad \forall\, m, n. \quad (9)$$

One proves in addition that if $\theta_1, \theta_2 \in \mathcal{N}$,

$$F(\theta_1 + \theta_2) \leq \{[F(\theta_1)]^{\frac{1}{2}} + [F(\theta_2)]^{\frac{1}{2}}\}^2, \quad (10)$$

where $F(\theta) = \sum x_k(1 - x_k)(1 - \cos \theta\lambda_k)$. From (9) and (10) one sees the following.

Remark 1: If $F(\theta)$ is uniformly bounded on an open interval of the real line, then it is uniformly bounded on every bounded interval.

We shall now prove the following

Proposition 2: If the (tensor product) representation is not discrete, U_θ cannot exist for all $\theta \in \mathcal{N}$ if $\mathcal{N}$ is an open set with Lebesque measure $\neq 0$.

It will be sufficient to prove this proposition for an open interval (a, b); due to the group structure of U_θ ($U_{\theta_1} U_{\theta_2} = U_{\theta_1 + \theta_2}$), one can take $a = 0$. Consider the (increasing) sequence of positive functions F_N defined by

$$F^N(\theta) = \sum_1^N x_k(1 - x_k)(1 - \cos \lambda_k \theta). \quad (11)$$

For each N, F^N is continuous and bounded, and therefore measurable on any Borel set. Assume that $F^N(\theta)$ converges [to $F(\theta)$] for all θ in $(0, c)$. [Since $F^{N+1}(\theta) \geq F^N(\theta)$, $F^N(\theta)$ converges to $F(\theta)$, for all θ for which $F(\theta)$ exists.] We want to prove that this leads to a contradiction.

By Osgood's theorem,[10] $F(\theta)$ is then continuous on a residual set,[11] since it is the pointwise limit of a sequence of continuous functions. Therefore, one can find $0 < a < b < c$ such that $F(\theta)$ is uniformly bounded for $a < \theta < b$. But then, by Remark 1, it is uniformly bounded on any bounded interval (a', b').

Consider now

$$G_N^{(a,b)} \equiv \int_a^b F^N(\theta) \alpha\theta = \sum_{k=1}^N x_k(1 - x_k)[b - a - k^{-1}(\sin \lambda_k b - \sin \lambda_k a)]. \quad (12)$$

Since $F^N(\theta)$ is uniformly bounded for all N, including $N = \infty$, and $F^N(\theta)$ converges pointwise to $F(\theta)$, it follows that the convergence is uniform on every bounded open interval and therefore that

$$G_N^{(a,b)} \xrightarrow[N\to\infty]{} G^{(a,b)} \equiv \int_a^b F(\theta)\alpha\theta.$$

From (12) one can see, however, that if $\{\lambda_k\}$ does not have 0 as a limit point, $G_N^{(a,b)}$ does not converge for any (a, b) when $N \to \infty$ [since $1 - \lambda_k^{-1} \sin \lambda_k > \delta(\epsilon)$ if $\lambda_k > \epsilon$].

We conclude therefore the following.

Proposition 3: If the representation is not discrete and if $\{\lambda_k\}$ does not have zero as limit point, the operator U_θ cannot exist for all θ in $\mathcal{N}$, if $\mathcal{N}$ is an open set of Lebesque measure $\neq 0$.

It should be remarked that, under the assumptions of Proposition 3, the set of θ's for which U_θ exists could be dense in the real line, as the following simple example shows.

Assume that $\lambda_k = k$ and $x_n(1 - x_n) = k^{-1}\delta_{n,k!}$, $\sum_n x_n(1 - x_n) = \infty$. If $\theta/2\pi$ is rational, one has

$$\sum_n x_n(1 - x_n)(1 - \cos \theta n) < \infty,$$

since only a finite number of terms in the sum are nonzero. Therefore, U_θ exists for all θ's for which $\theta/2\pi$ is rational. In fact, given a countable set Σ_0 in 0–2π, one can show that there exists a (nondiscrete) representation in which U_θ exists for all $\theta \in \Sigma_0$.

On the other hand, if

$$\sum_k \lambda_k^2 < \infty,$$

U_θ exists for all θ's in all tensor product representations; this is, in fact, a particular case of a result which will be established later.

Finally, combining (9), (10), and Osgood's theorem, one has

Proposition 4: If U_θ exists for a set of Lebesque measure different from zero, then U_θ is (strongly) continuous in θ on every open interval, and in particular on any neighborhood of the origin, and, therefore, by Stone's theorem, there exists a self-adjoint operator Γ such that $U_\theta = e^{i\Gamma\theta}$ (indeed, λ_k are its eigenvalues).

B. Tensor Product Representations of CCR's

All the results outlined so far hold also for the tensor product representations of canonical commutation relations. These are defined much in the same way as for the CAR's, but now the $\mathcal{H}_i$ are infinite-dimensional (separable) Hilbert spaces; $\hat{a}_i$ and $\hat{a}_i^*$ satisfy on $\mathcal{H}_i$ the relations

$$[\hat{a}_i, \hat{a}_i^*] = 1, \quad [\hat{a}_i, \hat{a}_i] = 0,$$

and a_i and a_i^* are defined by

$$a_i^{\#}\{\xi_1, \cdots, \xi_i, \cdots\} = \{\xi_1, \cdots, \hat{a}_i\xi_i, \cdots\}, \qquad a_i^{\#} = a_i \text{ or } a_i^*,$$

on

$$\mathcal{H} = \bigotimes_i^{\{\Omega_i\}} \mathcal{H}_i, \quad \Omega_i \in \mathcal{H}_i.$$

The mapping we are interested in is again

$$a_k \to a_k e^{i\lambda_k\theta}. \tag{2'}$$

One can also here reformulate the problem in terms of a free Bose system[3] over $\mathcal{E}(\mathbb{R}^3)$, where $\mathcal{E}$ is some subset of $\mathcal{L}^2(\mathbb{R}^3)$. The situation is, however, more complicated here, as compared to the Fermi system, since there is no "canonical" candidate for $\mathcal{E}(\mathbb{R}^3)$ and, if one chooses a particular basis $\{g_i\}$, $g_i \in \mathcal{E}$, $\forall\, i$, and the correspondence

$$a(g_i) \to a_i, \quad a^*(g_i) \to a_i^*,$$

then the results may depend on the basis chosen.

We shall set aside this problem for the time being and concentrate on the discrete formulation.

A necessary and sufficient condition for the existence of U_θ is found as in the case of the CAR's and is

$$\sum_{n,i,j} \beta_n^i \beta_n^j [1 - \cos(\theta_{n,i} - \theta_{n,j})] < \infty, \tag{7'}$$

where $\theta_{n,k} = k\theta\lambda_n$ and $\beta_n^k = |\alpha_n^k|^2$, $\sum_k \beta_n^k = 1$, $\forall\, n$, and the α_n^k are defined by

$$\Omega_n = \sum_k \alpha_n^k \xi_n^k, \quad a_n^* a_n \xi_n^k = k\xi_n^k, \quad |\xi_n^k| = 1.$$

(The last equations define ξ_n^k, modulo an inessential phase factor.)

A sufficient condition for the validity of (7') is

$$\sum_{\substack{n,i,j \\ i\neq j}} \beta_n^i \beta_n^j < \infty. \tag{8'}$$

The (product) representations for which (8') is satisfied are called discrete in the terminology of Ref. 8; indeed, (8') is equivalent to

$$\sum_n (1 - \beta_n^{k(n)}) < \infty \tag{8''}$$

for a (unique, modulo changments in a finite number of points) choice of the function k.

The equivalence of (8') and (8'') is easily established and will not be proven here. In a discrete representation, for every value of $\theta \in \mathbb{R}^1$ there exists a unitary

operator U_θ which induces the automorphism defined by (2'), and, in fact, the (group of) U_θ's is strongly continuous in θ.

We shall henceforth assume that (8') is not satisfied. Let

$$\gamma_{\xi,n} = \sum_n \beta_n^k \beta_n^{k+\xi}, \quad \sum_\xi \gamma_{\xi,n} = 1, \quad \forall\, n.$$

Let $\alpha \leftrightarrow (n, \xi)$, $n = 1, 2, \cdots$ and $\xi = 1, 2, \cdots$ be an enumeration of $Z^+ \times Z^+$; and let

$$\rho_\alpha \equiv \xi\lambda_n, \quad x_\alpha = \gamma_{\xi,n}.$$

Then (7') takes the form

$$\sum_{\alpha=1}^{\infty} x(1 - \cos \rho_\alpha\theta) < \infty, \quad \sum x_\alpha = \infty. \qquad (7'')$$

This is very similar to (7), however, since $\{\rho_\alpha\}$ can have ∞ as a limit point even if $\{\lambda_n\}$ does not, the analog of Proposition 1 does not hold here.[12] One has the analog of Proposition 2 and 3, and therefore:

Proposition 3': If a product representation of the CCR's is not discrete and if the sequence $\{\lambda_k\}$ does not have zero as limit point, an operator U_θ inducing (2') cannot exist for all $\theta \in \mathcal{N}$, if $\mathcal{N}$ is an open set of Lebesque measure $\neq 0$. As in the case of the CAR's, if $\sum x_\alpha(1 - \cos \rho_\alpha\theta) < \infty$, $\theta \in \mathcal{N}$, $\mathcal{N}$ an open set in $\mathbb{R}^1$, then $(\Omega, U_\theta\Omega)$ is continuous in θ in any neighborhood of the origin. Also $(\xi, U_\theta\xi)$ is continuous, where $\xi = \{\xi_k\}$, $\xi_k \neq \Omega_k$ only for a finite number of k's. As the set of such ξ is dense in $\otimes^\Omega \mathcal{H}_k$ and U_θ is unitary, we have the following.

Proposition 4': If (7') is satisfied for all $\theta \in \mathcal{A}$, $\mathcal{A}$ an open interval of $\mathbb{R}^1$, then the U_θ form a strongly continuous group of unitary operators, and a self-adjoint operator Γ exists, such that $U_\theta = e^{i\theta\Gamma}$.

C. Extensions to Other Representations of the CAR's and CCR's

The following results are easily established, and extend slightly the representations of the CAR's and of the CCR's for which τ_θ defined by (2) [resp. (2')] can be induced by a group of unitary operators.

Proposition 5: If the product representations studied in Secs. 2(A) and 2(B) are subrepresentations of a representation π acting on a Hilbert space $\mathcal{H}$ and if the hypotheses of Propositions 3 and 3' are satisfied, then one cannot find a continuous group of unitary operators U_θ on $\mathcal{H}$ such that

$$U_\theta\pi(A)U_\theta^{-1} = \pi(\tau_\theta(A)).$$

Indeed, let π_1 be a subrepresentation of π, and let the hypotheses of Propositions 3 and 3' be satisfied by π_1. Assume that for all $\theta \in \mathcal{N}$, $\mathcal{N}$ a neighborhood of the origin, one can find U_θ (on $\mathcal{H}$) such that

$$U_\theta\pi(a_k)U_\theta^{-1} = \pi(a_k)e^{i\theta\lambda_k}.$$

We will prove that this leads to a contradiction.

Since π_1 is a subrepresentation of π, there exist a projection $P \in \pi(\mathcal{A})'$ [the commutant of $\pi(\mathcal{A})$] such that

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2, \quad \mathcal{H}_1 = P\mathcal{H}, \quad \pi_1(\mathcal{A}) = \pi(\mathcal{A})\big|_{P\mathcal{H}}.$$

But then one can find operators U_{ij} from $\mathcal{H}_j$ to $\mathcal{H}_i$, $\|U_{ij}\| \leq 1$, such that

$$\pi_i(\tau_\theta(A))U_{ij} = U_{ij}\pi_j(A), \quad U_{11}U_{11}^* \in \pi_1(\mathcal{A})'. \quad (13)$$

Since $\pi_1(\mathcal{A})$ is irreducible, $U_{11}U_{11}^* \equiv c(\theta) \in \mathbb{R}^1$, $|c| \leq 1$.

If $U(\theta)$ is strongly continuous in θ, also $c(\theta)$ is continuous; but $c(\theta)$ takes the value 1 at $\theta = 0$. Therefore, one can find a neighborhood $\mathcal{N}$ of the origin such that $c(\theta) > \frac{1}{2}$, $\theta \in \mathcal{N}$.

It follows from (13) that the operators V_θ defined, for $\theta \in \mathcal{N}$, by $U_{11}(\theta) = c^{\frac{1}{2}}(\theta)V_\theta$ induce the automorphism τ_θ in the representation π_1; but this is excluded by Propositions 3 and 3'.

D. Further Extensions

We shall need an extension of the previous results to the case of tensor product representations (of the CAR's or of the CCR's) which are constructed as in Secs. 2(A) and 2(B), but with "a finite number of degrees of freedom" in each space $\mathcal{H}_i$. More precisely, taking, e.g., the case of CCR's, the representation is now defined on

$$\bigotimes_1^\Omega \mathcal{H}_i, \quad \Omega = \{\Omega_i, \|\Omega_i\| = 1\},$$

by operators a_i, a_i^*, b_i, and b_i^* defined by

$$a_i\{\xi_i\} = \{\xi_1, \cdots, \hat{a}_i\xi_i, \cdots), \text{ etc.},$$

where $\hat{a}_i, \cdots, \hat{b}_i^*$ are operators on $\mathcal{H}_i$ satisfying

$$[\hat{a}_i, \hat{a}_i^*] = [\hat{b}_i, \hat{b}_i^*] = 1, \quad \text{all other commutators} = 0.$$

The operators $a_i, \cdots, b_i^*$ thus satisfy

$$[a_i, a_j^*] = [b_i, b_j^*] = \delta_{ij}, \quad \text{all other commutators vanish.} \quad (14)$$

The automorphism τ_θ (of the algebra generated by a_i and b_i) is now characterized by

$$a_k \to a_k e^{i\theta\lambda_k}, \quad b_k \to b_k e^{-i\theta\lambda_k}. \qquad (15)$$

One could again reformulate everything in terms of a Bose (or Fermi) system over $\mathfrak{L}^2(\mathbb{R}^3) \oplus \mathfrak{L}^2(\mathbb{R}^3)$; it is

then immediately seen that the generalization introduced here is necessary if τ_θ has to acquire an interpretation as a gauge group of the second kind.

Also, for the representations of the algebra (14) and the action of the group of transformations (15), a result holds which is the analog of Propositions 3 and 3′; the place of the discrete representations is now taken by a somewhat more complex class. Let a_i and b_i be given as operators in $\mathcal{H}_i$; a tensor product representation of the algebra (14) will be called charge discrete if it is (unitarily equivalent to the one) defined by the construction outlined in Sec. 1(B), on the space $\otimes_i^{\xi} \mathcal{H}_i$, where the vector $\xi \equiv \{\xi_i\}$ (also denoted by $\otimes_i \xi_i$) is charge discrete in the following sense.

Definition: With the previous notation, let ξ_i^n, $n = 0, \pm 1, \pm 2, \cdots$, be the orthogonal projections of $\xi_i \in \mathcal{H}_i$ on the eigenspace of $a_i^* a_i - b_i^* b_i$ relative to the eigenvalue n. Then ξ is charge discrete if and only if

$$\sum_{n,i,j} |\xi_i^n|^2 \, |\xi_j^n|^2 < \infty \tag{16}$$

or, equivalently, if there exists a function $k \to n(k)$ such that

$$\sum_k (1 - |\xi_k^{n(k)}|^2) < \infty. \tag{16′}$$

(Loosely speaking, the ξ_i are, for all but a finite number of values of i, "almost" eigenvalues of $N_{a_i} - N_{b_i}$.)

With this definition, one has the following proposition, a proof of which can be obtained by a simple adaptation of the proof of Propositions 3 and 3′:

Proposition 6: The mapping τ_θ, defined by (15) on the algebra $\mathcal{A}$ generated by the operators in (14), can be implemented in a tensor product representation by unitary operators U_θ for all θ in a neighborhood of $\theta = 0$ (and then in all neighborhoods) if and only if the representation is charge discrete. When this condition is met, the U_θ can be taken to form a weakly continuous group, and the corresponding generator Γ has ξ in its domain if $\{\lambda_k\}$ does not have ∞ as a limit point (this condition is only sufficient).

3. RELATIVISTIC FREE FIELDS AND LOCAL CHARGES

A. Local Gauge Transformations

We shall now use the results of Sec. 2 to make statements about local gauge transformations of a relativistic free field. We shall consider only a charged scalar field, but the results hold for charged fields of any spin.

A free scalar field is (see, e.g., Ref. 13) a continuous linear functional on a space of (sufficiently smooth) solutions of the Klein–Gordon equation, with values in (unbounded) self-adjoint operators on a Hilbert space $\mathcal{H}$ such that

$$[\phi^*(f), \phi(g)] = (f, g)_{\mathrm{KG}} I,$$
$$[\phi(f), \phi(g)] = 0, \tag{17}$$

with

$$(f, g)_{\mathrm{KG}} \equiv \int_{t=0} \left(\bar{f}(x, t) \frac{\partial}{\partial t} g(x, t) - g(x, t) \frac{\partial}{\partial t} \bar{f}(x, t) \right) \alpha^3 x.$$

Another definition,[3,14,15] which avoids the domain questions which must be specified in (17), is, loosely speaking, an exponentiated version of (17), and consists in a mapping from a space of (smooth) solutions of the Klein–Gordon equation to pairs of unitary operators $W_1(f)$, $W_2(g)$ satisfying

$$W_1(f) W_1(g) = W_1(g) W_1(f),$$
$$W_2(f) W_2(g) = W_2(g) W_1(f) \exp[\tfrac{1}{2} \operatorname{Im} (f, g)_{\mathrm{KG}}] \tag{17′}$$

with the proviso that only those representations are considered in which

$$\lambda \to W_1(\lambda f)$$

is, for fixed f, strongly continuous in λ (to ensure the existence of the fields ϕ).

We shall, by way of example, consider the Fock representation, in which all domain problems are easily settled (and also the arbitrariness in the choice of the space of solutions of the KG equation does not effect the results), and we shall work in the unexponentiated form.

From the corresponding property of the solutions of the KG equation, it follows that a free field is identified by the restriction at some fixed time t_0 of the field and of its first time derivative. In conventional notations, this characterizes a field as a mapping

$$\mathfrak{L}^2(\mathbb{R}^3) \supset S \ni f \to (\phi(f, t_0), \pi(f, t_0)).$$

We have

$$\phi(f, t) = a(\tilde{f}\omega^{-\frac{1}{2}} e^{i\omega t}) + b^*(\tilde{f}\omega^{-\frac{1}{2}} e^{-i\omega t}),$$
$$i\pi(\tilde{f}, t) = a(f\omega^{\frac{1}{2}} e^{i\omega t}) - b^*(\tilde{f}\omega^{\frac{1}{2}} e^{-i\omega t}), \tag{18}$$

where $\tilde{f}$ is the Fourier transform of f and $\omega(k) = (k^2 + m^2)^{\frac{1}{2}}$.

In (18), $a(\chi)$ and $b(\chi)$ are representations of a Bose system [over $\mathfrak{L}^2(\mathbb{R}^3)$]; in the representation space there exists a (unique, up to a phase) vector Ω in the domain of $a(\cdot)$, $b(\cdot)$, such that $a(\chi)\Omega = b(\chi)\Omega = 0$, $\chi \in \mathfrak{L}^2(\mathbb{R}^3)$.

It should be noted that another natural representation of a Bose system can be obtained from (18) (always on Fock space) writing

$$\phi(f,t) = \alpha_t(f) + \beta_t^*(f),$$
$$i\pi(f,t) = \alpha_t(f) - \beta_t^*(f). \tag{19}$$

There exist now, in the Fock representation, no vector Ω' in the domain of α_t and α_t^* (or rather of their closure) such that $\alpha_t(\chi)\Omega' = \beta_t(\chi)\Omega' = 0$, for χ and some t. One could, of course, construct a representation in which such a vector Ω exists; this would be a natural representation for the description of gauge transformations, inasmuch as one can find unitary operators inducing any given transformation $f \to e^{i\phi}f$, at least if ϕ is sufficiently smooth.

However, in terms of the α and the β which appear in (19), the time evolution has now a complicated form, e.g.,

$$\alpha_t(f) = \alpha(f_t^{(1)}) + \beta^*(f_t^{(2)}), \tag{19'}$$

where

$$\tilde{f}_t^{(1)}(k) = \cos(t\omega_k) + \tfrac{1}{2}i(\omega_k^2 + \omega_k^{-2})\sin(t\omega_k),$$
$$\tilde{f}_t^{(2)}(k) = \tfrac{1}{2}i(\omega_k^2 - \omega_k^{-2}\sin(t\omega_k).$$

It is easily seen, using, e.g., the methods of Ref. 16, that for no value of t can one find a unitary operator $U(t)$ inducing (19'); in particular, a Hamiltonian does not exist.

This fact is at the basis of the difficulty in constructing "local charges" for a free relativistic field and, in general, for a nonfree field in the interaction representation. We are now ready to prove the following.

Proposition 7: Let $Sf \to (\phi(f), \pi(f))$ provide the ($t = 0$ description of the) Fock representation of a free scalar charged relativistic field; let B be any open set in $\mathbb{R}^3$ with the piecewise differentiable boundaries. Then there exists no real number θ, different from $2n\pi$, $n \in Z$, for which one can find a unitary operator U_θ such that

$$U_\theta \begin{Vmatrix} \phi(f) \\ \pi(f) \end{Vmatrix} U_\theta^{-1} = e^i \begin{Vmatrix} \phi(f) \\ \pi(f) \end{Vmatrix} \quad \text{if} \quad \operatorname{supp} f \subset B,$$
$$U_\theta \begin{Vmatrix} \phi(f) \\ \pi(f) \end{Vmatrix} U_\theta^{-1} = \begin{Vmatrix} \phi(f) \\ \pi(f) \end{Vmatrix} \quad \text{if} \quad \operatorname{supp} f \cap B = \phi. \tag{20}$$

The proof relies on the fact that,[15] under the hypotheses of Proposition 7, the representation space $\mathcal{H}_F$ can be written as

$$\bigotimes_i^{\Omega_i} \mathcal{H}_i;$$

on each $\mathcal{H}_i$ acts a quartet of operators $\hat{a}_i$, $\hat{a}_i'$, $\hat{b}_i$, $\hat{b}_i'$ such that $[\hat{a}_i, \hat{a}_i^*] = \cdots [\hat{b}_i', \hat{b}_i'^*] = 1$ and all other commutators vanish. The set $\hat{a}_i, \cdots, \hat{b}_i'^*$ is irreducible in $\mathcal{H}_i$. Denote by $a_i, \cdots, b_i'^*$ the images of the natural "lifting" of $\hat{a}_i, \cdots$ to $\otimes_1^\Omega \mathcal{H}_i$. The a_i and a_i' provide a representation of a (complex) free Bose system over

$$\mathcal{M}^- \equiv \{\omega^{-\frac{1}{2}}f, f \in S, \operatorname{supp} f \subset B\}^-$$

(the closure is taken in the topology provided by the positive bilinear form

$$[f,g] = (f, \omega^{-1}g)_{\mathfrak{L}^2(\mathbb{R}^3)}),$$

with the correspondence $a \leftrightarrow a(\xi_i)$, $a_i' \leftrightarrow a'(\xi_i)$, where the ξ_i, $i = 1, 2, \cdots$, are the eigenfunctions of a positive self-adjoint operator Λ, with totally discrete spectrum and eigenvalues λ_i^2, $\lambda_i \to \infty$ when $i \to \infty$. Similarly, b_i and b_i' are associated with a representation of a complex Bose system over

$$\mathcal{N}^- = \{\omega^{\frac{1}{2}}f, f \in D, \operatorname{supp} f \subset \mathbb{R}^3 - B\}^-.$$

The mapping defined in (20) derives from the following mapping on the set of functions on which $\phi(\cdot)$ and $\pi(\cdot)$ are defined:

$$f \to e^{i\theta}f \quad \text{if} \quad \operatorname{supp} f \subset B,$$
$$f \to f \quad \text{if} \quad \operatorname{supp} f \cap B = \phi; \tag{20'}$$

this induces on $\mathcal{M}$ the transformation $\xi \in \mathcal{M} \to e^{i\theta}\xi \in \mathcal{M}$ (which extends to a unitary transformation of $\mathcal{M}^-$) and on $\mathcal{N}$ (and therefore $\mathcal{N}^-$) the identity transformation.

Therefore, the mapping (20) takes the form

$$a_k \to a_k e^{i\theta}, \quad a_k' \to a_k' e^{-i\theta}, \quad b_i \to b_i, \quad b_i' \to b_i'. \tag{21}$$

It can be proved finally that

$$\alpha_i\Omega_i = \alpha_i'\Omega_i = \beta_i\Omega_i = \beta_i'\Omega_i = 0,$$

where

$$a_i = \alpha_i + \lambda_i(\beta_i - \beta_i^*), \quad b_i = \beta_i - \lambda_i(\alpha_i + \alpha_i'^*)$$

and similarly for a_i' and b_i'.

Since $\sum_i \lambda_i^2 = \infty$, the vector $\otimes_i \Omega_i$ is not charge-discrete relative to $\sum N_{a_i} - \sum N_{a_i'}$, and Proposition 6 (or rather an easy generalization thereof) completes the proof of Proposition 7. A similar line of arguments could be used to prove that in a Fock representation one cannot find a unitary operator inducing the following transformation:

$$\phi^i(f) \to T_j^i\phi^i(f) \quad \operatorname{supp} f \subset B,$$
$$\pi^i(f) \to T_j^i\pi^j(f),$$
$$\phi^i(f) \to \phi^i(f) \quad \operatorname{supp} f \cap B = \phi,$$
$$\pi^i(f) \to \pi^i(f),$$

where T is a unitary matrix, $i, j = 1, 2, \cdots, n$, $N = \infty$ allowed, if (ϕ^i, π^i), $i = 1, 2, \cdots, n$, form a set of n scalar-free relativistic fields and B is an open region in $\mathbb{R}^3$ with piece-wise differentiable boundaries.

B. Smooth Function on $\mathbb{R}^3$

The results of the preceding section are precise statements which go in the following intuitive direction: Even if $j_0(\mathbf{x}, t)$ is a putative charge density,

$$\int j(\mathbf{x}, t)\chi(\mathbf{x})d^3x$$

does not exist as an operator when χ is the characteristic function of the open set B. It would clearly be desirable to make statements in a direction corresponding to χ being any "smooth" function on $\mathbb{R}^3$. We shall now do so, but for this aim the results of Secs. 1 and 2 will no longer be sufficient. Let φ be infinitely many times differentiable and with support in an open region $B \subset \mathbb{R}^3$ (in short, $\varphi \in D_B$).

We shall study the transformation induced on the field by the mapping

$$\tau_\varphi : f \to e^{i\varphi} f \tag{22}$$

of the underlying space (in this case, S). Let $\mathcal{M}$ be defined as before; on $\mathcal{M}$, τ_φ induces the following transformation τ'_φ:

$$\xi \to \omega^{-\frac{1}{2}} e^{i\varphi} \omega^{\frac{1}{2}} \xi \equiv \tau'_\varphi \xi. \tag{23}$$

It can be shown that τ'_φ extends to a bounded operator $C(\varphi)$ on $\mathcal{M}^-$; its adjoint is the closure of

$$P_{\mathcal{M}^-} \omega^{\frac{1}{2}} e^{i\varphi} \omega^{-\frac{1}{2}} P_{\mathcal{M}^-},$$

where $P_{\mathcal{M}^-}$ is the orthogonal projection onto $\mathcal{M}^-$.

Let ξ_i be an orthonormal complete family of eigenfunctions of Λ, with eigenvalues λ_i, and $C_{ij}(\varphi)$ the matrix form of $C(\varphi)$ in this basis. Then, with $\phi_i \equiv \phi(f_i)$ and $\pi_i = \pi(f_i)$, the mapping (22) reads

$$\phi_i \to C_{ij}(\varphi)\phi_j, \quad \pi_i \to d_{ij}(\varphi)\pi_j,$$

where

$$d_{ij}(\varphi) = (\xi_i, C^*(-\varphi)\xi_j)$$

[notice that $C(\varphi_1)C(\varphi_2) = C(\varphi_1 + \varphi_2)$, $C(0) = I$].

The transformation properties of the operators $a_k, \cdots, b'^*_k$ are then

$$\begin{aligned} a^1_i &= \tfrac{1}{2}(c_{ik} + \alpha_{ik})a_k + \tfrac{1}{2}(c_{ik} - \alpha_{ik})a'^*_k \\ &\quad + (\alpha_{ik} - \delta_{ik})\lambda_k(b'^*_k - b_k), \\ b^1_i &= \tfrac{1}{2}(c_{ik} + \alpha_{ik})b_k - \tfrac{1}{2}(c_{ik} - \alpha_{ik})b'^*_k \\ &\quad + \lambda_i(c_{ik} - \delta_{ik})(a'^*_k - a_k), \end{aligned} \tag{24}$$

and similar ones for a'_k and b'_k.

Since $a_i, \cdots, b'^*_i$ provide an irreducible Fock representation of the CCR's, it is known (Ref. 16, Theorem 4.1) that the transformation (24) [and therefore the transformation (22)] can be induced by a unitary operator if and only if the operators $C(\varphi) - D(\varphi)$, $\Lambda(1 - C(\varphi))$, and $(D(\varphi) - 1)\Lambda$ are Hilbert–Schmidt.

It turns out, however, that $\Lambda(1 - C(\varphi))$ and $(1 - D(\varphi))\Lambda$ are not Hilbert–Schmidt for any $\varphi \in S$, supp $\varphi \in B$; we shall sketch the proof for $\Lambda(1 - C(\varphi))$. We shall, in fact, show that $\Lambda(1 - C(\varphi))$ is not bounded (this result does not depend on the number of space dimensions).

To prove that $\Lambda(1 - C(\varphi))$ is not bounded, it suffices to find $g \in C_\infty(B)$, such that $(1 - e^{i\varphi})g \in C_\infty(B)$, $(1 - e^{i\varphi})g \notin D_{\Lambda\omega^{-\frac{1}{2}}}$, where B is any open region which contains the support of φ, has piecewise differentiable boundaries and a complement $B^\perp$ (in $\mathbb{R}^3$) with nonempty interior.

Since $\varphi \in D$, we see that $1 - \cos\varphi$ is measurable, and one can find an open set $B_1 \subset B$ such that $(1 - \cos\varphi) \geq \epsilon$ on B_1. With $B_2 \Subset B_1$, assume that $h \in D(B_2)$. Then g, defined by $g = h(1 - e^{i\varphi})^{-1}$, has support in B_1, is C^∞, and is such that $(1 - e^{i\varphi})g = h$. The proof that $\Lambda(1 - c)$ is not bounded is now completed by showing[14,15] that one can choose $h \in D(B_2)$ so that $h \notin D_{\Lambda\omega^{-\frac{1}{2}}}$.

C. Conclusions

It is worth noting that one can provide alternative proofs of Proposition 8 and even extend it to any $\varphi \in S$ (not necessarily with compact support). In particular, one could notice that Eq. (24) holds, for any $\varphi \in S$, with $\lambda_i \equiv 0$ and C_{ik}, D_{ik} the matrix form, relative to a suitable basis ξ_i, of the closable bilinear forms defined on (a dense set of) $\mathfrak{L}^2(\mathbb{R}^3)$ by the operators $\omega^{-\frac{1}{2}} e^{i\varphi} \omega^{\frac{1}{2}}$ and $\omega^{\frac{1}{2}} e^{i\varphi} \omega^{-\frac{1}{2}}$, respectively. One can then show that no Hilbert–Schmidt operator K exists such that $(\xi_i, K\xi_j) = C_{ij} - D_{ij}$, and use the techniques of Ref. 16 to prove the nonexistence of U_φ. The analysis presented here goes somewhat beyond this result; in particular, it provides [see Eq. (21)], for any given region B, a large class of representations in which the total charge in B is defined. Also, if it can be proved that the operator $C(\varphi) - D(\varphi)$ [see Eqs. (23)] is Hilbert–Schmidt for[17] all $\varphi \in D(B)$, then Eq. (24) selects in a natural way a class of tensor product representations in which the charge density in θ is defined.

It is, of course, beyond the scope of this paper to answer the question whether there exists a representation (not of tensor product type) of a relativistic Bose field in which space–time translations *and* local

gauge transformations are induced by continuous groups of unitary operation.

ACKNOWLEDGMENTS

The author is indebted to M. Reed for stimulating discussions and to C. Kaysen for the hospitality at the Institute for Advanced Study, where the work was done.

* On leave from Istituto di Fisica Teorica, Universita di Napoli.
† Present address: Department of Physics, New York University, New York, New York 10012.

[1] H. Reeh, Fortschr. Physik **16**, 687 (1968).

[2] A similar analysis can be presumably carried out for other "simple" representations, so as to accomodate the case of interacting fields, at least in one-space one-time dimension; but this is beyond the purpose of this paper. This does not imply, of course, that one cannot find representations in which the gauge group is unitarily implemented; indeed, in the Fermi case, one can show that this occurs in a set of representations which is dense, in a suitable sense. What seems difficult to obtain is a representation which has generators both for the gauge group and for the Poincaré group, acting canonically on local fields. This is in contrast with the state of affairs if one gives up locality (as exemplified by nonrelativistic quantized fields).

[3] I. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, R.I., 1963).

[4] Typical choices for μ would be the Lebesque measure on $\mathbb{R}^3$ or the invariant measure on the positive branch of a timelike hyperboloid.

[5] J. von Neumann, Cenpositio Mat. **6**, 1 (1938).

[6] J. Fabrey, Commun. Math. Phys. **19**, 1 (1970).

[7] J. Klauder and J. McKenna, J. Math. Phys. **6**, 68 (1965).

[8] L. Gårding and A. Wightman, Proc. Natl. Acad. Sci. (U.S.) **40**, 617 (1959).

[9] M. Reed, Commun. Math. Phys. **18**, 65 (1970).

[10] J. Kelley and I. Namioka, *Linear Topological Spaces* (Van Nostrand, New York, 1963), p. 86.

[11] A subset Z of a topological space X is residual if the complement of Z in X is the union of a countable family of sets each of which is nowhere dense in X.

[12] For $\lambda_n = 1$, $\forall n$, the choice $\beta_n^k = \frac{1}{2}\delta_{k,0} + \frac{1}{2}\delta_{k,n!}$ provides a representation (nondiscrete) such that U_θ is defined (and unitary) for all rational θ, but is not defined in a complete neighborhood of $\theta = 0$.

[13] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965), pp. 31–37.

[14] H. Araki, J. Math. Phys. **1**, 1 (1969).

[15] G. F. Dell'Antonio, Commun. Math. Phys. **9**, 81 (1968).

[16] D. Shale, Trans. Am. Math. Soc. **103**, 149 (1962).

[17] One can prove that it is compact if the complement of B in $\mathbb{R}^3$ has a nonempty interior.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Bifurcate Space-Times

P. HAJICEK
Institute for Theoretical Physics, University of Berne, Berne, Switzerland

(Received 29 May 1970)

The notion of space-time is generalized in order to include the non-Hausdorff manifolds that do not admit any sort of pathological bifurcate curves. By means of these space-times, a geodesically complete extension of Taub and NUT space is constructed.

I. INTRODUCTION

In an earlier paper,[1] an extension of the tangent bundles of the Taub and NUT spaces has been constructed and shown to contain much more complete geodesics than the Taub–NUT space. A discussion of the paper with R. J. Geroch and S. W. Hawking revealed the following three facts:

(1) Cutting out points from the hypermanifold is not uniquely determined by the condition that the resulting space must be Hausdorff.

(2) The use of the bundle of frames instead of the tangent bundle when constructing the hypermanifold could be more effective in the sense that the condition of Hausdorffness would not then require cutting out any points.

(3) The example of non-Hausdorff extension of the Taub and the NUT spaces shows that there are non-Hausdorff spaces in which no curve has more than one end point. What is the difference between this sort of non-Hausdorff space and that with curves of more continuations?

In the present paper, this last question is going to be answered. By the same tools, the remaining points will be clarified.

2. BIFURCATE SURFACES IN E^3

Suppose we have two square sheets Q_1 and Q_2 of paper with some coordinates on them—for instance, the following:

$$Q_1: \quad -1 < x^1 < 1, \quad -1 < x^2 < 1, \qquad ds^2 = (dx^1)^2 + (dx^2)^2,$$

$$Q_2: \quad -1 < y^1 < 1, \quad -1 < y^2 < 1, \qquad ds^2 = (dy^1)^2 + (dy^2)^2.$$

Now, we glue Q_1 and Q_2 together in some way and wish to describe the construction. One way of doing this is to give the map $\varphi: A \to B$, $A \subset Q_1$ and $B \subset Q_2$, which associates the point $\varphi(p)$ of Q_2 glued on p to each such point p of Q_1. The sets A and B will be called the overlapping sets and the map φ the gluing map; they must have the following properties:

(1) A is open (or else the resulting space would not be a manifold),

(2) φ is a diffeomorphism of appropriate differentiability class of A onto B (or else Q_1 or Q_2 would be torn or the resulting space would not be smooth),

(3) φ is an isometry (or else Q_1 or Q_2 would be folded).

In Fig. 1, two cases of gluing together Q_1 and Q_2 are shown, the characteristics of which are given in Table I. We observe the following:

(1) In case I, every curve has exactly one end point. In case II, the geodesic

$$x^1 = s, \quad x^2 = 0, \quad y^1 = s, \quad y^2 = 0$$

is bifurcate.

(2) In case I, neither the gluing map nor its inverse has any continuous extension in the following sense: There is no connected set $A' \subset Q_1$ ($B' \subset Q_2$) containing a component C_A of A (C_B of B) as a proper subset and a map $\psi: A' \to Q_2$ ($\psi: B' \to Q_1$) such that (a) $\psi|_{C_A} = \varphi$ ($\psi|_{C_B} = \varphi^{-1}$) and (b) ψ is continuous.

On the other hand, in case II, there is such an extension $\psi: Q_1 \to Q_2$ given by $y^1 = x^1$, $y^2 = x^2$.

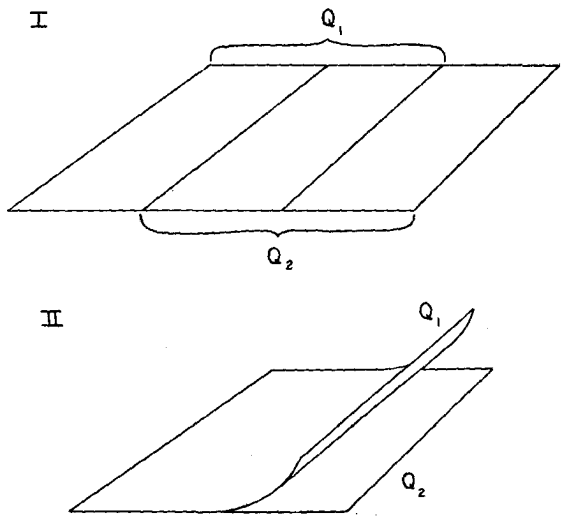


FIG. 1. Two possible cases of gluing together Q_1 and Q_2.

TABLE I. Description of the overlapping sets and the gluing maps corresponding to the two cases in Fig. 1.

Case	Overlapping sets	Gluing maps
I	$A \subset Q_1$: $0 < x^1 < 1, -1 < x^2 < 1$ $B \subset Q_2$: $-1 < y^1 < 0, -1 < y^2 < 1$	φ: $y^1 = x^1 - 1$ $y^2 = x^2$
II	$A \subset Q_1$: $-1 < x^1 < 0, -1 < x^2 < 1$ $B \subset Q_2$: $-1 < y^1 < 0, -1 < y^2 < 1$	φ: $y^1 = x^1$ $y^2 = x^2$

The gluings, such as in case I, will be called continuously inextendable. We shall see that the relation between the continuous extendability and occurrence of the bifurcate geodesics as suggested by the observations (1) and (2) is very simple and general.

3. BIFURCATE SPACE-TIMES

A space-time manifold is a 4-dimensional connected paracompact, not necessarily Hausdorff, manifold of some differentiability class C^k with pseudo-Riemannian metric of signature -2 and differentiability class C^l, $l < k$.

A space-time manifold which is Hausdorff will be simply called a Hausdorff space-time.

Next, we shall be interested in bifurcate curves; there are two kinds of them. Under a bifurcate curve, we understand a pair of curves C, C' in a space-time manifold M, $C: [0, 1] \to M$, $C': [0, 1] \to M$ such that

$$C = C' \quad \text{on} \quad [0, g], \quad C \neq C' \quad \text{on} \quad (g, 1], \qquad 0 < g < 1,$$

for the first kind and

$$C = C' \quad \text{on} \quad [0, g), \quad C \neq C' \quad \text{on} \quad [g, 1], \qquad 0 < g \leq 1,$$

for the second kind.

The first kind of bifurcate curve can be found in every space-time manifold. It is the innocuous kind because it never leads to bifurcate geodesics or other pathologies: The accelerations of the two curves C and C' in the point $C(g) = C'(g)$ must be different from each other (if defined at all). This is not the case for the second kind, which, however, can exist only in non-Hausdorff space-time manifolds, because in every neighborhood of $C(g)$ must lie some points of every neighborhood of $C'(g)$, namely some points of $C([0, g))$. The curves C and C' can already have the same acceleration or other invariant characteristics in every pair of points $C(t)$, $C'(t)$, $t \in [0, 1]$, not leading to any controversy with the uniqueness theorem for the corresponding system of differential equations.

Now, we can state our main definition and theorem.

Definition: A space-time is a space-time manifold which results from gluing together at most countable number of Hausdorff space-times. Each gluing with the gluing map φ and the overlapping sets A and B must have the following properties:

(a) A is open,

(b) φ is a diffeomorphism of class C^k of A onto B,

(c) φ is isometry,

(d) the gluing is continuously inextendable.

Space-times which are non-Hausdorff will be called bifurcate.

Theorem: The necessary and sufficient condition for a manifold constructed by gluing together Hausdorff manifolds to admit bifurcate curves of the second kind is that the gluing be continuously extendable.

Proof: (1) Every space-time manifold is locally Hausdorff because it is locally homeomorph to E^4. Whether the manifold as a whole is Hausdorff or not depends, therefore, entirely on the way in which its charts are glued together. Suppose that the gluing is continuously extendable for some two charts (U_1, h_1) and (U_2, h_2), say; let the overlapping sets be $A_1 \subset U_1$ and $A_2 \subset U_2$ and the gluing map be φ. Assume for the sake of simplicity that A_1 is connected. Then there is some connected set A_1' such that $A_1 \subset A_1' \subset U_1$, $A_1' - A_1 \neq \varnothing$, and the mapping $\psi: A_1' \to U_2$, $\psi|_{A_1} = \varphi$ is continuous. Choose a point $p \in \dot{A}_1 \cap A_1'$ where the dot denotes the topological boundary. As A_1 is open, a timelike or spacelike curve C always exists with one end point in p such that $\{C\} - p \subset A_1$. The second curve C' given by $C' = \psi \circ C$ is well defined because ψ is continuous. But then the pair C, C' is a curve with two end points, for $p \neq \psi(p)$, and this is a special case of a bifurcate curve of the second kind.

(2) Suppose that there is a bifurcate curve of the second kind in M. That is to say, we have a pair of curves C, $C': [0, 1] \to M$, identical on $[0, g)$ and different on $[g, 1]$. This is only possible if C lies in some (U_1, h_1), C' lies in some (U_2, h_2), and if both charts are glued together along some open sets A_1 and A_2 such that $C([0, g)) \subset A_1$, $C'([0, g)) \subset A_2$, $C([g, 1]) \cap A_1 = \varnothing$, and $C'([g, 1]) \cap A_2 = \varnothing$. Let the corresponding gluing map be $\varphi: A_1 \to A_2$. Then

we can construct a map $\psi:(A_1 \cup \{C\}) \to (A_2 \cup \{C'\})$ in the following way: $\psi|_{A_1} = \varphi$, $\psi|_{\{C\}} = C' \circ C^{-1}$. The set $A_1 \cup \{C\}$ is connected because A_1 and $\{C\}$ are and $A_1 \cap \{C\} \neq \varnothing$. The map ψ is well defined because $\varphi|_{A_1 \cap \{C\}} = C' \circ C^{-1}|_{A_1 \cap \{C\}}$, C being one-to-one and continuous because C^{-1}, C', and φ are. Thus, the gluing is continuously extendable. QED

4. MAXIMAL ANALYTIC EXTENSION OF THE TAUB–NUT SPACE

What remains to be shown is that the class of the bifurcate space-times is nonempty. This will be done by means of an example.

In Refs. 2 and 1, the extensions T_1, T_2, P_1, P_2, and P_3 of the Taub and NUT spaces were described. We repeat some information on them in appropriate coordinates.

T_1: topology $R \times S^3$,

coordinates $-\infty < z_1 < \infty$, $0 \le \xi_1 < 4\pi$,

$0 \le \theta_1 \le \pi$, $0 \le \varphi_1 < 2\pi$,

$$ds^2 = -(2l)^2[2(d\xi_1 + \cos\theta_1\, d\varphi_1)\, dz_1 + U(z_1)(d\xi_1 + \cos\theta_1\, d\varphi_1)^2 + \tfrac{1}{4}(4z_1^2 + 1)(d\theta_1^2 + \sin^2\theta_1\, d\varphi^2)];$$

T_2: topology $R \times S^3$,

coordinates $-\infty < z_2 < \infty$, $0 \le \xi_2 < 4\pi$,

$0 \le \theta_2 \le \pi$, $0 \le \varphi_2 < 2\pi$,

$$ds^2 = -(2l)^2[-2(d\xi_2 + \cos\theta_2\, d\varphi_2)\, dz_2 + U(z_2)(d\xi_2 + \cos\theta_2\, d\varphi_2)^2 + \tfrac{1}{4}(4z_2^2 + 1)(d\theta_2^2 + \sin^2\theta_2\, d\varphi^2)];$$

P_1: topology $[Z_2, \infty) \times S^3$,

coordinates $Z_2 \le u_1 < \infty$, $0 \le \xi_1 < 4\pi$,

$0 \le \eta_1 \le \pi$, $0 \le \psi_1 < 2\pi$,

$$ds^2 = (2l)^2[U^{-1}(u_1)\, du_1^2 - U(u_1)(d\xi_1 + \cos\eta_1\, d\psi_1)^2 - \tfrac{1}{4}(4u_1^2 + 1)(d\eta_1^2 + \sin^2\eta_1\, d\psi_1^2)];$$

P_2: topology $[Z_1, Z_2] \times S^3$,

coordinates $Z_1 \le u_2 \le Z_2$, $0 \le \xi_2 < 4\pi$,

$0 \le \eta_2 \le \pi$, $0 \le \psi_2 < 2\pi$,

$$ds^2 = (2l)^2[U^{-1}(u_2)\, du_2^2 - U(u_2)(d\xi_2 + \cos\eta_2\, d\psi_2)^2 - \tfrac{1}{4}(4u_2^2 + 1)(d\eta_2^2 + \sin^2\eta_2\, d\psi_2^2)];$$

P_3: topology $(-\infty, Z_1] \times S^3$,

coordinates $-\infty < u_3 \le Z_1$, $0 \le \xi_3 < 4\pi$,

$0 \le \eta_3 \le \pi$, $0 \le \psi_3 < 2\pi$,

$$ds^2 = (2l)^2[U^{-1}(u_3)\, du_3^2 - U(u_3)(d\xi_3 + \cos\eta_3\, d\psi_3)^2 - \tfrac{1}{4}(4u_3 + 1)(d\eta_3^2 + \sin^2\eta_3\, d\psi_3^2)].$$

Here $l \neq 0$ is a parameter, $Z_1 < Z_2$ are real numbers, and

$$U(x) = \frac{4(x - Z_1)(Z_2 - x)}{4x^2 + 1}.$$

Now, we can glue together all the spaces into a bifurcate space-time by means of the following gluing maps:

$$z_2 = z_1, \quad \theta_2 = \theta_1, \quad \varphi_2 = \varphi_1,$$

$z_1 < Z_1$: $\xi_2 = \xi_1 - 2z_1 + \tfrac{1}{2}(Z_2 - Z_1)^{-1}[-(4Z_2^2 + 1)\log(Z_2 - z_1) - (4Z_1^2 + 1)\log(Z_1 - z_1)]$,

$Z_1 < z_1 < Z_2$: $\xi_2 = \xi_1 - 2z_1 + \tfrac{1}{2}(Z_2 - Z_1)^{-1}[-(4Z_2^2 + 1)\log(Z_2 - z_1) + (4Z_1^2 + 1)\log(z_1 - Z_1)]$,

$Z_2 < z_1$: $\xi_2 = \xi_1 - 2z_1 + \tfrac{1}{2}(Z_2 - Z_1)^{-1}[(4Z_2^2 + 1)\log(z_1 - Z_2) + (4Z_1^2 + 1)\log(z_1 - Z_1)]$,

for the space-times T_1 and T_2, and

$Z_2 < z_1$: $u_1 = z_1$, $\eta_1 = \theta_1$, $\psi_1 = \varphi_1$,

$\xi_1 = \xi_1 - z_1 + \tfrac{1}{4}(Z_2 - Z_1)^{-1}[(4Z_2^2 + 1)\log(z_1 - Z_2) + (4Z_1^2 + 1)\log(z_1 - Z_1)]$,

$Z_1 < z_1 < Z_2$: $u_2 = z_1$, $\eta_2 = \theta_1$, $\psi_2 = \varphi_1$,

$\xi_2 = \xi_1 - z_1 + \tfrac{1}{4}(Z_2 - Z_1)^{-1}[-(4Z_2^2 + 1)\log(Z_2 - z_1) + (4Z_1^2 + 1)\log(z_1 - Z_1)]$,

$z_1 < Z_1$: $u_3 = z_1$, $\eta_3 = \theta_1$, $\psi_3 = \varphi_1$,

$\xi_3 = \xi_1 - z_1 + \tfrac{1}{4}(Z_2 - Z_1)^{-1}[-(4Z_2^2 + 1)\log(Z_2 - z_1) - (4Z_1^2 + 1)\log(Z_1 - z_1)]$,

for the space-times T_1 and P_1, T_1 and P_2, and T_1 and P_3, respectively.

This description determines the space which we wanted to construct sufficiently, the remaining gluing maps being superpositions of the given ones. The obtained non-Hausdorff manifold T is a bifurcate space-time because none of the gluing maps are extendable: They get singular on the boundaries of the overlapping sets. (See also Ref. 3.)

This example is of interest for itself. In fact, we

have arrived at a geodesically complete bifurcate space–time which extends the Taub space. The kind of extension is, of course, not quite clear because of the strange nature of the spaces P_1, P_2, and P_3 (see Ref. 1).

It is not difficult to show by the method used already in Ref. 1 that the bundle of frames of T is a Hausdorff space in accordance with the suggestion of Hawking as mentioned in the first section. Thus, the Taub–NUT hypermanifold constructed by means of bundles of *frames* is a *bundle* of frames, and in this aspect it has a simpler structure than that constructed by means of tangent bundle in Ref. 1. The definition of bifurcate space–times renders it possible to work directly with the space–time instead with some of its bundles. But the bundle of frames of T could remain interesting in connection with another question: What behavior will the metric due to Schmidt show on it?[4] The question is nontrivial and may contribute to the understanding of the spaces P_i.

Another observation is that almost each of the spacelike 3-dimensional hypersurfaces $z_1 = u_2 = z_2 = Z$, $Z_1 < Z < Z_2$, is compact, and the normals to it converge everywhere. In this way, the bifurcate space–times need not satisfy Theorem 1 of Hawking's paper.[5] (This is, of course, strongly conditioned by whether P_i are regarded as regular or not.) That is to say, the bifurcate space–times could be, in a sense, less singular than the Hausdorff ones.

ACKNOWLEDGMENTS

It is a pleasure to express the author's gratitude to Dr. R. Geroch and Dr. S. W. Hawking for their stimulating discussion. Special thanks go to Professor Dr. A. Mercier, who makes possible and encourages this research at his Institute.

[1] P. Hajicek, Commun. Math. Phys. **17**, 109 (1970).
[2] C. W. Misner and A. H. Taub, Zh. Eksp. Teor. Fiz. **55**, 233 (1968) [Sov. Phys.—JETP **28**, 122 (1969)].
[3] R. Geroch, J. Math. Phys. **9**, 450 (1968).
[4] B. G. Schmidt, Seminar on Bearings of Topology upon General Relativity, Gwatt at the Lake of Thoune, Switzerland, May, 1970.
[5] S. W. Hawking, Proc. Roy. Soc. (London) **A300**, 187 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Path Integrals and Product Integrals*

JOHN F. HAMILTON, JR.
Mathematics Department, Indiana University, Bloomington, Indiana 47401
AND
L. S. SCHULMAN†
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 9 June 1970)

The product integral is shown to be the appropriate mathematical tool for implementing the path integral for spin in its simplest form.

INTRODUCTION

Various proposals have been made for accommodating spin into the framework of Feynman's path integral formulation of quantum mechanics.[1–5] However, in our opinion, none of these theories is the sort of thing one would use to solve easy problems. From the standpoint of economy, the simplest path integral for spin should have no more new physical ideas than are necessary to go from the nonrelativistic Schrödinger equation without spin to that with spin.

The way to include spin was clear to Feynman at least as far back as his first paper on path integrals.[6] His suggestion was to put spinors in the Lagrangian and then, when path integrating, pay attention to the ordering of operators. The same formal prescription can also be applied to the Dirac equation.[7] However, the presumed complexity of the mathematical operations has prevented implementation of these ideas.[8]

In fact, work on the mathematical techniques necessary for this implementation first began in 1887. Volterra[9] studied infinite products of matrices, and, subsequently, Schlesinger,[10] Rasch,[11] and Masani[12] developed the theory of this "product integral." As we shall see, many of its properties are appropriate for physical application. The motivation for the study of the product integral was systems of coupled first-order linear equations. The nonrelativistic Schrödinger equation with spin is obviously an example of such a system. When spin and orbital motion are coupled, however, the unboundedness of the operators involved precludes fully rigorous statements (in this article anyway).

have arrived at a geodesically complete bifurcate space–time which extends the Taub space. The kind of extension is, of course, not quite clear because of the strange nature of the spaces P_1, P_2, and P_3 (see Ref. 1).

It is not difficult to show by the method used already in Ref. 1 that the bundle of frames of T is a Hausdorff space in accordance with the suggestion of Hawking as mentioned in the first section. Thus, the Taub–NUT hypermanifold constructed by means of bundles of *frames* is a *bundle* of frames, and in this aspect it has a simpler structure than that constructed by means of tangent bundle in Ref. 1. The definition of bifurcate space–times renders it possible to work directly with the space–time instead with some of its bundles. But the bundle of frames of T could remain interesting in connection with another question: What behavior will the metric due to Schmidt show on it?[4] The question is nontrivial and may contribute to the understanding of the spaces P_i.

Another observation is that almost each of the spacelike 3-dimensional hypersurfaces $z_1 = u_2 = z_2 = Z$, $Z_1 < Z < Z_2$, is compact, and the normals to it converge everywhere. In this way, the bifurcate space–times need not satisfy Theorem 1 of Hawking's paper.[5] (This is, of course, strongly conditioned by whether P_i are regarded as regular or not.) That is to say, the bifurcate space–times could be, in a sense, less singular than the Hausdorff ones.

ACKNOWLEDGMENTS

It is a pleasure to express the author's gratitude to Dr. R. Geroch and Dr. S. W. Hawking for their stimulating discussion. Special thanks go to Professor Dr. A. Mercier, who makes possible and encourages this research at his Institute.

[1] P. Hajicek, Commun. Math. Phys. **17**, 109 (1970).
[2] C. W. Misner and A. H. Taub, Zh. Eksp. Teor. Fiz. **55**, 233 (1968) [Sov. Phys.—JETP **28**, 122 (1969)].
[3] R. Geroch, J. Math. Phys. **9**, 450 (1968).
[4] B. G. Schmidt, Seminar on Bearings of Topology upon General Relativity, Gwatt at the Lake of Thoune, Switzerland, May, 1970.
[5] S. W. Hawking, Proc. Roy. Soc. (London) **A300**, 187 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Path Integrals and Product Integrals*

JOHN F. HAMILTON, JR.
Mathematics Department, Indiana University, Bloomington, Indiana 47401
AND
L. S. SCHULMAN†
Physics Department, Indiana University, Bloomington, Indiana 47401

(Received 9 June 1970)

The product integral is shown to be the appropriate mathematical tool for implementing the path integral for spin in its simplest form.

INTRODUCTION

Various proposals have been made for accommodating spin into the framework of Feynman's path integral formulation of quantum mechanics.[1–5] However, in our opinion, none of these theories is the sort of thing one would use to solve easy problems. From the standpoint of economy, the simplest path integral for spin should have no more new physical ideas than are necessary to go from the nonrelativistic Schrödinger equation without spin to that with spin.

The way to include spin was clear to Feynman at least as far back as his first paper on path integrals.[6] His suggestion was to put spinors in the Lagrangian and then, when path integrating, pay attention to the ordering of operators. The same formal prescription can also be applied to the Dirac equation.[7] However, the presumed complexity of the mathematical operations has prevented implementation of these ideas.[8]

In fact, work on the mathematical techniques necessary for this implementation first began in 1887. Volterra[9] studied infinite products of matrices, and, subsequently, Schlesinger,[10] Rasch,[11] and Masani[12] developed the theory of this "product integral." As we shall see, many of its properties are appropriate for physical application. The motivation for the study of the product integral was systems of coupled first-order linear equations. The nonrelativistic Schrödinger equation with spin is obviously an example of such a system. When spin and orbital motion are coupled, however, the unboundedness of the operators involved precludes fully rigorous statements (in this article anyway).

In the next section we present the formal path integral for spin referred to above. This will introduce the theory of the product integral. We will then show just how neatly the product integral specializes to spin, and give a worked example, intending to demonstrate thereby that in its present form the path integral can give a simple answer to a simple question. The final section is a discussion.

FORMAL PATH INTEGRAL FOR SPIN

The Green's function for the Schrödinger equation satisfies

$$\left(H - \frac{i\partial}{\partial t}\right) G(t, t') = -i\delta(t - t') \tag{1}$$

so that, if H is independent of t, $G(t, t') = \theta(t - t') \times \exp[-iH(t - t')]$. If $\Psi(t)$ is a solution of the Schrödinger equation with initial (t') condition $\Psi(t')$, then $\Psi(t) = G(t, t')\Psi(t')$. For $t \geq t' \geq t''$ we have

$$G(t, t'') = G(t, t')G(t', t''). \tag{2}$$

The basic fact on which the theory of path integration rests is that, for sufficiently small $t - t'$, the coordinate space matrix elements of $G(t, t')$ can be well approximated by

$$G(\mathbf{x}, t; \mathbf{x}', t') = \left(\frac{m}{2\pi i(t - t')}\right)^{\frac{1}{2}} \exp\left(\frac{im(\mathbf{x} - \mathbf{x}')^2}{2(t - t')} - iV(\mathbf{x})(t - t')\right), \tag{3}$$

where $\Psi(\cdot, t) \in L_2(R^3)$ and $H = -(\Delta/2m) + V(\mathbf{x})$. A given time interval $[t', t]$ is broken into subintervals $[t', t_1]$, $[t_1, t_2], \cdots, [t_{N-2}, t_{N-1}]$, $[t_{N-1}, t]$, and on each of these G is approximated by (3). Equation (2) is then used repeatedly and advantage taken of the fact that the product of exponentials is the exponential of the sum to obtain the formula

$$G(\mathbf{x}, t; \mathbf{x}', t') = \int \cdots \int d^3x_1 \cdots d^3x_{N-1} \left(\frac{m}{2\pi i\epsilon}\right)^{N/2} \times \exp\left[i \sum_{j=0}^{N-1} \left(\frac{m}{2\epsilon}(\mathbf{x}_{j+1} - \mathbf{x}_j)^2 - V(\mathbf{x}_{j+1})\epsilon\right)\right], \tag{4}$$

where $t' \equiv t_0$, $\mathbf{x}' \equiv \mathbf{x}_0$, $t \equiv t_N$, $\mathbf{x} \equiv \mathbf{x}_N$, $\epsilon = (t - t')/N$. The sum in the exponent is an approximating sum for the action evaluated along the broken line path through $\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_N$. The action is

$$S[x(\cdot)] = \int_{t'}^{t} L[x(\tau), \dot{x}(\tau)]\, d\tau$$

and $L = \frac{1}{2}m\dot{\mathbf{x}}^2 - V(\mathbf{x})$. By integrating over $\mathbf{x}_1, \cdots, \mathbf{x}_{N-1}$, we get a sum over (these broken line) paths of the quantity $Ae^{iS/\hbar}$, with A the constant given above. The finite-time Green's function is then the limit of (4) as $N \to \infty$. This sum over paths is sometimes written as

$$G = \int \mathfrak{D}x(t) \exp\{iS[x(t)]\}. \tag{5}$$

In the presence of a magnetic field a term $e\dot{\mathbf{x}} \cdot \mathbf{A}(\mathbf{x})$ is added to the Lagrangian L. However, when we approximate G as in (3) or (4), it is not adequate to evaluate $\mathbf{A}$ at $\mathbf{x}_j$ or $\mathbf{x}_{j+1}$ as for $V(\mathbf{x})$, but rather a combination such as the following must be taken:

$$e(\mathbf{x}_{j+1} - \mathbf{x}_j) \cdot \tfrac{1}{2}[\mathbf{A}(\mathbf{x}_j) + \mathbf{A}(\mathbf{x}_{j+1})]. \tag{6}$$

For spin-$\frac{1}{2}$ particles the wavefunction is a 2-component spinor, and the Hamiltonian is

$$H = (2m)^{-1}(\mathbf{p} - e\mathbf{A})^2 + V + \gamma\boldsymbol{\sigma} \cdot \mathbf{B}, \tag{7}$$

where γ is a constant related to the magnetic moment. Feynman observed that, by replacing (6) by

$$\tfrac{1}{2}e[(\mathbf{x}_{j+1} - \mathbf{x}_j) \cdot \boldsymbol{\sigma}][\mathbf{A}(\mathbf{x}_j) \cdot \boldsymbol{\sigma}] + \tfrac{1}{2}e[\mathbf{A}(\mathbf{x}_{j+1}) \cdot \boldsymbol{\sigma}][(\mathbf{x}_{j+1} - \mathbf{x}_j) \cdot \boldsymbol{\sigma}], \tag{8}$$

a particular value of γ is obtained.[6,13] However, it is not necessary to be so dramatic, and in addition to (6) one can simply add a term $i\gamma\boldsymbol{\sigma} \cdot \mathbf{B}(t)(t - t')$ to the argument of the exponential in (3).

Now, however, the product of exponentials need not be the exponential of the sum, and the noncommutativity of the spin matrices requires that the iterated Green's function remain a product. For example, where $\mathbf{B}$ is independent of $\mathbf{x}$, the spatial wavefunction can be ignored and the spin Green's function can be written

$$G(t, 0) = \prod_{n=0}^{N-1} \exp[-i\gamma\mathbf{B}(n\epsilon) \cdot \boldsymbol{\sigma}\epsilon] \tag{9}$$

with factors of lower n to the right of those with higher n.

This much was explicit or implicit in Feynman's original work.

PRODUCT INTEGRAL

The concept of the product integral can, in general, be developed for L_1 functions from a linearly ordered measure space to a Banach algebra. But for the purposes of this paper it is sufficient to consider the Riemann theory of product integrals as developed by Schlesinger,[14] who deals with functions from $[a, b] \subset R$ to the algebra of $n \times n$ matrices ($\mathfrak{X}$).

The Riemann product integral of a matrix-valued

function f is defined by

$$\int_a^{\breve{b}} \exp\,[f(x)\,dx]$$
$$= \lim_{\pi\downarrow} \prod_{k=1}^{n} \exp\,[f(t_k)\cdot|\Delta_k|]$$
$$= \lim_{\pi\downarrow} \{\exp\,[f(t_n)\cdot|\Delta_n|]\cdots\exp\,[f(t_1)\cdot|\Delta_1|]\}, \quad (10)$$

where π is a partition of $[a, b]$ into subintervals Δ_k and $t_k \in \Delta_k$.[15] Thus, the product integral of a matrix-valued function is itself matrix valued. Finally, one can make the product integral a function from R to $\mathfrak{X}$. Let

$$\breve{F}(t, a) = \int_a^{\breve{t}} \exp\,[f(x)\,dx] = \int_a^{\breve{b}} \exp\,[f(x)\cdot\chi_{[a,t]}\,dx], \quad (11)$$

where $\chi_{[a,t]}$ is the characteristic function of the interval $[a, t]$. It can be shown that this satisfies

$$\breve{F}(t, a) = 1 + \int_a^t f(x)\breve{F}(x, a)\,dx, \quad (12)$$

$$\breve{F}'(t, a) = f(t)\breve{F}(t, a), \quad (13)$$

where $f(x)$, $\breve{F}(x, a) \in \mathfrak{X}$. Equation (12) can be iterated to yield

$$\breve{F}(t, a) = 1 + \int_a^t f(t_1)\,dt_1 + \int_a^t f(t_1)\int_a^{t_1} f(t_2)\breve{F}(t_2, a)\,dt_2\,dt_1.$$

This scheme of substitution leads to the Peano series

$$\breve{F}(t, a) = 1 + \int_a^t f(t_1)\,dt_1 + \int_a^t f(t_1)\int_a^{t_1} f(t_2)\,dt_2\,dt_1 + \cdots. \quad (14)$$

Another important equation involves the "integration by parts" [12] identity for product integrals:

$$\int_a^{\breve{t}} \exp\,\{[f(x) + g(x)]\,dx\}$$
$$= \int_a^{\breve{t}} \exp\,[f(x)\,dx]\cdot\int_a^{\breve{t}} \exp\left[\left(\int_a^{\breve{x}} \exp\,[f(s)\,ds]\right)^{-1} \times g(x)\left(\int_a^{\breve{x}} \exp\,[f(s)\,ds]\right)dx\right]. \quad (15)$$

It is illustrative to look at the commutative case for the equations given above:

(a) $\breve{F}(t, a) = \int_a^{\breve{t}} \exp\,[f(x)\,dx] = \exp\,[\int_a^t f(x)\,ds]$.

(b) The Peano series yields the series expansion for $\exp\,[\int_a^t f(x)\,dx]$.

(c) Equation (15) simply verifies that the exponential rule now works, i.e., that

$$\exp\,(A + B) = \exp\,(A)\exp\,(B).$$

The resemblance of the product integral, as defined in Eq. (10), to the time-ordered product is not at all accidental and reflects their common use in the description of evolution. Formulas with a great similarity to those we are using can be found in Ref. 16, Appendix H, as can references to other work along these lines.

SUM OVER HISTORIES IN SPIN SPACE

While it is obvious that the path integral can be thought of as a kind of product integral, we emphasize that we are not making this identification for the space coordinates of the system. It would be very nice if for the $f(x)$ of Eq. (10) we could substitute $-iH$, where H is the usual Hamiltonian for the Schrödinger equation with or without spin. That is,

$$G(t, t') = \int_{t'}^{\breve{t}} \exp\,[-iH(\tau)\,d\tau]. \quad (16)$$

Certainly the "equation" (16) suggests that product integration may be useful in this context, and Kato[17] has employed similar techniques in studies of an evolution operator. However, the integration by parts formula [Eq. (15)] does not seem to have been established from this point of view. On the other hand, this formula is certainly not new to physics, and one may easily observe that it is formally identical with the expression for the propagator obtained from the interaction picture.[18]

The question to which we can rigorously address ourselves, however, involves a particle with spin in a spatially homogeneous magnetic field. This allows the separation of space and spin coordinates, and we discuss the latter.

The Green's function for the spin coordinates, given in Eq. (9), can be written

$$G(t, 0) = \int_0^{\breve{t}} \exp\,[-i\gamma\boldsymbol{\sigma}\cdot\mathbf{B}(\tau)\,d\tau]. \quad (17)$$

To interpret this as a sum over paths, we include the indices for the 2 × 2 matrices and indicate the limit:

$$G(t, 0)_{\alpha\alpha'} = \lim_{N\to\infty} \sum_{\{\alpha\}} \{\exp\,[-i\gamma\boldsymbol{\sigma}\cdot\mathbf{B}(t-\epsilon)\epsilon]\}_{\alpha\alpha_{N-1}} \times \{\exp\,[-i\gamma\boldsymbol{\sigma}\cdot\mathbf{B}(t-2\epsilon)\epsilon]\}_{\alpha_{N-1}\alpha_{N-2}} \times \cdots \times \{\exp\,[-i\gamma\boldsymbol{\sigma}\cdot\mathbf{B}(0)\epsilon]\}_{\alpha_1\alpha'}, \quad (18)$$

where $\epsilon = t/N$ and the sum is over $\alpha_1 = \uparrow, \downarrow$, $\alpha_2 = \uparrow, \downarrow, \cdots, \alpha_{N-1} = \uparrow, \downarrow$. If we imagine that this acts on

some initial state, then the meaning is clear. Each factor in each summand of (18) is an amplitude for going from one to another (possibly the same) position, and each summand is the amplitude for a specific history of possibilities. As $\epsilon \to 0$, the individual factors approach unity, but there is still some amplitude for going from one of the discrete states to the other.

Each "path" for this path integral is thus a sequence whose entries are the symbols $\uparrow$ and $\downarrow$. When the mesh in t is fine enough to neglect the time dependence of $\mathbf{B}(t)$, successive factors in each of the products in Eq. (18) commute with one another, and the integral over these stretches can be performed. If one nevertheless wishes to evaluate the sum over sequences of $\uparrow$, $\downarrow$ for constant $\mathbf{B}$, the problem is of an essentially combinatorial character and bears some resemblance to the "relativistic 1-dimensional particle" path integral described by Feynman and Hibbs.[19] In summing over paths in the discrete spin space, it turns out that, as the angle θ of magnetic field with the z axis approaches 0 or π, the most important contributions to the sum involve paths with the smallest number of switches ($\uparrow$ to $\downarrow$) or ($\downarrow$ to $\uparrow$) and conversely for θ near $\frac{1}{2}\pi$. This suggests a kind of classical interpretation, and in Footnote 20 we shall indicate how the classical top itself can be cast as a product integral.

For magnetic fields which are not uniform, space and spin coordinates do not separate, and one is essentially back to Feynman's formal prescription. If the sum over spin histories is performed first, one obtains

$$G(\mathbf{x}, t; \mathbf{x}', t') = \int \cdots \int d^3x_1 \cdots d^3x_{N-1} \left(\frac{m}{2\pi i\epsilon}\right)^{N/2} \times \exp\left(i \sum_{j=0}^{N-1} S_{j+1,j}\right) \int_{t'}^{\breve{t}} \exp\left[-i\gamma \boldsymbol{\sigma} \cdot \mathbf{B}(\tilde{\mathbf{x}}, t)\, dt\right],$$

where $S_{j+1,j}$ is the action evaluated along the line from $\mathbf{x}_j$ to $\mathbf{x}_{j+1}$ (in the time ϵ). The $\tilde{\mathbf{x}}$ appearing in $\mathbf{B}(\tilde{\mathbf{x}}, t)$ is the broken line connecting $\mathbf{x}'\mathbf{x}_1 \cdots \mathbf{x}_{N-1}\mathbf{x}$. The (spin) product integral can be evaluated (in principle) along this path and then the integral over $\mathbf{x}_1 \cdots \mathbf{x}_{N-1}$ performed.

One circumstance under which this may not be an entirely impractical procedure is in the semiclassical limit, where the most important contribution to the sum over paths is from the "classical path" (that obeying the classical equations of motion). In that case $\mathbf{B}[\mathbf{x}(t), t]$ would, in effect, be some definite function of t.

A WORKED EXAMPLE

We illustrate our claim that simple problems have simple solutions with the example of magnetic resonance.

A spin-$\frac{1}{2}$ particle is in the following time-dependent, spatially homogeneous magnetic field:

$$\mathbf{B}(t) = \hat{z} + \lambda(\hat{x} \cos \omega t - \hat{y} \sin \omega t). \tag{19}$$

The Green's function of Eq. (17) is computed by breaking up the integrand as in the "integration by parts" formula (15). To the physicist, particularly one who has solved this problem by other means,[21] this formula will be seen to be the transition to the interaction picture. We break up the integrand with the following identifications:

$$f = \tfrac{1}{2}i\omega\sigma_z,$$
$$g = -i\sigma_z(\gamma + \tfrac{1}{2}\omega) - i\gamma\lambda(\sigma_x \cos \omega t - \sigma_y \sin \omega t). \tag{20}$$

It may be verified that

$$\int_0^\tau \exp\left[f(x)\, dx\right] = \exp\left(-\tfrac{1}{2}i\omega\sigma_z\tau\right),$$

so that this term appearing in the second factor of the right-hand side of Eq. (15) essentially brings the rotating field to rest. Performing the remaining trivial integrations leads to

$$G(t, 0) = \exp\left(-i\omega\sigma_z t\right) \times \exp\left\{-i[\sigma_z(\gamma + \tfrac{1}{2}\omega) + \sigma_x\gamma\lambda]t\right\}. \tag{21}$$

The motion is in effect compounded of two rotations.[20]

In a magnetic resonance experiment, one would be most likely to encounter a linearly polarized field. The argument of the exponential would therefore contain an additional term

$$h = -i\gamma\lambda(\sigma_x \cos \omega t + \sigma_y \sin \omega t), \tag{22}$$

which would at first be included with the g of Eq. (20). After the transformation by f, this term would become

$$-i\gamma\lambda(\sigma_x \cos 2\omega t + \sigma_y \sin 2\omega t); \tag{23}$$

it would rotate at *twice* the frequency. The expression (23) would then be separated from the transformed g by another integration by parts to yield

$$G(t, 0) = G^{(0)}(t, 0) \int_0^{\breve{t}} \exp\left\{\exp\left(i\boldsymbol{\sigma} \cdot \tfrac{1}{2}\boldsymbol{\Omega}\tau\right) \times \left[-i\gamma\lambda(\sigma_x \cos 2\omega t + \sigma_y \sin 2\omega t)\right] \times \exp\left(-i\boldsymbol{\sigma} \cdot \tfrac{1}{2}\boldsymbol{\Omega}\tau\right) d\tau\right\}, \tag{24}$$

where $G^{(0)}(t, 0)$ is the expression given in (21) and $\mathbf{\Omega} = \hat{z}(\gamma + \frac{1}{2}\omega) + \hat{x}\gamma\lambda$. It is reasonable to expand the right-hand factor of (24) via the Peano series [Eq. (14)], which is just perturbation theory.

The foregoing manipulations are effectively identical to those described by Gottfried.[21] There is no new physics in this approach, and our only intention is to show the manageability, indeed convenience, of this formulation.

DISCUSSION

The product integral is a natural tool for many problems in physics because it is tailor made for processes in which there is some sort of evolution. Indeed, Arley[22] used this technique to study cascades induced by cosmic rays, and, while at this time mathematical difficulties prevent full exploitation of the obvious relation between path integration and product integration, it is nevertheless possible, as we have shown in this paper, to have a quite reasonable path integral for spin in terms of the product integral.

There are one or two obvious points to be made. First, although our calculations involved spin $\frac{1}{2}$, any other spin would have been just as good. Second, problems involving transitions among a finite number of levels could be handled by product integration if one supplied an effective Hamiltonian for the mutual interaction.

ACKNOWLEDGMENT

We wish to thank Professor P. Masani for his helpful suggestions.

* Supported in part by the National Science Foundation and U.S. Army Research Office (Durham).

† Address during academic year 1970–71: Physics Department, Technion-Israel Institute of Technology, Haifa, Israel.

[1] Y. Nambu, Progr. Theoret. Phys. (Kyoto) **7**, 131 (1952); see particularly p. 158.
[2] P. T. Matthews and A. Salam, Nuovo Cimento **2**, 120 (1955).
[3] W. Tobocman, Nuovo Cimento **3**, 1213 (1956).
[4] J. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960).
[5] L. S. Schulman, Phys. Rev. **176**, 1558 (1968).
[6] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[7] C. Morette (DeWitt), Phys. Rev. **81**, 848 (1951).
[8] R. P. Feynman, Phys. Rev. **84**, 108 (1951), Appendix D.
[9] V. Volterra, Rend. Accad. Lincei **3**, 393 (1887).
[10] L. Schlesinger, Math. Z. **33**, 33 (1931); **35**, 485 (1932).
[11] G. Rasch, J. Reine Angew. Math. **171**, 65 (1934).
[12] P. R. Masani, Trans. Am. Math. Soc. **61**, 147 (1947).
[13] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967), p. 78.
[14] L. Schlesinger, Math. Z. **33**, 33 (1931).
[15] Although Schlesinger uses

$$\lim_{\pi} \prod_{k=1}^{n} [I + f(\tau_k) \cdot |\Delta_k|]$$

in his definition (I is the identity matrix), he proves the definitions to be equivalent.
[16] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969). Rosen's exponentiated integral chronologically ordered, $T(\exp \int_a^b f(t)\, dt)$, in our notation is $\breve{\int}_a^b \exp [f(t)\, dt]$.
[17] T. Kato, J. Math. Soc. Japan **5**, 208 (1953); see, in particular, Theorem 2 and proof.
[18] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1962), Chap. 17.
[19] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), problem 2-6, p. 34.
[20] It is amusing that we have here a solution for the motion of a symmetric top. If the top position is given as a point $U(t)$ on the group manifold of $SU(2)$ (as in Ref. 5), then the angular velocity $\mathbf{\Omega}(t)$ can be given by $U(t + \epsilon) = \exp(-i\mathbf{\Omega}(t) \cdot \boldsymbol{\sigma}\epsilon/2)U(t)$ as $\epsilon \to 0$. Obviously, the position of the top can be written

$$U(t) = \left(\breve{\int}_0^t \exp\left(-\tfrac{1}{2}\mathbf{\Omega}(\tau) \cdot \boldsymbol{\sigma} d\tau\right) \right) U(0).$$

For the symmetrical top it is known that the angular velocity has a sinusoidal dependence, as in Eq. (19), so that (21) is the required propagation matrix. The significance of the two rotations is also clear.
[21] K.Gottfried, *Quantum Mechanics* (Benjamin, New York, 1966), Vol. I, p. 427.
[22] N. Arley, *On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation* (Wiley, New York, 1943).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Completely Orthogonalized Plane Waves

M. D. GIRARDEAU
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon

(Received 3 August 1968)*

Orthogonalized plane waves (OPW's) are orthogonal to the core states by construction, but they are not mutually orthogonal. Furthermore, the set of all core states plus all OPW's is overcomplete. In this paper we construct a set of "completely orthogonalized plane waves" (COPW's) which share with OPW's the properties of being orthogonal to the core states and simply related to plane waves, but are superior to OPW's in that the set of all core states plus all COPW's is both orthonormal and complete (not overcomplete).

1. INTRODUCTION

The use of OPW's in the 1-electron band theory of solids has a long history, extending back to their original invention for this purpose by Herring[1] and in recent years reaching a high level of sophistication in the pseudopotential method.[2] Although the lack of linear independence of OPW's and the resultant overcompleteness of the set of all core states plus all OPW's has apparently not been much of a problem in 1-electron applications, it can cause serious difficulties in applications to many-body problems. It has recently been shown[3] how these difficulties can be circumvented in the case of many-electron problems. However, the method proposed there[3] is rather complex algebraically, and furthermore is not readily applicable to many-boson problems.[4] In the present paper, we shall show how a simple modification of the usual OPW formalism leads to a complete orthonormal set of states which is readily applicable to 1-particle, many-fermion, and many-boson problems.

2. ORTHOGONALIZED PLANE WAVES

Suppose that we are provided with a set of "core" or "bound" states $\phi_\alpha(x)$ which are *orthonormal* but *not complete*, together with a *complete*, *orthonormal* set of "plane waves" $\psi_k(x)$; here, x stands for all spatial and spin variables of a single particle, and $\int dx$ will imply integration over spatial and summation over spin variables. The usual OPW's ϕ_k are defined by

$$\phi_k(x) = \psi_k(x) - \Sigma_\alpha(\alpha \mid k)\phi_\alpha(x), \tag{1}$$

where

$$(\alpha \mid k) = \int \phi_\alpha^*(x)\psi_k(x)\,dx. \tag{2}$$

Although orthogonal to the ϕ_α by construction, the ϕ_k are not mutually orthogonal. In fact, they are not even linearly independent.[5] As a result, the set of all ϕ_α plus all ϕ_k is overcomplete.[5] The geometrical reason is clear: There are as many ϕ_k as there are ψ_k, and when we add the ϕ_α we obtain a set with "b too many" functions, where b is the number of bound states ϕ_α. The apparent reason that this has not caused difficulties in applications to the band theory of solids is that if the set $\{\phi_k\}$ is truncated, as in approximating an infinite secular determinant by a finite one, then the resulting *finite* set of $\{\phi_\alpha\}$ and $\{\phi_k\}$ is no longer overcomplete, but undercomplete. Nevertheless, the secular determinant will necessarily become more and more "ill-conditioned" as the size of the basis is increased beyond some optimal size. Furthermore, there are serious difficulties of principle in attempting to use the set of all ϕ_α plus *all* ϕ_k as a basis in many-body calculations.

3. CONSTRUCTION OF COPW'S FOR THE CASE OF ONE BOUND STATE

Suppose there is only one ϕ_α; call it ϕ_b. The set consisting of ϕ_b plus all OPW's ϕ_k is overcomplete because it has "one too many" members. This suggests that it might be possible to construct a complete orthonormal set by "leaving out" one function. In applications there will usually be one of the plane waves ψ_k, say ψ_0, which "most resembles" ϕ_b. This suggests that we try to construct modified OPW's in which the overcompleteness and nonorthogonality problem is removed by leaving out the member corresponding to $k = 0$. We shall show that this can be done provided only that ψ_0 is not identical with ϕ_b:

$$(b \mid 0) \equiv \int \phi_b^*(x)\psi_0(x)\,dx \neq 1. \tag{3}$$

Define

$$f_0(x) = \phi_b(x),$$
$$f_k(x) = \psi_k(x) - c_k[\phi_b(x) - \psi_0(x)], \quad k \neq 0, \tag{4}$$

where c_k remains to be determined. The condition

$$(f_0, f_k) = 0, \quad k \neq 0, \tag{5}$$

uniquely determines c_k:

$$c_k = (b \mid k)/[1 - (b \mid 0)]. \tag{6}$$

Then it is easily verified that the f_k are orthonormal:

$$\begin{aligned}(f_k, f_{k'}) &= (\psi_k, \psi_{k'}) - c_{k'}(b \mid k)^* - c_k(b \mid k')^* \\ &\quad + c_k^* c_{k'}[2 - (b \mid 0) - (b \mid 0)^*] \\ &= (\psi_k, \psi_{k'}) = \delta_{kk'}, \quad k \neq 0, \quad k' \neq 0. \end{aligned} \tag{7}$$

Finally, one can verify that the set $\{f_k\}$ is complete,

$$\Sigma_k f_k(x) f_k^*(x') = \Sigma_k \psi_k(x) \psi_k^*(x') = \delta(x - x'), \tag{8}$$

with the aid of the relations

$$\begin{aligned}\Sigma_k' |c_k|^2 &= \frac{1 - |(b \mid 0)|^2}{|1 - (b \mid 0)|^2}, \\ \Sigma_k' c_k^* \psi_k(x) &= \frac{\phi_b(x) - (b \mid 0)^* \psi_0(x)}{1 - (b \mid 0)^*}, \end{aligned} \tag{9}$$

which follow from (6) and completeness of the set $\{\psi_k\}$; the primes on the summations (9) imply omission of $k = 0$.

The expressions (4) and (6) for the COPW's can be written in a more compact form by defining

$$\begin{aligned}\phi_b'(x) &= \psi_0(x) - \phi_b(x), \\ (b' \mid k) &= \int [\phi_b'(x)]^* \psi_k(x)\, dx = \delta_{k0} - (b \mid k). \end{aligned} \tag{10}$$

Then

$$f_k(x) = \psi_k(x) - \frac{(b' \mid k)}{(b' \mid 0)} \phi_b'(x), \tag{11}$$

both for $k \neq 0$ and for $k = 0$.

4. EXAMPLE

We shall present here a simple example which brings out the physical reasoning behind the mathematical ansatz (4). Suppose x stands for the position vector $\mathbf{x}$ of a single spinless particle, take the indices k to be wave vectors $\mathbf{k}$, and choose the $\psi_{\mathbf{k}}$ to be normalized plane waves with periodicity volume Ω:

$$\psi_{\mathbf{k}}(\mathbf{x}) = \Omega^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{12}$$

Define $\chi_b(\mathbf{x})$ and $\tilde{\chi}_b(\mathbf{k})$ by

$$\begin{aligned}\phi_b'(\mathbf{x}) &= \Omega^{-\frac{1}{2}} \chi_b(\mathbf{x}), \\ \tilde{\chi}_b(\mathbf{k}) &= \int \chi_b(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x. \end{aligned} \tag{13}$$

Then the COPW's (11) are

$$f_{\mathbf{k}}(\mathbf{x}) = \Omega^{-\frac{1}{2}} \left[e^{i\mathbf{k}\cdot\mathbf{x}} - \frac{\tilde{\chi}_b^*(\mathbf{k})}{\tilde{\chi}_b^*(0)} \chi_b(\mathbf{x}) \right]. \tag{14}$$

Suppose $\phi_b(\mathbf{x})$ differs from $\psi_0(\mathbf{x}) = \Omega^{-\frac{1}{2}}$ only by having a "hole cut out" in the neighborhood of the origin.[6] Then $\chi_b(\mathbf{x})$ will be unity in the neighborhood of the origin, falling to zero for x larger than the radius of the hole. It is then clear from (14) that not only f_0, but all $f_{\mathbf{k}}$ of sufficiently low k, will have the same hole cut out of them. The functions $f_{\mathbf{k}}$ are thus well suited to the physics of the suggested application[6]; they correctly represent the physical fact that He^4 atoms of low energy do not penetrate the "bubble" blown by the electron. This application was the origin of the ansatz (4).

5. CONSTRUCTION OF COPW'S FOR THE CASE OF MANY BOUND STATES

Return now to the case of many bound states ϕ_α, originally considered in Sec. 2. In many applications it is possible to pick out a subset of the "plane waves" ψ_k such that each member of this subset resembles, in certain respects, one of the ϕ_α. In this way, one sets up a 1-to-1 correspondence $\phi_\alpha \leftrightarrow \psi_{k_\alpha}$ between the bound states ϕ_α and certain of the plane waves. This suggests the following generalization of (4):

$$\begin{aligned}f_\alpha(x) &= \phi_\alpha(x), \\ f_k(x) &= \psi_k(x) - \Sigma_\alpha c_{\alpha,k}[\phi_\alpha(x) - \psi_{k\alpha}(x)], \\ &\qquad k \neq \text{any } k_\alpha. \end{aligned} \tag{15}$$

The conditions

$$(f_\alpha, f_k) = 0, \quad k \neq \text{any } k_\beta \tag{16}$$

lead to the equations

$$\Sigma_\beta M_{\alpha\beta} c_{\beta,k} = (\alpha \mid k) \tag{17}$$

for the determination of the $c_{\alpha,k}$, where

$$M_{\alpha\beta} = \delta_{\alpha\beta} - (\alpha \mid k_\beta). \tag{18}$$

The inhomogeneous equations (17) have a unique solution for the $c_{\alpha,k}$ provided only that the determinant of the matrix $(M_{\alpha\beta})$ does not vanish.[7] It is readily verified that the $c_{\alpha,k}$ satisfying (17) also lead to orthonormal COPW's:

$$(f_k, f_{k'}) = \delta_{kk'}, \quad k \text{ and } k' \neq \text{any } k_\alpha. \tag{19}$$

The completeness relation

$$\begin{aligned}&\Sigma_\alpha f_\alpha(x) f_\alpha^*(x') + \Sigma_k' f_k(x) f_k^*(x') \\ &\qquad = \Sigma_\alpha \psi_{k\alpha}(x) \psi_{k\alpha}^*(x') + \Sigma_k' \psi_k(x) \psi_k^*(x') \\ &\qquad = \delta(x - x'), \end{aligned} \tag{20}$$

where the prime on Σ_k' implies omission of all k_α, can be verified by tedious algebraic reductions based on the identities

$$\begin{aligned}c_{\alpha,k} &= \Sigma_\beta (M^{-1})_{\alpha\beta} (\beta \mid k), \\ \Sigma_k' (\beta \mid k)^* \psi_k(x) &= \phi_\beta(x) - \Sigma_\alpha (\beta \mid k_\alpha)^* \psi_{k\alpha}(x), \\ \Sigma_k' (\alpha \mid k)(\beta \mid k)^* &= \delta_{\alpha\beta} - \Sigma_\gamma (\alpha \mid k_\gamma)(\beta \mid k_\gamma)^*, \\ \Sigma_\beta (M^{-1})_{\alpha\beta} (\beta \mid k_\gamma) &= (M^{-1})_{\alpha\gamma} - \delta_{\alpha\gamma}, \end{aligned} \tag{21}$$

which follow from (17), (18), and completeness of the set $\{\psi_k\}$.

More explicit expressions for the COPW's can be derived in the special case that the symmetry of the ϕ_α and $\psi_{k\alpha}$ is such that

$$(\alpha \mid k_\beta) = 0, \quad \alpha \neq \beta. \tag{22}$$

Then it follows trivially from (17) and (18) that

$$c_{\alpha,k} = (\alpha \mid k)/[1 - (\alpha \mid k_\alpha)]. \tag{23}$$

As in Sec. 3, it is convenient to define

$$\phi'_\alpha(x) = \psi_{k_\alpha}(x) - \phi_\alpha(x),$$

$$(\alpha' \mid k) = \int [\phi'_\alpha(x)]^* \psi_k(x)\, dx = \delta_{k,k_\alpha} - (\alpha \mid k). \tag{24}$$

Then (15) can be written

$$f_k(x) = \psi_k(x) - \Sigma_\alpha \frac{(\alpha' \mid k)}{(\alpha' \mid k_\alpha)} \phi'_\alpha(x) \tag{25}$$

for all k, provided that we define

$$f_{k_\alpha}(x) = f_\alpha(x) = \phi_\alpha(x). \tag{26}$$

ACKNOWLEDGMENTS

I am indebted to Peter E. Parks for discussions which led to clarification and simplification of this formalism. This research was supported by the National Science Foundation under grant GP-7689.

* Delay in publication was due to a mix-up in the editorial office and was in no way the fault of the author.

[1] C. Herring, Phys. Rev. **57**, 1169 (1940); C. Herring and A. G. Hill, Phys. Rev. **58**, 132 (1940).

[2] See, e.g., W. A. Harrison, *Pseudopotentials in the Theory of Metals* (Benjamin, New York, 1966).

[3] M. D. Girardeau and M. L. Sage, Physica **32**, 1521 (1966).

[4] The theory of inhomogeneities (impurities, ions, vortices, etc.) in liquid He^4 suggests many applications to many-boson systems.

[5] Equation (4) ff of Ref. 3.

[6] This case is suggested by the theory of the negative ion (bubble) in liquid He^4, which is currently being investigated, using this formalism, by P. E. Parks and the author.

[7] This determinant would vanish in the trivial case that every ψ_{k_α} is identical with its correspondent ϕ_α, but for any reasonable nontrivial choice of the ϕ_α it will not vanish.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Equivalence of Basis-Dependent and Basis-Independent Approach to Canonical Field Operators

GERHARD C. HEGERFELDT
Institut für Theoretische Physik, Universität Göttingen, 34 Göttingen, Germany

(Received 3 February 1970)

The basis-independent approach to canonical commutation relations (CCR's), which allows arbitrary test-function spaces for smearing the field operators, is a generalization of the basis-dependent approach, in which the fields are smeared with an orthonormal system (or finite linear combinations thereof) to obtain an infinite set of q_k and p_k. Using recent results on continuity properties of representations of the CCR's, we show that every representation of the basis-independent type in a separable Hilbert space can be obtained by continuous extension of a suitable representation of the basis-dependent type where properties like irreducibility, cyclicity, etc., remain unaffected. In this sense, both approaches are equivalent, and the classification problem for CCR's is reduced from the simultaneous consideration of all representations for *all* possible test-function spaces to those for a single one (up to isomorphism).

1. INTRODUCTION

There exist two different approaches to the canonical commutation relations for infinitely many degrees of freedom or fields (equal-time commutation relations). The first one, as investigated by Gårding and Wightman,[1] starts from infinitely many q_k and p_k with the usual commutation relations. The second one, as formulated by Segal[2] and studied by Araki[3] and Lew,[4] starts from smeared field operators $\Phi(f)$ and $\Pi(g)$ where f and g are elements of some real test-function or some general real linear spaces $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$, respectively. One has a nondegenerate bilinear form (f, g). Heuristically,

$$\Phi(f) = \int \Phi(\mathbf{x}, t_0) f(\mathbf{x}) d^3\mathbf{x}$$

and similarly for $\Pi(g)$, so that in view of

$$[\Phi(\mathbf{x}), \Pi(\mathbf{x})] = i\delta^{(3)}(\mathbf{x} - \mathbf{y})$$

one demands

$$[\Phi(f), \Pi(g)] = i(f, g).$$

In this second "basis-independent" approach one

which follow from (17), (18), and completeness of the set $\{\psi_k\}$.

More explicit expressions for the COPW's can be derived in the special case that the symmetry of the ϕ_α and $\psi_{k\alpha}$ is such that

$$(\alpha \mid k_\beta) = 0, \quad \alpha \neq \beta. \tag{22}$$

Then it follows trivially from (17) and (18) that

$$c_{\alpha,k} = (\alpha \mid k)/[1 - (\alpha \mid k_\alpha)]. \tag{23}$$

As in Sec. 3, it is convenient to define

$$\phi'_\alpha(x) = \psi_{k_\alpha}(x) - \phi_\alpha(x),$$

$$(\alpha' \mid k) = \int [\phi'_\alpha(x)]^* \psi_k(x)\, dx = \delta_{k,k_\alpha} - (\alpha \mid k). \tag{24}$$

Then (15) can be written

$$f_k(x) = \psi_k(x) - \Sigma_\alpha \frac{(\alpha' \mid k)}{(\alpha' \mid k_\alpha)} \phi'_\alpha(x) \tag{25}$$

for all k, provided that we define

$$f_{k_\alpha}(x) = f_\alpha(x) = \phi_\alpha(x). \tag{26}$$

ACKNOWLEDGMENTS

I am indebted to Peter E. Parks for discussions which led to clarification and simplification of this formalism. This research was supported by the National Science Foundation under grant GP-7689.

* Delay in publication was due to a mix-up in the editorial office and was in no way the fault of the author.

[1] C. Herring, Phys. Rev. **57**, 1169 (1940); C. Herring and A. G. Hill, Phys. Rev. **58**, 132 (1940).

[2] See, e.g., W. A. Harrison, *Pseudopotentials in the Theory of Metals* (Benjamin, New York, 1966).

[3] M. D. Girardeau and M. L. Sage, Physica **32**, 1521 (1966).

[4] The theory of inhomogeneities (impurities, ions, vortices, etc.) in liquid He^4 suggests many applications to many-boson systems.

[5] Equation (4) ff of Ref. 3.

[6] This case is suggested by the theory of the negative ion (bubble) in liquid He^4, which is currently being investigated, using this formalism, by P. E. Parks and the author.

[7] This determinant would vanish in the trivial case that every ψ_{k_α} is identical with its correspondent ϕ_α, but for any reasonable nontrivial choice of the ϕ_α it will not vanish.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 12, NUMBER 1 JANUARY 1971

Equivalence of Basis-Dependent and Basis-Independent Approach to Canonical Field Operators

GERHARD C. HEGERFELDT
Institut für Theoretische Physik, Universität Göttingen, 34 Göttingen, Germany

(Received 3 February 1970)

The basis-independent approach to canonical commutation relations (CCR's), which allows arbitrary test-function spaces for smearing the field operators, is a generalization of the basis-dependent approach, in which the fields are smeared with an orthonormal system (or finite linear combinations thereof) to obtain an infinite set of q_k and p_k. Using recent results on continuity properties of representations of the CCR's, we show that every representation of the basis-independent type in a separable Hilbert space can be obtained by continuous extension of a suitable representation of the basis-dependent type where properties like irreducibility, cyclicity, etc., remain unaffected. In this sense, both approaches are equivalent, and the classification problem for CCR's is reduced from the simultaneous consideration of all representations for *all* possible test-function spaces to those for a single one (up to isomorphism).

1. INTRODUCTION

There exist two different approaches to the canonical commutation relations for infinitely many degrees of freedom or fields (equal-time commutation relations). The first one, as investigated by Gårding and Wightman,[1] starts from infinitely many q_k and p_k with the usual commutation relations. The second one, as formulated by Segal[2] and studied by Araki[3] and Lew,[4] starts from smeared field operators $\Phi(f)$ and $\Pi(g)$ where f and g are elements of some real test-function or some general real linear spaces $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$, respectively. One has a nondegenerate bilinear form (f, g). Heuristically,

$$\Phi(f) = \int \Phi(\mathbf{x}, t_0) f(\mathbf{x}) d^3\mathbf{x}$$

and similarly for $\Pi(g)$, so that in view of

$$[\Phi(\mathbf{x}), \Pi(\mathbf{x})] = i\delta^{(3)}(\mathbf{x} - \mathbf{y})$$

one demands

$$[\Phi(f), \Pi(g)] = i(f, g).$$

In this second "basis-independent" approach one

can, in principle, consider all possible test-function spaces $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$. In this respect, it is a generalization of the first approach which is obtained by choosing the particular test-function space $\mathfrak{V}_\Phi = \mathfrak{V}_\Pi = \mathfrak{V}_0$, where $\mathfrak{V}_0$ consists of all finite linear combinations of some orthonormal system $h_1, h_2, \cdots$; $\mathfrak{V}_0$ can be identified with the space of all finite sequences $(a_1, \cdots, a_n, 0, \cdots)$ of real numbers. One can define $q_k = \Phi(h_k)$ and $p_k = \Pi(h_k)$. This will be called the basis dependent (or Gårding–Wightman) approach.

The choice of specific test-function spaces, up to now, had to be justified by physical arguments. If $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$ were chosen too large, e.g., equal to the space L^2 of all square-integrable functions, one might lose some physically significant representations of the CCR's with a smaller test-function space, for, by restricting representations for L^2 to a subspace $\mathfrak{V} \subset L^2$, one need not obtain all representations for $\mathfrak{V}$. On the other hand, if the test-function spaces were chosen too small, this might destroy properties of the representation such as cyclicity, irreducibility, or what not; and in quantum field theory these properties have a direct physical significance (cf., e.g., Araki's results for the Hamiltonian[3]).

The purpose of this paper is to show that, without losing any of such properties, one can restrict oneself to $\mathfrak{V}_0$. It will turn out that every representation of the basis-independent approach (for any choice of $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$) in a separable Hilbert space can be obtained from a suitable Gårding–Wightman representation by extending the latter in the natural metric introduced by Hegerfeldt and Klauder[5] and identifying isomorphic test-function spaces. This reduces the problem of classifying all representations of the CCR's for all possible test-function spaces to the study of representations of the Gårding–Wightman type.

It is customary to define a representation of the CCR's to be a family of unitary operators $U(f)$ and $V(g)$ with $f \in \mathfrak{V}_\Phi$ and $g \in \mathfrak{V}_\Pi$ satisfying the Weyl relations

$$U(f_1 + f_2) = U(f_1)U(f_2),$$
$$V(g_1 + g_2) = V(g_1)V(g_2),$$
$$V(g)U(f) = \exp[i(f, g)]U(f)V(g).$$

One assumes ray continuity, i.e., weak continuity of $U(\lambda f)$ and $V(\lambda g)$ in λ for fixed f and g. The relationship to the fields is given by $U(f) = e^{i\Phi(f)}$ and $V(g) = e^{i\Pi(g)}$ through Stone's theorem. This Weyl form avoids domain questions. Taking $\mathfrak{V}_\Phi = \mathfrak{V}_\Pi = \mathfrak{V}_0$ one obtains the basis-dependent approach in the Weyl form.

Any representation of the CCR's is a direct sum of cyclic representations. A cyclic representation of the basis-dependent approach, i.e., with $\mathfrak{V}_\Phi = \mathfrak{V}_\Pi = \mathfrak{V}_0$, belongs necessarily to a separable Hilbert space, so that in this case nonseparable Hilbert spaces can enter only through uncountable direct sums. In the basis-independent approach this need not be so.[6] It is clear, therefore, that the above-mentioned separability condition is necessary.

A precise formulation of the main result is given in Sec. 4. Its proof is based on properties of the metric d which are of independent interest. They are derived in Secs. 2 and 3. Section 5 contains a discussion of the results and of its connection with Gårding domains and analytic vectors.

2. CONVERGENCE IN METRIC d IS CONVERGENCE IN MEASURE

In Ref. 5 it has been shown that each representation of the CCR's which is cyclic or a countable direct sum of cyclic representations induces a natural metric d_U on $\mathfrak{V}_\Phi$ and d_V on $\mathfrak{V}_\Pi$. The metric can be written as

$$d_U(f - f')^2 = \int_{-\infty}^{\infty} d\lambda e^{-\lambda^2} \|[U(\lambda(f - f')) - \mathbf{1}]\varphi_0\|^2, \qquad f, f' \in \mathfrak{V}_\Phi, \quad (2.1)$$

and similarly d_V for $\mathfrak{V}_\Pi$, where φ_0 is either a cyclic vector or a vector whose components with respect to the direct sum of cyclic subspaces are cyclic in each subspace. The metric defines the weakest vector topology on $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$ for which the maps $f \to U(f)$ and $g \to V(g)$ are strongly continuous. Furthermore, the representation can be extended by continuity to the completion of $\mathfrak{V}_\Phi$ and $\mathfrak{V}_\Pi$ in d_U and d_V, respectively. In the following, we mostly suppress the indices U and V and talk about the metric d.

In Ref. 7 a direct integral form of representations of the CCR's has been obtained which is, in particular, valid for a separable Hilbert space $\mathcal{H}$. Accordingly, one can write

$$\mathcal{H} = \int_{\mathfrak{V}'_\Phi}^{\oplus} \mathcal{H}(F)\, d\mu(F), \qquad (2.2)$$

where $\mathfrak{V}'_\Phi$ is the algebraic dual of $\mathfrak{V}_\Phi$, $F \in \mathfrak{V}'_\Phi$, and where μ is a finite $\mathfrak{V}_\Pi$-quasi-invariant measure (on the Borel sets of $\mathfrak{V}'_\Phi$, the σ-algebra generated by the weakly open cylinder sets in $\mathfrak{V}'_\Phi$). The action of $U(f)$ is given by

$$(U(f)\varphi)(F) = e^{i\langle f, F\rangle}\varphi(F). \qquad (2.3)$$

The form of $V(g)$ in this direct integral realization is not of interest here. Now the following connection between d and μ will be established.

Theorem 2.1: Assume $f_n \in \mathfrak{U}_\Phi$. Then $d(f_n) \to 0$ if and only if $(f_n, F) \to 0$ in μ-measure.

Proof: Let φ_0 correspond to the vector function $\varphi_0(F)$. According to the remark following Corollary 5.4 in Ref. 5, one can assume $\|\varphi_0(F)\| \equiv 1$. Inserting Eq. (2.3) into (2.1), one obtains

$$d(f)^2 = \int_{-\infty}^{\infty} d\lambda e^{-\lambda^2} \int_{\mathfrak{U}'_\Phi} d\mu(F) \, \|(e^{i\lambda(f,F)} - 1)\varphi_0(F)\|^2,$$

$$d(f)^2 = 2\int d\lambda e^{-\lambda^2} \int d\mu(F)\{1 - \cos[\lambda(f, F)]\}. \quad (2.4)$$

Using the properties of such measures, as, for instance, explained in Ref. 8, one can show that $\cos[\lambda(f, F)]$ is measurable on $R_1 \times \mathfrak{U}'_\Phi$. Therefore, since the integrand is positive, one can interchange the order of integration. Thus one obtains

$$d(f)^2 = 2\pi^{\frac{1}{2}} \int d\mu(F)(1 - e^{-\frac{1}{4}(f,F)^2}). \quad (2.5)$$

Now, for $d(f_n)^2 \to 0$, it is necessary that the integrand go to zero in measure. Thus $\exp[-\frac{1}{4}(f_n, F)^2]$ has to go to one in measure. But then (f_n, F) goes to zero in measure. Conversely, if this is the case, then the integrand goes to zero in measure. Since the integrand is bounded and since the measure is finite, the integral tends to zero, by Lebesgue's bounded convergence. QED

Interchanging the role of $U(f)$ and $V(g)$, one obtains a finite measure $\tilde{\mu}$ on $\mathfrak{U}'_\Pi$ and a result analogous to Theorem 2.1.

Each $f \in \mathfrak{U}_\Phi$ defines a measurable function on $\mathfrak{U}'_\Phi$ by $f(F) = (f, F)$. As for any finite measure space, the set of all (equivalence classes of all) μ-measurable functions on $\mathfrak{U}'_\Phi$ becomes a metrizable linear topological space if endowed with the topology of convergence in measure. $\mathfrak{U}_\Phi$ can be regarded as a subspace. Theorem 2.1 states that $\mathfrak{U}_\Phi$ with its metric d is homeomorphic to this subspace.

3. THE METRIC SPACES $\mathfrak{U}_\Phi$ AND $\mathfrak{U}_\Pi$ ARE SEPARABLE

In this section the following result will be proved:

Theorem 3.1: Let $U(f)$, $V(g)$, $f \in \mathfrak{U}_\Phi$ and $g \in \mathfrak{U}_\Pi$ be a representation of the CCR's in a *separable* Hilbert space $\mathcal{H}$. Then $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$ are separable in the metric d.

Proof: By Theorem 1 of Ref. 7, one can regard the space L^2_μ of all μ-square-integrable functions as a subspace of $\mathcal{H}$. This can also be seen directly from the integral decomposition (2.2), if one represents $\mathcal{H}(F)$ as a sequence space of dimension $n(F)$. Then L^2_μ is isometric to the space of all vector functions $\psi(F) \in \mathcal{H}(F)$ with all components, except possibly the first, vanishing. Thus $L^2_\mu \subset \mathcal{H}$ is separable.

This implies that the measure ring of all measurable sets in $\mathfrak{U}'_\Phi$ modulo μ-null sets is separable.[9] Let $\hat{E}_1$, $\hat{E}_2, \cdots$ be a dense set in the measure ring, let E_i be a representative for $\hat{E}_i$, and let $\chi_{E_i}(F)$ be the characteristic function of E_i, $i = 1, 2, \cdots$. For the metric on the ring, one has

$$\hat{\rho}(\hat{E}_i, \hat{E}_j) = \mu\{F: |\chi_{E_i}(F) - \chi_{E_j}(F)| > 0\}.$$

Now let $f(F)$ be a measurable function. Then there exists a sequence of simple functions $s_\nu(F)$ which converges to $f(F)$ everywhere. Hence, $s_\nu \to f$ in measure. The simple functions s_ν have the form

$$s_\nu(F) = \sum_{\alpha=1}^{N(\nu)} c_\alpha^{(\nu)} \chi_\alpha^{(\nu)}(F),$$

where the $\chi_\alpha^{(\nu)}$ are characteristic functions of measurable sets. Hence, for each $\chi_\alpha^{(\nu)}$, there is a particular $E_{i(\alpha,\nu)}$ of the above $\{E_i\}$ such that

$$\mu\{F: |\chi_{E_{i(\alpha,\nu)}}(F) - \chi_\alpha^{(\nu)}(F)| > 0\} < \epsilon/2N(\nu)r_\nu,$$

where ϵ is any positive number and where r_ν is the maximum over α of rational numbers $r_\alpha^{(\nu)}$ which are so chosen that

$$|c_\alpha^{(\nu)} - r_\alpha^{(\nu)}| < \epsilon/2N(\nu).$$

We put

$$\hat{s}(F) = \sum_{\alpha=1}^{N(\nu)} r_\alpha^{(\nu)} \chi_{i(\alpha,\nu)}(F). \quad (3.1)$$

Both s_ν and $\hat{s}$ are in L^1, and by the triangle inequality $\|s_\nu - \hat{s}\|_1 < \epsilon$. Hence for each ν there is a sequence $\hat{s}_{\nu,k}$ of the same form as in Eq. (3.1) converging to s_ν in L^1. But this implies $\hat{s}_{\nu,k} \to s_\nu$ in measure. Let ρ be the metric for convergence in measure. One can choose a function $\hat{s}_\nu$ from $\{\hat{s}_{\nu,k}\}$ such that $\rho(s_\nu - \hat{s}_\nu) < 2^{-\nu}$. Then

$$\rho(f - \hat{s}_\nu) \leq \rho(f - s_\nu) + \rho(s_\nu - \hat{s}_\nu) \to 0.$$

Since the set of functions of the form of Eq. (3.1) is countable, the space of measurable function with the topology of convergence in measure is separable. Now, every subset of a separable metric space is also separable. Hence, by Theorem 2.1, $\mathfrak{U}_\Phi$ with its metric d is separable.

Interchanging the role of $U(f)$ and $V(g)$, one obtains the same result for $\mathfrak{U}_\Pi$. QED

Although it is not needed for the following, we note in passing that the converse of Theorem 3.1 also holds.

Corollary 3.1: Let $U(f)$, $V(g)$ be a representation of the CCR's which decomposes into a countable direct sum of cyclic representations. Let both $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$ with their respective metric d be separable. Then the representation space $\mathcal{H}$ is separable.[10]

Proof: Consider a cyclic subspace $\mathcal{H}_1 \subset \mathcal{H}$ with cyclic vector φ_0. The set of linear combinations with rational coefficients of vectors of the form $U(f)V(g)\varphi_0$ is dense in $\mathcal{H}_1$. Now let $\{f_i,\ i = 1, 2, \cdots\}$ and $\{g_j,\ j = 1, 2, \cdots\}$ be dense sets in $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$ for d. For each $f \in \mathfrak{U}_\Phi$ and $g \in \mathfrak{U}_\Pi$, given $\epsilon > 0$, there is an f_i and a g_j such that

$$\|\{U(f) - U(f_i)\}V(g)\varphi_0\|$$

and

$$\|\{V(g) - V(g_j)\}\varphi_0\| < \tfrac{1}{2}\epsilon.$$

Hence

$$\|\{U(f)V(g) - U(f_i)V(g_j)\}\varphi_0\| < \epsilon.$$

Therefore, the set of linear combinations with rational coefficients of all $U(f_i)V(g_j)\varphi_0$ is dense in $\mathcal{H}_1$. Thus $\mathcal{H}_1$ is separable. As a countable direct sum of such subspaces, $\mathcal{H}$ is also separable. QED

4. THERE ARE ENOUGH REPRESENTATIONS OF THE BASIS-DEPENDENT APPROACH

The result announced in the introduction is now an immediate consequence of the following theorem, whose significance will be discussed further below.

Theorem 4.1: Let $U(f)$, $V(g)$, with $f \in \mathfrak{U}_\Phi$ and $g \in \mathfrak{U}_\Pi$, be a representation of the CCR's in a separable Hilbert space, and let d_U and d_V be the associated metrics on $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$. Then there are elements $h_1, h_2, \cdots$ in $\mathfrak{U}_\Phi$ and $h_1', h_2', \cdots$ in $\mathfrak{U}_\Pi$ such that:

(i) $(h_i, h_k') = \delta_{ik}$ for all i, k.

(ii) $\mathfrak{U}_\Phi^0 = \mathfrak{L}\{h_1, h_2, \cdots\}$

and

$$\mathfrak{U}_\Pi^0 = \mathfrak{L}\{h_1', h_2', \cdots\},$$

the finite linear spans of the $h_1, h_2, \cdots$ and $h_1', h_2', \cdots$, are dense in the metric spaces $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$, respectively. If $\mathfrak{U}_\Phi = \mathfrak{U}_\Pi = \mathfrak{U}$, one can choose $h_i = h_i'$ such that $\mathfrak{U}_\Phi^0 = \mathfrak{U}_\Pi^0$ is dense in $\mathfrak{U}$ for the metric $d_U + d_V$.

Proof: By Theorem 3.1, there exists a sequence of elements in $\mathfrak{U}_\Phi$ which is dense in $\mathfrak{U}_\Phi(d_U)$. We omit all elements which are linearly dependent on preceding ones. We call the resulting sequence $\{f_1, f_2, \cdots\}$, and denote by $\mathfrak{U}_\Phi^0$ the set of all finite linear combinations of the f_i. Then $\mathfrak{U}_\Phi^0$ is dense in $\mathfrak{U}_\Phi(d_U)$. We define $\{g_1, g_2, \cdots\}$ and $\mathfrak{U}_\Pi^0$ in the same way for $\mathfrak{U}_\Pi$.

Assume $f \in \mathfrak{U}_\Phi^0$ and $(f, g) = 0$ for each $g \in \mathfrak{U}_\Pi^0$. Then $f = 0$, since for any element $\tilde{g} \in \mathfrak{U}_\Pi$ there is a sequence $\{\tilde{g}_n\}$ such that $\tilde{g}_n \in \mathfrak{U}_\Pi^0$ and $\tilde{g}_n \to \tilde{g}(d)$. But then, by Theorem 1.5 of Ref. 5, $(f, \tilde{g}) = \lim (f, \tilde{g}_n) = 0$. Similarly $g = 0$ if $(f, g) = 0$ for all $f \in \mathfrak{U}_\Phi^0$. Hence the bilinear form (f, g) remains nondegenerate when restricted to $\mathfrak{U}_\Phi^0$ and $\mathfrak{U}_\Pi^0$.

It follows that, for fixed i, (f_i, g_ν) cannot vanish for all ν. With this observation one can construct $\{h_i\}$ and $\{h_i'\}$ by a three-step induction, a generalization of the Schmidt orthogonalization procedure.

We put $h_1 = f_1$ and let g_{k_1} be the first g_i such that $a_1 = (f_1, g_{k_1}) \neq 0$. We put $h_1' = a_1^{-1} g_{k_1}$. Assume that h_i, g_{k_i}, and h_i', $i = 1, n-1$, have been defined in such a way that

$$(h_i, h_k') = \delta_{ik},$$

$$\mathfrak{L}\{h_\nu, \nu = 1, \cdots, i\} = \mathfrak{L}\{f_\nu, \nu = 1, \cdots, i\},$$

$$\mathfrak{L}\{h_\nu', \nu = 1, \cdots, i\} = \mathfrak{L}\{g_{k_\nu}, \nu = 1, \cdots, i\},$$

$$(h_i, g_{k_i}) \neq 0, \quad (h_i, g_\alpha) = 0 \quad \text{for} \quad \alpha < k_i, \quad i = 1, \cdots, n-1.$$

Now we define

$$h_n = f_n - \sum_{i=1}^{n-1} (f_n, h_i') h_i.$$

Then $h_n \neq 0$, $(h_n, h_i') = 0$, and $\mathfrak{L}\{h_1, \cdots, h_n\} = \mathfrak{L}\{f_1, \cdots, f_n\}$. Let g_{k_n} be the first member of $\{g_1, \cdots\}$ which satisfies $a_n = (h_n, g_{k_n}) \neq 0$.[11] Define

$$h_n' = a_n^{-1}\left\{g_{k_n} - \sum_{i=1}^{n-1} (h_i, g_{k_n}) h_i'\right\}.$$

Then $(h_n, h_n') = 1$, $(h_i, h_n') = 0$ for $i < n$, and

$$\mathfrak{L}\{h_1', \cdots, h_n'\} = \mathfrak{L}\{g_{k_1}', \cdots, g_{k_n}'\}.$$

Obviously, the finite linear span of the h_i is $\mathfrak{U}_\Phi^0$. Furthermore, each g_i appears as a g_{k_n} for some n; for, otherwise, $(h_n, g_i) = 0$ for all n and thus $g_i = 0$. Since $g_{k_n} \in \mathfrak{L}\{h_1', \cdots, h_n'\}$, it follows that $\mathfrak{U}_\Pi^0 = \mathfrak{L}\{h_1', h_2', \cdots\}$. If $\mathfrak{U}_\Phi = \mathfrak{U}_\Pi = \mathfrak{U}$, then $\mathfrak{U}$ is separable for the metric $d_U + d_V$, and one simply has to orthogonalize a dense sequence. QED

The denseness of $\mathfrak{U}_\Phi^0$ in $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi^0$ in $\mathfrak{U}_\Pi$ with respect to the metrics d_U and d_V, respectively, means that every element $f \in \mathfrak{U}_\Phi$ is a limit of a sequence in $\mathfrak{U}_\Phi^0$, $f_n \to f(d)$, where at the same time, by Ref. 5, $U(\lambda f_n) \to U(\lambda f)$ strongly for every[12] λ; similarly for $\mathfrak{U}_\Pi$. Thus it suffices to know the operators $U(f)$ and $V(g)$ only for $f \in \mathfrak{U}_\Phi^0$ and $g \in \mathfrak{U}_\Pi^0$ since then the whole representation is determined. That is, one can confine oneself to the subrepresentation $U(f)$, $V(g)$, with $f \in \mathfrak{U}_\Phi^0$ and $g \in \mathfrak{U}_\Pi^0$. Since the metrics d_U^0 and d_V^0

induced by this representation on $\mathfrak{U}_\Phi^0$ and $\mathfrak{U}_\Pi^0$ coincide with the restriction of d to $\mathfrak{U}_\Phi^0$ and $\mathfrak{U}_\Pi^0$, it suffices to extend the representation $U(f)$, $V(g)$, with $f \in \mathfrak{U}_\Phi^0$ and $g \in \mathfrak{U}_\Pi^0$, in the metrics d_U^0 and d_V^0. To obtain all of $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$, it may, in general, not be necessary to go to the completion of $\mathfrak{U}_\Phi^0$ and $\mathfrak{U}_\Pi^0$ since $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$ are, in general, not complete in d.

Now, if one puts $U(\lambda h_k) = e^{i\lambda q_k}$ and $V(\lambda h_k') = e^{i\lambda p_k}$, one just obtains a basis-dependent representation; any $U(f), f \in \mathfrak{U}_\Phi^0$, is a finite product of some $U(\lambda_k h_k)$ and similarly for $g \in \mathfrak{U}_\Pi^0$. If one isomorphically identifies $h_n \leftrightarrow h_n'$ then $\mathfrak{U}_\Phi^0 = \mathfrak{U}_\Pi^0$. In this case, one has to bear in mind that in general the metrics d_U and d_V are different so that, to obtain $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$, one has to extend $\mathfrak{U}_0 = \mathfrak{U}_\Phi^0 = \mathfrak{U}_\Pi^0$ by d_U^0 and d_V^0, respectively.

This shows that the basis-dependent approach provides enough representations to obtain any representation of the *basis-independent* approach by a continuous extension of a suitably chosen *basis-dependent* Gårding–Wightman representation. This is summarized in the next corollary, a reformulation of Theorem 4.1 and the main result of this paper.

Corollary 4.1 (Sufficiency of Basis-Dependent Approach): For every representation of the basis-independent approach, $U(f), V(g)$, with $f \in \mathfrak{U}_\Phi$ and $g \in \mathfrak{U}_\Pi$, in a separable Hilbert space $\mathcal{H}$, there is a representation of the basis-dependent approach, $U_0(f_0)$, $V_0(g_0)$, with f_0, $g_0 \in \mathfrak{U}_0 = \mathfrak{L}\{h_1, h_2, \cdots\}$, $(h_i, h_k) = \delta_{ik}$, in $\mathcal{H}$, with the following properties: (i) $\mathfrak{U}_0$ can be embedded in $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$ in such a way that $U(f) = U_0(f_0)$ if $f \leftrightarrow f_0$ and $V(g) = V_0(g_0)$ if $g \leftrightarrow g_0$; (ii) $\mathfrak{U}_\Phi$ can be embedded in the completion of $\mathfrak{U}_0$ with respect to the metric d_{U_0}, and $\mathfrak{U}_\Pi$ can be embedded in the completion of $\mathfrak{U}_0$ with respect to the metric d_{V_0}, i.e., every $f \in \mathfrak{U}_\Phi$ can be identified with d_{V_0}-Cauchy sequence in $\mathfrak{U}_0$; (iii) if $f_n \in \mathfrak{U}_0$, $f \in \mathfrak{U}_\Phi$, and $f_n \to f(d_{U_0})$, then $U_0(\lambda f_n) \to U(\lambda f)$ strongly for each λ; if $g_n \in \mathfrak{U}_0$, $g \in \mathfrak{U}_\Pi$, and $g_n \to g(d_{V_0})$, then $V_0(\lambda g_n) \to V(\lambda g)$ strongly for each λ.[13]

5. DISCUSSION: ANALYTIC VECTORS AND GÅRDING DOMAINS

Let us discuss the above result for the physically important case of equal test-functions spaces for both fields, $\mathfrak{U}_\Phi = \mathfrak{U}_\Pi = \mathfrak{U}$. One has to determine *all* representations of the CCR's in separable Hilbert spaces for *all* $\mathfrak{U}$. Consider, first, Gårding–Wightman representations, i.e., for the space $\mathfrak{U}_0$ of all finite linear combinations of an orthonormal basis. Each such representation, D say, determines a metric $d_{U_0} + d_{V_0}$ on $\mathfrak{U}_0$, and D is strongly continuous in this metric. Denote by $\overline{\mathfrak{U}}_0^D$ the completion of $\mathfrak{U}_0$ in this metric. Then D can be extended to a representation $\bar{D}$ for the larger space $\overline{\mathfrak{U}}_0^D$ by continuity.[14] Let $\mathfrak{U}_0 \subset \mathfrak{U}_1 \subset \overline{\mathfrak{U}}_0^D$ and let the bilinear form (f, g) be nondegenerate on $\mathfrak{U}_1$. We denote by $D_{\mathfrak{U}_1}$ the restriction to $\bar{D}$ to $\mathfrak{U}_1$. Then the basic result of this paper is that all $D_{\mathfrak{U}_1}$ obtained in this way (for all possible D) give a complete solution for the above "basis-independent" approach for *all* $\mathfrak{U}$ (with a suitable identification of test-function spaces). Of course, the proof in Sec. 4 goes just the other way around.

For general test-function spaces $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$, the situation is similar. $\mathfrak{U}_0$ is the smallest space one can start with since finite linear combinations are always allowed. The completions of $\mathfrak{U}_0$ in d_{U_0} and d_{V_0} yield the largest possible test-function spaces,[14] and their subspaces yield all $\mathfrak{U}_\Phi$ and $\mathfrak{U}_\Pi$. This carries the idea[5] that "every representation determines its own test-function space" to its final conclusion.

As an immediate application, we note a considerable simplification in the realization of the CCR's by means of a measure.[3,7] The above theorem implies that any representation of the CCR's in a separable Hilbert space can be realized as a direct integral by means of a measure on $\mathfrak{U}_0'$, the space of all infinite sequences of real numbers. The measure is quasi-invariant not only for $\mathfrak{U}_0 \subset \mathfrak{U}_0'$, but also for $\mathfrak{U}_\Pi$ embedded in $\mathfrak{U}_0'$. This means that the in general extremely large space $\mathfrak{U}_\Phi'$ can be replaced by the much smaller space $\mathfrak{U}_0'$.

It may be worthwhile to point out that the above orthonormal bases have little to do with L^2 convergence of test functions. This is most easily seen from the following example, which at first sight seems to be paradoxical and inconsistent with the above results. Consider a representation of the CCR's characterized by the expectation functional

$$E(f, g) = \exp\{-[(f,f) + (g,g) + i\tilde{f}(0) + i\tilde{g}(0)]\},$$
$$\tilde{f}(0) = \int f(\mathbf{x})\, d^3\mathbf{x}, \quad \tilde{g}(0) = \int g(\mathbf{x})\, d^3\mathbf{x}, \tag{5.1}$$

which is well defined for any f and $g \in L^2 \cap L^1$. It was noted by Araki (cf. the remark in Ref. 15) that this representation coincides with the Fock representation when f and g is restricted to the subspace $\{f \in L^2 \cap L^1, \tilde{f}(0) = 0\}$. Therefore, if one chooses an orthonormal basis $\{h_n\}$ where $\tilde{h}_n(0) = 0$ for each n, then $\tilde{\mathfrak{U}}_0 = \mathfrak{L}\{h_1, h_2, \cdots\}$ is norm dense in $L^2 \cap L^1$, and the extension of the representation yields the Fock representation, not the original one.

The solution to this apparent paradox is quite

simple. The natural metric is, in this case,

$$d_U(f)^2 = 2\pi^{\frac{1}{2}}(1 - [1 + (f,f)]^{-1} \times \exp\{-\tilde{f}(0)^2/[1 + (f,f)]\}) \quad (5.2)$$

and the same expression for $d_V(g)$. Omitting the term with $\{\tilde{f}(0)\}^2$, one obtains the metric for the Fock representation. It is evident that the above subspace $\mathfrak{V}_0$ is not dense in $L^2 \cap L^1$ with respect to the metric d_U since convergence in d_U means that both $(f_n - f, f_n - f)$ and $\{\tilde{f}_n(0) - \tilde{f}(0)\}$ converge to zero. Hence, in the metric d_U, one can never approximate a function f with $\tilde{f}(0) \neq 0$ by functions from $\mathfrak{V}_0$.

The above results have an immediate application to the question of analytic vectors and Gårding domains. In a recent paper Reed[16] has shown that the generators q_k and p_k, $k = 1, 2, \cdots$, of a basis-dependent representation possess a common dense set of analytic vectors as a common invariant dense Gårding domain, on which they are essentially self-adjoint. Certain infinite linear combinations of q_k and of the p_k can also be defined on this domain; they leave it invariant and are essentially self-adjoint on it. Hence one has, by Theorem 4.1, the following.

Corollary 5.1: Let $U(f)$, $V(g)$, with $f \in \mathfrak{V}_\Phi$ and $g \in \mathfrak{V}_\Pi$, be a representation in a separable Hilbert space $\mathcal{H}$. Then there are subspaces $\hat{\mathfrak{V}}_\Phi$ of $\mathfrak{V}_\Phi$ and $\hat{\mathfrak{V}}_\Pi$ of $\mathfrak{V}_\Pi$, dense in the natural metric and a domain D, dense in $\mathcal{H}$, such that all generators $\Phi(f)$, $f \in \hat{\mathfrak{V}}_\Phi$, and $\Pi(g)$, $g \in \hat{\mathfrak{V}}_\Pi$, are defined and essentially self-adjoint on D, leave D invariant, and possess each $\psi \in D$ as an analytic vector.

One obviously has, with the $\mathfrak{V}^0_\Phi$ and $\mathfrak{V}^0_\Pi$ of Theorem 4.1, the inclusions $\mathfrak{V}^0_\Phi \subset \hat{\mathfrak{V}}_\Phi \subset \mathfrak{V}_\Phi$ and $\mathfrak{V}^0_\Pi \subset \hat{\mathfrak{V}}_\Pi \subset \mathfrak{V}_\Pi$. It would be interesting if one could show that one can always choose $\hat{\mathfrak{V}}_\Phi = \mathfrak{V}_\Phi$ and $\hat{\mathfrak{V}}_\Pi = \mathfrak{V}_\Pi$.

ACKNOWLEDGMENT

The author would like to thank J. R. Klauder for stimulating discussions.

[1] J. Gårding and A. S. Wightman, Proc. Natl. Acad. Sci. (U.S.) **40**, 622 (1954).
[2] I. E. Segal, Trans. Am. Math. Soc. **88**, 12 (1958).
[3] H. Araki, J. Math. Phys. **1**, 492 (1960).
[4] I. S. Lew, thesis, Princeton University, 1960.
[5] G. C. Hegerfeldt and J. R. Klauder, Commun. Math. Phys. **16**, 329 (1970).
[6] E.g., take $\mathfrak{V} = \mathfrak{V}_\Phi = \mathfrak{V}_\Pi$ corresponding to *uncountably* many degrees of freedom and a product representation.
[7] G. C. Hegerfeldt and O. Melsheimer, Commun. Math. Phys. **12**, 304 (1969).
[8] G. C. Hegerfeldt, J. Math. Phys. **10**, 1681 (1969) (cf. Sec. 3).
[9] P. R. Halmos, *Measure Theory* (Van Nostrand, Princeton, N.J., 1950), p. 177.
[10] The metric d is defined only for countable sums of cyclic representations. For these the associated topology is the weakest vector topology τ for which the maps $f \to U(f)$ and $g \to V(g)$ are strongly continuous. If one allows uncountable sums of cyclic representations and replaces d by τ, then the statement of the corollary is, in general, not true. Take, for instance, an uncountable sum of the Fock representation. The associated τ is then just equivalent to the metric of the Fock representation, i.e., separable, but the associated Hilbert space is nonseparable.
[11] Note that automatically $(h_n, g_{k_i}) = 0$ for $i < n$.
[12] Conversely, if $U(\lambda f_n) \to U(\lambda f)$ strongly for each λ, then $f_n \to f(d)$.
[13] Similarly, as in Footnote 12, the converse also holds.
[14] It may happen that the bilinear form becomes degenerate on the completions. If one wants to retain the nondegeneracy, one may have to restrict oneself to subspaces.
[15] E. J. Woods, Commun. Math. Phys. **17**, 1 (1970).
[16] M. C. Reed, Commun. Math. Phys. **14**, 336 (1969).